

From Trust Among Agents to Reputation of Abstract Arguments by Using Subjective Logic

Francesco Santini

Department of Mathematics and Computer Science, University of Perugia, Italy
francesco.santini@dmi.unipg.it

Abstract. Subjective Logic provides a standard set of logical operators intended for use in domains containing uncertainty. At the same time, the motivations behind the adoption of Argumentation in AI are rooted into reasoning and explanation in presence of incomplete and uncertain information. This work uses Subjective Logic as a means to represent the beliefs of different agents towards arguments and attacks, and to aggregate them with the purpose to have an overall reputation from all the considered agents in the considered community. Agents are also allowed to form their opinion from others' opinions by exploiting trust paths. Finally, the obtained beliefs can be used to compute the community-biased expectation that a set of (abstract) arguments satisfies a given semantics.

1 Introduction

An *Abstract Argumentation Framework* (AAF) [10] is an abstract structure consisting of a set arguments, whose origin, nature and possible internal organisation is not specified, and by a binary relation of attack on the set of arguments, whose meaning is not specified either: that is, an AAF can be represented as a pair $\langle \mathcal{A}, \mathcal{R} \rangle$, which in turn can be represented as a directed graph where nodes are arguments and $a \rightarrow b$ if $(a, b) \in \mathcal{R}$. As a classical example, argument a may stand for “Tomorrow will rain because the national weather forecast says so”, while argument b for “Tomorrow will not rain because the regional weather forecast says so”; the corresponding framework is $\langle \mathcal{A} = \{a, b\}, \mathcal{R} = \{(a, b), (b, a)\} \rangle$.

Given a framework, it is possible to examine the question on which set(s) of arguments can be accepted, hence collectively surviving the conflict defined by R . Answering this question corresponds to defining an argumentation semantics [10]. Considering the previous example, either $\{a\}$ or $\{b\}$ alone can be accepted, while $\{a, b\}$ cannot be accepted because of the internal conflict.

Subjective Logic (SL) [16] is a calculus for subjective opinions which in turn represent probabilities affected by degrees of uncertainty. In general, SL is suitable for modelling and analysing situations involving uncertainty and relatively unreliable sources. A subjective opinion can express trust in a source or it can express belief about events and propositions. A binomial opinion applies to a binary state variable, and can be represented as a *Beta PDF* (Probability Density Function) [16]. A multinomial opinion applies to a state variable of multiple possible values, and can be represented as a *Dirichlet PDF* [16]. SL has been already used for modelling subjective *trust networks* [17] and *structured Argumentation* [20].

Since arguments are often uncertain, it can be useful to quantify the uncertainty associated with each argument, as previously explored in other works in the literature [14, 13, 19, 26]. Do we believe more in national or regional weather forecast? How much are we certain about our belief? For this reason, we define *Subjective Logic-based AAFs* (sIAAFs), where both arguments and attacks are associated with a binomial opinion defined in SL, i.e., described in terms of *belief*, *disbelief*, and *uncertainty* values, i.e., $\langle b, d, u \rangle$. As shown in [24], sIAAFs can be straightforwardly reconnected to the *constellations* approach proposed in [19], but information is more granular due to the fact that a probability value can be derived from a triple $\langle b, d, u \rangle$. A *dogmatic* opinion, that is with $u = 0$, is equivalent to probabilities. An *absolute* opinion, that is $b = 1$, is equivalent to *true*. A *vacuous* opinion, that is $u = 1$ is equivalent to *undefined*.

Afterwards, with the purpose to find a framework to assign opinions to arguments and attacks, we introduce agents on top of sIAAFs. In this scenario, new with respect to [24], different opinions related to arguments and attacks between arguments come from different agents. Consequently, SL operators can be used to aggregate these subjective opinions together in a resulting opinion, which describes the belief/disbelief/uncertainty of the whole group of agents. This represents the reputation of an argument (or attack) in the considered *community*, which consist of individuals bounds together by social relationships. This reputation comes from all the direct subjective-opinions of agents, but also from (indirect) opinions of other agents in the same community, by considering transitive trust-relationships: if A trusts B who strongly believes in argument a (i.e., high belief and low uncertainty rating), then the *direct* opinion ω_a^A can be aggregated with ω_a^B through the opinion of A towards B : ω_B^A . If A has no opinion about a , then she can make one as just explained. By aggregating the beliefs of all the agents w.r.t. the same argument/attack, e.g., $\omega_a^A - \omega_a^B - \omega_a^C$, then we compute the reputation of a . The same example can be rephrased by computing the reputation towards an attack from a to b , i.e., (a, b) : $\omega_{(a,b)}^A - \omega_{(a,b)}^B - \omega_{(a,b)}^C$. Finally, these opinions can be used in the same way as in the *constellations* approach (Section 5), for instance to find the community-biased expectation that a set of arguments satisfies a given semantics.

This work extends the results in [24] by introducing *Trust Network-based sIAAF* as a way to connect trust in agents with trust in arguments and attacks. Concisely, the paper links a trust network among agents with the belief the same agents have in the components of the considered AAF.

The paper is organised as follows: in Section 2 we summarise the background notions behind SL. Section 3 proposes Subjective Logic-based AAFs and how to work with them by computing the expectations of semantics and argument acceptance by using opinions instead of probability values. Then, in Section 4 we embed a trust model in top of sIAAFs: we describe how trust paths among agents can be used to compute an indirect opinion on arguments and attacks. Section 5 and Section 6 ends the paper with related work and conclusions respectively.

2 Subjective Logic

A subjective opinion expresses belief about states of a state space called a “*frame of discernment*”, or “*frame*” for short. In practice, a state in a frame can be regarded

as a statement or proposition, so that a frame contains a set of statements. Let $X = \{x_1, x_2, \dots, x_k\}$ be a frame of cardinality k , where x_i ($1 \leq i \leq k$) represents a specific state. An opinion distributes belief mass over the reduced power-set of the frame denoted as $\mathcal{R}(X)$ defined as:

$$\mathcal{R}(X) = \mathcal{P}(X) \setminus \{X, \emptyset\}, \quad (1)$$

where $\mathcal{P}(X)$ denotes the powerset of X and $|\mathcal{P}(X)| = 2^k$. All proper subsets of X are states of $\mathcal{R}(X)$, but the frame X and the empty set \emptyset are not states of $\mathcal{R}(X)$, in line with the hyper-Dirichlet model [12]. $\mathcal{R}(X)$ has cardinality $\kappa = 2^k - 2$.

An opinion is a composite function consisting of belief masses, uncertainty mass and base rates. It applies to a frame, also called a state space, and can have an attribute that identifies the belief owner. An opinion is a composite function that consists of a belief vector \mathbf{b} ,¹ an uncertainty parameter u , and base rate vector \mathbf{a} ,² which take values in the interval $[0, 1]$. An opinion satisfies the following additivity requirements.

$$\text{Belief additivity: } u_X + \sum_{x_i \in \mathcal{R}(X)} \mathbf{b}_X(x_i) = 1. \quad (2)$$

$$\text{Base rate additivity: } \sum_{i=1}^k \mathbf{a}_X(x_i) = 1, \text{ where } x_i \in X. \quad (3)$$

A subjective (hyper) opinion of user A over the frame X is denoted as $\omega_X^A = (\mathbf{b}_X, u_X, \mathbf{a}_X)$, where \mathbf{b}_X is a belief vector over the states of $\mathcal{R}(X)$, u_X is the complementary uncertainty mass, and \mathbf{a}_X is a base rate vector over X , all seen from the viewpoint of belief owner A. The belief vector \mathbf{b}_X has $(2^k - 2)$ parameters, whereas the base rate vector \mathbf{a}_X only has k parameters. The uncertainty parameter u_X is a simple scalar. Thus, a general opinion contains $(2^k + k - 1)$ parameters and hence it is a hyper opinion. However, given that Eq.(2) and Eq.(3) remove one degree of freedom each, opinions over a frame of cardinality k only have $(2^k + k - 3)$ degrees of freedom. The probability projection of hyper opinions is the vector denoted as E_X in Eq.(4).

$$E_X(x_i) = \sum_{x_j \in \mathcal{R}(X)} \mathbf{a}_X(x_i/x_j) \mathbf{b}_X(x_j) + \mathbf{a}_X(x_i) u_X, \quad (4)$$

where $x_i \in \mathcal{R}(X)$ and $\mathbf{a}_X(x_i/x_j)$ denotes relative base rate, i.e. the base rate of subset x_i relative to the base rate of (partially) overlapping subset x_j , defined as follows:

$$\mathbf{a}_X(x_i/x_j) = \frac{\mathbf{a}_X(x_i \cap x_j)}{\mathbf{a}_X(x_j)}, \quad \forall x_i, x_j \in \mathcal{R}(X). \quad (5)$$

Equivalent probabilistic representations of opinions, e.g. as Beta pdf (probability density function) or a Dirichlet pdf, offer an alternative interpretation of subjective opinions in terms of traditional statistics [18]. There is no simple visualisation of hyper

¹ A belief vector \mathbf{b} specifies the distribution of belief masses over the elements of $\mathcal{R}(X)$.

² Base rate generally refers to the (base) class probabilities unconditioned on featural evidence, frequently also known as *prior probabilities*. The concept of base rates is central in the theory of probability. Base rates are for example useful for default and for conditional reasoning.

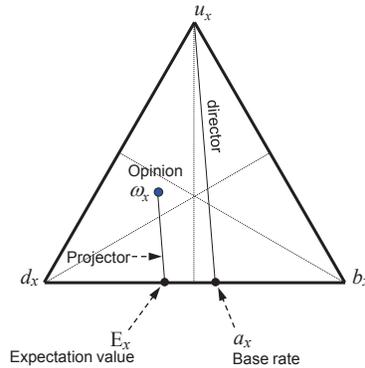


Fig. 1. Binomial opinion point in triangle.

opinions, but simple visualisations can be used for binomial and multinomial opinions as explained below.

Binomial opinions, which be extensively used in the remainder of the paper, apply to binary frames and have a special notation as described below. Let $X = \{x, \bar{x}\}$ be a binary frame, then a binomial opinion about the truth of state x is the ordered quadruple $\omega_x = \langle b, d, u, a \rangle$ where:

- b , *belief*: belief mass in support of x being true;
- d , *disbelief*: belief mass in support of \bar{x} (NOT x);
- u , *uncertainty*: uncertainty about probability of x ;
- a , *base rate*: non-informative prior probability of x .

The special case of Eq.(2) in case of binomial opinions is expressed by Eq.(6).

$$b + d + u = 1. \quad (6)$$

Similarly, the special case of the probability expectation value of Eq.(4) in case of binomial opinions is expressed by Eq.(7).

$$E_x = b + au. \quad (7)$$

A binomial opinion can be visualised as a point inside an equal sided triangle as shown in Figure 1, where the belief, disbelief and uncertainty axes go perpendicularly from each edge to the opposite vertex indicated by b_x , d_x and u_x . The base rate a_x shows on the base line, and the probability expectation value E_x is determined by projecting the opinion point to the base line in parallel with the base rate director.

In case the opinion point is located at the left or right corner of the triangle, i.e. with $d = 1$ or $b = 1$ and $u = 0$, the opinion is equivalent to boolean TRUE or FALSE, then SL becomes equivalent to binary logic. Moreover, where $b + d = 1$ a binomial opinion is equivalent to a traditional probability, where $b + d < 1$ it expresses degrees of uncertainty, and where $b + d = 0$ it expresses total uncertainty.

Most operators in Table 1 are generalisations of binary logic and probability operators. For example, addition is simply a generalisation of addition/union of probabilities,

Subjective Logic operators	Operator notation
Addition	$\omega_{x \cup y}^A = \omega_x^A + \omega_y^A$
Subtraction	$\omega_{x \setminus y}^A = \omega_x^A - \omega_y^A$
Multiplication	$\omega_{x \wedge y}^A = \omega_x^A \cdot \omega_y^A$
Division	$\omega_{x / y}^A = \omega_x^A \setminus \omega_y^A$
Comultiplication	$\omega_{x \vee y}^A = \omega_x^A \sqcup \omega_y^A$
Codivision	$\omega_{x \bar{\vee} y}^A = \omega_x^A \sqcap \omega_y^A$
Complement	$\omega_{\bar{x}}^A = \neg \omega_x^A$
Deduction	$\omega_{y x}^A = \omega_x^A \odot (\omega_{y x}^A, \omega_{y \bar{x}}^A)$
Abduction	$\omega_{y x}^A = \omega_x^A \odot (\omega_{y x}^A, \omega_{x y}^A, a_y)$
Transitivity / discounting	$\omega_x^{A:B} = \omega_B^A \otimes \omega_x^B$
Cumulative fusion / consensus	$\omega_x^{A \circ B} = \omega_x^A \oplus \omega_x^B$
Averaging fusion	$\omega_x^{A \ominus B} = \omega_x^A \oplus \omega_x^B$
Constraint fusion	$\omega_x^{A \& B} = \omega_x^A \odot \omega_x^B$

Table 1. Some SL operators. For a more detailed explanation of them refer to [16]. The superscripts A and B are attributes that identify the respective belief sources or belief owners (e.g., two agents uttering arguments); x is a state in the considered frame of discernment (e.g., an argument or an attack).

while multiplication is conjunction/and. Other operators, e.g., deduction, abduction, discounting, are not related to logic instead. For the mathematical details of the operators in Table 1, refer to [16]. Some of the operators are only meaningful for combining binomial opinions, but some also apply to multinomial opinions. Most of the operators in Table 1 are binary, but complement is unary, deduction is ternary and abduction is quaternary.

3 SL-based Abstract Argumentation Frameworks

In this section we redefine the *constellations* approach in *probabilistic argumentation* (see Section 5) by using SL instead of plain probability values on arguments and attacks (as accomplished in the standard definition of the constellations approach instead). All the results in this section are background information taken from [10] (for what concerning AAFs) and [24] (concerning slAAF). We start by recalling the classical definitions behind AAFs:

Definition 1 (Abstract Argumentation Frameworks [10]). *An Abstract Argumentation Framework (AAF) is a pair $\langle A, R \rangle$ of a set A of arguments and a binary relation R on A , called attack relation. $\forall a_i, a_j \in A$, $R(a_i, a_j)$ means that a_i attacks a_j (R is asymmetric).*

A semantics specifies how to derive a set of *extensions* from an AAF, where an extension $B \subseteq A$ is a subset of “collectively” acceptable arguments.

Definition 2 (Semantics [10]). *Let $F = \langle A, R \rangle$ be an AAF. A set $B \subseteq A$ is conflict-free, denoted $B \in cf(F)$, iff there are no $a, b \in B$, such that $R(a, b)$. An argument $a \in A$ is*

defended by a set $B \subseteq A$ if for each $b \in A$, such that $R(b, a)$, there is $c \in B$ s.t. $R(c, b)$. A conflict-free set is also admissible, that is $S \in \text{adm}(F)$, if each $a \in B$ is defended by B . Given a conflict-free B , the semantics originally defined in [10] are:

complete: $B \in \text{com}(F)$, if $B \in \text{adm}(F)$ and for each $a \in A$ defended by B , $a \in B$ holds;
preferred: $B \in \text{prf}(F)$, if $B \in \text{adm}(F)$ and there is no $C \in \text{adm}(F)$ with $B \subset C$;
stable: $B \in \text{stb}(F)$, if for each $a \in A \setminus B$, $\exists b \in B$ s.t. $R(b, a)$;
grounded: $B = \text{grd}(F)$ if $B \in \text{com}(F)$ and there is no $C \in \text{com}(F)$ with $C \subset B$.

The acceptance state of a single argument can be conceived in terms of its extension membership.

Definition 3 (Argument acceptance [23]). Given one of the semantics $\sigma \in \{\text{com}, \text{stb}, \text{prf}\}$ and a framework F , an argument a is i) justified iff $\forall B \in \sigma(F)$, $a \in B$, ii) a is defensible if $\exists B \in \sigma(F)$, $a \in B$ and a is not justified, iii) a is overruled iff $\nexists B \in \sigma(F)$, $a \in B$.

A SL-based AAF extends Dung’s argument framework by associating an opinion with each argument and attacks in the original AAF.

Definition 4 (SL-based Argumentation Frameworks). A SL-based Abstract Argumentation framework (sIAAF) is a tuple $\langle \mathcal{A}, \mathcal{R}, \mathcal{O}_{\mathcal{A}}, \mathcal{O}_{\mathcal{R}} \rangle$ where $\langle \mathcal{A}, \mathcal{R} \rangle$ is a Dung’s AAF (Definition 1), $\mathcal{O}_{\mathcal{A}} : \mathcal{A} \rightarrow \Omega_{\mathcal{A}}$ and $\mathcal{O}_{\mathcal{R}} : \mathcal{R} \rightarrow \Omega_{\mathcal{R}}$, where $\Omega_{\mathcal{A}}$ and $\Omega_{\mathcal{R}}$ respectively are the set of binomial opinions on each argument and each attack.

Hence, given $\mathcal{A} = \{a_1, a_2, \dots, a_n\}$, for each $a_i \in \mathcal{A}$ we have that $X_{a_i} = \{a_i, \bar{a}_i\}$ represents a binary frame where, with an abuse of notation, a_i indicates that “argument a_i is trustworthy” and where \bar{a}_i states that “argument a_i is not trustworthy”. The same considerations hold for $\mathcal{R} = \{(a_i, a_j), \dots, (a_l, a_k)\}$: $X_{(a_i, a_j)} = \{(a_i, a_j), \overline{(a_i, a_j)}\}$ represents a binary frame where (a_i, a_j) indicates “attack (a_i, a_j) is trustworthy”, and $\overline{(a_i, a_j)}$ states “ $\overline{(a_i, a_j)}$ is not trustworthy”.³ Therefore, our framework collects a binomial opinion $\omega_{a_i} = \langle b, d, u, a \rangle$ for each $a_i \in \mathcal{A}$, and $\omega_{(a_i, a_j)} = \langle b, d, u, a \rangle$ for each $(a_i, a_j) \in \mathcal{R}$.

Remark 1. In this paper, we suppose agents trust an argument if they generically believe in that argument: for instance, if they believe its premises are true, and if they believe the consequence of the claim is logically sound. Therefore, an agent trusts an argument if it believes it is both *valid* and *sound*. Indeed this evaluation is subjective: some agents might not catch a statement is a *fallacy* instead,⁴ or the fact some of the premises are just false instead of true. Different agents possess different knowledge about the same facts. Similarly, agents can differently judge whether two arguments are in conflict or not, if such arguments do not exactly negate each other: for example, “doing a ” or “doing b ” in the same time interval are not in conflict if they there is time to do both of them in sequence.

³ Note that i can be equal to j in case we have a self attack $R(a_i, a_i)$.

⁴ A fallacy is the use of invalid or otherwise faulty reasoning in the construction of an argument. For instance, *hasty generalization* is making assumptions about a whole group or range of cases based on a sample, e.g., “graduated students are nerd”.

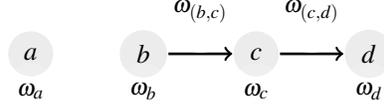


Fig. 2. An example of Subjective Logic-based Abstract Argumentation Framework (slAAF). We use the same example proposed in [19] in order to better show the differences between the two approaches (i.e., probability values and SL).

Table 2. Opinions for the slAAF in Figure 3.

Table 3. Opinions on arguments.

opinion	b	d	u
ω_a	1	0	0
ω_b	0.6	0.2	0.2
ω_c	1	0	0
ω_d	0.5	0.2	0.3

Table 4. Opinions on attacks.

opinion	b	d	u
$\omega_{(b,c)}$	1	0	0
$\omega_{(c,d)}$	1	0	0

In Figure 2 we show an example of slAAF; we use the same example used in [19], in order to better show the differences between the original constellations approach in [19], and by using SL instead. In Table 2 we provide the values for the tuples $\omega_{a_i} = \langle b, d, u, a \rangle$ and $\omega_{w_{(a_i, a_j)}} = \langle b, d, u, a \rangle$ with respect to the AAF in Figure 2.

As a reminder from Section 2, the base rate a is the prior probability of the proposition in the absence of specific belief or disbelief. The default value is the relative atomcity, i.e., 0.5 for a binary state space containing the proposition and its negation. For this reason, a is not reported in Table 2, and quadruples are in the following simplified as triples $\langle b, d, u \rangle$. The opinion related to the complement, e.g., $\omega_{\bar{b}}$, is not reported because it can be obtained easily from $\omega_b = \langle 0.6, 0.2, 0.2 \rangle$ as $\langle 0.2, 0.6, 0.2 \rangle$, by exchanging belief with disbelief (see complement operator in Table 1).

A slAAF represents the set of all Dung's classical frameworks that can potentially be created from it. Similarly to [19], we call this creation process the inducement of an AAF from a slAAF. All arguments and attacks with a probability expectation of 1 will be found in the induced AAF, which can also contain additional arguments and attacks, as specified in Definition 5.

Definition 5 (Inducing an AAF from a slAAF). A Dung's framework $AAF = \langle A, R \rangle$ is said to be induced from a slAAF $= \langle \mathcal{A}, \mathcal{R}, \mathcal{O}_{\mathcal{A}}, \mathcal{O}_{\mathcal{R}} \rangle$ iff the remainder holds:

- $A \subseteq \mathcal{A}$,
- $R \subseteq (\mathcal{R} \cap (A \times A))$,
- $\forall a \in \mathcal{A}$ such that $\omega_a = \langle 1, 0, 0 \rangle$, then $a \in A$,
- $\forall (a_i, a_j) \in \mathcal{R}$ such that $\omega_{(a_i, a_j)} = \langle 1, 0, 0 \rangle$ and $\omega_{a_i} = \omega_{a_j} = \langle 1, 0, 0 \rangle$, then $(a_i, a_j) \in R$.

Moreover, we write $\mathbb{I}(slAAF)$ to represent the set of all AAFs that can be induced from a slAAF.

Given Definition 5, an AAF induced from a slAAF contains a subset of the arguments found in the source slAAF, together with a subset of attacks in the slAAF, subject to these defeats containing only arguments found within the induced AAF.

In practice, the process described in Definition 5 splits the uncertainty expresses in a slAAF into constellations (see Section 5) of different possible worlds, each with a different probability. For instance, given the slAAF in Figure 2 and Table 2, then $\mathbb{I}(slAAF)$ is equivalent to the following set of four derived frameworks:

$$\begin{aligned} F_1 &= \langle \{a, c\}, \emptyset \rangle & F_2 &= \langle \{a, b, c\}, \{(b, c)\} \rangle \\ F_3 &= \langle \{a, c, d\}, \{(c, d)\} \rangle & F_4 &= \langle \{a, b, c, d\}, \{(b, c), (c, d)\} \rangle \end{aligned}$$

This allows us to compute the expectation of some AAF being induced from a slAAF. Informally, such expectation value can be computed via the joint expectations of the arguments and attack relations appearing in the considered slAAF. In order to formalise such a concept compactly, we first need to identify the set of attacks that may appear in an induced AAF, as accomplished in [19]. We call this set R_A :

$$R_{\mathcal{A}} = \{(a_i, a_j) \mid a_i, a_j \in A \text{ and } (a_i, a_j) \in \mathcal{R}\}$$

Hence, it is possible to compute the expectation of some AAF being induced from a slAAF, as defined in Definition 6. The expectations E_{a_i} and $E_{(a_i, a_j)}$ are computed from the opinions returned by $\mathcal{O}_{\mathcal{A}}(a_i)$ and $\mathcal{O}_{\mathcal{R}}(a_i, a_j)$ respectively, for each $a_i \in \mathcal{A}$ and $(a_i, a_j) \in \mathcal{R}$. As a remainder from Section 2, expectations for binomial opinion is given by $b + au$ from (b, d, u) , with $a = 0.5$ (see Eq. 7 in Section 2).

Definition 6 (Expectation of an induced AAF). *Given $slAAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{O}_{\mathcal{A}}, \mathcal{O}_{\mathcal{R}} \rangle$, the expectation of $F = \langle A, R \rangle \in \mathbb{I}(slAAF)$ can be computed as in Eq. 8:*

$$E_F^{\mathbb{I}} = \prod_{a_i \in A} E_{a_i} \prod_{a_i \in (\mathcal{A} \setminus A)} (1 - E_{a_i}) \prod_{(a_i, a_j) \in R} E_{(a_i, a_j)} \prod_{(a_i, a_j) \in (R_{\mathcal{A}} \setminus R)} (1 - E_{(a_i, a_j)}) \quad (8)$$

We can then list the expectation value for all the four induced AAFs: $E_{F_1}^{\mathbb{I}} = 0.105$, $E_{F_2}^{\mathbb{I}} = 0.245$, $E_{F_3}^{\mathbb{I}} = 0.195$, $E_{F_4}^{\mathbb{I}} = 0.455$. For example, $E_{F_1}^{\mathbb{I}} = (1 \times 1) \times ((1 - 0.7) \times (1 - 0.65)) = 0.105$; no attack is considered in the computation because $F_1 = \langle \{a, c\}, \emptyset \rangle$.

Hence, also the semantics change over these two AAFs: in F_1 the set $\{a, c\}$ satisfies the grounded and stable semantics (no attack is present), while F_2 returns different extensions: $stb(F_2) = \{\{a, b\}\}$, and $grd(F_2) = \{a, b\}$.

Similarly to [19], we can derive the following property:

Proposition 1. *The sum of all the expectation values of all the AAFs that can be induced from a slAAF is 1:*

$$\sum_{F_i \in \mathbb{I}(slAAF)} E_{F_i}^{\mathbb{I}} = 1$$

The proof simply derives from exhaustively considering all the possible worlds; in our running example, $0.105 + 0.245 + 0.195 + 0.455 = 1$.

We can now define the expectation of some set of arguments satisfying one of the semantics σ in the literature, for example the properties introduced in Definition 2, i.e. $\sigma \in \{com, prf, stb, grd\}$ (notice that other semantics have been successively in the literature [2, Ch. 2]). For this reason, we define a function $v : (\sigma, B, F) \rightarrow \{false, true\}$ that returns *true* if and only if the set of arguments B represents one of the extensions satisfying σ given a framework F : that is, $v(\sigma, B, F)$ is *true* if and only if $B \in \sigma(F)$, *false* otherwise.

Definition 7 (Semantics expectation). *Given a sIAAF $\langle \mathcal{A}, \mathcal{R}, \mathcal{O}_{\mathcal{A}}, \mathcal{O}_{\mathcal{R}} \rangle$, the expectation that a given set of arguments $B \in \mathcal{A}$ satisfies a semantics σ is:*

$$E_{\sigma}^{\mathbb{I}}(B, sIAAF) = \sum_{F_i \in \mathbb{I}(sIAAF)} E_{F_i}^{\mathbb{I}} \quad \text{where } v(\sigma, B, F_i) = true \quad (9)$$

For instance, the expectation $E_{grd}^{\mathbb{I}}(\{a, c\}, sIAAF) = 0.105 + 0.195 = 0.3$: the set $\{a, c\}$ represents a grounded extension in F_1 and F_3 , whose expectation is respectively 0.105 and 0.195. $E_{stb}^{\mathbb{I}}(\{a, b, d\}, sIAAF) = 0.455$ since the set $\{a, b, d\}$ is a stable extension only in F_4 , whose expectation is 0.455.

In the same way, we can compute the expectation of acceptance of an argument w.r.t. $\mathbb{I}(sIAAF)$ and σ : the same argument can be justified/defensible/overruled (i.e., j/d/o, see Definition 3) in multiple generated worlds. We take advantage of a function $z : (\sigma, acpt, a_i, F) \rightarrow \{false, true\}$, which returns *true* if argument a_i is accepted as requested ($acpt \in \{j, d, o\}$) in F , given a semantics σ .

Definition 8 (Acceptance expectation). *Given a sIAAF $\langle \mathcal{A}, \mathcal{R}, \mathcal{O}_{\mathcal{A}}, \mathcal{O}_{\mathcal{R}} \rangle$, the expectation that an argument $a \in \mathcal{A}$ is justified/defensible/overruled ($acpt \in \{j, d, o\}$) w.r.t. semantics σ is:*

$$E_{\sigma, acpt}^{\mathbb{I}}(a, sIAAF) = \sum_{F_i \in \mathbb{I}(sIAAF)} E_{F_i}^{\mathbb{I}} \quad \text{where } z(\sigma, acpt, a_i, F_i) = true \quad (10)$$

For instance, the expectation $E_{adm, j}^{\mathbb{I}}(a, sIAAF) = 1$ (argument a is accepted in F_1, F_2, F_3 , and F_4), while $E_{adm, d}^{\mathbb{I}}(c, sIAAF) = 0.105 + 0.195 = 0.3$: argument c is accepted in F_1 (expectation 0.105) and F_3 (expectation 0.195).

Note that generating all the possible worlds in the constellations and then enumerate all the extensions for each of them can lead to computational issues: the number of worlds exponentially grows in the size of the considered sIAAF. Even if the state of the art of argumentation solvers is quite advanced [4], the exact expectation value need to be approximated [19].

4 From Trust Between Agents to Belief in Arguments

The work in [15] describes a method for trust network analysis using subjective logic (TNA-SL). It provides a simple notation for expressing transitive trust relationships, and

defines a method for simplifying complex trust networks so that they can be expressed in a concise form and be computationally analysed. Trust measures are expressed as beliefs, and Subjective Logic operators are used to compute trust between arbitrary parties in the network.

In this section we outline a computational framework where we use TNA-SL to compute the reputation in the attacks and arguments uttered in a public (for a community) debate, for which we suppose not all of the agents in that community have had the opportunity to attend to. Alternatively, some agents could have attended but have not been able to form an opinion because of impediments as, for instance, cultural differences in the audience, the education level, or cognitive limitations in general [25]. Hence, some of the agents form their *derived* opinion from friends and acquaintances by using trust relationships (derived opinions represent recommendations). At the same time, also agents who have a *direct* opinion are influenced by other parties they know [6]. The proposed approach follows these steps:

1. direct and indirect opinions of the same agent can be aggregated in order to produce a single belief for the same argument or attack, and
2. aggregated opinions of single agents can be further aggregated with the purpose to produce a reputation for an argument or attack, which reflects the belief of all the considered community;
3. finally, the obtained slAAF, where each argument and attack is weighed with an opinion as derived from items 1 and 2, can be studied by using the constellations approach proposed in Section 3.

We first define Trust Network-based slAAFs.

Definition 9 (Trust Network-based slAAF). A *Trust Network-based slAAF* (abbreviated to TN-slAAF) is formed by a slAAF $\langle \mathcal{A}, \mathcal{R}, \mathcal{O}_{\mathcal{A}}, \mathcal{O}_{\mathcal{R}} \rangle$ (see Definition 4), and a Trust Network represented as $\langle P, T \rangle$, where P is the set of agents (we require $\mathcal{A} \cap P = \emptyset$) and T is a binary trust relation on P : $\forall p_i, p_j \in P, T(p_i, p_j)$ means that p_i trusts p_j (T is asymmetric). Moreover, there is a further binary relation N of direct and derived binomial opinions (see Sec. 2), where each element $(p, x) \in N$ relates an agent $p \in P$ with $x \in \mathcal{A}$ or $x \in \mathcal{R}$.

The next definition is used to describe trust paths in a TN-slAAF.

Definition 10 (Trust path in TN-slAAF). Given a TN-slAAF, a *trust path* is always rooted in $p \in P$ and either ends in $a \in \mathcal{A}$ or $(a, b) \in \mathcal{R}$.

In the remainder we will use capital letters A, B, \dots for agent names in P , and lowercase letters for arguments (i.e., a, b, \dots). Moreover, when describing trust paths, the symbol “:” will be used to denote the transitive connection of two consecutive trust arcs to form a transitive trust path. The “ \diamond ” symbol visually resembles a simple graph of two parallel paths between a pair of agents [15]. With no restrictions on the possible trust arcs, trust paths from a given source X to a given target y can contain cycles, which could result in inconsistent calculative results. Cycles in the trust graph must therefore be controlled when applying calculative methods to derive measures of trust between

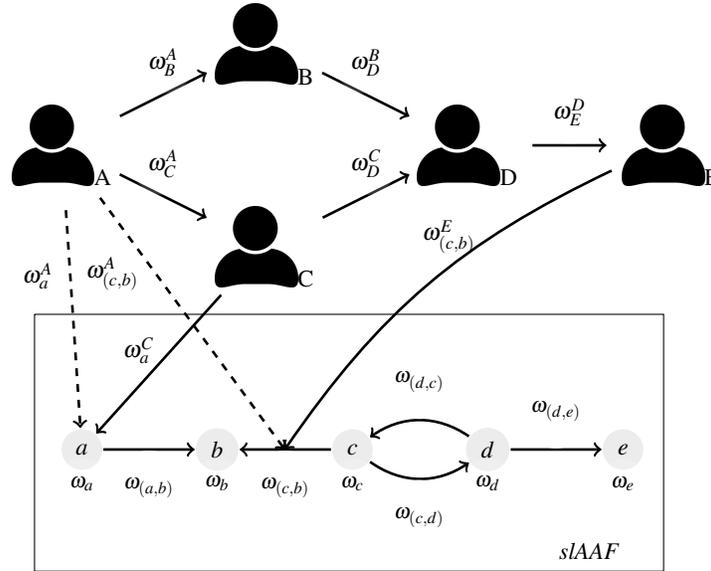


Fig. 3. An example of Trust Network-based sIAAF, with a community of agents interacting on it through trust paths. Dashed edges represent *derived* opinions.

two parties. *Normalisation* and *simplification* are two different control approaches, and the trust model presented in this paper can take advantage from these techniques as introduced in [15]. For the sake of brevity, we point the reader to [15] for a more exhaustive explanation.

An example of TN-sIAAF is reported in Figure 3: the upper part of the figure represents all the agents in a community, while the lower part shows the considered debate in the form of a sIAAF as described in Section 3. A Trust Network is thus tied to an AAF, and the opinions related to arguments and attacks are represented and aggregated in SL. Arguments are detailed in Example 1.

Example 1. We detail the sIAAF arguments in Figure 3, taking into consideration a discussion in favour/against the legalisation of Marijuana. *A, B, C, D, E* in Figure 3 are the audience of a debate concerning this topic.

- *a*: Official report from rating agencies say the financial crisis dramatically impacted on the overall financial budget.
- *b*: The budget allocated to healthcare for light drugs needs to be increased, because statistics say the number of light drugs users suffering from effects is increasing and treatments are expensive.
- *c*: Marijuana should not be legalised because it would rise healthcare expenses.
- *d*: Marijuana should be legalised because prisons are overcrowded and a large part of prisoners is in custody because they are marijuana users.

Table 5. Some opinions for the TN-slAAF in Figure 3.

ω_B^A	ω_C^A	ω_D^B	ω_D^C	ω_E^D	ω_a^C	$\omega_{(c,b)}^E$
$\langle 0.9, 0, 0.1 \rangle$	$\langle 0.3, 0, 0.7 \rangle$	$\langle 0.3, 0.5, 0.2 \rangle$	$\langle 0.5, 0.1, 0.4 \rangle$			

- e : To overcome prison overcrowding, most of the people think new prisons need to be built. Then, we should do that.

In this example, we focus on agent A in Figure 3 (the complete network can be more complex than Figure 3), and we derive an indirect opinion towards argument a along the trust path $[A, C] : [C, a]$, and an indirect opinion towards attack (c, b) along the path

$$([A, E]) = ([A, B] : [B, D]) \diamond ([A, C] : [C, D]) : [D, E] : [E, (c, b)].$$

The *discounting* [16] operator $\omega_x^{A:B} = \omega_B^A \otimes \omega_x^B$ in Table 1 can be used to compute transitive trust along a trust path, while the *consensus* [16] operator $\omega_x^{A \circ B} = \omega_x^A \oplus \omega_x^B$ in Table 1 can be used to fuse two beliefs into one, thus composing parallel paths together. Formally, $\omega_x^{A:B} = \langle b_B^A b_x^B, d_B^A d_x^B, d_B^A + u_B^A + b_B^A u_x^B \rangle$, and $\omega_x^{A \circ B} = \langle b_B^A b_x^B, d_B^A d_x^B, d_B^A + u_B^A + b_B^A u_x^B \rangle$. With the *cumulative fusion* operator, i.e. \otimes , the observations are supposed as independent; the cumulative rule is equivalent to a posteriori updating of Dirichlet distributions.⁵

The effect of discounting in a transitive path is to increase uncertainty, that is to reduce the confidence in the expectation value. The effect of the consensus operator is to reduce uncertainty, that is to increase the confidence in the expectation value. Then, we can compute ω_a^A and $\omega_{(c,b)}^A$ as

$$\omega_a^A = \omega_C^A \otimes \omega_a^C$$

$$\omega_{(c,b)}^A = ((\omega_B^A \otimes \omega_D^B) \oplus (\omega_C^A \otimes \omega_D^C)) \otimes \omega_E^D \otimes \omega_{(c,b)}^E.$$

Given the beliefs in Table 5, we can compute $\omega_a^A = \langle 0.28, 0.48, 0.24 \rangle$ and $\omega_{(c,b)}^A = \langle 0.44, 0.09, 0.48 \rangle$. Finally, these beliefs can be assigned to dashed edges in Figure 3.

As previously advance, we can use SL to aggregate direct and derived beliefs of the same agent, and to aggregate beliefs of different agents. This is visually described in two small TN-slAAF examples, respectively in Figure 4 and in Figure 5. In Figure 4, agent A can aggregate its direct opinion with two derived ones, which are obtained by two trust paths (not shown in the figure) as previously introduced in this section: hence, $\omega_a^A = \tilde{\omega}_a^A \oplus \tilde{\omega}_a^A \oplus \hat{\omega}_a^A$. In Figure 5 the reputation of argument a can be computed via the consensus operator, i.e., $\omega_a = \omega_a^A \oplus \omega_a^B \oplus \omega_a^C$; ω_a can be then directly used in the computational framework presented in Section 3.

5 Related Work

We revise the literature about *probabilistic argumentation* (i.e., constellations and epistemic approach), and also from the point of view of trust sources and systems.

⁵ More details on this operator, and how to compute it, can be found in [16].

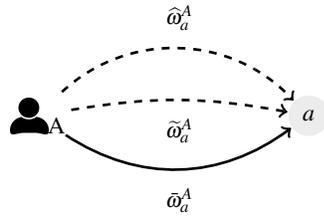


Fig. 4. An example of one direct and two derived opinions (through one trust path each).

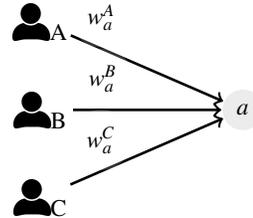


Fig. 5. An example of multiple opinions from different agents to be aggregated and reach a final reputation for a .

Constellations. In the constellations approach, uncertainty in the topology of the graph (probabilities on arguments and attacks) is used to make probabilistic assessments on the acceptance of arguments. The authors of [11] provided the first proposal to extend abstract argumentation with a probability distribution over sets of arguments which they use with a version of assumption-based argumentation in which a subset of the rules are probabilistic rules. In [19] a probability distribution over the sub-graphs of the argument graph is introduced, and this can then be used to give a probability assignment for a set of arguments being an admissible set or extension of the argument graph. In [7] the authors characterise the different semantics from the approach of [19] in terms of probabilistic logic with the purpose of providing an uniform logical formalisation and also pave the way for future implementations.

Epistemic. In the epistemic approach instead, the topology of the graph is fixed but probabilistic assessments on the acceptance of arguments are evaluated w.r.t. the relations of the arguments in the graph. For instance, in [3] the authors cast epistemic probabilities in the context of de Finetti's theory of subjective probability, and they analyse and revise the relevant rationality properties in relation with de Finetti's notion of coherence. However, most of the work in this directions is authored by M. Thimm [26] and A. Hunter [14]. In the first work, the authors proposes a probabilistic semantics for Abstract Argumentation in order to assign probabilities or degrees of belief to individual arguments. The presented semantics generalise the classical notions of semantics [10]. In the second work, the author starts from considering logic-based argumentation with uncertain arguments, but ends showing how this formalisation relates to uncertainty of abstract arguments. The two authors join their efforts in [13].

Trust and Argumentation. Trust and Argumentation are two strictly related concepts, as the florid literature of the last years proves. In [22] the authors investigate the combination of trust measures on agents and the use of argumentation for reasoning about belief, thus combining an existing system for reasoning about trust and an existing system of argumentation. In [1] the authors study how the different arguments interact and how an agent may decide to trust another source and thus to accept information coming

from that source. The system also deals with graded trust (like agent i trusts to some extent agent j). Trust of sources is also studied in [27], together with a model to trust in a trustworthy way. In [21] the authors identify two types of argumentative relevance: internal relevance, i.e. the extent to which a premise has a bearing on its purported conclusion (thus considering *structured* arguments), and external relevance, i.e. a measure of how much a whole argument is pertinent to the matter under discussion. Two more works on Trust and Argumentation are [8] and [9].

6 Conclusion

The aim of this paper is to encompass Trust Network Analysis [15] and to irradiate the effect of direct and derived trust among the agents in a network towards a Subjective Logic-based AAF. Hence, entities can form their opinion by considering their direct belief, and the beliefs of parties through trust paths linking them together. In addition, all these subjective opinions can be fused into a reputation score related to each argument and attack. Such a score represents how much the studied community of agents evaluates their credibility. Finally, the resulting slAAF can be studied using the constellations approach as in related works [19].

In the future, we would like to extend this study along two different lines. The first one concerns the argumentation side of our proposal: for instance, we are interested in deal with slAAFs from the point of view of the epistemic approach (see Section 5). The second line concerns the trust analysis of the network among agents. Future goals are to enrich the framework by taking into consideration ageing factors: agents (and in particular human agents) may change their behaviour over time, so it is desirable to give greater weight to more recent ratings using longevity factors. In addition, we would like to enrich the picture with distrust besides trust, also by exploring other computational frameworks as [5].

References

1. Amgoud, L., Demolombe, R.: An argumentation-based approach for reasoning about trust in information sources. *Argument & Computation* 5(2-3), 191–215 (2014)
2. Arrow, K.J., Sen, A., Suzumura, K.: *Handbook of Social Choice and Welfare*. North-Holland (2002)
3. Baroni, P., Giacomin, M., Vicig, P.: On rationality conditions for epistemic probabilities in abstract argumentation. In: *Computational Models of Argument - Proceedings of COMMA. FAIA*, vol. 266, pp. 121–132. IOS Press (2014)
4. Bistarelli, S., Ross, F., Santini, F.: Not only size, but also shape counts: abstract argumentation solvers are benchmark-sensitive. *J. Log. Comput.* 28(1), 85–117 (2018)
5. Bistarelli, S., Santini, F.: On merging two trust-networks in one with bipolar preferences. *Mathematical Structures in Computer Science* 27(2), 215–233 (2017)
6. Campbell-Meiklejohn, D.K., Bach, D.R., Roepstorff, A., Dolan, R.J., Frith, C.D.: How the opinion of others affects our valuation of objects. *Current Biology* 20(13), 1165–1170 (2010)
7. Doder, D., Woltran, S.: Probabilistic argumentation frameworks - A logical approach. In: *Scalable Uncertainty Management - 8th International Conference, SUM. Lecture Notes in Computer Science*, vol. 8720, pp. 134–147. Springer (2014)

8. Dondio, P., Longo, L.: Trust-based techniques for collective intelligence in social search systems. In: *Next Generation Data Technologies for Collective Computational Intelligence, Studies in Computational Intelligence*, vol. 352, pp. 113–135. Springer (2011)
9. Dondio, P., Longo, L.: Computing trust as a form of presumptive reasoning. In: *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*. pp. 274–281. IEEE Computer Society (2014)
10. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence* 77(2), 321–357 (1995)
11. Dung, P.M., Thang, P.M.: Towards (probabilistic) argumentation for jury-based dispute resolution. In: *Computational Models of Argument: Proceedings of COMMA, FAIA*, vol. 216, pp. 171–182. IOS Press (2010)
12. Hankin, R.K.: A Generalization of the Dirichlet Distribution. *Journal of Statistical Software* 33(11), 1–18 (February 2010)
13. Hunter, A., Thimm, M.: Probabilistic reasoning with abstract argumentation frameworks. *J. Artif. Intell. Res.* 59, 565–611 (2017)
14. Hunter, A.: A probabilistic approach to modelling uncertain logical arguments. *Int. J. Approx. Reasoning* 54(1), 47–81 (2013)
15. Jøsang, A., Hayward, R., Pope, S.: Trust network analysis with subjective logic. In: *Computer Science 2006, Twenty-Ninth Australasian Computer Science Conference (ACSC), CRPIT*, vol. 48, pp. 85–94. Australian Computer Society (2006)
16. Jøsang, A.: *Subjective Logic - A Formalism for Reasoning Under Uncertainty*. *Artificial Intelligence: Foundations, Theory, and Algorithms*, Springer (2016)
17. Jøsang, A., Bhuiyan, T.: Optimal trust network analysis with subjective logic. In: *Proceedings of the Second International Conference on Emerging Security Information, Systems and Technologies, SECURWARE*. pp. 179–184. IEEE Computer Society (2008)
18. Jøsang, A., Hankin, R.K.: Interpretation and Fusion of Hyper Opinions in Subjective Logic. In: *Proceedings of the 15th International Conference on Information Fusion (FUSION 2012)* (2012)
19. Li, H., Oren, N., Norman, T.J.: Probabilistic argumentation frameworks. In: *Theorie and Applications of Formal Argumentation - First International Workshop, TAFE, LNCS*, vol. 7132, pp. 1–16. Springer (2011)
20. Oren, N., Norman, T.J., Preece, A.D.: Subjective logic and arguing with evidence. *Artif. Intell.* 171(10-15), 838–854 (2007)
21. Paglieri, F., Castelfranchi, C.: Trust, relevance, and arguments. *Argument & Computation* 5(2-3), 216–236 (2014)
22. Parsons, S., T., Y., Sklar, E., McBurney, P., Cai, K.: Argumentation-based reasoning in agents with varying degrees of trust. In: *10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. pp. 879–886. IFAAMAS (2011)
23. Prakken, H., Vreeswijk, G.: Logics for defeasible argumentation. In: *Handbook of philosophical logic*, pp. 219–318. Springer (2002)
24. Santini, F., Jøsang, A., Pini, M.S.: Are my arguments trustworthy? abstract argumentation with subjective logic. In: *21st International Conference on Information Fusion, FUSION*. pp. 1982–1989. IEEE (2018)
25. Tajfel, H.: Social and cultural factors in perception. *Handbook of social psychology* 3, 315–394 (1969)
26. Thimm, M.: A probabilistic semantics for abstract argumentation. In: *ECAI - 20th European Conference on Artificial Intelligence, FAIA*, vol. 242, pp. 750–755. IOS Press (2012)
27. Villata, S., Boella, G., Gabbay, D.M., van der Torre, L.W.N.: A socio-cognitive model of trust using argumentation theory. *Int. J. Approx. Reasoning* 54(4), 541–559 (2013)