A Structure-guided Approach to Capturing Bayesian Reasoning about Legal Evidence in Argumentation

Sjoerd T. Timmer

John-Jules Ch. Meyer

Henry Prakken

Silja Renooij

Bart Verheij

A Structure-guided Approach to Capturing Bayesian Reasoning about Legal Evidence in Argumentation

Sjoerd T. Timmer, John-Jules Ch. Meyer,
Henry Prakken, Silja Renooij & Bart Verheij

February 18, 2015

Abstract

Over the last decades the rise of forensic sciences has led to an increase in the availability of statistical evidence. Reasoning about statistics and probabilities in a forensic science setting can be a precarious exercise, especially so when independences between variables are involved. To facilitate the correct explanation of such evidence we investigate how argumentation models can help in the interpretation of statistical information. In this paper we focus on the connection between argumentation models and Bayesian belief networks, the latter being a common model to represent and reason with complex probabilistic information. We introduce the notion of a *support graph* as an intermediate structure between Bayesian networks and argumentation models. A support graph disentangles the complicating graphical properties of a Bayesian network and enhances its intuitive interpretation. Moreover, we show that this model can provide a suitable template for argumentative analysis. Especially in the context of legal reasoning, the correct treatment of statistical evidence is important.

1 Introduction

The currently ongoing debate about the role of probabilities in legal cases provokes an interest in legal argumentation theory about statistical evidence in court. The availability of forensic methods, such as DNA analysis and fingerprint matching, has increased over the last decades. Most of these methods yield results that are accompanied by graded uncertainties. These uncertainties are often expressed numerically by forensic experts, but lawyers, judges and other legal experts have notorious difficulty interpreting such results. Forensic experts usually have a scientific background where probabilistic reasoning is more common. As a result, they suffer less from common statistical reasoning fallacies than legal experts, who are trained in—and well-adjusted to—argumentative reasoning [8, 2, 5, 21]. The differences in background between legal and forensic experts creates a communication gap, which we believe can be bridged if we develop a better understanding of the relation between the two kinds of reasoning.

In this paper we introduce a novel approach to the explanation of Bayesian networks (BNs) and we apply this method to a legal setting. We introduce the notion of a *support graph*, which provides insight into the relations modelled in a Bayesian network. We show how these support graphs can be constructed, and how they can be used for the argumentative analysis of a case modelled by a Bayesian network. In particular, we show how the nodes of a support graph can be labelled to obtain a graph with a more argumentative interpretation. By adding numerical labels to the nodes as well, we can even maintain probabilistic information that was present in the BN model.

Bayesian Networks have been used as a tool to model forensic evidence, and even complete legal cases [9, 6, 14, 16, 15, 26, 27]. However, attempts to explain probabilistic inferences have usually focused on visual or textual explanations so far. See for instance the work of Lacave and Díez [13, 12], Koiter [11] and Druzdzel [3]. All of these methods attempt to explain what is modelled in a Bayesian network by visually displaying the different interactions between variables

or by verbally presenting the relations among variables. However, none of these methods investigate the structural information in the BN and neither do they construct structured argumentation.

Vreeswijk [28] has proposed a method to construct rules from BNs to form arguments but this approach only respects the independence properties of a BN under a number of limiting constraints on the design of the network under which, for example, no inter-causal interaction can be modelled. Suermondt [20] has introduced a framework in which paths of inference are identified in a BN graph. A method to formalise structured BN reasoning is *argument diagram extraction*, an approach introduced by Keppens [10]. An argument diagram is a graphical structure that informally represents Bayesian argumentation, but does not allow one to identify possible counter arguments. In argumentation diagrams, therefore, only one side of the story is highlighted.

Timmer et al. [22, 23] have studied a method to extract inference rules from Bayesian networks to construct arguments. One of the disadvantages of that method is the exponential explosion of the number of redundant arguments in terms of the size of the BN.

In this paper we present a novel approach that yields fewer redundant argumentative results. In general, the BN model enables a number of probabilistic inference rules. In argumentation, these rules are chained and combined to form arguments. The exponential number of ways to combine rules can lead to a combinatorial explosion of the number of arguments. In the current paper this issue is addressed by making a fundamental observation about the nature of aggregation in probabilistic reasoning. See Section 3.3 for details.

The rest of this paper is structured as follows. In Section 2 we discuss the relevant features of both probabilistic and argumentative legal reasoning. We informally introduce support graphs in Section 3 motivated by some fundamental observation on the nature of probabilistic reasoning. We formally define the approach in Section 4 where we also introduce an algorithm for constructing support graphs. We can use this model to introduce a number of argumentatively meaningful labelling of the support graph, which we show in Section 5. Finally we discuss the advantages and disadvantages of the new approach.

2 Background

2.1 Probabilistic Forensic Reasoning

Probabilistic models, such as Bayesian networks, are based on the notion of random variables. A random variable has a number of mutually exclusive and collectively exhaustive outcome states. In general, the number of outcomes per variable is not limited, but for simplicity we will assume throughout the rest of this paper that variables are binary.

Random variables can represent legal information. At the forensic source level variables merely represent forensic facts, such as whether a trace originates from a particular source or not. More generally, variables can also represent, for instance, the guilt or the motive of a suspect. Kadane and Schum [9] have modelled the famous Sacco and Vanzetti case as a Bayesian network containing variables about events as well as intentions and beliefs held by the suspects.

Random variables thus model parts of the world. Some of the outcomes of a variable may be actually observed, which influences the probabilities of other variables. Forensic results often report these correlations between variables. The best-known way to relate variables numerically is through the so called *likelihood ratio*. The likelihood ratio of evidence e pertaining to two competing hypotheses h_1 and h_2 is:

$$\text{LR} = \frac{P(e|h_1)}{P(e|h_2)}$$

After observing evidence e the likelihood ratio can be used to compute the posterior odds of the hypotheses from the prior odds of these hypotheses:

$$\frac{P(h_1|e)}{P(h_2|e)} = \frac{P(h_1)}{P(h_2)} * \text{LR}$$
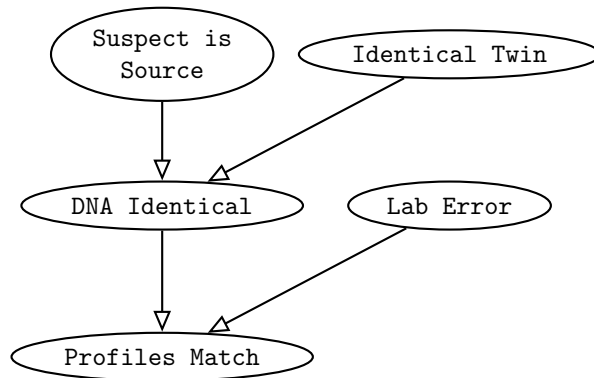
(a) The graphical structure of the BN.

Suspect is Source	
true	0.1
false	0.9

Identical Twin	
true	0.001
false	0.999

Lab Error	
true	0.001
false	0.999

DNA Identical

Suspect is Source	true		false	
Identical Twin	true	false	true	false
true	1	1	0.5	0
false	0	0	0.5	1

Profiles Match

DNA Identical	true		false	
Lab Error	true	false	true	false
true	0.5	0.999	0.5	0.002
false	0.5	0.001	0.5	0.998

(b) The conditional probability tables. These fictitious numbers serve to illustrate the methodology and are not based on real statistics.

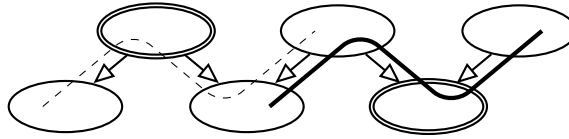Figure 1: Example of a Bayesian network.

Figure 2: An example of an active (bold) and an inactive (dashed) trail in a BN. Double lined nodes represent observed variables.

A well-known method to represent complex probabilistic information is the Bayesian network (BN) model. A BN is a graphical model that represents a probability distribution but also graphically provides some information on the independence relation between the variables. Every random variable is associated with a node in an acyclic directed graph. Figure 1 displays an example of a Bayesian network. This particular network models how a matching DNA profile correlates with the probability that a particular suspect is the source of a sample taken from the crime scene. The identity of the suspect is modelled by the 'Suspect is Source' variable at the top. The possibility of the presence of an unknown identical twin is explicitly modelled at the top right. The network can be interpreted causally to the extent that the suspect being the source 'causes' the DNA to match, which in turn 'causes' the profiles to match. The latter can also be influenced by the presence of lab errors as modelled by the 'Lab Error' variable.

Besides the graph, a Bayesian network specifies for every node the probability of the outcomes of the associated variable given the values of the parents of the corresponding node in the graph (see Figure 1b). From these numbers any (conditional) probability or likelihood ratio concerning these variables can be calculated. Bayesian networks exploit independence between variables to minimise the number of parameters that are needed. Precisely this graphical representation of independence is what makes the BN model very powerful but also sometimes hard to interpret. Two variables in the graph are *independent* when there is no *active* trail (undirected path) between them. A trail becomes inactive when a node on the trail blocks it. A head-to-head node (a node with two incoming edges on a particular trail, such as the bottom rightmost node in Figure 2) blocks the trail when none of its descendants (including the node itself) are observed. Other nodes do not block the trail unless they are themselves observed. Throughout this paper observed/instantiated variables will be displayed with a double outline in our figures.

To complicate things further, the arrows in the BN have no significance beyond their use to describe independence as just explained. A clear intuition of what an arrow represents and specifically what the direction of an arrow signifies in the context of the application is missing. A BN can be used to model causal relations, but it is not necessary to model the direction of the causality by the direction of the edges. Often many different graphical models can represent exactly the same independence relation.

Head-to-head nodes can capture inter-causal interactions, which are important to take into account in reasoning under uncertainty [18]. To illustrate this, suppose we consider a lab error and identical DNA as two causes of a profile match. This corresponds to the bottom three nodes in Figure 1a. The two causes are independent except for the fact that they share a common effect. Assume for now that this shared effect has been observed. Either of the two causes (the lab error and the actual match) could explain the presence of the effect (the observed match). Therefore, the observation of the effect increases our belief in either of the two causes. If a lab error is then observed as well, the profile match has been explained and the belief in identical DNA drops. What happens is that one cause *explains away* the other.

2.2 Argumentative Legal Reasoning

Argumentative models of legal proof can be traced back to the graphical representations by Wigmore [29] in the beginning of the 20th century. Since then, many formalisms of argumentative reasoning have been introduced.

Formal research on argumentation is inspired by work on logic [24]. Central to the argumentation
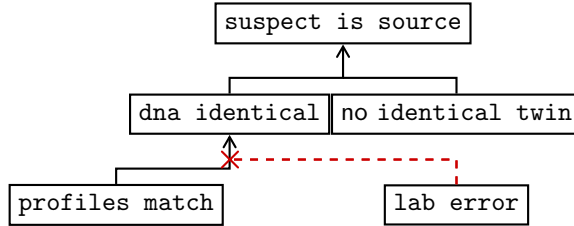
Figure 3: A visual representation of some arguments about the source of a DNA trace.

approach is the use of inference rules. A rule relates antecedents (premises) with consequents (conclusions). In classical logic it is the case that if all premises are true and the inference rule is sound, then the conclusion must be true. However, problems with strict rules have been identified and defeasible reasoning [19] has become an important aspect in many approaches to argumentation. Defeasibility means that rules, in contrast to deductive logic, can have exceptions. This means that the truth of the antecedents does not guarantee the truth of the consequent. Instead, a rule of the form $a \Rightarrow b$ has an interpretation that reads something like a 'usually implies' b. For example, let us imagine a defeasible rule 'Profiles Match' $\Rightarrow$ 'DNA Identical'. It states that if the forensic expert reports a profile match, then usually the DNA from the suspect is the same as that from the person leaving the trace. This constitutes a defeasible inference rule: if no exceptional circumstances occur we must conclude that the suspect and the source share the same DNA. One of the exceptional circumstances under which the rules loses its evidential value could arise when a mistake was made in the lab. If errors were made during the scientific analysis, then the rule does not apply any more. In that case the conclusion—that the suspect and the perpetrator are genetically equal—may or may not hold, but this can no longer be inferred from the reported match.

Defeasibility is a powerful tool for reasoning under uncertainty. A question that arises is how conflicts between defeasible arguments have to be resolved. Dung [4] has proposed an influential framework to model argument defeat and how acceptability of arguments follows from a defeat relation according to a number of different semantics. This work is entirely on an abstract level (without looking at the structure of rules inside arguments). For this internal structure many different formalisms have been proposed. See for instance ASPIC+ [17], Carneades [7] or DefLog [25]. A common feature of most approaches to argumentation is that arguments form a tree- or graph-like structure in which nodes are statements and links represent rules. A situation like that is often graphically represented as in Figure 3. In this example the final conclusion ('suspect is source') is shown together with one intermediate conclusion ('dna identical') and two premises ('profiles match' and 'no identical twin'). The process of identification is split in two separate phases. First, from the profile match it is inferred that the suspect has the same DNA as the person leaving the trace and, secondly, from this it is argued that the persons must be the same individual. This corresponds to two defeasible rules:

- 'profiles match' $\Rightarrow$ 'dna identical'

- 'dna identical' *and* 'no identical twin' $\Rightarrow$ 'suspect is source'

For the first rule, an *undercutting* [19] counter argument has been shown; in case of a lab error the rule does not apply.

3 Introducing support graphs

Before we can define our approach to modelling BN information in an argumentative setting we make a number of fundamental observations and informally show how support graphs can be used to explain what is modelled by the structure of a BN. We start with a discussion on the nature of inference and attack and argue that they are very similar in a probabilistic setting. To capture

both concepts in BNs we introduce the notion of *support*. Based on this notion we informally introduce the support graph framework.

3.1 Inference, attack and support

We observe that there is a remarkable similarity between inference in the argumentative sense and the probabilistic sense. Measures of probabilistic strength have therefore been investigated and compared [1]. The likelihood ratio, for instance, can be used as a measure of inferential strength for a defeasible inference rule.

$$M(e_1, \ldots, e_n \Rightarrow h) = \frac{P(e_1, \ldots, e_n | h)}{P(e_1, \ldots, e_n | \neg h)}$$

If this measures is greater than one,we know that the premises e_i have a positive effect on the conclusion. Indeed, if this likelihood ratio is bigger than one, observing e_1 through e_n will increase the odds of h.

If we were to add an observation to the list of premises of a rule in such a way that the likelihood ratio tips over and drop below one, the collective set of evidence has a negative effect on the conclusion. This is in a way similar to attack in argumentation. We observe, therefore, that the likelihood ration can model both inference and attack. Indeed, for binary variables, decreasing the probability of one outcome inherently increases the probability of the other outcome proportionally. This motivates our choice to abstract from outcomes, inference and attack and instead build an argumentative tool that focusses on the structural aspects of probabilistic reasoning.

We say that a variable *supports* another variable if observing the former changes the probability distribution of the latter. We emphasise that this is not a positive or negative notion, because, by the mutually exclusive nature of variable outcomes, any influence is positive for at least one of the outcomes and negative for another.

If we are to use probabilistic support to construct arguments, we need to consider how to efficiently represent the support between variables. A Bayesian network is a computationally very efficient representation of a probability distribution, but it lacks a clear and intuitive way to model argumentative support.

We could extract from the BN rules with premises and conclusions by enumerating subsets of all possible variable instantiations, but this is not computationally feasible if we do not take guidance from the graphical structure present in the Bayesian network to identify interesting rules. One of the problems is that a probability distribution models complex (in the sense that the strength of an inference often depends on other variables) interactions between variables and that we therefore have to choose the premises from subsets of the variables if we want to account for these interactions. Enumerating subsets is an exponential process, which makes this a computationally complex approach. This justifies our search for a method that uses the structural properties of the BN to build a template for the argumentation before actually considering rules.

The approach we take is to abstract away from the outcomes of variables and instead represent the structural aspects of how a particular conclusion can be supported. We can do so because these outcomes influence the strengths of the inference but not the structure of the arguments that we build.

3.2 Support factors

The question that naturally arises is which variables to consider when finding support for a particular conclusion variable. As we have just argued, we are not so much interested in every variable that could change the outcome of a node since many of these influences are indirect and only act through intermediate variables. A variable is directly influenced by its neighbours in the BN graph, but often also indirectly by other variables connected to those neighbours. What we aim to do is minimise the number of variables that we have to consider as possible premises by only looking at the variables that 'directly' influence the variable of interest. To define this more meaningful subset of variables to consider, we observe that evidence in BNs propagates through
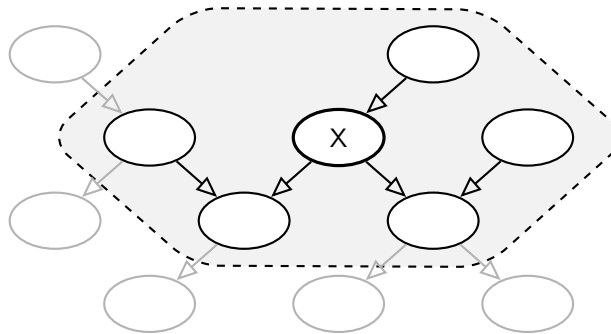
Figure 4: The Markov blanket of X shown in a dashed outline.

active trails. We can therefore limit the set of support factors by looking at the last two steps of those trails. We distinguish two different types of interactions. The first is a direct correlation with a neighbour in the BN graph. This is what is used when we reason about a parent and a child in the BN graph either in the direction of the edge or against it. The second form of interaction originates from the nature of the head-to-head connection in BNs. When a variable has more than one incoming edge, it creates an inter-causal interaction between the parents. This means that if a variable X has a child C with another parent P, then this other parent can explain certain outcomes of X away or in. This interaction is not captured if we only look at pairwise correlations between neighbouring variable pairs P–C and C–X. We will therefore define the set of support factors as the set of parents, children and parents of children of a node. These nodes together form the so-called Markov blanket, which has the well-known property that a node is independent of any other node in the graph given its Markov blanket. This means that, for our purposes, this set is exactly the minimal set of directly influencing variables that we want to include as possible premises for a rule. See Figure 4 for an example.

Let us now look back at our example BN. The node of interest in this example is the 'Suspect is Source' node, because that is what we hope to prove. Given that we look at the Markov blanket we can find the supporters 'DNA Identical' and 'Identical Twin'. We can visualise this graphically as in Figure 5a. The first supporter—'Identical Twin'—supports 'Suspect is Source' by an inter-causal interaction; it is a parent of a child in the graph, but it is also intuitive to see that the presence of a twin—explaining away the evidence that the DNA is identical—decreases the probability of the suspect being the source. The second supporter—'DNA Identical'—supports the conclusion via a direct correlation.

We can proceed by extending the support graph one step further. The 'DNA Identical' variable has a Markov blanket that contains all other nodes in the BN graph, so we could support it with any other variable. However, we argue that only the 'Lab Error' and 'Profiles Match' are appropriate supporters in this case as shown in Figure ??. We will formalise this later, but the intuition is as follows. We do not want the 'Suspect is Source' variable to support something that itself supports that very variable. A construction like that models circular reasoning. If we were to allow further support of 'DNA Identical' by 'Identical Twin', we would create a support path that follows the head-to-head connection ('Suspect is Source'-'DNA Identical'-'Identical Twin') in the BN graph. Argumentatively, this path would model the following argumentation (depending on how outcomes are correlated): there is an identical twin, therefore the DNA is more likely to be identical, therefore the suspect is more likely to be the source. We see that we make two consecutive inferences that are probabilistically valid on their own, but the combination (that the suspect is more likely to be the source because he has an identical twin) is not valid. In this example, the opposite would be more likely and indeed we have modelled that already by allowing the 'Identical Twin' to support the 'Suspect is Source' variable directly. In the next section we will show a formalisation that enables us to filter exactly these two types of undesirable supporters from the graph.
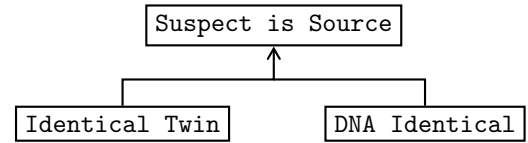
(a) First step in the support graph construction.
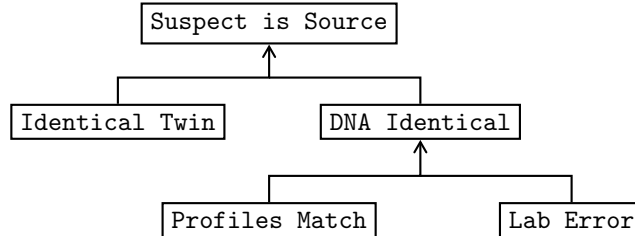
(b) Support graph with two levels of support.

Figure 5: Examples of support graphs for the BN in Figure 1.

3.3 Compulsory aggregation

We have seen that, in argumentation theory, arguments are built by combining rules into inference trees and that those rules can be identified by a positive strength measure in a probabilistic setting. Suppose we are interested in a particular conclusion variable. We have argued that we should look at the Markov blanket as the possible supporters for a conclusion about this variable. As we have also noted, the supporters can interact with each other. A particular outcome of a supporter can strengthen or weaken the effect of some other supporter, or even completely eliminate it. We can imagine that an exhaustive enumeration of all subsets of possible factors yields a combinatorial explosion, not just for the number of probabilistic rules to consider but also for the number of arguments that can later be constructed from those rules.

In the legal setting, but arguably also in a wider perspective, this problem can be addressed by observing that we should include as much of the available evidence as possible in the premises. We need to consider the case as a whole, which means that we cannot eliminate any of the factors for which there is some evidence. One cannot afford not to take any weakening factor into account for which there is even the smallest bit of evidence, since the adversarial party is surely going to hold it against you in court. On the other hand, you will want to make an argument as strong as possible by including anything that contributes to the belief in the conclusion, so you do not want to omit any positively contributing factors either.

Recalling the DNA matching example, what good would an argument be that ignores evidence about an identical twin if such evidence is available? If you have some evidence, for instance from a population database, that the suspect has no identical twins, then you make an unnecessarily weak argument by leaving that fact out. If, on the other hand, you know that the suspect has an identical twin, you have to check whether the DNA match still supports the hypothesis that the suspect is the source and to what degree. What this means is that we should include in the support a maximal subset of support factors for which we can find evidence.

4 Support graphs: formalisation

Until now we have only informally discussed how to build support graphs. We will now detail how support graphs can be constructed by introducing a formal definition and an algorithmic approach to the construction of such a graph.
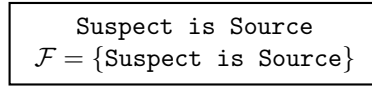
(a) The first step in support graph generation is to create a node for the variable of interest.
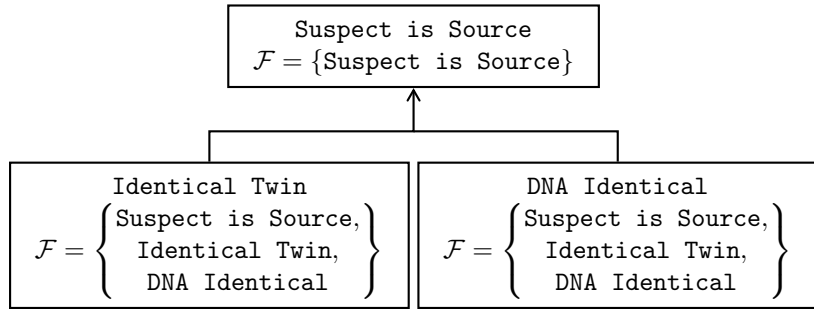
(b) Expanding the only support node in Figure 6a yields two possible supporters.
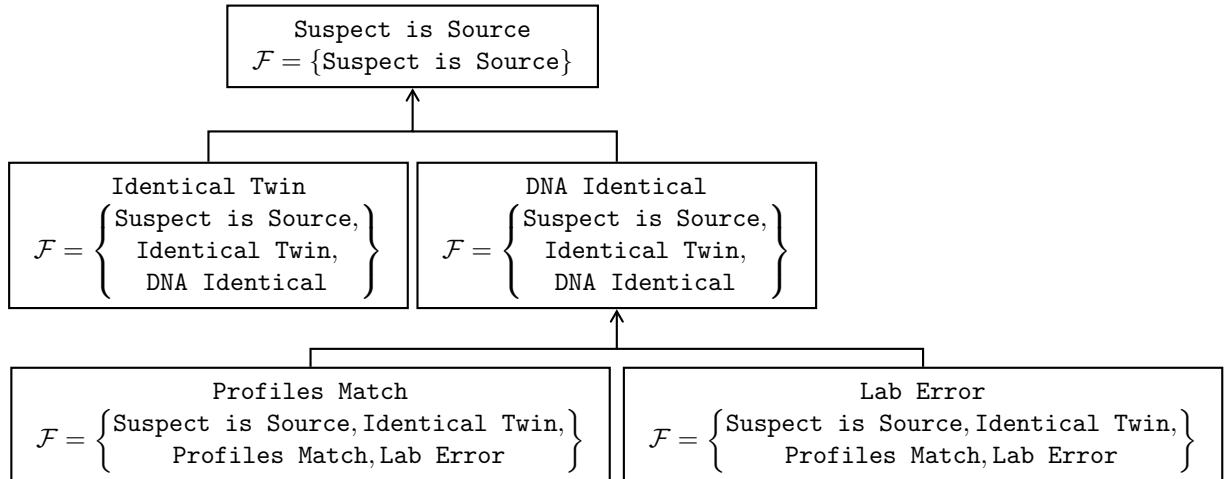
(c) The final support graph.

Figure 6: The step-by-step construction of a support graph for the BN in Figure 1.

Figure 7: The support factors (shown as the sources of the double arrows) of X provide exactly everything that can potentially support X either through direct inferences (B, F and G) or inter-causal interactions (C and D).

4.1   Support factors

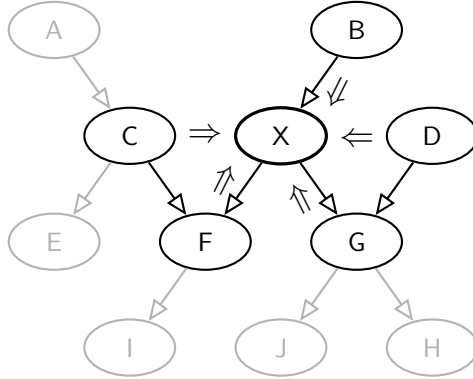In order to build a support graph, we will need to define how a variable can be supported. We define for every node X in the BN a set of *support factors*, which contains the variables that potentially support X. Assignments to these variables can later act as premises of possible rules for conclusions pertaining to X. As we have already informally argued, nodes from the Markov blanket of X completely determine the outcome of X and are therefore together the minimal set of supporters that we can consider without omitting any possibly influential variables:

Definition 1 (Support factors) *The set of support factors of a node X in a BN equals the Markov blanket of that node.*

For any node X we will consider nodes from this set when we want to find support for it. See Figure 7 for an example of this. Based on a set of observations we can later identify which of these potential supporters actually provide reasons for X. For instance, G in Figure 7 is a support factor of X, but if none of $\{G, H, J\}$ is observed it will not provide any actual support. The fact that it is a support factor merely indicates that it has the potential to influence X for some set of evidence.

4.2   Support graph construction

We now detail how the support graph follows from the support factors. As we have argued in Section 3, we have to take some special precautions to preclude circular argumentation and support via head-to-head connections since the former is a logical fallacy and the latter a probabilistic violation.

A support graph for a BN with variables $\mathbf{V}$ is a directed graph in which every node from $\mathbf{V}$ can occur multiple times. Edges in such a graph represent possible support. The supporters of a node will be aggregated as much as possible given the available evidence later. The construction of such a graph starts by creating a node for the conclusion of interest and recursively adding sub-supporters. In this way we build a tree of potential supporters. For brevity, we will use the name of the variable to denote the support graph node when there is no risk of confusion. Sub-supporters are taken from the set of support factors, which we defined to be the Markov blanket of the node for which we are adding support. However, we prevent chaining of inferences via head-to-head connections or circular reasoning by excluding some of the support factors. This can be achieved by maintaining (for every recursive step in the support graph) a *forbidden set* ($\mathcal{F}$) of variables that can no longer be used in further support because of these constraints.

We build the support graph recursively by adding potential supporters for X from the Markov blanket of X minus the set of forbidden nodes $\mathcal{F}$ in the node that we are finding support for. The forbidden set is inherited in every support step from the node that is being supported. In addition some variables are added to $\mathcal{F}$ depending on the type of inference. This means that the set can only grow along a branch in the support graph.

By adding to every node's forbidden set, the supported variable itself we effectively forbid a variable to support itself, eliminating the possibility of cyclic reasoning. The other variables that we add to the forbidden set depend on the connection between the supporting variable and the supported variable in the BN as defined in Definition 2.

Definition 2 (Forbidden set) *The forbidden set $\mathcal{F}$ of a support graph nodes for a BN node X (supporting itself a parent for Y with a forbidden set $\mathcal{F}'$) is:*

$$\mathcal{F} = \begin{cases} \{X\} & \textit{for the root of the support graph.} \\ \mathcal{F}' \cup \{X\} & \textit{otherwise, if X is a parent of Y in the BN.} \\ \mathcal{F}' \cup \{X\} \cup Par(X) & \textit{otherwise, if X is a child of Y in the BN.} \\ \mathcal{F}' \cup \{X\} \cup \mathbf{C} & \textit{otherwise, where $\mathbf{C}$ are the common children of X and Y.} \end{cases}$$

The third case shows that when we support a node with a rule against the direction of a BN edge we exclude any future inference along the direction of an edge. This effectively removes the possibility to argue along and then against BN edges in that order and therefore prevents exactly the probabilistic violation that we discussed earlier. Note that we do not make such an exclusion in the second case, and indeed it is probabilistically speaking not a problem to make an inference against the direction of an edge followed by an inference along the direction of an edge.

This exclusion mechanism is best explained by demonstrating it on the example BN from Figure 1. The initial support graph contains just one node associated with the variable of interest as shown in Figure 6a. Since the forbidden set contains only the variable itself, the whole Markov blanket is used to construct supporters. The first supporter (left in Figure 6b) is for the variable 'Identical Twin'. The forbidden set is constructed using the last case from Definition 2 and we add (in addition to the forbidden set of the root node) the new support node ('Identical Twin'), and the common child ('DNA Identical'). For the other supporter—'DNA Identical'—we obtain exactly the same forbidden set, but here we use the third case of the definition. We support a parent in the BN graph with a child. To prevent rule chaining through head-to-head connections we add other parents of the child ('Identical Twin' in this case) to the forbidden set.

We can proceed by expanding the support graph even further. The 'Identical Twin' variable has a Markov blanket that contains the 'Profiles Match' and the 'Suspect is Source' variables. Both of these are already in the forbidden set and therefore we cannot add further supporters. The 'DNA Identical' variable, on the other hand, has a Markov blanket that contains all other nodes in the BN graph. Since 'Identical Twin', 'Suspect is Source', and 'DNA Identical' itself are already in the forbidden set, we add support from 'Lab Error' and 'Profiles Match' (see Figure 6c). This is also the final support graph since none of the newly added supporters can be supported itself.

We can implement this construction as a recursive expansion of the support graph as presented in Algorithm 1. To define the expand operation we first need to introduce a few additional notational conventions. Let $Par(X)$ be the set of all parents of nodes X and $Chd(X)$ the set of all children of nodes X in the BN. Let $Var(S)$ denote the BN variable associated with the support graph node S. As a notational shortcut we will use $Chd(S)$ and $Par(S)$ (for a node S in the support graph) to denote children and parents in the BN of the variable associated with the support node.

```
function main():
    Data: BN
    Data: V_i, the variable of interest
    root = new support graph node for V_i; F(root) = {V_i}
    expand(root)
function expand(S):
    Data: S, the support graph node to expand
    foreach child in Chd(S) except F(S) do
        S' = new support graph node for child
        F(S') = F(S) ∪ {child, Par(child)}
        expand(S')
        add S' to supporters of S
    end
    foreach parent in Par(S) except F(S) do
        S' = new support graph node for parent
        F(S') = F(S) ∪ {parent}
        expand(S')
        add S' to supporters of S
    end
    foreach child in Chd(S) except F(S) do
        foreach parent in Par(child) except F(S) and Par(S) and Chd(S) do
            S' = new support graph node for parent
            F(S') = F(S) ∪ {parent, child}
            expand(S')
            add S' to supporters of S
        end
    end
```

Algorithm 1: Recursive algorithm to construct a support graph while maintaining a forbidden set $\mathcal{F}$ to exclude chaining of head-to-head connections and cyclic arguments.

4.3 Merging identical branches

Under some conditions sub-trees can be merged such that the tree resulting from Algorithm 1 becomes a graph. This operation makes the graph more intuitive to interpret because some duplicate reasoning steps are presented only once. In our running example every node occurs exactly once in the support graph, but this is a side effect of having a singly connected BN graph. In general a variable can occur more than once in the support graph. This happens when there are different ways in which a variable can support the variable of interest. In particular this means that there must be multiple trails in the BN. In that case it can happen that during the expansion of the support graph a node for the same variable is constructed twice with the same set of forbidden nodes. In such cases the sub-trees underneath those support graph nodes will be exactly identical. We can then merge these identical branches and gain a more insightful support graph.

5 Support graph labellings

The support graph provides insight in the allowed reasoning steps in a probabilistic model. An added benefit lies in its capability to function as an argumentation template. We will now describe how we can label the support graph with outcomes such that it becomes a more argumentative graph. Since the support graph model is strongly based on the BN model we can also incorporate the numerical aspects of the BN in those argumentation graphs using probabilistic measures of strength. There are multiple measures that can be used for this purpose, and we show that the support graph formalisms allows us to compare results while providing some argumentative
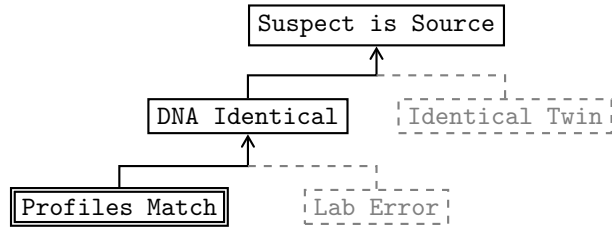
```
              ┌─────────────────────┐
              │   Suspect is Source │
              └─────────────────────┘
                    ↑        ┆- - - - - - ┆
          ┌─────────────────┐  ┆ Identical Twin ┆
          │   DNA Identical │  └- - - - - - - ┘
          └─────────────────┘  ┆- - - - - ┆
      ┌════════════════┐  ┆ Lab Error ┆
      ║  Profiles Match ║  └- - - - - ┘
      └════════════════┘
```

Figure 8: Observational pruning of the support graph. This is the support graph of the running example pruned to evidence for 'Profiles Match'. Non-observationally supported nodes are shown in dashed, grey outlines.

```
              ┌─────────────────────┐
              │   Suspect is Source │
              └─────────────────────┘
                    ↑        ┆- - - - - - ┆
          ┌═════════════════┐  ┆ Identical Twin ┆
          ║   DNA Identical ║  └- - - - - - - ┘
          └═════════════════┘
              ┆- - - - - - - - - - ┆
      ┌════════════════┐  ┆ Lab Error ┆
      ║  Profiles Match ║  └- - - - - ┘
      └════════════════┘
```
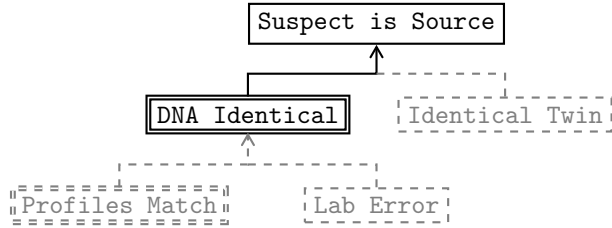
Figure 9: Further pruning of the support graph given a direct observation of 'DNA Identical'.

intuitions on the interpretation of these measures.

5.1 Observational support

The support graphs so far are generic models in the sense that they do not take into account a set of observed nodes. They can be instantiated with any set of observations. As a side effect, observations can be added in an interactive manner without having to recompute the underlying support graph. As we have discussed in Section 3 we want to use compulsory aggregation to enforce that every support factor that can be derived, is present in the premises for a conclusion node. We can do this by labelling the observed nodes as *observationally supported* and propagating this information upwards in the tree. Suppose that in our example the node 'Profiles Match' is observed. We can then label the nodes 'Profiles Match', 'DNA Identical' and 'Suspect is Source' as observationally supported. This is shown in Figure 8 by greying out the other nodes. We essentially retain the argument that infers from a profile match that the DNA must be equal and from there that the suspect must be the source. This is the maximally supported argument that we can make. The two branches that were not observed can be retained as 'anchor' places for future strengthening or weakening of the argument. In this way we can present arguments with possible counter arguments. These counter arguments have an inherently different status than the arguments in the observationally supported part of the support graph. Upon actually entering new evidence, the old argument is invalidated and superseded by a new argument that includes the new evidence.

There is another way in which branches can become irrelevant given some evidence. Consider a different set of observations containing only the 'DNA Identical' variable itself (ignoring for the sake of argument that in this example it is unlikely that this is directly observable). An observation is, in a sense, the ultimate kind of support we can find for a node, and therefore further support is not required. Even though there may be evidence lower in the sub-tree, we can safely ignore the whole branch under the observed variable. This is shown in Figure 9.

5.2 Adding instantiations

So far, we have added observed variables to disregard or highlight certain parts of the support graph, but we have not yet discussed how we can add assignments to variables to form instantiated

arguments. We recall that arguments consist of a number of combined rule applications, and rules consist of premises and conclusions,which are variable assignments. If we can assign outcomes to the variables in a support graph, the graph gains an argumentative interpretation in the sense that it then connects statements by inferences. The evidentially supported sub-graph forms an argument graph very similar to arguments in, for instance, ASPIC+.

For the observed variables it should be clear that the observed outcome is also the outcome that we assign in the support graph. For other nodes we propose a method based on probabilistic measures of inferential strength and we use two such measures to show that the support graph formalism can be used to compare different approaches to probabilistic reasoning. The first measure is inspired by the aforementioned likelihood ratio $(P(e|h)/P(\overline{e}|h))$ approach and emphasises the incremental effect of evidence. The second measure takes the posterior odds $(P(h|e)/P(\overline{h}|e))$ as the leading numerical valuation of a variable.

To calculate a probabilistic measure of inferential strength, a choice has to be made regarding the inclusion of nodes in the set of premises e. We choose to include the *observed leaves* under the conclusion h in the support graph because we want to capture the effect of these leaves on the conclusion.

Definition 3 (Numerical support) *We assign to every non-observed node S in the support graph an outcome s_i and a numerical support $M(s_i, \mathbf{E}_S)$ such that this measure $M()$ is maximal. Here, $\mathbf{E}_S$ denotes the set of observed outcomes of all observed leaves under S in the support graph and $\overline{s_i}$ is the other outcome of S. If the measure of both outcomes are equal, the outcome is not determined.*

To illustrate this, we apply two different measures of strength but other measures are imaginable and have been compared before [1]. We now show how such measures can be used to generate support graph labellings. Let us first consider the likelihood ratio approach, where the measure of strength equals:

$$M_{LR}(s_i, \mathbf{E}_S) = \frac{P(\mathbf{E}_S|s_i)}{P(\mathbf{E}_S|\overline{s_i})}$$

When we apply this labelling to the running example with the evidence that 'Profiles Match' is 'true' we obtain the argument as in Figure 10a. Since the likelihood ratio expresses a change in odds, the reported outcomes do not correspond to the *best* conclusion. These labels merely indicate a change towards those outcomes.

What we can see in this example is that the rule from the profile match to a DNA match is quite strong ($LR = 400$), and that the uncertainty about the existence of a twin slightly reduces the strength of this conclusion, but that, all evidence considered, there is a high likelihood ratio in favour of the hypothesis that the suspect is the source of the sample.

If we further investigate the case and find that the suspect has an identical twin we can update the labelling of the support graph as shown in Figure 10b. What happens is that the evidence still supports the conclusion, but the strength has reduced a lot. The evidence still points in the direction of the same hypothesis.
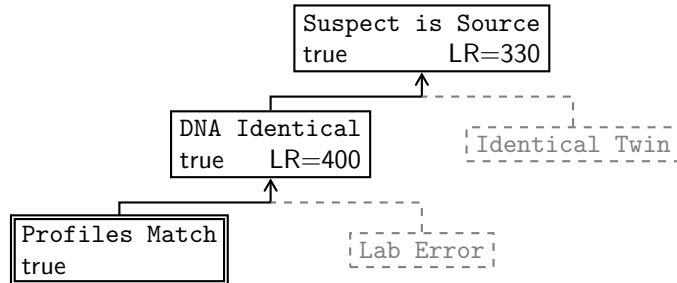
The likelihood ratio only reports the incremental force of the evidence. The posterior probability of the conclusion can have increased twofold but it can still be very low. To investigate this, we instantiate the same example using the posterior odds measure:

$$M_{odds}(s_i, \mathbf{E}_S) = \frac{P(s_i|\mathbf{E}_S)}{P(\overline{s_i}|\mathbf{E}_S)}$$

The result is shown in Figure 11. Note that the outcome of the 'DNA Identical' has not changed (it is highly likely to be true) with respect to Figure 10b but the outcome of the 'Suspect is Source' has changed to the 'false' value. Even though the likelihood of the evidence points in the direction of the 'true' outcome, it is still more likely that the suspect is not the source.

These experiments show that the different measures of strength convey a different argumentative interpretation. In the likelihood ratio approach the emphasis is on the change in belief. The

```
┌─────────────────────────┐
│ Suspect is Source       │
│ true          LR=330    │
└─────────────────────────┘
           ↑
┌──────────────────┐        ┌ ─ ─ ─ ─ ─ ─ ─ ─ ┐
│ DNA Identical    │          Identical Twin
│ true     LR=400  │        └ ─ ─ ─ ─ ─ ─ ─ ─ ┘
└──────────────────┘
        ↑
┌╥─────────────────╥┐    ┌ ─ ─ ─ ─ ─ ┐
║ Profiles Match   ║      Lab Error
║ true             ║    └ ─ ─ ─ ─ ─ ┘
└╨─────────────────╨┘
```

(a) The first piece of evidence is entered and the outcomes propagates through the graph.

```
┌─────────────────────────┐
│ Suspect is Source       │
│ true          LR=2.0    │
└─────────────────────────┘
           ↑
┌──────────────────┐        ┌╥────────────────╥┐
│ DNA Identical    │        ║ Identical Twin  ║
│ true     LR=400  │        ║ true            ║
└──────────────────┘        └╨────────────────╨┘
        ↑
┌╥─────────────────╥┐    ┌ ─ ─ ─ ─ ─ ┐
║ Profiles Match   ║      Lab Error
║ true             ║    └ ─ ─ ─ ─ ─ ┘
└╨─────────────────╨┘
```
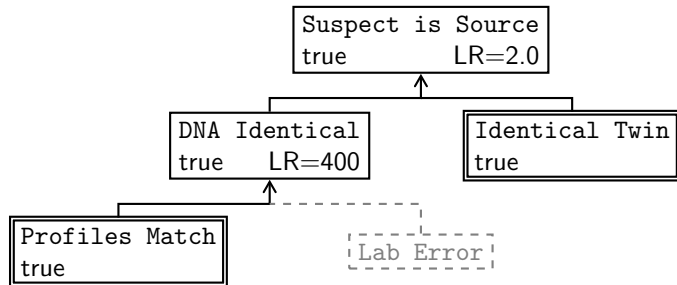
(b) When a second piece of evidence is entered, the support graph does not need to be recomputed. We can simply redo the labelling.

Figure 10: Support graphs labelled with instantiations and outcomes according to the likelihood ratios in the BN model.

```
┌─────────────────────────┐
│ Suspect is Source       │
│ false       odds=4.51   │
└─────────────────────────┘
           ↑
┌──────────────────┐        ┌╥────────────────╥┐
│ DNA Identical    │        ║ Identical Twin  ║
│ true  odds=44.6  │        ║ true            ║
└──────────────────┘        └╨────────────────╨┘
        ↑
┌╥─────────────────╥┐    ┌ ─ ─ ─ ─ ─ ┐
║ Profiles Match   ║      Lab Error
║ true             ║    └ ─ ─ ─ ─ ─ ┘
└╨─────────────────╨┘
```
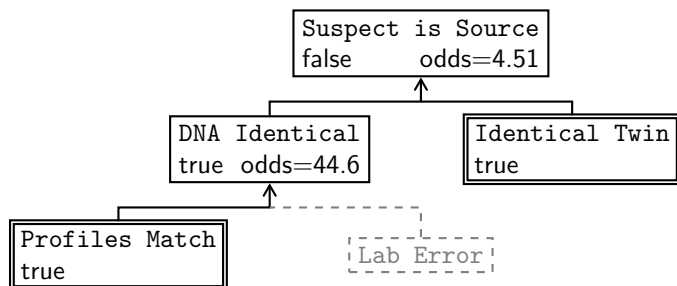
Figure 11: The same evidence, now evaluated using the odds measure of strength.

posterior odds approach, in contrast, more clearly shows the degrees of belief in the conclusions. However, the degrees to which these beliefs are based on the evidence is not as clear as in the other approach.

6 Discussion and future research

In this paper we have presented support graphs as a novel tool in the explanation of Bayesian network models of evidence. These models allow for a more argumentatively clear way of presenting probabilistic information. We have shown that they function as a template to construct arguments about the case that is modelled in the BN. Such argumentative models can be used to explain the information modelled in, or the conclusions issued by Bayesian network models. We emphasise that, in particular in the case of legal reasoning, the interpretation of statistical evidence can be hard. Argumentative analysis of the probabilistic evidence can facilitate the correct interpretation of such evidence.

One of the potential shortcomings of this method is that, given a particular set of evidence, the support graph generates only one argument for the variable of interest. This means that the weighing of pros and cons—what argumentation is usually used for—is shifted from the argumentation to the calculations on the support graph. What we do maintain from the argumentation perspective is the possibility to identify possible attackers. In future research we hope to establish a fully formalised translation to formally well defined arguments such that we can investigate how probabilistic arguments attack and support each other.

In this paper we have shown that the support graph model can be used to analyse different ways of probabilistic reasoning. A more extensive analysis of these results and an exploration of possible other approaches to argumentatively label these graphs can prove interesting. One suggestion for future research is to investigate the role of prior probabilities. We have seen that the support graph abstraction is a useful tool in the analysis of probabilistic argumentation. In this paper, we have demonstrated the flexibility of the support graph model using the likelihood ratio and the posterior odds. We suggest that, to construct strong arguments, we might have to include a high prior as a special kind of support factor.

To summarise, this paper has introduced support graphs as a useful tool in the argumentative analysis of Bayesian networks. As an intermediate model, support graphs abstract away from actual evidence and outcomes. We have shown that, given the support graph, we can reconstruct different argumentative results about the case.

Acknowledgments

References

[1] V. Crupi, K. Tentori, and M. Gonzales. On Bayesian measures of evidential support: Theoretical and empirical issues. *Philosophy of Science*, 74(2):229–252, 2007.

[2] P. Diaconis and D. Freedman. The persistence of cognitive illusions. *Behavioral and Brain Sciences*, 4(3):333–334, 1981.

[3] M. J. Druzdzel. Qualitative verbal explanations in Bayesian belief networks. *Artificial Intelligence and Simulation of Behaviour Quarterly*, 94:43–54, 1996.

[4] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 2005.

[5] I. Evett. Avoiding the transposed conditional. *Science & Justice*, 35(2):127–131, 1995.

[6] N. E. Fenton, M. Neil, and D. A. Lagnado. A general structure for legal arguments about evidence using Bayesian Networks. *Cognitive Science*, 37(1):61–102, 2013.

[7] T. F. Gordon and D. Walton. Proof burdens and standards. In I. Rahwan and G. R. Simari, editors, *Argumentation in Artificial Intelligence*, pages 239–258. Springer, 2009.

[8] U. Hahn and M. Oaksford. The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review*, 114(3):704–732, 2007.

[9] J. B. Kadane and D. A. Schum. *A Probabilistic Analysis of the Sacco and Vanzetti Evidence*. Wiley, New York, 1996.

[10] J. Keppens. Argument diagram extraction from evidential Bayesian Networks. *Artificial Intelligence and Law*, 20(2):109–143, 2012.

[11] J. R. Koiter. Visualizing inference in Bayesian Networks. Master's thesis, Delft University of Technology, 2006.

[12] C. Lacave and F. J. Díez. A review of explanation methods for Bayesian Networks. *Knowledge Engineering Review*, 17(2):107–127, 2002.

[13] C. Lacave, M. Luque, and F. J. Díez. Explanation of Bayesian Networks and influence diagrams in Elvira. *Systems, Man, and Cybernetics, Part B*, 37(4):952–965, 2007.

[14] D. A. Lagnado, N. E. Fenton, and M. Neil. Legal idioms: a framework for evidential reasoning. *Argument and Computation*, 4(1):1–18, 2012.

[15] K. B. Laskey and S. M. Mahoney. Network fragments: Representing knowledge for constructing probabilistic models. In *Uncertainty in Artificial Intelligence*, pages 334–341. Morgan Kaufmann, 1997.

[16] T. S. Levitt and K. B. Laskey. Computational inference for evidential reasoning in support of judicial proof. *Cardozo Law Review*, 22:1691–1731, 2000.

[17] S. Modgil and H. Prakken. A general account of argumentation with preferences. *Artificial Intelligence*, 195:361–397, 2013.

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

[19] J. L. Pollock. Defeasible reasoning. *Cognitive Science*, 11:481–518, 1987.

[20] H. J. Suermondt. *Explanation in Bayesian Belief Networks*. PhD thesis, Departments of Computer Science and Medicine, Stanford University, Stan- ford, 1992.

[21] W. C. Thompson and E. L. Schumann. Interpretation of statistical evidence in criminal trials: the prosecutor's fallacy and the defense attorney's fallacy. *Law and Human Behavior*, 11(3):167–187, 1987.

[22] S. T. Timmer, J.-J. C. Meyer, H. Prakken, S. Renooij, and B. Verheij. Inference and attack in Bayesian networks. In K. Hindriks, M. de Weerdt, B. van Riemsdijk, and M. Warnier, editors, *Proceedings of the 25th Benelux Conference on Artificial Intelligence*, pages 199–206, 2013.

[23] S. T. Timmer, J.-J. C. Meyer, H. Prakken, S. Renooij, and B. Verheij. Extracting legal arguments from forensic Bayesian networks. In R. Hoekstra, editor, *Legal Knowledge and Information Systems. JURIX 2014: The Twenty-seventh Annual Conference*, volume 217, pages 71–80, 2014.

[24] F. H. van Eemeren, B. Garssen, E. C. W. Krabbe, A. F. S. Henkemans, B. Verheij, and J. H. M. Wagemans. Argumentation and artificial intelligence. In *Handbook of Argumentation Theory*. Spinger, 2014.

[25] B. Verheij. Deflog: on the logical interpretation of prima facie justified assumptions. *Journal of Logic and Computation*, 13(3):319–346, 2003.

[26] C. Vlek, H. Prakken, S. Renooij, and B. Verheij. Modeling crime scenarios in a Bayesian network. In B. Verheij, E. Francesconi, and A. Gardner, editors, *Proceedings of the 14th International Conference on AI and Law*, pages 150–159, 2013.

[27] C. S. Vlek, H. Prakken, S. Renooij, and B. Verheij. Building Bayesian networks for legal evidence with narratives: a case study evaluation. *Artificial Intelligence and Law*, 22(4):375–421, 2014.

[28] G. A. Vreeswijk. Argumentation in Bayesian belief networks. In I. Rahwan, P. Moraïtis, and C. Reed, editors, *Argumentation in Multi-Agent Systems*, volume 3366 of *Lecture Notes in Computer Science*, pages 111–129. Springer Berlin / Heidelberg, 2005.

[29] J. H. Wigmore. *The Principles of Judicial Proof*. Little, Brown and Company, Boston, 1913.