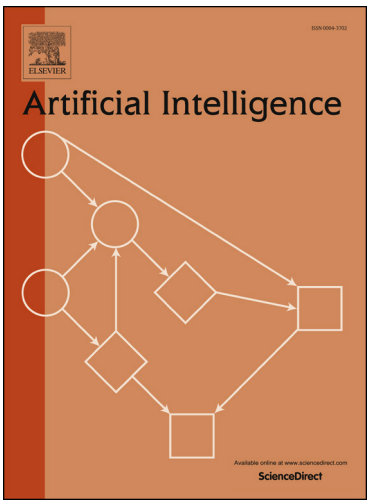
# Accepted Manuscript




Arguing about informant credibility in open multi-agent systems

Sebastian Gottifredi, Luciano H. Tamargo, Alejandro J. García, Guillermo R. Simari

PII: S0004-3702(18)30077-8
DOI: https://doi.org/10.1016/j.artint.2018.03.001
Reference: ARTINT 3059

To appear in: *Artificial Intelligence*

Received date: 22 December 2016
Revised date: 5 January 2018
Accepted date: 5 March 2018

Please cite this article in press as: S. Gottifredi et al., Arguing about informant credibility in open multi-agent systems, *Artif. Intell.* (2018), https://doi.org/10.1016/j.artint.2018.03.001

# Arguing about informant credibility in open multi-agent systems

*Email addresses:* sg@cs.uns.edu.ar (Sebastian Gottifredi), lt@cs.uns.edu.ar (Luciano H. Tamargo), ajg@cs.uns.edu.ar (Alejandro J. García), grs@cs.uns.edu.ar (Guillermo R. Simari)

# Arguing about informant credibility in open multi-agent systems

Sebastian Gottifredi[a], Luciano H. Tamargo[a], Alejandro J. García[a], Guillermo R. Simari[a]

[a]*Institute for Computer Science and Engineering (UNS–CONICET)*
*Department of Computer Science and Engineering, Universidad Nacional del Sur, Argentina*

**Abstract**

This paper proposes the use of an argumentation framework with recursive attacks to address a trust model in a collaborative open multi-agent system. Our approach is focused on scenarios where agents share information about the credibility (informational trust) they have assigned to their peers. We will represent informants' credibility through credibility objects which will include not only trust information but also the informant source. This leads to a recursive setting where the reliability of certain credibility information depends on the credibility of other pieces of information that should be subject to the same analysis. Credibility objects are maintained in a credibility base which can have information in conflict. In this scenario, we will formally show that our proposal will produce a partially ordered credibility relation; such relation contains the information that can be justified by an argumentation process.

*Keywords:* Argumentation, Multi-agent system, Trust, Credibility orders

## 1. Introduction

Open multi-agent systems have been characterized [20, 31] as multi-agent systems where the agents have different aims and goals and they can freely join and leave the system. In such system, it is possible to assume that agents display important characteristics: agents can be assumed to be self-interested and unreliable, agents lack a global perspective of the system, and there is no central control over the agents behavior that could facilitate the prediction of the interactions with other agents.

In this paper, we will consider a set of deliberative agents that participate in a collaborative open multi-agent system in which each agent plays the role of an *informant* for other agents, and can receive information from multiple sources,

*Email addresses:* sg@cs.uns.edu.ar (Sebastian Gottifredi), lt@cs.uns.edu.ar (Luciano H. Tamargo), ajg@cs.uns.edu.ar (Alejandro J. García), grs@cs.uns.edu.ar (Guillermo R. Simari)

*i.e.*, several agents could contribute information to another agent. It is clear that some form of trust model is needed when the adoption of a critical decision depends on the credibility (*informational trust*) assigned to the information received from other agents. In multi-agent systems, interaction is a crucial activity and, through interaction, agents share different kinds of information. In our approach, agents will share information about the credibility or informational trust they have assigned to their peers and then, through this interaction, the credibility assigned to their peers could change. Since the credibility information received may be conflicting, we propose an argumentative formalism for handling decisions about the credibility information that an agent (as presented in [9], the *trustor*) stores about their informants (as presented in [9], the *trustees*). Therefore, when an agent is faced with conflictive information, the proposed argumentation framework can be used to decide which information is more credible, and hence, which information prevails.

In multi-agent systems, representing the credibility associated with a piece of information and making the evaluation of this credibility possible are two important aspects when the agents have their own beliefs and can obtain new information from other sources [12]. Furthermore, as it was mentioned in [28], agents acting in open environments depend on reputation and trust mechanisms to evaluate the behavior of potential partners. Research in this area has increased considerably, and reputation and trust mechanisms have become key elements in the design of multi-agent systems [25, 33].

In [28] and [32], a set of relevant aspects to classify trust models is proposed. In this proposal, we will consider only two of those aspects: *information sources* and a *trust reliability measure*. In [32] the authors suggest that "...Sometimes, as important as the trust/reputation value itself is to know how reliable is that value and the relevance it deserves in the final decision making process...". In our approach, as in [36, 37], we introduce this type of information through the use of agent identifiers that represent the information sources, avoiding in that way the need of a separate data structure to maintain the measure of reliability. This leads to a recursive setting in which the reliability of certain credibility information depends on the credibility of other pieces of information that should be subject to the same analysis.

This research is motivated by two key factors. To begin with, the existence of large amounts of available data that intelligent agents and information system can access nowadays from different sources such as social networks, open data servers, and other similarly available origins. Additionally, the need of having a way to consider the reliability of that data. Indeed, many actual applications, such as booking services, buying and selling portals, renting locations pages or tourist attractions recommending services, etc. request and use their customers' evaluations to produce rankings of the products they handle. Moreover, nowadays there is also a tendency of having a reciprocal evaluation of the informants' reliability. The combination of these two factors produces a significant amount of information regarding the reliability of those informants, and these assessments are becoming available. Since that information could come from different sources (*e.g.*, social media, open data), it is clear that substantial inconsistency

in the obtained data could arise; therefore, there is a need for devising a way of deciding which information prevails. In this scenario, since the sources of that information are dynamic, we consider that the classical approaches of revising databases eliminating information for the sake of maintaining consistency would not be appropriated because that approach loses valuable information. In our proposal, we examine a scenario where the two factors mentioned above are present: we propose a formalization that allows to represent the evaluation of informant agents and to investigate the possibility of dealing with information from different sources that, when put together, the result could be contradictory.

Below, we will introduce an example in the stock market domain that considers the credibility that a particular agent called Tory assigns to its informants in this topic. Our aim of using this example is to provide a more intuitive introduction of the ideas in order to help the reader follow the presentation. Furthermore, it will be employed as a running example for the rest of the paper, though in Section 5 other applications domains will be discussed. Here, we will restrict our research to considering credibility information for a single topic; multi-topic or multi-context credibility orders are left as future work. Intuitively, the notion of credibility should be irreflexive and transitive.

**Example 1** (Running example)**.** *Let* Alex, Barbara, *and* Carla *(abbreviated* $A$*,* $B$*,* $C$*) be stock market experts, and let* Harry, John, *and* Kate *(abbreviated* $H$*,* $J$*,* $K$*) be journalists on this topic. Agent* Tory *(*$T$*) has all the agents mentioned above as its informants, and* $T$ *has collected the following information about their credibilities in the stock market topic. According to* Kate, Alex *is more credible than* Carla *(represented* $A > C$*); and according to* Harry, Carla *is more credible than* Barbara, $(C > B)$*. However, for* John, Barbara *is more credible than* Alex, $(B > A)$*. Finally, according to the editor of a newsletter called X-market, which specializes on this topic, the journalist* Kate *is more credible than* John, $(K > J)$*.*

*Now, assume that* Tory *has to decide whether Alex is more credible than* Barbara*. Tory can conclude* $B > A$ *using the information obtained from John but, assuming the credibility relation is transitive, Tory can also infer* $A > B$ *from the information obtained from* Kate $(A > C)$ *and* Harry $(C > B)$*. In this situation,* Tory *can use the information that she has about the credibility of the journalists* John, Harry *and* Kate *to decide whether* $A > B$ *or* $B > A$ *prevails. Since the editor of* X-market *reports that* $K > J$*, then* Tory *has a reason to consider* Kate *more credible than* John *on this topic and therefore, a reason to support that* Alex *is more credible than* Barbara*.*

The example above describes a situation where an agent receives information from different sources and shows how the agents' subjective attribution of credibility to a particular informant can be related to the credibility attributed to others. In our proposal, we will deal with information that an agent stores about the credibility of its informants, and also we will assume that the information received from an agent is as credible as the agent that supplies it.

In decision-making problems where information coming from multiple agents is involved it is possible to use the credibility of the informants to help in making

a decision. Similar to [1, 12, 34], we will favor the use of the word *credibility* to refer to this characteristic of informant agents as this particular word carries an intuitive sense that helps to understand the related problems. Here, the credibility of informant agents will be represented in a *credibility base* that will store the credibilities that an agent assigns to its informants and also its reliability.

In the literature, credibility has been represented using elements of possibility theory, see [6, 12], or adopting a symbolic approach, as in [8, 13, 18, 24, 35, 36, 37]. Some of these symbolic approaches, for instance [24, 35], propose to use total orders whereas others [36, 37] adopt partial orders providing the capability of representing cases where the credibility of two agents cannot be compared because of different reasons, *e.g.*, the relation between them has not been established. In particular in [36], following belief revision techniques, the authors formalized change operators over a credibility base to maintain it as non-contradictory. Although those change operators have been designed using the principle of informational economy (*i.e.*, when accepting a new piece of information the agent should aim at a minimal change of its previously held beliefs), in general, when maintaining consistency some information is lost. We will follow a different tack in this work: our proposal gives the capability of adding multiple beliefs simultaneously (*e.g.*, to join two or more bases) without having to eliminate the possible contradictions that may arise.

Since obtaining information is a very expensive process, in our approach an agent will keep all the information it receives thus allowing the possibility of having an inconsistent credibility base. In this manner, in dynamic scenarios with many interactions where the trust model uses *witness information* [28, 32], our proposal will cover cases in which beliefs that in the current situation are less important, but may become so in the future, are not lost. It is also important to note that, since we keep all the information, the order in which new information is added to the credibility base will not affect the result as occurs in [36]. In this paper, we will propose how to equip agents with a mechanism that will allow them to coherently infer information from their credibility base despite possible inconsistencies. Therefore, the main contribution of our proposal is to define an argumentation formalism which will provide the capability to coherently infer strict partial orders from such bases. An argument in this formalism will provide a reason to support the fact that an informant is more credible than other. As arguments can be challenged, we will introduce different types of attacks. Given the recursive nature of the attacks in our system, in order to define which arguments will be accepted, and thus which conclusions will be obtained, we will use the Argumentation Framework with Recursive Attacks (AFRA) introduced in [3, 4], which allows attacks to the attack relation. We will use acceptability semantics (in particular preferred extensions) to determine which are the accepted arguments in our system, and then, we will formally show that the conclusions of an extension from a credibility base constitute a partial order with respect to the credibility relation.

The rest of this paper is structured as follows. In Section 2, we will introduce how to represent the credibility of informants. In Section 3, an argumentation

framework with three different type of attacks is proposed. Then, in Section 4 we will define how to obtain credibility partial orders from potentially inconsistent credibility bases using our argumentation approach. In Section 5 we discuss how our approach can be applied to some real world domains. Section 6 analyzes related work, and finally, in Section 7 conclusions are offered and ideas for future work are given. All the proofs for the results will be included in the corresponding sections.

## 2. Representing Informant's Credibility

In this proposal, we consider a finite set $\mathbb{A}$ of identifiers for naming informant agents and each agent will be labeled with a unique name from that set. Agents' identifiers will be denoted with uppercase italic letters that can have subscripts and each identifier will represent a unique agent; thus $\mathbb{A} = \{A, B, \ldots, Z, A_1, \ldots, Z_n\}$. Since our approach assumes a collaborative environment, it is unnecessary to make provisions to prevent identity theft. We also assume that each agent already has a repository of the informational trust of other agents, *i.e.*, we do not require any particular bootstrapping mechanism. Following [36], the credibilities an agent assigns to its informant agents is stored in a *Credibility Base* as defined next.

**Definition 1** (Credibility object - Credibility Base)**.** *Let $\mathbb{A}$ be a set of agent identifiers and $P, Q, S \in \mathbb{A}$. A credibility object is a pair $[P > Q, S]$ which represents that agent $P$ is strictly more credible than $Q$, and the agent $S$ is the information source of the credibility element $P > Q$. A credibility base is a finite set $\mathbb{C}$ of credibility objects.*

A credibility base stores credibility objects that are pairs of credibility elements together with their associated information source which will represent the reliability of the associated credibility element. Observe that, as in [36], the credibility base itself stores the information for considering the reliability of each element. Thus, there is no need for a separate data structure maintaining the measure of reliability of credibility elements. We include below a simple example that will be used for introducing the main concepts of our approach.

**Example 2.** *Consider the credibility base:*

$$\mathbb{C}_2 = \{[A > C, K], [C > B, H], [B > A, J], [K > J, X]\}.$$

*This set represents the information introduced above for our running example: the agent $K$ has informed that $A$ is more credible than $C$ ($A > C$), the agent $H$ has informed that $C > B$, $J$ has informed that $B > A$, and $X$ (the editor of the newsletter) has informed that $K > J$.*

Consider the credibility base $\mathbb{C}_2$ from Example 2 where the following two credibility elements $A > C$ and $C > B$ are explicitly stored. It is clear that from $\mathbb{C}_2$, $A > B$ can be inferred, which is in contradiction with the credibility element

$B > A$ informed by $J$ that is explicitly stored in $\mathbb{C}$. Next, we will introduce three functions: $\mathsf{ce}(\mathbb{C})$, $\mathsf{cl}(\mathbb{C})$, and $\mathsf{rel}(\mathbb{C})$ that will be used in the formalism proposed below. The first function returns all the credibility elements of a credibility base, *i.e.*, $\mathsf{ce}(\mathbb{C}) = \{P > Q : [P > Q, S] \in \mathbb{C}\}$. Note that $\mathsf{ce}(\mathbb{C})$ is a finite set that corresponds to the projection of $\mathbb{C}$ with respect to those credibility elements that are explicitly represented in $\mathbb{C}$. Then, $\mathsf{cl}(\mathbb{C})$ will include all credibility elements that can be inferred from $\mathsf{ce}(\mathbb{C})$ by the transitive closure of $\mathsf{ce}(\mathbb{C})$. Considering the credibility base from Example 2, $\mathsf{ce}(\mathbb{C}_2) = \{A > C, C > B, B > A, K > J\}$ and $\mathsf{cl}(\mathbb{C}_2) = \{A > C, C > B, B > A, A > B, C > A, B > C, K > J, A > A, B > B, C > C\}$.

In a credibility object $[A > C, K]$, $K$ is the source of $A > C$, and $K$ represents the reliability of $A > C$. Given a credibility base $\mathbb{C}$, the function $\mathsf{rel}(\mathbb{C})$ returns the set of agent identifiers that are sources of credibility elements in $\mathbb{C}$, *i.e.*, $\mathsf{rel}(\mathbb{C}) = \{S : [P > Q, S] \in \mathbb{C}\}$. This function will be used to determine the set of agent identifiers that represents the reliability of an argument. As we will show below, an argument will be composed of a set of credibility objects, and to obtain its reliability the sources of those objects will be considered.

Observe that given a credibility base $\mathbb{C}$, $\mathsf{cl}(\mathbb{C})$ can have contradictory credibility elements, *e.g.*, $A > B$ and $B > A$ both belong to $\mathsf{cl}(\mathbb{C}_2)$. Also note that both elements were obtained from different sources and, for that reason, their reliability could be different. Therefore, in the following sections we will propose an argumentation formalism that given a credibility base it will find a non-contradictory subset of $\mathsf{cl}(\mathbb{C})$ that has the information that is more reliable. We will also show below that this subset is a strict partial order of credibility elements.

## 3. Arguments and attacks

Each credibility element that can be inferred from a credibility base $\mathbb{C}$ will have an argument associated that supports it, and such argument will contain the reliability information for that inference. This reliability information will be used for comparing arguments supporting contradictory conclusions. An argument is a minimal set of credibility objects as defined next.

**Definition 2** (Argument)**.** *Let $\mathbb{C}$ be a credibility base, $\mathbb{A}$ be a set of agents, and $B, C \in \mathbb{A}$ where $B \neq C$, an argument $\mathcal{X}$ for the conclusion $B > C$ from $\mathbb{C}$ is a set of credibility objects $\mathcal{X} \subseteq \mathbb{C}$, such that $B > C \in \mathsf{cl}(\mathcal{X})$ and there is no $\mathcal{X}' \subset \mathcal{X}$ holding that $B > C \in \mathsf{cl}(\mathcal{X}')$. The set of all arguments that can be built from $\mathbb{C}$ will be denoted as $\mathsf{Args}_{\mathbb{C}}$. If $\mathcal{X}$ is an argument for the conclusion $B > C$, sometimes we will use the notation $\langle \mathcal{X}, B > C \rangle$.*

Observe that it is not possible to have arguments for a conclusion such as $A > A$. This is concordant with the intuition that credibility is an irreflexive relation since agents cannot be more credible than themselves. An argument is a minimal set with respect to set inclusion and, therefore, an argument cannot have a cycle. Note that even though an argument $\mathcal{X}$ can have one or more

credibility objects, each argument has only one conclusion. Any proper subset of an argument is itself an argument for a different conclusion, and represents an intermediate reasoning step in the argument in which it is contained and, trivially, an argument is a subargument of itself. This notion is captured in the subargument relation defined next.

**Definition 3** (Subargument)**.** *Let $\mathbb{C}$ be a credibility base and $\mathcal{X}, \mathcal{Z} \in \mathsf{Args}_{\mathbb{C}}$, then $\mathcal{Z}$ is a subargument of $\mathcal{X}$ if and only if $\mathcal{Z} \subseteq \mathcal{X}$.*

The following proposition shows that the argument construction is monotonic. That is, the addition of credibility objects to a credibility base does not preclude the construction of an argument.

**Proposition 1.** *Let $\langle \mathcal{X}, A > B \rangle$ be an argument that can be built from the credibility base $\mathbb{C}$ and $\mathbb{C} \subseteq \mathbb{C}'$ then $\langle \mathcal{X}, A > B \rangle$ can also be built from $\mathbb{C}'$.*

*Proof.* Straightforward from Definition 2.

**Example 3.** *Consider again the credibility base of our running example introduced in Example 2:*

$$\mathbb{C}_2 = \{[A > C, K], [C > B, H], [B > A, J], [K > J, X]\}.$$

*There are seven arguments that can be built from $\mathbb{C}_2$. For the sake of clarity, in this example we will present three of them, nevertheless the remaining arguments will be presented when the full example is studied in greater depth in Section 4. For instance, from $\mathbb{C}_2$ it is possible to build the argument $\mathcal{A}_2 = \{[A > C, K], [C > B, H]\}$ for the conclusion $A > B$, the argument $\mathcal{A}_1 = \{[A > C, K], [B > A, J]\}$ for the conclusion $B > C$, and $\mathcal{A}_3 = \{[B > A, J]\}$ for the conclusion $B > A$. Note that $\mathcal{A}_3$ is a subargument of $\mathcal{A}_1$. On the other hand, there are several subsets of $\mathbb{C}_2$ that are not arguments. For instance, the set $\mathcal{N} = \{[A > C, K], [C > B, H], [B > A, J]\}$ is not an argument because for any potential conclusion in $\mathsf{cl}(\mathcal{N})$ the set $\mathcal{N}$ is not minimal. A similar situation occurs with the set $\{[B > A, J], [K > J, X]\}$ and with the set $\{[A > C, K], [K > J, X]\}$*

Figure 1 depicts the three arguments introduced in Example 3. There, each argument is represented as a table of one column with its name and conclusion at the top, and below them, the set of credibility objects used for obtaining the conclusion. Given an argument $\mathcal{X}$, the set $\mathsf{rel}(\mathcal{X})$ contains all the agent identifiers which are sources of the credibility objects in $\mathcal{X}$, and hence $\mathsf{rel}(\mathcal{X})$ represents the *reliability* of $\mathcal{X}$. Considering the arguments from Example 3, we have $\mathsf{rel}(\mathcal{A}_1) = \{K, J\}$, $\mathsf{rel}(\mathcal{A}_2) = \{K, H\}$, and $\mathsf{rel}(\mathcal{A}_3) = \{J\}$.

Observe that in Example 3 the argument $\mathcal{A}_2$ provides a reason to conclude that $A$ is more credible than $B$, whereas $\mathcal{A}_3$ concludes the contrary, *i.e.*, $B$ is more credible than $A$. We next introduce the definition of arguments with contradictory conclusions that will be used in our formalization below.

**Definition 4** (Contradictory conclusions)**.** *Let $\mathcal{X}$,$\mathcal{Z}$ be two arguments in $\mathsf{Args}_{\mathbb{C}}$. If $\mathcal{X}$ concludes $B > C$ and $\mathcal{Z}$ concludes $C > B$, for some $B, C \in \mathbb{A}$, then $\mathcal{X}$ and $\mathcal{Z}$ have contradictory conclusions.*

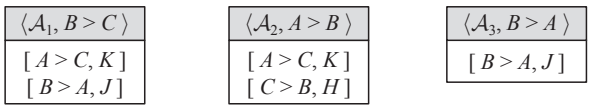



Figure 1: Arguments from Example 3

Clearly, arguments with contradictory conclusions are in conflict. As we will discuss in detail below, we will study three types of conflicts that will lead to three different attack relations. The first type, called *trust-attacks*, captures the type of conflicts related to contradictory conclusions. The second type of attacks, called *reliability-attacks*, captures how an argument can challenge the reliability of an argument that trust-attacks another. The third type, called *indirect-reliability-attacks*, captures the indirect conflict that is implicit between two reliability-attacks that challenge opposing positions.

These three attack relations together will characterize an AFRA's recursive attack relation, and that framework will be used to determine the arguments that are finally accepted. For convenience, given a credibility base $\mathbb{C}$ we will denote with $\mathsf{TAtts}_{\mathbb{C}}$ the set of trust-attacks from $\mathbb{C}$, with $\mathsf{RAtts}_{\mathbb{C}}$ the set of reliability-attacks from $\mathbb{C}$, and with $\mathsf{IAtts}_{\mathbb{C}}$ the set of indirect-reliability-attacks from $\mathbb{C}$. Below we will explain the intuitions that motivate each type of attack and present their definitions.

### *3.1. Trust-Attacks*

A *trust-attack* captures the conflict between arguments that conclude contradictory credibilities. Intuitively, argument $\mathcal{Y}$ trust-attacks another $\mathcal{X}$ if either they have contradictory conclusions, or there is a subargument of $\mathcal{X}$ having a contradictory conclusion with $\mathcal{Y}$; that is, either $\mathcal{Y}$ contradicts the credibility concluded by $\mathcal{X}$ or $\mathcal{Y}$ contradicts an intermediate reasoning step of $\mathcal{X}$. Next, we formalize this notion.

**Definition 5** (Trust-Attack)**.** *Let $\mathcal{X}, \mathcal{Y}$ be two arguments in $\mathsf{Args}_{\mathbb{C}}$, we will say that $(\mathcal{Y}, \mathcal{X}) \in \mathsf{TAtts}_{\mathbb{C}}$ if there exists a subargument $\mathcal{Z}$ of $\mathcal{X}$ such that $\mathcal{Y}$ and $\mathcal{Z}$ have contradictory conclusions. This subargument $\mathcal{Z}$ is referred as the disagreement subargument of $(\mathcal{Y}, \mathcal{X})$ and denoted $\mathsf{dis}((\mathcal{Y}, \mathcal{X})) = \mathcal{Z}$.*

Recall that every argument is trivially a subargument of itself. Thus, an argument $\mathcal{A}$ will be attacked by (and will also attack) any other argument $\mathcal{B}$ such that $\mathcal{A}$ and $\mathcal{B}$ have contradictory conclusions. Therefore, intuitively an argument can trust-attack other argument at its conclusion or at an intermediate point (subargument). Figure 2 shows the three arguments introduced in Example 3 and the trust-attacks between them; these attacks are depicted with solid arrows. Each arrow is labeled with a lowercase Greek letter, *e.g.*, there exists a trust-attack $\alpha = (\mathcal{A}_2, \mathcal{A}_3)$ because $\mathcal{A}_2$ and $\mathcal{A}_3$ have contradictory conclusions. An analogous situation holds for $\beta = (\mathcal{A}_3, \mathcal{A}_2)$. The trust-attack $\lambda = (\mathcal{A}_2, \mathcal{A}_1)$ occurs because $\mathcal{A}_1$ has a subargument concluding $B > A$ and $\mathcal{A}_2$

concludes $A > B$. Finally, the trust-attack $\pi = (\mathcal{A}_1, \mathcal{A}_2)$ occurs because $\mathcal{A}_1$ and a subargument of $\mathcal{A}_2$ have contradictory conclusions.

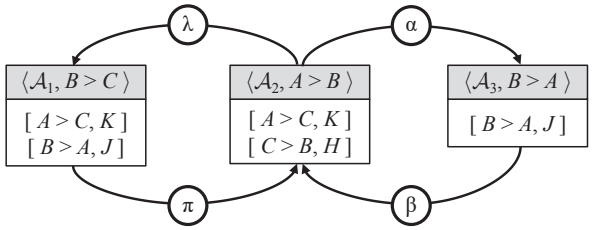



Figure 2: Trust-attacks between arguments from Example 3

In the case shown in Figure 2, where only three arguments are considered, a kind of blocking situation occurs because all the arguments are trust-attacked by another argument. Nevertheless, in the credibility base $\mathbb{C}_2$ there is more information that can be used. In the next subsection, we will show how an attack can be attacked in order to decide which information prevails; but first, we will show some results that are related to the previous definitions and that characterize our framework.

The following result shows that any argument built from a credibility base has internal coherence. That is, an argument should be consistent by itself, and thus it will not attack itself.

**Proposition 2.** *There is no* $\mathcal{X} \in \mathit{Args}_{\mathbb{C}}$ *such that* $(\mathcal{X}, \mathcal{X}) \in \mathit{TAtts}_{\mathbb{C}}$.

*Proof.* Suppose that $(\mathcal{X}, \mathcal{X}) \in \mathsf{TAtts}_{\mathbb{C}}$, then there is an argument $\mathcal{Z}$ which is a subargument of $\mathcal{X}$ such that $\mathcal{X}$ and $\mathcal{Z}$ have contradictory conclusions. That is, if $\mathcal{X}$ concludes $B > C$ then $\mathcal{Z}$ concludes $C > B$. Then, by Definition 2 it holds that $\{B > C, C > B\} \subseteq \mathsf{cl}(\mathcal{X})$. Therefore, there exists $\mathcal{X}' \subset \mathcal{X}$ such that $B > C \in \mathsf{cl}(\mathcal{X}')$, contradiction.

The following proposition shows that trust-attacks happen in pairs. That is, if an argument trust-attacks another argument, the attacking argument will, in turn, also be trust-attacked.

**Proposition 3.** *Let* $\mathcal{X}, \mathcal{Y} \in \mathit{Args}_{\mathbb{C}}$. *If* $(\mathcal{Y}, \mathcal{X}) \in \mathit{TAtts}_{\mathbb{C}}$ *then there is* $\mathcal{Z} \in \mathit{Args}_{\mathbb{C}}$ *such that* $(\mathcal{Z}, \mathcal{Y}) \in \mathit{TAtts}_{\mathbb{C}}$.

*Proof.* If $(\mathcal{Y}, \mathcal{X}) \in \mathsf{TAtts}_{\mathbb{C}}$ then by Definition 5 there is a subargument $\mathcal{Z}$ of $\mathcal{X}$ such that $\mathcal{Y}$ and $\mathcal{Z}$ have contradictory conclusions. Then, given that any argument is trivially a subargument of itself, by Definition 5 it holds that $(\mathcal{Z}, \mathcal{Y}) \in \mathsf{TAtts}_{\mathbb{C}}$.

The following corollary establishes the notion of symmetry of trust-attacks between arguments with contradictory conclusions.

**Corollary 1.** *Let $\alpha = (\mathcal{Y}, \mathcal{X}) \in \mathsf{TAtts}_{\mathbb{C}}$. If $\mathsf{dis}(\alpha) = \mathcal{X}$ then $(\mathcal{X}, \mathcal{Y}) \in \mathsf{TAtts}_{\mathbb{C}}$.*

*Proof.* Straightforward from Definition 5 and Proposition 3.

*3.2. Reliability-Attacks*

We introduce now the possibility of representing an attack to an attack called *reliability-attack*. Below we will explain how with this form of attack the reliability of the information can be used to decide which argument prevails when arguments for contradictory conclusions are obtained.

Recall that given an argument $\mathcal{X}$ the set $\mathsf{rel}(\mathcal{X})$ represents the *reliability* of $\mathcal{X}$ and contains all the agent identifiers that are sources of the credibility objects in $\mathcal{X}$. For instance, as shown for the arguments from Example 3, $\mathsf{rel}(\mathcal{A}_1) = \{K, J\}$, $\mathsf{rel}(\mathcal{A}_2) = \{K, H\}$ and $\mathsf{rel}(\mathcal{A}_3) = \{J\}$.

Consider again the credibility base $\mathbb{C}_2$ and the Figure 2 where trusts-attacks $\alpha = (\mathcal{A}_2, \mathcal{A}_3)$ and $\beta = (\mathcal{A}_3, \mathcal{A}_2)$ are depicted. From $\mathbb{C}_2$ the argument $\mathcal{A}_4 = \{[K > J, X]\}$ concluding $K > J$ can be built. Observe that the argument $\mathcal{A}_2$ has $K$ as one of the sources of its information ($K \in \mathsf{rel}(\mathcal{A}_2)$), and $\mathcal{A}_3$ has $J$ as the source of its information ($J \in \mathsf{rel}(\mathcal{A}_3)$). In this scenario, we can use the argument $\mathcal{A}_4$ to establish a preference of $\mathcal{A}_2$ over $\mathcal{A}_3$, given that $\mathcal{A}_2$ has sources that can be deemed as more credible than some of the sources in $\mathcal{A}_3$. That is, $\mathcal{A}_4$ establishes that $\mathcal{A}_2$ is based on more reliable information than $\mathcal{A}_3$. Following this line of reasoning, argument $\mathcal{A}_4$ challenges the validity of the trust-attack $\beta$ (see Figure 3 where this challenge is depicted with a dashed line).

This form of conflict will be captured by an attack to an attack that will be introduced below in Definition 6; but first, we will introduce two auxiliary notions: the *origin* of an attack $\alpha$, denoted $\mathsf{og}(\alpha)$, and the *target* of an attack, denoted $\mathsf{tg}(\alpha)$. Given an attack $\alpha = (\mathcal{A}, T)$, the first element of the pair is the origin of the attack, *i.e.*, $\mathsf{og}(\alpha)$=$\mathcal{A}$, while the second element of the pair is the target of the attack, *i.e.*, $\mathsf{tg}(\alpha)$=$T$. Since in our framework there can be attacks to arguments and attacks to attacks, the target $T$ could be either an argument or an attack.

Reliability-attacks, as we have discussed above, capture the preference of some arguments considering how reliable they are in light of their sources. This notion will formalize the intuition of challenging attacks.

Next, we will formally introduce the basic form of reliability-attacks, that will challenge trust-attacks in $\mathsf{TAtts}_{\mathbb{C}}$ (as was exemplified above). Once we have introduced the indirect-reliability-attacks, we will present the recursive version of reliability-attacks capable of targeting such form of attack.

**Definition 6** (Basic Reliability-attack)**.** *Let $\mathcal{X}$ be an argument in $\mathsf{Args}_{\mathbb{C}}$ concluding $C > B$ and $\beta \in \mathsf{TAtts}_{\mathbb{C}}$ such that no agent in the conclusion of $\mathsf{og}(\beta)$ appears in $\mathsf{rel}(\mathcal{X})$. We say that $(\mathcal{X}, \beta) \in \mathsf{RAtts}_{\mathbb{C}}$ if and only if $B \in \mathsf{rel}(\mathsf{og}(\beta))$ and $C \in \mathsf{rel}(\mathsf{dis}(\beta))$.*

**Example 4.** *Consider the credibility base of our running example:*

$$\mathbb{C}_2 = \{[A > C, K], [C > B, H], [B > A, J], [K > J, X]\}.$$

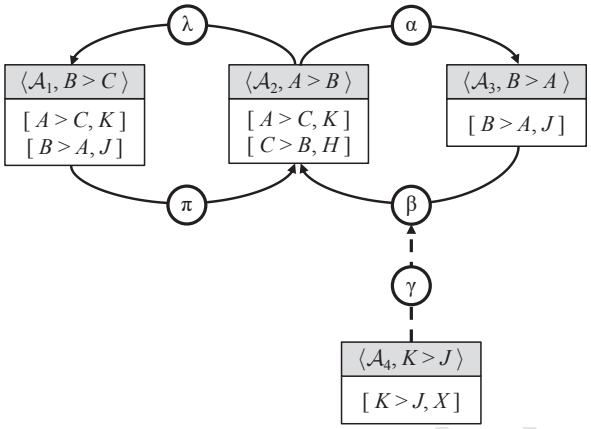



Figure 3: Reliability-attack described in Example 4

*In Example 3, we have shown how from $\mathbb{C}_2$ it is possible to build the arguments $\mathcal{A}_2, \mathcal{A}_3$, and $\mathcal{A}_1$, and Figure 2 depicts a situation where the trust-attacks $\alpha$, $\beta$, $\gamma$, and $\pi$ appear. Note that the argument $\mathcal{A}_4 = \{[K > J, X]\}$ can also be built from $\mathbb{C}_2$. Figure 3 shows that $\mathcal{A}_4$ effects a reliability-attack to $\beta$, denoted $\gamma = (\mathcal{A}_4, \beta)$, since we have that $\mathsf{og}(\beta) = \mathcal{A}_3$, neither $A$ nor $B$ are in $\mathsf{rel}(\mathcal{A}_4) = \{X\}$, $J \in \mathsf{rel}(\mathcal{A}_3)$, the disagreement subargument of $\beta$ is $\mathsf{dis}(\beta) = \mathcal{A}_2$, $K \in \mathsf{rel}(\mathcal{A}_2)$, and $K > J$ is the conclusion of $\mathcal{A}_4$.*

For convenience, a reliability-attack will be depicted using dashed arrows (*e.g.*, $\gamma$ in Figure 3), and if we consider only the arguments that can be obtained from $\mathbb{C}_2$, there is no other reliability-attack. In particular, it is worth to mention that $\mathcal{A}_4$ does not produce a reliability-attack to $\pi$. To understand why, observe that $J \in \mathsf{rel}(\mathsf{og}(\pi))$ but the disagreement subargument for $\pi$ is $\mathsf{dis}(\pi) = \{[C > B, H]\}$ and $K \notin \mathsf{rel}(\mathsf{dis}(\pi))$.

Another important remark concerning Definition 6 is that it forbids circular defenses using reliability-attacks. Observe that the argument producing a reliability-attack, should not contain as part of its reliability (*i.e.*, its sources) any of those agents involved in the trust conflict which it aims to settle. In other words, an argument will not be able to challenge the reliability of a trust-attack if some of its sources are the ones that cause such trust-attack.

Although in Figure 3 the reliability-attack $\gamma$ breaks the blocking situation between arguments $\mathcal{A}_2$ and $\mathcal{A}_3$, there are more arguments and attacks in this scenario and more elements need to be considered to decide which arguments prevail. The complete analysis of the situation will be shown in Section 4.

The following proposition shows that if an argument challenges a trust-attack which models the conflict between two arguments, it will also challenge every other trust-attack that involves any superargument of such arguments.

**Proposition 4.** *Let $\mathbb{C}$ be a credibility base, $\alpha = (\mathcal{W}, \mathcal{V}) \in \mathsf{TAtts}_{\mathbb{C}}$ and $(\mathcal{X}, \alpha) \in$*

*$\mathsf{RAtts}_{\mathbb{C}}$. For every trust-attack $\beta = (\mathcal{W}, \mathcal{Y}) \in \mathsf{TAtts}_{\mathbb{C}}$ where the argument $\mathcal{V}$ is a subargument of $\mathcal{Y}$ it holds that $(\mathcal{X}, \beta) \in \mathsf{RAtts}_{\mathbb{C}}$.*

*Proof.* Since $\gamma = (\mathcal{X}, \alpha) \in \mathsf{RAtts}_{\mathbb{C}}$, by Definition 6 it holds that if $\mathcal{X}$ concludes $A > B$ then $B \in \mathsf{rel}(\mathsf{og}(\alpha))$ and $A \in \mathsf{rel}(\mathsf{dis}(\alpha))$. Also, by Definition 5 it holds $\mathsf{dis}(\alpha) = \mathsf{dis}(\beta)$, then $A \in \mathsf{rel}(\mathsf{dis}(\beta))$. Therefore, by Definition 6 it holds that $(\mathcal{X}, \beta) \in \mathsf{RAtts}_{\mathbb{C}}$.

### *3.3. Indirect-Reliability-Attacks*

To motivate this new form of attack, we will extend our running example with more information objects that will introduce more arguments (see Example 5).

**Example 5.** *Consider the credibility base*

$$\mathbb{C}_5 = \{[A>C, K], [C>B, H], [B>A, J], [K>J, X], [J>H, Y]\}$$

*which is in fact $\mathbb{C}_2 \cup \{[J>H, Y]\}$. The new credibility object represents that the newsletter Y-invest (abbreviated $Y$) has informed that agent $J$ is more credible than $H$. Note that from $\mathbb{C}_5$ we have the same arguments and attacks shown in Example 4 and, in addition, the argument $\mathcal{A}_5 = \{[J > H, Y]\}$ can be built. This argument produces a reliability-attack $\delta = (\mathcal{A}_5, \alpha)$ because $\mathsf{og}(\alpha) = \mathcal{A}_2$, $H \in \mathsf{rel}(\mathcal{A}_2)$, the disagreement subargument of $\alpha$ is $\mathsf{dis}(\alpha) = \mathcal{A}_3$, $J \in \mathsf{rel}(\mathcal{A}_3)$, and $\mathcal{A}_5$ has $J > H$ as conclusion. Furthermore, note that $\mathcal{A}_5$ also produces a reliability-attack $\theta$ to $\lambda$, since $H \in \mathsf{rel}(\mathsf{og}(\lambda))$ and $J \in \mathsf{rel}(\mathsf{dis}(\lambda))$. The situation is shown in Figure 4.*

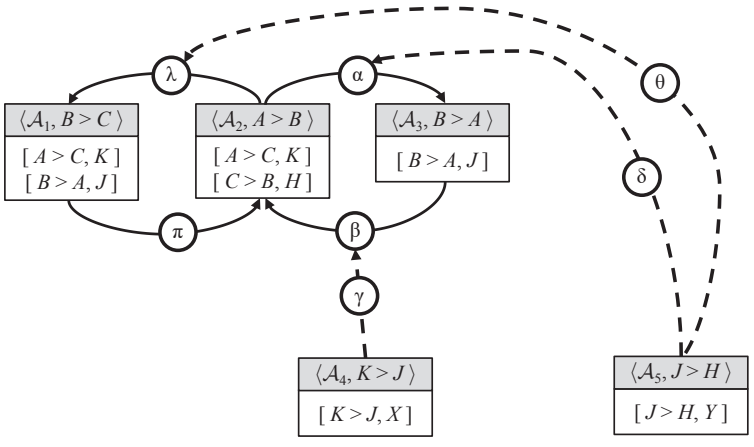



Figure 4: Reliability-attacks between arguments from Example 5

Indirect-reliability-attacks will be used to capture an implicit situation of conflict that can occur between two reliability-attacks. Observe in Figure 4

that the reliability-attack $\gamma$ challenges the trust-attack $\beta$ and, simultaneously, the reliability-attack $\delta$ challenges the trust-attack $\alpha$. Therefore, considering $\delta$ and $\gamma$ it could be seen that $\gamma$ is representing a preference of $\mathcal{A}_2$ over $\mathcal{A}_3$, meanwhile $\delta$ represents the contrary, and thus this means an implicit conflict between reliability-attacks such as $\delta$ and $\gamma$ exists. Also note that, if we suppose that the reliability-attacks $\delta$ and $\gamma$ are successful, then $\alpha$ and $\beta$ will be ineffective. This leads to an undesirable situation where arguments with contradictory conclusions, such as $\mathcal{A}_2$ and $\mathcal{A}_3$, could hold together.

To capture a conflict as the one between $\delta$ and $\gamma$ discussed above, we will introduce *indirect-reliability-attacks* which originate from an argument and target a reliability-attack. The argument producing the indirect-reliability-attack will be the same that produces the reliability-attack which is in conflict with the reliability-attack that is the target of the indirect-reliability-attack. As we did for the basic reliability-attack, we will present the basic indirect-reliability-attacks that target basic reliability-attacks next.

**Definition 7** (Basic Indirect-Reliability-attack)**.** *Let $\alpha, \beta \in \mathsf{RAtts}_{\mathbb{C}}$ where $\mathsf{og}(\alpha) = \mathcal{X}$ and $\mathsf{tg}(\alpha), \mathsf{tg}(\beta) \in \mathsf{TAtts}_{\mathbb{C}}$. We say that $(\mathcal{X}, \beta) \in \mathsf{IAtts}_{\mathbb{C}}$ if and only if either $\mathsf{og}(\mathsf{tg}(\alpha)) = \mathsf{dis}(\mathsf{tg}(\beta))$ or $\mathsf{og}(\mathsf{tg}(\alpha)) = \mathsf{tg}(\mathsf{tg}(\beta))$*

**Example 6.** *Consider the credibility base $\mathbb{C}_5$ from Example 5. In this scenario, given that $\gamma$ is a reliability-attack and $\mathsf{og}(\gamma) = \mathcal{A}_4$, we have that $\mathcal{A}_4$ produces the indirect-reliability-attack $\phi$ to $\delta$. This is so because $\mathsf{tg}(\gamma) = \beta$, $\mathsf{og}(\beta) = \mathcal{A}_3$ and because $\mathsf{tg}(\delta) = \alpha \in \mathsf{TAtts}_{\mathbb{C}_5}$, $\mathsf{tg}(\alpha) = \mathcal{A}_3$. An analogous situation holds for the indirect-reliability-attack $\varepsilon$: $\mathsf{tg}(\delta) = \alpha$, $\mathsf{og}(\alpha) = \mathcal{A}_2$ and because $\mathsf{tg}(\gamma) = \beta \in \mathsf{TAtts}_{\mathbb{C}_5}$, $\mathsf{tg}(\beta) = \mathcal{A}_2$. These indirect-reliability-attacks, together with previously described items, are depicted using dotted arrows in Figure 5. In addition, note that there is no indirect-attack to $\theta$.*

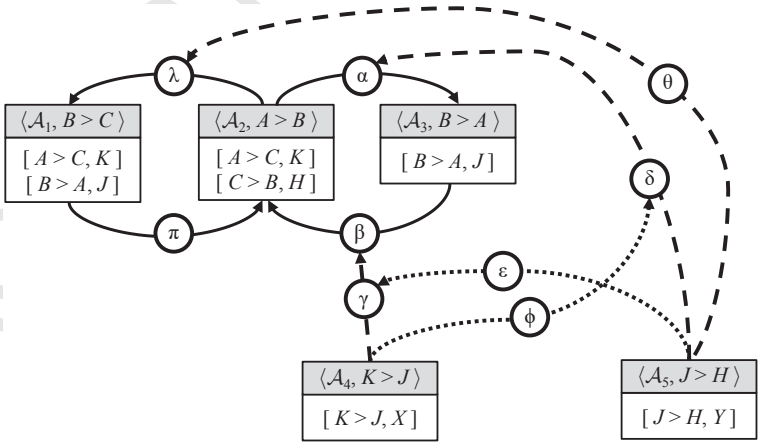



Figure 5: Indirect-reliability-attacks described from Example 6

As we have stated above Definition 6, reliability-attacks can also challenge indirect-reliability-attacks, and the intuition is similar to how these attacks challenged trust-attacks: a reliability-attack poses a preference over the arguments involved in a conflict. Following this line of reasoning, indirect-reliability-attacks should be capable of targeting these reliability-attacks, and in turn such indirect-reliability-attacks can also be challenged by reliability-attacks. Clearly, this naturally leads to a mutually recursive definition where both types of attacks can attack each other.

**Definition 8** (Reliability-attack / Indirect-Reliability-attack)**.** *Let $\mathcal{X}$ be an argument in $\mathsf{Args}_{\mathbb{C}}$ with conclusion $C > B$. It will hold that:*

- $(\mathcal{X}, \beta) \in \mathsf{RAtts}_{\mathbb{C}}$ *iff* $\beta \in \mathsf{IAtts}_{\mathbb{C}}$ *is such that no agent in the conclusion of* $\mathsf{og}(\beta)$ *is in* $\mathsf{rel}(\mathcal{X})$*,* $B \in \mathsf{rel}(\mathsf{og}(\beta))$*, and* $C \in \mathsf{rel}(\mathsf{og}(\mathsf{tg}(\beta)))$*.*
- $(\mathcal{X}, \beta) \in \mathsf{IAtts}_{\mathbb{C}}$ *iff* $\beta \in \mathsf{RAtts}_{\mathbb{C}}$ *and there is* $\alpha \in \mathsf{RAtts}_{\mathbb{C}}$ *such that* $\mathsf{og}(\alpha) = \mathcal{X}$ *and* $\mathsf{og}(\mathsf{tg}(\alpha)) = \mathsf{og}(\mathsf{tg}(\mathsf{tg}(\beta)))$*.*

We extend in Example 7 our running example to show a scenario where there exists a reliability-attack targeting an indirect-reliability-attack (which corresponds to the second condition from Definition 8).

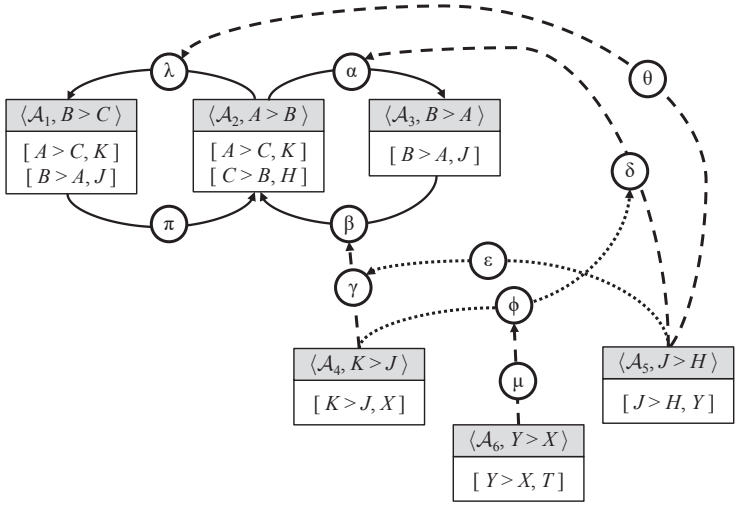



Figure 6: Indirect-reliability-attacks described in Example 7

**Example 7.** *Consider the credibility base*

$$\mathbb{C}_7 = \{[A > C, K], [C > B, H], [B > A, J], [K > J, X], [J > H, Y], [Y > X, T]\}$$

*that corresponds to* $\mathbb{C}_5 \cup \{[Y > X, T]\}$*. Here, the new credibility object is expressing that for the agent Tory the newsletter Y-invest is more credible than*

*X-market. Note that from* $\mathbb{C}_7$ *we have the same arguments and attacks shown in Example 6 and, in addition, the argument* $\mathcal{A}_6 = \{[Y > X, T]\}$ *can be built. The argument* $\mathcal{A}_6$ *provides information to prefer the indirect-reliability-attack* $\varepsilon$ *over the indirect-reliability-attack* $\phi$*; this is so because* $\mathcal{A}_6$ *establishes that* $\mathcal{A}_5$ *(which is the argument producing* $\varepsilon$*) is based on more reliable sources than* $\mathcal{A}_4$ *(which is the argument that produces* $\phi$*). This preference is captured by a reliability-attack from* $\mathcal{A}_6$ *to the indirect-reliability-attack* $\phi$*, as we depict in Figure 6; this intuitions are formalized in Definition 6. In this particular case, we have that* $\mu = (\mathcal{A}_6, \phi)$ *is a reliability-attack because* $\mathsf{og}(\phi) = \mathcal{A}_4$*,* $X \in \mathsf{rel}(\mathcal{A}_4)$*, and* $\mathsf{tg}(\phi) = \delta$*,* $\mathsf{og}(\delta) = \mathcal{A}_5$ *and* $Y \in \mathsf{rel}(\mathcal{A}_5)$*.*

Next, we will present another example to show indirect-reliability-attacks that target reliability-attacks, which in turn target indirect-reliability-attacks.

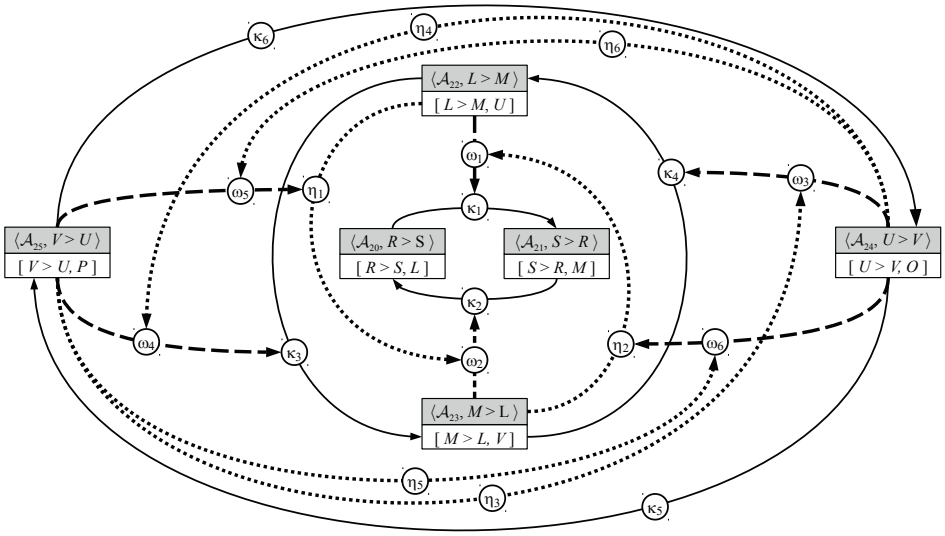



Figure 7: Arguments and attacks from Example 8

**Example 8.** *Consider the following credibility base:*

$$\mathbb{C}_8 = \{[R > S, L], [S > R, M], [L > M, U], [M > L, V], [U > V, O], [V > U, P]\}.$$

*Figure 7 shows the arguments and attacks that can be built from* $\mathbb{C}_8$*. There are six trust-attacks between these arguments:* $\kappa_1, \kappa_2$ *capture the conflict between* $\mathcal{A}_{20}$ *and* $\mathcal{A}_{21}$*,* $\kappa_3, \kappa_4$ *the conflict between* $\mathcal{A}_{22}$ *and* $\mathcal{A}_{23}$*, and* $\kappa_5, \kappa_6$ *the conflict between* $\mathcal{A}_{24}$ *and* $\mathcal{A}_{25}$*. There are four basic reliability-attacks:* $\omega_1$*,* $\omega_2$*,* $\omega_3$ *and* $\omega_4$*; and there are four basic indirect-reliability-attacks:* $\eta_1$*,* $\eta_2$*,* $\eta_3$ *and* $\eta_4$*. Note that there are two non basic reliability-attacks:* $\omega_5$ *and* $\omega_6$*, and two non basic indirect-reliability-attacks:* $\eta_5$ *and* $\eta_6$*. In particular, observe that* $\eta_5 = (\mathcal{A}_{25}, \omega_6)$ *occurs*

*because there is* $\omega_5$ *such that* $\mathsf{og}(\omega_5) = \mathcal{A}_{25}$*, and it also holds that* $\mathsf{tg}(\omega_5) = \eta_1$*,* $\mathsf{og}(\eta_1) = \mathcal{A}_{22}$ *and* $\mathsf{tg}(\omega_6) = \eta_2$*,* $\mathsf{tg}(\eta_1) = \omega_1$*,* $\mathsf{og}(\omega_1) = \mathcal{A}_{22}$*. An analogous situation occurs for* $\eta_6$*.*

In the following section, we will introduce an argumentation formalism for analyzing which arguments and attacks prevail.

## 4. Obtaining the credibility relation through argumentation

As described above, arguments can attack each other, but they can challenge attacks as well. It is clear that conflicting elements cannot be accepted together and their status need to be evaluated, and in such evaluation it is necessary to consider their acceptability status. Intuitively, an argument should be accepted if it is able to survive the attacks it receives, and should be rejected otherwise [30]. Since in our formalization attacks can also receive attacks, the evaluation process should also be capable of deciding whether an attack can be regarded as effective or not. In our proposal, the goal of the evaluation process is that after everything is considered, the accepted arguments will provide the credibility elements that will be justified from the credibility base.

In this work, we will use the evaluation approach of argumentation semantics in abstract argumentation frameworks [14]. Such semantics are abstracted from argument internal structures and are focused on establishing which arguments can be regarded as accepted considering just the attacks. In particular, given the recursive nature of the attacks in our system, we will use an AFRA introduced in [3, 4], which extends the classical approach of [14] by allowing attacks to the attack relation.

An AFRA is expressed as a set of arguments and a set of attacks, where an attack is a pair $(\mathcal{A}, \mathsf{T})$ such that $\mathcal{A}$ is an argument and $\mathsf{T}$ is either an argument or an attack [4]. To instantiate an AFRA, we will use arguments and attacks obtained from a credibility base, and then we will analyze which elements are accepted from this framework. Thus, each credibility base will characterize a particular AFRA which will be called its *associated AFRA*.

**Definition 9** (Associated AFRA)**.** *Let* $\mathbb{C}$ *be a credibility base, the associated AFRA of* $\mathbb{C}$ *is* $(\mathsf{Args}_{\mathbb{C}}, \mathsf{Atts}_{\mathbb{C}})$ *where* $\mathsf{Args}_{\mathbb{C}}$ *is the set of arguments that can be built from* $\mathbb{C}$*, and* $\mathsf{Atts}_{\mathbb{C}} = \mathsf{TAtts}_{\mathbb{C}} \cup \mathsf{RAtts}_{\mathbb{C}} \cup \mathsf{IAtts}_{\mathbb{C}}$ *with* $\mathsf{TAtts}_{\mathbb{C}}$ *its set of trust-attacks,* $\mathsf{RAtts}_{\mathbb{C}}$ *its set of reliability-attacks and* $\mathsf{IAtts}_{\mathbb{C}}$ *its set of indirect-reliability-attacks.*

It is important to remark that the sets of attacks are pairwise disjoint by their own nature, *i.e.*, every attack belongs to only one form of attack.

**Example 9.** *Consider for instance the credibility base* $\mathbb{C}_7$ *from Example 7. The asociated AFRA* $(\mathsf{Args}_{\mathbb{C}_7}, \mathsf{Atts}_{\mathbb{C}_7})$ *of that credibility base is depicted in Figure 8. Observe that* $\mathsf{Args}_{\mathbb{C}_7}$ *has ten arguments: six of them were shown in previous examples, and there are other four arguments:* $\mathcal{A}_7$*,* $\mathcal{A}_8$*,* $\mathcal{A}_9$*, and* $\mathcal{A}_{10}$ *(see Figure 8). Note that* $\mathcal{A}_3$ *is a subargument of* $\mathcal{A}_1$ *and* $\mathcal{A}_7$*;* $\mathcal{A}_9$ *is a subargument of* $\mathcal{A}_2$

*and $\mathcal{A}_7$; $\mathcal{A}_8$ is a subargument of $\mathcal{A}_1$ and $\mathcal{A}_2$. There are twelve trust-attacks in $\mathsf{TAtts}_{\mathbb{C}_7}$, fifteen reliability-attacks in $\mathsf{RAtts}_{\mathbb{C}_7}$ and six indirect-reliability-attacks in $\mathsf{IAtts}_{\mathbb{C}_7}$.*

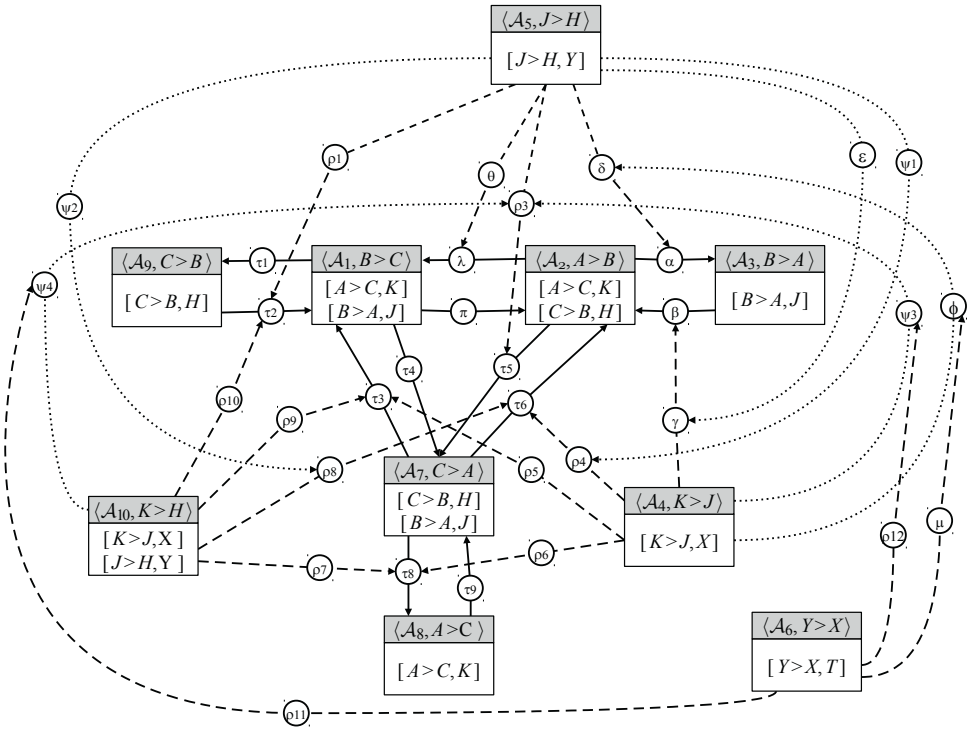



Figure 8: Associated AFRA of $\mathbb{C}_7$

Following [4], the first step towards determining the accepted arguments from an associated AFRA $(\mathsf{Args}_{\mathbb{C}}, \mathsf{Atts}_{\mathbb{C}})$ is to establish the existing attacks and consequent defeats. A defeat to an argument or an attack in an AFRA is determined by analyzing a recursive attack relation [4]. In addition, argument defeats are also propagated to the attacks they originate. In what follows, we will slightly adapt the formalization of defeats in an AFRA to use them in an associated AFRA of a credibility base. In the definitions below we will use the functions $\mathsf{og}(\cdot)$ and $\mathsf{tg}(\cdot)$ that were introduced in Section 3.2. The reader should note that these two notions correspond to *src* and *trg* defined in [3, 4] which we renamed to $\mathsf{og}$ and $\mathsf{tg}$ in order to avoid confusion with our notion of *source* of a credibility element.

**Definition 10** (Defeat in asociated AFRA)**.** *Let $\langle \mathsf{Args}_{\mathbb{C}}, \mathsf{Atts}_{\mathbb{C}} \rangle$ be the asociated AFRA of a credibility base $\mathbb{C}$, $\alpha, \beta \in \mathsf{Atts}_{\mathbb{C}}$ and $X \in \mathsf{Args}_{\mathbb{C}} \cup \mathsf{Atts}_{\mathbb{C}}$:*

- *$\alpha$ directly defeats $X$ iff $\mathsf{tg}(\alpha) = X$.*

- $\alpha$ indirectly defeats $\beta$ *iff* $\mathsf{tg}(\alpha) = \mathsf{X}$, *and* $\mathsf{X} = \mathsf{og}(\beta)$.

*It is said that* $\alpha$ *defeats* $\mathsf{X}$, *iff* $\alpha$ *directly or indirectly defeats* $\mathsf{X}$.

**Example 10.** *In the associated AFRA* $(\mathsf{Args}_{\mathbb{C}_7}, \mathsf{Atts}_{\mathbb{C}_7})$ *depicted in Figure 8, for instance, we have that* $\alpha$ *directly defeats* $\mathcal{A}_3$, $\alpha$ *indirectly defeats* $\beta$, $\beta$ *directly defeats* $\mathcal{A}_2$, $\beta$ *indirectly defeats* $\alpha$, $\delta$ *directly defeats* $\alpha$, $\gamma$ *directly defeats* $\beta$, $\phi$ *directly defeats* $\delta$, *and* $\mu$ *directly defeats* $\phi$.

Note that attacks are the only elements of the framework capable of defeating arguments and other attacks. This is also coherent with the fact that an attack can be rendered ineffective by either attacking the attack itself or the argument where originates.

In abstract argumentation, the formal definitions of declarative methods ruling the argument evaluation process are called *argumentation semantics*. In an AFRA, a semantics definition specifies how to obtain a set of extensions. An extension $E$ of an AFRA is simply a subset of arguments and attacks that represents a set of elements which can *survive together* or are *collectively acceptable* [4].

We will introduce the basic elements from which the different semantics are defined [5]. The notion of *conflict-free set* captures the idea that an extension is a set of elements that "can stand together"; in other words, the attacks in the set do not defeat elements of the set. The notion of *acceptability* establishes that an argument or an attack $\mathsf{X}$ is acceptable with respect to (or defended by) a set of elements if this set defeats every attack defeating $\mathsf{X}$. Finally, the notion of *admissibility* captures the intuition that an extension is a set of elements that "can stand on its own", *i.e.*, it is a conflict-free set and it defends all its elements.

**Definition 11** (Conflict-freeness, Acceptability, Admissibility)**.** *Let* $\langle \mathsf{Args}_{\mathbb{C}}, \mathsf{Atts}_{\mathbb{C}} \rangle$ *be the asociated AFRA of a credibility base* $\mathbb{C}$, $S \subseteq \mathsf{Args}_{\mathbb{C}} \cup \mathsf{Atts}_{\mathbb{C}}$, *and* $\mathsf{X} \in \mathsf{Args}_{\mathbb{C}} \cup \mathsf{Atts}_{\mathbb{C}}$. *Then:*

- $S$ *is* conflict-free *iff there is no* $\alpha, \mathsf{Y} \in S$ *s.t.* $\alpha \in \mathsf{Atts}_{\mathbb{C}}$ *and* $\alpha$ *defeats* $\mathsf{Y}$.
- $\mathsf{X}$ *is* acceptable *w.r.t.* $S$ *iff for all* $\alpha \in \mathsf{Atts}_{\mathbb{C}}$ *such that* $\alpha$ *defeats* $\mathsf{X}$, *there exists* $\beta \in S$ *such that* $\beta$ *defeats* $\alpha$.
- $S$ *is an* admissible set *iff* $S$ *is conflict-free and each element in* $S$ *is acceptable w.r.t.* $S$.

It is worth mentioning that the concepts presented by Definition 11 have a remarkable difference with those used in classical abstract argumentation frameworks. In such frameworks, these concepts are only defined in terms arguments [14]; instead, as we are in the context of the AFRA, here, these notions are defined in terms of arguments and attacks. This is because in an AFRA not only arguments can be defeated, but also attacks (for a more detailed discussion see [4]).

**Example 11.** *In the associated AFRA* $(\mathsf{Args}_{\mathbb{C}_7}, \mathsf{Atts}_{\mathbb{C}_7})$ *of* $\mathbb{C}_7$ *depicted in Figure 8, we have that the set* $\{\mathcal{A}_2, \mathcal{A}_3, \beta\}$ *is not conflict free because* $\beta$ *defeats* $\mathcal{A}_2$*, and the set* $\{\mathcal{A}_3, \beta, \gamma\}$ *is not conflict free since* $\gamma$ *defeats* $\beta$*. Also note that* $\mathcal{A}_3$ *is not acceptable with respect to* $\{\mathcal{A}_5, \mathcal{A}_1\}$ *since this set has no element that defeats* $\alpha$*. However,* $\mathcal{A}_3$ *is acceptable with respect to* $\{\mathcal{A}_5, \mathcal{A}_1, \pi\}$ *given that* $\pi$ *indirectly defeats* $\alpha$*. Finally, observe that the set* $\{\mathcal{A}_5, \mathcal{A}_3, \delta\}$ *is not admissible since* $\delta$ *is not acceptable with respect to the set (it is not defended form* $\phi$*), whereas* $\{\mathcal{A}_5, \mathcal{A}_3, \mu, \delta\}$ *and* $\{\mathcal{A}_1, \rho_1, \theta, \rho_5\}$ *are admissible.*

If we look further into the admissible set $\{\mathcal{A}_1, \rho_1, \theta, \rho_5\}$ from Example 11, it can be noted that adding elements such as $A_5$ or $\mu$ to that set will also result into an admissible set. We can keep adding elements and at some point we will reach a maximal admissible set such that adding any other element will result into a non admissible set. Such maximal admissible sets will hold together all the elements of the framework that can stand by their own, and thus should be regarded as accepted. This idea of maximizing accepted elements is expressed by the *preferred semantics*. Next, the formalization of these intuitions based in [3, 4] are presented.

**Definition 12** (Preferred Extension)**.** *Let* $\langle \mathsf{Args}_{\mathbb{C}}, \mathsf{Atts}_{\mathbb{C}} \rangle$ *be the asociated AFRA of a credibility base* $\mathbb{C}$ *and* $S \subseteq \mathsf{Args}_{\mathbb{C}} \cup \mathsf{Atts}_{\mathbb{C}}$*. The set* $S$ *is a* preferred extension *of* $\langle \mathsf{Args}_{\mathbb{C}}, \mathsf{Atts}_{\mathbb{C}} \rangle$ *iff it is a maximal (w.r.t.* $\subseteq$*) admissible set.*

**Example 12.** *Consider the credibility base* $\mathbb{C}_7$ *and its associated AFRA as depicted in Figure 8. This credibility base has only one preferred extension* $E_{\mathbb{C}_7} = \{\mathcal{A}_6, \mathcal{A}_5, \mathcal{A}_3, \mathcal{A}_1, \mathcal{A}_4, \mathcal{A}_8, \mathcal{A}_{10}, \mu, \rho_{11}, \rho_{12}, \rho_1, \theta, \delta, \rho_3, \psi_1, \psi_2, \epsilon, \tau_1, \pi, \tau_4, \beta, \rho_{10}, \rho_9, \rho_7, \tau_9, \rho_5, \rho_6\}$.

There are several argumentation semantics defined for AFRAs [3, 4]; nevertheless, since we do not aim to contrast the results of every conceivable semantics applied to our approach, in this work we will only consider the preferred semantics. This semantics provides a sensible approach to show how the argumentation machinery can be used for dealing with the different conflicts that we have described in the previous sections and for deciding which elements prevail. In Example 12 we have shown that in same cases one single preferred extension holds; however, as shown in Example 13, when there are mutual attacks between unchallenged arguments several extensions hold (one for each mutually exclusive option).

**Example 13.** *Consider the credibility base* $\mathbb{C}_8$ *introduced in Example 8. The associated AFRA of this base is characterized by the arguments and attacks that were depicted in Figure 7. From such associated AFRA there are two preferred extensions:* $E_{\mathbb{C}_8 1} = \{\mathcal{A}_{20}, \mathcal{A}_{23}, \mathcal{A}_{25}, \kappa_1, \kappa_4, \kappa_6, \omega_2, \omega_4, \omega_5, \eta_2, \eta_3, \eta_5\}$ *and* $E_{\mathbb{C}_8 2} = \{\mathcal{A}_{21}, \mathcal{A}_{22}, \mathcal{A}_{24}, \kappa_2, \kappa_3, \kappa_5, \omega_1, \omega_3, \omega_6, \eta_1, \eta_4, \eta_6\}$.

As it was mentioned, the justification state of a credibility element (*e.g.*, $A > B$) in a base $\mathbb{C}$ will depend on the status of the argument that supports that credibility element in the associated AFRA $\langle \mathsf{Args}_{\mathbb{C}}, \mathsf{Atts}_{\mathbb{C}} \rangle$; in turn, this status

depends on the membership to an extension of such an argument. Note that, given a credibility base $\mathbb{C}$, there always exists an associated AFRA from which it is possible to obtain the extensions containing the acceptable arguments; hence, we will introduce to our framework the notion of extension of a credibility base that considers preferred extensions which satisfy a particular constraint. There can be AFRA extensions that are not extensions for a credibility base.

**Definition 13** (Extension of a credibility base)**.** *Let $\mathbb{C}$ be a credibility base and $E$ a preferred extension of its associated AFRA $\langle \mathsf{Args}_{\mathbb{C}}, \mathsf{Atts}_{\mathbb{C}} \rangle$. We will say that $E$ is an extension of a credibility base $\mathbb{C}$ iff for any argument $\mathcal{X} \in \mathsf{Args}_{\mathbb{C}}$ if every subargument $\mathcal{Z} \subset \mathcal{X}$ is in $E$ then $\mathcal{X}$ is in $E$.*

**Definition 14** (Justified credibilities)**.** *Let $\mathbb{C}$ be a credibility base and $E$ an extension of $\mathbb{C}$, then the set of justified credibilities of $E$ is $\mathsf{J}_E = \{A > B : \langle \mathcal{X}, A > B \rangle \in E\}$.*

For instance, the credibility base $\mathbb{C}_7$ has only one extension that is the only preferred extension $E_{\mathbb{C}_7}$ of its associated AFRA $\langle \mathsf{Args}_{\mathbb{C}_7}, \mathsf{Atts}_{\mathbb{C}_7} \rangle$, which was shown in Example 12. Then, the justified credibilities of $\mathbb{C}_7$ are $\mathsf{J}_{E_{\mathbb{C}_7}} = \{X > Y, J > H, B > A, B > C, K > J, A > C, K > H\}$. Note that in this case the credibility information in $\mathbb{C}_7$ is enough for deciding in every conflictive situation. In contrast, consider now the credibility base $\mathbb{C}_m$ introduced in Example 13; in that example, we have shown that its associated AFRA has two preferred extensions $E_{\mathbb{C}_m 1}$ and $E_{\mathbb{C}_m 2}$. Observe that both are extensions of $\mathbb{C}_m$, therefore there are two possible sets of justified credibilities, $\mathsf{J}_{E_{\mathbb{C}_{m1}}} = \{L > M, R > S\}$ and $\mathsf{J}_{E_{\mathbb{C}_{m2}}} = \{M > L, S > R\}$. Having more than one justified credibility set means that the credibility information in the base is not enough for deciding in every conflictive situation. In this work, we do not propose any particular mechanism to chose among multiple extensions. Nevertheless, a skeptical approach could involve to consider the intersection of all the extensions, *i.e.*, only those arguments that appear in every extension (see [15] for more details on the implication of such approach).

Next, with the following example we show that not every extension of the associated AFRA of a credibility base is an extension of that credibility base.

**Example 14.** *Consider a credibility base $\mathbb{C}_s = \{[O > P, X], [P > Q, Y], [Q > O, Z]\}$. From this credibility base we can build six arguments: $\mathcal{A}_{31}$ for $O > P$, $\mathcal{A}_{32}$ for $P > Q$, $\mathcal{A}_{33}$ for $Q > O$, $\mathcal{A}_{34}$ for $O > Q$, $\mathcal{A}_{35}$ for $P > O$ and $\mathcal{A}_{36}$ for $Q > P$. In Figure 9 we depict the associated AFRA of $\mathbb{C}_s$. The preferred extension from such AFRA are $E_{\mathbb{C}_s 1} = \{\mathcal{A}_{31}, \mathcal{A}_{34}, \mathcal{A}_{32}, \alpha 4, \alpha 13, \alpha 10, \alpha 5, \alpha 2\}$, $E_{\mathbb{C}_s 2} = \{\mathcal{A}_{31}, \mathcal{A}_{36}, \mathcal{A}_{33}, \alpha 4, \alpha 7, \alpha 9, \alpha 6, \alpha 1\}$, $E_{\mathbb{C}_s 3} = \{\mathcal{A}_{35}, \mathcal{A}_{33}, \mathcal{A}_{32}, \alpha 3, \alpha 12, \alpha 8, \alpha 6, \alpha 2\}$, $E_{\mathbb{C}_s 4} = \{\mathcal{A}_{31}, \mathcal{A}_{33}, \mathcal{A}_{32}, \alpha 4, \alpha 6, \alpha 2\}$. Observe that $E_{\mathbb{C}_s 1}$, $E_{\mathbb{C}_s 2}$ and $E_{\mathbb{C}_s 3}$ are also extensions of $\mathbb{C}_s$, whereas $E_{\mathbb{C}_s 4}$ is not. This is because $E_{\mathbb{C}_s 4}$, for instance, includes $\mathcal{A}_{31}$ and $\mathcal{A}_{32}$ but it does not include the superargument $\mathcal{A}_{34}$. In addition, note that if $E_{\mathbb{C}_s 4}$ was considered an extension of $\mathbb{C}_s$, the set of justified credibilities for such extension would had been $\{O > P, P > Q, Q > O\}$ which clearly is not coherent with what we expect from a set of justified credibilities.*

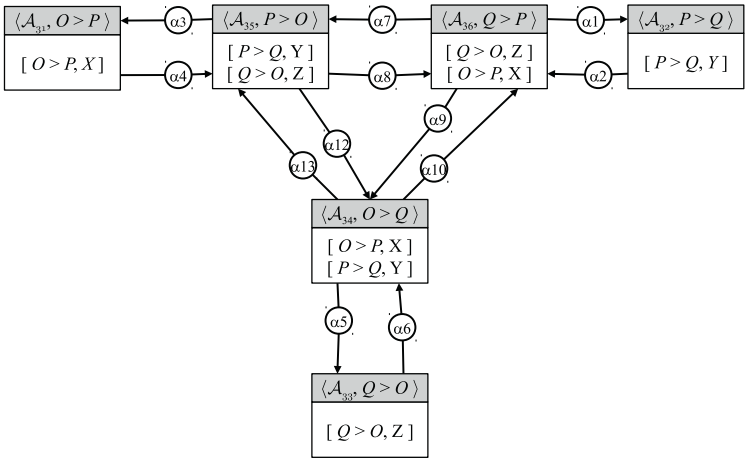



Figure 9: Associated AFRA of $\mathbb{C}_s$

It is worth mentioning that using the AFRA as the basis for the acceptance calculus allows us to apply any existing algorithm for preferred extension enumeration for Dung's classical abstract argumentation frameworks. Even though it is known that preferred extension enumeration is a costly process, the argumentation community has recently developed several methods that show promising empirical results, as it can be seen in the International Competition on Computational Models of Argumentation [40]. To use such methods, it is possible to translate an AFRA into a classical argumentation framework, as proposed in [4]. This transformation can be polynomially computed and there is a bijection between the preferred extensions of the resulting framework and those of the original AFRA.

Once the preferred extensions are calculated, we can easily discard those AFRA's extensions that are not extensions of the credibility base and then pick the conclusion of every argument in each of the remaining extensions to build the sets of justified credibilities. Both tasks can be polynomially computed. The former requires just to check for every pair of arguments in the extension if the super-argument that can be built by combining them is also in the extension. The latter can be trivially computed by iterating among the arguments in the extension.

Next, we include results that characterize our framework. The following proposition shows that an extension of a credibility base is coherent with respect to the argument composition. That is, if an argument is accepted then every part of it will be accepted as well.

**Proposition 5.** *Let $\mathbb{C}$ be a credibility base and $E$ an extension of $\mathbb{C}$. It holds that if an argument $\mathcal{X}$ is in $E$ then every subargument $\mathcal{Z}$ of $\mathcal{X}$ belongs to $E$.*

*Proof.* Assume that $\mathcal{X}$ is in a extension $E$ of $\mathbb{C}$ and suppose that there is a

subargument $\mathcal{Z}$ of $\mathcal{X}$ such that $\mathcal{Z} \notin E$; note that $\{\mathcal{Z}\} \cup E$ is conflict free. Then, by Definition 12, it holds that $\mathcal{Z}$ is not acceptable w.r.t. $E$. Therefore, since $\mathcal{Z} \notin E$ there must exist $\alpha$ defeating $\mathcal{Z}$ and there is no $\beta \in E$ such that $\beta$ defeats $\alpha$. Also, by Definition 5, 6, 7, 8 and 9, arguments are only attacked by trust-attacks, thus $\alpha \in \mathsf{TAtts}_{\mathbb{C}}$. Since $\mathcal{Z}$ is a subargument of $\mathcal{X}$, by Definition 5 there is $\gamma \in \mathsf{TAtts}_{\mathbb{C}}$ such that $\mathsf{tg}(\gamma) = \mathcal{X}$, $\mathsf{dis}(\gamma) = \mathsf{dis}(\alpha)$ and $\mathsf{og}(\gamma) = \mathsf{og}(\alpha)$. By Proposition 4 and Definition 6, we know that if there is no defeater in $E$ for $\alpha$ then there is no defeater in $E$ for $\gamma$. Hence, $\mathcal{X} \notin E$ in contradiction with the original assumption.

As we have shown, reliability-attacks can defeat trust-attacks, turning the latter ineffective. Given that trust-attacks model the conflicts between arguments with contradictory conclusions, it is important to show that even in the presence of reliability-attacks, in our formalism no extension will contain such arguments together. This is formalized in the following Lemma.

**Lemma 1.** *Let $\mathbb{C}$ be a credibility base and $E$ an extension of $\mathbb{C}$, then there is no pair of arguments $\mathcal{X}$ and $\mathcal{Y}$ in $E$, such that $\mathcal{X}$ and $\mathcal{Y}$ have contradictory conclusions.*

*Proof.* Suppose that there are $\mathcal{X}$, $\mathcal{Y}$ in $E$ such that $\mathcal{X}$ and $\mathcal{Y}$ have contradictory conclusions. By Definition 5 and Corollary 1 it holds that there are $\phi = (\mathcal{X}, \mathcal{Y})$ and $\psi = (\mathcal{Y}, \mathcal{X})$, such that $\phi, \psi \in \mathsf{TAtts}_{\mathbb{C}}$. By Definition 12, we have that $E$ is conflict free, then $\phi, \psi \notin E$. Given that by Definition 12 $\mathcal{X}$ and $\mathcal{Y}$ are acceptable with respect to $E$, there are $\alpha, \beta \in E$ such that $\alpha$ defeats $\phi$ and $\beta$ defeats $\psi$. By Definitions 5, 6, and 7 we have that $\alpha, \beta \in \mathsf{RAtts}_{\mathbb{C}}$. Then, by Definition 7 there are $\gamma, \delta \in \mathsf{IAtts}_{\mathbb{C}}$ such that $\gamma$ defeats $\alpha$ and $\delta$ defeats $\beta$. Thus, by Definition 12 there should be $\epsilon, \omega \in E$ such that $\epsilon$ defeats $\delta$ and $\omega$ defeats $\gamma$. In particular, those defeats should be direct defeats; otherwise, $\epsilon$ would also defeat $\alpha$ and $\omega$ would also defeat $\beta$. Then, by Definition 8, it holds that $\epsilon, \omega \in E$. Then, given that for $\epsilon$ and $\omega$ are in $E$ we have an analogous case to that of $\alpha$ and $\beta$. This leads us to an infinite amount of attacks, which is a contradiction.

As we mentioned in the introduction of this article, one of the main goals of our system was to determine from a potentially conflicting credibility base a strict partial order representing the credibilities that can be coherently justified. As reported in [36], a set of justified credibilities from a credibility base is sound if it is a strict partial order. The following theorem shows that our approach is sound, *i.e.*, every set of justified credibilities obtained from a credibility base is sound.

**Theorem 1.** *Let $\mathbb{C}$ be a credibility base with $\mathsf{J}_E$ the set of justified credibilities of an extension $E$ of $\mathbb{C}$. It holds that $\mathsf{J}_E$ is a strict-partial order.*

*Proof.* We have to prove that the relation expressed by the set $\mathsf{J}_E$ is irreflexive, asymmetric, and transitive.

- Irreflexive: Suppose that $\mathsf{J}_E$ is not irreflexive, then there is an $A$ such that $A > A \in \mathsf{J}_E$. Thus, there must exist an argument $\mathcal{X}$ such that its conclusion is $A > A$, which is not possible.

- Asymmetric: Straightforward from Lemma 1 and Definition 14.

- Transitive: If $A > B$ and $B > C$ are in $\mathsf{J}_E$ then we have two arguments $\mathcal{X}$ and $\mathcal{Y}$ in $E$ concluding $A > B$ and $B > C$ respectively. By Definition 2, it holds that there is an argument $\mathcal{Z} = \mathcal{X} \cup \mathcal{Y}$ concluding $A > C$. Then $\mathcal{X}$ and $\mathcal{Y}$ are subarguments of $\mathcal{Z}$ and, by Definition 13 it holds that $\mathcal{Z} \in E$. Therefore, $A > C \in \mathsf{J}_E$.

Finally, in Proposition 6 we show that when there are no conflicts in the credibility base every element in that base will be a justified credibility under any semantics. Then, in Proposition 7 we show that in a conflicting credibility base there are elements in the base that will not be justified under any semantics.

**Proposition 6.** *Let $\mathbb{C}$ be a credibility base with $\mathsf{J}_E$ as the set of justified credibilities of an extension $E$ of $\mathbb{C}$. If $\mathsf{TAtts}_{\mathbb{C}} = \emptyset$, it holds that $\mathsf{J}_E = \mathsf{cl}(\mathbb{C})$*

*Proof.*

- If $A > B \in \mathsf{J}_E$, by Definition 14 it holds that there is an argument $\langle \mathcal{X}, A > B \rangle$ in $E$, thus $\langle \mathcal{X}, A > B \rangle$ in $\mathsf{Args}_{\mathbb{C}}$. By Definition 2 we have that $\mathcal{X} \subseteq \mathbb{C}$ and $A > B \in \mathsf{cl}(\mathcal{X})$. Then $A > B \in \mathsf{cl}(\mathbb{C})$.

- If $A > B \in \mathsf{cl}(\mathbb{C})$, then there is $\mathcal{X} \subseteq \mathbb{C}$ such that $A > B \in \mathsf{cl}(\mathcal{X})$ and $\mathcal{X}$ is the minimal set holding that. Then by Definition 2, $\langle \mathcal{X}, A > B \rangle \in \mathsf{Args}_{\mathbb{C}}$. Additionally, given that $\mathsf{TAtts}_{\mathbb{C}} = \emptyset$, it holds that $\mathsf{RAtts}_{\mathbb{C}} = \emptyset$ and $\mathsf{IAtts}_{\mathbb{C}} = \emptyset$, then the associated AFRA of $\mathbb{C}$ is $\langle \mathsf{Args}_{\mathbb{C}}, \emptyset \rangle$. Therefore, by Definition 12 it holds that $E = \mathsf{Args}_{\mathbb{C}}$. Then $\langle \mathcal{X}, A > B \rangle \in E$, and by Definition 14 holds that $A > B \in \mathsf{J}_E$.

**Proposition 7.** *Let $\mathbb{C}$ be a credibility base with $\mathsf{J}_E$ the set of justified credibilities of an extension $E$ of $\mathbb{C}$. If $\mathsf{TAtts}_{\mathbb{C}} \neq \emptyset$ then it holds that $\mathsf{J}_E \subset \mathsf{cl}(\mathbb{C})$.*

*Proof.*

- If $A > B \in \mathsf{J}_E$ then, by Definition 14, there is an argument $\langle \mathcal{X}, A > B \rangle$ in $E$. By Definition 2 it holds that $\mathcal{X} \subseteq \mathbb{C}$ and $A > B \in \mathsf{cl}(\mathcal{X})$. Then $A > B \in \mathsf{cl}(\mathbb{C})$.

- Since $\mathsf{TAtts}_{\mathbb{C}} \neq \emptyset$, there are two arguments with contradictory conclusions, for instance, $\langle \mathcal{X}, A > B \rangle$ and $\langle \mathcal{Y}, B > A \rangle$. Then, by Definition 2 it holds that $A > B \in \mathsf{cl}(\mathcal{X})$, $B > A \in \mathsf{cl}(\mathcal{Y})$, $\mathcal{X} \subseteq \mathbb{C}$, and $\mathcal{Y} \subseteq \mathbb{C}$; thus, $A > B$ and $B > A$ are in $\mathsf{cl}(\mathbb{C})$. By Definition 5 $(\mathcal{X}, \mathcal{Y})$ and $(\mathcal{Y}, \mathcal{X})$ are in $\mathsf{TAtts}_{\mathbb{C}}$; so, by Definitions 10 and 12, if $\mathcal{X} \in E$ it holds that $\mathcal{Y} \notin E$ and, therefore, by Definition 14 $B > A \notin \mathsf{J}_E$. On the other hand, if $\mathcal{Y} \in E$ it holds that $\mathcal{X} \notin E$, then by Definition 14 results that $A > B \notin \mathsf{J}_E$.

## 5. Applying our approach to real world domains

A possible application of our approach is to use the proposed mechanism to estimate and maintain the credibility associated with individuals in a social network. The idea is to feed our framework with the information emerging from the interaction between the users of the network to obtain a (possibly partial) ordering of said users that will reflect their credibility. This credibility relation can then, in turn, be used by one of the users to help in deciding between contradictory opinions/positions of two other users.

Let us consider for instance Twitter, which is an online news and social networking service that allows users to post and exchange short messages known as "tweets". A user in Twitter can propagate a tweet from another user by "retweeting" it. A retweeted message contains a reference to the user who originally posted it. The strength of Twitter as a medium for information dissemination is based in large part on its speed and number of retweets. Retweeting is often used as an indication that the original information was of high value or significant interest for the retweeting user. There are studies showing that retweeting indicates, not only interest in a message, but also trust in the message and the originator [2, 26]. Therefore, retweets can be seen as indicators of trust based on the propagation of information. That is, if a user $A$ retweets messages from a user $B$ often, then it can be considered that user $A$ implicitly trusts user $B$ [39].

We can use our formalism to measure user trust resulting of the retweet behavior of a group of Twitter users. For this, based on the intuition mentioned above, we can use the retweets as a measure of credibility among users. The idea is to use the retweets from a user as an indicator of how credible are other users for her. Then, intuitively, if user $A$ retweets more from user $B$ than from another user $C$ there is an indication that $A$ will find $B$ more credible than $C$. Using that information, we will be able to construct a credibility object $[B > C, A]$. Each user will have their own credibility assessment according to their retweeting profile, and it is possible that two or more users have discrepancies in such valuation. For instance, if we have that user $D$ retweets more posts from user $C$ than posts from user $B$, then there is an indication that $D$ finds more credible $C$ than $B$, contrasting with the assessment of user $A$ discussed above. In this scenario, we will construct $[C > B, D]$. Using our formalism, if we consider the credibility objects for $A$ and $D$, we will have two conflicting arguments (they will trust-attack each other).

Another consideration regarding measuring user trust in this scenario is that we have to deal with situations where users can retweet posts among each other leading to potentially recursive valuation. Our formalism provides a tool to reason with the above-mentioned conflicts in the context of such recursivity, as we have shown through the paper. In particular, using reliability attacks, most reliable sources will help decide among conflicts of assessment. Then, a user $X$ that is the one that has been the most retweeted (that is, simply counting how much she has been retweeted) in our approach will not necessarily be the most credible user. To decide this, a qualitative argumentation analysis will be carried

out considering how reliable are the users that report $X$ as the most credible user and how reliable are the users estimating the contrary. Also, contrasting with the existing methods for user trust assessment the argumentation reasoning mechanism provides an intelligible explanation of the trust valuation process from a human point of view [17, 16]. Therefore, using our mechanism to measure the credibility of Twitter users based on their retweeting profiles, would give a user a qualitative and explanatory tool to choose when she has to make a decision regarding the information that other users provide.

In the last years, the use of trust and reputation systems for enhancing online service provision has been widely used (*e.g.*, Amazon, eBay, GoogleMaps, Trip Advisor, AirBnB, Stack Exchange). As stated in [21], those systems use the aggregated ratings about a given party to derive a trust or reputation score, which can assist other parties in deciding whether or not to transact with that party in the future. A natural side effect is that it also provides an incentive for good behavior, and therefore tends to have a positive effect on market quality. Any system that facilitates interaction between humans depends on how they respond to it and people appear to respond well to online services that have a reputation component despite some of them having drawbacks or being somewhat primitive.

Note that in all the systems mentioned in the paragraph above, users rank other users. Hence, our proposed approach can be used as a complement to enhance the way in which reputation is assessed in those systems. Consider for instance Stack Exchange which is a network of question-and-answer websites on topics in varied fields, each site covering a specific topic in which questions, answers, and users are subject to a reputation award process. Each users' reputation score goes up when others users vote up questions, answers, and edits (*e.g.*, currently, +5 for a question voted up, +10 for an answer voted up, +15 when an answer is accepted, and +2 for an edit approved). Once the reputation score of a user reaches certain level, the user unlocks privileges like the ability to vote, comment, and edit other users posts and also vote down (that costs one reputation point). For instance, with a reputation score of 15 the user has the privilege to vote, with 50 the privilege to leave comments, and with 125 to vote down. At the highest levels, users have access to special moderation tools to keep the site focused and helpful.

Similarly to the application described above for Twitter, when users vote other user's questions or answers, not only they are increasing their partners' reputation but also they are setting their preferences among them. That is, if a user A gives more positive feedback (votes, etc.) to the user B than the user C, then there is an indication that A finds B more credible than C. Also, if for a query there are several answers given for instance by users G, H, and I and user A votes for G's answer, then a preference can be inferred: that for A user G is more credible than H or I. Note that no preference of A over H or I can be entailed. That is, in terms of our proposed notation $[G > H, A]$ and $[G > I, A]$. In all the websites of the Stack Exchange community users rank other users, then, following with this example, it can also be the case that for a different user B, it can be inferred that $[H > G, B]$. Consider finally that from the same

website it can be inferred that $[A > B, X]$, then our proposed approach can be applied to decide between two contradictory answers given by users G and H, using more information than the rankings of these answers.

## 6. Related work

As we mentioned before in this article, and also in [36, 37], trust kept as a credibility order can present contradictory information, that is, from a credibility base it is possible to infer that an agent is more credible than other and vice versa. To address this issue, in [36] the authors formalized change operators over a credibility base following belief revision techniques aiming to maintain credibility bases without contradictory information. Maintaining consistency using those belief revision operators will lead to lose some information. Since obtaining information is an expensive process, in contrast to [23], in order to avoid losing information to maintain consistency, in our approach, we have proposed to keep all beliefs even in situations in which the credibility bases are contradictory. In dynamic scenarios where the trust model uses *witness information* [28, 32] with many interactions, our approach covers cases in which beliefs that are currently not important may become so in the future.

In [36] the credibility bases were defined in a similar manner as in the present work. There, in order to determine which information prevails when contradictory information arises, a reliability function was used to obtain a set of agent identifiers which represents the credibility of a given credibility element, and it considers all the agent identifiers involved in every minimal proof of that credibility element. They follow a cautious approach: the function first obtains the set with the least credible agent identifiers from each proof, and then, if there exist more than one proof, the most credible identifiers of the resulting set. Therefore, to compute the reliability of a credibility element, they use a function *min* and a function *max*. Thus, based on a credibility base, they define a function such that when given a credibility element in the transitive closure of the credibility base will return a set of agent identifiers that represents the reliability of the credibility element with respect to the credibility base. In our work, where credibility bases may contain information in conflict, it becomes unsound to compute reliability using the functions *min* and *max* because this can lead to potential cycles. For this reason, in this article, we defined an argumentation formalism with recursive attacks which provides the capability to infer a strict partial order from a credibility base (without contradictory information). In addition, in [36], the possibility that an agent receives multiple beliefs simultaneously, was not considered; for that reason, in that approach the order in which the beliefs are acquired will affect which information is retained by the agent. In contrast to [23], the proposal introduced here gives the capability of adding multiple beliefs simultaneously. This choice allows to address cases where it is possible to join two or more bases without worrying about the possible contradictions that may arise.

The works [27] and [38] propose an argumentation formalism that can be used to reason using information about trust. This formalism is described as a

set of graphs, and to determine agent's beliefs the authors propose a model which considers the trust in the information that is used for building arguments. Like ours, this approach is intended for a multi-agent setting and informant agents can have different credibilities. In contrast to our work, where each agent has its own credibility order, they use a centralized notion of trust that is codified in a shared trust network. This global network holds information about how agents trust each other and can be used to obtain an agent-centric trust network that represents the viewpoint of a particular agent. Although from these graphs it is possible to determine a credibility order for each agent, these orderings are strongly dependent on the connections in the global network. In contrast, in our work each agent has its own credibility order which is completely independent from the credibility order of any other agent. Another significant difference is that they use numerical values to establish the trust relation among agents, leading to a total order on the set of agents. In contrast, our approach uses symbolic information and an argumentation system to infer a strict partial order. Similarly to our formalism, each piece of information is linked to an agent that determines how credible this information is, and the formalism in [27, 38] uses an argumentation inference mechanism to deal with a potentially contradictory belief base. Nevertheless, unlike in our approach, in theirs there are no inconsistencies in the trust networks and their beliefs do not involve information about trust. Their argumentation system uses trust information from these networks to reason with non-trust referring beliefs while, in contrast, ours uses credibility information to reason about the potential conflicts in the credibility.

In [18, 19], in a manner similar to ours, the authors use a symbolic approach to model credibility using two global relations: the trust relation and the distrust relation. These relations together with the set of agents constitute a trust system. A pair $(a, b)$ in the trust (distrust) relation determines that agent $a$ trusts (distrusts) agent $b$. Their formalism aims to determine whether an agent trusts another taking into account the potential conflicts that may appear when the trust and distrust relations are analyzed together in the trust system. To do this, they follow an argumentation approach in which arguments represent a position for an agent to either trust or distrust a peer. Additionally, when considering an advanced version of their system, each agent is also provided with a partial order defined among its peers, using this order to codify the efficacy in which this agent trusts its peers (aiming to model a grade of trustworthiness or reputation). Even if this can be seen as similar to the credibility base of our proposal, their efficacy measure is a partial order while our bases, as we have shown in various examples, are not necessarily ordered. Therefore, their efficacy is a consistent measure while a credibility base can have credibility conflicts. In addition, they use these efficacy orders to provide strength to their arguments in order to decide if an agent trusts another or not, while we use the credibility base to build arguments and reason about the credibility itself. In this sense, the goal of the argumentation systems differ. They aim to decide if an agent trusts another or not from the trust and distrust relation, while our formalism aims to determine a partial order form a potentially contradictory credibility base. Nevertheless, similarly to us, their proposal aims to enrich the

trust mechanisms by allowing the representation of conflicting trust notions and use an argumentation mechanism to decide which information prevails. It could be interesting to study how to integrate our credibility bases with their efficacy measure and how our formalism could consider trust and distrust relations as theirs.

In our system, given the nature in which reliability-attacks are generated, a weakest link approach is used to decide between conflicting arguments (between trust-attacks, specifically). With this we aim to capture the notion that an argument for a credibility element is as weak as the weakest (least reliable) source that provided information to build such an argument. Intuitively, using this approach in a conflicting situation, the more sources used to build an argument the more chances that the argument ends up defeated. Nevertheless, it is worth to mention that, in our model, we do not explicitly deal with the problem of assessing the trustworthiness of a piece of information considering the amount of sources needed to infer it. There are more comprehensive approaches to deal with this problem like [11, 22]. In [22], an argumentation mechanism based on trust is used as a layer of a belief revision process for agents dealing with (potentially conflicting) opinions about their pairs. In that argumentation approach, trust is used to build a preference ordering amongst arguments. First, they aggregate the information of the different opinions regarding the same proposition. Then, using these aggregated propositions, they build arguments whose trustworthiness is assessed using a conjunctive fusion operator over the opinions forming the argument. This assessment considers the number of agents and information pieces which where needed for building the argument. The work of [11] is similar to [22] in the sense that they use the argumentation mechanism as part of a belief revision process to deal with potentially conflicting information obtained from different sources. Agents transmit complete arguments and, to calculate trustworthiness, they use the most trustworthy source that communicated an argument in an iterative fuzzy labeling process. Even though these works do not have explicit mechanisms for considering agents that provide information about the trust relation on their peers, it could be interesting to adapt some of their ideas to our system. For that purpose, we can think of two alternatives: either extend the notion of basic reliability-attack to consider the above mentioned strategies, or extend the AFRA to consider preferences codifying these strategies.

In [12], the authors propose a framework over multi-agent systems which uses trust to make decisions. Like ours, the agents interact exchanging information and these agents consider the sources to compute the trust of the information; however, they perform the computation of credibilities through a possibilistic model which they use to determine the acceptance of information in a framework to support belief revision. In our framework, we adopt a symbolic approach and the credibility of a piece of information is evaluated considering all the credibility objects using an argumentation framework. Furthermore, their proposal works in scenarios that are not necessarily collaborative using trust and distrust, where the credibility of a piece of information represents the capability of the agent to evaluate the tenability of the piece of information with respect to its own

competence and the ones of the sources.

In [29], a qualitative bipolar argumentative modeling of trust is proposed. Similar to our proposal, their approach is qualitative and only a finite number of levels is assumed in the trust scale. In contrast with our proposal they use a bipolar argumentative approach where trust and distrust can be independently assessed. In their approach, an agent can evaluate its trust into an object $X$ (that can be either a source or another agent) on the basis of two types of information: the observed behaviour of $X$, and the reputation of $X$ according to the other agents. Reputation information is viewed as an input information used for revising or updating its own trust evaluation based on its perception. Two kinds of arguments in favour of trusting an agent (either pointing out that a good point is reached or a bad point is avoided), and also two kinds of arguments against trusting an agent can be constructed (either pointing out that a bad point is satisfied or that a good point is not reached). These four kinds of arguments are based on an inference rule and the trust evaluation of the agent, that is represented with an interval $[t-, t+]$ over a discrete scale $S$, with the intended meaning that the trust is not larger than $t+$ and not smaller than $t-$. Although the paper indicates some basic mechanisms leading to revision of trust values, it is mainly focused on trust evaluation rather than trust dynamics in a multiple-agent world.

The work reported in [23], describes an approach that enables personalized communication about trust. The proposal is based on certain capabilities an agent must have. Namely, the ability to adapt its trust model to personalize its evaluations to the other agent's needs, the capability of communicating its criteria for evaluating trust together with the beliefs and goals that led to these criteria, and the willingness and ability to change its trust model when and if it is persuaded that it is wrong. A difference with the work presented here is that their proposal considers aggregation and a dialogue between agents to obtain conclusions regarding trust. We have concentrated on producing a system that addresses the problem of receiving trust information where the argumentation process is circumscribed to that information, while Koster et al. consider the beliefs and goals of the participants to build the arguments. It is interesting to note that it could be possible to integrate both systems in a sufficiently complex framework; this effort is outside the current goals of this research.

## 7. Conclusions and future work

The importance of having trust models has been widely emphasized in the literature. In multi-agent systems, representing and making possible the evaluation of the credibility associated with a piece of information is important, especially when the agents have their own beliefs and can obtain new information from other sources [12]. As stated in [28, 32], two elements have contributed to substantially increase the interest on trust in this area: the dissemination of the multi-agent paradigm to implement distributed systems and the dramatic evolution of e-commerce. The study of trust has many applications in Information and Communication Technologies; *e.g.*, trust has been recognized as a

key factor for successful e-commerce adoption. These systems are used by autonomous software agents as a mechanism to search for trustworthy exchange partners and as an incentive in decision-making about whether or not to honor contracts. Our proposal can be applied to any system requiring that trust or credibility of informants be taken into account or in approaches in which users can review or provide opinions about other users, as we mentioned in Section 5.

In this work, we have proposed an argumentative framework with recursive attacks to address a trust model in a collaborative open multi-agent system. We have focused in scenarios where agents share information about the credibility or informational trust they have assigned to their peers (referred to as *witness information* in the literature). We have represented informants' credibility through credibility objects which include not only trust information (credibility element) but also the informant source. These objects are maintained in a credibility base which can store information in conflict. Thus, when an agent needs to determine the credibility of its peers to make decisions, the agent needs an ordered set of informants without information in conflict.

Here, we have presented a system capable of building a partially ordered credibility relation from a potentially contradictory credibility base by using argumentative inference. Such relation is not any arbitrary subset of partially ordered credibility elements obtained from the credibility base; actually, the output of our system is a subset of credibility elements that can be justified after an argumentation process. In this process, we aimed to select the most reliable information when conflicts are analyzed. In the formalism, an argument is a reason why an agent may infer that an informant is preferred over other; these reasons are tentative and can be challenged for other reasons through attacks. As we have shown, the reliability measure in our approach depends only on the credibility relation, leading to a recursive scenario where this credibility can be also in conflict or challenged. To capture these recursive relations, we have introduced different forms of attacks to attacks. The key element for this are the reliability-attacks, whose goal is to capture the intuition that an argument or an attack based on potentially more credible sources can not be defeated by another with lesser credible sources. These attacks are considered by the argumentation mechanism to decide which arguments are accepted, which in turn are the arguments that support the credibility elements that we consider as justified.

To reason argumentatively with the recursive interactions, we instantiated an AFRA [3, 4] with the arguments and attacks of a credibility base, and used the notions of acceptability semantics already defined for such framework. This framework was selected because it was conceived and equipped with tools that naturally deal with recursive attacks and regards attacks as defeasible entities. As pointed out in [7] the defeasibility of attacks can be a useful modeling tool when we are dealing with meta-argumentation such as in our case: an argument producing a reliability-attack to a trust-attack is arguing about the acceptability status of arguments involved in the trust-attacks. Alternatively, we could have used another argumentation framework such as [14]; however, that would have lead us to use an artificial encoding of the recursive relations using special

arguments and to formalize the intuition of the acceptance status of these special arguments.

We used the notion of acceptability semantics to determine which are the acceptance status of the arguments in our system; in particular, we committed to the extensional approach of the preferred semantics. We have formally shown that extensions are sound, *i.e.*, there is no pair of arguments holding contradictory conclusions in an extension, and we have shown that the conclusions of an extension from a credibility base constitute a partial order with respect to the credibility relation. We have also shown that when a credibility base has several extensions, each one corresponds to a choice of arguments and attacks involved in conflicts that could not be resolved using reliability.

As future work, we intent to develop a complete implementation of our proposal. Our idea is to develop tools for generating the arguments and attacks of a credibility base and integrate them with some of the recent methods for extension enumeration in abstract argumentation (such as jArgSemSAT [10]). For such an integration, as previously mentioned in Section 4, we could use the existing methods for transforming AFRAs into classical argumentation frameworks. Having such implementation will allow us to throughly evaluate our approach using real world data, in the context of the applications described in Section 5. We also want to integrate the notion of distrust along to the credibility relation, in non collaborative scenarios. Our idea in this regard is to use another type of attack to model the inherent conflicts between these two notions, and to consider the problem of identity theft. Also we will study how to integrate our credibility bases with the efficacy measure used in [18, 19] to reason with trust and distrust relations. In addition, we want to study how our approach can be combined with other argumentation frameworks that use trust as a decision support mechanism, such as [38]. Finally, it is also clear that a bootstrapping mechanism is needed; one alternative to effect this process could be to use direct experience [32]. Such alternative is left as future work.

## Acknowledgments

This work has been partially supported by EU H2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 690974 for the project MIREL: MIning and REasoning with Legal texts, and by funds provided by CONICET, Universidad Nacional del Sur, Argentina, by PGI-UNS (grants 24/ZN32, 24/N040, 24/N035).

## References

[1] Sibel Adali. *Modeling Trust Context in Networks*. Springer Briefs in Computer Science. Springer, 2013.

[2] Sibel Adali, Robert Escriva, Mark K. Goldberg, Mykola Hayvanovych, Malik Magdon-Ismail, Boleslaw K. Szymanski, William A. Wallace, and Gregory Todd Williams. Measuring behavioral trust in social networks. In

*IEEE International Conference on Intelligence and Security Informatics, ISI 2010, Vancouver, BC, Canada, May 23-26, 2010, Proceedings*, pages 150–152, 2010.

[3] Pietro Baroni, Federico Cerutti, Massimiliano Giacomin, and Giovanni Guida. Encompassing attacks to attacks in abstract argumentation frameworks. In *10th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Verona, Italy*, pages 83–94. Lecture Notes in Artificial Intelligence, vol. 5590. Springer, Germany, 2009.

[4] Pietro Baroni, Federico Cerutti, Massimiliano Giacomin, and Giovanni Guida. AFRA: Argumentation Framework with Recursive Attacks. *International Journal of Approximate Reasoning*, 52(1):19–37, 2011.

[5] Pietro Baroni and Massimiliano Giacomin. Semantics of abstract argument systems. In *Argumentation in Artificial Intelligence*, pages 25–44. Springer, 2009.

[6] Salem Benferhat, Didier Dubois, Henri Prade, and Mary-Anne Williams. A practical approach to revising prioritized knowledge bases. *Stud. Log.*, 1(70):105–130, 2002.

[7] Guido Boella, Dov M. Gabbay, Leendert W. N. van der Torre, and Serena Villata. Meta-argumentation modelling I: methodology and techniques. *Studia Logica*, 93(2-3):297–355, 2009.

[8] John Cantwell. Resolving conflicting information. *Journal of Logic, Language and Information*, 7(2):191–220, 1998.

[9] Christiano Castelfranchi and Rino Falcone. *Trust theory: A socio-cognitive and computational model*, volume 18. John Wiley & Sons, 2010.

[10] Federico Cerutti, Mauro Vallati, and Massimiliano Giacomin. An Efficient Java-Based Solver for Abstract Argumentation Frameworks: jArgSemSAT. *International Journal on Artificial Intelligence Tools*, 26(2):1–26, 2017.

[11] Célia da Costa Pereira, Andrea Tettamanzi, and Serena Villata. Changing one's mind: Erase or rewind? In *IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011*, pages 164–171, 2011.

[12] Célia da Costa Pereira, Andrea G. B. Tettamanzi, and Serena Villata. A belief-based approach to measuring message acceptability. *Proceedings of the 10th International Conference on Scalable Uncertainty Management (SUM 2016)*, pages 140–154, September 2016.

[13] Aldo Dragoni, Paolo Giorgini, and Paolo Puliti. Distributed belief revision versus distributed truth maintenance. In *Proceedings of the Sixth IEEE International Conference on Tools with Artificial Intelligence (TAI 94)*, pages 499–505. IEEE Computer Society Press, 1994.

[14] Phan M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.

[15] Phan Minh Dung, Paolo Mancarella, and Francesca Toni. Computing ideal sceptical argumentation. *Artif. Intell.*, 171(10-15):642–674, 2007.

[16] Xiuyi Fan and Francesca Toni. On computing explanations in argumentation. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA.*, pages 1496–1502, 2015.

[17] Alejandro Javier García, Carlos Iván Chesñevar, Nicolás D. Rotstein, and Guillermo Ricardo Simari. Formalizing dialectical explanation support for argument-based reasoning in knowledge-based systems. *Expert Syst. Appl.*, 40(8):3233–3247, 2013.

[18] William T. Harwood, John A. Clark, and Jeremy L. Jacob. Networks of trust and distrust: Towards logical reputation systems. In *Logics in Security*, 2010.

[19] William T. Harwood, John A. Clark, and Jeremy L. Jacob. A perspective on trust, security and autonomous systems. In *New Security Paradigms Workshop*, 2010.

[20] Trung Dong Huynh, Nicholas R. Jennings, and Nigel R. Shadbolt. An integrated trust and reputation model for open multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 13(2):119–154, 2006.

[21] Audun Jøsang, Roslan Ismail, and Colin Boyd. A survey of trust and reputation systems for online service provision. *Decision Support Systems*, 43(2):618–644, 2007.

[22] Andrew Koster, Ana L. C. Bazzan, and Marcelo de Souza. Liar liar, pants on fire; or how to use subjective logic and argumentation to evaluate information from untrustworthy sources. *Artif. Intell. Rev.*, 48(2):219–235, 2017.

[23] Andrew Koster, Jordi Sabater-Mir, and W. Marco Schorlemmer. Personalizing communication about trust. In *International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2012, Valencia, Spain, June 4-8, 2012 (3 Volumes)*, pages 517–524, 2012.

[24] Patrick Krümpelmann, Luciano H. Tamargo, Alejandro J. García, and Marcelo A. Falappa. Forwarding credible information in multi-agent systems. *Proceedings of the 3rd International Conference on Knowledge Science, Engineering and Management (KSEM 2009)*, 5914/2009:41–53, November 2009.

[25] Michael Luck, Peter McBurney, Onn Shehory, and Steve Willmott. Agent technology: computing as interaction (a roadmap for agent based computing). 2005.

[26] Panagiotis Takis Metaxas, Eni Mustafaraj, Kily Wong, Laura Zeng, Megan O'Keefe, and Samantha Finn. What do retweets indicate? results from user survey and meta-review of research. In *Proceedings of the Ninth International Conference on Web and Social Media, ICWSM 2015, University of Oxford, Oxford, UK, May 26-29, 2015*, pages 658–661, 2015.

[27] Simon Parsons, Elizabeth Sklar, and Peter McBurney. Using argumentation to reason with and about trust. In *Argumentation in Multi-Agent Systems - 8th International Workshop, ArgMAS 2011, Taipei, Taiwan, May 3, 2011, Revised Selected Papers*, pages 194–212, 2011.

[28] Isaac Pinyol and Jordi Sabater-Mir. Computational trust and reputation models for open multi-agent systems: a review. *Artif. Intell. Rev.*, 40(1):1–25, 2013.

[29] Henri Prade. A qualitative bipolar argumentative view of trust. In *Scalable Uncertainty Management, First International Conference, SUM 2007, Washington, DC, USA, October 10-12, 2007, Proceedings*, pages 268–276, 2007.

[30] Iyad Rahwan and Guillermo R. Simari. *Argumentation in Artificial Intelligence*. Springer, Heidelberg, Germany, 2009.

[31] Sarvapali D. Ramchurn, Trung Dong Huynh, and Nicholas R. Jennings. Trust in multi-agent systems. *The Knowledge Engineering Review*, 19(1):1–25, 003 2004.

[32] Jordi Sabater-Mir and Carles Sierra. Review on computational trust and reputation models. *Artificial Intelligence Review*, 24(1):33–60, 2005.

[33] Jordi Sabater-Mir and Laurent Vercouter. *Multiagent Systems. Second Edition*, chapter Book chapter 9: Trust and Reputation in Multiagent Systems, pages 381–420. The MIT Press, 2013.

[34] Jesse R. Sparks and David N. Rapp. Unreliable and anomalous: How the credibility of data affects belief revision. In Laura A. Carlson, Christoph Hölscher, and Thomas F. Shipley, editors, *Proceedings of the 33th Annual Meeting of the Cognitive Science Society, CogSci 2011, Boston, Massachusetts, USA, July 20-23, 2011*. cognitivesciencesociety.org, 2011.

[35] Luciano H. Tamargo, Alejandro J. García, Marcelo A. Falappa, and Guillermo R. Simari. Modeling knowledge dynamics in multi-agent systems based on informants. *The Knowledge Engineering Review (KER)*, 27(1):87–114, 2012.

[36] Luciano H. Tamargo, Alejandro Javier García, Marcelo A. Falappa, and Guillermo R. Simari. On the revision of informant credibility orders. *Artificial Intelligence*, 212:36–58, 2014.

[37] Luciano H. Tamargo, Sebastian Gottifredi, Alejandro J. García, and Guillermo R. Simari. Sharing beliefs among agents with different degrees of credibility. *Knowledge and Information Systems (KAIS)*, 47(43):1–33, 2016.

[38] Yuqing Tang, Kai Cai, Peter McBurney, Elizabeth Sklar, and Simon Parsons. Using argumentation to reason about trust and belief. *J. Log. Comput.*, 22(5):979–1018, 2012.

[39] Mozhgan Tavakolifard, Kevin C. Almeroth, and Jon Atle Gulla. Does social contact matter?: modelling the hidden web of trust underlying twitter. In *22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013, Companion Volume*, pages 981–988, 2013.

[40] Matthias Thimm, Serena Villata, Federico Cerutti, Nir Oren, Hannes Strass, and Mauro Vallati. Summary report of the first international competition on computational models of argumentation. *AI Magazine*, 37(1):102, 2016.