Social Argumentation Systems

Antonis Bikakis
Department of Information Studies, University College London, UK
a.bikakis@ucl.ac.uk

Giorgos Flouris
Institute of Computer Science, FORTH, Heraklion, Greece.
fgeo@ics.forth.gr

Joao Leite
NOVA LINCS, NOVA University Lisbon, Portugal
jleite@fct.unl.pt

Theodore Patkos
Institute of Computer Science, FORTH, Heraklion, Greece.
patkos@ics.forth.gr

Abstract

While a lot of research has been conducted on understanding and formalising the interplay of arguments within the context of Computational Argumentation, this research is not fully applicable to the types of arguments that populate the Social Web. In that context, arguments usually have the form of comments, opinions or reviews, and are the main ingredients of online discussion forums, social networks, online rating and review sites, debate portals and other online communities - the electronic version of word-of-mouth communication. As a result, voting and other forms of reaction to the provided comments or arguments (other than just "attacks") are allowed, features that are not normally considered in the classical literature on Computational Argumentation. In this chapter, we study extensions of argumentation frameworks that have been proposed to describe and understand the more complex types of interactions among arguments that can be found in the Social Web, and present the current state-of-the-art, as well as open problems.

Vol. 12 No. 3 2025
Journal of Applied Logics — IfCoLog Journal of Logics and their Applications

1 Introduction

The use of the Web is constantly evolving. Although users were originally expected to be merely consumers of Web information, in recent years we experienced a proliferation of portals and online systems allowing users to become also producers of information. As a result, the social aspect of the Web has been flourishing, allowing users to post opinions, comments and reviews populating a wide range of online systems, from social media and online discussion forums to news sites and product review sites [15]. The impact of this user-generated information is clearly evident, especially on consumer behaviour and businesses [30]. Due to the vast number of comments that users need to search through to locate the most helpful ones, their proper ranking, filtering and recommendation become critical functionalities. Apart from the open-ended discussions on the Web, more goal-oriented debates are also becoming popular, e.g., in debate portals for active citizenship[1], or in decision support systems, such as issue-based information systems (IBIS) [28], [10], [14].

In this setting, it is not enough for applications to provide functionality for opinion or argument exchange. In order to help users reach informed, well-justified and sensible conclusions or decisions, such applications also share the need to *evaluate arguments* based on quantitative methods. Towards this goal, various methods have been used to rate user arguments or comments, which vary from voting mechanisms, such as like/dislike counters, and expert ratings, to combinations of these with user responses in the form of counter (attacking) or follow-up (supporting) arguments. Methodologies from computational argumentation have been proposed as a powerful tool for a more accurate evaluation of an argument's acceptance, and a number of formal frameworks have emerged that properly adapt argumentation algorithms to the needs of the Social Web (e.g., [29], [37], [24]) or decision support systems (e.g., [14], [41]).

This chapter focuses mainly on argumentation frameworks that treat two different types of reaction to an argument: verbal responses, which are commonly modelled as arguments that are related to the original one (e.g., via an attack or support relationship); and (positive or negative) votes, which express someone's approval or disapproval of the argument. We call such frameworks *social argumentation systems*. Their main aim is to provide a way to evaluate the social acceptance of an argument (which in most of these frameworks is referred to as *strength*, or *score*) taking into account the responses and the votes it has received and the strengths of such responses. Also, in many cases, the computed strength is also affected by a *base score*, which reflects the strength of an argument before any reactions are

[1] https://www.kialo.com, https://www.kialo-edu.com, https://www.createdebate.com

considered. The base score may reflect different intuitions, such as an a priori assessment by experts, argument popularity, or other features that are supported by the underlying application. The strength of an argument, as computed by a social argumentation system, usually quantifies the degree of acceptance of the argument, although some social argumentation systems may also evaluate other aspects of arguments, such as their acceptability, quality, relevance, objectivity and others (see, for example, Section 4 in this chapter and the related papers [37], [35]).

Note that, under this viewpoint, the argument evaluation process in a social argumentation system differs from the one in the standard Abstract Argumentation Frameworks (AAFs) of Dung [22] and most of their extensions. In AAFs and similar frameworks, an argument is either in the extension, or not. On the other hand, social argumentation systems employ *gradual semantics*, i.e., the strength of an argument is expressed in terms of a numerical value, enabling a more fine-grained evaluation compared to the two- or three-valued acceptability semantics of most argumentation frameworks. Having said that, one can view extensions as a special case of a numerical assignment, where the only values allowed are 0 (not in the extension) and 1 (in the extension).

The aim of this chapter is *to provide an overview of applications of formal argumentation to the Social Web, in the form of Social Argumentation Systems.* Towards this, we start by presenting a number of relevant principles for such systems (Section 2). Then, we present two examples of social argumentation systems, namely Social Argumentation Frameworks (Section 3) and s-mDiCE (Section 4). We then briefly survey other social argumentation systems, including some examples of extensions or applications of existing argumentation frameworks designed or used to model other aspects of discussions in social media, such as the semantic relations among posts, the social relevance of a post, the influence of users and multi-topic discussions (Section 5). We conclude in Section 6.

2 Principles and properties

As will become obvious by our presentation in the following sections of this chapter, methodologies for computing different notions of argument strength in a social context abound. Thus, a question naturally arises: which one is best? The answer, of course, depends on the needs of the application, but even so, how can we evaluate and compare these methodologies in the context of any given application?

To address this question, many authors have proposed several different *principles*, i.e., logical constraints, or axioms, that formalise certain intuitive properties that a "good" social argumentation system should satisfy. In this section, we study

some of these principles, with emphasis on the ones that are most relevant to social argumentation.

2.1 The principle-based approach in argumentation

The *principle-based approach* in argumentation, sometimes called the *axiomatic approach* [43], is a methodology for developing principles for assessing the "quality" or "relevance" of different semantics for specific applications. Importantly, such principles are not necessarily ubiquitous or indisputable, and different principles may be relevant to different applications: as a matter of fact, for any given application, some principles may be desirable, others irrelevant, and others even undesirable. But this is exactly the strength of the principle-based approach: by choosing the principles that are (un)desirable for a certain application, the designer of the application can immediately identify the semantics that are (in)appropriate for the application at hand, and therefore make an informed choice.

Thus, the principle-based approach can be viewed as a methodology for choosing the most appropriate semantics to use for a particular application. Also, it has been argued that this approach provides a systematic way of viewing semantics and their properties, guiding the search for novel interesting argumentation semantics [43], and allowing the identification and definition of new relevant principles [12].

The principle-based approach has been used for many different types of argumentation frameworks. As expected, the bulk of the related work deals with the development of principles for the standard Abstract Argumentation Frameworks of Dung [22], and, thus, are not entirely relevant to the gradual argumentation semantics that interest us here. We start by briefly presenting some of the studies that adopt the principle-based approach in settings other than gradual argumentation semantics, before analysing in more detail the principles that are relevant to gradual argumentation semantics.

2.2 Proposed principles

In [11], the authors present the principle-based approach and apply it to abstract argumentation. The work is comprehensive, evaluating 13 different principles against 11 different semantics. A very similar discussion appears in [43], which evaluates 8 principles against 15 proposed semantics. Note that there is some overlap in the considered principles and semantics (and thus in some of the results as well) between these two studies. Importantly, both approaches apply to the standard semantics of abstract argumentation frameworks, and not for gradual argumentation, and thus their relevance to the present chapter is limited.

A similar work appears in [49] for abstract agent argumentation frameworks, i.e., frameworks that extend Dung's AAFs with agents. The idea in this setting is that each argument is associated with certain agents, and this association may affect the semantics of the framework and the extensions that it has. Obviously, it also affects the potentially desirable principles that such semantics should satisfy. In that work, the authors examine 52 agent semantics and 17 principles under this setting, all of them considering the standard AAF semantics.

A work that is more closely related to this chapter is [16], which examines the principle-based approach for *ranking-based semantics*, in which the argument evaluation process results to a ranking determining whether an argument is "more acceptable" than another. That work evaluates 7 ranking-based semantics against 18 principles.

In [19], the author examines a set of properties related to how argumentation frameworks (and their associated semantics) behave when used for reasoning. Thus, the principles presented in this paper apply to the output of the argumentation process (i.e., the conclusions), rather than the (accepted) arguments themselves.

In [26], the authors propose a set of principles for bipolar argumentation. Their approach is based on the use of labels that are defined in an abstract manner based on an appropriate algebra. Depending on how the labels and the respective algebra are defined, they could be used to represent different things, including the "strength" of an argument, under various different notions of what "strength" may mean. Thus, this work can be considered to fall within the scope of gradual and/or social argumentation. In fact, the authors of [26] explicitly mention the possible application of this work to social platforms.

Other studies define principles that are specific to the formalism at hand. For example, [25] propose 4 principles, most of which are adaptations of AAF principles for the formalism proposed in that paper (Abstract Argumentation Frameworks with Domain Assignments – AAFDs), and show which AAFD semantics satisfy such principles.

More in relation to the present chapter, numerous principles for gradual and social argumentation have appeared in various papers [26], [5], [6], [16], [31], [41], [3], [7], [14], [29], [42], [4], [8]. Although these principles often use different notation and terminology, they capture similar intuitions. Still, their comparison is difficult. This problem was identified and addressed by [13], resulting in an organisation of the principles that will be the focus of the following subsection.

2.3 Organising the gradual argumentation principles

Perhaps the most prominent work presenting principles for gradual argumentation appeared in a series of papers [39], [12], [13], where the authors presented an effort to organise the numerous principles proposed in other studies under a single unifying umbrella of flexible definitions. The focus of this chapter on the work of [13] is justified by the fact that it essentially unifies many previously-expressed principles in other papers (such as those described above).

In particular, [13] collected several principles that have appeared in the literature, noticing that they have common conceptual roots and are based on a small set of common patterns. Then, they formalised these patterns into 11 *group properties* (GPs), which are essentially generic principles that correspond to most of the principles that have appeared in previous work. Further, these 11 GPs can be viewed as instantiations of 4 novel parametric principles.

The above organisation of principles has many benefits. First, it provides a systematic way to view principles relevant to gradual argumentation, thereby revealing the existence of novel principles that have so far not been studied in the literature. Secondly, it provides a simplifying and unifying terminology to study and discuss the relevant principles, ending the polyphony that hindered the comparison of principles appearing in previous papers. In other words, it provided a unifying substrate with the use of which most of the principles in the gradual argumentation literature can be expressed, compared against, and studied.

From a formal perspective, the authors of [13] base their analysis on an argumentation model called *Quantitative Bipolar Argumentation Framework (QBAF)*, which is a tuple $(X, \mathcal{R}^-, \mathcal{R}^+, \tau)$, where X is a finite set of arguments, $\mathcal{R}^-, \mathcal{R}^+$ is the attack and support relation (respectively) between arguments, and $\tau : X \mapsto \mathbb{I}$ is a total function mapping each argument with a *base score* in $\mathbb{I}$, where $\mathbb{I}$ is a set equipped with a preorder (usually $\mathbb{I} = [0, 1]$, i.e., a real number in the $0 \dots 1$ range, but other options are also possible). The base score corresponds to the argument's evaluation before considering its relationships (attack, support) with other arguments, and is a common feature in social argumentation systems. In [13], the use of τ is overloaded, and exploited as a convenient abstraction to hide a possibly complex method of computing the base score, based, e.g., on experts' assessment, argument popularity, votes, various types of non-verbal reactions on arguments, or other features that are supported by the underlying application. Note also that either $\mathcal{R}^-$ or $\mathcal{R}^+$ could be empty, leading to special cases of QBAFs where only support (or only attack, respectively) relations are allowed.

The *strength* (or *score*) of an argument is computed using another function, $\sigma : X \mapsto \mathbb{I}$. The function σ corresponds to the final arguments' assessment, after taking

into account both the base score and all relationships (attack, support) between arguments. The principles associated with gradual argumentation frameworks (i.e., QBAFs, in the terminology of [13]) are all meant to restrict the behaviour of σ in ways that make intuitive sense.

Using this formalisation, [13] presented 11 intuitive group properties, and their formal formulation as a GP, as well as 4 principles, which were shown to imply the GPs and thus provide a more general intuition behind them. These 4 principles are called *balance*, *strict balance*, *monotonicity*, and *strict monotonicity*. Balance and strict balance capture the intuition that any difference between the score of an argument ($\sigma(a)$) and its base score ($\tau(a)$) must correspond to some imbalance between the scores of its attackers and supporters. Monotonicity and strict monotonicity capture the intuition that each of the factors that affect an argument's score (base score, attackers, supporters) has a monotonic effect on the argument's score.

Further, the authors of [13] parameterise the QBAF model using the following five features:

- Whether it is required that $\mathcal{R}^- = \emptyset$ or $\mathcal{R}^+ = \emptyset$.
- The exact definition of $\mathbb{I}$ and its associated preorder $\leq$ (as well as its strict counterpart, $<$).
- A special relation $\ll$ between elements of $\mathbb{I}$, such that $< \subseteq \ll \subseteq \leq$.
- Whether $\mathbb{I}$ has a bottom element and whether arguments whose strength equals that bottom element are (or should be) considered in the evaluation or not.
- The definition of τ.

This parameterisation essentially allows different frameworks that have appeared in the literature to be recast as a QBAF. Equally importantly, it allows the parameterisation of principles in order to capture different intuitions.

A comprehensive evaluation of the principles that different argumentation frameworks and their semantics satisfy, appears in [13] (see Table 5 in that paper). In particular, the authors recast 19 different argumentation frameworks in the QBAF terminology, using the parameterisation options described above, and then showed which of the 4 main principles are satisfied by each of these 19 approaches. As a corollary (and combined with other results in the same paper), one can easily identify the GPs and other properties that are satisfied by each of these approaches.

2.4 Principles for dynamic frameworks

The aforementioned principles all deal with static frameworks, i.e., given a framework, they suggest how the scoring function σ should behave. An interesting exception is the principle of *smoothness*, which was informally defined in [35] and formally in the extended version of that paper [36]. Unlike the other principles, smoothness deals with how the scoring function behaves under *changes in the framework*. For this reason, it was not considered by [13], despite the generality and comprehensiveness of the principles presented there.

In particular, smoothness, as the name implies, guarantees that the scoring function of a gradual argumentation framework will behave "smoothly", i.e., small changes in the framework (e.g., a new relationship, or a small change in the base score of an argument) cannot have large effects on the overall evaluation of arguments. This is an essential feature for any rating framework, in order to be adopted by the public, as the effect of an action on the framework should be commensurate with the importance of the action itself; big leaps that are not justified by the underlying changes may seem counterintuitive to users, causing them to lose their trust in the objectivity of the rating algorithms.

The mathematical notion that is closest to the intuition presented above is the notion of the derivative of a function: in functions over real numbers, the derivative determines the rate at which the function changes at each point. However, for the considered setting this notion must be adapted to apply over more complex (non-continuous) domains, because changes in our case are not necessarily continuous (e.g., the addition of a new relationship between arguments). Thus, to define smoothness over arbitrary functions and sets, *semi-metrics* were used to determine the "rate" of change:

Definition 2.1. *Given a set S, a function $d_S : S \times S \mapsto \mathbb{R}$ is called a semi-metric for S iff for all $x, y \in S$: $d_S(x, y) \geq 0$, $d_S(x, y) = d_S(y, x)$ and $d_S(x, y) = 0$ iff $x = y$.*

Definition 2.2. *Consider two sets S, T equipped with semi-metrics d_S, d_T. A function $f : S \mapsto T$ is called ℓ-smooth (for d_S, d_T) iff $d_T(f(x), f(y)) \leq \ell \cdot d_S(x, y)$ for all $x, y \in S$.*

The value of ℓ in Definition 2.2 determines the "smoothness" of the function: a large ℓ implies that the function has at least some "abrupt" points, i.e., in our setting, that there are cases where simple (small) actions by the users would lead to major changes in the assessment result of the related arguments. On the other hand, small ℓ guarantees that a large number of changes are required to achieve a large effect on the assessment results, thus making the function more reluctant to

change. Given a function f, when there is no ℓ such that f is ℓ-smooth, we will say that f is ∞-smooth. Moreover, we will say that f is *exactly ℓ-smooth* when it is ℓ-smooth and there is no $\ell' < \ell$ such that f is ℓ'-smooth.

In the considered context, smoothness should be applied over the argument rating function (σ in the [13] terminology), to determine how quickly the ratings change when the framework changes. Note that the input to σ is an argument, therefore, to apply smoothness we technically need to also include the QBAF itself as an input to σ, i.e., define σ as $\sigma_F : X \mapsto \mathbb{I}$, where F is the QBAF under consideration. Now smoothness studies "how much" a change in F affects the output $\sigma_F(a)$ for $a \in X$. To apply smoothness here, one should use the semi-metric d_S to quantify the effect of each possible change in F (e.g., the addition/deletion of arguments, the addition/deletion of attack/support relationships, and the modification of one or more base scores), and d_T to quantify the effect on the actual score (given that $\mathbb{I}$ is usually a numerical domain, such as $[0, 1]$, a reasonable choice for d_T is the simple difference, i.e., $d_T(x, y) = |x - y|$).

Note that different applications may have different requirements regarding smoothness. As an example, using functions with high sensitivity to input (i.e., less smooth) will allow applications that attract few users to lay more emphasis on maintaining the liveness of discussions by having users' feedback cause reasonable, yet evident, effects in the course of the discussions. On the contrary, applications that lay emphasis on the reliability of the outcome of a debate, such as product/services rating sites, probably want to disallow small changes to significantly impact the outcome, in order to secure reliable results, therefore requiring functions that are less sensitive to user input. In this respect, smoothness is no different than other principles, in that it is application-dependent, and parameterisable (using ℓ, as well as the definitions of d_S, d_T). For applications of this principle, see Section 4.

3 Social argumentation frameworks

Justified by the fact that a growing percentage of users were giving up on the social web for lack of intellectually stimulating discussions, [29] argued that a Social Network should facilitate:

- More *open participation* where users with different levels of expertise are able to easily express their arguments, even without knowing formal argumentation and any formal rules of debate.

- More *flexible participation* where debates are not restricted to a pair of users arguing for antagonistic sides, but where there may be more than just two sides,

more users can propose arguments for each side, and each user is allowed to contribute with arguments for more than one side of the debate.

- More *detailed participation* where users are allowed to express their opinions by voting on individual arguments and on argument relations, instead of just on the overall debate's outcome.

- Appropriate *feedback* to users so that they can easily assess the strength of each argument, taking into account not only the logical consequences of the debate, but also the popular opinion and all its subjectiveness.

To that end, they envisioned a self-managing online debating system capable of accommodating two archetypal levels of participation.

On the one hand, experts, or enthusiasts, would be provided with simple mechanisms to specify their arguments and also a way to specify which arguments attack which other arguments. When engaging in a debate, users always propose arguments for specific purposes, like making a claim central to the issue being discussed, or defeating arguments supporting an opposing claim. Thus, the envisioned system should allow users to describe an abstract argument, capable of attacking other arguments, simultaneously with its natural language (or image, video, link, etc.) representation. Therefore, the formal specification of arguments and attacks becomes a natural by-product of the users' intent when proposing new arguments. To make this process as painless and easy as possible, and enable more people to participate, no particularly deep knowledge (such as logic) can be required. It is natural that a new argument might attack a previously proposed argument - indeed, that was likely the object of its creation. However, it is also possible that an older argument attacks the new argument as well. Therefore, the system should allow users to add this new attack relation to the system.

On the other hand, less expert users who prefer to take a more observational role, and do not wish to engage in proposing arguments or attacks, should also be accommodated in the system through a less complex participation scheme. These users may simply read the arguments in natural language (or image, video, link, etc.) and state whether they agree with them. This induces a voting mechanism similar to what is found in current social networks. There are alternatives, such as having argument's social trustworthiness be based on people's opinions of who proposed it. Voting on arguments seems to offer the path of least resistance for being the closest to current social networks. Additionally, it is apparent that not all attacks bear the same weight. Some attacks might have an obvious logical foundation (e.g., undercuts or rebuts), thus gaining trust from the more perceptive users. Other attacks might be less obvious or downright senseless, especially in open online contexts, making

users doubt or wish to discard them. Thus, extending the ability to vote to attacks becomes eminently desirable. Not only does voting on attacks more accurately represent a crowd's opinion in a variety of situations, but it also allows the system to self-regulate by letting troll-attacks be "down-voted" to irrelevance.

The system should also be able to autonomously and continuously provide an up to date view of the outcome of the debate, e.g., by assigning a value to each argument that somehow represents its social strength, taking the structure of the argumentation framework (arguments and attacks) and the votes into account. A nice GUI, e.g., depicting arguments with a size and/or colour proportional to these values would make the debate easier to follow, bringing forward relevant (socially) winning arguments, while downgrading unsound, unfounded (even troll) arguments. So that users may understand and follow a debate, small changes in the underlying argumentation framework and its social feedback (i.e., votes) should result in small changes to the formal outcome of the debate. If a single new vote entirely changes the outcome of a debate, users cannot gauge its evolution and trends, and are likely to lose interest.

In addition, any debating system as the one envisioned should also ensure, as argued in [29], that a few crucial properties be satisfied.

- *There should always be at least one solution to a debate.* From a purely logical standpoint, one may consider that some debates simply contain inconsistencies that make it impossible to assign them meaningful semantics. However, we are dealing with the Social Web, where inconsistency is the *norm*. If the system is incapable of providing solutions to every debate, then there is too much risk involved in using it. We believe that most of its users would prefer a system that would, nevertheless, provide them some valuation of the arguments that is somehow justifiable, instead of telling them that the debate is inconsistent.

- *There should always be at most one solution to a debate.* Logicians and mathematicians find it perfectly natural for there to be multiple, or even infinite, solutions to a given problem. However, in a social context as far-reaching as the Internet, it is disingenuous to assume that the general user-base, which likely covers a large portion of the educational spectrum, shares these views with the same ease. It is very hard for someone who has invested personal effort into a debate to accept that all arguments are in fact true (in a multitude of models)!

- *Argument outcomes should thus be represented very flexibly.* In particular, to accurately represent the opinions of thousands of voting users, arguments should be evaluated using degrees of acceptability, or gradual acceptability.

Two-valued or three-valued semantics risk grossly under-representing much of the user-base.

- *Formal arguments and attacks must be easy to specify.* For example, assuming knowledge of first-order logic for specifying structured arguments would alienate many potential users when the present goal is to include as many as possible. Moreover, simpler frameworks would make implementing and deploying such a system in different contexts (web forums, blogs, social networks, etc) much easier.

- *Argument strength should be limited by popular opinion, and every vote should count.* In a true social system, there should be no arguments of authority, nor votes without effect. Argument strength can be weaker than its direct support base, since arguments may be attacked by other arguments, but the direct opinion expressed by the votes should act as an upper bound on the strength of the argument. Also, positive (resp. negative) votes should increase (resp. decrease) the strength of the argument/attack on which they are cast (how much can depend on many factors).

- *Computing and updating debate outcomes should be highly efficient.* With the increasing speed of social interactions on the Web 2.0, users would grow impatient if new arguments and votes would not have an almost immediate effect on the debate outcome.

In the remainder of this section, we describe Social Argumentation Frameworks [29; 23], which can serve as the underlying formal backbone of an online debating system as the one described above.

Social Argumentation Frameworks use abstract arguments in the sense of Dung [22][2], but add the possibility of associating pro and con votes on both arguments and attacks, and a (family of) semantics that goes beyond the classical accepted/defeated valuations assigning each argument a value from an ordered set of values (e.g. the $[0, 1]$ real interval, a set of colours, textures, etc.). One particular semantics – based on the popular product T-norm and probabilistic sum T-co-norm and assigning arguments values from the $[0, 1]$ real interval – deserves particular attention because of its formal properties, namely guaranteeing the existence and uniqueness of a model, and also because of the existence of an algorithm that can effectively and efficiently compute the debate outcome. Social Argumentation Frameworks provide

[2] Abstract Argumentation[22] and Argumentation Theory in general grounds debates in solid logical foundations and has in fact been shown to be applicable in a multitude of real-life situations (c.f. [32]).

the theoretical foundations on which to build interaction tools that will provide more robust, flexible, pervasive and interesting social debates than those currently available.

In the following, we start by describing Social Argumentation Frameworks (SAF) and their semantics, which we subsequently illustrate with a very simple example. We then discuss some important formal properties of SAF. Finally, we present an efficient algorithm for computing debate outcomes.

3.1 Framework and semantics

We start by describing a Social Argumentation Framework. First introduced in [29] and later extended in [23], it is an extension of Dung's AAF, composed of arguments and an attack relation to which we add an assignment of votes to each argument and each attack.

Definition 3.1 (Social Argumentation Framework)**.** *A* social argumentation framework *is a triple* $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$*, where*

- $\mathcal{A}$ *is a set of arguments,*
- $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ *is a binary attack relation between arguments,*
- $V : \mathcal{A} \cup \mathcal{R} \to \mathbb{N} \times \mathbb{N}$ *is a total function mapping each argument and each attack to its number of positive and negative votes.*

The notion of a semantic framework is used to aggregate operators representing the several parametrisable components of a semantics:

Definition 3.2 (Semantic Framework)**.** *A social argumentation semantic framework is a 6-tuple* $\langle L, \curlywedge_{\mathcal{A}}, \curlywedge_{\mathcal{R}}, \curlyvee, \neg, \tau \rangle$ *where:*

- L *is a totally ordered set with top element* $\top$ *and bottom element* $\bot$*, containing all possible valuations of an argument.*
- $\curlywedge_{\mathcal{A}}, \curlywedge_{\mathcal{R}} : L \times L \to L$ *are two binary algebraic operation on argument valuations used to determine the valuation of an argument based on its valuation given by the votes and how weak its attackers are (*$\curlywedge_{\mathcal{A}}$*), and to determine the strength of an attack given the votes on the attack and the valuation of the attacking argument (*$\curlywedge_{\mathcal{R}}$*).*
- $\curlyvee : L \times L \to L$ *is a binary algebraic operation on argument valuations used to determine the valuation of a combined attack;*

- $\neg : L \to L$ *is a unary algebraic operation on argument valuations used to determine how weak an attack is.*
- $\tau : \mathbb{N} \times \mathbb{N} \to L$ *is a vote aggregation function which produces a valuation of an argument based on its votes.*

The definition of a semantic framework imposes very little on the behaviour of the operators. As such, many specific semantic frameworks could result in systems whose behaviour would be far from intuitive – a semantic framework where an increase in the strength of the attacking arguments would result in an increase in the strength of the attacked argument would make little sense. There are several basic properties that the operators should obey so that the resulting semantics is adequate for its purpose. For example, $\neg$ should be antimonotonic, continuous, $\neg\bot = \top$, $\neg\top = \bot$ and $\neg\neg a = a$; $\curlywedge_{\mathcal{A}}, \curlywedge_{\mathcal{R}}$ should be continuous, commutative, associative, monotonic w.r.t. both arguments and have $\top$ as their identity element; $\curlyvee$ should be continuous, commutative, associative, monotonic w.r.t. both arguments and have $\bot$ as its identity element; and τ should be monotonic w.r.t. the first argument and antimonotonic w.r.t. the second argument.

Continuity of operators guarantees small changes in the social inputs result in small changes in the models. Were this not the case, outcomes of debates would be very unstable, hard to follow and more easily exploited by trolls. The remaining algebraic properties simply state that the order in which arguments are attacked makes no difference; that an argument's valuation is proportional to its crowd support; that aggregated attacks are proportional to the attacking arguments; and so forth.

Notice also that the valuation set L of arguments (often denoted as $\mathbb{I}$ in other frameworks, as mentioned in the previous section) is parametrisable. L could be $[0, 1] \subseteq \mathbb{R}$, but it could also be any finite, countable or uncountable set of values such as booleans, colours, textures, or any other set that is deemed appropriate for users of the final application, so long as it is totally ordered.

One particular semantic framework has received great attention because of its properties. It uses a simple vote aggregation function and is based on the well known product T-norm and probabilistic sum T-conorm, which combine the desirable properties discussed above, i.e. they are continuous, commutative, associative and monotonic[3]. It is dubbed the Simple Product Semantics and is defined as follows:

[3]Besides other uses, product t-norm and its dual probabilistic sum t-conorm are the standard semantics for conjunction and disjunction, respectively, in Fuzzy Logic[45]

Definition 3.3 (Simple Product Semantics)**.** *A simple product semantic framework is* $\mathcal{S}_\epsilon^{\cdot} = \langle [0,1], \curlywedge^{\cdot}, \curlywedge^{\cdot}, \curlyvee^{\cdot}, \neg, \tau_\epsilon \rangle$ *where*

- $x \curlywedge^{\cdot} y = x \cdot y$*, i.e. the product T-norm,*
- $x \curlyvee^{\cdot} y = x + y - x \cdot y$*, i.e. the T-conorm dual to the product T-norm,*
- $\neg x = 1 - x$*,*
- $\tau_\varepsilon\left(v^+, v^-\right) = \frac{v^+}{v^+ + v^- + \varepsilon}$*, with* $\epsilon > 0$.[4]

The heart of the semantics is in the definition of a model, which combines the operators of a semantic framework $\mathcal{S}$ into a system of equations, one for each argument, that must be satisfied.

Definition 3.4 (Social Model)**.** *Let* $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$ *be a social argumentation framework and* $\mathcal{S} = \langle L, \curlywedge_{\mathcal{A}}, \curlywedge_{\mathcal{R}}, \curlyvee, \neg, \tau \rangle$ *a semantic framework. A total mapping* $M : \mathcal{A} \to L$ *is a* social model of F under semantics $\mathcal{S}$*, or* $\mathcal{S}$-model of F*, if*

$$M(a) = \tau(a) \curlywedge_{\mathcal{A}} \neg \bigcurlyvee_{a_i \in \mathcal{R}^-(a)} \left(\tau\left((a_i, a)\right) \curlywedge_{\mathcal{R}} M\left(a_i\right)\right) \qquad \forall a \in \mathcal{A}$$

where $\mathcal{R}^-(a) \triangleq \{a_i : (a_i, a) \in \mathcal{R}\}$*,* $\curlyvee\{x_1, x_2, ..., x_n\} \triangleq ((x_1 \curlyvee x_2) \curlyvee ... \curlyvee x_n)$ *and* $\tau(x) \triangleq \tau(v^+, v^-)$ *whenever* $V(x) = (v^+, v^-)$*. We refer to* $M(a)$ *as the social strength, or value, of* a *in* M*, dropping the reference to* M *whenever unambiguous.*

Each equation encodes the contribution of votes and attacks to the social strength of an argument, for which we now proceed to provide further intuition.

Whenever an argument a_i attacks another argument a, then the strength of the attack is the valuation of the attacking argument a_i reduced by the social support of the attack: no argument's attack is stronger than either its own valuation or the social support of the attack itself. We use $\curlywedge_{\mathcal{R}}$ to restrict these values.

$$\tau\left((a_i, a)\right) \curlywedge_{\mathcal{R}} M\left(a_i\right)$$

As an argument may have multiple attackers, all of their attack strengths must be aggregated to form a stronger combined attack value, using operator $\curlyvee$.

$$\bigcurlyvee_{a_i \in \mathcal{R}^-(a)} \left(\tau\left((a_i, a)\right) \curlywedge_{\mathcal{R}} M\left(a_i\right)\right)$$

[4]The meaning of ϵ is explained in [29] and, in practice, it should be a sufficiently small value with no significant influence on the result of the voting aggregation function.

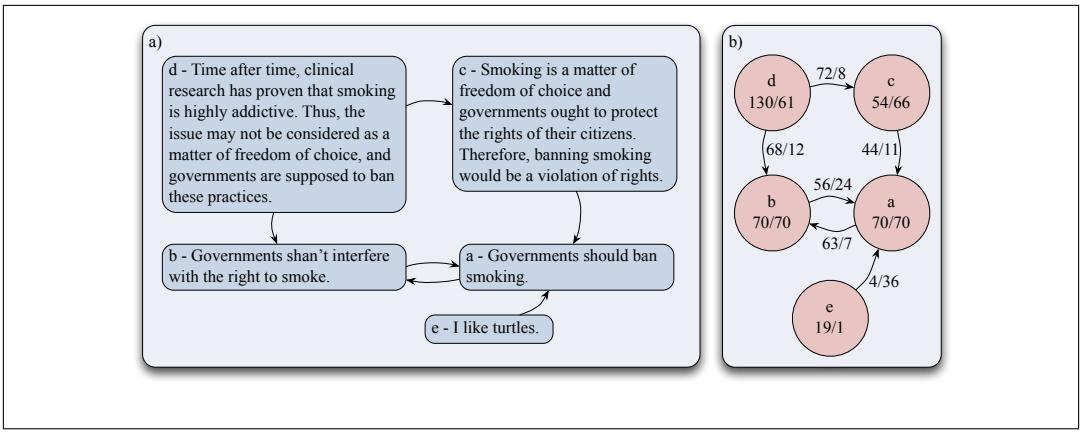

Figure 1: Social Argumentation Framework: a) arguments and attacks; b) votes

The above equation results in a combined attack strength that must be turned into a restricting value, representing how permissive or weak the attack is, using the $\neg$ operator.

$$\neg \bigvee_{a_i \in \mathcal{R}^-(a)} \left(\tau\left((a_i, a)\right) \curlywedge_{\mathcal{R}} M\left(a_i\right)\right)$$

In a social context where the crowd has given its direct opinion on argument a through the votes, it seems clear that a's valuation should never turn out higher than a's social support $\tau(a)$. Thus, an argument's valuation is given by restricting $\tau(a)$ with the value of the aggregated attack using the final operator $\curlywedge_{\mathcal{A}}$.

$$\tau(a) \curlywedge_{\mathcal{A}} \neg \bigvee_{a_i \in \mathcal{R}^-(a)} \left(\tau\left((a_i, a)\right) \curlywedge_{\mathcal{R}} M\left(a_i\right)\right)$$

Throughout the remainder of the section, $\mathcal{S}$ stands for the Simple Product Semantics.

3.2 Illustrative example

Consider a social interaction inspired by [44] where several participants, while arguing about the role of the government in what banning smoking is concerned, set forth the arguments and attack relations depicted in Fig. 1 a).

Note that these arguments are structurally different: a and b are unsupported claims, c and d contain multiple premises and a conclusion, while e, despite being rather consensual (who doesn't like turtles?), seems to be totally out of context and

can hardly be seen as an attack on a (here, the attack by e on a is meant to represent a troll attack). Our goal is to show that SAFs' level of abstraction allows meaningful arguments to be construed out of most forms of participation – in fact, with suitable GUIs, arguments could even be built from videos, pictures, links, etc. – while the participation through voting will help deal with mitigating the disturbing effect of unsound arguments and poorly specified (troll) attacks.

After a while, the arguments and attacks garner the pro/con votes depicted in Fig. 1 b). Arguments a and b obtain the same direct social support as expressed by the 70 *pro* and *con* votes. Meanwhile, a's attack on b is deemed stronger than its counterpart, judging from their votes. One might speculate that this is a consequence of a delivering a more direct message. Whereas argument c does not get much love from the crowd (a vote ratio of 54/66), its attack on a is still supported by the community (44/11). Perhaps initially there was a better sentiment towards c but the introduction of d, which amassed a decent amount of support itself (130/61), turned the odds against c. Both of d's attacks on b and c materialise to be strong enough, the former being slightly weaker (72/8 versus 68/12). Lastly, argument e received just a mere number of votes, most being positive (19/1). However, there seems to have been a significant effort from the users to discredit the attack on a by e (4/36). Note that e is a perfectly legitimate argument. Indeed the crowd endorses the fondness for turtles – it's the attack, not the argument, that is not logically well-founded.

With the abstract argumentation framework and the votes on arguments and attacks in hand, we can turn our attention to the valuation of the arguments.

If we consider the social support of each argument, i.e. its value considering only the votes it obtained while ignoring attack relations, we obtain the following values:[5] $\tau(a) = 0.50$, $\tau(b) = 0.50$, $\tau(c) = 0.45$, $\tau(d) = 0.68$ and $\tau(e) = 0.95$, as depicted in Fig. 2 a) (where the size of each node is proportional to its value).

The original SAF semantics [29], which considers attacks between arguments but not the votes on attacks, assigns the following values to arguments: $M(a) = 0,02$, $M(b) = 0,16$, $M(c) = 0,14$, $M(d) = 0,68$ and $M(e) = 0,95$, as depicted in Fig. 2 b). As expected, d and e retain their initial social support values, since they are not attacked, while the remaining arguments see a decrease in their social support value. Argument a decreases the most while b and c maintain a reasonable fraction of their initial strength. Since two of a's attackers – b and c – are attacked by d, which is a non-attacked argument with strong social support, their value is weakened, so their effect on a is lessened. Thus, we can conclude that the main cause for the downfall in a's value is e's attack.

[5] We will consider the Product Semantics as in Def.3.3, with a neglectable low ϵ

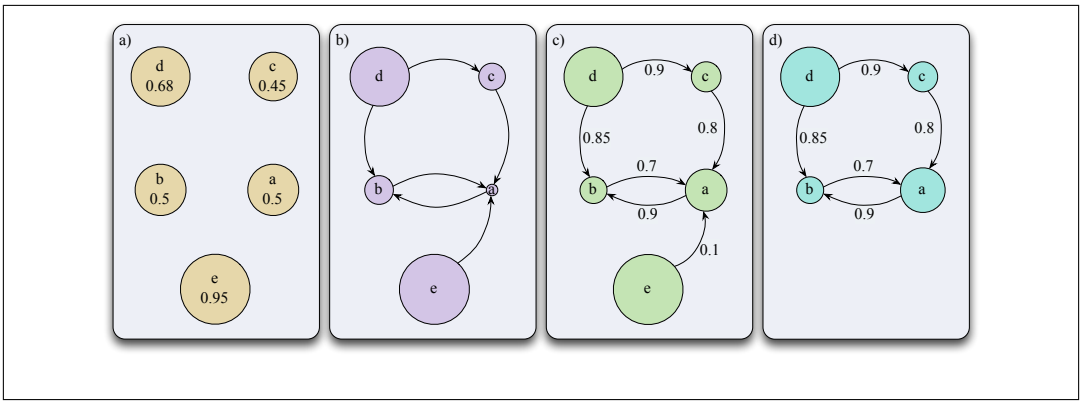

Figure 2: Model of the Social Abstract Argumentation Framework: a) considering social support only; b) considering attacks but not their strength; c) considering attack strength; d) considering attack strength, without argument e.

We can now turn our attention to the model that also takes votes on attacks into consideration, which assigns the following values to arguments: $M(a) = 0,35$, $M(b) = 0,14$, $M(c) = 0,17$, $M(d) = 0,68$ and $M(e) = 0,95$, as depicted in Fig. 2 c). The value assigned to a by the model increases from 0.02 to the more plausible level of 0.34, mostly due to e's weakened capability to attack a. Indeed, the crowd's overwhelming con votes on the (troll) attack of e on a essentially neutralised it. To confirm, we compare it with the model obtained if argument e was simply removed, depicted in Fig. 2 d), whose valuations of $M(a) = 0,39$, $M(b) = 0,14$, $M(c) = 0,17$ and $M(d) = 0,68$ are very similar to those obtained in the presence of e but with a very weakened attack on a, which allows us to conclude the success of the model in discounting attacks that are socially deemed unsound, such as troll attacks. Since the weights of the remaining attacks are relatively high and also close to each other at the same time, their impact is somewhat minimal.

3.3 Algorithms

The problem of finding a model according to the simple product semantics can be cast to the problem of finding a solution to a nonlinear system where variables represent the arguments and equations encode their attacks, with the following generic form:

Definition 3.5. *A Social Abstract Argumentation System is a square nonlinear*

system with n variables $\{x_1, \ldots, x_n\}$ *and n equations:*

$$x_i = \tau_i \prod_{j \in A_i} (1 - \tau_{ji} x_j) \qquad 1 \leq i \leq n \tag{1}$$

where $\tau_i, \tau_{ji} \in \ ]0,1[$ *and* $A_i \subseteq \{1, \ldots, n\}$.[6]

Contrary to the linear case, systems of nonlinear equations cannot be solved exactly using a finite number of elementary operations. Instead, iterative algorithms are usually used to generate a sequence $(\mathbf{x}^{(k)})_{k \in N_0}$ of approximate solutions. These algorithms start with an initial guess $\mathbf{x}^{(0)}$ and, to generate the approximating sequence, follow an iteration scheme of the form $\mathbf{x}^{(k+1)} = \mathbf{g}(\mathbf{x}^{(k)})$ where the fixed-points for $\mathbf{g}$ are solutions $\mathbf{x}^*$ of the nonlinear system.

The success of iterative algorithms depends on their convergence properties. Given a domain of interest, an iterative method that converges for any arbitrary initial guess is called globally convergent. If convergence is only guaranteed when the initial approximation is already close enough to the solution, the algorithm is called locally convergent. In the case of Social Abstract Argumentation Systems the domain of interest is $]0,1[^n$ thus the iterative algorithm must converge to a solution $\mathbf{x}^* = (x_1^*, \ldots, x_n^*) \in \ ]0,1[^n$.

Two classical algorithms that can be used to approximate the solution of such a system are the Iterative Fixed-Point Algorithm (IFP) where the iteration scheme is directly obtained from the equations (1), and the Iterative Newton-Raphson Algorithm (INR).[7]

Unfortunately, IFP is only locally convergent and often divergent, even for systems with a reduced number of variables, while INR, also only locally convergent, requires the computation of a Jacobian matrix at each iteration, which is prohibitive for large systems.

Based on the Iterative Successive Substitutions Algorithm (ISS) previously proposed for Social Argumentation Frameworks without votes on attacks [20] – itself an adaptation of the Gauss-Seidel method for systems of nonlinear equations –, here we present an adaptation to also admit votes on attacks.

Definition 3.6 (ISS)**.** *The ISS algorithm uses the iteration rule:*

$$x_i^{(k+1)} = \tau_i \prod_{j<i, j \in A_i} \left(1 - \tau_{ji} x_j^{(k+1)}\right) \prod_{j \geq i, j \in A_i} \left(1 - \tau_{ji} x_j^{(k)}\right) \tag{2}$$

[6] We can exclude $\tau_i, \tau_{ji} \in \ \{0,1\}$ because arguments and attacks (x) with $\tau(x) = 0$ have no effect in the system while $\tau(x) = 1$ can never occur, according to the simple product semantics in Def. 3.3, because $\epsilon > 0$.

[7] A comprehensive treatment of methods for solving nonlinear systems of equations with some recent developments on iterative methods can be found in [34; 9].

From the initial guess $\mathbf{x}^{(0)}$, elements of $\mathbf{x}^{(k+1)}$ are computed sequentially using forward substitution until the stopping criterion is attained.

Following a similar strategy as [20], we can prove the global convergence of the algorithm.

The algorithm performs as well as its original version [20]. For example, it is able to approximate the model of debates with 5000 arguments and an attack density of 0.1 (i.e. 10% of all pairs of arguments are related through an attack) in well under 1 second.[8] A thorough analysis of the original ISS algorithm can be found in [20]. Additionally, just as with the original ISS, we can exploit the structure of the debate to obtain considerable gains in efficiency.

4 A multi-aspect comment evaluation framework

Another generic framework that formalises the most commonly used features found in online debate platforms is s-mDiCE. This approach also introduces a set of novel features, which serve diverse purposes of debate platforms.

As already discussed, almost all online debate platforms implement some form of voting mechanism, such as positive/negative votes, like/dislike counters, star-based rating etc. s-mDiCE formalises votes, which is a generic enough mechanism that enables other types of rating to be transformed to votes rather easily. In addition, it also incorporates the notion of a *base score* (or intrinsic strength) BS, which is often used in decision-making systems (e.g., in [41] or in [13], where it is denoted as τ, as explained in Section 2). The base score offers an one-time, prior evaluation of an argument; as the dialogue evolves other users may influence this initial evaluation, positively or negatively, through their arguments or votes. It can be used to represent various notions, such as to capture an expert's initial rating over an opinion, before any debate has taken place. In some platforms, the base score may also obtain a more personalised flavor, representing for instance the trust that a user attributes to the person who issues an argument, regardless of its content.

In addition to voting mechanisms, many platforms, especially those intending to implement structured debates under the ranking-based semantics, appoint a characterisation of users' opinions as being in favour of or against other opinions or topics. According to such semantics, the strength of an opinion, or more accurately an argument, as is often referred to in these platforms, is determined by the type, number and strength of the arguments that respond to it, which are taken into account by s-mDiCE.

[8] The higher the attack density, the slower the convergence of the algorithm. However, an attack density of 0.1 seems to be a rather high value.

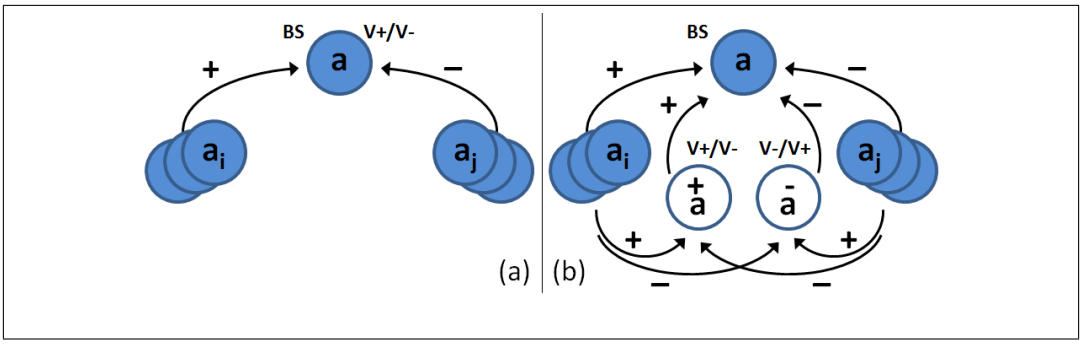

Figure 3: (a) A debate graph with votes, base score and user-generated supporting and attacking arguments, (b) The same debate graph as transformed with blank nodes.

One novel feature of s-mDiCE is the fact that it treats positive and negative votes as arguments. The idea is that a positive vote signifies a person that is happy to submit exactly the same opinion as the one stated in the original argument; this means that all arguments that support or attack the original argument also place a support or attack to that vote (see Fig. 3). Since the arguments that represent the votes do not carry any content of their own, these are called *blank arguments*, and their strength is associated with the degree with which people identified themselves with the original opinion. The symmetricity exhibited among positive and negative blank arguments may seem redundant, but it helps promote the intuitiveness of the model, as explained later on.

Another novelty is the utilisation of a set of *aspects* (dimensions), upon which an argumentative opinion is evaluated. For example, and depending on the domain of interest, such aspects may concern how relevant an argument is to the topic of discussion, how complete its justification is, how objective it is, or others. This way, users can choose to specify the aspects of an opinion they agree on and the ones they disagree, affecting the final strength accordingly. Such a scheme also assists in better understanding the intentions of users, which is difficult to capture in many of the existing platforms (for instance, often it is very difficult to interpret a down vote as representing an objection to the position stated or to the explanation given).

Finally, the framework enables a debate platform to correlate all the aforementioned features, in order to calculate two distinct scores to characterize an argument, namely *acceptance* and *quality* scores. The former aims to represent how strongly the position expressed by the argument is supported by the community, whereas the latter represents how well the position is justified; in the vast majority of existing frameworks, those two metrics are unified leading to misconceptions about what a

Functions	Description
$g^{vot} : \mathbb{N}^0 \times \mathbb{N}^0 \to \mathbb{I}$	Aggregates positive and negative votes
$g^{set} : (\mathbb{N}^0)^{\mathbb{I}} \to \mathbb{I}$	Aggregates arguments of the same polarity
$g^{diff} : \mathbb{I} \times \mathbb{I} \to \mathbb{I}'$	Aggregates arguments of different polarity
$g^{dlc} : \mathbb{I} \times \mathbb{I}' \to \mathbb{I}$	Aggregates the base score, and the strength of support and attack arguments
$g^{\mathcal{ACC}} : \mathbb{I}^N \to \mathbb{I}$	Aggregates the acceptance score of all aspects
$g^{\mathcal{QUA}} : \mathbb{I}^N \to \mathbb{I}$	Aggregates the quality score of all aspects
$\mathcal{ACC} : \mathcal{A} \to \mathbb{I}$	Returns the acceptance score of an argument
$\mathcal{QUA} : \mathcal{A} \to \mathbb{I}$	Returns the quality score of an argument

Table 1: Overview of the s-mDiCE Generic Functions

user's reaction to an opinion signifies, e.g., when one agrees with a comment that is irrelevant to a given discussion.

In the remainder of this section, a formal definition of the main functions that transform features into comparable strength scores is given, along with a study of some of their properties. An algorithm that can apply these generic functions is then presented, with the goal of clarifying how s-mDiCE works.

4.1 Formalization

s-mDiCE is a generic formal framework that enables the evaluation of the strengths of arguments considering one or more aspects. To do so, it relies on a number of functions that help quantify and aggregate the strength of the various features. These functions are shown in Table 1 and are formally defined next.

Definition 4.1. *An s-mDiCE (symmetric multi-Dimensional Comment Evaluation) framework is an (N+1)-tuple* $\langle \mathcal{A}, \mathcal{D}^*_{d1}, \ldots, \mathcal{D}^*_{dN} \rangle$*, where* $\mathcal{A}$ *is a finite set of arguments and* $\mathcal{D}^*_{d1}, \ldots, \mathcal{D}^*_{dN}$ *are aspects (dimensions), under which an argument is evaluated.*

Definition 4.2. *An aspect* $\mathcal{D}^*_x$ *corresponding to an argument set* $\mathcal{A}$ *is a 5-tuple* $\langle \mathcal{R}^{supp}_x, \mathcal{R}^{att}_x, BS_x, V^+_x, V^-_x \rangle$*, where* $\mathcal{R}^{supp}_x \subseteq \mathcal{A} \times \mathcal{A}$ *is a binary acyclic support relation on* $\mathcal{A}$*,* $\mathcal{R}^{att}_x \subseteq \mathcal{A} \times \mathcal{A}$ *is a binary acyclic attack relation on* $\mathcal{A}$*, and* $BS_x : \mathcal{A} \to \mathbb{I}$*,* $V^+_x : \mathcal{A} \to \mathbb{N}^0$ *and* $V^-_x : \mathcal{A} \to \mathbb{N}^0$ *are total functions mapping each argument to a base score (*$\mathbb{I}$ *can be any totally ordered set), a number of positive and a number of negative votes relative to this aspect, respectively.*

The attack and support relations are acyclic, since a new argument can only support/attack previously added comments. The $\mathbb{I}$ set is parameterisable, similar to the L set discussed in the previous section; in the sequel, it is assumed that $\mathbb{I} = [0, 1]$,

which is most frequently used range in similar frameworks. The base score is a fixed value assigned to each argument before any computation takes place. If no value is given, the default value can be set to a value that neutralises its effect.

Votes as blank arguments. The set of votes on any argument in an s-mDiCE framework is transformed into a pair of supporting and attacking blank arguments (Fig. 3). Before formally defining blank arguments, some convenient notation is needed. Let $\widetilde{\mathcal{A}}$ denote the set of user-generated arguments and $\mathring{\mathcal{A}}$ refer to the set of blank arguments of an s-mDiCE framework $\mathcal{F}$, such that $\mathcal{A} = \mathring{\mathcal{A}} \cup \widetilde{\mathcal{A}}$. Given an aspect $\mathcal{D}_x^* = \langle \mathcal{R}_x^{supp}, \mathcal{R}_x^{att}, BS_x, V_x^+, V_x^- \rangle$, the set of direct supporters of an argument $a \in \mathcal{A}$ is defined as $\mathcal{R}_x^+(a) = \{a_i : (a_i, a) \in \mathcal{R}_x^{supp}\}$. Similarly, the set of direct attackers of a is defined as $\mathcal{R}_x^-(a) = \{a_i : (a_i, a) \in \mathcal{R}_x^{att}\}$.

Definition 4.3. *Let $\mathcal{F}$ be an s-mDiCE framework and $\mathcal{D}_x^* = \langle \mathcal{R}_x^{supp}, \mathcal{R}_x^{att}, BS_x, V_x^+, V_x^- \rangle$ be an aspect of $\mathcal{F}$. For each argument $a \in \widetilde{\mathcal{A}}$, we define two new arguments $\overset{+}{a}$ and $\overset{-}{a}$, called the supporting blank and attacking blank argument of a respectively, such that*

- $(\overset{+}{a}, a) \in \mathcal{R}_x^{supp}$,
- $V_x^+(\overset{+}{a}) = V_x^+(a)$, $V_x^-(\overset{+}{a}) = V_x^-(a)$,
- *for all $(a_i, a) \in \mathcal{R}_x^{supp}$ it also holds that $(a_i, \overset{+}{a}) \in \mathcal{R}_x^{supp}$,*
- *for all $(a_j, a) \in \mathcal{R}_x^{att}$ it also holds that $(a_j, \overset{+}{a}) \in \mathcal{R}_x^{att}$*

Similarly,

- $(\overset{-}{a}, a) \in \mathcal{R}_x^{att}$,
- $V_x^+(\overset{-}{a}) = V_x^-(a)$, $V_x^-(\overset{-}{a}) = V_x^+(a)$,
- *for all $(a_i, a) \in \mathcal{R}_x^{supp}$ it also holds that $(a_i, \overset{-}{a}) \in \mathcal{R}_x^{att}$,*
- *for all $(a_j, a) \in \mathcal{R}_x^{att}$ it also holds that $(a_j, \overset{-}{a}) \in \mathcal{R}_x^{supp}$.*

Aggregation of votes. The positive and negative votes that each argument receives over time need to be aggregated into a single value.

Function definition 4.1. *The generic score function $g^{vot} : \mathbb{N}^0 \times \mathbb{N}^0 \to \mathbb{I}$ aggregates positive and negative votes into a single strength score.*

There are many different ways to instantiate g^{vot}, in order to represent an estimate of the community's stance towards that argument. Some are rather simplistic, e.g., averaging their population, while others provide more insights. Many frameworks often rely on the mean of the Wilson Score Interval [46], which is a popular choice for systems that need more accurate estimations, as it assesses the probability that the next vote will be of a certain polarity:

$$g^{vot}(v^+, v^-) = \frac{2 \cdot v^+ + z^2}{2 \cdot (v^+ + v^- + z^2)} \tag{3}$$

where $z = 1.96$ for a confidence level of 95%. When no votes are placed, the initial score is 0.5. One can implement other instantiations instead, considering the particular requirements of the domain, in order to control for instance the rate of convergence as an argument is populated by more votes. Clearly, the above definition of g^{vot} has the desirable property of being increasing with respect to the number of positive votes, and decreasing with respect to the number of negative votes.

To determine the smoothness of the g^{vot} function, d_S is defined as the number of votes that were added or deleted, i.e., $d_S(\langle v_1^+, v_1^- \rangle, \langle v_2^+, v_2^- \rangle) = |v_1^+ - v_2^+| + |v_1^- - v_2^-|$ (essentially, the Manhattan distance for 2-dimensional vectors), and d_T as the difference in the output, i.e., $d_T(x, y) = |x - y|$ (also the Manhattan distance, for 1-dimensional vectors). Under these definitions, it can be shown that g^{vot} is exactly $\frac{1}{2 \cdot (1+z^2)}$-smooth, and that this extreme is reached only for the first positive vote placed; all subsequent votes have strictly smaller effects. Note how the parameter z can be used to enforce different smoothness properties on g^{vot}.

Aggregation of the strength of arguments with the same polarity. Supporting and attacking arguments form a set that collectively affects the strength of the target argument. This combined support or attack can take into account, for instance, the strongest argument in the set or it can average the strength of all members in the set. Such schemes can be captured by the $g^{set} : (\mathbb{N}^0)^{\mathbb{I}} \to \mathbb{I}$ function[9] in s-mDiCE.

Function definition 4.2. *The generic score function $g^{set} : (\mathbb{N}^0)^{\mathbb{I}} \to \mathbb{I}$ aggregates the strength of a set of (supporting or attacking) arguments into a single strength score.*

For most purposes, the T-CoNorm function $\perp_{sum}: \mathbb{I} \times \mathbb{I} \to \mathbb{I}$, also known as the probabilistic sum, is a convenient choice, as it satisfies a number of useful properties,

[9] g^{set} is meant to take as input a multiset over elements of $\mathbb{I}$.

especially when $\mathbb{I} = [0, 1]$ [27]:

$$\perp_{sum}(x_1, x_2) = x_1 + x_2 - x_1 \cdot x_2 \tag{4}$$

For a multiset S of natural numbers, it is defined:

$$\perp^*_{sum}(S) = \begin{cases} 0, \text{ if } S = \emptyset \\ \perp_{sum}(x_1, \perp^*_{sum}(\{x_2, ..., x_n\})), \\ \quad \text{if } S = \{x_1, x_2 ..., x_n\} \text{ with } n > 0 \end{cases} \tag{5}$$

Consequently, the g^{set} function for the multiset of argument strengths can be instantiated as follows:

$$g^{set}(S) = \perp^*_{sum}(\{x_i : x_i \in S \text{ and } x_i \geq \vartheta\}) \tag{6}$$

Here, it is assumed that the inputs x_i represent the strength (score) of each supporting (or attacking) argument. Constant ϑ can be used to discard arguments that fall below a given threshold, rendering them ineffective in changing the strength score of other arguments. This way, irrelevant or troll arguments can easily be neutralized.

The following monotonicity properties can be easily shown for g^{set}:

- If $A, B \in (\mathbb{N}^0)^{\mathbb{I}}$, $A \subseteq B$, then $g^{set}(A) \leq g^{set}(B)$, where $\subseteq$ should be interpreted as the subset relationship for multisets.
- If $A, B, C \in (\mathbb{N}^0)^{\mathbb{I}}$, and $g^{set}(A) \leq g^{set}(B)$, then $g^{set}(C \uplus A) \leq g^{set}(C \uplus B)$, where $\uplus$ stands for "union" for multisets.

The first condition guarantees that the addition of arguments cannot decrease the combined strength of a set of arguments. The second condition ensures that the addition of "stronger" arguments has a more powerful effect than the addition of "weaker" ones. It also becomes clear from the above that when an argument's acceptance score increases, this has a negative effect on the acceptance score of all the arguments it attacks, and a positive effect on the acceptance score of all the arguments it supports. This effect propagates along the tree of arguments using this pattern.

To determine the smoothness of g^{set}, we need to define a semi-metric for $(\mathbb{N}^0)^{\mathbb{I}}$; our notion of distance will be based on the strength (based on g^{set}) of the symmetric difference between the sets compared, in particular: $d_{(\mathbb{N}^0)^{\mathbb{I}}}(X, Y) = g^{set}(X \setminus Y \uplus Y \setminus X)$. This can be viewed as a special type of edit distance, where the importance of the "edits" $(X \setminus Y, Y \setminus X)$ is judged by the g^{set} function itself. For the range of g^{set}, we will use, as usual, the semi-metric $d_T(x, y) = |x - y|$. Under these assumptions, it can be shown that g^{set} is exactly 1-smooth.

Aggregation of the strength of arguments with opposite polarity. In addition to the cumulative effect of arguments that jointly support or attack another argument, s-mDiCE needs to calculate how to balance the antagonistic effect of the supporting and attacking sets.

Function definition 4.3. *The generic score function* $g^{diff} : \mathbb{I} \times \mathbb{I} \to \mathbb{I}'$ *aggregates the strength of supporting and attacking arguments into a single strength score.*

There are different instantiations that combine these sets to specify the overall attitude. The polynomial $g^{diff} : \mathbb{I} \times \mathbb{I} \to [-1, 1]$ is often used, which offers a convenient solution for many domains:

$$g^{diff}(x_s, x_a) = x_s^n - x_a^n - x_a \cdot x_s^n + x_s \cdot x_a^n \tag{7}$$

Apparently, this function is increasing with respect to x_s and decreasing with respect to x_a, for any $n \geq 1$ (under the assumption that $\mathbb{I} = [0, 1]$). For $n = 1$ in particular, this equation results in the difference between the two values (as suggested, e.g., in [41]). In decision-making systems, where reaching reliable conclusions becomes critical, the effect of arguments should begin to matter only when they obtain a substantial strength. This behavior can be obtained for larger values of n, which force the system to react very slowly initially, but when some clear tendency for/against an opinion has appeared, the system quickly achieves a steeper increase in confidence.

In particular, it is n that determines the smoothness of g^{diff}, as g^{diff} is exactly n-smooth. The maximum effect of a change in the inputs of g^{diff} is reached when the current sentiment is very positive (close to 1) or very negative (close to 0) and someone casts an opposing argument. This means that it is easier to cast doubts on the strength of a strong argument, than to quickly trust a doubtful one. This aims at promoting the liveness of the discussion without damaging the credibility of conclusions. One can easily adapt this behavior by changing the degree of the polynomial in Eq. (7) or by applying a different function altogether. Note that lower degrees (n) in g^{diff} lead to more smooth functions; this is due to the fact that higher-degree polynomials tend to change very fast in the limit cases (i.e., when the current sentiment is close to 0 or 1), and more slowly in the intermediate points, whereas lower-degree polynomials are more uniform in their behaviour.

Of course, other schemes may also be applied according to the domain requirements. For example, one may decide to assess the informative quality of an argument by applying a scheme which increases with the strength of votes and decreases with the strength of both positive and negative arguments. This is based on the rationale that the "ideal" comment would attract only positive votes and no supporting

arguments. In other words, in an ideal setting, supporting arguments are only asserted to add material or to explain better the opinion stated, thus giving a sense of discomfort related to the quality of the target argument.[10]

Dialectical strength. The formalization so far incrementally builds the strength of an argument taking into consideration the strength of the support and opposition it attracts, including the votes, which are represented as blank arguments. These parameters are finally aggregated with the base score in the $g^{dlc} : \mathbb{I} \times [-1, 1] \to \mathbb{I}$ function, in order to provide the overall dialectical strength of an argument.

Function definition 4.4. *The generic score function* $g^{dlc} : \mathbb{I} \times [-1, 1] \to \mathbb{I}$ *valuates the dialectical strength of an argument, considering the aggregation of the base score, and the strength of the supporting and/or attacking arguments.*

As with the previous functions, different instantiations of the dialectical strength can be devised. A popular choice is to trust the base score more when the other scores do not converge to a positive or a negative value, e.g.,:

$$\begin{aligned} g^{dlc}(x_b, d) &= x_b \cdot (1 - |d|) + \frac{d + |d|}{2} \\ where\ \ d &= g^{diff}(x_s, x_a) \end{aligned} \tag{8}$$

In the equations above, it is assumed that x_b is the base score and d is the strength of the combined support and/or opposition that it has attracted. Yet another example could be to give more credit to the base score initially, and, as the argument attracts more votes and/or arguments, let the strength of the latter start weighing more in the final score. This way, rather than the supporting and attacking sets balancing each other out when they have equal strength, such a scheme will manage to capture the increasing confidence obtained as the discussion progresses.

The function defined in equation 8 is apparently increasing with respect to each of its inputs, i.e., both the base score x_b and the strength of the combined support and/or opposition that it has attracted (d). For smoothness, using the Manhattan distance for both the input and the output of the function, we can show that g^{dlc} is exactly 1-smooth. This value applies in two cases. The first is when the aggregated strength of the argument's responses is equal (or close) to 0, in which case modifications to the base score have a linear effect on the result of the function. The

[10] Of course, in practice we often see arguments, such as "I totally agree", "True!", which offer support but no valuable content. The use of aspects in s-mDiCE can help identify such arguments, reducing their quality, without affecting the supporting or attacking effect they have on the target argument.

second is when the base score is in one of the two extremes (very low, i.e., close to 0 or very high, i.e., close to 1), and only when the aggregated score d indicates an opposite stance by the community. For example, for an argument with a base score of 1, an overall negative stance by the community will more quickly lower its score, compared to the case where we had a lower base score, or a positive stance by the community. This applies also symmetrically to the opposite case.

Acceptance and quality Scores. The aforementioned aggregation functions compute various metrics regarding the strength of an argument, considering a single aspect. In addition, s-mDiCE allows for more refined valuations of a single argument. For instance, an argument may have a different set of votes regarding its relevance, another set for its objectivity and a third one for its informativeness. As a result, the final score of the argument needs to be calculated by taking into consideration the scores obtained on each individual aspect.

Function definition 4.5. *Let $\mathcal{F} = \langle \mathcal{A}, \mathcal{D}^*_{d1}, ..., \mathcal{D}^*_{dN} \rangle$ be an s-mDiCE framework. The generic score functions $g^{\mathcal{ACC}} : \mathbb{I}^N \to \mathbb{I}$ and $g^{\mathcal{QUA}} : \mathbb{I}^N \to \mathbb{I}$ can be used to aggregate the dialectical strength of each individual aspect. Eventually, the functions $\mathcal{ACC} : \mathcal{A} \to \mathbb{I}, \mathcal{QUA} : \mathcal{A} \to \mathbb{I}$ are used to denote the acceptance and quality scores of an argument $a \in \mathcal{A}$, respectively.*

A simple, weight-based solution that determines the influence of each aspect is often a sufficient solution. In the following equation, w_i quantifies the weight assigned by the system moderator on aspect $\mathcal{D}^*_{di}$ based on other metrics or by experience (where x_i will be the dialectical strength for the given aspect):

$$g^{\mathcal{QUA}}(x_1, ..., x_n) = \sum_{i=1}^{n} w_i^{QUA} \cdot x_i \text{ , with } \sum_{i=1}^{n} w_i^{QUA} = 1, w_i^{QUA} \geq 0 \tag{9}$$

$$g^{\mathcal{ACC}}(x_1, ..., x_n) = \sum_{i=1}^{n} w_i^{ACC} \cdot x_i \text{ , with } \sum_{i=1}^{n} w_i^{ACC} = 1, w_i^{ACC} \geq 0 \tag{10}$$

Given that $w_i^{QUA} \geq 0$ and $w_i^{ACC} \geq 0$ for all i, $g^{\mathcal{QUA}}$ and $g^{\mathcal{ACC}}$ are both increasing with respect to each of its inputs. Similarly, the smoothness of $g^{\mathcal{ACC}}$ and $g^{\mathcal{QUA}}$ is determined by (i.e., is equal to) the weight with the maximum value. The smoothness with respect to each aspect in particular (i.e., if we assume that the scores of all other aspects remain constant) is determined by its respective weight. An interesting conclusion from this is that "balanced" functions (where w_i are close to each other, or equal) are smoother, i.e., less sensitive to input changes. Moreover, as expected,

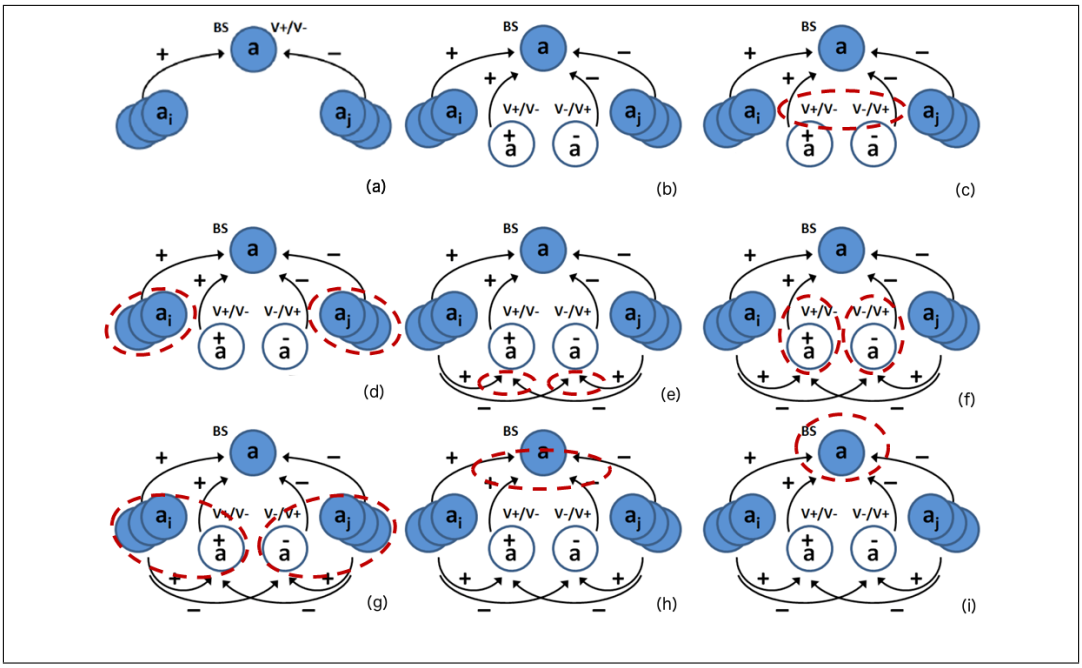

Figure 4: Computation steps of Algorithm 1.

the largest effects appear when the changed aspects are those that have the largest weights.

In the following section, we demonstrate how $\mathcal{ACC}$, $\mathcal{QUA}$ can be computed for a given argument, by successively applying all aggregate functions mentioned above. Given the monotonicity properties of their constituent functions, we observe that the effects of each aspect on the argument's acceptance and/or quality score depend on its respective weight: all aspects have a monotonically increasing effect on the outcome of the functions $\mathcal{ACC}$ and $\mathcal{QUA}$, whereas their smoothness is also determined by these weights (just like in the case of $g^{\mathcal{ACC}}$, $g^{\mathcal{QUA}}$ above).

4.2 Computation loop in s-mDiCE

Given the aforementioned aggregation functions, Algorithm 1 below presents the steps that can be followed to calculate the acceptance and quality scores of an argument in a debate graph, as the one shown in Fig. 4 (a). Despite the procedural presentation of the algorithm, not all steps need to be executed in a sequential manner.

Algorithm 1 takes as input an s-mDiCE framework consisting of one or more

Algorithm 1 *CalcScores($\mathcal{F}, a$)*: Calculate Acceptance / Quality Score

INPUT: an s-mDiCE framework $\mathcal{F} = \langle \mathcal{A}, \mathcal{D}^*_{d1}, \ldots, \mathcal{D}^*_{dN} \rangle$ and an argument $a \in \mathcal{A}$
OUTPUT: All strength scores of a

for each aspect $\mathcal{D}^*_x = \langle \mathcal{R}^{supp}_x, \mathcal{R}^{att}_x, BS_x, V^+_x, V^-_x \rangle$ of $\mathcal{F}$ **do**
  % *Generate blank arguments*
  Create $\overset{+}{a}$ and $\overset{-}{a}$, according to Definition 4.3

  % *Calculate vote strength of blank arguments*
  let $g^{vot}_x(\overset{+}{a}) = g^{vot}(V^+_x(\overset{+}{a}), V^-_x(\overset{+}{a}))$
  let $g^{vot}_x(\overset{-}{a}) = g^{vot}(V^+_x(\overset{-}{a}), V^-_x(\overset{-}{a}))$

  % *Calculate strength without blank arguments*
  let $s^{blSuppSet}_x(a) = g^{set}(\{g^{dlc}_x(a_i) : a_i \in \mathcal{R}^+_x(a) \cap \widetilde{\mathcal{A}}\})$
  let $s^{blAttSet}_x(a) = g^{set}(\{g^{dlc}_x(a_i) : a_i \in \mathcal{R}^-_x(a) \cap \widetilde{\mathcal{A}}\})$

  % *Calculate combined strength for blank arguments*
  let $g^{diff}_x(\overset{+}{a}) = g^{diff}(s^{blSuppSet}_x(a), s^{blAttSet}_x(a))$
  let $g^{diff}_x(\overset{-}{a}) = g^{diff}(s^{blAttSet}_x(a), s^{blSuppSet}_x(a))$

  % *Calculate blank arguments strength*
  let $g^{dlc}_x(\overset{+}{a}) = g^{dlc}(BS_x(\overset{+}{a}), g^{diff}_x(\overset{+}{a}))$
  let $g^{dlc}_x(\overset{-}{a}) = g^{dlc}(BS_x(\overset{-}{a}), g^{diff}_x(\overset{-}{a}))$

  % *Calculate strength with blank arguments*
  let $s^{suppSet}_x(a) = g^{set}(\{g^{dlc}_x(a_i) : a_i \in \mathcal{R}^+_x(a)\})$
  let $s^{attSet}_x(a) = g^{set}(\{g^{dlc}_x(a_i) : a_i \in \mathcal{R}^-_x(a)\})$

  % *Calculate combined support/attack strength*
  let $g^{diff}_x(a) = g^{diff}(s^{suppSet}_x(a), s^{attSet}_x(a))$

  % *Calculate acceptance/quality for one aspect*
  let $g^{dlc}_{acc,x}(a) = g^{dlc}(BS_x(a), g^{diff}_x(a))$
  let $g^{dlc}_{qua,x}(a) = g^{dlc}(BS_x(a), g^{diff}_x(a))$
end for
% *Calculate overall acceptance/quality*
let $\mathcal{ACC}(a) = g^{\mathcal{ACC}}(g^{dlc}_{d1}(a), ..., g^{dlc}_{dN}(a))$
let $\mathcal{QUA}(a) = g^{\mathcal{QUA}}(g^{dlc}_{d1}(a), ..., g^{dlc}_{dN}(a))$

aspects, and an argument, whose acceptance and quality scores are to be calculated. The majority of computations are executed for each aspect individually. The first step is to generate the blank nodes of a given aspect (Fig. 4 (b) and line 3 in Algorithm 1), followed by the computation of the vote strength of each (Fig. 4 (c) and lines 6, 7 in Algorithm 1).

Before calculating their overall strength, one needs to determine the support and attack they receive (Fig. 4 (d) and lines 10, 11 in Algorithm 1). Notice that the supporting (resp. attacking) set of a blank argument includes only the user-generated arguments that exist in the supporting (resp. attacking) set of the input argument. Their strength is then aggregated (Fig. 4 (e) and lines 14, 15 in Algorithm 1), producing all values needed to calculate the dialectical strength of the blank arguments (Fig. 4 (f) and lines 18, 19 in Algorithm 1).

The rest of the algorithm continues in a similar style, in order to compute the corresponding scores for the input argument. It starts by calculating the strength of the supporting and attacking sets, which now include the corresponding blank arguments (Fig. 4 (g) and lines 22, 23 in Algorithm 1), and then their aggregated strength (Fig. 4 (h) and line 26 in Algorithm 1).

These values are then aggregated to compute the dialectical strength of the target argument for the given aspect (Fig. 4 (i)). Notice that the algorithm computes two different values, the acceptance score (line 29 in Algorithm 1) and the quality score (line 30 in Algorithm 1). As discussed in the previous section, a feature may weigh differently in each case, which can lead to different instantiations of the dialectical aggregation function.

Finally, the overall acceptance and quality scores are computed by aggregating all corresponding scores of each individual aspect (lines 33, 34 in Algorithm 1)

As can be seen from the above, the algorithm is recursive, triggering the computation of the dialectical strength of the arguments that exist one level below the input argument in the debate graph (see lines 10, 11, 22, 23), Based on the assumption of having a debate graph without cycles, the recursion is guaranteed to terminate.

To conclude, a note is needed to justify the symmetry in computations for the supporting and the attacking blank arguments, which is evident in Algorithm 1. In particular, one can see that having computed the dialectical strength of the one, one can easily compute the strength of the other (that is, $g_x^{dlc}(\overset{+}{a}) = 1 - g_x^{dlc}(\overset{-}{a})$), which raises the question of whether it is necessary to have both blank arguments in the framework. Indeed, it is possible to substitute this pair with a single blank argument, which according to its strength, it is assigned either to the supporting or the attacking set (following a pendulum pattern). A problematic situation would

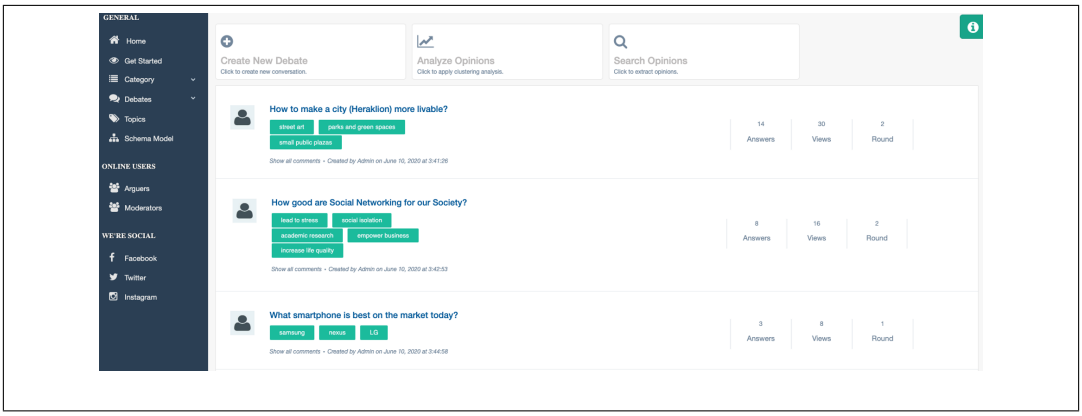

Figure 5: The Apopsis Debate Platform.

arise though when the strength of this argument is divided exactly in the middle, denoting for instance that the votes are shared among the participants in the debate. Omitting the argument altogether in this case would be counterintuitive, as its score should affect the supporting/attacking set to capture this dichotomy of opinions. For such cases, and for promoting clarity in the presentation of the model, s-mDiCE relies on a symmetric modeling of blank arguments.

4.3 Application to discussion platforms

The s-mDiCE framework was first deployed in the Apopsis platform[11] (Figure 5), a web-based debating platform that aims to motivate online users to participate in well-structured discussions by raising issues and posting ideas or comments that support or attack other opinions [48]. The main goal of the system is to offer an automated opinion analysis that determines and extracts the most useful and strongest opinions expressed in a dialogue, eventually assisting decision-makers in understanding the opinion exchange process. The aspects chosen to determine the acceptance and quality scores of arguments within Apopsis are correctness, relevance, and sufficiency of evidence (Figure 6). Users can specify which of these aspects they consider inadequate when voting against a particular opinion, optionally adding a counter-argument to support their claim. The s-mDiCE framework calculates the different opinion scores as the dialogue progresses, in order to pinpoint the strongest opinions in a debate, but also in the generation of different types of analytics, such as clustering users with similar opinions or preferences.

[11] https://demos.isl.ics.forth.gr/apopsis/

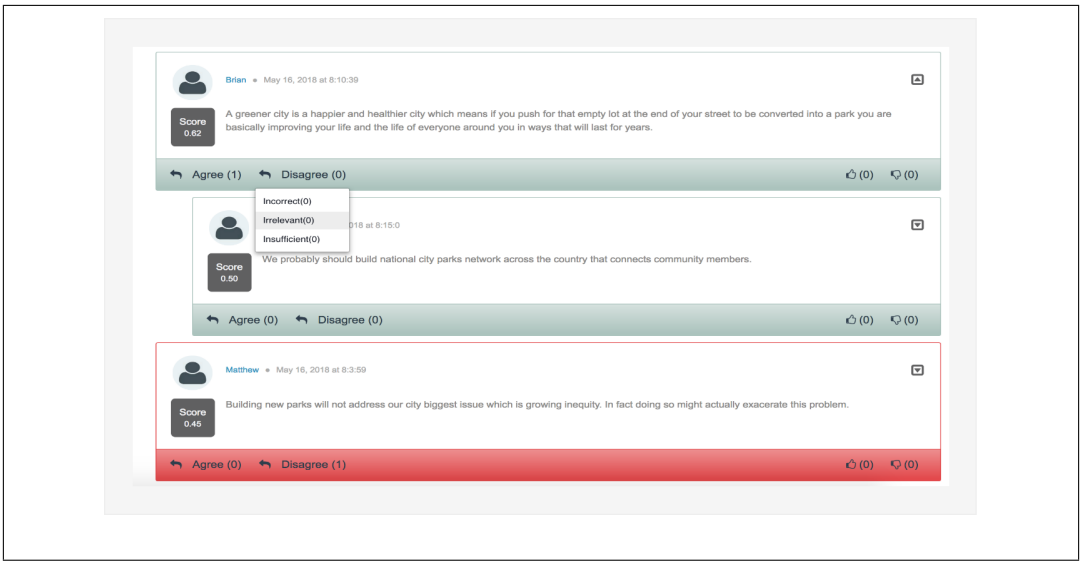

Figure 6: Rating opinions in the Apopsis platform.

Recently, the framework was also used in the Argument Navigator[12] tool, developed in the context of the DebateLab project [47]. The project developed a suite of tools and services that can assist the work of the professional journalist in accomplishing everyday tasks (e.g., writing, archiving, retrieving articles), as well as the activity of the ordinary Web user (reader) who wishes to be well-informed about topics or entities of interest. The Article Navigator in particular is a search engine that can be used to explore, visualise and rank arguments on the Web. Among other functionalities, the user can vote for or against an argument with respect to certain aspects, such as informativeness, validity, and relevance. The argument mining process is accomplished automatically, based on a token classification / sequence labeling approach for extracting segments of argumentative discourse units, while diverse Deep Learning techniques and gradient-based modeling are applied to perform argument relation and stance classification.

5 Other approaches to social argumentation

In this section we review other approaches to social argumentation. We classify them into two categories: (a) other social argumentation systems, which like SAF and s-mDiCE integrate arguments with social votes aiming to model and reason with

[12] https://debatelab.ics.forth.gr/tools/

online debates; (b) extensions or applications of existing argumentation frameworks aiming to model other aspects of discussions in social media, such as the semantic relations among posts, the social relevance of a post, controversy and multi-topic discussions.

5.1 Frameworks integrating arguments and votes

Starting with the first category, *Quantitative Argumentation Debate for Voting* (QuAD-V) frameworks were developed to support collaborative debates and deliberation within e-Democracy [40]. QuAD-V evolved from *Quantitative Argumentation Debate* (QuAD) frameworks [14; 41], which incorporate attack and support argument relations and intrinsic strengths of arguments[13]. QuAD-V extend QuAD frameworks with a set of users and their votes on arguments, based on which the base score (which in the rest of this section we refer to as *social support*) of an argument can be computed using the following formula:

$$\tau_v(a) = \begin{cases} 0.5 & if\ |\mathcal{U}| = 0 \\ 0.5 + (0.5 \times \frac{V^+(a) - V^-(a)}{|\mathcal{U}|}) & if\ |\mathcal{U}| \neq 0 \end{cases} \tag{11}$$

where $\mathcal{U}$ denotes the set of users, $V^+(a)$ the number of positive votes, and $V^-(a)$ the number of negative votes for argument a.

The following function is used to calculate the strength of an argument:

$$v_a = \begin{cases} \tau_v(a) \cdot (1 - |g(\mathcal{R}^+(a)) - g(\mathcal{R}^-(a))|) & if\ g(\mathcal{R}^-(a)) \geq g(\mathcal{R}^+(a))) \\ \tau_v(a) + (1 - \tau_v(a)) \cdot |g(\mathcal{R}^+(a)) - g(\mathcal{R}^-(a))| & otherwise \end{cases} \tag{12}$$

where g is a function that calculates the aggregated strength of a set of arguments given the strength of each argument in the set:

$$g(\{a_1, \cdots, a_n)\} = 1 - \prod_{i=1}^{n}(1 - v_{a_i}) \tag{13}$$

The QuAD-V frameworks exhibit the following properties. (*i*) $\tau_v(a)$ (the social support of a) is monotonically increasing (decreasing) with respect to the number of positive (negative) votes for a; (*ii*) v_a (the strength of a) is monotonically non-decreasing (non-increasing) with respect to the aggregated strength of the supporters (attackers) of a and the number of positive (negative) votes for a; (*iii*) an argument with stronger (weaker) attackers than supporters has a strength lower (higher) than

[13]Note that QBAFs (studied in section 2.3) are yet another extension/evolution of QuADs

the argument's social support, provided that the social support is not already minimal (maximal); (*iv*) for an argument to have the minimum (maximum) strength, either the supporters (attackers) have the minimum value and the attackers (supporters) the maximum or all votes for the argument are negative (positive) with its attackers (supporters) at least as strong as its supporters (attackers); (*v*) v_a is continuous with respect to $g(\mathcal{R}^+(a))$ and $g(\mathcal{R}^+(a))$, i.e. the aggregated strength of the attackers (resp., supporters) of a.

A novelty of this work is the characterisation of users as rational/irrational taking into account their votes on each argument, its attackers and its supporters. A user is considered irrational in the following two cases: (a) (s)he agrees with (votes positively for) an argument, agrees also with one of its attackers but does not agree with any of its supporters; (b) (s)he disagrees with (votes negatively for) an argument, agrees with one of its supporters but does not agree with any of its attackers. Based on the concept of rationality, they also introduce a methodology (*QuAD-V opinion polling*) for evolving polls, which aims at highlighting and eradicating irrationalities in user's opinions through a series of dynamic questions to irrational users, making the polls more informative to the pollster.

A similar approach to integrating votes and arguments was developed and implemented in [24] to support *Quaestio-it*, a web-based Q&A debating platform. The main underlying idea was the same: the strength of an argument is determined by the (positive and negative) votes it receives and the strength of its attacking and supporting arguments. Specifically, they define two functions (f_{att}, f_{supp}) that calculate the strength of an argument taking into account its social support (i.e. positive and negative votes it has received) and the aggregated strength of its attackers (resp. supporters):

$$f_{att}(a) = \tau_v(a) \cdot (1 - g(\mathcal{R}^-(a)) \tag{14}$$

$$f_{supp}(a) = \tau_v(a) \cdot (1 + g(\mathcal{R}^+(a)) \tag{15}$$

The strength of a set of arguments is calculated recursively using the following formula:

$$g(\{a_1, a_2, \ldots, a_n\}) = v_{a_1} + (1 - v_{a_1}) \cdot g(\{a_2, \ldots, a_n\}) \tag{16}$$

where v_{a_1} is the strength of argument a_1 and $g(\emptyset) = 0$.

The social support for an argument is calculated using the lower bound of the Wilson Score Interval [46]:

$$ws(x, y) = \frac{n}{n + z^2}\left[\hat{p} + \frac{z^2}{2n} - z\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2}{4n^2}}\right] \tag{17}$$

where $n = x + y$, $\hat{p} = x/n$ and $z = 1.96$ for a confidence level of 95%. The social support of an argument a is given by:

$$\tau_v(a) = ws(V^+(a), V^-(a)) \tag{18}$$

where $V^+(a)$ is the number of positive votes, and $V^-(a)$ the number of negative votes for argument a.

The following function is used to calculate the strength of a:

$$v_a = \begin{cases} \tau_v(a) & if\ \mathcal{R}^-(a) = \mathcal{R}^+(a) = \emptyset \\ f_{supp}(a) & if\ \mathcal{R}^-(a) = \emptyset \\ f_{att}(a) & if\ \mathcal{R}^+(a) = \emptyset \\ (f_{supp}(a) + f_{att}(a))/2 & otherwise \end{cases} \tag{19}$$

Similarly with s-mDiCE, this framework is symmetric with respect to supporting and attacking arguments, i.e., a supporting argument increases the value of an argument's strength by the same amount by which an equivalent attack would decrease it. However, the aggregated strength of a set of arguments is defined in a way that induces discontinuity in certain cases.

5.2 Modelling other aspects of social web debates

In this section we review studies that apply models and methods from formal argumentation to model and/or reason with various aspects of social media discussions, such as the semantic relations among posts, the social relevance of a post, controversy and multi-topic discussions. One study of this type is presented in [1]; its main aim is to analyse discussions in Twitter, specifically to identify the social accepted tweets and measure the controversy between the users participating in a discussion. To achieve this aim, they model a discussion in Twitter as a Value-based Argumentation Framework $\mathcal{F} = \langle T, attacks, R, W, Valpref \rangle$, where T is the set of tweets, $attacks = \{(t_1, t_2) \mid t_1, t_2 \in T \text{ and } t_1 \text{ criticises } t_2\}$, R is a non-empty set of ordered values used to model the social relevance of tweets, $W : A \rightarrow R$ is a function that assigns a value from R to each tweet and $Valpref \subseteq R \times R$ is the ordering relation over R. The ideal extension of $\mathcal{F}$ is the *accepted set of tweets* in the discussion. They consider three different ways to define W (i.e., to quantify the social relevance of a tweet), each of which takes into account a different type of information: the number of followers of the author of the tweet; the number of retweets of the tweet; and the number of favourites for the tweet. They also present an analysis discussion system, which consists of two components: the Discussion Retrieval component, which retrieves relevant information from a discussion, i.e. the tweets,

their semantic relations, and the number of followers, retweets and favourites; and the Reasoning component, which computes the accepted set of tweets. Finally, they present two measures for controversy in a discussion: the *controversy degree*, which is the number of rejected tweets (i.e., tweets that are not in the ideal extension of the corresponding VAF) that criticise an accepted tweet, and the *controversy depth*, which is the length of the longest controversial path (sequence of tweets connected via the *attacks* relation) in a discussion.

A similar approach is used in [2] to model and reason with debates in Reddit. They model a Reddit debate Γ with root comment r as a *Debate Tree* $\mathcal{T} = \langle C, r, E, L\rangle$ such that for every comment in Γ there is a node in C, r is the root of $\mathcal{T}$, if c_1 answers c_2 in Γ then there is a directed edge (c_1, c_2) in E, and $L : E \to [-2, 2]$ assigns a value to each edge denoting the sentiment of the corresponding answer, from highly negative (-2) to very positive (2). They then prune the Debate Tree by discarding neutral comments (based on their sentiment values) and their subtrees. To find the accepted comments in a debate, they map the corresponding pruned debate tree to a Value-based Argumentation Framework: each comment is mapped to an argument, each answer to a comment is mapped to an attack from the answer to the comment if the sentiment of the answer is negative, and the score of each comment is mapped to a natural number that represents its social acceptance. The set of accepted comments in a Reddit debate is the ideal extension of the corresponding VAF. They also propose measures for quantifying the users' influence, the controversy that they generate throughout a debate, their contribution to the polarisation of the debate and their social acceptance. In all such measures, they use the notion of *filtered tree*, which results from a pruned debate tree by removing the comments of a given user (excluding the root comment) and all the comments in the subtrees rooted by the user's comments. Some of these measures are: (*i*) the *debate engaging degree* of a user u, which is used to quantify the interactions of the user; (*ii*) the *influence degree* of a user u, which is used to quantify the comments that change their status (from accepted to rejected or vice versa) if we disregard the comments from u; (*iii*) the *polarisation degree* of a solution S, which is a measure of the bias of S towards comments in favour of the root comment and comments against the root comment; (*iv*) the *rebalancing degree* of a user u, which quantifies the influence of the user on the polarisation of a debate solution; and (*v*) the *social acceptance* of a user u, which sums up the scores of the user's comments;

An argumentation framework that integrates the notion of *topic tags* or *hashtags* used in social media applications, such as Facebook and Twitter, was proposed in [17]. They introduce the notion of *hashtagged argument*, which they define as a pair $\langle a, \mathcal{H}_a\rangle$, where $\mathcal{H}_a$ denotes a set of hashtags associated with argument a. To model the relations among topics, they define *hashtag graphs*; a vertex in such a

graph denotes a hashtag, and an undirected edge between two vertices denotes some relationship between the corresponding hashtags. The distance between two vertices in a hashtag graph is the number of edges in the shortest path connecting them. The distance between two hashtagged arguments can be defined in several ways, for example as the minimum or the maximum or the average distance between the hashtags of the two arguments. A *hashtagged argumentation framework* Ω is defined as a pair $\langle \Phi, \mathcal{G} \rangle$, where Φ is an AAF consisting of hashtagged arguments and $\mathcal{G}$ is a hashtag graph. To reason with such frameworks they adjust the standard acceptability semantics of AAFs to take into account the hashtags of arguments and their relations. Specifically, they redefine acceptability as follows: a hashtagged argument a is ϵ-*acceptable* w.r.t. a set of hashtagged arguments S when for every hashtagged argument b that attacks a there is a hashtagged argument $c \in S$ that attacks b and $d_\Omega(a, c) \leq \epsilon$, where d_Ω is a distance function for Ω and ϵ a user-defined threshold. Admissible, complete, grounded and preferred semantics are then defined in the standard way. They also provide an alternative definition for acceptability semantics, which takes into account both the distance between an argument and its defenders but also the distance between the arguments in an extension.

An earlier study explored the use of Bipolar Argumentation Frameworks for modelling and reasoning with online debates [18]. Specifically, they proposed the use of textual entailment for classifying the relation between two sentences in one of the following types: *entailment*, i.e. the meaning of one of the two sentences can be inferred from the other; *contradiction*, i.e., the two sentences cannot be simultaneously true; and *unknown*, i.e., the truth of one sentence cannot be verified on the basis of the other. Using an empirical study they found a high correlation between entailment and support, i.e., in most (61.6%) of the cases where annotators identified that a sentence a supports another sentence b, they also identified that a entails b, and an even higher correlation between contradiction and attack, i.e., in most (71.4%) of the cases where the annotators identified that a sentence a attacks another sentence b, they also identified that a contradicts b. They also verified with another experiment that the correlation between attack and contradiction also holds for other types of attacks that can be deduced from a bipolar argumentation framework, i.e., *supported*, *secondary*, *mediated* and *extended* attacks.

Finally, as mentioned in 2, Labeled Bipolar Argumentation Frameworks [26] is another formalism that can be applied to social argumentation systems. Similar to s-mDiCE, it enables the valuation of an argument with respect to different dimensions (argument features) taking into account the strength of its attackers and supporters. A distinctive characteristic of this formalism is that it allows assigning ranges of values to an argument, which is useful when there is uncertainty in the original valuation of an argument. On the other hand, it does not explicitly handle

social votes; in their running example from the domain of social platforms, they represent the social rating of an argument as one of its features and assume that the original argument valuation for this feature are given. A common limitation of both approaches compared to Social Argumentation Frameworks (discussed in Section 3) is their inability to treat cycles in the argument graph.

6 Conclusion

The recent trend in Web usage has elevated users from pure consumers of information to "prosumers", i.e., both consumers and producers of information. A large part of this trend is attributed to websites that allow users to contribute their opinions on any conceivable topic, reviews on physical or digital products and services, as well as commentaries on events, people, ideas, or things. This trend often has the form of a discussion, with arguments that support one's opinions, as well as responses or other reactions to such opinions by other users. In such cases, making sense out of a (possibly long and complex) debate is important for users, and part of the sense-making process is the ability to automatically evaluate arguments.

The original argumentation theory is not fully suitable to cope with this evaluation procedure because the arguments appearing in these kinds of debates are rarely totally accepted or totally rejected; instead, a numerical assessment under the so-called *gradual semantics* is more suitable. Furthermore, the original argumentation theory has no support for the temporal dimension (i.e., the order in which the arguments are presented), or for other types of reactions (such as votes) that one can typically use in such systems.

In this chapter, we surveyed *social argumentation systems*, i.e., various systems and frameworks specifically designed to model, analyse or enable these kinds of debates. This includes theories and principles that such systems should satisfy, as well as specific technical solutions that address these issues and the properties that such solutions satisfy. Our aim is to provide an overview for interested researchers and practitioners in choosing the most suitable solution for their purposes, and/or in developing alternative methods for argument analysis or evaluation in such settings.

Future avenues of research in this area include (*i*) the further development or extension of social argumentation systems to take into account the characteristics of users (user profiles, expertise or popularity of users, etc,); (*ii*) the evaluation of social argumentation systems using data from online debates in social networks or debate websites; and (*iii*) the development of social web applications that fully exploit the capabilities of such frameworks to facilitate and analyse online debates (some examples of such applications are already available and are described in another

chapter of this volume).

To the best of the authors' knowledge, social argumentation systems focus on abstract arguments. Recasting those ideas in the context of structured argumentation (e.g., ASPIC+ [38], [33] or ABA [21]) is another future work direction with significant potential. In particular, the additional information provided by the arguments' structure may be exploited both to allow more specific user input (e.g., votes relating only to a particular premise of an argument, or to the argument's reasoning), and a more fine-grained evaluation of the argument that takes into account its structure.

References

[1] Teresa Alsinet, Josep Argelich, Ramón Béjar, Cèsar Fernández, Carles Mateu, and Jordi Planes. Weighted argumentation for analysis of discussions in twitter. *Int. J. Approx. Reason.*, 85:21–35, 2017.

[2] Teresa Alsinet, Josep Argelich, Ramón Béjar, and Santi Martínez. Measuring user relevance in online debates through an argumentative model. *Pattern Recognit. Lett.*, 133:41–47, 2020.

[3] L. Amgoud and J. Ben-Naim. Ranking-based semantics for argumentation frameworks. In *Proceedings of the Seventh International Conference on Scalable Uncertainty Management (SUM)*, pages 134–147, 2013.

[4] L. Amgoud and J. Ben-Naim. Axiomatic foundations of acceptability semantics. In *Proceedings of the Fifteenth International Conference on Principles of Knowledge Representation and Reasoning (KR-16)*, pages 2–11, 2016.

[5] L. Amgoud and J. Ben-Naim. Evaluation of arguments from support relations: Axioms and semantics. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)*, pages 900–906, 2016.

[6] L. Amgoud and J. Ben-Naim. Evaluation of arguments in weighted bipolar graphs. In *Proceedings of the Fourteenth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU-17)*, pages 25–35, 2017.

[7] L. Amgoud, J. Ben-Naim, D. Doder, and S. Vesic. Ranking arguments with compensation-based semantics. In *Proceedings of the Fifteenth International Conference on Principles of Knowledge Representation and Reasoning (KR-16)*, pages 12–21, 2016.

[8] L. Amgoud, J. Ben-Naim, D. Doder, and S. Vesic. Acceptability semantics for weighted argumentation frameworks. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)*, pages 56–62, 2017.

[9] I. K. Argyros and F. Szidarovszky. *The Theory and Applications of Iteration Methods*. Systems Engineering. Taylor & Francis, 1993.

[10] P. Baroni, M. Romano, F. Toni, M. Aurisicchio, and G. Bertanza. An argumentation-based approach for automatic evaluation of design debates. In *CLIMA XIV: Proceedings*

of the 14th International Workshop on Computational Logic in Multi-Agent Systems - Volume 8143, pages 340–256, 2013.

[11] Pietro Baroni and Massimiliano Giacomin. On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence*, 171(10):675–700, 2007.

[12] Pietro Baroni, Antonio Rago, and Francesca Toni. How many properties do we need for gradual argumentation? In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, 2018.

[13] Pietro Baroni, Antonio Rago, and Francesca Toni. From fine-grained properties to broad principles for gradual argumentation: A principled spectrum. *Int. J. Approx. Reason.*, 105:252–286, 2019.

[14] Pietro Baroni, Marco Romano, Francesca Toni, Marco Aurisicchio, and Giorgio Bertanza. Automatic evaluation of design alternatives with quantitative argumentation. *Argument and Computation*, 6(1):24–49, 2015.

[15] Antonis Bikakis, Giorgos Flouris, Theodore Patkos, and Dimitris Plexousakis. Sketching the vision of the web of debates. *Journal of the Frontiers of AI, research topic on Computational Argumentation: a foundation for Human-centric AI*, 6, 2023.

[16] Elise Bonzon, Jérôme Delobelle, Sébastien Konieczny, and Nicolas Maudet. A comparative study of ranking-based semantics for abstract argumentation. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence*, 2016.

[17] Maximiliano Celmo David Budán, Maria Laura Cobo, Diego C. Martínez, and Guillermo Ricardo Simari. Proximity semantics for topic-based abstract argumentation. *Inf. Sci.*, 508:135–153, 2020.

[18] Elena Cabrio and Serena Villata. A natural language bipolar argumentation approach to support users in online debate interactions†. *Argument Comput.*, 4(3):209–230, 2013.

[19] Martin Caminada. Rationality postulates: Applying argumentation theory for non-monotonic reasoning. In *Handbook of Formal Argumentation*. College Publications, 2018.

[20] Marco Correia, Jorge Cruz, and João Leite. On the efficient implementation of social abstract argumentation. In Torsten Schaub, Gerhard Friedrich, and Barry O'Sullivan, editors, *ECAI 2014 - 21st European Conference on Artificial Intelligence, 18-22 August 2014, Prague, Czech Republic - Including Prestigious Applications of Intelligent Systems (PAIS 2014)*, volume 263 of *Frontiers in Artificial Intelligence and Applications*, pages 225–230. IOS Press, 2014.

[21] K. Cyras, X. Fan, C. Schulz, and F. Toni. Assumption-based argumentation: Disputes, explanations, preferences. In *Handbook of Formal Argumentation*. College Publications, 2018.

[22] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–357, September 1995.

[23] Sinan Egilmez, João G. Martins, and João Leite. Extending social abstract argumentation with votes on attacks. In Elizabeth Black, Sanjay Modgil, and Nir Oren, editors,

Theory and Applications of Formal Argumentation - Second International Workshop, TAFA 2013, Beijing, China, August 3-5, 2013, Revised Selected papers, volume 8306 of *Lecture Notes in Computer Science*, pages 16–31. Springer, 2013.

[24] Valentinos Evripidou and Francesca Toni. Quaestio-it.com: a social intelligent debating platform. *Journal of Decision Systems*, 23(3):333–349, 2014.

[25] Giorgos Flouris, Theodore Patkos, Antonis Bikakis, Alexandros Vassliades, Nick Bassiliades, and Dimitris Plexousakis. Theoretical analysis and implementation of abstract argumentation frameworks with domain assignments. *International Journal of Approximate Reasoning (IJAR)*, 161C, 2023.

[26] Melisa G. Escañuela Gonzalez, Maximiliano C. D. Budán, Gerardo I. Simari, and Guillermo R. Simari. Labeled bipolar argumentation frameworks. *Journal of Artificial Intelligence Research (JAIR)*, 70, 2021.

[27] Erich Peter Klement, Radko Mesiar, and Endre Pap. *Triangular Norms*. Springer, 1 edition, 2000.

[28] W. Kunz and H. Rittel. *Issues as elements of information systems. Working Paper 131*. Institute of Urban and Regional Development, University of California, Berkeley, California, 1970.

[29] João Leite and João G. Martins. Social abstract argumentation. In Toby Walsh, editor, *IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011*, pages 2287–2292. IJCAI/AAAI, 2011.

[30] Michael Luca. Reviews, reputation, and revenue: The case of yelp.com. Technical report, Technical Report 12-016, Harvard Business School, 2011.

[31] P. Matt and F. Toni. A game-theoretic measure of argument strength for abstract argumentation. In *Proceedings of the Eleventh European Conference on Logics in Artificial Intelligence (JELIA-08)*, pages 285–297, 2008.

[32] S. Modgil, F. Toni, F. Bex, I. Bratko, C. I. Ches nevar, W. Dvořák, M. A. Falappa, X. Fan, S. A Gaggl, A. J. García, M. P. González, T. F. Gordon, J. Leite, M. Možina, C. Reed, G. R. Simari, S. Szeider, P. Torroni, and S. Woltran. The added value of argumentation. In S. Ossowski, editor, *Agreement Technologies*, volume 8 of *Law, Governance and Tech. Series*, pages 357–403. Springer, 2013.

[33] Sanjay Modgil and Henry Prakken. Abstract rule-based argumentation. *Journal of Applied Logics - IfCoLog Journal of Logics and their Applications (FLAP)*, 4, 2017.

[34] J. M. Ortega and W.C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2000.

[35] Theodore Patkos, Antonis Bikakis, and Giorgos Flouris. A multi-aspect evaluation framework for comments on the social web. In Chitta Baral, James P. Delgrande, and Frank Wolter, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference, KR 2016, Cape Town, South Africa, April 25-29, 2016*, pages 593–596. AAAI Press, 2016.

[36] Theodore Patkos, Antonis Bikakis, and Giorgos Flouris. Rating comments on the socialweb using a multi-aspect evaluation framework. Technical report, TR-463, Institute of Computer Science, Foundation for Research and Technology - Hellas, 2016.

[37] Theodore Patkos, Giorgos Flouris, and Antonis Bikakis. Symmetric multi-aspect evaluation of comments - extended abstract. In Gal A. Kaminka, Maria Fox, Paolo Bouquet, Eyke Hüllermeier, Virginia Dignum, Frank Dignum, and Frank van Harmelen, editors, *ECAI 2016 - 22nd European Conference on Artificial Intelligence, 29 August-2 September 2016, The Hague, The Netherlands - Including Prestigious Applications of Artificial Intelligence (PAIS 2016)*, volume 285 of *Frontiers in Artificial Intelligence and Applications*, pages 1672–1673. IOS Press, 2016.

[38] Henry Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1:93–124, 2010.

[39] Antonio Rago, Pietro Baroni, and Francesca Toni. On instantiating generalised properties of gradual argumentation frameworks. In *Scalable Uncertainty Management*, 2018.

[40] Antonio Rago and Francesca Toni. Quantitative argumentation debates with votes for opinion polling. In Bo An, Ana L. C. Bazzan, João Leite, Serena Villata, and Leendert W. N. van der Torre, editors, *PRIMA 2017: Principles and Practice of Multi-Agent Systems - 20th International Conference, Nice, France, October 30 - November 3, 2017, Proceedings*, volume 10621 of *Lecture Notes in Computer Science*, pages 369–385. Springer, 2017.

[41] Antonio Rago, Francesca Toni, Marco Aurisicchio, and Pietro Baroni. Discontinuity-free decision support with quantitative argumentation debates. In Chitta Baral, James P. Delgrande, and Frank Wolter, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference, KR 2016, Cape Town, South Africa, April 25-29, 2016*, pages 63–73. AAAI Press, 2016.

[42] M. Thimm and G. Kern-Isberner. On controversiality of arguments and stratified labelings. In *Proceedings of Computational Models of Argument (COMMA-14*, 2014.

[43] Leendert van der Torre and Srdjan Vesic. The principle-based approach to abstract argumentation semantics. In *Handbook of Formal Argumentation*. College Publications, 2018.

[44] D. Walton. Argumentation theory: A very short introduction. In I. Rahwan and G. R. Simari, editors, *Argumentation in Artificial Intelligence*, pages 1–22. Springer, 2009.

[45] Haohao Wang, Wei Li, and Bin Yang. An extension of several properties for fuzzy t-norm and vague t-norm. *Journal of Intelligent & Fuzzy Systems*, 46(3):6881–6891, 2024.

[46] Edwin B. Wilson. Probable Inference, the Law of Succession, and Statistical Inference. *Journal of the American Statistical Association*, 22(158):209–212, 1927.

[47] E. Ymeralli, G. Flouris, V. Efthymiou, K. Papantoniou, T. Patkos, G. Petasis, N. Pittaras, G. Roussakis, and E Tzortzakakis. Representing online debates in the context of e-journalism. In *The Sixteenth International Conference on Advances in Semantic Processing (SEMAPRO 2022)*, 2022.

[48] Elisjana Ymeralli, Giorgos Flouris, Theodore Patkos, and Dimitris Plexousakis. Apop-

sis: A web-based platform for the analysis of structured dialogues. In *On the Move to Meaningful Internet Systems. OTM 2017 Conferences: Confederated International Conferences: CoopIS, C&TC, and ODBASE 2017, Rhodes, Greece, October 23-27, 2017, Proceedings, Part II*, page 224–241, 2017.

[49] Liuwen Yu, Dongheng Chen, Lisha Qiao, Yiqi Shen, and Leendert van der Torre. A principle-based analysis of abstract agent argumentation semantics. In *Proceedings of the 18th International Conference on Principles of Knowledge Representation and Reasoning*, pages 629–639, 2021.

Received December 2024