

Dispute Trees as Explanations in Quantitative (Bipolar) Argumentation

Kristijonas Čyras¹, Timotheus Kampik² and Qingtao Weng¹

¹Ericsson Research, Sweden and USA

²Umeå University, Sweden

Abstract

We present an approach to explaining inference in Quantitative Bipolar Argumentation Graphs (QBAGs): we propose the notion of a Quantitative Dispute Tree (QDT) that effectively collects the QBAG's directed paths between a topic argument and its children, labelled as proponent and opponent. A QDT is intended to be interpreted as a dispute that is sufficient to establish the direction of the change of strength of the topic argument. We propose to define pro and con arguments by using contribution functions that quantify a given argument's contribution to the strength of another argument. We advance some principles for contribution functions as well as gradual semantics to ensure some reasonable properties of QDTs.¹

Keywords

Explainable AI, Quantitative argumentation, Dispute trees

1. Introduction


In AI research focused on reasoning, formal argumentation has emerged as a promising facilitator of eXplainable AI (XAI) [1, 2]. However, to properly foster explainability formal argumentation needs to be explainable in itself. One popular way to explain argumentation-based inference is the construction of Dispute Trees (DTs) [3, 4]: given a topic argument in an argument graph, DTs collect arguments *pro* and *con* the topic argument, expanding along the edges. With an intuitive interpretation of the expansion and labelling of arguments as proponents and opponents, DTs facilitate explainability of (non-)acceptance of arguments in argument graphs (see [1] for an overview). Previously, DTs have been introduced to abstract, structured and, recently, bipolar argumentation approaches [3, 4, 5]. We here introduce DTs to quantitative bipolar argumentation which features both support and attack relations as well as weighted nodes (argument strengths) in argument graphs. Quantitative argumentation is of interest to the community as a family of argumentative reasoning methods that use numerical information [6]. It is also of interest as a potential bridge to connectionist AI approaches and explainability thereof, e.g. when interpreting feed-forward neural networks [7] or used in explainable recommender systems [8].

¹At the time of writing this paper, Kristijonas Čyras was partially unaffiliated, but the paper is based on the research carried out with Ericsson Research in Sweden. Qingtao Weng is affiliated with Ericsson in Sweden.

1st International Workshop on Argumentation for eXplainable AI (ArgXAI, co-located with COMMA '22), September 12, 2022, Cardiff, UK

✉ kcyras@gmail.com (K. Čyras); tkampik@cs.umu.se (T. Kampik); qingtao.weng@ericsson.com (Q. Weng)

🆔 0000-0002-4353-8121 (K. Čyras); 0000-0002-6458-2252 (T. Kampik); 0000-0001-7574-8951 (Q. Weng)

 © 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

In this spirit of aiming at explainable reasoning approaches that can incorporate numerical (statistical or otherwise) information, we advance Quantitative Dispute Trees (QDTs) for explainability of argument strengths in quantitative (bipolar) argumentation. We specifically tackle the problem of explaining the change from initial to final argument strength in quantitative bipolar argumentation graphs (QBAGs) evaluated using (numerical) gradual semantics. We will define QDTs as formal graph-based structures that help to explain or justify why the (numerical) strength of a given topic argument increases or decreases when a QBAG is evaluated, and how other arguments contribute to this increase/decrease. The following example provides an intuition of this paper’s main contribution.

Example 1.1. Consider the QBAG G depicted in Figure 1a (we give formal definitions of QBAGs and their semantics in Preliminaries). There are six arguments a, b, c, d, e, f with initial strengths given in brackets, and final strengths (given in boldface) that result after evaluating G using the Quadratic Energy (QE) gradual semantics. We consider a to be the topic argument whose strength increase (from 0.5 to 0.519) is to be explained. We aim to answer these questions:

- How much does each argument contribute to the final strength of a ?
- Which arguments are pro and which con (a getting stronger)?
- Which arguments when disclosed are sufficient to guarantee that the final strength of a will not decrease below the initial strength even when further available arguments are disclosed?

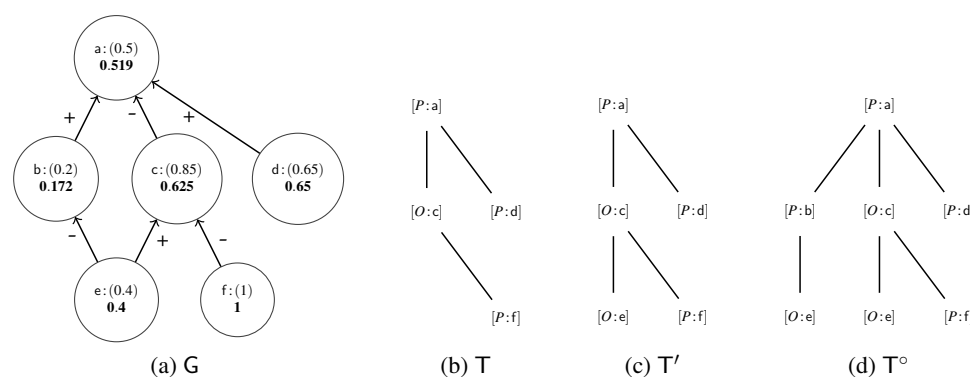


Figure 1: Subfigure 1a shows a QBAG G . Here, a node labelled $\begin{matrix} x:(i) \\ \mathbf{f} \end{matrix}$ carries argument x with initial strength $\tau(x) = i$ and final strength $\sigma(x) = \mathbf{f}$. Edges labelled $+$ and $-$ respectively represent attack and support. Subfigures 1b, 1c and 1d show quantitative dispute trees (QDTs) constructed from G , to be defined and discussed in Section 4.

It is not immediate how much each argument contributes to a being stronger – we will propose some ways to quantify that. It is also not clear which arguments contribute positively or negatively, e.g. f , as an attacker of an attacker, can be said to be pro a getting stronger. But it would not be so straightforward if f also attacked d – for simplicity, this situation is not included in this example, but our approach will allow to qualify that as well. It is further not clear whether all the arguments are needed to establish that a gets stronger. Perhaps only some will suffice, e.g. the strong con argument c and the pro arguments d and f , so that whether b or e are considered would not change the fact that a only gets stronger. We will advance QDTs (some depicted in Figure 1) that capture these considerations.

In this paper we will propose a generic notion of argument *contribution* to another argument that is meant to both qualify whether one argument is pro or con another, and to quantify how much one arguments contributes to the final strength of another one. We will advance a few concrete instantiations of argument contribution functions applicable to quantitative argumentation and suggest a few reasonable principles that contribution functions should arguably satisfy. We will then propose a definition of QDTs that use proponent and opponent labels defined by a contribution function for an interpretation as a kind of dispute regarding the change in the strength of a given topic argument. Our QDTs will effectively amount to taking a connected sub-graph of a QBAG and collecting the directed paths between the topic argument and its children into a tree that is in the end sufficient to establish the topic argument getting stronger or weaker.

This work is a very preliminary step towards defining QDTs for QBAGs. We will see that there are many ways of going about defining argument contributions and subsequent explanations of argument strength changes, and rather than proposing a definite approach, we would like to initiate a discussion around the possible merits and drawbacks of different formalisations.

2. Preliminaries

This section introduces the formal preliminaries of our work. Let \mathbb{I} be a real interval. Typically, $\mathbb{I} = [0, 1]$ is the unit interval. A *quantitative bipolar argumentation graph* contains a set of arguments related by binary *attack* and *support* relations, and assigns an *initial strength* in \mathbb{I} to the arguments. The (initial) strength can be thought of as initial credence in, or importance of, arguments. Typically, the greater the strength in \mathbb{I} , the more credible or important the argument is.

Definition 2.1 (Quantitative Bipolar Argumentation Graph (QBAG) [9, 6]). A *Quantitative Bipolar Argumentation Graph (QBAG)* is a quadruple $(Args, \tau, Att, Supp)$ consisting of a set of arguments $Args$, an *attack* relation $Att \subseteq Args \times Args$, a *support* relation $Supp \subseteq Args \times Args$ and a total function $\tau : Args \rightarrow \mathbb{I}$ that assigns the *initial strength* $\tau(x)$ to every $x \in Args$.

Henceforth, we assume as given a fixed but otherwise arbitrary QBAG $G = (Args, \tau, Att, Supp)$, unless specified otherwise. We also assume that $Args$ is finite.

Given a $a \in Args$, the set $Att_G(x) := \{z \mid z \in Args, (z, x) \in Att\}$ is the set of attackers of x and each $z \in Att_G(x)$ is an *attacker* of x ; the set $Supp_G(x) := \{y \mid y \in Args, (y, x) \in Supp\}$ is the set of supporters of x and each $y \in Supp_G(x)$ is a *supporter* of x . We may drop the subscript G when the context is clear.

Reasoning in QBAGs amounts to updating the initial strengths of arguments to their final strengths, taking into account the strengths of attackers and supporters. Specifically, given a QBAG, a strength function assigns final strengths to arguments in the QBAG. Different ways of defining a strength function are called gradual semantics.

Definition 2.2 (QBAG Semantics and Strength Functions [6, 9]). A *gradual semantics* σ defines for $G = (Args, \tau, Att, Supp)$ a (possibly partial) *strength function* $\sigma_G : Args \rightarrow \mathbb{I} \cup \{\perp\}$ that assigns the *final strength* $\sigma_G(x)$ to each $x \in Args$, where \perp is a reserved symbol meaning ‘undefined’.

Note that we restrict ourselves to numeric strength functions by stipulating \mathbb{I} to be an interval, instead of a generic set of elements with a preorder. This is typical of many gradual semantics and will simplify our exposition by allowing arithmetic operations with strengths.

We may abuse the notation and drop the subscript G so that σ denotes the strength function.

A gradual semantics can define a strength function as a composition of multivariate real-valued functions that determines the strength of a given argument by aggregating the strengths of its attackers and supporters, taking into account the initial strengths [9]. A strength function so defined is recursive and generally takes iterated updates to produce a sequence of strength vectors, whence the final strengths are defined as the limits (or fixed points) if they exist. However, for *acyclic* QBAGs – namely, QBAGs without directed cycles – defining a semantics and computing the final strengths can be more straightforward: in the topological order of an acyclic QBAG as a graph, start with the leaves,¹ set their final strengths to equal their initial strengths, and then iteratively update the strengths of parents whose all children already have final strengths defined.

Table 1 gives a list of common influence and aggregation functions. Table 2 shows some examples of gradual semantics.

Aggregation	Functions	
Sum	$\alpha_v^{\Sigma} : [0, 1]^n \rightarrow \mathbb{R}$	$\alpha_v^{\Sigma}(s) = \sum_{i=1}^n v_i \times s_i$
Product	$\alpha_v^{\Pi} : [0, 1]^n \rightarrow [-1, 1]$	$\alpha_v^{\Pi}(s) = \prod_{i:v_i=-1} (1 - s_i) - \prod_{i:v_i=1} (1 - s_i)$
Top	$\alpha_v^{max} : [0, 1]^n \rightarrow [-1, 1]$	$\alpha_v^{max}(s) = M_v(s) - M_{-v}(s),$ where $M_v(s) = \max\{0, v_1 \times s_1, \dots, v_n \times s_n\}$
Influence	Functions	
Linear(k)	$l_w^l : [-k, k] \rightarrow [0, 1]$	$l_w^l(s) = w - \frac{w}{k} \times \max\{0, -s\} + \frac{1-w}{k} \times \max\{0, s\}$
Euler-based	$l_w^e : \mathbb{R} \rightarrow [w^2, 1]$	$l_w^e(s) = 1 - \frac{1-w^2}{1+w \times e^s}$
p-Max(k) for $p \in \mathbb{N}$	$l_w^p : \mathbb{R} \rightarrow [0, 1]$	$l_w^p(s) = w - w \times h(-\frac{s}{k}) + (1-w) \times h(\frac{s}{k}),$ where $h(x) = \frac{\max\{0, x\}^p}{1+\max\{0, x\}^p}$

Table 1

Common aggregation α and influence l functions [9, pp. 1724 Table 1; with a fixed typo for p-Max(k)]. Parameter s represents the strength of an argument at that state, w the initial strength, and s_i and $v_i \in \{-1, 1\}$ the strengths and relationships, respectively, of the argument’s attackers/supporters.

Semantics	Aggregation	Influence
QuadraticEnergyModel	Sum	2-Max(1)
SquaredDFQuADModel	Product	1-Max(1)
EulerBasedTopModel	Top	EulerBased
EulerBasedModel	Sum	EulerBased
DFQuADModel	Product	Linear(1)

Table 2

Examples of gradual semantics. In this paper we use QuadraticEnergyModel for illustrations.

While many gradual semantics can be defined for QBAGs in general, their convergence is not always guaranteed in a particular QBAG. For several semantics, convergence is however guaranteed in acyclic QBAGs. (See e.g. [9] for a neat exposition of convergence results under various semantics.) We restrict our attention to QBAGs for which a fixed but otherwise arbitrary gradual semantics is well-defined. In other words, our study applies to the setting where a gradual semantics σ defines a total strength function σ_G assigning the final strengths to all arguments

¹Here, leaves are nodes without incoming edges.

of a given G . We specifically use acyclic QBAGs and the Quadratic Energy (QE; see Table 2) semantics [10] (computable in accordance with a topological ordering of an acyclic QBAG). We however note that both the formal definitions and theoretical analysis given in this paper apply to the more general setting with well-defined gradual semantics giving total strength functions.

Example 2.1. In Example 1.1, QE semantics defines the following strength function σ :

$$\sigma(x) = \tau(x) + (1 - \tau(x)) \cdot \frac{\max\{E(x), 0\}^2}{1 + \max\{E(x), 0\}^2} - \tau(x) \cdot \frac{\max\{-E(x), 0\}^2}{1 + \max\{-E(x), 0\}^2}$$

where

$$E(x) = \sum_{Supp(y,x)} \sigma(y) - \sum_{Att(z,x)} \sigma(z).$$

Concretely, $\sigma(f) = \tau(f) = 1$, $\sigma(e) = \tau(e) = 0.4$ and $\sigma(d) = \tau(d) = 0.65$. For $\sigma(b)$, note that $E(b) = -\sigma(e)$, so $\sigma(b) = \tau(b) + (1 - \tau(b)) \cdot \frac{\max\{-\tau(e), 0\}^2}{1 + \max\{-\tau(e), 0\}^2} - \tau(b) \cdot \frac{\max\{\tau(e), 0\}^2}{1 + \max\{\tau(e), 0\}^2} = \tau(b) - \tau(b) \cdot \frac{\tau(e)^2}{1 + \tau(e)^2} = 0.172$ (all numbers are rounded to the 3rd decimal in this example).² Similarly, $E(c) = \sigma(e) - \sigma(f)$, so that $\sigma(c) = \tau(c) - \tau(c) \cdot \frac{(\tau(f) - \tau(e))^2}{1 + (\tau(f) - \tau(e))^2} = 0.625$. Finally, $\sigma(a) = 0.519$. The final strengths thus obtained are given in bold within nodes in Figure 1a.

3. Argument Contributions

To begin with, in order to determine if one argument positively or negatively contributes to the strength of another argument – akin to being a proponent or an opponent in a fictitious dispute regarding the final strength of arguments – we may consider different ways to define such contributions. For instance, we could ask how different the final strengths of a given topic argument would be in the absence of the contributing argument and/or relevant relationships – in other words, in various sub-graphs of a QBAG. A useful notion for talking about QBAGs as sub-graphs of one another is that of a restriction of a QBAG to a set of arguments, as follows.

Definition 3.1 (QBAG Restriction). Given a QBAG $G = (Args, \tau, Att, Supp)$ and a set of arguments $A \subseteq Args$, we define the *restriction of G to A* as the QBAG $G \downarrow_A := (A, \tau \cap (A \times \mathbb{I}), Att \cap (A \times A), Supp \cap (A \times A))$.

We next summarise some ways to *attribute* some argument's $x \in Args$ contribution to the final strength $\sigma_G(a)$ of a given argument $a \in Args$ in G . We will denote such contribution by $Ctrb_{G,a}(x)$.

- Compute $\sigma_G(a) - \sigma_{G'}(a)$, where G' is obtained from G by removing x :

$$Ctrb_{G,a}(x) = \sigma_G(a) - \sigma_{G \downarrow_{Args \setminus \{x\}}}(a) \quad (1)$$

This is the basic idea found in [11], but the authors therein stipulate that it ignores the fact that the children of x contribute to the strength of x and hence to the strength of a , and that this

²These computations can be done using the code at github.com/kcyras/QDT and found on Jupyter notebook github.com/kcyras/QDT/blob/main/exampleQDT.ipynb.

definition is thus, arguably, not the right one. We could argue against that: after all, x “absorbs” the contributions from its children, and so its mere existence enables those to propagate further to a . We, however, do not intend to settle whether this is a reasonable definition of attributing contributions, but merely give it as an option, following [11].

- Similarly following [11], we can compute $\sigma_{G^-}(a) - \sigma_{G'}(a)$ where G^- is obtained from G by removing direct relations to x and G' is obtained from G by removing x entirely:

$$\text{Ctrb}_{G,a}(x) = \sigma_{(Args, \tau, Att \setminus \{(y,x) | (y,x) \in Att\}, Supp \setminus \{(y,x) | (y,x) \in Supp\})}(a) - \sigma_{G \downarrow_{Args \setminus \{x\}}}(a) \quad (2)$$

In other words, we determine the strength of a with attackers and supporters of x removed (thus considering only the “direct contribution” of x) and take away the strength of a with x removed.

- Following the game-theoretic approaches to capture the degree of influence on inconsistency [12] or of inputs on outputs in decision making systems [13, 14, 15, 16, 17], we can for instance use Shapley(-like) values:

$$\text{Ctrb}_{G,a}(x) = \sum_{X \subseteq Args \setminus \{x\}} \frac{|X|! \cdot (|Args|! - |X| - 1)!}{|Args|!} \left(\sigma_{G \downarrow_{Args \setminus (X \cup \{x\})}}(a) - \sigma_{G \downarrow_{Args \setminus X}}(a) \right) \quad (3)$$

This measure is in general computationally intractable, but can nevertheless serve as an option.

- Many attribution techniques in numerical analysis aim to quantify which variable a given function f is sensitive to: *gradient saliency*, for instance, calculates the derivative of f ’s outputs with respect to its inputs [18]. In our setting, assuming G is acyclic, we can traverse G in its topological order and write $\sigma(a) = f(\tau(a), \tau(b_1), \dots, \tau(b_K))$, where $\{b_1, \dots, b_K\}$ are all the arguments from which there is a path³ to a in G and f is a composition of aggregation α and influence ι functions. Then, assuming f is differentiable, compute the partial derivative $\frac{\partial f}{\partial \tau(x)}$ of f with respect to the initial strength of x and evaluate it at the point of all the initial strengths.

$$\text{Ctrb}_{G,a}(x) = \left. \frac{\partial f}{\partial \tau(x)} \right|_{(\tau(a), \tau(b_1), \dots, \tau(x), \dots, \tau(b_K))} \quad (4)$$

Example 3.1. In Example 1.1, G is acyclic and we have spelled out in Example 2.1 e.g. $\sigma(b)$ in terms of all the other arguments connected to b via a direct path, namely e . So we can use the gradient saliency method given in Equation 4 to determine the contributions. For instance,

$$\frac{\partial \sigma(b)}{\partial \tau(e)} = \frac{\partial}{\partial \tau(e)} \left(\tau(b) - \tau(b) \cdot \frac{\tau(e)^2}{1 + \tau(e)^2} \right) = \frac{2\tau(b)\tau(e)^3}{(1 + \tau(e)^2)^2} - \frac{2\tau(b)\tau(e)}{1 + \tau(e)^2}.$$

So $\text{Ctrb}_{G,b}(e) = \frac{2\tau(b)\tau(e)^3}{(1 + \tau(e)^2)^2} - \frac{2\tau(b)\tau(e)}{1 + \tau(e)^2} \Big|_{\tau(b)=0.2, \tau(e)=0.4} \approx -0.119$.

We will be interested in explaining the strength change of argument a in Example 1.1, so we compute the following contributions: $\text{Ctrb}_a(b) = 0.158$, $\text{Ctrb}_a(c) = -0.134$, $\text{Ctrb}_a(d) = 0.183$, $\text{Ctrb}_a(e) = -0.123$, $\text{Ctrb}_a(f) = 0.101$.

Whatever the specific method for determining an argument’s contribution to the final strength of another argument, we will rely on a generic notion of contribution that we stipulate as follows.

³There is a directed path in G from x to a iff $\exists (xR_1b_1 \dots b_nR_n a)$ with $R_k \in \{Att, Supp\}$ and $b_k \in Args \forall k \in \{1, \dots, n\}, n \geq 1$.

Definition 3.2 (Contribution). Given a QBAG G , a *topic* argument $a \in \text{Args}$ and any $x \in \text{Args}$, we let $\text{Ctrb}_{G,a}(x) \in \mathbb{R}$ (or simply $\text{Ctrb}_a(x)$ where the context is clear) denote the *contribution of x to the topic argument's final strength $\sigma_G(a)$* . We call the function $\text{Ctrb} : \text{Args} \times \text{Args} \rightarrow \mathbb{R}$ defined by $\text{Ctrb}(a,x) = \text{Ctrb}_a(x)$ a *contribution function*.

Note that the first three contribution function examples given above are partial functions; in particular, they do not define the topic argument's contribution to itself $\text{Ctrb}_a(a)$, because $\sigma_{G'}(a)$ is undefined when a is not in (the set of arguments of) G' . The gradient saliency contribution function (Equation 4), however, defines contributions for all arguments to all others.

We expect some intuitive behaviours from a contribution function and can formalise these expectations as contribution function principles, analogously to principles that have been defined for gradual semantics [6]. In the case of contribution functions, principle satisfaction obviously depends on the properties of the semantics used. One intuitive requirement is that some contributions exist whenever the final strength of the topic argument changes.

Principle 3.1 (Contribution Existence). A contribution function Ctrb satisfies the *contribution existence principle* w.r.t. a gradual semantics σ iff for every QBAG $G = (\text{Args}, \tau, \text{Att}, \text{Supp})$, for every $a \in \text{Args}$ it holds that whenever $\sigma_G(a) \neq \tau(a)$, then $\exists x \in \text{Args}$ s.t. $\text{Ctrb}_a(x) \neq 0$.

Not all contribution functions given above always satisfy the contribution existence principle:

Example 3.2 (Violation of Contribution Existence). Assume a gradual semantics for acyclic graphs where the final strength of an argument depends on the *maximal* absolute final strength of its attackers and supporters: $\sigma(x) = \tau(x) - \max\{\tau(y) \mid y \in \text{Att}(x)\}$ (behaviour that reflects one of the basic ideas in abstract argumentation [19]). Consider the following QBAG with a topic argument a having exactly two attackers b and c : $(\{a, b, c\}, \{(a, 0), (b, 1), (c, 1)\}, \{(b, a), (c, a)\}, \{\})$. We have $\sigma(a) = -1 \neq \tau(a)$. Using Equation 1 to define Ctrb , since removing only one attacker still leaves the same final strength of a , we get $\text{Ctrb}_a(b) = \text{Ctrb}_a(c) = 0$. Ctrb thus violates the contribution existence principle w.r.t. semantics σ .

On the other hand, the gradient saliency contribution function together with a differentiable strength function such as given by the QE semantics does satisfy Principle 3.1 (basic calculus).

Another intuitive principle, which we call *directionality*, stipulates that an argument can only contribute to another argument if there is a directed path from the former to the latter. As a shorthand, we denote the existence of a directed path from x to a (in G) by $r_G(x, a)$, and non-existence of any such path by $\neg r_G(x, a)$.

Principle 3.2 (Directionality). A contribution function Ctrb satisfies the *directionality principle* w.r.t. a gradual semantics σ iff for every QBAG $G = (\text{Args}, \tau, \text{Att}, \text{Supp})$, for each $a, x \in \text{Args}$ it holds that whenever $\neg r_G(x, a)$, then $\text{Ctrb}_a(x) = 0$.

We can imagine a gradual semantics that makes an argument's final strength dependent on the arguments it attacks or supports (somewhat reflecting the idea of *range-based* semantics in abstract argumentation [20]). For such a semantics, all of the contribution functions introduced above would violate the directionality principle. For instance, consider the strength function $\sigma(x) = \tau(x) - \sum_{\{y \mid x \in \text{Att}(y)\}} \tau(y)$, the QBAG $(\{a, b\}, \{(a, 1), (b, 1)\}, \{(b, a)\}, \{\})$ with b attacking

a , and Ctrb defined using Equation 1. There is no directed path from a to b , but removing a changes the final strength of b from 0 to 1, whence the contribution of a to b is $0 - 1 \neq 0$.

We can also observe that for instance the gradient saliency contribution function with QE semantics satisfies Principle 3.2 (as the derivative on a variable not in the function equals 0).

We now want to propose a use of argument contributions towards explaining argument strength changes in QBAGs by means of labelling quantitative dispute trees that we define next.

4. Quantitative Dispute Trees

Given a topic argument a in G , a quantitative dispute tree (QDT) can be defined as a tree T consisting of alternating proponent and opponent arguments (according to their contributions $\text{Ctrb}_a(\cdot)$) and undirected links among them following the relationships from G . When the final strength of a increases (or at least does not decrease) compared to its initial strength $\tau(a)$, we will be interested in QDTs that correspond to sub-graphs of G in which the final strength of a increases (or at least does not decrease) and does not oscillate around $\sigma(a)$ upon addition of further arguments. Analogously for the case when the final strength of a decreases in G compared to $\tau(a)$. We first define QDTs and then formalise the desired condition.

Definition 4.1 (Quantitative Dispute Tree (QDT)). A *quantitative dispute tree* (QBDT) for the topic argument $a \in \text{Args}$ in an acyclic QBAG $G = (\text{Args}, \tau, \text{Att}, \text{Supp})$ is an in-tree⁴ T , such that:

1. every node of T is of the form $[L : x]$, with $L \in \{P, O\}$ and $x \in \text{Args}$: the node holds argument named x and is
 - either a proponent (P) node, denoted $[P : x]$, if $\text{Ctrb}_a(x) \geq 0$,
 - or an opponent (O) node, denoted $[O : x]$, if $\text{Ctrb}_a(x) < 0$;
2. the root of T is $[P : a]$ – a P node holding a ;
3. every non-root node $[C : c]$ is a child of **exactly one** $[L : x]$ in T such that either $c \in \text{Att}_G(x)$ or $c \in \text{Supp}_G(x)$.

Note that the above P and O are “graph-global” notions of proponent and opponent, given topic a , rather than relative to attackers and/or supporters of a given argument, as is common in dispute trees defined for non-quantitative argumentation [3, 4, 5].

For the sake of representational simplicity, we avoid explicit naming of nodes in QDTs and implicitly force nodes holding the same argument to appear as multiple distinct nodes. In particular, if some $c \in \text{Args}$ relates to (attacks or supports) distinct $x, y \in \text{Args}$, then $[L : c]$ can appear as children of both nodes holding x and y in T only as implicitly distinctly named copies. We likewise keep the argument relationships implicit without labelling edges of QDTs.

We first note that QDTs always exist:

Proposition 4.1 (Existence of QDTs). *For every acyclic QBAG $G = (\text{Args}, \tau, \text{Att}, \text{Supp})$, for every $a \in \text{Args}$, there exists a QDT T for a .*

Proof. By definition of a QDT, using only condition 2, one can always construct the trivial QDT with exactly one node $[P : a]$. Still, observe that non-trivial QDTs exist as well: take the

⁴A directed rooted tree with edges oriented towards the root.

unique connected component of G containing a and map each of its directed paths from every $x \in \text{Args} \setminus \{a\}$ into a branch of QDT T by traversing the path from a to x in the opposite direction of attacks and supports. This procedure is well-defined due to conditions 1 and 3. \square

Note that the P and O labels of nodes do not determine the construction of a QDT, but are intended to help to “read” the QDT, by providing interpretation as to which nodes are pro and con, as determined by argument contributions.

In Example 1.1, Figure 1d depicts QDT T° that is constructed from the whole of G , using contributions from Example 3.1 to label nodes of T° as proponent or opponent.

Now, observe that there is a natural correspondence between a QDT T constructed for some topic argument in G and a unique sub-graph of G that has the same arguments as those held by the nodes of T (with the same argument held by multiple nodes in T , if any, mapping to the same node in G) and the relations recovered using condition 3 of Definition 4.1. So with an abuse of notation we will denote by $\sigma_T(x)$ the final strength of x in the sub-graph of G to which the QDT T thus corresponds. With this, we are ready to define the desired property of QDT sufficiency.

Definition 4.2 (Sufficiency). A QDT T is *sufficient* for a if **for any** (non-necessarily proper) super-tree T' of T satisfying the QDT conditions 1, 2 and 3 from Definition 4.1, it holds that:

- $\sigma_{T'}(a) \geq \tau(a)$, in case $\sigma_G(a) \geq \tau(a)$;
- $\sigma_{T'}(a) < \tau(a)$, in case $\sigma_G(a) < \tau(a)$.

Note that the sufficiency condition entails that T itself must satisfy $\sigma_T(a) \geq \tau(a)$, in case $\sigma_G(a) \geq \tau(a)$ (respectively, $\sigma_T(a) < \tau(a)$, in case $\sigma_G(a) < \tau(a)$). Intuitively, a QDT witnesses the (non-)decrease of the topic argument’s strength when compared to the initial one.

In Example 1.1, T° given in Figure 1d (as before, the proponent and opponent labels are determined using contributions from Example 3.1) is sufficient for a : it is a QDT as observed after Proposition 4.1; it satisfies $\sigma_{T^\circ}(a) = \sigma_G(a) \geq \tau(a)$; and it has no proper super-trees.

While QDTs for a exist in general (Proposition 4.1), existence of sufficient QDTs may need some reasonable restrictions of the strength function. We identify one such property of gradual semantics next, which says that the final strength of any argument depends only on the arguments from which there is a directed path to the argument in question.

Principle 4.1 (Directional Connectedness). A gradual semantics σ satisfies the *directional connectedness principle* iff for every QBAG $G = (\text{Args}, \tau, \text{Att}, \text{Supp})$, for any $x \in \text{Args}$ it holds that $\sigma_G(x) = \sigma_{G \downarrow_{\{x\} \cup \{y \in \text{Args} \mid r_G(y,x)\}}}(x)$.

We claim that our directional connectedness principle is different from the existing gradual argumentation principles such as General Principles (GP) 1 and 6 in [6]. Roughly, GP1 insists that the final strength of an argument differs from its initial strength only if it has direct attackers or supporters; however, it does not exclude contributions from “downstream” (i.e. children of attackers/supporters) or unconnected arguments. GP6 insists that arguments of equal initial strength, with equally strong attackers and supporters, are of equal final strength; however, this does not preclude equal contribution of all arguments to all other arguments, roughly speaking.

Assuming directional connectedness, sufficient QDTs are guaranteed to exist:

Proposition 4.2 (Sufficient QDTs). *If σ satisfies the directional connectedness principle, then for every QBAG $G = (\text{Args}, \tau, \text{Att}, \text{Supp})$, for every $a \in \text{Args}$, there exists a sufficient QDT T for a .*

Proof. Let G^* be the sub-graph of G containing a and all and only the children of a , i.e. $G^* = G \downarrow_{\{a\} \cup \{y \in \text{Args} \mid r_G(y,a)\}}$. Using Principle 4.1, $\sigma_G(a) = \sigma_{G^*}(a)$. As per the proof of Proposition 4.1, construct a QDT T using the whole of G^* . Then G^* trivially corresponds to T , so that $\sigma_{G^*}(a) = \sigma_T(a)$. Since T has no proper super-trees, T is sufficient for a . \square

In practice, we can construct a sufficient QDT by inspecting how adding various arguments and relationships, starting with the topic a , affects the final strength of a in the (QBAG corresponding to the) QDT being thus constructed.

Example 4.1. We claim that QDT T given in Figure 1b from Example 1.1 is sufficient for a . First, we have $\sigma_T(a) = 0.524 > \tau(a)$. Extending T with a proponent b would only increase the final strength of a . But even extending T with only the opponent e yields a super-tree T' with $\sigma_{T'}(a) = 0.5003 > \tau(a)$; see Figure 1c. So all super-trees of T (including itself) satisfy the sufficiency condition given in Definition 4.2.

Note that T is actually a minimal sufficient QDT for a . Indeed, without $[P:f]$, i.e. in the sub-tree with nodes $\{[P:a], [O:c], [P:d]\}$, the final strength of a would be 0.481. This rules out both such a sub-tree and the sub-tree with nodes $\{[P:a], [P:d]\}$ from satisfying sufficiency. Similarly, without $[P:d]$, the final strength of a in the sub-tree of T with arguments $\{[P:a], [O:c], [P:f]\}$ would be 0.424, which with the above rules out any proper sub-tree of T from satisfying sufficiency.

Once can also check that in this example T is actually a unique minimal sufficient QDT for a . For instance, the candidate tree with nodes $\{[P:a], [P:b], [O:c], [P:d]\}$ (and any of its sub-trees) are ruled-out by the fact that in the tree T° without $[P:f]$ the final strength of a is 0.499. Similarly, without d and e , the final strength of a is 0.476, so that trees with subsets of nodes $\{[P:a], [P:b], [O:c], [O:e], [P:f]\}$ are ruled out from satisfying sufficiency as well.

All in all, super-trees of T are all and only QDTs sufficient for a , with T being the unique minimal such. T can thus be seen as a minimum dispute that suffices to establish that the final strength of a will be above $\tau(a)$, increasing towards $\sigma_G(a)$.

Our running example above is with respect to the gradient saliency contribution function. Using different contribution functions, including those offered in Section 3, may result in different QDTs. This would mean that using QDTs as a form of justification or explanation of argument strength changes in QBAGs depends and requires agreement on the notion of argument contribution to begin with. We think investigations of QDT dependence on argument contribution quantification is an interesting line of research that we leave for the future. We can nonetheless observe that assuming all the above proposed principles for reasonable strength and contribution functions, there is at least one sufficient QDT with a non-trivially contributing argument:

Proposition 4.3. *If a gradual semantics σ satisfies the directional connectedness principle and a contribution function Ctrb satisfies both the contribution existence and directionality principles w.r.t. σ , then for every acyclic QBAG $G = (\text{Args}, \tau, \text{Att}, \text{Supp})$, for every $a \in \text{Args}$ with $\sigma(a) \neq \tau(a)$, there exists a QDT T sufficient for a with a node $[L:x]$ such that $\text{Ctrb}_a(x) \neq 0$.*

Proof. Using Principle 4.1, we can construct QDT T sufficient for a from the sub-graph of G containing all and only the children of a , as in the proof of Proposition 4.2. Using Principle 3.1, there is $x \in \text{Args}$ with $\text{Ctrb}_a(x) \neq 0$. Using Principle 3.2, by contraposition, there is a directed path from x to a in G . Thus, using Principle 4.1 again, T has a node $[L:x]$, as required. \square

5. Discussion

Our work adds to the growing body of research on argumentative explainability [1, 2]. Our focus lies in formally explaining argumentative inference using dispute trees in gradual argumentation. DTs are a well-established concept in formal argumentation and play a crucial role in research on formal – and in particular, abstract – argumentation semantics [4]. More recently, DTs have been positioned as facilitators of argumentative explainability, supporting both abstract and structured argumentation [21]. However, DTs have so far not been introduced to quantitative argumentation. Generally, with the exception of the recent work (by the first two authors of this paper) on change explainability in QBAGs [22], no research appears to exist that formally studies the explainability of gradual argumentation-based inference. Instead, most works on quantitative argumentation and explainability focus on the application of the former to facilitate the latter, e.g. in the context of recommender systems [8, 23]. We speculate that a better formal understanding of gradual argumentation explainability can serve as a facilitator of its application potential.

The formal work on quantitative dispute trees (QDTs) we present in this paper has plenty of limitations and thus space for open discussion. For instance, different contribution functions may adhere to different desirable principles: the contribution function given by Equation 1 may satisfy that given two attackers/supporters of an argument, the one with the higher *absolute* final strength has the higher contribution; whereas the gradient saliency contribution function may satisfy that given two attackers/supporters of an argument, the one whose final strength is higher *proportionally* to its initial strength has the higher contribution; this different behaviour can be inspected in Example 3.1. Contribution function dependency on gradual semantics is also an important issue: for instance, if an attacker of the topic argument is “unreasonably” assigned positive contribution, then the attacker may “unintuitively” appear as a proponent node in a QDT. Delineating further desirable properties of contribution functions and characterising their satisfaction with respect to various gradual semantics would be an interesting research direction.

The current work can also potentially be extended in various other directions. One is to explain why an argument’s final strength exceeds or falls below some threshold value or to explain the differences in final strengths of two or more arguments (we are studying definitions of multi-topic quantitative dispute graphs). Another is a comparison of quantitative and abstract dispute trees. And as with most explainability research, the usefulness of QDTs as explanations in realistic applications and with user studies should be investigated. We hope to pursue this in the future.

References

- [1] K. Čyras, A. Rago, E. Albin, P. Baroni, F. Toni, Argumentative XAI: A Survey, in: Z.-H. Zhou (Ed.), 30th International Joint Conference on Artificial Intelligence, IJCAI, Montreal, 2021, pp. 4392–4399. doi: [10.24963/ijcai.2021/600](https://doi.org/10.24963/ijcai.2021/600). arXiv: 2105.11266.
- [2] A. Vassiliades, N. Bassiliades, T. Patkos, Argumentation and explainable artificial intelligence: a survey, *The Knowledge Engineering Review* 36 (2021) e5. doi: [10.1017/S0269888921000011](https://doi.org/10.1017/S0269888921000011).
- [3] P. M. Dung, R. Kowalski, F. Toni, Dialectic Proof Procedures for Assumption-Based, Admissible Argumentation, *Artificial Intelligence* 170 (2006) 114–159. doi: [10.1016/j.artint.2005.07.002](https://doi.org/10.1016/j.artint.2005.07.002).
- [4] P. M. Dung, P. Mancarella, F. Toni, Computing Ideal Sceptical Argumentation, *Artificial Intelligence* 171 (2007) 642–674. doi: [10.1016/j.artint.2007.05.003](https://doi.org/10.1016/j.artint.2007.05.003).

- [5] Q. Zhong, X. Fan, X. Luo, F. Toni, An explainable multi-attribute decision model based on argumentation, *Expert Systems with Applications* 117 (2019) 42–61. doi: [10.1016/j.eswa.2018.09.038](https://doi.org/10.1016/j.eswa.2018.09.038).
- [6] P. Baroni, A. Rago, F. Toni, From Fine-Grained Properties to Broad Principles for Gradual Argumentation: A Principled Spectrum, *International Journal of Approximate Reasoning* 105 (2019) 252–286. doi: [10.1016/j.ijar.2018.11.019](https://doi.org/10.1016/j.ijar.2018.11.019).
- [7] N. Potyka, Interpreting neural networks as quantitative argumentation frameworks, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 2021, pp. 6463–6470.
- [8] A. Rago, O. Cocarascu, F. Toni, Argumentation-based recommendations: Fantastic explanations and how to find them, in: *Proceedings of the 27th International Joint Conference on Artificial Intelligence, IJCAI'18*, AAAI Press, 2018, p. 1949–1955.
- [9] N. Potyka, Extending Modular Semantics for Bipolar Weighted Argumentation, in: N. Agmon, E. Elkind, M. E. Taylor, M. Veloso (Eds.), *18th International Conference on Autonomous Agents and MultiAgent Systems, IFAAMAS*, Montreal, 2019, pp. 1722–1730.
- [10] N. Potyka, Continuous Dynamical Systems for Weighted Bipolar Argumentation, in: Michael Thielscher, F. Toni, F. Wolte (Eds.), *Principles of Knowledge Representation and Reasoning: 16th International Conference*, 2018, pp. 148–157.
- [11] J. Delobelle, S. Villata, Interpretability of Gradual Semantics in Abstract Argumentation, in: G. Kern-Isberner, Z. Ognjanovic (Eds.), *15th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, volume 11726 LNAI, Springer, Belgrade, 2019, pp. 27–38.
- [12] A. Hunter, S. Konieczny, On the measure of conflicts: Shapley Inconsistency Values, *Artificial Intelligence* 174 (2010) 1007–1026. doi: [10.1016/j.artint.2010.06.001](https://doi.org/10.1016/j.artint.2010.06.001).
- [13] A. Datta, S. Sen, Y. Zick, Algorithmic Transparency via Quantitative Input Influence: Theory and Experiments with Learning Systems, in: *2016 IEEE Symposium on Security and Privacy*, IEEE, 2016, pp. 598–617. doi: [10.1109/SP.2016.42](https://doi.org/10.1109/SP.2016.42).
- [14] S. M. Lundberg, S.-I. Lee, A Unified Approach to Interpreting Model Predictions, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, Curran Associates, 2017, pp. 4765–4774. [arXiv:1705.07874](https://arxiv.org/abs/1705.07874).
- [15] C. Labreuche, S. Fossier, Explaining Multi-Criteria Decision Aiding Models with an Extended Shapley Value, in: J. Lang (Ed.), *27th International Joint Conference on Artificial Intelligence, IJCAI*, Stockholm, 2018, pp. 331–339. doi: [10.24963/ijcai.2018/46](https://doi.org/10.24963/ijcai.2018/46).
- [16] T. Yan, A. D. Procaccia, If You Like Shapley Then You'll Love the Core, in: *35th Conference on Artificial Intelligence*, AAAI Press, 2021, pp. 5751–5759.
- [17] I. E. Kumar, S. Venkatasubramanian, C. Scheidegger, S. A. Friedler, Problems with Shapley Value-based Explanations as Feature Importance Measures, in: *37th International Conference on Machine Learning*, 2020, pp. 5491–5500. [arXiv:2002.11097](https://arxiv.org/abs/2002.11097).
- [18] A. Davies, P. Veličković, L. Buesing, S. Blackwell, D. Zheng, N. Tomašev, R. Tanburn, P. Battaglia, C. Blundell, A. Juhász, M. Lackenby, G. Williamson, D. Hassabis, P. Kohli, Advancing mathematics by guiding human intuition with AI, *Nature* 600 (2021) 70–74. doi: [10.1038/s41586-021-04086-x](https://doi.org/10.1038/s41586-021-04086-x).
- [19] P. M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, *Artificial intelligence* 77 (1995) 321–357.
- [20] B. Verheij, Two approaches to dialectical argumentation: admissible sets and argumentation stages, *Proc. NAIC 96* (1996) 357–368.
- [21] X. Fan, F. Toni, On computing explanations in argumentation, in: *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI'15*, AAAI Press, 2015, p. 1496–1492.
- [22] T. Kampik, K. Čyras, Explaining Change in Quantitative Bipolar Argumentation (to appear), in: F. Toni (Ed.), *Computational Models of Argument*, IOS Press, 2022.
- [23] A. Rago, O. Cocarascu, C. Bechlivanidis, D. Lagnado, F. Toni, Argumentative explanations for interactive recommendations, *Artificial Intelligence* 296 (2021) 103506. doi: <https://doi.org/10.1016/j.artint.2021.103506>.