Explaining Legal Bayesian Networks Using Support Graphs

Sjoerd T. TIMMER [a], John-Jules Ch. MEYER [a], Henry PRAKKEN [a,b], Silja RENOOIJ [a], Bart VERHEIJ [c]

[a] *Department of Information and Computing Sciences, Utrecht University*
[b] *Faculty of Law, University of Groningen*
[c] *Institute of Artificial Intelligence, University of Groningen*

Abstract. Legal reasoning about evidence can be a precarious exercise, in particular when statistics are involved. A number of recent miscarriages of justice have provoked a scientific interest in formal models of legal evidence. Two such models are presented by Bayesian networks (BNs) and argumentation. A limitation of argumentation is that it is difficult to embed probabilities. BNs, on the other hand, are probabilistic by nature. A disadvantage of BNs is that it can be hard to explain what is modelled and how the results came about. Assuming that a forensic expert presents evidence in a way that is either already a BN or expressed in terms that easily map to a simple BN, we may wish to express the same information in argumentative terms. We address this issue by translating Bayesian networks to arguments. We do this by means of an intermediate structure, called a *support graph*, which represents the variables from the Bayesian network, maintaining independence information in the network, but connected in a way that more closely resembles argumentation. In the current paper we test the support graph method on a Bayesian network from the literature. We argue that the resulting support graph adequately captures the possible arguments about the represented case. In addition, we highlight strengths and limitations of the method that are revealed by this case study.

Keywords. Argumentation, Bayesian networks, Explanation, Legal evidence, Legal reasoning, Probabilistic reasoning, Support graphs.

1. Introduction

A number of formal models for reasoning under uncertainty—such as legal reasoning about evidence—have been introduced. We study argumentation and Bayesian networks (BNs). With the introduction of scientific methods in the investigation of legal evidence, a number of challenges have arisen. Most notable is the question how to reason with probabilistic evidence. Since forensic evidence is often presented in the form of probabilities, it has become necessary to find models of legal evidence and legal proof that can accommodate probabilities. This is illustrated by a number of recent miscarriages of justice, such as in the infamous cases of Sally Clark in the UK and Lucia de Berk in the Netherlands. We argue that problems originate from a communication barrier between scientific and legal experts. Where scientific experts are usually well-versed in mathematics and probability theory, lawyers and judges are often more accustomed to arguments.

In previous work we have introduced support graphs as an intermediate structure between arguments and BNs [17,18]. In this paper we test the support graph method on a BN from the literature, which provides an evaluation of the method. We do not aim to present a normative system but rather a descriptive system that can be used to explain the evidence in order to explain what is modelled by that BN and which arguments are allowed by the information captured in it. We find that the resulting arguments are as expected, but we also highlight a limitation of the approach, which is that when a conclusion is supported by multiple pieces of evidence the individual contributions are not clearly distinguishable.

In the next section we review relevant background knowledge and we illustrate the support graph method by means of a small toy example. In Section 3 we present our main case study. In Section 4 we discuss the results and provide references to related and possible future research.

2. Background

2.1. Argumentation

Argumentation [19,3] is the study of arguments, counter arguments and argument evaluation. Formal argumentation is based on logic, and a number of systems that implement this coexist [14,7,5]. Typical for argumentation models is that arguments are built by combining inference rules. Starting with some premises, a rule with some of those premises as its antecedents can be applied. Further rules can be applied if their antecedents are satisfied by either premises or conclusions of rules that were already applied. In this fashion a sequence (or in fact a tree) of inferences can be constructed which we call an argument.

In common sense reasoning, inferences are usually not strict but *defeasible*. A defeasible inference can have exceptional circumstances under which the inference does not apply. An argument that uses a defeasible inference can be attacked by another argument that concludes that the exceptional circumstances hold. Such an attacking argument is called an *undercutter*. Another kind of attack is *rebuttal*, which tries to prove the contrary of a statement rather than attacking the link between statements [15].

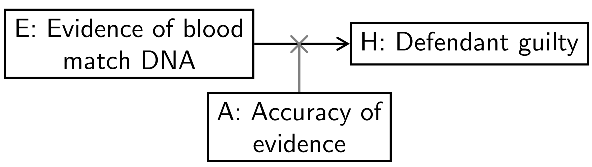
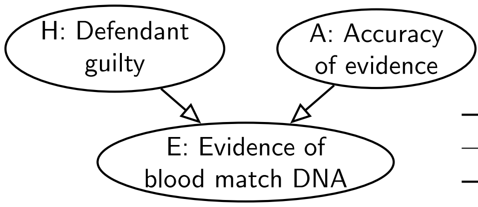


Figure 1. Examples of arguments. H is supported by E but this inference is undercut by A.

Figure 1 portrays a very small set of arguments about a DNA matching procedure. A match has evidential value for the hypothesis that the defendant is guilty. As an undercutter to this inference an argument is shown that says that the evidence is not accurate and therefore this conclusion cannot be drawn from that premise.

2.2. Bayesian networks

A Bayesian network (BN) is a model of a probability distribution in which a directed acyclic graph $G = (\mathbf{V}, \mathbf{E})$ consisting of variables $\mathbf{V}$ and directed edges $\mathbf{E}$ is used to express knowledge of a conditional independence relation. For a variable V we will use $\mathrm{Par}(V)$ to denote the set of parents in the graph and $\mathrm{Cld}(V)$ for the set of children. Figure 2a shows a small example, which was taken from Fenton et al. [4]. Three variables are shown that represent a minimalistic model of DNA identification. With every variable is associated a conditional probability table that specifies the probabilities of the outcomes of that variable conditional on any configuration of outcomes of parents of that node in the graph (see Figure 2b). Together these tables specify the full joint probability over all variables. Observations can be entered into the model and posterior probabilities of other nodes conditioned on those observations can be calculated. For instance, $P(H = true|E = true)$ can be calculated (to be 0.161).



(a) graphical structure

H:

false	0.99
true	0.01

A:

false	0.1
true	0.9

E:

H	false		true	
A	false	true	false	true
false	0.5	$1 - 10^{-6}$	0.5	0.0
true	0.5	10^{-6}	0.5	1.0

(b) conditional probability tables

Figure 2. Small example BN. From Fenton et al. [4].

One of the aspects of BNs that can be hard to explain to legal experts who are not trained in probability theory, is that the graphical structure of the BN encodes probabilistic independences. When two adjoining edges converge (such as in Figure 2a) this is called a *head-to-head* connection. A node on a trail (a simple path in the undirected version of the graph) is said to *block* inference through that trail if it is observed but not a head-to-head connection, or if it is an unobserved head-to-head node without observed descendants. A trail is called active when it is not blocked by any of the nodes on the trail. If there are no active trails between two nodes, they are conditionally independent given the evidence.

A head-to-head-connection can model *explaining away*, which occurs when two parents of a node both have a positive correlation with their common child but feature a negative correlation with each other conditioned on this child.

The fact that new evidence can sometimes remove dependencies (when it blocks the last remaining path) and sometimes add new possible dependencies (when it unblocks a head-to-head connection) can be the source of confusion.

We observe that the step-wise propagation of evidence through a BN is similar to making inferential steps in argumentation. There are three distinct ways in which an atomic inferential step can be made in a BN. Two simple methods are to follow an edge from the BN graph and reason from one side to the other. This is allowed in both directions so we obtain two possible kinds of inferences, which we will treat slightly differently. Note that although these inferences are on their own all acceptable, there are some restrictions when combining inferential steps, which we will discuss below. The third way to make an inference is to reason from one parent of a node to another parent of that same node. This is possible because *explaining away* captures a more complex interaction than the combination of one inference along the direction of an edge with an inference against the direction of another edge. In fact, this combination is exactly the only thing that is not allowed when combining inferential steps. Instead the *explaining away* inference between parents is the only allowed way to reason between these variables.

2.3. Support graphs

To explain inference in a BN in argumentative terms, we have introduced an intermediate model that we have called a support graph [17,18]. A support graph is a compact representation of all possible inferential steps in a given BN. It maintains the independence information from the BN but presents the variables in a graphical structure that more closely matches an argumentative interpretation. It does this in a way that correctly reflects the complex interactions between parents of a common child. It is not an argument graph because it presents the variables and not specific variable assignments.

A support graph is constructed by iteratively deepening it. We start with a variable of interest $V^\star$ and check for all of the three kinds of possible inferences that we mentioned above: parents, children and parents of children. In fact this collection is often referred to as the Markov blanket (MB) of a node. We maintain for every branch a set of variables $\mathcal{F}$ that contains *forbidden* variables that cannot be used in further support. We use this set to preclude a number of prohibited combinations of inferences. In every step further down in the support graph, $\mathcal{F}$ is copied and possibly enlarged. This is described by Definition 1 and clarified in Figure 3.

Definition 1 (Support graph)**.** *Given a BN with graph* $G = (\mathbf{V}, \mathbf{E})$ *and a variable of interest* $V^\star$*, a* support graph *is a tuple* $\langle \mathcal{G}, \mathcal{V}, \mathcal{F} \rangle$ *where* $\mathcal{G}$ *is a directed graph* $(\mathbf{N}, \mathbf{L})$*, consisting of nodes* $\mathbf{N}$ *and edges* $\mathbf{L}$*,* $\mathcal{V} : \mathbf{N} \mapsto \mathbf{V}$ *assigns variables to nodes, and* $\mathcal{F} : \mathbf{N} \mapsto \mathcal{P}(\mathbf{V})$ *assigns sets of variables to each node, such that* $\mathcal{G}$ *is the smallest graph containing the node* $N^\star$ *(for which* $\mathcal{V}(N^\star) = V^\star$ *and* $\mathcal{F}(N^\star) = \{V^\star\}$*) closed under the following expansion operation:*

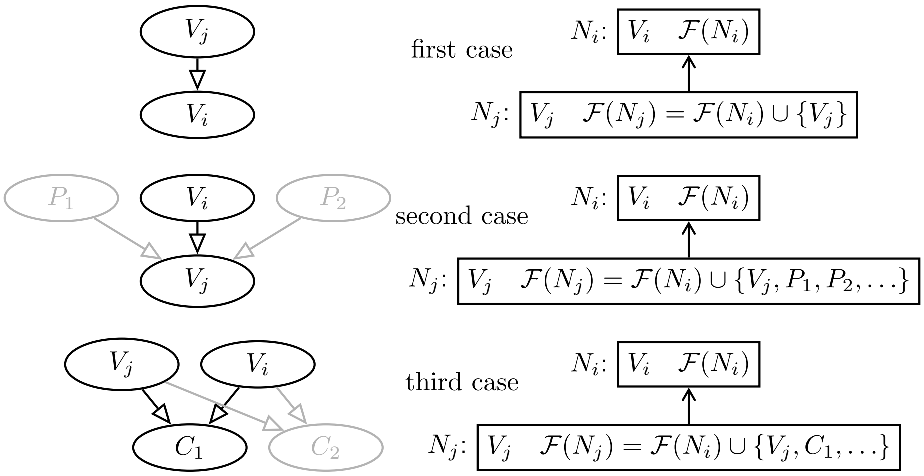


Figure 3. Visual representation of the three cases in Definition 1. A support node for variable V_i can obtain support in three different ways from a variable V_j, depending on its graphical relation to V_i.

A supporter N_j *with variable* $\mathcal{V}(N_j) = V_j$ *is added as a parent to a node* N_i *(with* $V_i = \mathcal{V}(N_i)$*), whenever:*

1. $V_j \in \mathrm{Par}(V_i) \setminus \mathcal{F}(N_i)$*, or*
2. $V_j \in \mathrm{Cld}(V_i) \setminus \mathcal{F}(N_i)$*, or*
3. $V_j \in \mathrm{Par}(\mathrm{Cld}(V_i) \setminus \mathcal{F}(N_i)) \setminus \mathcal{F}(N_i)$

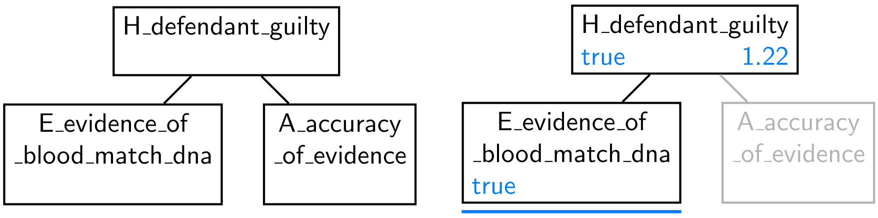
The forbidden set $\mathcal{F}(N_j)$ *of the new support node is respectively:*

1. $\mathcal{F}(N_i) \cup \{V_j\}$
2. $\mathcal{F}(N_i) \cup \{V_j\} \cup \{V_k \in \mathrm{Par}(V_j) | V_i \rightarrow V_j \leftarrow V_k \textit{ in } \mathbf{E} \textit{ but not } V_i - V_j\}$
3. $\mathcal{F}(N_j) \cup \{V_j\} \cup (\mathrm{Cld}(V_i) \cap \mathrm{Cld}(V_j))$

If a support node with this forbidden set and the same $\mathcal{V}(N_j)$ *already exists, that node is added as the parent of* N_i*, otherwise a supporting node* N_j *is created.*

Informally, support for a variable V_i is taken from the Markov blanket. Support from a neighbour V_j is omitted if it is in the forbidden set. Support through head-to-head connections is omitted if the common children are in the forbidden set.

We illustrate the support graph construction process with the example BN presented in Figure 2. As the variable of interest we take “H: Defendant guilty”. The first case from the support graph definition does not apply because H does not have parents. The second case does apply, so we create a support for H from E. The $\mathcal{F}$-set for the new support node contains H and E, but also A because it is another parent of E and reasoning between parents through the child is prohibited. This means that this branch cannot be supported by further nodes. However, we can retrace our steps and try to find more support for the variable of interest. We have not yet invoked the third case on that node, so we add a support for H from A. We can see from the definition that not only are H and A added to $\mathcal{F}$ of this node but also E because common children are added to this set. This means that this



(a) support graph for BN in Figure 2.

(b) Augmented support graph that has been labelled with likelihood ratios and best supported outcomes.

Figure 4. First example BN. From Fenton et al. [4].

branch is now also finished and the support graph construction is completed. The resulting support graph is shown in Figure 4a. We note that the term *support* in this case means that one of the outcomes supports either outcome of the supported variable and not necessarily that the truth of one variable supports the truth of the other. A support graph can, already on its own, be a valuable explanation tool to show how inference propagates through a BN. It can, however, also be used as a skeleton for argument construction. In [17] we have shown how different labellings can be applied to turn support graphs into arguments. The support graph method offers a number of ways in which the correct assignment labels can be added to the nodes. For this, a probabilistic measure of inferential strength is used. A well-known measure of strength [1] of evidence on a hypothesis is the likelihood ratio (LR). We have added outcome assignments for the outcome with the highest LR in Figure 4b together with that LR. The resulting argument graph shows that the evidence (DNA match) has a likelihood ratio of 1.22 on the truth outcome of the hypothesis. The second branch of the support graph has been greyed out because no observation is available for the accuracy. We still show the accuracy node to indicate that this variable can provide further support (if the match turns out to be accurate) or attack (if the match is not accurate). These are exactly the argument and the possible undercutter that one would expect from this case.

3. The Lulu case study

As a case study we examine a larger network, which originates from Taroni et al. [16, page 50]. The BN is shown in Figure 5. This model represents a case in which a suspect (Jack) is charged with the stabbing of Lulu. The available evidence in this case is (E) that a DNA profile match is found between his blood and a sample taken from the crime scene (CS) and (W) that a witness (John) testifies that he saw Jack near Lulu's house shortly after the time of the stabbing. The right part of the network models that Jack may have been in love with Lulu, causing John to be jealous of Jack which could pose an alternative explanation for John's testimony. An arrow is drawn between nodes J and F because it is more likely that Jack was near Lulu's place if he was in love with her.

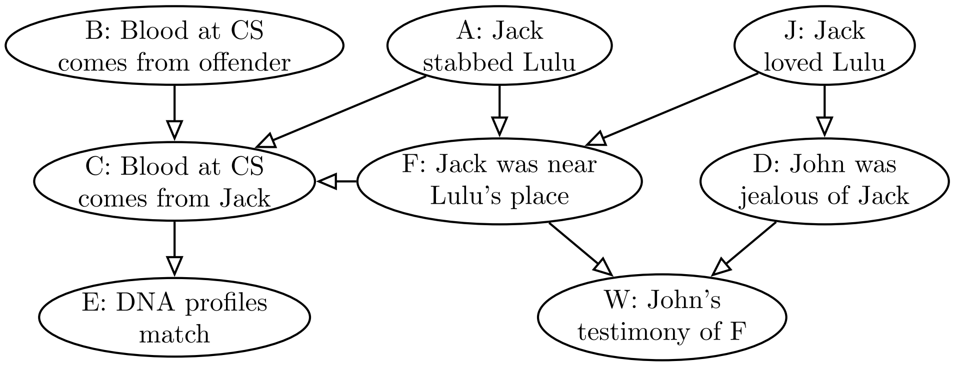


Figure 5. The Lulu network [16]. We have omitted the conditional probability tables for brevity.

Figure 6 shows the resulting support graph labelled with likelihood ratios and corresponding outcomes. Branch 1 shows the argument from the DNA match (E), via the sample identity C, to the conclusion that Jack indeed stabbed Lulu. I.e., the profiles match, therefore the blood at the crime scene is from Jack and therefore it must have been Jack who stabbed Lulu. This is the most important argument that we expect to find in this network. However, we also expect to find the undercutter that states that if the blood at the crime scene is not from the offender, then Jack's DNA at the crime scene is no proof that he stabbed her. We indeed recognise this as the greyed out node on the far right of the graph.

We identify a second incriminating argument in branch 2. It states that Jack was near Lulu's place. For this two possible sub-arguments are identified. Firstly, to leave a bloodstain, one needs to be present at the crime scene so a DNA match is evidence that Jack was present. And secondly, there is a testimony by John. For either sub-supporter we see a possible attacker. For the former sub-argument we see again that if the blood at the crime scene is not from the offender is does not hold. For the latter sub-argument we see that it may not hold (or be as strong) if John was jealous of Jack.

A third argument is found (branch 3) that weakens the conclusion that Jack stabbed Lulu. It states that John's testimony against Jack could be explained by John's jealousy which in turn could be explained by the fact that Jack loved Lulu (otherwise John would not need to be jealous). The latter explains why Jack may have been near Lulu's house, even if he did not stab her.

Looking at the LR measures of inferential strength, we observe that branch 3 poses, by far, the least support. The likelihood ratio valuation of that evidence is only 1.26. By comparison, the support from the DNA match is a lot stronger ($\text{LR} = 498$). However, if we examine branch 2 of the support graph, we observe that it has a collective evidential strength of 594. This is even larger than the DNA matching evidence. This is the result of the fact that our support graph construction method has identified that the DNA match is not just a supporter for the hypothesis that Jack stabbed Lulu, but also for the fact that he was near her place. Although this is not incorrect, it shows that our current method (only presenting collective probabilistic strength) may be too crude because in this branch we can not distinguish from which observation the stronger support originates.

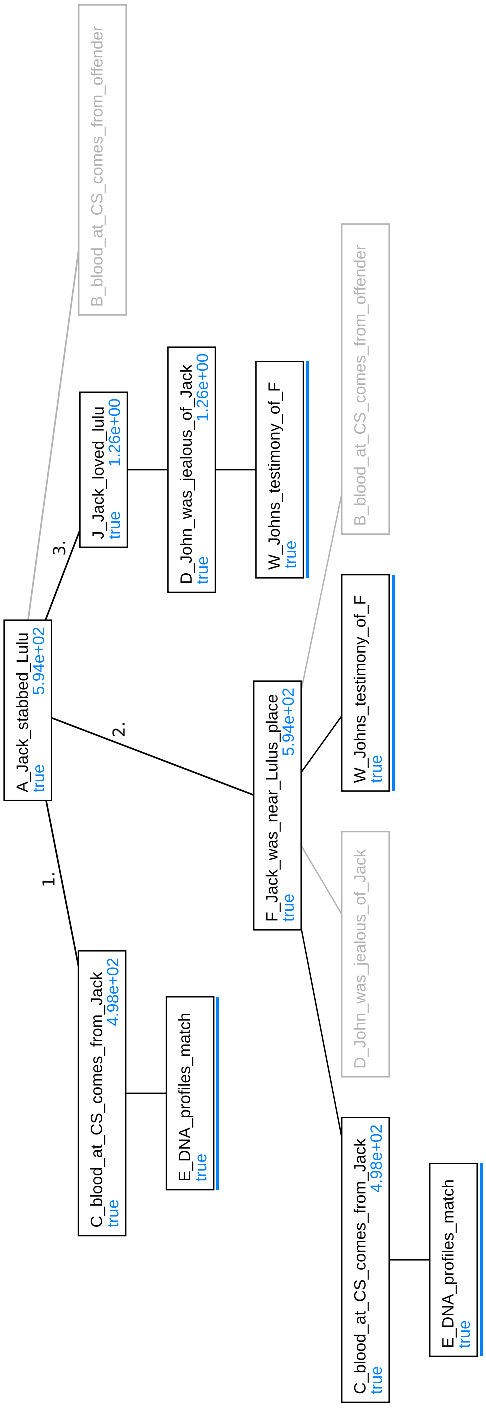


Figure 6. Support graph for the Lulu networks. Branches from the root have been numbered 1-3 for easy reference from the text.

4. Related research

BNs have been used as a tool to model forensic evidence, and even complete legal cases [6,4,11,13,12,20,21]. 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 [10,9], Koiter [8] and Druzdzel [2]. All of these methods attempt to explain what is modelled in a BN by visually displaying the different interactions between variables or by verbally presenting the relations among variables. However, none of these methods use argumentation.

Vreeswijk [22] 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. A method to extract argument that follow BN reasoning based on numeric information is *argument diagram extraction* [7]. An argument diagram is a graphical structure that informally represents Bayesian argumentation, but does not allow one to identify possible counter arguments. In argument diagrams, therefore, only one side of the story is highlighted.

5. Conclusion

We have applied the support graph method to an example BN from the existing literature to identify argumentative structures in BNs for explanation purposes. This serves to illustrate the approach and suggests that support graphs indeed capture exactly the inferences that are probabilistically valid and are sufficiently concise to provide a good overview of a case. We take these results to be a good indication that support graphs are indeed suitable models to explain inference in BNs. To conclude that they have good explanatory value in real world applications, however, requires further investigation.

We suggest that in future research it should be investigated how individual strengths of contributions to support can be identified. For now we have presented the collective likelihood ratio of all branches under the current node as a measure of its argumentative strength. However, some supporting branches may contribute more than others. It may even be the case that some decrease the probability of the conclusion but are outweighed by others. This is valuable information when we want to construct arguments or explanations about the case.

Acknowledgement

This work is part of the project "Designing and Understanding Forensic Bayesian Networks with Arguments and Scenarios", supported by the Forensic Science research program financed by the Netherlands Organisation for Scientific Research (NWO). See `http://www.ai.rug.nl/~verheij/nwofs/`.

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

[3] 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.

[4] 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.

[5] T. F. Gordon, H. Prakken, and D. Walton. The carneades model of argument and burden of proof. *Artificial Intelligence*, 171(10–15):875–896, 2007.

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

[7] J. Keppens. Argument diagram extraction from evidential Bayesian networks. *Artificial Intelligence & Law*, 20:109–143, 2012.

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

[9] C. Lacave and F. J. Díez. A review of explanation methods for Bayesian networks. *Knowledge Engineering Review*, 17:107–127, 2002.

[10] 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.

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

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

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

[14] S. Modgil and H. Prakken. The ASPIC+ framework for structured argumentation: a tutorial. *Argument and Computation*, 5(1):31–62, 2014.

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

[16] F. Taroni, C. Aitken, P. Garbolino, and A. Biedermann. *Bayesian Networks and Probabilistic Inference in Forensic Science*. John Wiley & Sons, Ltd, 2006.

[17] S. T. Timmer, J.-J. C. Meyer, H. Prakken, S. Renooij, and B. Verheij. Explaining Bayesian networks using argumentation. In S. Destercke and T. Denoeux, editors, *Proceedings of the 13th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, volume 9161 of *Lecture Notes in Artificial Intelligence*, pages 83–92. Springer, 2015.

[18] S. T. Timmer, J.-J. C. Meyer, H. Prakken, S. Renooij, and B. Verheij. A structure-guided approach to capturing Bayesian reasoning about legal evidence in argumentation. In *Proceedings of the 15th International Conference on Artificial Intelligence and Law*, pages 109–118. ACM, 2015.

[19] F. H. van Eemeren, B. Garssen, E. C. W. Krabbe, A. F. S. Henkemans, B. Verheij, and J. H. M. Wagemans. *Handbook of Argumentation Theory*. Springer, Dordrecht, 2014.

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

[21] 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.

[22] 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.