An Investigation of Argumentation Framework Characteristics

Josh Murphy, Isabel Sassoon, Michael Luck, Elizabeth Black

Department of Informatics, King's College London

firstname.lastname@kcl.ac.uk

Abstract. We investigate the relationship between the structural properties of argumentation frameworks and their argument-based characteristics, examining the characteristics of structures of Dung-style frameworks and two generalisations: extended argumentation frameworks and collective-attack frameworks. Our results show that the structural properties of frameworks have an impact on the size of extensions produced, on the proportion of subsets of arguments that determine some topic argument to be acceptable, and on the likelihood that the addition of some new argument will affect the acceptability of an existing argument, all characteristics that are known to affect the performance of argumentation-based technologies. We demonstrate the applicability of our results with two case studies.

1 Introduction

Argumentation is a key sub-field of AI that provides an intuitive reasoning mechanism for dealing with inconsistent, uncertain and incomplete knowledge. A set of arguments and the relationships between them can be represented as a directed graph (referred to as an *argumentation framework*) to which one of a number of *semantics* can be applied to determine which arguments it is coherent to accept [12]. While progress has been made in the development of argumentation-based technologies (*e.g.*, argument solvers [5] and real-world applications [18]) realistic evaluations of such technologies is difficult, due to the shortage of repositories of argumentation frameworks that are representative of real-world domains [10]; typically, argument technologies are evaluated on randomly generated frameworks, with little understanding of how the *structure* of such frameworks impacts on performance. It has been shown that structural differences in argumentation frameworks can affect the performance of argumentation-based technologies, such as dialogue systems [3] and argument solvers [2]. We argue here that a better understanding of these effects can not only allow for a more thorough evaluation, but can also inform development of technologies that are optimised for specific framework structures.

In order to explore the characteristics of argumentation frameworks with different structural properties, we consider the classic Dung-style argumentation frameworks (which represent attacks between arguments) [12] and two generalisations of these that each have their own particular structural traits: extended

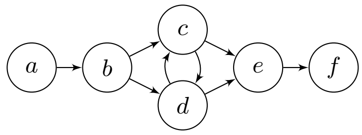

Fig. 1. An example DAF.

argumentation frameworks (which allow an argument to attack the attack between two arguments) [17] and collective-attack frameworks (which allow attacks from sets of arguments) [22]). We investigate three key characteristics.

1. The size of the set of acceptable arguments generated by the grounded and preferred sceptical semantics. This characteristic is known to affect performance of argument solvers [8];
2. The proportion of argument subsets of the framework that determine some topic argument to be acceptable. This characteristic is known to be a factor in the effectiveness of strategies for persuasion [3];
3. Whether the addition of a new argument to the framework results in a change of acceptability of some topic argument. This is a type of dynamic argumentation, another factor in the effectiveness of dialogue strategies [1], and also may be a key property for improving the computational efficiency of a variety of other argument technologies [16].

We demonstrate applicability of our results with two case studies: a Dung-style framework from a decision-making tool for aggregating the effects of medical treatment [15], and an extended framework from a statistical model selection tool in a clinical domain [25].

2 Argumentation frameworks (AFs)

Since Dung's seminal work [12], the dominant approach to argumentation-based reasoning is to represent arguments as abstract entities in an argumentation framework that captures the relationships between them, and then to apply one of several argumentation semantics to determine which subsets of arguments it is rational to present as a coherent set. We now define *Dung-style argumentation frameworks* [12] (DAFs), which capture attacks between arguments.

Definition 1. *A* ***Dung-style argumentation framework (DAF)*** *is a pair $\langle A, R \rangle$ s.t. A is a finite set of arguments and $R \subseteq A \times A$ is a set of attacks. $(x, y) \in R$ means x attacks y.*

Argumentation semantics are based on the intuitive principles that it is not rational to accept any two conflicting arguments, and that an argument which is attacked can only be accepted if all of its attacking arguments are themselves attacked by an accepted argument [12].

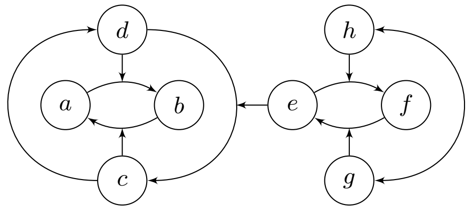

Fig. 2. An example EAF, which is also a HEAF.

Definition 2. *Let $\langle A, R\rangle$ be a DAF and $S \subseteq A$.*
- *S is* **conflict-free** *iff $\forall a, b \in S$: $(a, b) \notin R$.*
- *$a \in A$ is* **acceptable** *w.r.t. S iff $\forall b$ s.t. $(b, a) \in R$: $\exists c \in S$ s.t. $(c, b) \in R$.*
- *S is* **admissible** *iff S is conflict-free and each argument in S is acceptable w.r.t. S.*

There are a range of different semantics that build on these principles and determine sets of arguments that can rationally be presented as coherent, known as **extensions**. Here we consider two semantics: an argument is acceptable under the **preferred sceptical semantics** if it is part of all maximal admissible sets; an argument is acceptable under the **grounded semantics** if it is in the smallest set S such that every argument that is acceptable w.r.t. S is in S. In the DAF shown in Figure 1, a and f are the only arguments that are acceptable under the preferred sceptical semantics, while a is the only argument acceptable under the grounded semantics.

Example 1. Considering the DAF in Figure 1, the only argument acceptable under the grounded semantics is a, whereas the arguments a and f are acceptable under the preferred sceptical semantics.

Though Dung-style argumentation frameworks are expressive, many generalisations have been proposed which provide explicit representation of relationships other than attacks between arguments, seeking to intuitively capture particular aspects of argumentation [4]. *Extended argumentation frameworks* (EAFs) allow the representation of arguments that attack attack relations [17]. Thus, given an argument a which attacks b, an argument c may attack the attack between a and b. In this way, an EAF may be used to capture (possibly conflicting) preference relations between arguments. For example, see Figure 2 in which c represents a preference for a over b, which conflicts with d representing a preference for b over a. EAFs are an especially expressive model as they represent preferences as defeasible arguments, allowing agents to argue about their preferences and, powerfully, about preferences over other preferences.

Definition 3. *An* ***extended argumentation framework (EAF)*** *is a tuple $\langle A, R, D\rangle$ s.t. A is a finite set of arguments, $R \subseteq A \times A$ is a set of attacks,*
- *$D \subseteq A \times R$ is a set of attacks on attacks, and*
- *if $(z, (x, y)), (z', (y, x)) \in D$ then $(z, z'), (z', z) \in R$.*

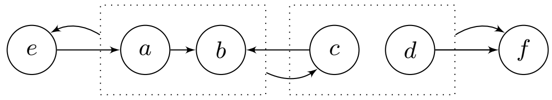

Fig. 3. An example CAF.

We especially consider here *hierarchical EAFs* (HEAFs), a particularly interesting class of EAFs that can be used to formalise practical reasoning [17].

Definition 4. *An EAF $\langle A, R, D\rangle$ is a* ***hierarchical extended argumentation framework (HEAF)*** *iff there exists a partition $P = [\langle\langle A_1, R_1\rangle, D_1\rangle, ..., \langle\langle A_j, R_j\rangle, D_j\rangle, ...]$ s.t.:*

- *$A = \cup_{i=1}^{\infty} A_i, R = \cup_{i=1}^{\infty} R_i, D = \cup_{i=1}^{\infty} D_i$, and for $i = 1, ..., \infty$, $\langle A_i, R_i\rangle$ is a DAF.*
- *If $(z, (x, y)) \in D_i$ then $(x, y) \in R_i, z \in A_{i+1}$*

We refer to an argument a as being in a lower *partition than an argument b if $a \in A_p$, $b \in A_q$, and $p < q$.*

The arguments in Figure 2 can be partitioned into 4 levels: $\{a, b\}$, $\{c, d\}$, $\{e, f\}$, and $\{g, h\}$, where $\{a, b\}$ is the lowest partition and $\{g, h\}$ is the highest.

EAF argumentation semantics are defined equivalently as for DAFs, with the following adjustments [17].

Definition 5. *Let $\langle A, R, D\rangle$ be an EAF and $S \subseteq A$.*

- *a* **defeats**$_S$ *b (also written as $a \rightarrow^S b$) iff $(a, b) \in R$ and $\nexists c \in S$ s.t. $(c, (a, b)) \in D$.*
- *S is* **conflict-free** *iff $\forall a, b, \in S$: if $(a, b) \in R$ then $(b, a) \notin R$ or $\exists c \in S$ s.t. $(c, (a, b)) \in D$.*
- *$R_S = \{x_1 \rightarrow^S y_1, \dots, x_n \rightarrow^S y_n\}$ is a* **reinstatement set** *for $c \rightarrow^S b$ iff: (i) $c \rightarrow^S b \in R_S$; (ii) $\forall i \in \{1, \dots, n\}$: $x_i \in S$, and (iii) $\forall x \in R_s$, $\forall y'$ s.t. $(y', (s, y)) \in D$: $\exists x' \rightarrow^S y' \in R_S$.*
- *$a \in A$ is* **acceptable** *w.r.t. S iff $\forall b$ s.t. $b \rightarrow^S a$: $\exists c \in S$ s.t. $c \rightarrow^S b$ and there is a reinstatement set for $c \rightarrow^S b$.*

Collective-attack frameworks (CAFs) allow the representation of sets of arguments that attack an argument [22]. They can allow for a more intuitive representation of common-sense reasoning and human dialogues and have been shown to be useful in practical applications of argumentation [21,23]. See Figure 3, in which there are three collective attacks: the set of arguments $\{a, b\}$ attacks the argument c, $\{a, b\}$ attacks e, and $\{c, d\}$ attacks f.

Definition 6. *A* ***collective-attack framework (CAF)*** *is a pair $\langle A, R\rangle$ s.t. A is a finite set of arguments, and $R \subseteq (2^A \backslash \{\emptyset\}) \times A$ is a set of attacks where $(X, y) \in R$ is an attack from the set of arguments X to the argument y.*

Similarly to EAFs, CAF argumentation semantics are defined equivalently as for DAFs but with the following adjustments [22].

Definition 7. *Let* $\langle A, R\rangle$ *be a CAF and* $S \subseteq A$.
- S *is* **conflict-free** *iff* $\nexists a \in S$ *s.t.* $\exists S' \subseteq S$ *s.t.* $(S', a) \in R$.
- $a \in A$ *is* **acceptable** *w.r.t.* S *iff* $\forall B \subseteq A$ *s.t.* $(B, a) \in R$: $\exists b \in B$, $\exists S' \subseteq S$ *s.t.* $(S', b) \in R$.

3 Structural properties of AFs

There are many different structural properties of DAFs, HEAFs and CAFs we could investigate. Here we consider the DAF attack density, the distribution of arguments across the different levels of a HEAF, and the restriction on the number of arguments that may appear in a CAF collective-attack set. Our analysis of the characteristics of these different structural properties (Section 4) provides valuable insights for understanding their impact on the performance of argument technologies such as argument solvers or dialogue systems, particularly for domains or applications in which the structural properties we consider here are typical. Our case studies (Section 5) demonstrate the applicability of two of the structural classes we consider. More generally, our results show there is significant difference in the characteristics of different structural classes of AFs, which it can be important to consider when developing argument technologies or selecting the most appropriate AF representation (*e.g.*, [9]).

3.1 DAF attack density

Attack density of a DAF is the ratio of attack relations to the number of arguments. A framework with many attacks with respect to the number of arguments is *dense*, while a framework with fewer attacks is *sparse*.

Definition 8. *An* ***n-sparse DAF (n-DAF)*** *is a DAF* $\langle A, R\rangle$ *s.t.* $|R| = \frac{|A|}{n}$, *where* $n \in [0, 1]$.

We investigate 0.25-DAFs, 0.5-DAFs and 0.75-DAFs. Note that as n increases, the framework becomes more sparse. Note also that the number of attacks in the framework is linearly related to the number of arguments in the frameworks. We found in initial testing that if the number of attacks is tied instead to the number of possible attacks in the graph (which increases exponentially with the number of arguments) small changes in sparseness value produce very sharp changes in the characteristics of that structural class of DAF; linearly relating the number of attacks to arguments allows us to explore this relationship more finely.

We also consider a class of DAFs that correspond to *minimum-spanning trees* (mst-DAFs), which is a fully connected DAF in which the number of attacks is linearly related to the number of arguments ($|R| = |A| - 1$).

Definition 9. *A* **mst-DAF** *is a DAF* $\langle A, R\rangle$ *such that* $\langle A, R\rangle$ *is a* minimum-spanning tree *of* $\langle A, A \times A\rangle$.

3.2 Distributed HEAFs

In some domains, particularly human dialogues, it seems reasonable to assume that the number of arguments will be higher than the number of preferences over those arguments, which will be higher than the number of preferences over preferences, *etc.* We consider two different distributions of the proportion of arguments that appear in the different HEAF partitions. For *normally-distributed HEAFs* (nEAFs), we use the binomial coefficient to approximate the normal distribution (continuous) over a finite number of partitions (discrete), and thus the proportions with which to assign arguments to each partition. We use the number of partitions relative to the number of arguments in the graph that allows for the best fit with the normal distribution (computed with Sturges' formula [26]). The choice of normal distribution provides the desired trend of decreasing proportions, and is somewhat common in nature [14].

Definition 10. *The discrete normal distribution over l partitions is given by the formula* $\mathsf{norm\_dist}(l) = [d_0, d_1, ..., d_{l-1}]$ *s.t.:*
- $n = 2l - 1$
- $d_k = \frac{n!}{k!(n-k)!}$

The proportional weights of the partitions are thus given by the formula $\mathsf{norm\_prop}(l) = [p_0, p_1, ..., p_{l-1}]$ *such that* $p_i = 2(d_i) \div 2^n$*.*

We can then use this definition of a normal distribution over partitions to define normally-distributed HEAFs. Note that the HEAF in Figure 2 us a normally-distributed HEAF.

Definition 11. *A* **normally-distributed HEAF (nEAF)** *is a HEAF* $\langle A, R, D\rangle$ *with a partition* $P = [\langle\langle A_1, R_1\rangle, D_1\rangle, ..., \langle\langle A_m, R_m\rangle, D_m\rangle]$ *s.t.:*
- $A = \cup_{i=1}^m A_i, R = \cup_{i=1}^m R_i, D = \cup_{i=1}^m D_i$*, and for* $i = 1, ..., m$*,* $\langle A_i, R_i\rangle$ *is a DAF,*
- *if* $(z, \langle x, y\rangle) \in D_i$ *then* $(x, y) \in R_i, z \in A_{i+1}$*,*
- $m = \lfloor \log_2 |A| \rfloor + 1$ *(Sturges' formula), and*
- $|A_j| = \lfloor (p_{l-j} \times |A|) + 1 \rfloor$ *where* $\mathsf{norm\_prop}(m) = [p_0, p_1, ..., p_{l-1}]$.

We also consider *evenly-distributed HEAFs* (eEAFs), in which each level of the partition has an equal number of arguments. We consider eEAFs to be an interesting corner-case to investigate. Again, we use Sturges' formula to compute an appropriate number of partitions.

Definition 12. *An* ***evenly-distributed HEAF (eEAF)*** *is a HEAF* $\langle A, R, D\rangle$ *with a partition* $P = [\langle\langle A_1, R_1\rangle, D_1\rangle, ..., \langle\langle A_m, R_m\rangle, D_m\rangle]$ *such that:*
- $A = \cup_{i=1}^m A_i, R = \cup_{i=1}^m R_i, D = \cup_{i=1}^m D_i$*, and for* $i = 1, ..., m$*,* $\langle A_i, R_i\rangle$ *is a DAF.*
- *If* $(z, (x, y)) \in D_i$ *then* $(x, y) \in R_i, z \in A_{i+1}$
- $m = \lfloor \log_2 |A| \rfloor + 1$
- *For* $i = 0, ..., m$*,* $|A_i| = \lceil (|A| \div m \pm 1) \rceil$

3.3 Capped CAFs

We consider two structures of CAF: those in which the size of any collective-attack set is no greater than (*capped at*) 3 and CAFs in which there is no restriction on the size of collective-attacks sets. We refer to capped frameworks as *cCAFS*, and those which are uncapped as *uCAFs*.

Definition 13. *A* ***capped collective-attack framework (cCAF)*** *is a CAF* $\langle A, R\rangle$ *s.t.* $\forall (S, a) \in R : |S| \leq 3$.

Note, in the rest of this paper, to emphasise the distinction with capped collective-attack frameworks, we refer to collective-attack frameworks where it is not necessarily the case that there is an upper bound of 3 on the size of the attacking sets as **uncapped collective-attack frameworks, (uCAFs)**.

4 Characteristics of structural classes of AF

We ran experiments with the following structural classes of AF: 0.25-DAF, 0.5-DAF, 0.75-DAF, mst-DAF, eEAF, nEAF, cCAF and uCAF. We consider specifically the size of the grounded and the preferred sceptical extensions (known to affect the performance of argument solvers [8]), the proportion of argument subsets that determine some topic argument to be acceptable (a factor in the effectiveness of dialogue strategies for persuasion [3]), and whether the addition of a new argument to the framework results in a change of acceptability of some topic argument (also a factor in the effectiveness of dialogue strategies [1] and intrinsic to a variety of other argument technologies [16]). To investigate these properties empirically, we generate random instances of the specified structures.

When generating DAFs, we ensure that each possible weakly-connected DAF with the specified density is equally likely to be generated, only excluding frameworks that contain self-attacking arguments. For EAFs, we begin by generating each partition as a 0.5-DAF (in the same manner as described above), where the number of arguments in the partitions depends on the distribution of the EAF (e.g. whether it is a eEAF or nEAF). Then, we add one random preference relation from each argument (excluding those in the lowest partition), to a random attack relation in the partition directly below it; preference relations are generated one at a time, ensuring that the EAF has a valid HEAF structure (specifically maintaining the property in Definition 3, bullet 2). Finally, for CAFs, we begin by generating a random 0.25-DAF (in the manner described above); the attacks generated form the singleton attacks of the CAF. We then add attacks from sets of more than one argument so that the total number of attacks in the resulting CAF is the same as the number of attacks in a 0.5-DAF with the same number of arguments. We begin by first randomly selecting the size of the attacking set (for cCAFs either 2 or 3, for uCAFs from 2 to $|A| - 1$), we then randomly select a set of arguments of that size and then randomly select an argument to be attacked by that set; we repeat until we have the required number of attacks.

Our experiments were implemented in Java, partly using the Tweety library [27]. Our code is available at `github.com/joshlmurphy`. Experiments were run on an Intel i5 3.20GHz CPU, with 4GB RAM.

4.1 Size of extension

The argument solver competition [8], in which argument solvers attempt to complete a set of tasks related to computational argumentation as efficiently as possible (such as computing an extension, or determining whether a particular argument is acceptable) used three benchmark sets of DAFs. Two of these benchmarks were characterised by the size of their extensions: frameworks with large grounded extensions and frameworks with large preferred extensions. Most solvers were slower when tasked with frameworks with a large preferred extension compared to those frameworks with a large grounded extension. This indicates that the size of the extensions of a framework is an important consideration when employing an argument solver for certain tasks. We investigated how the average size of both the grounded and the preferred sceptical extensions differs between our chosen framework classes.

For each framework class, we randomly generated at least 1,000 frameworks with n number of arguments, where $n = 12, 24, 36$. Figure 4 shows the average size of both the grounded and the preferred sceptical extension of the frameworks we generated. For DAFs, we observe a trend for both semantics that the more dense the DAF, the smaller the size of the extension. We also observe that the larger the framework, the larger the extension.

Interestingly, CAFs reverse this trend when using the grounded semantics: the larger a uCAF/cCAF, the smaller (on average) the grounded extension. This surprising result can be explained by the intuition that as you increase the number of arguments in a CAF, this increases the *proportion* of group attacks, and thus the more arguments that are part of a collective-attack relation, leading to a higher number of attack cycles (the more arguments in a set S that collectively attack an argument a, the higher the chance that a will attack at least one argument in S, causing a cycle) and the more attack cycles in a framework the smaller the grounded extension is likely to be. This is supported by the observation that uCAFs have on average a smaller grounded extension than cCAFs, which, we conclude, is due to more arguments being part of a collective-attack relation in uCAFs (as there is no cap on the number of arguments in the attack relation). When using the preferred semantics, cycles are less of a factor in the size of the extension, and so we observe that the size of the preferred sceptical extension increases as the size of CAF increases.

We find that eEAFs are more likely to have a larger grounded extension than nEAFs, but have similar sized preferred sceptical extensions. We reason that in EAFs, the more arguments that attack an attack between two arguments that exist in a framework, the more likely attack relations in the partition *below* will be defeated. This effectively lowers the attack density in lower partitions. So in the frameworks with more preferences on average (eEAFs) there will be a lower overall attack density. As we observe in DAFs, the lower the attack density of

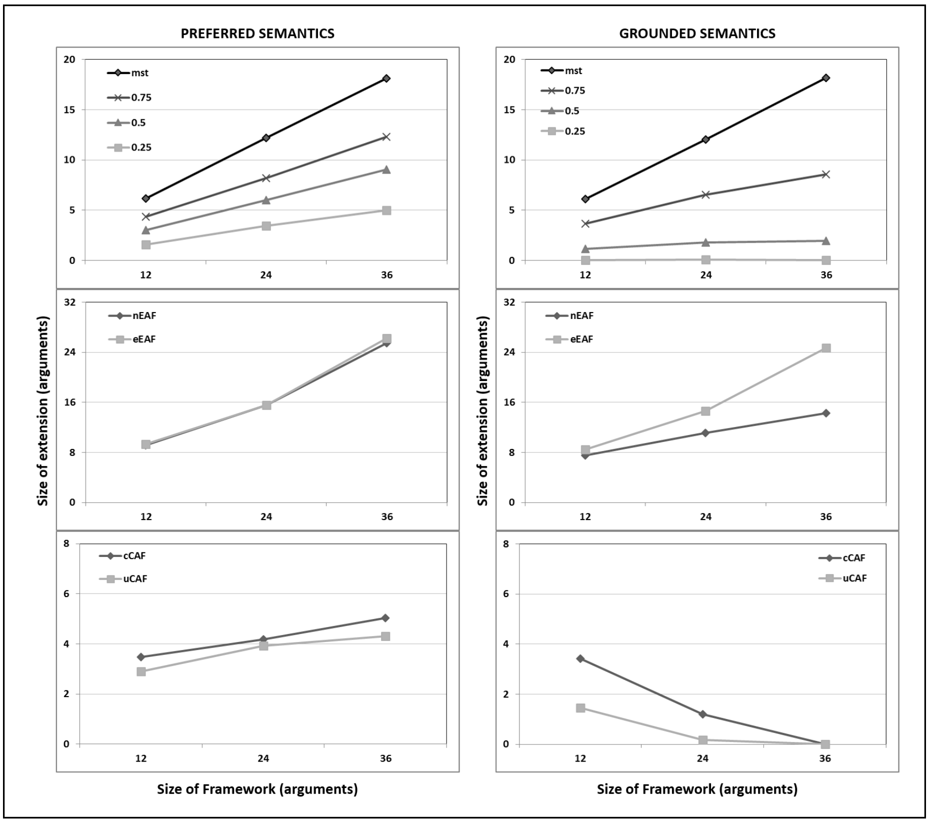

Fig. 4. The size of the grounded and preferred sceptical extensions.

the framework, the larger the extension — this is reflected in the results for the grounded extension.

4.2 Subsets that determine a topic acceptable

Some particular *topic* argument will be determined acceptable by some subsets of arguments, but not others. Any topic argument will be acceptable in at most 50% of the subsets, since it will not exist in half of the subsets of the power set (an argument is deemed unacceptable in a framework it is not a part of). We refer to the proportion of subsets in which the topic argument is determined to be acceptable as *SA*. This property has been found to be an important factor in the success of persuasion dialogues [3]: the lower SA, the more difficult it is to persuade an agent that the topic argument is acceptable.

We investigated whether average SA differs between the selected framework classes. Our implementation is naive, exhaustively checking whether some topic

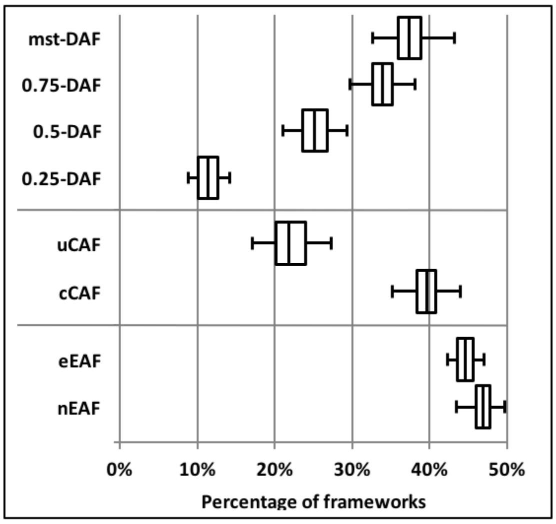

Fig. 5. The percentage of argument subsets in which a particular argument is acceptable under the grounded semantics.

is acceptable in every set in the power set of arguments. The time for these experiments is very high due to the exponential growth in the size of the power set. To feasibly compute the results we used the grounded semantics (which are faster to compute) and limited the framework size to 12 arguments. We generated at least 1,000 random instances of each framework class with 12 arguments, each time randomly selecting a topic argument.

Using the analysis of variance test (ANOVA, a collective of tests used to analyse the difference between the means of multiple groups [13]) we find that the different argumentation framework classes have significantly different SA (apart from nEAF and eEAF which are distinct from other classes but not from each other), and thus that each class is a distinct population ($p < 0.05$ for each class); this implies that the framework class is a significant factor in determining SA. The largest difference between two classes is between *mst-DAF* and *nEAF*, with a 36.06 percentage points difference between means.

In the different classes of DAF, we observe a clear trend that the more dense a framework class, the lower SA is for that class (see Figure 5). This follows the trend observed in Figure 4, where the more dense a DAF, the smaller the grounded extension. Similarly, uCAF frameworks typically have a smaller grounded extension than cCAF frameworks, and this trend is repeated for SA. For nEAF and eEAF frameworks of 12 arguments, there is little difference between the size of grounded extensions, and this trend is again shown for SA, in which eEAF and nEAF were not found to have significantly different SA. When using the grounded semantics it appears that the size of the extension and SA are linked.

Table 1. The percentage of frameworks that are resistant.

Framework class	12 args	24 args	36 args
mst-DAF	80.9	85.8	90.7
0.75-DAF	81.9	89.5	93.4
0.5-DAF	87.5	90.8	95.1
0.25-DAF	92.3	93.1	96.0
nEAF	87.2	91.7	96.8
eEAF	89.9	93.4	98.5
cCAF	76.7	84.1	88.8
uCAF	69.2	74.3	79.6

4.3 Resistance of AFs

Argumentation is an inherently dynamic process, with arguments and attack relations changing as new knowledge becomes available. The dynamic nature of argumentation can potentially be exploited for computational efficiency [16] as well as for strategic advantage [6]. Amgoud and Vesic consider whether the addition of a new argument to a framework changes the acceptability of a specific argument (termed the *topic argument*) [1]. If the addition of a new argument does not cause a change in the topic argument's acceptability we say the framework is *resistant*, otherwise it is *susceptible*. To investigate whether there is a difference in their resistance, for each framework class, we randomly generated at least 1,000 frameworks with n number of arguments, where $n = 12, 24, 36$, selecting both a topic argument and a test argument at random, determining the AF to be resistant if the acceptability of the topic argument is unaffected by the inclusion of the test argument. Table 1 shows the percentage of the framework instances we generated that are resistant.

For all classes we observe that the larger the framework, the more likely it is to be resistant. Intuitively, the more arguments in a framework, the more likely it is that an argument is topographically further away from the topic, and therefore the less likely the test argument will change the acceptability of the topic (this relationship can be used as a heuristic to inform an argument dialogue strategy [20]).

In a cCAF, a new argument can alter the acceptability of arguments both through introducing new argument-argument attacks as well as new collective-attacks. This is also true in uCAFs, though they have a greater chance of introducing collective-attacks: since the size of a collective attack is uncapped, each argument is in more collective attack relations on average. Thus, when we add a new argument to a uCAF it is likely to result in more changes in the acceptability of arguments, and this is why cCAFs are more resistant.

We see that eEAFs are more resistant than nEAFs, indicating that the higher the proportion of preference arguments to arguments, the more resistant the EAF will be. This is because an argument cannot alter the acceptability of an argument in a partition higher than its own partition, since all attack relations are either to arguments in the same partition or to arguments in the partition directly below. Therefore, if the topic argument is in a higher partition than the test argument, the framework is guaranteed to be resistant. In eEAFs it is more likely that the topic will be in a higher partition (since it is randomly selected and there are more arguments in higher partitions than in a nEAF) and thus the less likely it is that the test argument will affect the topic's acceptability.

5 Case studies

We examine two case study frameworks, obtained from argument technologies deployed on real-world data. These motivate the relevance of the classes of framework structure we investigate (showing that the results of our experiments on randomly generated AFs map to the properties of our case-study frameworks) and also allow us to demonstrate how our results can be used to inform argument technologies.

5.1 Trial aggregation

As evidence-based decision-making becomes increasingly important, clinical trials can provide an important source of information to inform healthcare professionals. Hunter and Williams propose an argument-based approach for aggregating the positive and negative effects of potential treatments, which has been shown to produce recommendations that align with published clinical guidelines [15]. The approach performs a type of meta-analysis on a range of clinical literature, producing a DAF (very sparse, almost a mst-DAF in structure) on which reasoning about possible treatment options is done. We use such a framework as our first case study.

Table 2 shows the number of arguments present in our trial aggregation case study DAF (|Args|), the size of its grounded and preferred sceptical extensions (|Gr| and |Pr|), the average SA over all possible topic arguments (SA) and the percentage of cases that were resistant over each possible topic argument with a randomly selected test argument (Res). We see that the results correlate with the results obtained from mst-DAF presented earlier in this paper, with the size of extensions, SA, and resistance being within the expected ranges of mst-DAFs. This evidences the relevance of the structures we investigate.

We consider particularly the resistance of this case study framework to demonstrate how our results may be used to inform a specific application. The resistance of the trial aggregation framework is exceptionally high (*97.2%*). This indicates that new arguments added in the future, in this case by the addition of new clinical studies, are unlikely to change the acceptability of other arguments in the framework. This implication of this is that new studies are unlikely to have an affect on the recommended treatment, meaning there can be confidence in the current recommendation. If a framework produced by the trial aggregation approach had a low resistance, new studies would be likely to change the recommended treatment, and this would imply that the recommendation is not yet reliable.

Table 2. Case study results.

	|Args|	|Gr|	|Pr|	SA	Res
Trial aggregation	34	9	9	41.9	97.2
Model selection	13	7	7	46.5	89.1

5.2 Statistical model selection

Clinicians without statistical training often need support to select a suitable model to correctly analyse and reason about their data. Sassoon *et al.* propose a tool that uses argumentation to aid in the process of deciding which statistical model is most suited to a user's research question, data and preferences [25]. The requirements and preferences of the user, as well as preferences from applicable context domains, are captured in an EAF, which can then inform the user of the most suitable model to use. We use a framework produced by using this tool with real-world data from a study involving clinicians (originally presented in [25]) as our second case study. The framework is an eEAF, being a HEAF with the same number of arguments at each level of the hierarchy.

Table 2 shows the number of arguments present in our statistical model selection case study eEAF (column |Args|), the size of its grounded and the preferred sceptical extensions (column |Gr| and column |Pr|), the average SA over all possible topic arguments (column SA) and the percentage of cases that were resistant over each possible topic argument with a randomly selected test argument (column Res). We see that the results correlate with our experiments over randomly generated eEAFs. Perhaps the most interesting result from this case study is the high SA of the framework (*46.5%*). Empirical investigations have demonstrated that the higher SA, the easier it is for a persuader to convince a persuadee of the acceptability of some argument [3,19]. Therefore, we would expect the persuasion of a user to use a particular statistical model to be likely to be successful when the underlying AF is an eEAF, as in this case study we consider here.

6 Discussion

We have shown that the type of AF and its structural properties have a significant effect on the size of the grounded and preferred sceptical extensions, on the proportion of subsets that determine some topic argument to be acceptable, and on the resistance of the framework; these characteristics are known to be important factors in the performance of different argument technologies. Understanding these relationships is therefore important when considering how to evaluate such technologies. Furthermore, it can allow technologies to be optimised for specific domains in which certain structures of AF are known to be typical (such as our case study domains). For example, solvers can be developed to be faster for particular classes of framework, or a dialogue strategy can be effective for particular knowledge domains.

Related work considers how graph-theoretic properties of DAFs can be used to predict the "best" argument solver for a particular DAF [7], specifically the work considers how fast solvers are for DAFs with structures based on social networks. In contrast, we consider a range of general argument-based characteristics that are known to impact on various argumentation-technologies, including argument solvers. The structures we investigate are based on those derived from generalised argumentation frameworks commonly used in argument-technology, and our case-studies demonstrate the relevance of these structures.

We could also examine structures of framework derived from natural human-style argumentation (such as recent work by Rosenfield and Kraus [24]). Argument mining offers the possibility of obtaining large datasets of frameworks from real-world human-based argumentation, and can be applied to a vast array of domains, providing a range of framework structures related to human-reasoning. However, representing human reasoning in a formal argumentation framework is a challenging task; detecting arguments can be difficult in human dialogues because conflict tends to be hidden [11]. Nevertheless, investigating the properties of structural patterns that may be emerging in representations of human reasoning is a possible direction for future work.

References

1. L. Amgoud and S. Vesic. On revising argumentation-based decision systems. In *10th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 71–82. Springer LNAI 5590, 2009.
2. S. Bistarelli, F. Rossi, and F. Santini. A comparative test on the enumeration of extensions in abstract argumentation. *Fundamenta Informaticae*, 140(3-4):263–278, 2015.
3. E. Black, A. Coles, and S. Bernardini. Automated planning of simple persuasion dialogues. In *15th International Workshop on Computational Logic in Multi-Agent Systems*, pages 87–104. Springer LNAI 8624, 2014.
4. G. Brewka, S. Polberg, and S. Woltran. Generalizations of dung frameworks and their role in formal argumentation. *IEEE Intelligent Systems*, 29(1):30 – 38, 2014.
5. R. Brochenin, T. Linsbichler, M. Maratea, J. Wallner, and S Woltran. Abstract solvers for Dung's argumentation frameworks. In *3rd International Workshop on Theory and Applications of Formal Argumentation*, pages 40–58. Springer LNAI 9524, 2016.
6. C. Cayrol, F. de Saint-Cyr, and M. Lagasquie-Schiex. Change in abstract argumentation frameworks: Adding an argument. *Journal of Artificial Intelligence Research*, 38:49–84, 2010.
7. F. Cerutti, M. Giacomin, and M. Vallati. Algorithm selection for preferred extensions enumeration. In *5th International Conference on Computational Models of Argument*, pages 221–232. IOS Frontiers in AI and Applications Vol. 266, 2014.
8. F. Cerutti, N. Oren, H. Strass, M. Thimm, and M. Vallati. A benchmark framework for a computational argumentation competitions. In *5th International Conference on Computational Models of Argument*, pages 459–460. IOS Frontiers in AI and Applications Vol. 266, 2014.

9. F. Cerutti, A. Palmer, A. Rosenfeld, J. Snajder, and F. Toni. A pilot study in using argumentation frameworks for online debates. In *1st International Workshop on Systems and Algorithms for Formal Argumentation*, pages 63–74, 2016.
10. G. Charwat, W. Dvorak, S. Gaggl, J. Wallner, and S. Woltran. Methods for solving reasoning problems in abstract argumentation: A survey. *Artificial Intelligence*, 220:28 – 63, 2015.
11. S. Concannon, P Healey, and M. Purver. How natural is argument in natural dialogue? In *16th International Workshop on Computational Models of Natural Argument*, 2016.
12. P. 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, 1995.
13. R. Fisher. *Statistical methods for research workers.* Genesis Publishing, 1925.
14. S. Frank. The common patterns of nature. *Journal of Evolutionary Biology*, 22(8):1563–1585, 2009.
15. A. Hunter and M. Williams. Aggregating evidence about the positive and negative effects of treatments. *Artificial Intelligence in Medicine*, 56(3):173 – 190, 2012.
16. B. Liao, L. Jin, and R. Koons. Dynamics of argumentation systems: A division-based method. *Artificial Intelligence*, 175(11):1790–1814, 2011.
17. S. Modgil. Reasoning about preferences in argumentation frameworks. *Artificial Intelligence*, 173(9):901 – 934, 2009.
18. S. Modgil, F. Toni, et al. The added value of argumentation. In S. Ossowski, editor, *Agreement Technologies*, pages 357–403. Springer, 2013.
19. J. Murphy, E. Black, and M. Luck. Arguing from similar positions: an empirical analysis. In *3rd International Workshop on the Theory and Applications of Formal Argumentation*, pages 177–193. Springer LNAI 9524, 2016.
20. J. Murphy, E. Black, and M. Luck. A heuristic strategy for persuasion dialogues. In *6th International Conference on Computational Models of Argument*, pages 411–417. IOS Frontiers in AI and Apps. Vol. 287, 2016.
21. S. Nielsen and S. Parsons. An application of formal argumentation: Fusing bayes nets in mas. In *1st International Conference on Computational Models of Argument*, pages 33–44. IOS Frontiers in AI and Applications Vol. 144, 2006.
22. S. Nielsen and S. Parsons. A generalization of Dung's abstract framework for argumentation: Arguing with sets of attacking arguments. In *3rd International Workshop on Argumentation in Mutli-Agent Systems*, pages 54–73. Springer LNAI 4766, 2006.
23. N. Oren, T. Norman, and A. Preece. Loose lips sink ships: A heuristic for argumentation. In *3rd International Workshop on Argumentation in Multi-Agent Systems*, pages 121–134. Springer LNAI 4049, 2006.
24. A. Rosenfield and S. Kraus. Strategical argumentative agent for human persuasion. In *22nd European Conference on Artificial Intelligence*, pages 320–328. IOS Frontiers in AI and Applications Vol. 285, 2016.
25. I. Sassoon, J. Keppens, and P. McBurney. Preferences in argumentation for statistical model selection. In *6th International Conference on Computational Models of Argument*, pages 53–60. IOS Frontiers in AI and Applications Vol. 287, 2016.
26. H. Sturges. The choice of a class interval. *Journal of the American Statistical Association*, 21(153):65–66, 1926.
27. M. Thimm. Tweety - A comprehensive collection of Java libraries for logical aspects of artificial intelligence and knowledge representation. In *14th International Conference on Principles of Knowledge Representation and Reasoning*, pages 528–537. AAAI Press, 2014.