Probabilistic Semantics for the Carneades Argument Model Using Bayesian Networks

Matthias Grabmair [a] Thomas F. Gordon [b] and Douglas Walton [c]
[a] *Intelligent Systems Program, University of Pittsburgh, USA*
[b] *Fraunhofer FOKUS, Berlin, Germany*
[c] *University of Windsor, Canada*

Abstract. This paper presents a technique with which instances of argument structures in the Carneades model can be given a probabilistic semantics by translating them into Bayesian networks. The propagation of argument applicability and statement acceptability can be expressed through conditional probability tables. This translation suggests a way to extend Carneades to improve its utility for decision support in the presence of uncertainty.

Keywords. Carneades, argumentation and probability, Bayesian Networks

1. Introduction

The *Carneades* argument model [9] was developed with the goal of formalizing *Walton*'s theory of argumentation structure and evaluation. The acceptability of a proposition at issue is determined by aggregating pro and con arguments using proof standards. Several proof standards in the legal domain have been formally modeled, including preponderance of the evidence, clear and convincing evidence and beyond a reasonable doubt [8,9]. A Carneades argument graph is a bipartite, directed, acyclic graph, with statement nodes and argument nodes. Each statement node represents two literals, an atomic proposition, P, and its logical complement, $\neg P$. An argument pro P is an argument con $\neg P$ and vice versa. The acceptability of a literal P is determined by using proof standards to aggregate the applicable arguments pro and con P. An argument is applicable only if, recursively, its premises are acceptable and its negative premises and exceptions are not acceptable.

This paper uses the structural similarities between Carneades argument graphs and Bayesian networks (BN) to demonstrate how the Carneades method of evaluating arguments using proof standards can be simulated with BNs by translating an argument structure into a *Carneades Bayesian Network* (CBN). The translation relies on the Carneades formalism to define a class of BNs which can be instantiated to an autonomous BN given an instance of an argument evaluation structure. The simulation gives Carneades an alternative, potentially useful probabilistic semantics and suggests a way to extend Carneades, by modeling assumptions as probability distributions over possible statuses of literals and inferring probability distributions over the acceptability of arguments and derivability of other literals. The benefit for some applications could be that probabilities seamlessly interface to established decision-support techniques and allow a qualitative model of argument to be extended to support quantitative methods for reasoning about uncertainty.

2. Bayesian Networks

A Bayesian network (BN) is a model of random variables and the conditional probabilities between them based on a directed acyclic graph [17]. BNs allow for a structured representation of probabilistic causal relationships by exploiting conditional independence relations to reduce the number of probabilistic parameters over the full joint distribution, thereby also reducing the computational complexity of the inference calculation. One can compute the probability of a certain variable having a certain status given information about the status of other variables by multiplying the respective Bayesian probabilities.

Definition 1 (Bayesian Network) *A **Bayesian Network (BN)** is an ordered triple $\langle V, D, \mathcal{P} \rangle$, where V is a set of nodes in the network which correspond to random variables. D is the set of directed edges between the nodes represented as ordered pairs (v_1, v_2) where $v_1, v_2 \in V$. $\mathcal{P}$ is a set of probability functions, each mapping a certain node and a configuration of its immediate parent nodes onto a probability value. Further, the Markov assumption needs to be met, i.e. each random variable needs to be conditionally independent of its non-descendant variables given its parent variables.*

3. The Carneades Model

This section gives an introduction to the Carneades argument model as it has been defined in previous work in greater detail, e.g. [9,10]. Our point of departure is the concept of an argument. Put informally, an argument consists of a set of premises which have to hold in order for the argument to be applicable.

Definition 2 (argument) *Let $\mathcal{L}$ be a propositional language. An **argument** is a tuple $\langle P, E, c \rangle$ where $P \subset \mathcal{L}$ are its **premises**, $E \subset \mathcal{L}$ are its **exceptions** and $c \in \mathcal{L}$ is its **conclusion**. For simplicity, c and all members of P and E must be literals, i.e. either an atomic proposition or a negated atomic proposition. Let p be a literal. If p is c, then the argument is an argument **pro** p. If p is the complement of c, then the argument is an argument **con** p.*

Argumentation takes place in a dialogue, which Carneades divides into *states* [10]. Each state contains a "momentary instance" of the arguments brought forth.

Definition 3 (argumentation process) *An **argumentation process** is a sequence of **states** where each state is a set of arguments. In every chain of arguments, $a_1, \ldots a_n$, constructable from the arguments in a state by linking the conclusion of an argument to a premise or exception of another argument, a conclusion of an argument a_i may not be a premise or exception of an argument a_j, if $j < i$. A set of arguments which violates this condition is said to contain a* cycle *and a set of arguments which complies with this condition is called* cycle-free.

Arguments are assessed using the concept of an *audience*, which has certain factual assumptions and a set of subjective weights it assigns to the arguments presented [2,1].

Definition 4 (audience) *An* audience *is a structure $\langle \Phi, f \rangle$, where $\Phi \subset \mathcal{L}$ is a consistent set of literals assumed to be acceptable by the audience and f is a partial function mapping arguments to real numbers in the range $0.0 \ldots 1.0$, representing the relative weights assumed to be assigned by the audience to the arguments.*

The actual evaluation of arguments presented at a given state to an audience takes place in an *evaluation structure*. It assigns a proof standard to each statement in the argument, representing the argumentative threshold for establishing a statement.

Definition 5 (argument evaluation structure) *An* ***argument evaluation structure*** *is a tuple $\langle \Gamma, \mathcal{A}, g \rangle$, where Γ is a state in an argumentation process, $\mathcal{A}$ is an audience and g is a total function mapping propositions in $\mathcal{L}$ to their applicable proof standards in the process. A* ***proof standard*** *is a function mapping tuples of the form $\langle p, \Gamma, \mathcal{A} \rangle$ to the Boolean values true and false, where p is a literal in $\mathcal{L}$.*

In a given evaluation context, a statement is acceptable if its proof standard has been met according to the assessment of the audience.

Definition 6 (acceptability) *Let $\mathcal{S} = \langle \Gamma, \mathcal{A}, g \rangle$ be an argument evaluation structure, where $\mathcal{A} = \langle \Phi, f \rangle$. A literal p is* ***acceptable*** *in $\mathcal{S}$ if and only if $g(p)(p, \Gamma, \mathcal{A})$ is true.*

In addition to the acceptability, we further make use of the derivability of literals as defined in Gordon & Ballnat [11].

Definition 7 (derivability) *Let $\mathcal{S} = \langle \Gamma, \mathcal{A}, g \rangle$ be an argument evaluation structure, where $\mathcal{A} = \langle \Phi, f \rangle$. A literal p is* ***in*** *S, denoted $(\Gamma, \Phi) \vdash_{f,g} p$, if and only if*

- *$p \in \Phi$ or*
- *$(\neg p \notin \Phi$ and p is acceptable in $S)$*

Otherwise p is ***out****, denoted $(\Gamma, \Phi) \nvdash_{f,g} p$.*

Consequently, an argument is applicable once it has been put forward and each of its premises has either been assumed to be acceptable or is acceptable. Further, none of the exceptions of the argument may be acceptable or assumed to be acceptable.

Definition 8 (argument applicability) *Let $\langle state, audience, standard \rangle$ be an argument evaluation structure. An argument $\langle P, E, c \rangle$ is* ***applicable*** *in this argument evaluation structure if and only if*

- *the argument is a member of the arguments of the state,*
- *every proposition $p \in P$, the premises, is an assumption of the audience or, if neither p nor $\overline{p}$ is an assumption, is acceptable in the argument evaluation structure and*
- *no proposition $p \in E$, the exceptions, is an assumption of the audience or, if neither p nor $\overline{p}$ is an assumption, is acceptable in the argument evaluation structure.*

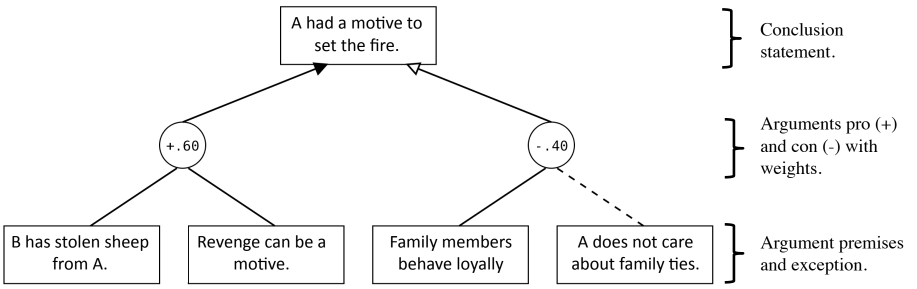

Figure 1. The arson example argument in the current Carneades visualization method.

For reasons of brevity, this paper only explains one proof standard, namely *preponderance of the evidence*. It states that a literal or its complement is acceptable depending on which side has the stronger applicable argument.

Definition 9 (preponderance of the evidence) *Let $\langle \Gamma, \mathcal{A}, g \rangle$ be an argument evaluation structure and let p be a literal in $\mathcal{L}$. $pe(p, \Gamma, \mathcal{A}) = true$ if and only if*

- *there is at least one applicable argument pro p in Γ and*
- *the maximum weight assigned by the audience $\mathcal{A}$ to the applicable arguments pro p is greater than the maximum weight of the applicable arguments con p.*

Fig. 1 shows an exemplary Carneades argument graph. Imagine an attorney representing farmer A in a civil proceeding. A has been accused of having set the farm of his brother B on fire and has been sued for compensation. The evidence speaks against A and the only element still at issue is the motive with which A had set the fire. The claimant has the burden of proof and argues (pro-argument, black arrowhead) that A had a motive (Literal M) as it has been established that B had stolen a sheep from A some time ago (Literal S) and revenge is a plausible motive for the arson (Literal R). The defense argument (con-argument, white arrowhead) is that it is less likely that A would commit the arson as they are members of the same family, hence A would likely be loyal to his sibling (Literal L). This argument would be applicable unless the claimant could show that A is not a "family person", i.e. someone who cares about family ties (Literal $\neg F$, exception shown as dashed line). Under the preponderance of evidence standard, one assigns greater weight to the pro-argument (0.6 vs. 0.4, see fig. 1), because one thinks the notion of revenge weighs heavier with the jury than loyalty among siblings. For the pleading to the jury, consequently, one sketches out the (admittedly simplified) argument graph shown in fig. 1. While one could expand each leaf node literal into a subtree of arguments or add additional arguments relating, for example, family loyalty and revenge, we leave the graph at this level of complexity for purposes of the illustration.

4. The Translation

The BN produced by our translation is not functionally equivalent to a Carneades argument structure in that it can be arbitrarily modified and still reflect a valid Carneades argument graph. Instead, the translation produces a BN *conditioned* on a specific configuration of an argument evaluation structure (see def. 5) with a set of probability functions emulating the status propagation of the just presented Carneades formalism. We thereby show how to simulate Carneades using a homomorphic BN. We do not claim that Carneades argument evaluation structures and BNs are isomorphic, which would allow, in the reverse direction, arbitrary BNs to be simulated using Carneades.

We justify the translation by explanation and by presenting an example argument graph and its corresponding BN. We do not have a full formal proof of functional correctness of the translation yet. We expect to be able to publish it in the near future.

We begin by introducing random variables for literals and arguments. Statement variables represent the *in*- or *out*-status of a literal given its proof standard and connecting arguments. Recall that in a Carneades graph structure, p and $\neg p$ are treated as a unit. We do not capture this unity in the BN translation. Instead, for each literal p in the argument structure, the BN contains two variables, one corresponding to p and the other one to $\neg p$. An argument depending on the literal as a premise or exception can then be linked to the respective variables as needed. This also means that every argument variable always connects to two statement variables, namely to one for its conclusion as a pro argument and to another one for its negated conclusion as a con argument. Since statement and argument variables represent two different kinds of random variables, our BN is effectively bipartite. This is distinct from Vreeswijk's interpretation of an existing BN as containing argument subtrees [20] as well as from Muecke & Stranieri [15], who manually translate generic argument graphs into BNs.

Definition 10 (statement and argument variables) *A **statement variable** is a random variable representing the probability of a given literal having a certain derivability status. It can take the values $\{in, out\}$. An **argument variable** is a random variable representing the probability of a certain argument being applicable. It can take the values $\{applicable, not\ applicable\}$.*

We can now translate an argument evaluation structure into a BN structure of variables, edges and probability functions. We commence by translating literals and arguments into their respective kind of variables. Notice that we need to create a statement variable not only for every literal used in the state, but also for its complement.

Definition 11 (variable translation) *If $\mathcal{S} = \langle \Gamma, \mathcal{A}, g \rangle$ is an argumentation evaluation structure, then $V_\mathcal{S} = L \cup A$ is the set of statement and argument variables corresponding to the literals in Γ. Let $literals(\Gamma)$ be a set of literals consisting of both positive and negated versions of all literals used by arguments in Γ. Let sv_l be a statement variable representing literal l and av_a be an argument variable representing argument a.*

- *Literals:* $L = \{sv_l | l \in literals(\Gamma)\}$
- *Arguments:* $A = \{av_a | a \in \Gamma\}$

We can now connect the variables using directed, labeled edges. We create two outgoing edges for arguments, one for its conclusion and one for its complement.

Definition 12 (edge translation) *If $\mathcal{S} = \langle \Gamma, \mathcal{A}, g \rangle$ is an argumentation evaluation structure, then $D_\mathcal{S} = D_{Pr} \cup D_X \cup D_C$ is a set of directed, labeled edges connecting the variables of $V_\mathcal{S}$ represented as three-element tuples. Let $v_x \in V_\mathcal{S}$ be the random variable associated with statement/argument x.*

- *Premises:* $D_{Pr} = \{(v_p, v_a, premise) | a = \langle P, E, c\rangle \in \Gamma, p \in P\}$
- *Exceptions:* $D_X = \{(v_e, v_a, exception) | a = \langle P, E, c\rangle \in \Gamma, e \in E\}$
- *Conclusions:* $D_C = \{(v_a, v_c, pro) | a = \langle P, E, c\rangle \in \Gamma\}$
$\cup \{(n_a, n_{\neg c}, con) | a = \langle P, E, \neg c\rangle \in \Gamma\}$

The propagation functionality is added through the set of probability functions, which associates with each variable a function that determines the probability of the given random variable having a certain value conditioned on the values of the immediate parent variables and, as explained above, the elements of the argument evaluation structure: assumptions, proof standards and argument weights. For an argument variable, the parent variables are its premises and exceptions, while the parents of a statement variable represent arguments pro and con the statement. As we simulate the value propagation of the Carneades model, all conditional probabilities assigned are either 0 or 1, making the nodes in the network behave deterministically in accordance with the argument formalism. Notice that all information about the adjudication of competing applicable arguments (i.e. proof standards and argument weights) flows into the probability functions.

Definition 13 (probability functions) *If $\mathcal{S} = \langle \Gamma, \mathcal{A}, g\rangle$ is an argument evaluation structure where $\mathcal{A} = \langle \Phi, f\rangle$, $V_{\mathcal{S}} = S \cup A$ consists of a set of statement variables S and a set of argument variables A, and $D_{\mathcal{S}}$ is the set of connecting edges, then $\mathcal{P}_{\mathcal{S}} = \mathcal{P}_a \cup \mathcal{P}_s$ is a set of probability functions defined as follows:*

If $a \in A$ is an argument variable and $s_1, ..., s_n$ are its parent statement variables, then P_a is the probability function for the variable a in the BN.

$$P_a(a = \textit{applicable} | s_1, ..., s_n) = \begin{cases} 1, & \textit{if for every } s_i \textit{ it holds that either:} \\ & ((s_i, a, premise) \in D_{\mathcal{S}} \textit{ and } s_i = in) \\ & \textit{or } ((s_i, a, exception) \in D_{\mathcal{S}} \textit{ and } s_i = out) \\ 0, & \textit{otherwise} \end{cases}$$

$$P_a(a = \textit{not applicable} | s_1, ..., s_n) = 1 - P_a(a = \textit{applicable} | s_1, ..., s_n)$$

If $s \in S$ is a statement variable representing literal l and $a_1 ... a_n$ are its parent argument variables, then P_s is the probability function for the variable s:

$$P_s(s = in | a_1, ..., a_n) = \begin{cases} 1, & \textit{if } (\Gamma, \Phi) \vdash_{f,g} p \\ 0, & \textit{if } (\Gamma, \Phi) \nvdash_{f,g} p \end{cases}$$

$$P_s(s = out | a_1, ..., a_n) = 1 - P_a(s = in | a_1, ..., a_n)$$

Let $\mathcal{P}_a = \{P_a | a \in A\}$ and $\mathcal{P}_s = \{P_s | s \in S\}$.

By itself, the translation defines a class of BNs conditioned on an argument evaluation structure. One can instantiate a single autonomous CBN (variables, edges, probability functions) given a specific argument evaluation structure.

Definition 14 (Carneades Bayesian network) *A **Carneades Bayesian Network** (CBN) is an ordered tuple $\langle V_{\mathcal{S}}, D_{\mathcal{S}}, \mathcal{P}_{\mathcal{S}}\rangle\rangle$, where $V_{\mathcal{S}}$ is a set of statement and argument variables in the network, $D_{\mathcal{S}}$ is a set of directed edges and $\mathcal{P}_{\mathcal{S}}$ is the set of probability functions*

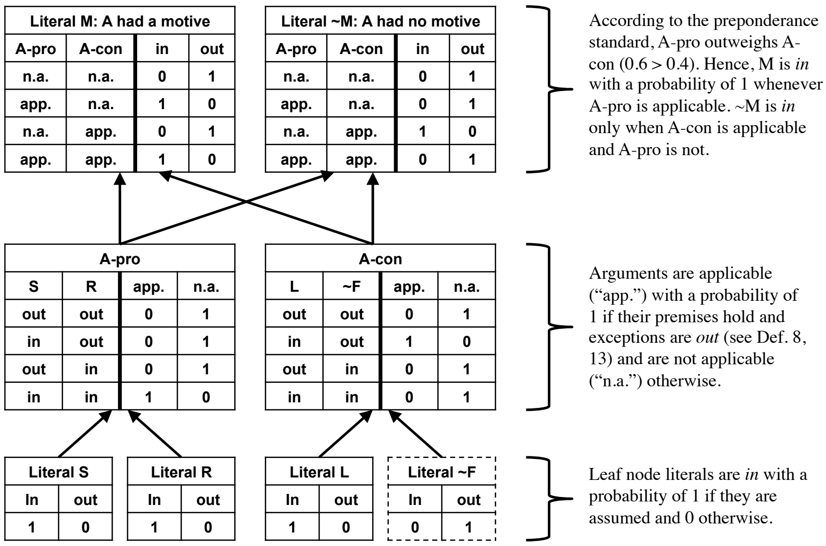

Figure 2. The example argument as a BN. The dashed box represents a negative literal. Negation is shown as $\sim$. Unused literals have been omitted. For the propositions of the abbreviated literals, see end of section 3.

conditioned on the information in a given argument evaluation structure $\mathcal{S}$. Further, the class of CBNs shall be defined as:

$$\mathcal{C} = \{\langle V_{\mathcal{S}}, D_{\mathcal{S}}, \mathcal{P}_{\mathcal{S}} \rangle\rangle | \mathcal{S} \textit{ is a Carneades argument evaluation structure}\}$$

5. An Example Translation

To give an illustration of a full reduction, we show our Carneades argument in the arson example as a BN in fig. 2 with the conditional probability tables attached to each variable. We begin by assuming that all premises of the two arguments hold, i.e. S, R and L are accepted and F is rejected. Given this configuration, the network behaves like a Carneades argument graph and will conclude that the presence of a motive will be deemed acceptable by the audience with a probability of 1 (i.e. $P(M = in) = 1$), because although both arguments are applicable, the pro-argument trumps the con-argument by virtue of it having a higher weight. This is an example of the patterns of 0s and 1s in the probability tables being contingent on the proof standard of each statement as well as the assumptions and weights attributed to the audience. If one were to change the weights so that the con-argument outweighed the pro-argument, the probability tables for *A-pro* and *A-con* in the depicted BN would change according to the translation.

6. The Characteristics and Utility of Probabilistic Semantics

6.1. Extension for Probability Distributions over Leaf Node Literals

In the example, we have modeled the assumptions about the audience in such a way that both arguments are applicable. The final binary determination of whether it will conclude that there is a motive is a matter of the proof standard and therefore about the weights which the audience is believed to assign to the arguments. However, the much more intuitive approach is to speak of the audience being more or less *likely* to accept or reject a certain statement, thereby introducing a notion of uncertainty into the concept of assumptions. Here, we can benefit from probabilistic semantics and extend the CBN model by introducing probability distributions over the *in*- and *out*-status of leaf literals (i.e. literals without arguments for or against them).

Up to now, we have modeled the literals l and $\neg l$ as distinct nodes in the network. Hence, such an extension conflicts with the consistency requirement for literals assumed by the audience. If both l and $\neg l$ are *in* with a certain nonzero probability each, the product of these probabilities is the probability of the assumptions being inconsistent with regard to that pair of literals. In order to remedy this issue, our extension needs to enforce the exlusiveness of the *in*-status. We achieve this by connecting positive/negated literal leaf pairs to a new common parent leaf variable with three possible values, which we shall refer to as a *probabilistic assumption variable*. If l is *accepted*, it is *in* and $\neg p$ is *out*. l being *rejected* is the inverse case. If l is *questioned*, both l and $\neg l$ are *out*.

Definition 15 (probabilistic assumption variable) *A **probabilistic assumption variable** is a random variable taking the possible values $M = \{$accepted, rejected, questioned$\}$ representing the probability of a certain literal being accepted, rejected or left undecided by the audience.*

Definition 16 (enhanced probability functions for leaf literals) *Assume $c = \langle V_{\mathcal{S}}, D_{\mathcal{S}}, \mathcal{P}_{\mathcal{S}}\rangle$ is a CBN. Let $P_l \in \mathcal{P}_{\mathcal{S}}$ be the probability function for a literal l's statement variable v_l in $V_{\mathcal{S}}$ and let L be the set of positive leaf literals in the argument. Let B be a set containing a probabilistic assumption variable b_l for each literal l in L. Let $\mathcal{P}_{ass}$ be set of a probability functions $P_{ass}(b_l = m)$ determining the probability with which $b_l \in B$ has status $m \in M$ such that $\sum_{m \in M} P_{ass}(b_l = m) = 1$ for all $b_l \in B$. Then an enhanced CBN $c' = \langle N'_{\mathcal{S}}, D'_{\mathcal{S}}, \mathcal{P}'_{\mathcal{S}}\rangle$ is constructed from the following components:*

$$V'_{\mathcal{S}} = V_{\mathcal{S}} \cup B$$

$$D'_{\mathcal{S}} = D_{\mathcal{S}} \cup \{(b_l, v_l, ass), (b_l, v_{\neg l}, ass) | b_l \in B; v_l, v_{\neg l} \in V'_{\mathcal{S}}\}$$

$$P'_l(v_l = in | b_l) = \begin{cases} 1, & \textit{if } b_l = \textit{accepted} \\ 0, & \textit{otherwise} \end{cases}$$

$$P'_l(v_l = out | b_l) = 1 - P'_l(v_l = in | b_l)$$

$$P'_{\neg l}(v_{\neg l} = in | b_l) = \begin{cases} 1, & \textit{if } b_l = \textit{rejected} \\ 0, & \textit{otherwise} \end{cases}$$

$$P'_{\neg l}(v_{\neg l} = out | b_l) = 1 - P'_{\neg l}(v_{\neg l} = in | b_l)$$

$$\mathcal{P}_{old} = \{P_l, P_{\neg l} | l \in L\} \ ; \ \mathcal{P}'_l = \{P'_l | l \in L\} \ ; \ \mathcal{P}'_{\neg l} = \{P'_{\neg l} | \neg l \in L\}$$

$$\mathcal{P}'_{\mathcal{S}} = (\mathcal{P}_{\mathcal{S}} \backslash \mathcal{P}_{old}) \cup \mathcal{P}'_l \cup \mathcal{P}'_{\neg l} \cup \mathcal{P}_{ass}$$

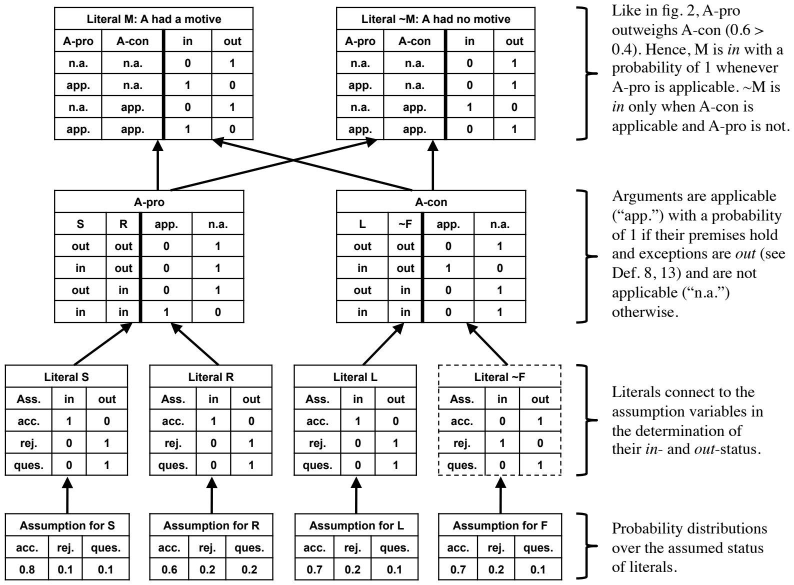

Figure 3. The example BN with extended literals and some possible probability distribution over assumed statements. Literals not used as premises or exceptions have been omitted.

This functionality connects Carneades to tasks of practical reasoning and planning. In the example, the claimant has the burden of proof and must show that $M = in$. The defendant's attorney does not need to establish $\neg M = in$, but only that $M = out$. She cannot tell for certain that the jury is going to accept a certain statement, but she can formulate her intuition about whether it is more likely to accept than to reject it. We extend the example BN and assign some probability distribution to the assumptions as shown in fig. 3. The diagram shows a distribution where the audience is likely to accept literals S (stolen sheep), L (loyalty among siblings) and F (A cares about family ties), but only moderately likely to accept R (revenge as a plausible motive), because burning the farm as revenge for a stolen sheep appears to be disproportionate. One can calculate the probability of successfully arguing that a motive has not been established: $P(M = out) = 0.52$. Modifying the assumption distributions or argument weights and recalculating the probabilities will yield insight into how best to persuade the jury and improve the chances. For example, the attorney can try to further weaken the jury's belief in revenge as a plausible motive (decrease of the probability the assumption for R to, for example, $P(\text{Assumption for R} = acc.) = 0.4$), thereby decreasing the probability of *A-pro* being applicable and achieving an overall success chance of $P(M = out) = 0.68$. Alternatively, she can try to convince the audience that *A-con* should weigh more heavily than *A-pro* (alteration of the conditional probability tables of M and $\neg M$ so that *A-con* wins over *A-pro* if both are applicable). This would increase her probability of success to $P(M = out) = 0.79$. Based on the calculations, it appears to be strategically better to plead for the family loyalty argument to have a higher weight than the revenge argument versus arguing that revenge is not a plausible motive.

We can see that this probabilistic form of Carneades allows some argumentation strategy decisions like the ones illustrated in the example to be cast into a modification of either the argument weights (thereby altering the probability functions) or the probabilistic assumption variables. After the modification, the network needs to be reinstantiated because of the change in the probability tables. The new BN can then be used to calculate the probability of successfully establishing the conclusion, which will improve, stay the same or worsen, thereby spreading open a space of possible actions, each associated with a certain payoff in argumentative certainty. We plan to pursue this possible utility of the BN translation in future work, specifically with regard to goal selection (see Gordon & Ballnat [11]) and dialogue games. Concerning the latter, our extension is similar to Riveret et al. [18], where probabilities were attached to premises of arguments to reflect the chance that a judge would accept the premise.

6.2. Probabilities, Proof Standards and Weights

In previous work on Carneades [10], it was argued that proof standards should not be modeled probabilistically because of (1) the difficulty in quantifying the necessary probabilities in argumentation scenarios, and (2) because multiple arguments can not typically be assumed to be independent of each other. The latter makes it difficult to generalize (conditional) independence between literals as one does not know whether they semantically connect to each other in a way not represented in the argument graph. In this paper, we use BNs to model the probability of a literal being accepted by the audience or satisfying its proof standard, i.e. the probability of the literal being *in* our *out*, not the probability of the it being true given the evidence. The probability of a proof standard being satisfied does not depend on an assumption that the arguments are independent.

Our model distinguishes between the strength of an argument by virtue of its weight in the eyes of the audience and the strength of an argument by virtue of its premises being highly likely to be fulfilled (i.e. the argument being applicable). The former is taken into account by the conditional probability tables which simulate the Carneades model. The latter is determined via propagating the probability distributions of the assumptions across the BN. Eventually, both quantities are synthesized into a single probability value representing how likely it is that the conclusion is successfully established through the given argumentation. A strong-weighted argument with unlikely premises may have less of an impact on the probability of success than a weaker-weighted argument which is applicable with a high probability. The new probabilistic semantics hence add an additional functional dimension to argument weights and proof standards as they allow the probabilities to reveal and quantify the tradeoffs involved in assessing the *value* of a certain argument a given context.

7. Computational Complexity

Despite the benefits of BNs over full joint distributions, exact inference of probabilities from a BN is NP-hard [5]. To remedy this, well-performing iterative sampling-based (so-called *Monte-Carlo*) approximation methods are commonly used. More efficient methods may be possible which exploit the structure of Carneades argument graphs. This is not the focus of this paper and we hence leave it for future work. Notice that computational complexity issues are only relevant when Carneades is extended as described in Section 6.1. Computational complexity issues are irrelevant for the general conceptual aspects of a probabilistic interpretation of existing argument models and for the main result of this paper, i.e. providing Carneades with a probabilistic semantics (section 4).

8. Further Related Work

Williams & Williamson have experimented with inferring argument weights for a model graph from existing data using BN learning methods [19]. By contrast, our approach constructs BNs from an existing model graph. Also, our model does not interpret argument weights in a probabilistic fashion. Freeman's transferral [8] of Cohen's concept of ampliative probability [4] to the Toulmin model is distinct in that it infers argument strength from the probability of its premises being fulfilled. Our model perceives argument strength as the weight assigned to an argument by a subjective audience given the fulfillment of its premises, the latter of which is computed probabilistically. This is conceptually similar to distinguishing different notions of argument strength as done by Krause et al.[14] to produce a confidence measure with regard to a proposition. The argumentation system of Zukerman et al. [21] distinguishes between an argument being normatively strong according to domain knowledge and it being persuasive to an audience by mapping two separate BNs onto each other. By contrast to our work, the probability measure of the BN is again per se equated with strength of belief in a goal proposition.

Carofiglio [3] proposes a multi-partite BN version of a Toulmin-like model with which probability distributions over the beliefs can be calculated. While this approach is similar to ours with regard to the translation, our translation handles the richer Carneades model, including proof standards. A quantitative probabilistic representation for belief in uncertain facts (explored in detail by Kadane & Schum [13]) has been criticized by Parsons [16]. He distinguishes different kinds of uncertainty and argues for the greater adequacy of qualitative methods in dealing with problems of uncertainty. We present a probabilistic interpretation of an existing argument model and hence do not take a substantive position with regard to these issues but intend to address them in future research.

Howard & Matheson [12] have abstracted BNs into *influence diagrams*, integrating both probabilistic inferences and decision problems into a graphical model. In influence diagrams, deterministic nodes are used to represent certain outcomes, similar to our approach of simulating the deterministic Carneades model through conditional probabilities. Also in the decision-theory context, Druzdzel & Suermondt [7] surveyed methods to identify random variables in a network relevant to a query given some evidence, thereby also relying on deterministic nodes. GeNIe & SMILE [6] provide a solid implementation of graphical decision-theoretic models including functionality for deterministic nodes.

9. Conclusions and Future Work

We have simulated Carneades argument evaluation structures using a specially constructed class of BNs and instances thereof. The translation centers around argument- and statement-specific probability functions which are used to compute conditional probability tables. The translation into BNs suggests a way to extend the Carneades model to handle argumentation with uncertain premises, where the premises are assigned a priori probabilities. Future research will focus on evaluating the utility of the extended Carneades model for quantitative decision support systems, reducing inference complexity, as well as on the implications for argumentation theory of viewing claims and decisions as having probabilities.

10. Acknowledgments

The authors thank Prof. Kevin Ashley, Collin Lynch and Saeed Amizadeh (Intelligent Systems Program, University of Pittsburgh) as well as Prof. Henry Prakken (Computer Science Department, Utrecht University; Law Faculty, University of Groningen) for very helpful suggestions and feedback during the process of writing this paper.

References

[1] T. J. Bench-Capon, S. Doutre, and P. E. Dunne. Audiences in argumentation frameworks. *Artificial Intelligence*, 171(42-71), 2007.

[2] T. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.

[3] V. Carofiglio. Modelling Argumentation with Belief Networks. 4th Workshop on Computational Models of Natural Argument, ECAI 2004.

[4] L. J. Cohen. An Introduction to the Philosophy of Induction and Probability. Clarendon Press, 1989.

[5] G. F. Cooper. Probabilistic Inference using Belief Networks is Np-Hard. Paper No. SMI-87-0195, Knowledge Systems Laboratory, Stanford University. Stanford, CA, USA, 1987.

[6] See the GeNIe & SMILE software: http://genie.sis.pitt.edu

[7] M. J. Druzdzel and H. J. Suermondt. Relevance in Probabilistic Models: "Backyards" in a "Small World". Working notes of the AAAI–1994 Fall Symposium Series: Relevance, 60-63, 1994.

[8] J. B. Freeman. Argument Strength, the Toulmin Model, and Ampliative Probability. In: van Eemeren, Garssen, B. (eds.), Pondering on Problems of Argumentation, 2009.

[9] T. F. Gordon, H. Prakken, and D. Walton. The Carneades model of argument and burden of proof. Artificial Intelligence 171, 10-11, 875-896, 2007.

[10] T. F. Gordon and D. Walton. Proof Burdens and Standards. Argumentation in Artificial Intelligence, I. Rahwan and G. Simari, Eds. Springer-Verlag, Berlin, Germany, 2009.

[11] T.F. Gordon and S. Ballnat. Goal Selection in Argumentation Processes. Proceedings of COMMA 2010.

[12] R. A. Howard and J. E. Matheson. Influence diagrams. Readings on the Principles and Applications of Decision Analysis, pages 721-762. Strategic Decisions Group, 2003.

[13] J. B. Kadane and D. A. Schum. A Probabilistic Analysis of the Sacco and Vanzetti Evidence. Wiley & Sons, 1996.

[14] P. Krause, S. Ambler, M. Elvang-Goransson, and J. Fox. A Logic of Argumentation for Reasoning under Uncertainty. Computational Intelligence 11, 113-131, 1995.

[15] N. Muecke and A. Stranieri. An argument structure abstraction for bayesian belief networks: just outcomes in on-line dispute resolution. Proceedings of the fourth Asia-Pacific Conference on Conceptual Modeling, pp. 35-40, 2007.

[16] S. Parsons. Qualitative Methods for Reasoning under Uncertainty. MIT Press, 2009.

[17] J. Pearl. Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning. (UCLA Technical Report CSD-850017). Proceedings of the 7th Conference of the Cognitive Science Society, 1985.

[18] R. Riveret, H. Prakken, A. Rotolo and G. Sartor. Heuristics in Argumentation: A Game-Theoretical Investigation. Proceedings of COMMA 2008, pp. 324-335, 2008.

[19] M. Williams and J. Williamson. Combining Argumentation and Bayesian Nets for Breast Cancer Prognosis. Journal of Logic, Language and Information 15: 155-178, 2006.

[20] G. A. Vreeswijk. Argumentation in Bayesian Belief Networks. Argumentation in Multi-Agent Systems, LNAI 3366, pp. 111-129, 2005.

[21] I. Zukerman, R. Mcconachy, and K. B. Korb. Bayesian Reasoning in an Abductive Mechanism for Argument Generation and Analysis. Proceedings of the Fifteenth National Conference on Artificial Intelligence, pp. 833-838, 1998.