Exploiting Causality in Constructing Bayesian Network Graphs from Legal Arguments

Remi WIETEN [a], Floris BEX [a,b], Henry PRAKKEN [a,c] and Silja RENOOIJ [a]

[a] *Information and Computing Sciences, Utrecht University, The Netherlands*
[b] *Institute for Law, Technology and Society, Tilburg University, The Netherlands*
[c] *Faculty of Law, University of Groningen, The Netherlands*

Abstract. In this paper, we propose a structured approach for transforming legal arguments to a Bayesian network (BN) graph. Our approach automatically constructs a fully specified BN graph by exploiting causality information present in legal arguments. Moreover, we demonstrate that causality information in addition provides for constraining some of the probabilities involved. We show that for undercutting attacks it is necessary to distinguish between causal and evidential attacked inferences, which extends on a previously proposed solution to modelling undercutting attacks in BNs. We illustrate our approach by applying it to part of an actual legal case, namely the Sacco and Vanzetti legal case.

Keywords. Bayesian networks, legal reasoning, argumentation, causality

1. Introduction

Bayesian networks (BNs) are probabilistic reasoning tools that are being applied in many complex domains where uncertainty plays a role, including forensics and law [1]. A BN consists of a graph, which captures the probabilistic independence relation among the modelled domain variables, and locally specified (conditional) probability distributions that collectively describe a joint probability distribution. BNs are well-suited for reasoning about the uncertain consequences that can be inferred from the evidence in a case. However, especially in data-poor domains, their construction needs to be done mostly manually, which is a difficult, time-consuming and error-prone task. Domain experts such as crime analysts and legal experts, therefore, typically resort to using qualitative reasoning tools, such as argument diagrams and mind maps [2] and Wigmore charts [1].

The benefits of BNs can be exploited if their construction can be facilitated by extracting information specified by experts using the tools and techniques they are familiar with. In previous research, Bex and Renooij [3] contributed to this idea by deriving constraints on a BN given structured arguments [4]. Their approach suffices for automatically constructing an undirected graph. However, for setting arc directions, as required for the directed BN graph, the authors resort to the standard approach used in BN construction: the BN modeller and the domain expert together specify the arc directions using the notion of causality as a guiding principle [5]. The resulting graph then has to be verified and refined in terms of the independence relation it represents.

From a legal perspective, it is, however, justified to assume that information regarding causality is present in the domain expert's original argument-based analysis [6, 7]. In this paper, we therefore propose to make causality information explicit in the argument-based analysis and to exploit this information in automatically constructing a completely directed BN graph. Moreover, we demonstrate that causality information in addition provides for constraining several conditional probability distributions. We illustrate our approach by applying it to the Sacco and Vanzetti legal case and compare our results to a previous BN modelling of the same case [1].

2. Preliminaries

2.1. Argumentation

Throughout this paper, we assume that the domain expert's analysis of a case is captured in an *argument graph*, in which claims are substantiated by chaining inferences from the observed case evidence; an example is depicted in Figure 1a. Argument graphs, which we define below, are closely related to Wigmore charts [1], mind maps and argument diagrams [2] with which many crime analysts and legal experts are familiar. We note that our formalism is an abstraction of existing argumentation formalisms (an overview is provided by [8]). Specifically, the evaluation of arguments is not taken into account, as it is not needed for our current purposes. By embedding our formalism in other existing formalisms, the dialectical status of arguments can be accounted for (cf. [9]).

Formally, an argument graph is a directed graph $AG = (\mathbf{P}, \mathbf{A})$, where $\mathbf{P}$ is a set of nodes representing propositions and $\mathbf{A}$ is a set of directed (hyper)arcs. Nodes $\mathbf{Ev} \subseteq \mathbf{P}$ corresponding to the (observed) *case evidence* are shaded root nodes in the argument graph. We assume that for every $p \in \mathbf{Ev}$, there is no node representing proposition p's negation $\neg p$ elsewhere in the graph. The set of directed (hyper)arcs $\mathbf{A}$ is comprised of three pairwise disjoint sets $\mathbf{S}$, $\mathbf{R}$ and $\mathbf{U}$, which are sets of support arcs, rebuttal arcs and undercutter arcs, respectively. A *support arc* is a solid (hyper)arc $s\colon p_1, \ldots, p_n \to p \in \mathbf{S}$, indicating an inference step from one or more propositions $p_1, \ldots, p_n \in \mathbf{P}$ (the *tails* of the arc) to a single proposition $p \in \mathbf{P}$ (the *head* of the arc). Two support arcs s_1 and s_2 form a *support chain* (s_1, s_2) in case the head of s_1 is a tail of s_2.

There are two types of attack arcs. A *rebuttal arc* $r \in \mathbf{R}$ is a bidirectional dashed arc in the argument graph that is drawn between two nodes in $\mathbf{P}$ iff these nodes represent a proposition p and its negation $\neg p$. An *undercutter arc* $u \in \mathbf{U}$ is a dashed hyperarc directed from a node $p \in \mathbf{P}$ (the undercutter of the inference) to a support arc $s \in \mathbf{S}$. Figure 3a depicts an example including undercutter arcs. Informally, a rebuttal is an attack on a proposition, while an undercutter attacks an inference by providing exceptional case circumstances under which the inference is not applicable.

In reasoning about evidence, a distinction can be made between causal and evidential inferences [6, 10]. Causal inferences are of the form '*c is a cause for e*' (e.g. fire causes smoke), whereas evidential inferences are of the form '*e is evidence for c*' (e.g. smoke is evidence for fire). We assume that every support arc $s \in \mathbf{S}$ is annotated with a causal 'c' or an evidential 'e' label; $\mathbf{S}$ then falls into two disjoint sets $\mathbf{S^C}$ and $\mathbf{S^E}$ of causal and evidential support arcs, respectively. We consider the tails of two causal arcs with the same head to be competing (but not necessarily mutually exclusive) causes for the common effect expressed by the head. Similarly, in case two evidential support arcs have distinct heads and a single, identical tail, we consider the heads to be competing causes

for the common effect expressed by the tail. In case a causal or evidential support arc has multiple tails, we assume that the tails are not in competition, as the arc expresses that only the tails together allow us to infer the head.

As noted by Pearl [10], the chaining of a causal inference and an evidential inference can lead to undesirable results. Consider the example in which a causal inference states that a smoke machine causes smoke and an evidential inference states that smoke is evidence for fire. If we now know that there is a smoke machine, then this is evidence that there is a fire, which is clearly undesirable. To this end, we assume that an argument graph does not contain a support chain (s_1, s_2) where $s_1 \in \mathbf{S^C}$, $s_2 \in \mathbf{S^E}$.

2.2. Bayesian Networks

A BN [5] is a compact representation of a joint probability distribution $\Pr(\mathbf{V})$ over a finite set of discrete random variables $\mathbf{V}$. The variables are represented as nodes in a directed acyclic graph $G = (\mathbf{V}, \mathbf{E})$, where $\mathbf{E} \subseteq \mathbf{V} \times \mathbf{V}$ is a set of directed arcs $V_i \to V_j$ from parent V_i to child V_j. Each node describes a number of mutually exclusive and exhaustive values; in this paper, we assume all nodes to be Boolean. The BN further includes, for each node, a conditional probability table (CPT) specifying the probabilities of the values of the node conditioned on the possible joint value combinations of its parents. A node is called instantiated iff it is set to a specific value. Given a set of instantiations, or evidence, the probability distributions over the other nodes in the network can be updated using probability calculus [5]. An example of a BN graph and one of its CPTs is depicted in Figures 1b and 1c, where ovals denote nodes and instantiated nodes are shaded.

The BN graph G captures the independence relation among its variables. Let a chain be defined as a sequence of distinct nodes and arcs in the BN graph. A node V is called a *head-to-head node* on a chain c if it has two incoming arcs on c. A chain c is *blocked* iff it includes a node V such that (1) V is an uninstantiated head-to-head node on c without instantiated descendants; or (2) V is instantiated and has at most one incoming arc on c. A chain is inactive if it is blocked; otherwise it is called *active*. If no active chains exist between V_1 and V_2 given instantiations of nodes in $\mathbf{Z}$, then they are considered conditionally independent given $\mathbf{Z}$. After constructing an initial BN graph, it should be verified that this graph is acyclic and that the graph correctly captures the (conditional) independencies. If the graph does not yet exhibit these properties, arcs should be reversed, added or removed by the BN modeller in consultation with the domain expert. We will refer to this step as the *'initial validation step'*.

In case the two parents of a head-to-head node are seen as causes of a common effect, then instantiation of the head-to-head node or one of its descendants will induce an active chain between the causes. If one of the causes is now observed, then the probability of the other cause being present as well can either increase, decrease or stay the same upon updating, depending on the probabilities in the CPT for the head-to-head node. In case the probability of the other cause decreases, this is called the *'explaining away'* effect [11]. To achieve the explaining away effect between two causes $V_1 = true$ (denoted v_1) and $V_3 = true$ (v_3) of $V_2 = true$ (v_2), the CPT for V_2 needs to be constrained such that V_1 and V_3 exhibit a *negative product synergy wrt* v_2: $\Pr(v_2 \mid v_1, v_3) \cdot \Pr(v_2 \mid \neg v_1, \neg v_3) \leq \Pr(v_2 \mid v_1, \neg v_3) \cdot \Pr(v_2 \mid \neg v_1, v_3)$. A *zero product synergy* is a special case of a negative product synergy, where the inequality in the above equation is replaced by an equality. In this case, no intercausal reasoning can occur between causes V_1 and V_3 of v_2.

3. Exploiting Causality in Constructing BN Graphs from Legal Arguments

In this section, we first present the steps of our structured approach for automatically constructing BN graphs from argument graphs, and then illustrate and explain the steps with several examples. Let $node\colon \mathbf{P} \to \mathbf{V}$ be an operator which maps every proposition p or $\neg p \in \mathbf{P}$ to a variable $node(p) = node(\neg p) \in \mathbf{V}$ that describes values p and $\neg p$. For a given argument graph $AG = (\mathbf{P}, \mathbf{A})$, a BN graph $G = (\mathbf{V}, \mathbf{E})$ is constructed as follows:

1. For every p or $\neg p \in \mathbf{P}$, the BN graph includes $node(p)$. For every $p \in \mathbf{Ev} \subseteq \mathbf{P}$, the set of observed nodes $\mathbf{Ev} \subseteq \mathbf{V}$ includes $node(p)$.
2. For every support arc $s\colon p_1, \ldots, p_n \to p \in \mathbf{S^C}$, $\mathbf{E}$ includes arcs $node(p_1) \to node(p), \ldots, node(p_n) \to node(p)$.
3. For every support arc $s\colon p_1, \ldots, p_n \to p \in \mathbf{S^E}$, $\mathbf{E}$ includes arcs $node(p) \to node(p_1), \ldots, node(p) \to node(p_n)$.
4. For every undercutter arc $u \in \mathbf{U}$ directed from a $p \in \mathbf{P}$ to support arc $s\colon q_1, \ldots, q_n \to q \in \mathbf{S^E}$, $\mathbf{E}$ includes arcs $node(p) \to node(q_1), \ldots, node(p) \to node(q_n)$.
5. For every undercutter arc $u \in \mathbf{U}$ directed from a $p \in \mathbf{P}$ to support arc $s\colon q_1, \ldots, q_n \to q \in \mathbf{S^C}$, $\mathbf{E}$ includes arc $node(p) \to node(q)$.
6. Verify the properties of the BN graph by performing the initial validation step.

We propose the following constraints on the CPTs of the BN under construction:

7a) For every pair of support arcs $s_1\colon p_1, \ldots, p_n \to p, s_2\colon q_1, \ldots, q_m \to p \in \mathbf{S^C}$, the CPT for $node(p)$ should be constrained such that $node(p_i)$ and $node(q_j)$ exhibit a negative product synergy wrt value p of $node(p)$ for $1 \leq i \leq n, 1 \leq j \leq m$.

7b) For every pair of support arcs $s_1\colon p \to p_1, s_2\colon p \to p_2 \in \mathbf{S^E}$, the CPT for $node(p)$ should be constrained such that $node(p_1)$ and $node(p_2)$ exhibit a negative product synergy wrt value p of $node(p)$.

7c) For every support arc $s\colon p_1, \ldots, p_n \to p \in \mathbf{S^C}$, the CPT for $node(p)$ should be constrained such that $node(p_i)$ and $node(p_j)$ exhibit a zero product synergy wrt value p of $node(p)$ for $1 \leq i \leq n, 1 \leq j \leq n, i \neq j$.

7d) For every undercutter arc $u \in \mathbf{U}$ directed from a $p \in \mathbf{P}$ to support arc $s\colon q_1, \ldots, q_n \to q \in \mathbf{S^E}$, the CPT for $node(q_i)$ should be constrained such that $node(p)$ and $node(q)$ exhibit a negative product synergy wrt value q_i of $node(q_i)$ for $1 \leq i \leq n$. For every pair of undercutter arcs $u_1, u_2 \in \mathbf{U}$ directed from $p_1 \in \mathbf{P}$ respectively $p_2 \in \mathbf{P}$ to s, the CPT for $node(q_i)$ should be further constrained such that $node(p_1)$ and $node(p_2)$ exhibit a negative product synergy wrt value q_i of $node(q_i)$ for $1 \leq i \leq n$.

In Section 3.1, we illustrate that steps 1-3 of our approach suffice for constructing BN graphs from argument graphs without undercutter arcs, where the CPTs of the BN under construction should be constrained according to steps 7a-7c. We then illustrate in Section 3.2 that the BN under construction needs to be further constrained in case undercutter arcs are present in the argument graph; this is accounted for in steps 4, 5 and 7d of our approach. Finally, we note that, while our approach exploits the domain knowledge captured in the argument graph, the argument graph may not contain all the information needed to resolve conflicts such as cycles and unwarranted (in)dependencies in the obtained BN graph; hence, step 6 of our approach.
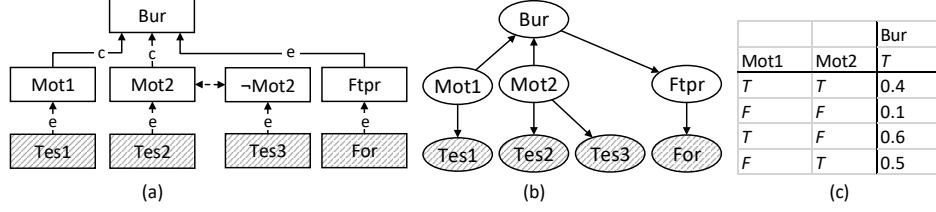
Figure 1. An argument graph involving two competing causes Mot1 and Mot2 for Bur (a); the corresponding BN graph constructed by our approach (b); a possible CPT for node Bur (c).

3.1. Argument Graphs without Undercutter Arcs

We first describe the main idea behind our approach. As an argument graph describes the inferences that can be made between a set of proposition nodes and a BN graph describes an independence relation over a set of variables, a transformation step from propositions to variables needs to occur upon constructing a BN graph; this is accounted for in step 1 of our approach. By the same step, two propositions involved in a rebuttal are captured as two mutually exclusive values of the same node. In steps 2 and 3, arcs in the BN graph are then directed using the *notion of causality*, that is, for every causal support arc $s \in \mathbf{S^C}$, arcs in the BN graph are directed from the nodes corresponding to the tails of s to the node corresponding to the head of s, and vice versa for an evidential support arc. This formalises the approach typically taken in the manual construction of BN graphs [5].

In comparing the knowledge captured in a BN graph to the knowledge captured in the original argument graph, it is important to note that a BN graph necessarily represents more than an argument graph, as for every proposition $p \in \mathbf{P}$ a variable is created which describes both values p and $\neg p$. Furthermore, without instantiating variables, a BN in itself is a joint probability distribution which does not model directionality; only when instantiating variables do reasoning patterns arise in the form of induced active chains. As the notion of an active chain is a symmetrical concept (an active chain exists between variables A and B given $\mathbf{Ev}$ iff an active chain exists between B and A given $\mathbf{Ev}$), a BN graph will also capture reasoning patterns in the opposite direction from the support arcs present in the argument graph. Our current focus will be on verifying whether the (chains of) support arcs in the argument graph are present in the BN graph in the form of active chains given the relevant case evidence. As the observed case evidence $\mathbf{Ev} \subseteq \mathbf{P}$ is indicated in the argument graph, we instantiate only the corresponding variables $\mathbf{Ev} \subseteq \mathbf{V}$. Here, we note that in determining the influence of a variable $E_i \in \mathbf{Ev}$, active chains given the variables $\mathbf{Ev} \setminus \{E_i\}$ need to be considered.

We illustrate our approach through the example depicted in Figure 1a. Suppose that a burglary has taken place and that we are interested in whether some suspect is in fact the burglar (Bur). Forensic analysis (For) shows there is a match between a pair of shoes owned by the suspect and footprints found near the crime scene, which provides us with evidence that the suspect left his footprints at the crime scene (Ftpr). According to witness testimony (Tes1), the suspect had a motive to commit this burglary, namely to commit theft (Mot1); however, according to another testimony (Tes2), the suspect had a different possible motive, namely to harm the owner (Mot2). These motives are expressed as competing causes for Bur in the argument graph. Finally, the suspect denied in his testimony (Tes3) that he had intentions to harm the owner (¬Mot2), which rebuts Mot2.

By following steps 1-3 of our approach, the BN graph of Figure 1b is constructed. For every support chain (s_1, s_2) in the argument graph, there exists an active chain be-

tween the nodes corresponding to the tails of s_1 and the head of s_2 in the BN graph given **Ev**. Specifically, given the case evidence **Ev** = {Tes1, Tes2, Tes3, For}, there exist active chains between Tes1 and Bur, Tes2 and Bur, Tes3 and Mot2, and For and Bur in the BN graph. Generally, we note that for any given argument graph, no head-to-head node is formed in the node corresponding to the head of s_1 for any support chain (s_1, s_2) in the argument graph by steps 1-3 of our approach. Since propositions in **Ev** are root nodes in an argument graph, corresponding instantiated nodes are root nodes in the BN graph; chains in the BN graph corresponding to support chains in the argument graph are, therefore, always active given **Ev**. Manual verification of whether support chains are captured in the form of active chains in the BN graph can, therefore, be skipped.

Besides structural constraints, information regarding causality in the argument graph also allows us to derive constraints on the CPTs of the BN. In the argument graph of Figure 1a, Mot1 and Mot2 are competing causes for the common effect Bur in that presence of one of the causes makes the other cause less likely. We propose to link this type of intercausal interaction in argument graphs to the explaining away effect in BNs. Specifically, as proposed in step 7a of our approach, the CPT for Bur should be constrained such that Mot1 and Mot2 exhibit a negative product synergy wrt value Bur = *true*. For example, the CPT for Bur can be chosen as in Figure 1c, as in this case $0.4 \cdot 0.1 \leq 0.6 \cdot 0.5$. Similarly, the heads of two evidential support arcs with the same tail are in competition for the effect expressed by the tail. Suppose that the two causal support arcs Mot1 $\rightarrow$ Bur and Mot2 $\rightarrow$ Bur in Figure 1a are replaced by two evidential support arcs Bur $\rightarrow$ Mot1 and Bur $\rightarrow$ Mot2. By following steps 1-3 of our approach, again the BN graph of Figure 1b is constructed. Competition between the causes can again be captured by constraining the CPT for Bur accordingly, as proposed in step 7b of our approach. We note that, in case two evidential support arcs have multiple (shared) tails, intercausal interactions will not in general be of some predetermined type that constrains the CPTs.

By following steps 2 and 3 of our approach, two competing causes automatically form a head-to-head connection in the node corresponding to the common effect for any given argument graph; interaction between two competing causes in an argument graph can, therefore, always be directly captured in the CPT for the node corresponding to the common effect. We note that the intended explaining away effect is only active in case the node corresponding to the common effect is instantiated or has an instantiated descendant. This is the case when in the corresponding argument graph the common effect is an element of **Ev** or in case there exists a chain of evidential support arcs for the common effect. For example, in Figure 1b the explaining away effect is active between Mot1 and Mot2, as Bur has instantiated descendant For; this is the case as there exists a chain of evidential support arcs from For to Bur in Figure 1a.

Figure 2a depicts an adjustment of the argument graph of Figure 1a, where the single causal support arc between Mot, Opp and Bur indicates that the presence of motive and opportunity together caused the suspect to commit the burglary. According to steps 1-3 of our approach, the BN graph of Figure 2b is constructed. Similar to the BN graph of Figure 1b, an active chain exists between Mot and Opp given **Ev**; however, as Mot and Opp are not in competition for Bur in this example, we need to assure that no intercausal reasoning can occur between Mot and Opp for value Bur = *true*. This can be achieved by constraining the CPT for Bur such that Mot and Opp exhibit a zero product synergy wrt value Bur = *true*, as proposed in step 7c of our approach. Note that the argument graph only informs us that there should be a zero product synergy between Mot and Opp
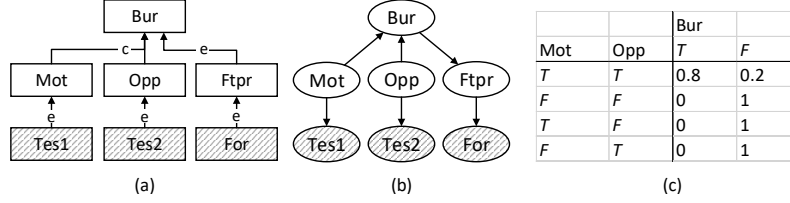
Figure 2. An argument graph involving two non-competing causes Mot and Opp for Bur (a); the corresponding BN graph constructed by our approach (b); a possible CPT for node Bur (c).

wrt value Bur = *true*; it does not inform us whether this should also hold wrt value Bur = *false*. Figure 2c depicts a possible CPT for Bur, where Mot and Opp exhibit a zero product synergy wrt value Bur = *true* as $0.8 \cdot 0 = 0 \cdot 0$. However, as $0.2 \cdot 1 \leq 1 \cdot 1$, Mot and Opp also exhibit a negative product synergy wrt value Bur = *false*. Care should be taken, therefore, in eliciting the involved probabilities.

3.2. Argument Graphs including Undercutter Arcs

Next, undercutting attacks are considered. Bex and Renooij [3] interpret undercutting attack as explaining away and propose constraints on the BN graph and CPTs accordingly. We demonstrate that their solution only has the desired effect in case the undercut support arc is evidential. The solution of Bex and Renooij is captured by steps 4 and 7d of our approach; step 5 then accounts for undercut causal support arcs.

In Figure 3a, an example of an argument graph is depicted in which both an evidential and a causal support arc are undercut. The evidential support arc Tes2 → Mot is undercut by proposition Lie, which states that person *x* who testified to Tes2 had reason to lie when giving his testimony. Since Tes2 is either the result of the true motive or due to a lie, Mot and Lie can be seen as competing causes of *x*'s testimony. Generally, undercutters of evidential support arcs can be considered competing causes for the common effects expressed by the tails of the support arc; for undercut evidential support arcs, we therefore follow the approach of [3]. The authors specify that the nodes corresponding to the propositions involved in an undercutting attack should form head-to-head nodes in the tails of the undercut support arc. By step 3 of our approach, the BN graph under construction includes arc Mot → Tes2. A head-to-head node can, therefore, be formed in node Tes2 by adding additional arc Lie → Tes2 to the BN graph; this is captured by step 4 of our approach. The explaining away effect can then be achieved by constraining the CPT for Tes2 such that Lie and Mot exhibit a negative product synergy wrt value Tes2 = *true*; this is captured by step 7d of our approach. For example, the CPT for Tes2 can be chosen as in Figure 3c, as in this case $0.2 \cdot 0.01 \leq 0.8 \cdot 0.3$.

We note that, in case there are multiple undercutters of an evidential support arc, the CPTs for the nodes corresponding to the tails need to be further constrained such that the explaining away effect occurs between the nodes corresponding to each pair of undercutters of the undercut support arc, as each undercutter expresses an additional competing cause; this is also captured by step 7d of our approach.

In the argument graph of Figure 3a, the causal support arc Mot → Bur is undercut by proposition ¬Opp. In contrast with the undercut evidential support arc, this undercutter cannot be considered a competing cause for the tail of the undercut support arc; the absence of opportunity cannot be considered a cause for motive. Instead, it can be considered a causal explanation for the fact that the suspect is not the burglar (¬Bur). For undercut causal support arcs, we therefore propose to form a head-to-head node in
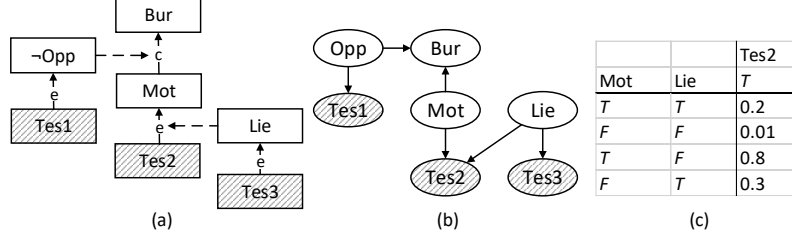
Figure 3. An argument graph involving both an undercutter of an evidential support arc and a causal support arc (a); the corresponding BN graph constructed by our approach (b); a possible CPT for node Tes2 (c).

the node corresponding to the head of the support arc as opposed to in the node corresponding to the tail of the support arc as proposed by [3]. By step 2 of our approach, the BN graph under construction includes arc Mot $\rightarrow$ Bur. A head-to-head node can, therefore, be formed in Bur by adding additional arc Opp $\rightarrow$ Bur to the BN graph; this is captured by step 5 of our approach. As $node$(Bur) describes both values Bur and ¬Bur, possible interactions, if any, between Mot and ¬Opp can be captured in the CPT for this node. We note that, in this case, intercausal interactions will not in general be of some predetermined type that constrains the CPT.

4. Case Study: The Sacco and Vanzetti Case

In this section, we apply our approach to part of an actual legal case, namely the well-known Sacco and Vanzetti case. The case concerns Sacco and Vanzetti, who were convicted for shooting and killing a payroll guard during a robbery. Kadane and Schum [1] performed a probabilistic analysis of this case by first constructing Wigmore charts, which are a type of (evidential) argument graph. The authors then manually constructed corresponding BNs by assessing the modelled independence relation and eliciting the necessary (conditional) probabilities. The authors did not provide a systematic approach for constructing BN graphs from Wigmore charts. Specifically, they constructed BN graphs by compiling and aggregating the evidence present in multiple Wigmore charts, where the choice which evidence to aggregate and compile was decided upon a case-by-case basis. In this section, we illustrate for one of their Wigmore charts that our approach provides a more structured way of constructing BN graphs.

In this case study, we only consider the evidence regarding Sacco's consciousness of guilt. During their arrest, Sacco and Vanzetti were armed. According to the two arresting officers, Connolly and Spear, Sacco and Vanzetti made suspicious hand movements, from which the prosecution concluded that Sacco and Vanzetti intended to draw their concealed weapons in order to escape their arrest. This suggests that Sacco and Vanzetti were conscious of having committed a criminal act.

In Figure 4, an argument graph concerning Sacco's consciousness of guilt corresponding to the Wigmore chart from Kadane and Schum [1, pp. 330–331] is depicted, along with the corresponding key list which indicates for every number in the argument graph to which evidence it corresponds. For reasons of space, the original Wigmore chart from Kadane and Schum is omitted. In formalising the Wigmore chart from Kadane and Schum, we have followed the approach of Bex and colleagues [7]. As noted by Kadane and Schum [1][p. 74–76], the natural direction of reasoning in their Wigmore charts is from the evidence to the ultimate conclusion, as all the evidence is known to them. All support arcs in the argument graph of Figure 4 are, therefore, evidential.
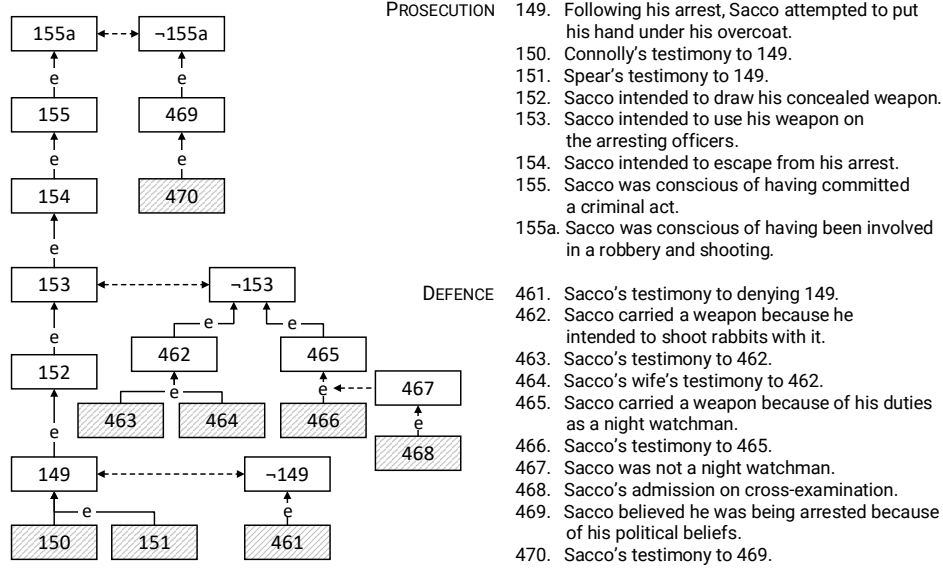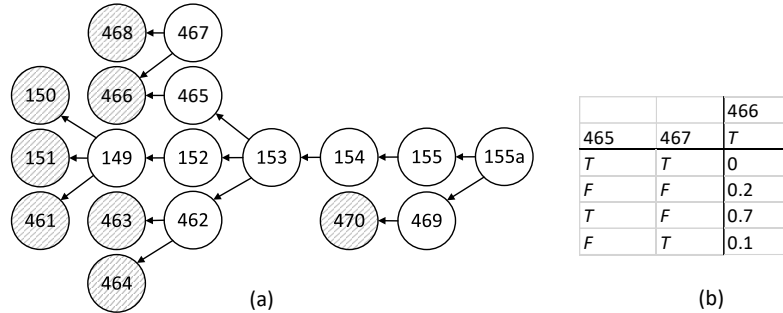
Figure 4. An argument graph corresponding to the Wigmore chart that Kadane and Schum constructed concerning Sacco's consciousness of guilt, along with the corresponding key list, adapted from [1, pp. 330–331].

By applying our construction approach, the BN graph of Figure 5a is obtained. Rebutting propositions are captured as two values of the same node by step 1 of our approach; arcs in the BN graph are then directed in the opposite direction of the evidential support arcs by step 3. Additional arc $467 \rightarrow 466$ is then added to the BN graph according to step 4. The obtained graph is largely identical to the BN graph that Kadane and Schum manually constructed for this part of the case [1, p. 232]. The only difference is that Kadane and Schum aggregated nodes 466, 467 and 468 into a single node; possible intercausal effects between 467 and 465 can, therefore, not be explicitly captured in their BN graph. While aggregation reduces the number of conditional probabilities to be elicited, we prefer to explicitly capture all elements of the argument graph in the corresponding BN graph to prevent loss of information. We note that, by step 7d of our approach, a constraint on the CPTs of the BN under construction is automatically obtained, which partially simplifies the elicitation procedure. By this step, nodes 465 and 467 should exhibit a negative product synergy wrt value $466 = true$. For example, the CPT for 466 can be chosen as in Figure 5b, as in this case $0 \cdot 0.2 \le 0.7 \cdot 0.1$.

5. Conclusion

In this paper, we have studied how legal arguments, including causality information, can be used to inform the construction of BN graphs. We have proposed a structured approach that allows domain experts to automatically construct a BN graph corresponding to their initial argument-based analysis which captures similar reasoning patterns as present in the original argument graph. We have illustrated that for undercutting attacks it is necessary to distinguish between causal and evidential attacked inferences. As a result, we extend on Bex and Renooij's [3] previously proposed solution to modelling undercutting attacks in BNs. In addition, we have identified when constraints on the CPTs of the BN are required to capture the desired effect of (induced) dependencies. In related research, Grabmair and colleagues [12] claim (without proof) that the Carneades argument model

Figure 5. The BN graph corresponding to the argument graph of Figure 4, as constructed by our approach (a); a possible CPT for node 466 (b).

can be given probabilistic semantics using BNs. However, the authors do not discuss issues related to forensics and causality.

In this paper, we have assumed that all support arcs are labelled with a causal or evidential label. In our future research, we intend to study the construction of BN graphs from partially labelled argument graphs. Furthermore, we have made no assumptions regarding inference strength; within our proposed constraints, the CPTs, therefore, need to be elicited manually. As arcs in the BN graph are directed from cause to effect, the necessary conditional probabilities can be elicited in the form of likelihood ratios, which are ratios of probabilities of observing the evidence under two mutually exclusive hypotheses. These can be directly entered in the CPT for each node, as is also done by e.g. Kadane and Schum [1]. In our future research, we will focus on deriving more probabilistic constraints, for example, by taking inference strength into account.

References

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

[2] P.A. Kirschner, S.J. Buckingham Shum, and C.S. Carr, eds., *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making.* Springer, 2003.

[3] F. Bex and S. Renooij. From arguments to constraints on a Bayesian network. In P. Baroni, T.F. Gordon, T. Scheffler, and M. Stede, eds., *Computational Models of Argument: Proceedings of COMMA 2016*, volume 287, pages 95–106. IOS press, 2016.

[4] H. Prakken. An abstract framework for argumentation with structured arguments. *Argument & Computation*, 1(2): 93–124, 2010.

[5] F.V. Jensen and T.D. Nielsen. *Bayesian Networks and Decision Graphs.* Springer, 2nd ed., 2007.

[6] F. Bex. An integrated theory of causal stories and evidential arguments. *Proceedings of the 15th International Conference on Artificial Intelligence and Law*, pages 13–22. ACM Press, 2015.

[7] F. Bex, H. Prakken, C.A. Reed, and D. Walton. Towards a formal account of reasoning about evidence: argumentation schemes and generalisations. *Artificial Intelligence and Law*, 11(2-3): 125–165, 2003.

[8] H. Prakken. Historical overview of formal argumentation. In P. Baroni, D. Gabbay, M. Giacomin, and L. van der Torre, eds., *Handbook of Formal Argumentation*, pages 73-141. College Publications, 2018.

[9] F. Bex, S. Modgil, H. Prakken, and C.A. Reed. On logical specifications of the argument interchange format. *Journal of Logic and Computation*, 23(5): 951–989, 2013.

[10] J. Pearl. Embracing causality in default reasoning. *Artificial Intelligence*, 35(2): 259–271, 1988.

[11] M.P. Wellman and M. Henrion. Explaining "explaining away". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(3): 287–292, 1993.

[12] M. Grabmair, T.F. Gordon, and D. Walton. Probabilistic semantics for the Carneades argument model using Bayesian networks. In P. Baroni, F. Cerutti, M. Giacomin, and G.R. Simari, eds., *Computational Models of Argument: Proceedings of COMMA 2010*, volume 216, pages 255–266. IOS Press, 2010.