Computational Models of Argument
P. Baroni et al. (Eds.)

doi:10.3233/978-1-61499-686-6-95

From Arguments to Constraints on a Bayesian Network

Floris BEX [a], Silja RENOOIJ [a]
[a] *Information and Computing Sciences, Utrecht University, The Netherlands*

Abstract. In this paper, we propose a way to derive constraints for a Bayesian Network from structured arguments. Argumentation and Bayesian networks can both be considered decision support techniques, but are typically used by experts with different backgrounds. Bayesian network experts have the mathematical skills to understand and construct such networks, but lack expertise in the application domain; domain experts may feel more comfortable with argumentation approaches. Our proposed method allows us to check Bayesian networks given arguments constructed for the same problem, and also allows for transforming arguments into a Bayesian network structure, thereby facilitating Bayesian network construction.

1. Introduction

Bayesian networks, graphical representations of probability distributions, are very well suited to epistemic reasoning because they capture the probabilistic (in)dependencies between variables in the domain of discourse. They have found a number of applications in domains such as medicine, forensics and the law [1]. However, constructing a Bayesian network and understanding the modelled influences between variables requires knowledge of the Bayesian network formalism, which means that domain experts (doctors, lawyers) can often only construct a network with the help of a modeller with the relevant mathematical background. In contrast, argumentation approaches can be said to more closely follow the reasoning of the domain experts, especially in legal or organizational contexts which are less mathematically inclined [2, 3]. Our aim is to bridge the communication gap between domain experts and Bayesian network analysts by developing a better understanding of the relation between the two kinds of reasoning.

The relations between arguments and Bayesian networks (BNs) can be considered from two directions. For the first direction, BNs are transformed into arguments or argument diagrams [4–6], which allows the knowledge captured in the BN to be understood more easily by domain experts accustomed to argumentative reasoning. This does not directly help domain experts to construct a BN based on argumentative reasoning, for which we need to transform in the other direction, that is, from arguments to BNs. A classic example is [3], who use Wigmore graphs, which are very similar to evidential arguments [7], as the basis for BNs. Schum and Kadane, however, do not provide a formal definition of their transformation. Such a definition is given in [8], in which Carneades argument evaluation structures are transformed into BNs, thus allowing existing BNs to be extended with the information contained in the arguments. The main focus in [8], however, is on simulating the Carneades method of evaluating arguments through BNs. In contrast, the main aim of this article is to explore how domain knowledge expressed as argument structures can inform the construction of BNs.

One of the difficulties of interpreting generic argument structures as a BN is that theoretical assumptions have to be made about, for example, the causal direction [9] and strength of the inference rules [10, 11]. From a practical point of view, it is also an issue that different BNs can represent the same reasoning. Consider for example a case where a forensic specialist constructs a BN for a part of the case, while a judge constructs an argument about the same part of the case. If we directly translate this argument, we might not end up with the exact same network as the forensic specialist, but this does not mean that the judge and the forensic specialist disagree, as different BNs can represent the same probability distribution. Because a set of evidential arguments therefore do not – and cannot – uniquely define a BN about the same part of the case, it makes sense to transform the arguments to a set of constraints that can be met by multiple possible BNs.

We hence distinguish two cases for transforming arguments to (constraints on) BNs. In the first case, a network already exists – hand-crafted by experts or learned from data – for the same case as the arguments and we want to check if the BN complies with the arguments. Here, the arguments indicate a number of basic *constraints* a BN has to adhere to. In the second case, we use the identified constraints to construct a new BN on the basis of the available arguments. Because the constraints are not exhaustive (i.e., given a set of constraints based on an argument there will be multiple BNs that adhere to these constraints), we propose a general *heuristic* for transforming arguments into BNs.

Sections 2 and 3 briefly discuss the formal preliminaries: a structured argumentation framework and Bayesian networks. In Section 4.1, we then discuss possible constraints on the graph of a BN given structured arguments, and in Section 4.2 we propose a heuristic for transforming arguments to a BN. Note that initially, the focus is on constraints on the *structure* (i.e. the graph) of the Bayesian network. Deriving constraints on the (conditional) probabilities of a network from structured arguments is more difficult, because it depends on probabilistic interpretations of argumentation which are more contentious. In Section 5 we briefly discuss some possible constraints on the (conditional) probabilities in the Bayesian network, and how they can be used to expand our heuristic.

2. Structured Argumentation

In this section, we define a simple propositional system for argumentation based on the ASPIC$^+$ framework [12], which captures the basic elements of structured argumentation. Because the idea is that domain experts have to be able to relatively easily construct arguments, we want to impose as few formal constraints as possible on these arguments. As has been shown [7], ASPIC$^+$ allows for arguments that are very similar to the Wigmore graphs [3] with which many lawyers and judges are familiar.

Definition 2.1 [**Argumentation systems**] An *argumentation system* is a triple $AS = (\mathcal{L}, \mathcal{R}, n)$ where:

- $\mathcal{L}$ is a logical language with contrariness function $^- : \mathcal{L} \to 2^{\mathcal{L}}$.
- $\mathcal{R}_s$ and $\mathcal{R}_d$ are two disjoint sets of strict ($\mathcal{R}_s$) and defeasible ($\mathcal{R}_d$) inference rules of the form $\varphi_1, \ldots, \varphi_n \to \varphi$ and $\varphi_1, \ldots, \varphi_n \Rightarrow \varphi$ respectively (where φ_i, φ are meta-variables ranging over well-formed formulas in $\mathcal{L}$).

- n is a naming convention for defeasible rules, which to each rule r in $\mathcal{R}_d$ assigns a well-formed formula φ from $\mathcal{L}$ (written as $n(r) = \varphi$).

Here, $\mathcal{L}$ is a propositional language, where φ is a *contrary* of ψ if $\varphi \in \overline{\psi}$ and $\psi \notin \overline{\varphi}$ (i.e. asymmetrical conflict) and φ is a *contradictory* of ψ, denoted by '$\varphi = -\psi$', if $\varphi \in \overline{\psi}$ and $\psi \in \overline{\varphi}$ (i.e. symmetrical conflict). Furthermore, $\mathcal{R}_s$ contains the inference rules of classical logic. Whereas in evidential reasoning we can distinguish between causal defeasible rules (fire causes smoke) and evidential defeasible rules (smoke is evidence for fire) [9], for now we make no assumptions as to the type of rule used. Finally, note that informally, $n(r)$ is a wff in $\mathcal{L}$ which says that rule $r \in R$ is applicable.

Definition 2.2 [**Knowledge bases**] A *knowledge base* in an $AS = (\mathcal{L}, \mathcal{R}, n)$ is a set $\mathcal{K} \subseteq \mathcal{L}$ consisting of two disjoint subsets $\mathcal{K}_e$ (the *evidence*) and $\mathcal{K}_p$ (the *ordinary premises*).

The evidence in $\mathcal{K}_e$ is similar to axiom premises [12], which cannot be denied or attacked. If, for example, a witness testimony is presented as evidence, then the existence of the testimony cannot be denied; of course we can still question its veracity. Ordinary premises can be undermined by other arguments (see Definition 2.4).

Arguments can be constructed from knowledge bases by chaining inference rules into trees. Here, for any argument A, the function `Sub` returns all sub-arguments of A; `Prop` returns all the formulas in A; `Prem` returns all the formulas of $\mathcal{K}$ (called *premises*) used to build A, `Conc` returns A's conclusion, `Rules` returns all inference rules in A and `TopRule` returns the last inference rule used in A.

Definition 2.3 [**Arguments**] An *argument* A on the basis of a knowledge base $\mathcal{K}$ in an argumentation system $AS = (\mathcal{L}, \mathcal{R}, n)$ is:

1. φ if $\varphi \in \mathcal{K}$ with: $\texttt{Prem}(A) = \{\varphi\}$; $\texttt{Conc}(A) = \varphi$; $\texttt{TopRule}(A)$ = undefined; $\texttt{Prop}(A) = \{\varphi\}$; $\texttt{Sub}(A) = \{\varphi\}$.

2. $A_1, \ldots, A_n \rightarrow/\Rightarrow \psi$ if $A_1, \ldots, A_n$ are arguments such that there exists a strict/defeasible rule $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow/\Rightarrow \psi$ in $\mathcal{R}_s/\mathcal{R}_d$, with:
   $\texttt{Prem}(A) = \texttt{Prem}(A_1) \cup \ldots \cup \texttt{Prem}(A_n)$;
   $\texttt{Conc}(A) = \psi$;
   $\texttt{TopRule}(A) = \texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow/\Rightarrow \psi$;
   $\texttt{Prop}(A) = \texttt{Prop}(A_1) \cup \ldots \cup \texttt{Prop}(A_n) \cup \{\psi\}$,
   $\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n) \cup \{A\}$;
   $\texttt{Rules}(A) = \texttt{Rules}(A_1) \cup \ldots \cup \texttt{Rules}(A_n) \cup \{\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow/\Rightarrow \psi\}$

Arguments can be attacked in essentially two ways: on their conclusion (rebutting attack) or on a defeasible inference step (undercutting attack).

Definition 2.4 [**Attack**] *A attacks B* iff *A undercuts* or *rebuts*, where:

- *A undercuts* argument B (on r) iff $\texttt{Conc}(A) \in \overline{n(r)}$ for some $B' \in \texttt{Sub}(B)$ such that $\texttt{TopRule}(B) = r$ and r is defeasible.
- *A rebuts* argument B (on φ) iff $\texttt{Conc}(A) \in \overline{\varphi}$ and $\texttt{Conc}(B') = \varphi$ for some $B' \in \texttt{Sub}(B)$ where either $\texttt{TopRule}(B)$ is defeasible or $\varphi \in \mathcal{K}_p$.

Argumentation systems plus knowledge bases form argumentation theories, which induce structured argumentation frameworks. Note that we do not include an ordering on the arguments, as this is not needed for current purposes. Furthermore, because we only use the structure of arguments and not take the evaluation of arguments into account for our constraints on Bayesian networks, we will not discuss any of the possible argumentation semantics for ASPIC$^+$ as given in [12].

Definition 2.5 [**Structured Argumentation Frameworks**] Let AT be an *argumentation theory* $(AS, \mathcal{K})$. A *structured argumentation framework (*SAF*)* defined by AT, is a pair $\langle \mathcal{A}, \mathcal{C} \rangle$ where $\mathcal{A}$ is the set of all finite arguments constructed from $\mathcal{K}$ in AS and $(X, Y) \in \mathcal{C}$ iff X attacks Y.

Example 2.6 As an example of an argument, suppose that a burglary has taken place and that we are interested in whether some suspect is guilty of committing the burglary (`Bur`). Forensic analysis (`For`) shows a match between a pair of shoes owned by the suspect and footprints (`Ftpr`) found near the crime scene. However, there is also evidence that there may have been a mix up at the forensic lab (`Mix`): the exact history of the footprints from crime-scene to lab has not been properly documented. This suspect had a motive (`Mot`) to commit this burglary, which is confirmed by at least one reliable testimony (`Tes1`). Furthermore, it is argued that the suspect also had the opportunity (`Opp`) to commit this burglary, but this is denied (`-Opp`) by a further testimony of the suspect himself (`Tes2`). We can now build arguments based on the evidence, where $\mathcal{K}_e = \{\texttt{For}, \texttt{Mix}, \texttt{Tes1}, \texttt{Tes2}\}$ and $\mathcal{K}_p = \{\texttt{Opp}\}$. Furthermore, $\texttt{Mix} \in \overline{r_{for}}$, where $r_{for} : \texttt{For} \Rightarrow \texttt{Ftpr}$ is the rule applied in A'_1.

A_1: `For`	A_3: `Tes1`	B_1: `Tes2`	C_1: `Mix`
A'_1: $A_1 \Rightarrow$ `Ftpr`	A'_3: $A_3 \Rightarrow$ `Mot`	B'_1: $B_1 \Rightarrow$ `-Opp`	
A_2: `Opp`	A_4: $A'_1, A_2, A'_3 \Rightarrow$ `Bur`		

Figure 1 shows the arguments, where inferences are modelled as dashed arrows (all the inferences in this example are defeasible) and attacks as thick black arrows. Grey propositions are in $\mathcal{K}_e$.

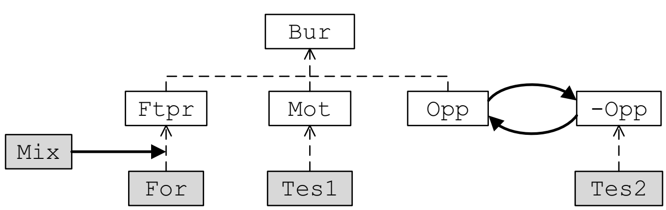

Figure 1. An argument in the example case

3. Bayesian Networks

In this section we briefly review Bayesian networks which combine a graph with conditional probability tables (CPTs) to compactly represent a joint probability distribution over a set of stochastic variables [13].

Definition 3.1 [**Bayesian network**] A BN is a triple $\mathcal{B} = (\mathbf{V}, \mathbf{A}, \mathcal{P})$ where:

- $G = (\mathbf{V}, \mathbf{A})$ is an acyclic directed graph with nodes $\mathbf{V}$ and arcs $\mathbf{A} \subset \mathbf{V} \times \mathbf{V}$ (arc (V_i, V_j) is directed from V_i to V_j);
- $\mathcal{P} = \{\mathrm{Pr}_V \mid V \in \mathbf{V}\}$ where each Pr_V is a set of (conditional) distributions $\mathrm{Pr}(V \mid \mathrm{par}(V))$ over variables $V \in \mathbf{V}$, one for each combination of values for the parents $\mathrm{par}(V)$ of V in graph G; these distributions are typically represented as tables (CPTs).

Note from the above definition that in a BN there is a one-to-one correspondence between nodes and stochastic variables. Moreover, a BN allows for defining a joint probability distribution that respects the independences portrayed by its digraph (see below).

Proposition 3.2 The BN $\mathcal{B} = (\mathbf{V}, \mathbf{A}, \mathcal{P})$ uniquely defines the following joint probability distribution $\mathrm{Pr}(\mathbf{V})$:

$$\mathrm{Pr}(\mathbf{V}) = \prod_{V \in \mathbf{V}} \mathrm{Pr}(V \mid \mathrm{par}(V))$$

A BN thus allows for computing any probability of interest over its variables; a typical query of interest is the probability $\mathrm{Pr}(h \mid \mathbf{e})$ of some hypothesis h given a combination of observations $\mathbf{e}$ for a set of observed variables $\mathbf{E}$. The computation of probabilities from the network specification is called inference and the Bayesian network framework includes various algorithms to this end.

Although the directed arcs in a BN graph may suggest that the represented relations are causal, the arcs in fact only have meaning in combination with other arcs: together they capture the independences among the represented variables by means of the graphical *d-separation criterion* [13].

Definition 3.3 [**d-separation**] Consider three sets of nodes $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ in graph G. In addition, consider a simple chain s in G.

- Chain s is said to be *blocked*, or *inactive*, given $\mathbf{Z}$ if the chain contains a node with two incoming arcs on the chain (a head-to-head node) which is not in $\mathbf{Z}$ and has no descendants in $\mathbf{Z}$, or it contains a node in $\mathbf{Z}$ that has at most one incoming arc on the chain; a chain that is not blocked by $\mathbf{Z}$ is said to be *active* given $\mathbf{Z}$.
- $\mathbf{X}$ and $\mathbf{Y}$ are said to be *d-separated* by $\mathbf{Z}$ if all possible chains s between nodes in $\mathbf{X}$ and $\mathbf{Y}$ are *inactive* given $\mathbf{Z}$.
- If $\mathbf{X}$ and $\mathbf{Y}$ are d-separated by $\mathbf{Z}$, then the two corresponding (sets of) variables $\mathbf{X}$ and $\mathbf{Y}$ are probabilistically independent given the third set $\mathbf{Z}$.

Note from the latter statement that the values of the variables in $\mathbf{Z}$ are assumed to be actually observed and therefore known. We assume that two nodes that are directly connected by an arc are not d-separated.

Example 3.4 Figure 2 shows the graph of a Bayesian network, where observed variables have been shaded; the conditional probabilities are for now left implicit. Note that, compared to the argument graph in Figure 1 there is an additional variable, `Rel`, representing the reliability of the witness that gave testimony 1. Note that, for example, given evidence

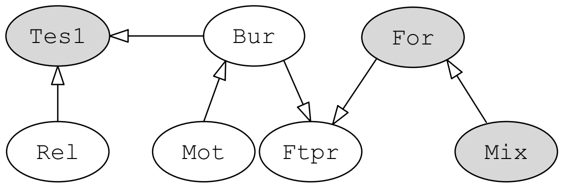

Figure 2. An Bayesian network graph for the example case

for `Tes1`, nodes `Rel` and `Bur` are not d-separated: given the witness testimony, their reliability will influence the probability of the suspect having committed the burglary. On the other hand, `Mix` and `Ftpr` are d-separated due to the observation for `For`.

4. From Argumentation to (Constraints on) Bayesian Network Graphs

A structured argumentation framework in ASPIC$^+$ can be used as the basis for (constraints on) the structure of a Bayesian network. First, we will discuss a number of generic constraints that can be put on a Bayesian network given an argumentation framework. After this, we will provide a heuristic for building a Bayesian network graph given these constraints based on arguments.

4.1. Constraints on the Structure of a Bayesian Network

First, we need to interpret all the elements of an argumentation framework in terms of the structure of a BN graph. The graph of a BN conveys the (in)dependencies between variables represented by the nodes in the graph. With structural constraints we denote both properties of nodes (which variables must be represented, and what are their values) and chains of arcs. In this paper we assume that all variables V are binary-valued, with possible values $\texttt{Val}(V) = \{\mathit{true}, \mathit{false}\}$. Since (in)dependencies are dynamic and can change depending on the observed variables, we will also include observations for variables in the structural constraints.

Constraint 4.1 [**Nodes and values**] Given a $\mathit{SAF} = \langle \mathcal{A}, \mathcal{C} \rangle$, the following constraints can be put on the nodes $\mathbf{V}$ of a Bayesian network.

- For every atomic proposition v or $-v$ in $\texttt{Prop}(A)$, where $A \in \mathcal{A}$, there exists a single node $V \in \mathbf{V}$ such that $v \equiv V = \mathit{true}$ and $-v \equiv V = \mathit{false}$;
- Every atomic proposition e_i or $-e_i$ in $\mathcal{K}_e$ is taken to represent the observed value of node E_i from the set of observed variables $\mathbf{E} \subset \mathbf{V}$.

We say that propositions v and $-v$ are *associated with* node V.

We now address the (chains of) arcs $\mathbf{A}$ that capture the (in)dependencies between variables in a BN. Whenever a rule is applied to infer some conclusion from a set of premises, we should be able to reason from each node associated with the premises to the node associated with the consequence in the BN. That is, given the context of the evidence upon which an argument is built, there should exist active chains between the nodes associated with the applied rules.

Constraint 4.2 [**Inference chains**] Given a *SAF* $= \langle \mathcal{A}, \mathcal{C} \rangle$, the following constraints based on inferences can be put on the chains of arcs $\mathbf{A}$ of a Bayesian network.

- For every rule $r : \varphi_1, \ldots, \varphi_n \rightarrow \psi$ or $r : \varphi_1, \ldots, \varphi_n \Rightarrow \psi$ such that $r = \texttt{Rules}(A)$ and $A \in \mathcal{A}$, there exist active chains between each of the nodes associated with $\varphi_1, \ldots, \varphi_n$ and the node associated with ψ, given the observed nodes $\mathbf{E}$ associated with evidence in $\mathcal{K}_e$.

An argumentation framework also includes a set of attack relations $\mathcal{C}$. As for inference, it makes sense to assume that whenever two propositions in arguments are in conflict (i.e. contrary or contradictory), that there is an influence between the values representing these propositions in the BN. Because contradictory propositions φ and $-\varphi$ are captured as two values of a single variable (see Constraint 4.1), not all of the propositions will translate to separate nodes in the BN, and hence the influences between the values associated with φ and $-\varphi$ will be captured in the (conditional) probabilities involving these values. In the case of φ being contrary to ψ – which occurs in rebutting attacks – the influence should be captured as an active chain between the two associated nodes given the evidence. Similarly, if a proposition undercuts the inference from one proposition to another then, given the evidence, there should be active chains between the nodes associated with this undercutting attack.

Constraint 4.3 [**Attack chains**] Given a *SAF* $= \langle \mathcal{A}, \mathcal{C} \rangle$, the following constraints based on attacks can be put on the chains of arcs $\mathbf{A}$ of a Bayesian network. For every attack relation $(A, B) \in \mathcal{C}$:

- if $\texttt{Conc}(A) = \varphi$ is a *contrary* of $\texttt{Conc}(B) = \psi$, then there exists an active chain between the node associated with φ and the node associated with ψ, given the observed nodes $\mathbf{E}$ associated with evidence in $\mathcal{K}_e$.
- if A *undercuts* B on r, where $\texttt{Conc}(A) = \chi$ and $r : \varphi_1, \ldots, \varphi_n \Rightarrow \psi$, then there exist active chains between the node associated with χ and the nodes associated with $\varphi_1, \ldots, \varphi_n, \psi$, given the observed nodes $\mathbf{E}$ associated with evidence in $\mathcal{K}_e$.

Example 4.4 In our case, a Bayesian network expert with limited domain knowledge builds the network in Figure 2, and a judge builds the argument in Figure 1. The question is now whether, given the evidence, the network conforms to the constraints imposed by the arguments. In the argumentation framework of the case, there is an argument about the suspect having the opportunity to commit the burglary (`Opp`), and a counterargument based on the suspect's testimony (`Tes2`). The BN, however, does not include variables representing `Opp` or `Tes2` and it thus violates some of the node constraints posed by the argumentation framework. Note that the BN also includes additional information that is not captured in the argumentation framework: there is a variable representing the reliability of a witness testimony (`Rel`) whereas none of the arguments includes a proposition about witness reliability. Remember that our current aim is to use the argumentation framework to put constraints on the BN and not vice versa, so any additional knowledge or reasoning in the BN is fine as long as it does not lead to violation of the constraints.

In order to determine the active chains in the BN of Figure 2, we have to consider the set of observations $\mathbf{E}$ based on the evidence in $\mathcal{K}_e$ and the nodes in the BN, so $\mathbf{E} = \{\texttt{For} = \mathit{true}, \texttt{Mix} = \mathit{true}, \texttt{Tes1} = \mathit{true}\}$. With respect to inference chains, we now see that there are active chains from Ftpr to For (inference in A_1'), Tes1 to Mot (inference in A_3'), Ftpr to Bur and Mot to Bur (both part of the inference in A_4'). Because the node representing Opp is not in the network, there is no chain from Opp to Bur as per the constraint based on the top rule application in A_4. Furthermore, because Tes2 is not a node in the network, there is no active chain from Tes2 to Opp, which is required by both the inference chain constraint and the first attack chain constraints. Finally, the second attack chain constraint says that, because Mix undercuts the application of r_{for} : For $\Rightarrow$ Ftpr in A_1', there should be active chains from Mix to For and from Mix to Ftpr. While these chains are both in the BN, the latter one, from Mix to Ftpr is blocked by the observation For $= true$, so the constraint is violated.

4.2. Transforming Arguments into Bayesian Network Graphs

In this section we propose a heuristic for the construction of a BN from arguments. Our BN graph construction heuristic starts by building an undirected skeleton in which all nodes that should be connected by active chains, according to the constraints from section 4.1, are connected directly by an edge. The resulting undirected graph will typically be densely connected and represent only few independencies. After this, it has to be decided which connections to retain and how to direct the edges. Recall that arguments are typically constructed by performing successive inference steps based on evidence. The rules behind these inferences can be either causal (fire causes smoke) or evidential rules (smoke is evidence for fire) [9]. In BNs – though the graph is just a representation of an independence relation – people tend to attach a causal interpretation to the arcs, and the notion of causality is typically used as a heuristic to guide the construction of the graph with the help of domain experts [13]. Hence, we propose to use the same heuristic for choosing the direction of the arcs. Furthermore, undercutting can be seen as a form of intercausal reasoning called *explaining away* (see Section 5). In BNs head-to-head nodes (see Definition 3.3) are explicitly used to model induced intercausal dependencies, such as exploited in explaining away, and we therefore propose to use head-to-head connections among the nodes involved in an undercutting attack. Our heuristic is then as follows:

BN graph construction heuristic

1) construct nodes according to Constraint 4.1;
2) each active chain dictated by Constraints 4.2 and 4.3 is implemented as a direct, *un*directed edge;
3) for all undercutting attacks: if a rule $r : \varphi_1, \ldots, \varphi_n \Rightarrow \psi$ is undercut by $\chi \in \overline{n(r)}$ then remove edge (χ, ψ) (or (ψ, χ)) and turn the other edges involved into the following directed ones: (χ, φ_i) and (ψ, φ_i) for all φ_i;
4) for all remaining undirected edges, choose the causal direction if a causal interpretation is possible, and an arbitrary direction, otherwise;
5) verify that the graph is acyclic and that all chains that should be active indeed are; otherwise remove or reverse appropriate arcs in consultation with a domain expert.

With respect to directing the edges we remark that BN graphs which share the same skeleton (undirected edges) and the same immoralities (head-to-head nodes for which the parents are not directly connected) are said to be Markov equivalent and capture the same independence relation. Arc reversal therefore does not change the modelled independence relation as long as the resulting graph is Markov equivalent.

Example 4.5 Using the heuristic described above on the arguments depicted in Figure 1 we can first construct an undirected graph G' (Figure 3). We then apply the third step of the heuristic: `Mix` – which in the argumentation framework is modelled as an undercutter of the rule `For` $\Rightarrow$ `Ftpr` – is now captured in the BN as a cause of `For`, together with `Ftpr`. The observation of head-to-head node `For` makes the chain between `Mix` and `Ftpr` active, so evidence for a mix-up will now change the probability of the footprints being the cause of the forensic results – further probabilistic constraints (see Section 5) could ensure that this intercausal effect is indeed "explaining away". For the fourth step of the heuristic, we can quite sensibly interpret the edges causally: having a motive and opportunity cause the suspect to commit the burglary, which in turn causes the footprints to be found. Furthermore, the evidential observations are caused by the events that led to them. Now it can be verified that all chains that should be active given $\mathbf{E} = \{\texttt{Mix}, \texttt{For}, \texttt{Tes1}, \texttt{Tes2}\}$ indeed are, and we end up with graph G in Figure 3.

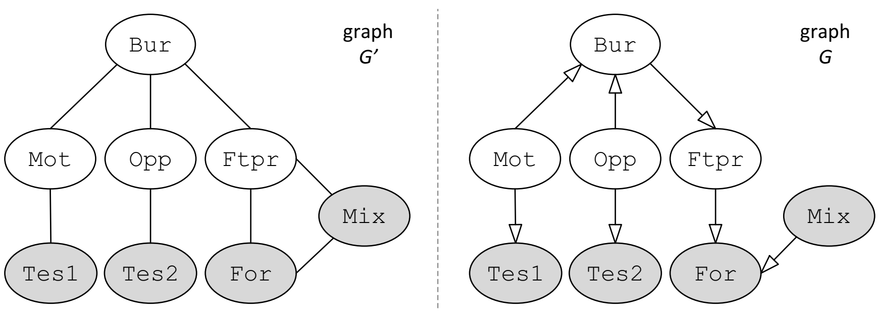

Figure 3. Undirected and directed graphs based on the argument in Figure1

The structure of the final graph G is quite close to the structure of the original arguments in Figure 1. This structural similarity can also be seen in [3]: in a Wigmore diagram, the arrows denote evidential influences, which are then reversed in the causal direction to form a BN graph. A difference between the informal heuristic of [3] and ours is that we do not assume all influences in the argument to be evidential.

5. Constraints on the Probabilities of a Bayesian Network

Since a BN represents a probability distribution with its independence relation, an important part of the BN is formed by the (conditional) probabilities over the variables in a network. Putting constraints on these probabilities based on the evidential arguments in a case requires a probabilistic interpretation of such arguments, which is not straightforward (see, a.o., [8, 10, 11, 14]). A full exploration of probabilistic constraints is therefore beyond the scope of this paper. Rather, we will show a few possibilities of transforming argument structures into probabilistic constraints, and briefly discuss some of the issues concerning a probabilistic interpretation of structured arguments.

As for structural constraints, conditional probability constraints can be derived from the inferences and attacks in an argumentation framework. The interpretation of a strict rule, for example, is fairly straightforward: the conclusion is necessarily true given the premises. However, with respect to defeasible rules different ideas exist on how they should be modelled probabilistically.

Constraint 5.1 [**Inference probabilities**] Given a *SAF* $= \langle \mathcal{A}, \mathcal{C} \rangle$, the following constraints can be put on the distribution Pr defined by a Bayesian network.

- For every strict rule $r : \varphi_1, \ldots, \varphi_n \rightarrow \psi$ such that $r = \texttt{Rules}(A)$, $A \in \mathcal{A}$, we have that $\Pr(\psi \mid \varphi_1, \ldots, \varphi_n) = 1$;
- For every defeasible rule $r : \varphi_1, \ldots, \varphi_n \Rightarrow \psi$ such that $r = \texttt{Rules}(A)$, $A \in \mathcal{A}$, we have that $\Pr(\psi \mid \varphi_1, \ldots, \varphi_n) > 0$.

Note that all probabilities are taken to hold in the context of the available evidence, that is, each probability should also be conditioned on the subset of evidence for the nodes not associated with the rule. The above constraint for defeasible rules, which has been adapted from [11], is fairly weak. It can, for example, be argued that the premise should make it more likely than not that the conclusion is true ($\Pr(\psi \mid \varphi) > 0.5$), or that the premises are only a relevant reason for the conclusion if they increase the belief in the conclusion (e.g. $\Pr(\psi \mid \varphi) > \Pr(\psi)$ [10]). The point is that there are many measures of support [14], and choosing exactly which ones to use to interpret inference is not trivial.

As for attacks, it makes sense to interpret rebuttal using classical negation, as rebutting propositions φ and $-\varphi$ are values of one variable. Undercutting, where the application of a rule is attacked, can be seen as a form of *explaining away*, where the influence of a certain observation should be removed, or at least weakened, given an additional observation [13] (for more on the relations between undercutting evidential arguments and alternative causal explanations also see [9]). In BNs, explaining away is considered to be a type of intercausal reasoning often found between two causes (say ψ and χ) of a common effect (say φ): if the effect is observed, both causes become more likely; however, if we subsequently know which cause actually occurred (e.g. χ), the probability of the other cause (ψ) being present as well decreases.

Constraint 5.2 [**Attack probabilities**] Given a *SAF* $= \langle \mathcal{A}, \mathcal{C} \rangle$, the following constraints based on attacks can be put on the conditional probabilities of a BN. For every attack relation $(A, B) \in \mathcal{C}$:

- if $\texttt{Conc}(A) = \varphi$ and $\texttt{Conc}(B) = \psi$ and $\varphi \in \overline{\psi}$, then $\Pr(\psi \mid \varphi) = 0$.
- if $\texttt{Conc}(A) = \chi \in \overline{n(r)}$ and $r : \varphi_1, \ldots, \varphi_n \Rightarrow \psi$ such that $r = \texttt{Rules}(B)$, then $\Pr(\psi \mid \varphi_1, \ldots, \varphi_n, \chi) < \Pr(\psi \mid \varphi_1, \ldots, \varphi_n)$.

A stronger version of the second constraint says that $\Pr(\psi \mid \varphi, \chi) = 0$ [11], which fits with the idea that undercutting arguments always defeat the argument they attack [12]. Similarly, in [8] an argument is not applicable given an exception.

Example 5.3 Again consider the arguments in Figure 1 and the BN in Figure 2. Consider the two defeasible rules r_{for} : $\texttt{For} \Rightarrow \texttt{Ftpr}$ and r_{mot} : $\texttt{Mot} \Rightarrow \texttt{Bur}$, and the undercutting attack (C_1, A_1'). In the context of the relevant evidence $\mathbf{e}'$, our choice of probabilistic interpretation now gives the following constraints:

$$\Pr(\texttt{Ftpr} = \mathit{true} \mid \texttt{For} = \mathit{true}, \mathbf{e}') > 0$$

$\Pr(\texttt{Ftpr} = \mathit{true} \mid \texttt{For} = \mathit{true}, \texttt{Mix} = \mathit{true}, \mathbf{e}') < \Pr(\texttt{Ftpr} = \mathit{true} \mid \texttt{For} = \mathit{true}, \mathbf{e}')$

From Figure 2 we see that given For, the effect of a mix-up on footprints is blocked, which means that the last constraint is violated since due to the blocking influence of Mix it holds that $\Pr(\texttt{Ftpr} = \mathit{true} \mid \texttt{For} = \mathit{true}, \texttt{Mix} = \mathit{true}, \mathbf{e}') = \Pr(\texttt{Ftpr} = \mathit{true} \mid \texttt{For} = \mathit{true}, \mathbf{e}')$.

In addition to rules and attacks, there are also other aspects of argumentation frameworks that can be interpreted probabilistically. For example, ASPIC$^+$ includes the option to define preferences over ordinary premises in $\mathcal{K}_p$ and defeasible rules in $\mathcal{R}_d$, which translate more or less directly to constraints on probabilities: if premise φ is preferred over premise ψ, then $\Pr(\varphi) > \Pr(\psi)$. Furthermore, it is also possible to capture the evaluation of arguments probabilistically, as in [8], where the (conditional) probabilities depend on the status of propositions in the Carneades argument evaluation structure.

Once we have a probabilistic interpretation of an argumentation framework, we can use the probability constraints in an elicitation procedure for obtaining the required local probability distributions [15]. However, the above discussion shows that there are different ways to interpret arguments probabilistically, and that different aspects of argumentation frameworks can be incorporated as constraints on probabilities. The difficulty here lies in the fact that the exact probabilistic interpretation of arguments and evidence, and hence the various types of constraints on a BN, is a contentious issue. In fact, one way to deal with discussions surrounding constraints on probabilities is to allow arguments *about* the various probabilistic constraints in a BN [16].

6. Conclusion

In this paper we proposed a method for establishing constraints on Bayesian networks from structured arguments. Using our method, the type of arguments typically constructed by, for example, legal experts can be used to check BNs and in particular the knowledge contained in them. Observed differences between the arguments and a BN may carry useful information, as they point to possible differences in the reasoning of a Bayesian network expert and the domain expert. Hence, the communication gap between these two types of experts can be bridged.

In addition to deriving constraints for BNs from a set of arguments, we have also designed a heuristic for constructing a BN graph based on these constraints. In this way, the constraints can help with the construction of a BN: domain experts can build evidential arguments, which are then transformed into BNs or BN skeletons, which can in turn be interpreted and further extended by Bayesian network experts in conjunction with the domain experts. The proposed heuristic thus expands the toolbox for building a BN relatively easily, and can be used together with existing techniques for building BNs such as idioms, recurrent BN structures based on standard arguments such as inductive arguments or arguments based on evidence [1].

One of the key issues when transforming arguments into BNs is the probabilistic interpretation of the elements of argumentation. It turns out that there are numerous possible interpretations of evidential support and argument strength [10, 11, 14], which obviously lead to different constraints on the probability distribution represented by a BN.

It is at present not clear if and how the choice of constraints influences the outcomes of the reasoning in BNs. Furthermore, the choice of interpretation also depends on the (implicit) assumptions regarding evidential argument strength the domain experts that build structured arguments have. We would like to address these questions in future research.

References

[1] N. Fenton and M. Neil. *Risk assessment and decision analysis with Bayesian networks*. CRC Press, 2012.
[2] V. Franqueira, T.T. Tun, Y. Yu, R. Wieringa, and B. Nuseibeh. Risk and argument: a risk-based argumentation method for practical security. In *19th IEEE International Requirements Engineering Conference*, pages 239–248. IEEE, 2011.
[3] J.B. Kadane and D.A. Schum. *A probabilistic analysis of the Sacco and Vanzetti evidence*, volume 773. John Wiley & Sons, 2011.
[4] J. Keppens. Argument diagram extraction from evidential Bayesian networks. *Artificial Intelligence and Law*, 20(2):109–143, 2012.
[5] S.T. Timmer, J.-J. Ch. Meyer, H. Prakken, S. Renooij, and B. Verheij. Extracting legal arguments from forensic Bayesian networks. In *JURIX 2014: The Twenty-seventh Annual Conference*, volume 217, pages 71–80, 2014.
[6] G.A.W. Vreeswijk. Argumentation in Bayesian belief networks. In *Argumentation in Multi-Agent Systems*, volume 3366 of *Lecture Notes in Computer Science*, pages 111–129, 2005.
[7] F.J. Bex, H. Prakken, C.A. Reed, and D.N. Walton. Towards a formal account of reasoning about evidence: Argumentation schemes and generalisations. *Artificial Intelligence and Law*, 11(2/3):125–165, 2003.
[8] M. Grabmair, T.F. Gordon, and D. Walton. Probabilistic semantics for the Carneades argument model using Bayesian networks. In *Proceedings of COMMA*, pages 255–266, 2010.
[9] F.J. Bex. An integrated theory of causal stories and evidential arguments. In *Proceedings of the 15th International Conference on Artificial Intelligence and Law*, pages 13–22. ACM, 2015.
[10] U. Hahn and J. Hornikx. A normative framework for argument quality: argumentation schemes with a Bayesian foundation. *Synthese*, pages 1–41.
[11] B. Verheij. Arguments and their strength: Revisiting Pollock's anti-probabilistic starting points. In *Proceedings of COMMA 2014, Computational Models of Argument*, pages 433–444. IOS Press, 2014.
[12] H. Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1:93–124, 2010.
[13] F. V. Jensen and T. D. Nielsen. *Bayesian Networks and Decision Graphs*. Springer, Berlin, 2007.
[14] V. Crupi, K. Tentori, and M. Gonzalez. On Bayesian measures of evidential support: Theoretical and empirical issues. *Philosophy of Science*, 74(2):229–252, 2007.
[15] M.J. Druzdzel and L.C. van der Gaag. Elicitation of probabilities for belief networks: Combining qualitative and quantitative information. In *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, pages 141–148, 1995.
[16] J. Keppens. On modelling non-probabilistic uncertainty in the likelihood ratio approach to evidential reasoning. *Artificial Intelligence and Law*, 22(3):239–290, 2014.