Abstract. We address the problem of efficient generation of structured argumentation systems. We consider a simplified variant of an ASPIC argumentation system and provide a backward chaining mechanism for the generation of structured argumentation graphs. We empirically compare the efficiency of this new approach with existing approaches (based on forward chaining) and characterise the benefits of using backward chaining for argumentation-based query answering.

Keywords. Argumentation, Backward chaining

1. Introduction

Logic-based argumentation is a powerful approach to reasoning with conflicting pieces of information. While argumentation traditionally generates arguments, most other approaches take a defeasible theory (DT) as input, together with a query. The output of the reasoning process is whether the query is, or is not, accepted by justified conclusions of the saturated DT, taking into account the interactions between conflicting facts. Most approaches to argumentation first compute all arguments, and detect conflicts between the arguments, thereby generating an argumentation graph [14, 19, 17]. Abstract argumentation semantics [8, 6] are then used to compute sets of justified arguments (called extensions), whose conclusions are compared to the query [7]. This follows the intuitions of what is known as forward chaining (FC) [15, 1]. The departure point of this work is the observation that FC is inappropriate for certain applications. Generating the entire set of arguments and identifying its extensions is computationally expensive [19, 18, 9], and one might want to answer a single query, in which case only arguments relevant to it should be generated. Here, we propose the use of backward chaining (BC) to do so, focusing on one argumentation system (AS) based on a variant of ASPIC [4].

The focus of our approach is computational efficiency, though we recognise that in the worst case, the entire set of arguments may need to be generated to answer a query. However, we show in our empirical evaluation that this rarely occurs. We also analyse how the rules interact, proposing sufficient conditions to determine when, for a specific set of rules, fewer arguments will be generated by our approach.

1 Corresponding Author: bruno.yun@abdn.ac.uk.
Our core contributions are as the introduction of (1) a new BC mechanism for ASPIC-like arguments; (2) a combinatorial structure that characterises the benefits of using BC over FC and (3) an empirical evaluation demonstrating its potential impact.

The paper is structured as follows: In Section 2, we introduce the background notions. In Section 3, we introduce the Graph of Rule Interaction to capture how rules trigger each other. In Section 4, we describe how to generate arguments with BC. In Section 5, we show that the AS obtained via BC satisfies desirable properties and confirm them using an empirical evaluation. Finally, Section 6 summarises our approach.

2. Background

We begin by providing a brief overview of ASs built upon DTs. The notions presented here are used in the remainder of the paper. An argumentation formalism is usually built around an underlying logical language \( \mathcal{L} \). We assume that \( \mathcal{L} \) is made up of a set of literals and that we possess classical negation, i.e., we have a function \( \neg \) s.t. \( \neg \psi = \phi \iff \psi = \neg \phi \) and \( \neg \neg \psi = \psi \). A strict rule is then an expression \( \mu : \phi_1, \ldots, \phi_n \rightarrow \psi \) and a defeasible rule is an expression \( \mu : \phi_1, \ldots, \phi_n \Rightarrow \psi \) with \( n \geq 0 \) and \( \{ \phi_1, \ldots, \phi_n, \psi \} \subseteq \mathcal{L} \). For a rule \( r = \phi_1, \ldots, \phi_n \Rightarrow \psi \), where \( \Rightarrow \in \{ \rightarrow, \Rightarrow \} \), we denote by \( \text{Body}(r), \text{Head}(r), \text{Name}(r) \) and \( \text{Imp}(r) \) the set \( \{ \phi_1, \ldots, \phi_n \} \), the literal \( \psi \), the literal \( \mu \), and \( \Rightarrow \) respectively. A rule \( \phi_1, \ldots, \phi_n \Rightarrow \psi \) is said to be applicable on a set of literals \( \mathcal{P} \subseteq \mathcal{L} \) iff \( \text{Body}(r) \subseteq \mathcal{P} \). A set \( \mathcal{P} \subseteq \mathcal{L} \) is said to be consistent iff there is no \( \phi, \psi \in \mathcal{P} \) s.t. \( \phi \Rightarrow \neg \psi \). The closure of \( \mathcal{P} \) under a set of strict rules \( \mathcal{S} \), denoted by \( \text{Cl}_{\mathcal{S}}(\mathcal{P}) \), is the minimal set s.t. (1) \( \mathcal{P} \subseteq \text{Cl}_{\mathcal{S}}(\mathcal{P}) \) and (2) for all strict rules \( r = \phi_1, \ldots, \phi_n \Rightarrow \psi \in \mathcal{S} \) s.t. \( r \) is applicable to \( \text{Cl}_{\mathcal{S}}(\mathcal{P}) \), it holds that \( \psi \in \text{Cl}_{\mathcal{S}}(\mathcal{P}) \). A DT is \( \mathcal{T} = (\mathcal{S}, \mathcal{D}) \) where \( \mathcal{S} \) is a set of strict rules and \( \mathcal{D} \) is a set of defeasible rules ².

Example 1. We consider a DT containing the following information about John, a patient in a hospital: \( r_1 \): “John has a prostate cancer” \( (\rightarrow c) \); \( r_2 \): “John is following a treatment for his cancer” \( (\rightarrow t) \); \( r_3 \): “John is a male patient” \( (\rightarrow m) \); \( r_4 \): “Studies show that there is no correlation between treating prostate cancer and bowel disorders” \( (\rightarrow \neg r_7) \); \( r_5 \): “John does not have abdominal pains” \( (\Rightarrow a) \); \( r_6 \): “If John does not have abdominal pain then he may not suffer from bowel disorder” \( (a \Rightarrow \neg b) \); \( r_7 \): “A patient with prostate cancer that is under treatment may suffer from bowel disorders” \( (c \Rightarrow b) \). The DT is modelled by \( \mathcal{T} = (\mathcal{S}, \mathcal{D}) \) s.t. \( \mathcal{S} = \{ r_1, r_2, r_3, r_4 \} \) and \( \mathcal{D} = \{ r_5, r_6, r_7 \} \).

Our AS is based on a version of ASPIC [4]. Here, a rule is a rule application in a deductive form [3] built upon other arguments. Given \( \mathcal{T} = (\mathcal{S}, \mathcal{D}) \), an argument \( \mathcal{A} \) is of the form \( \mathcal{A}_1, \ldots, \mathcal{A}_n \Rightarrow \psi \), where \( \{ \mathcal{A}_1, \ldots, \mathcal{A}_n \} \) is a minimal set of arguments s.t. there exists an \( r \in \mathcal{S} \cup \mathcal{D} \), where \( r \) is applicable to \( \{ \text{Conc}(\mathcal{A}_1), \ldots, \text{Conc}(\mathcal{A}_n) \} \), \( \Rightarrow \Rightarrow \text{Imp}(r) \) and \( \psi = \text{Head}(r) \). Let \( \mathcal{A} = \mathcal{A}_1, \ldots, \mathcal{A}_n \Rightarrow \psi \). Then the conclusion of \( \mathcal{A} \), denoted by \( \text{Conc}(\mathcal{A}) \), is \( \psi \) and the set of sub-arguments of \( \mathcal{A} \) is \( \text{Sub}(\mathcal{A}) = \text{Sub}(\mathcal{A}_1) \cup \cdots \cup \text{Sub}(\mathcal{A}_n) \cup \{ \mathcal{A} \} \). The top rule of \( \mathcal{A} \) is \( \text{TR}(\mathcal{A}) = \text{Conc}(\mathcal{A}_1), \ldots, \text{Conc}(\mathcal{A}_n) \Rightarrow \psi \). Given \( \mathcal{T} = (\mathcal{S}, \mathcal{D}) \), Name : \( \mathcal{S} \cup \mathcal{D} \rightarrow \mathcal{L} \) returns the name of each rule and provides a handle for rules to prevent other rule applications.

²Please note that in our formalism, axioms (resp. ordinary premises) [14] are represented using strict (resp. defeasible) rules with empty bodies (c.f., ASPIC+ [5])
We consider that we are given a binary, total, reflexive and transitive preference relation ≤ on arguments that reflects the quality of the pieces of information involved in it. One way of computing such a relation is to consider the type of rules (defeasible or strict) that are used in an argument and/or that defeasible rules have an associated strength, and lifting these preferences from rules to arguments [14]. We write \( A \prec B \) iff \( A \preceq B \) and \( B \not\preceq A \), and \( A \sim B \) iff \( A \preceq B \) and \( B \preceq A \). A defeats B iff \( B \preceq A \) and at least one of the following conditions is satisfied: (Rebutting) there exists \( B' \in \text{Sub}(B) \) such that \( \text{Conc}(B') = \neg\text{Conc}(A) \) and \( \text{Imp}(TR(B')) \) is \( \Rightarrow \) or (Undercutting) \( \text{Conc}(A) = \neg\text{Name}(TR(B)) \). An AS for a \( \mathcal{T} \), denoted \( \mathcal{A}S_{\mathcal{T}} \), is \( (\mathcal{A}, \text{Def}) \) where \( \mathcal{A} \) is the set of arguments generated and \( \text{Def} \subseteq \mathcal{A} \times \mathcal{A} \) is the defeat relation introduced above. Let \( \mathcal{A}S = (\mathcal{A}, \text{Def}) \), we say that \( \mathcal{A}S = (\mathcal{A'}', \text{Def}') \) is a sub-system of \( \mathcal{A}S \) iff \( \mathcal{A} \subseteq \mathcal{A}' \) and \( \text{Def} \subseteq \mathcal{A}' \times \mathcal{A}' \cap \text{Def} \).

Example 3 (Cont’d). If we assume that \( A_7 \prec A_6 \) and \( A_6 \preceq A_5 \) then \( A_6 \) is defeated by \( A_5 \) (undercutting) and \( A_6 \) defeats \( A_7 \) (rebutting) \( A_7 \) but \( A_7 \) does not defeat \( A_6 \). The AS corresponding to \( \mathcal{T} \) is \( \mathcal{A}S_{\mathcal{T}} = ((A_1, \ldots, A_7), \{(A_5, A_6), (A_6, A_7)\}) \).

3. The Graph of Rule Interaction

We define a new combinatorial structure over a DT called the Graph of Rule Interaction (GRI) that generalises of the Graph of Rule Dependency [2]. The GRI captures both how rules trigger each other and identifies the possible conflicts between them. There are three main elements to the GRI: (1) a set of nodes representing the rules, (2) a set of support links representing how rules can activate each other and (3) a set of attack links representing conflicts amongst rules. While attacks links are binary, support links are many-to-one relationships as multiple rules can be necessary to trigger a rule.

Definition 1 (Graph of Rule Interaction). Let \( \mathcal{T} = (\mathcal{P}, \mathcal{D}) \). The GRI of \( \mathcal{T} \) is \( \text{GRI}_{\mathcal{T}} = (\mathcal{N}, \mathcal{R}, \mathcal{R}_d) \), where: (A) \( \mathcal{N} = \mathcal{P} \cup \mathcal{D} \cup \{\emptyset\} \) represents the rules under consideration; (B) \( \mathcal{R} \subseteq \mathcal{N} \times \mathcal{N} \) s.t. \( \forall n_1 \in \mathcal{N} \) and \( \forall n \subseteq \mathcal{N} \), \( (n, n_1) \in \mathcal{R} \) iff \( \|n\| = \|\text{Body}(n_1)\| \) and \( \bigcup_{n \subseteq \mathcal{N}} \text{Head}(n) = \text{Body}(n_1) \). \( \mathcal{R} \) captures the support between rules and (C) \( \mathcal{R}_d \subseteq \mathcal{N} \times \mathcal{N} \) s.t. \( \forall n_1, n_2 \in \mathcal{N} \), \( (n_1, n_2) \in \mathcal{R}_d \) iff at least one of the following conditions holds: (1) \( \text{Head}(n_1) = \neg\text{Head}(n_2) \) and \( \text{Imp}(n_2) \Rightarrow \) or (2) \( \text{Head}(n_1) = \neg\text{Name}(n_2) \). \( \mathcal{R}_d \) captures potential defeats between rules. Note that \( \text{Head}(\emptyset) = \text{Body}(\emptyset) = \text{Name}(\emptyset) = \emptyset \).

Chains of rules can form where one rule is required for another to be applied, meaning that multiple rules can support others. We formalise this notion within the GRI.

Definition 2 (Support path). Let \( \mathcal{T} = (\mathcal{P}, \mathcal{D}) \) and \( \text{GRI}_{\mathcal{T}} = (\mathcal{N}, \mathcal{R}, \mathcal{R}_d) \). The sequence \( (\mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_k) \) is a support path to \( n \in \mathcal{P} \cup \mathcal{D} \) in \( \text{GRI}_{\mathcal{T}} \) iff all the following conditions are satisfied: (1) \( \forall i \in \{1, \ldots, k\}, \mathcal{N}_i \subseteq 2^\mathcal{N} \), (2) \( \mathcal{N}_1 = \{\emptyset\} \) and \( \mathcal{N}_k = \{n\} \) and (3) \( \forall i \in \{2, \ldots, k\}, \forall N_j \in \mathcal{N}_i \) and \( \forall n' \in \mathcal{N}_j \), there exists \( N' \in \mathcal{N}_{i-1} \) s.t. \( (N', n') \in \mathcal{R}_d \).

An activated rule is a rule with a support path to it. The underlying idea is that such a rule’s body can be obtained from other rules via the support path to it.

Note that we use restricted rebut and we did not evaluate unrestricted rebut due to space limitations.
Example 4. Backward Chaining for Argumentation

Let us first define the notion of an AS for a literal to characterise how the BC procedure for a literal is defined for the formalism. An argument for \( l \in \mathcal{L} \) is an argument that concludes \( l \). This notion is needed to define what an AS for a literal is.

Definition 5 (Argument for a literal). Let \( l \in \mathcal{L} \) and \( \mathcal{F} = (\mathcal{L}, \mathcal{D}) \). An argument for \( l \in \mathcal{L} \) is an argument for \( l \) iff \( \text{Conc}(A) = l. \mathcal{A} \subseteq \mathcal{L} \) is the set of arguments for \( l \) in \( \mathcal{A} \).

Definition 6 (AS for a literal). Let \( l \in \mathcal{L} \) and \( \mathcal{A} \subseteq \mathcal{L} \). We say that the sub-system \( \mathcal{A}S_{\mathcal{F}} = (\mathcal{A}', \text{Def}') \) of \( \mathcal{A}S_{\mathcal{F}} \) is the AS for \( l \) (w.r.t. \( \mathcal{F} \)) iff \( \mathcal{A}' \) is minimal (w.r.t. \( \subseteq \)) s.t. all the following are satisfied: (1) \( \mathcal{A}' \subseteq \mathcal{A} \), (2) If \( A \in \mathcal{A} \) and \( B \in \mathcal{A} \) s.t. \( (B,A) \in \text{Def} \) then \( B \in \mathcal{A}' \), (3) If \( A \in \mathcal{A} \) then \( \text{Sub}(A) \subseteq \mathcal{A}' \) and (4) \( \text{Def}' = \{(A,B) \in \text{Def} | A,B \in \mathcal{A}'\} \).

Note that for a given DT \( \mathcal{F} \), there is a unique AS for \( l \in \mathcal{L} \) w.r.t. \( \mathcal{F} \). Moreover, since \( \mathcal{A}S_{\mathcal{F}} = (\mathcal{A}', \text{Def}') \) is a sub-system of \( \mathcal{A}S_{\mathcal{F}} = (\mathcal{A}, \text{Def}) \), it has at most the same number of arguments and attacks, i.e. \( |\mathcal{A}'| \geq |\mathcal{A}| \) and \( |\text{Def}'| \geq |\text{Def}| \).

Example 5. Let \( \mathcal{A}S_{\mathcal{F}} = (\mathcal{A}', \text{Def}') \) where \( \mathcal{A}' = \{A_1,A_2,A_4,A_5,A_6,A_7\} \) and \( \text{Def}' = \{(A_5,A_6), (A_6,A_7)\} \). \( \mathcal{A}S_{\mathcal{F}} \) is the AS for \(-b \) w.r.t. \( \mathcal{F} \).

We now focus on the generation of arguments from a DT, describing a two-step procedure for generating the AS using BC, i.e. from \( \mathcal{F} \), we use BC to generate \( \mathcal{A}S_{\mathcal{F}} \), for any \( l \in \mathcal{L} \). Our approach makes use of two algorithms: AL and ASG (see Figure 1).

AL: This algorithm generates \( \mathcal{A} \). This procedure gathers all the rules that conclude \( l \) in \( R \) (line 5). It then constructs all the possible arguments that have a rule of \( R \) as top rule. (lines 6–9). Note that the parameter \( \text{seen} \) is used to avoid the generation of an infinite number of sub-arguments in case of rule cycles (for instance \( p \rightarrow q \) and \( q \rightarrow p \)).
This algorithm generates \( \mathcal{A}_f \). It starts by generating \( \mathcal{A}_f \), after which it successively adds sub-arguments and attacking arguments to the existing ones (lines 10-12). Once the arguments have been generated, \( \text{DEF-generate} \) computes the defeat relation (see Section 2) by comparing pairs of arguments. As attacking arguments are not always defeaters (depending on the \( \succeq \) relation chosen), we have to filter unnecessary arguments that do not fit the conditions of Definition 6 by using \( \text{AS-filter} \) (line 18).

5. Backward Chaining Evaluation

Let us first recall the usual argumentation semantics defined in Dung’s seminal paper [8]. Let \( \mathcal{A}_x \) be an AF and \( \mathcal{E} \subseteq \mathcal{A}_x \). We say that \( \mathcal{E} \) is conflict-free iff there is no \( a, b \in \mathcal{E} \) s.t. \( (a, b) \in \text{DEF} \). \( \mathcal{E} \) defends \( a \) iff for every \( b \in \mathcal{A}_x \) s.t. \( (b, a) \in \text{DEF} \), there exists \( c \in \mathcal{E} \) s.t. \( (c, b) \in \text{DEF} \). \( \mathcal{E} \) is admissible iff it is conflict-free and defends all its arguments. \( \mathcal{E} \) is a complete extension iff \( \mathcal{E} \) is admissible and contains all the arguments it defends. \( \mathcal{E} \) is a preferred extension iff it is a maximal (w.r.t. \( \subseteq \)) admissible set. \( \mathcal{E} \) is the grounded extension iff \( \mathcal{E} \) is a minimal (w.r.t. \( \subseteq \)) complete extension. We denote by \( \text{Ext}(x, y) \) the function that returns the set of extensions of the AS \( x \) w.r.t. \( y \), where \( y \in \{pr, gr\} \) and \( pr \) (resp. gr) stands for the preferred (resp. grounded) semantics. Likewise, \( \text{Acc}(x, y) \) returns the accepted arguments of the AS \( x \) w.r.t. the semantics \( y \).

**Definition 7** (Status). Let \( \mathcal{A}_x = (\mathcal{A}_x, \text{DEF}) \) and \( a \in \mathcal{A}_x \). \( a \) is accepted w.r.t. the preferred semantics (resp. grounded semantics) if \( \forall E \in \text{Ext}(\mathcal{A}_x, pr) \) (resp. \( \text{Ext}(\mathcal{A}_x, gr) \)), \( a \notin E \). \( a \) is rejected w.r.t. the preferred semantics (resp. grounded semantics) if \( \forall E \in \text{Ext}(\mathcal{A}_x, pr) \) (resp. \( \text{Ext}(\mathcal{A}_x, gr) \)), \( a \notin E \) and \( a \) is undecided if it is neither accepted nor rejected.

**Definition 8** (Acceptability of a literal). Let \( l \in \mathcal{L} \), \( \mathcal{T} \) and \( \mathcal{A}_x = (\mathcal{A}_x, \text{DEF}) \). \( l \) is accepted w.r.t. the preferred (resp. grounded) semantics and \( \mathcal{A}_x \) iff there exists an \( a \in \mathcal{A}_x \) s.t. \( \text{Conc}(a) = l \) and \( a \in \text{Acc}(\mathcal{A}_x, pr) \) (resp. \( a \in \text{Acc}(\mathcal{A}_x, gr) \)). Otherwise, \( l \) is rejected.
Example 6 (Cont’d), \( \text{Ext}(\mathcal{A}S_F^y, pr) = \text{Ext}(\mathcal{A}S_F^{gr}) = \{ \{A_1, A_2, A_3, A_4, A_5, A_7\} \} \) and \( \text{Ext}(\mathcal{A}S_F^y, pr) = \{ \{A_1, A_2, A_4, A_5, A_7\} \} \). Since \( A_7 \in \text{Acc}(\mathcal{A}S_F^y, pr) \), \( \neg b \in \text{Acc}(\mathcal{A}S_F^y, pr) \). Note that \( m \) is accepted w.r.t. preferred/grounded in \( \mathcal{A}S_F \) but rejected in \( \mathcal{A}S_F^y \).

Proposition 1. Let \( l \in \mathcal{L} \), \( \mathcal{A}S_F \) be the AS for \( \mathcal{T} \) and \( \mathcal{A}S_F^y \) be the AS for \( l \). It holds that \( l \in \text{Acc}_{\mathcal{T}}(\mathcal{A}S_F^y, y) \) iff \( l \in \text{Acc}_{\mathcal{T}}(\mathcal{A}S_F, y) \), where \( y \in \{pr, gr\} \).

We now show that (1) in specific DTs (characterised via a sufficient condition) the AS for a literal has strictly fewer arguments than the corresponding original AS, (2) the rules that are not activated are not taken into account when constructing the arguments of an AS and (3) the GRI can be used to filter a DT \( \mathcal{T} \) prior to the generation of \( \mathcal{A}S_F \).

Proposition 2. Let \( \mathcal{T} = (\mathcal{A}, \mathcal{D}) \), \( l \in \mathcal{L} \), \( \mathcal{A}S_F = \{\mathcal{A}, \mathcal{DEF}\} \), \( \mathcal{A}S_F^y = \{\mathcal{A}', \mathcal{DEF}'\} \) be the AS for \( l \). If there exists \( r \in \mathcal{T} \cup \mathcal{D} \) such that:

- \( r \) is activated and not potentially necessary for \( l \) then |\( \mathcal{A}' \)| < |\( \mathcal{A} \)|.
- \( r \) is not activated then \( \mathcal{A}S_{\mathcal{T}} = \mathcal{A}S_{\mathcal{D}} \), where \( \mathcal{T}' = (\mathcal{T} \setminus \{r\}, \mathcal{D} \setminus \{r\}) \).
- \( r \) is not potentially necessary for \( l \) then \( \mathcal{A}S_{\mathcal{T}} = \mathcal{A}S_{\mathcal{D}} \), where \( \mathcal{A}S_{\mathcal{T}} \) is the AS for \( l \) w.r.t. \( \mathcal{T}' = (\mathcal{T} \setminus \{r\}, \mathcal{D} \setminus \{r\}) \).

The third item of Proposition 2 shows that filtering DTs to only keep potentially necessary rules is possible when generating the AS for a literal. If \( \mathcal{T} \) contains rules that are not potentially necessary for a literal, this filtering reduces the time taken to answer a query. Moreover, the GRI only has to be computed once, and can then be stored in memory and reused for multiple queries. The proposed framework is inspired from previous approaches based on DT pre-processing [18].

5.1. Empirical Evaluation

To test our approach, we use existing benchmarks and DTs to compare the effectiveness of reasoning using BC and FC in the context of argumentation. To this end, we use existing DTs [12] as we are not aware of other standard benchmarks for instantiated argumentation. Due to space constraints, we only considered four theories from that work (tree, level, levels and teams). In tree(n,k), the rules form a k-branching tree of depth n where the literal \( p_0 \) is the root. In these theories, every literal occurs only once. In level(n), there is a cascade of \( n \) disputed conclusions, i.e., there are rules \( \Rightarrow p_i \) and \( p_{i+1} \Rightarrow \neg p_i \), for \( 0 \leq i \leq n \). In levels(n), for odd \( i \), the latter rule has a superior strength when compared to even rules. Finally, in teams(n), every literal is disputed with two rules for \( p_i \) and two rules for \( \neg p_i \), and the rules for \( p_i \) are superior to the rules for \( \neg p_i \).

To obtain the \( \leq \), we use the last-link principle described in [14].

For the FC procedure, we generated all arguments (5 times) using a breadth-first naive approach. For the BC procedure, for all theories, we randomly selected ten literals and generated the ASs for those literals. An upper limit of 200 minutes was set for all runs. Table 1 is split in three parts: FC, BC and DT filtration (by removing non-potentially necessary rules). In the Forward columns, we depict the mean time, the number of arguments generated, and the number of defeats of the graph\(^4\). In the Backward columns,
we show, across all literals on non timed out instances, the mean time for the generation of arguments, the mean number of arguments in the AS for the literal and the number of successful (non timeout) instances. In the Filter columns, we show the number of rules after the filtration, the percentage of the number of rules filtered, and the mean time used for obtaining the filtered DT. We make three important observations: (1) For the tree and teams DTs, all the runs were successful and the BC was significantly faster in generating the arguments than the FC. It is worth noting that while the FC procedure times out after \( n = 9 \), the BC procedure is able to provide an answer in less than 5 minutes. (2) For the level DTs, the BC takes longer than the FC (even if it generates fewer arguments). From \( n = 5000 \) onward, most instances timeout. Note that the BC takes longer than the FC for these instances because it checks and generates all the arguments that can potentially attack the existing arguments. (3) In the tree and teams DTs, we obtain fewer arguments with the BC compared to the FC. The gap in the number of arguments means that the process of verification does not cause a serious overhead in the computation time.

6. Discussion and Future Work

We introduced the notion of BC argumentation and illustrated our approach with an ASPIC-style structured AS. We analysed the links between the AS generated using the BC procedure and the FC procedure w.r.t. argumentation semantics and showed an empirical comparison of the time needed to generate the arguments for both procedures.

Our work is motivated by the need for efficient query answering frameworks that do not need to generate the whole set of arguments [19, 18]. Our work relates with existing BC-based works such as DeLP [10] or ABA [16]. However, our focus is explicitly on ASPIC-like systems. There are also similarities between BC and proof dialogues [13], though most such dialogues operate on abstract ASs.

We have identified several potential avenues of future work. First, we intend to create additional benchmarks for instantiated ASs by replicating the properties of existing DTs [11]. Second, we recognise that there are similarities between the process we use, and

<table>
<thead>
<tr>
<th>Theory</th>
<th>Mean time (s)</th>
<th># args</th>
<th># defeats</th>
<th>Mean time (s)</th>
<th>Mean # arguments</th>
<th># Succ. instances</th>
<th>Mean # rules</th>
<th>% Filtration</th>
<th>Mean time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Srebnik</td>
<td>( n = 8, k = 3 )</td>
<td>5712.7</td>
<td>59349</td>
<td>0</td>
<td>19.6</td>
<td>15.8</td>
<td>10</td>
<td>11</td>
<td>99.93%</td>
</tr>
<tr>
<td>Timeout</td>
<td>1968.50</td>
<td>0</td>
<td>53.5</td>
<td>3.5</td>
<td>10</td>
<td>3.5</td>
<td>99.99%</td>
<td>810.5</td>
<td></td>
</tr>
<tr>
<td>Level(n)</td>
<td>( n = 10 )</td>
<td>6.7</td>
<td>1999</td>
<td>2996</td>
<td>0.03</td>
<td>10.8</td>
<td>10</td>
<td>10.8</td>
<td>43.16%</td>
</tr>
<tr>
<td>( n = 1000 )</td>
<td>181.7</td>
<td>9999</td>
<td>14996</td>
<td>3927.2</td>
<td>602.3</td>
<td>3</td>
<td>3386</td>
<td>66.14%</td>
<td>302.30</td>
</tr>
<tr>
<td>( n = 10000 )</td>
<td>678.9</td>
<td>19999</td>
<td>29996</td>
<td>Timeout</td>
<td>-</td>
<td>0</td>
<td>8423</td>
<td>57.88%</td>
<td>490.40</td>
</tr>
<tr>
<td>Level(n)</td>
<td>( n = 10 )</td>
<td>6.7</td>
<td>1999</td>
<td>1998</td>
<td>1245.3</td>
<td>841.2</td>
<td>10</td>
<td>841.2</td>
<td>57.92%</td>
</tr>
<tr>
<td>( n = 5000 )</td>
<td>155.9</td>
<td>9999</td>
<td>9998</td>
<td>96542.02</td>
<td>3702</td>
<td>2</td>
<td>5984.4</td>
<td>41.95%</td>
<td>461.00</td>
</tr>
<tr>
<td>( n = 10000 )</td>
<td>696.8</td>
<td>19999</td>
<td>19999</td>
<td>428.5</td>
<td>163</td>
<td>1</td>
<td>10782.6</td>
<td>46.0%</td>
<td>555</td>
</tr>
<tr>
<td>Team(n)</td>
<td>( n = 3 )</td>
<td>0.4</td>
<td>176</td>
<td>272</td>
<td>0.26</td>
<td>3.1</td>
<td>10</td>
<td>3.1</td>
<td>97.89%</td>
</tr>
<tr>
<td>( n = 4 )</td>
<td>1.6</td>
<td>736</td>
<td>1568</td>
<td>1.62</td>
<td>19.9</td>
<td>10</td>
<td>17.1</td>
<td>97.1%</td>
<td>131.2</td>
</tr>
<tr>
<td>( n = 5 )</td>
<td>26.8</td>
<td>3008</td>
<td>8256</td>
<td>5.28</td>
<td>23.8</td>
<td>10</td>
<td>20.6</td>
<td>99.14%</td>
<td>198.3</td>
</tr>
<tr>
<td>( n = 6 )</td>
<td>539.7</td>
<td>12160</td>
<td>254335</td>
<td>18.35</td>
<td>3.8</td>
<td>10</td>
<td>3.8</td>
<td>99.96%</td>
<td>369.2</td>
</tr>
<tr>
<td>( n = 7 )</td>
<td>11613.2</td>
<td>48896</td>
<td>1401159</td>
<td>84.07</td>
<td>14.1</td>
<td>10</td>
<td>12.9</td>
<td>99.97%</td>
<td>866.5</td>
</tr>
</tbody>
</table>

Table 1. Summary of the empirical evaluation
different search algorithms. We intend to evaluate these different strategies, as well as heuristics for guiding the expansion process, and their effects on performance. Finally, integrating lifting rules for preferences (e.g. weakest link, elitist or democratic orderings [14]) could provide optimisations regarding argument expansion.

References