On Extracting Arguments from Bayesian Network Representations of Evidential Reasoning

Jeroen Keppens
Department of Informatics
King's College London
Strand, London WC2R 2LS
jeroen.keppens@kcl.ac.uk

ABSTRACT

Bayesian networks are a predominant approach to analyse the findings of forensic scientists. In part, this is due to the way the Bayesian approach fits the scientific method employed in forensic practice. The design of Bayesian networks that accurately and comprehensively represent the relationships between investigative hypotheses and evidence remains difficult and sometimes contentious, however. Recent research has shown that argumentation can inform the construction of Bayesian networks. But argumentation is a distinct approach to evidential reasoning with its on representation formalisms. This issue could be alleviated if it were easy to represent Bayesian networks as argumentation diagrams. This position paper presents an investigation into the similarities, differences and synergies between Bayesian networks and argumentation diagrams and shows a first version of an algorithm to extract argumentation diagrams from Bayesian networks.

Keywords

Evidential reasoning, Bayesian Networks, Argumentation

1. INTRODUCTION

Argumentation Diagrams (ADs) and Bayesian Networks (BNs) are the two predominant approaches to legal evidential reasoning. Both approaches provide a means to identify and evaluate the hypotheses that may have produced the available evidence in a case and to assess their plausibility [2].

Though they can also be employed as a means to model legal arguments [5], BNs are used in evidential reasoning to quantify the strength of support of evidence for alternative hypotheses, based on a combination of domain knowledge and quantitatively expressed beliefs. In other words, a BN summarises ones understanding of relationships the evidence in a case and its hypothesised explanations into numerical value. This feature of BNs enables two important types of application. Firstly, it provides a means to condense complex technical or scientific reasoning about the strength of inferences between two propositions. For instance, Mortera et. al. employ a BN approach based on first principles to assess the probative force of evidence of a partial match between a person's DNA and that found in a pool of biological material to which multiple persons may have contribute [16]. Secondly, it provides a means to estimate how much more strongly evidence supports one hypothesis over another, and how strongly the addition or removal of evidence would affect the relative level of support for alternative hypothesis, with a view to suggest the investigative actions that are most suitable for differentiating between the discovered alternative hypotheses and to assess the amount of information that could be gained by an investigative action or by the discovery of a particular piece of additional evidence [14]. This allows the BN approach to be employed for managing investigations efficiently [7].

An AD approach to evidential reasoning visualises and relates inferences made from evidence, identifies their constituent elements and depicts the different ways in which these inferences are supported and the ways in which they can be undermined. As ADs are a means of visualisation, they normally employ far richer and diverse representation formalisms than that of BNs. These representation formalisms aim to differentiate the different kinds of information and knowledge that are part of evidential reasoning. As such, ADs are used to marshal all the information pertaining to a particular case and to scrutinise the ways in which they are related in evidential reasoning. Like BNs, ADs can be employed to guide inquiries, but with a view to the validity of evidential reasoning inferences and testing those inferences for potential flaws [18].

Although ADs and BNs serve different purposes in evidential reasoning about a particular case, both approaches can also be considered as offering a different perspective on the same case. Therefore, there may be scope for ADs and BNs to inform one another. This paper is concerned with the way the AD perspective could inform the construction of BNs. This idea was formulated by Hepler et. al. [12], whose detailed examination of a case study shows how content from an AD can be incorporated into a BN. It is developed further in this paper by proposing a means to help compare the content of a BN with that of ADs, by means of a novel method to extract ADs from a BN. The proposed approach is illustrated by means of a number of BNs from the forensic science literature. It aims to inspire the development of future tools to design BNs for evidential reasoning and export ADs based on forensic BNs for legal reasoning. Because ADs contain information that BNs do not (and vice versa), generating arguments from BNs cannot be fully automated. Any attempt to do so will require a limited AD representation formalism, which is what the work presented in this paper relies on.

The remainder of this paper is structured as follows. Section 2

ICAIL '11, Jun 06-10 2011, Pittsburgh, PA, USA
Copyright 2011 ACM 978-1-4503-0755-0/11/06...$10.00.

introduces evidential reasoning by means of BNs and ADs. This leads to an examination of the similarities and differences in the information contained in evidential reasoning BNs and ADs in Section 3. Section 3 focusses on information contained in ADs that is not readily available in BNs and on the elements of a BN that may inform their extraction. Section 4.2 employs this analysis to propose an algorithm to extract arguments from BNs, under certain simplifying assumptions, and Section 5 evaluates the proposed approach by applying it to BNs that have been taken from the Forensic Science literature.

2. BACKGROUND

2.1 Evidential Bayesian networks

A Bayesian network (BN) is a representation that facilitates the representation and calculation of complex joint probability distributions. In evidential reasoning, it is used to assess the probability of certain states and plausible observations in hypothetical situations.

A BN consists of a directed acyclic graph (DAG) $(\mathbf{V}_b, \mathbf{E}_b)$, where $\mathbf{V}_b$ is a set of vertices or nodes and $\mathbf{E}_b$ is a set of edges or arcs, and a set of conditional probability tables (CPTs), one for each vertex. Each vertex $V \in \mathbf{V}_b$ corresponds to a variable with a domain of mutually exclusive values $\mathbf{D}_V$. As such, the term vertex and variable of a BN can be used interchangeably. The situation in any possible world is described by assigning each variable V in the BN exactly one of the values $v_i \in \mathbf{D}_V$ of its domain (hereafter denoted $V : v_i$). The edges of the DAG of the BN define independence relations between vertices, by means of an assumption known as the Markov condition: given truth values for the immediate parents of any vertex V in the BN, V is independent from any combination of other vertices in the network excluding its own descendants [8].

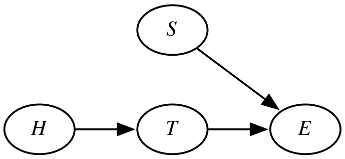

Figure 1: DAG of a simple BN

Figure 1 illustrates these idea with a simple BN of evidential reasoning. It contains a DAG with four nodes labelled H, T, S and E describing features relevant to a case where a suspect is accused of breaking a window. H represents the hypothesis that the suspect is guilty of the crime, T the transfer of glass fragments from window to the suspect's clothes, S whether sufficient long period of time has elapsed since the crime in which the glass fragments could have been shed from the suspect's clothes, and E the discovery of glass fragments matching the window in the suspect's clothes. All variables have boolean domains ({true,false}). Where boolean domains are used, the assignment of V : true will be denoted as v and the V : false as $\overline{v}$. While this example is simplistic (for illustrative purposes), it is representative of an evidential reasoning BN. The hypothesis under investigation is represented by a root node (H) in the BN, and the evidence by a leaf node (E). The edges represent causal relations in that committing the crime (H) causes glass fragments to end up in the perpetrator's clothes (T), and this may be discovered as evidence (E), even though the likelihood of the latter is reduced if a substantial amount of time has elapsed between crime and evidence collection.

	h	$\overline{h}$
$P(t\|H)$	0.9	0.01
$P(\overline{t}\|H)$	0.1	0.99

	t		$\overline{t}$	
	s	$\overline{s}$	s	$\overline{s}$
$P(e\|T,S)$	0.3	0.9	0	0
$P(\overline{e}\|T,S)$	0.7	0.1	1	1

Table 1: Sample CPTs for the simple BN example

As mentioned above, a CPT is associated with each variable. A CPT defines the probability distributions of the variable it is associated with, one for each combination of value assignments of its parents. In combination with the Markov condition mentioned earlier, this can be employed to calculate a joint probability distribution:

$$\begin{aligned} P :&\mathbf{D}_{V_1} \times \dots \mathbf{D}_{V_n} \mapsto [0,1] : \\ &(v_1, \dots, v_n) \rightarrow P(V_1 : v_1, \dots, V_n : v_n) \end{aligned} \quad (1)$$

where $\mathbf{V}_b = \{V_1, \dots, V_n\}$ and $v_i \in \mathbf{D}_{V_i}$. Thus, a BN can be defined by a tuple $\langle \mathbf{V}_b, \mathbf{E}_b, P \rangle$, where $(\mathbf{V}_b, \mathbf{E}_b)$ defines a DAG, each and $V \in \mathbf{V}_b$ possesses a domain $\mathbf{D}_V$ and P defines a probability distribution as in (1).

Sample CPTs for the ongoing example are shown in Table 1. CPTs facilitate the calculation of conditional probabilities considerably. For example:

$$\begin{aligned} P(e|h,s) =& \sum_T P(e|T,s) \times P(T|h) \\ =& 0.3 \times 0.9 + 0 \times 0.1 = 0.27 \end{aligned}$$

Similarly,

$$\begin{aligned} P(e|\overline{h},s) =& \sum_T P(e|T,s) \times P(T|\overline{h}) \\ =& 0.3 \times 0.01 + 0 \times 0.99 = 0.003 \end{aligned}$$

Bayesian analysis of forensic evidence involves the careful formulation of two hypothesis (H_1 and H_2) that are to be contrasted with one another by means of the available evidence (E). Typically, these correspond to explanations for the evidence put forward by the prosecution and the defence, but they may also be a working hypothesis put forward by investigators and the best alternative explanation. Next, the likelihood ratio LR is calculated, which compares the probability of the evidence under H_1 with that under H_2:

$$LR = \frac{P(E|H_1)}{P(E|H_2)} \quad (2)$$

If LR results in very high values above 1, say in the 100s, 1,000s or 10,000s, then it is reported that the evidence is moderately to very strongly "*consistent with*" hypothesis 1, compared to hypothesis 2. If LR results in values that are very close to 0, say 0.01, 0.001 or 0.0001, then it is reported that the evidence is moderately to very strongly "*consistent with*" hypothesis 2, compared to hypothesis 1. Thus, because:

$$\frac{P(e|h,s)}{P(e|\overline{h},s)} = \frac{0.27}{0.003} = 90$$

it can be argued that, according to the sample BN specified in Figure 1 and Table 1, the discovery of glass fragments in a suspect's clothes a substantial time after a perpetrator has broken a window, is moderately more consistent with the hypothesis that the suspect is guilty than the hypothesis that the suspect is innocent.

2.2 Argument diagrams

Argument diagrams (ADs) are visual representation of reasoning structure [17]. In their most basic form, they are directed graphs in which the vertices correspond to premises and conclusions and the edges to inferences between premises and conclusions. The premises and conclusions of ADs are propositions – statements that are deemed to be either true or false – rather than variables. As such, in an AD, the author commits to a particular truth value for the premises and conclusions of arguments.

The vertices and edges may be annotated with further information. Toulmin diagrams, for instance, will attach a warrant to specify how the inference was made, a backing to references support for the warrant and a qualifier to describe confidence in the strength of the inference [22]. Wigmore and Schum annotate their vertices and edges with symbols and text indicating the role of these elements in the argument's structure [18, 25]. These can extend the amount of information depicted by such ADs considerably. For example, edges in Wigmore and Schum's diagrams not only suggest inferential strength but also types of inferential support. Much of this argument specific information is not present in a BN. Therefore, this paper will employ a very simple AD representation formalism in which the edges are only annotated with a Toulmin diagram-like qualifier. Figure 2 illustrates this with a simple AD derived from the sample BN introduced in Section 2.1. In this diagram, the propositions e, t and h correspond to the assignments E : true, T : true and H : true respectively.

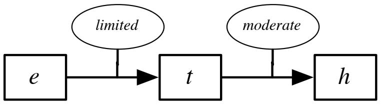

Figure 2: A simple AD

Generally speaking, arguments concerning evidence tend to depict what is known or observed, including the evidence, as the premises to inferences and the direction of reasoning is towards the hypothesis. The focus of the representation is on the individual inferences and on the relationship of their premises and/or conclusions to related arguments. More precisely, within the context of an individual inference, absolute truth values are assumed for the premises, including any major premise that warrants the inference itself, and the conclusion. But in the context of the broader diagram, these are related to reasons to reject them, such as conflicting propositions and alternative hypotheses. As demonstrated by Hepler et. al., this presentation of evidential can reveal useful knowledge that informs the design of BNs.

Note that this evidence-to-hypothesis-oriented, argument focussed representation is not a requirement of the application of ADs to evidential reasoning, as Bex et. al. have shown [2]. But it is the approach that this paper will adopt with a view to complement the story-based representation of BNs.

3. REPRESENTATION FORMALISM

3.1 Vertices

As explained above, the vertices of BNs and ADs represent different types of entities.

3.1.1 Vertex content

The vertices of BNs represent variables, each of which is associated with a domain of two or more values. A possible world is described by assigning each variable of the BN one value from its domain. Thus, the vertices of a BN represent multiple, mutually exclusive, possible worlds. The vertices of ADs represent propositions, usually representing a particular feature or property of possible worlds, and each vertex of an AD corresponds to only those possible worlds that share the particular feature referred to by the proposition. Consequently, while it is relatively straightforward to map propositions to boolean variables in extracting a BN from an AD [11], it is not easy to automatically map vertices of BNs to vertices of ADs.

Generally speaking, propositions can be derived from the variables and associated domains of a BN by constraining the value assignment of said variables. Categorical variables, which possess domains of values that cannot be meaningfully compared with one another, can give rise to propositions based on variable assignments and combinations thereof as defined by well-formed formulae of propositional calculus. Variables with ordered domains can also give rise to propositions based on lower and upper bounds as specified with comparative operators. In situations where arguments are to be extracted from very precise BNs, such as ones based on first principles of physics, for example, Davis's work on traffic accident reconstruction [9], fuzzy sets may need to guide the synthesis of relevant propositions to build arguments with. In such complex situations, the synthesis of propositions itself may become part of the arguments.

As such, the mapping of variables to propositions with a view to extract arguments from a BN is difficult to automate. While this constitutes an interesting research question, it is one that will not be addressed in this paper but left for future work. The remainder of this paper will assume that a BN only uses boolean variables. Propositions are derived from such variables, simply by committing a truth value assignment to the variable. Note that despite this simplifying assumption, the proposed approach still covers a substantial proportion of the BNs proposed in the Forensic Science literature. In what follows, a simplified notation of assignments will be employed whenever a variable V with a boolean domain is encountered: V : true will be denoted as v and V : false as $\overline{v}$.

3.1.2 Vertex types

Part of the explanatory value of an argumentation diagram lies in its extensive labelling scheme of propositions. BNs employ a comparatively much simpler labelling scheme. In a BN, a node corresponds to an observed or an unobserved variables. An observed variables amounts to a proposition that is known to be true, whereas unobserved variables to propositions with unknown truth values. An unobserved variable is said to be a hypothesis if it represents an explanation that is being proposed for the observations. Naturally, the hypothesis of a BN corresponds to the ultimate probandum of an AD and the observed variables to the evidence of the case. Hypotheses that are assessed by means of BNs are routinely decomposed into so-called partitioning hypotheses, each corresponding to a feature of the hypothesis of interest [4]. Such partitioning hypotheses of a BN correspond to penultimate probanda.

3.2 Edges

As the edges relate vertices to one another, these represent different kinds of relationships as well. Strictly speaking, the edges of a BN define conditional independence relations according to the Markov condition, as explained earlier. In practice, however, they represent influences between variable. In the example of Figure 1, for instance, the probability of the suspect incident (H) increases the likelihood (i.e. influences) of transfer of glass fragments (T),

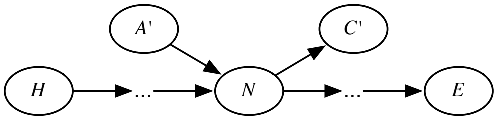

Figure 3: Supporting edges

which, in turn affects the likelihood of the retrieving glass fragments from the suspect (E).

The edges of ADs represent implications in a reasoning process. The focus is not so much on explaining the evidence but on assessing how the explanations where arrived at and whether the individual components are safe. For that reason, the direction of the representation is routinely, though not necessarily, inverse to the causal direction. In other words, a directed path of causal influences from a plausible hypothesis to evidence of BN can be represented as a set of arguments by inverting the path, starting at the evidence and ending at the hypothesis.

There are also edges that are relevant to arguments supporting certain hypotheses by means of a given set of evidence and that are not part of a directed path from hypothesis to evidence in the BN from which arguments are being extracted. Consider a directed path $H \rightarrow \ldots \rightarrow N \rightarrow \ldots \rightarrow E$ defined by a BN , where $H, N, E \in \mathbf{V}_b$. Let A' and C' be nodes of $\mathbf{V}_b$ such that $A' \rightarrow N \in \mathbf{E}_b$, $N \rightarrow C' \in \mathbf{E}_b$ and A' and C' are not part of a path between H and E. This situation is depicted in Figure 3. An AD proposed based on this BN by means of the aforementioned will at least contain a directed path $e \rightarrow \ldots \rightarrow n \rightarrow \ldots \rightarrow h$, where e, n and h are propositions derived from E, N and H respectively. Under the simplifying assumption committed to herein, the proposition n corresponds to assigning a value to the variable N. By doing so, and assuming that C' does not lead to evidence that is to be considered, it follows from the Markov condition that the probability distribution of C' becomes independent from the rest of the network and, therefore, can be ignored if merely for the sake of arguing how e might support h. The edge $A' \rightarrow N$ should be considered, however, and its role in an AD depends on effects in the CPT of N.

3.2.1 Second order influences

Second order influences are influences that affect the likelihood of the consequent variable, but only by modifying the effect or another influence (called a "first-order influence) on the consequent variable [13]. In the BN depicted in Figure 1, the edge $S \rightarrow E$ is an example of a second order influence. Here, the passing of time (S) does not cause the discover of glass fragments on the suspect (E) by itself. Instead, it is the transfer of glass fragments onto the suspect (T) that causes the glass fragment evidence (E). But if the transfer has occurred, then time will "inhibit" the causal effect of T on E. Therefore, $S \rightarrow E$ is deemed to be a second order influence.

Two types of second order influence can be distinguished in BNs with boolean or ordered domains: those that reduce the strength of effect of a first-order influence (called "inhibitors") and those that increase the strength of effect of a first-order influence (called "amplifiers"). An inhibitor that removes the effect of a first-order influence completely is known as a "disabler", and an amplifier that is so important that it is a requirement for a first-order influence to take effect is known as an "enabler" [13].

The CPTs of a BN can provide a strong indication as to whether influences are first-order or second-order. Let $A_1,\ldots,A_j,A_{j+1},\ldots,A_k,A_{k+1},\ldots,A_l,N$ be vertices with boolean domain in $\mathbf{V}_b$ such that the nodes A_i are parent nodes to N: $\forall i, A_i \rightarrow N \in \mathbf{E}_b$. To test whether influence of $A_{j+1}, \ldots, A_k$ on N is second-order to the influence of $A_1, \ldots, A_j$ on N, consider any combination of variable assignments $\mathbf{A}_c$, $\overline{\mathbf{A}_c}$, $\mathbf{A}_i$, $\overline{\mathbf{A}_i}$, $\overline{\overline{\mathbf{A}_i}}$, $\mathbf{A}_e$ and $\overline{\mathbf{A}_e}$ such that:

- $\mathbf{A}_c$ and $\overline{\mathbf{A}_c}$ are assignments of $A_1, \ldots, \underline{A_j}$ where $\mathbf{A}_c$ is any combination of value assignments and $\mathbf{A}_c$ is a combination of value assignments such that $\wedge_{a \in \overline{\mathbf{A}_c}} a =$ false.
- $\underline{\mathbf{A}_i}$, $\overline{\mathbf{A}_i}$ and $\overline{\overline{\mathbf{A}_i}}$ are assignments of $A_{j+1}, \ldots, A_k$ where $\wedge_{a \in \underline{\mathbf{A}_i}} a =$ true, $\wedge_{a \in \overline{\mathbf{A}_i}} a =$ false and $\wedge_{a \in \overline{\overline{\mathbf{A}_i}}} a =$ false.
- $\mathbf{A}_e$ and $\overline{\mathbf{A}_e}$ are assignments of $A_{k+1}, \ldots, A_l$ where $\mathbf{A}_e$ is any combination of value assignments and $\wedge_{a \in \overline{\mathbf{A}_e}} a =$ false

$A_{j+1}, \ldots, A_k$ are set to constitute an inhibitor with respect to the effect of $A_1, \ldots, A_j$ on N if:

$$\begin{aligned} P(n|\mathbf{A}_c \cup \overline{\mathbf{A}_i} \cup \mathbf{A}_e) &= P(n|\mathbf{A}_c \cup \overline{\overline{\mathbf{A}_i}} \cup \mathbf{A}_e) \geq P(n|\mathbf{A}_c \cup \underline{\mathbf{A}_i} \cup \mathbf{A}_e) \\ P(n|\overline{\mathbf{A}_c} \cup \overline{\mathbf{A}_i} \cup \overline{\mathbf{A}_e}) &= P(n|\overline{\mathbf{A}_c} \cup \overline{\overline{\mathbf{A}_i}} \cup \overline{\mathbf{A}_e}) \\ &= P(n|\overline{\mathbf{A}_c} \cup \underline{\mathbf{A}_i} \cup \overline{\mathbf{A}_e}) \leq \epsilon \end{aligned}$$

where ϵ is a small probability. $A_{j+1}, \ldots, A_k$ are set to constitute an amplifier with respect to the effect of $A_1, \ldots, A_j$ on N if:

$$\begin{aligned} P(n|\mathbf{A}_c \cup \overline{\mathbf{A}_i} \cup \mathbf{A}_e) &= P(n|\mathbf{A}_c \cup \overline{\overline{\mathbf{A}_i}} \cup \mathbf{A}_e) \leq P(n|\mathbf{A}_c \cup \underline{\mathbf{A}_i} \cup \mathbf{A}_e) \\ P(n|\overline{\mathbf{A}_c} \cup \overline{\mathbf{A}_i} \cup \overline{\mathbf{A}_e}) &= P(n|\overline{\mathbf{A}_c} \cup \overline{\overline{\mathbf{A}_i}} \cup \overline{\mathbf{A}_e}) \\ &= P(n|\overline{\mathbf{A}_c} \cup \underline{\mathbf{A}_i} \cup \overline{\mathbf{A}_e}) \leq \epsilon \end{aligned}$$

where ϵ is a small probability.

In proposing ADs from BNs, it important to be able to identify inhibitors and amplifiers. If an argument is based on the premise that the consequence of a first-order effect is true, then a possible inhibitor to the effect adds nothing to the argument. Even if the second-order effect of the inhibitor was a factor, it did not change the outcome in this case. Indeed, this is the reason why, in the sample AD of Figure 2, s or $\overline{s}$ need not be included. Similarly, if an argument is based on the premise that the consequence of a first-order effect is false, then a possible amplifier adds nothing to the argument and need not be included in the proposed AD.

Conversely, if an argument is based on the premise that the consequence of a first-order effect is false (true), then an inhibitor (amplifier) can justify that premise, and needs to be incorporated in the AD. Given that such second-order effects are used as justifications in such situations, the causal direction of the BN will be maintained in the proposed ADs.

3.2.2 Other influences

Influences to nodes N on a directed path from hypothesis to evidence in a BN that are not second-order to influences on such paths need to be treated differently. in general, such influences may constitute alternative plausible explanations for the proposition derived from N and need to be treated accordingly in an AD. This implies that they should be represented as consequences of the proposition derived from N in the AD.

It is interesting to note that, as alternative explanations, these alternative explanation provide hooks for the synthesis of counterarguments. As the latter is beyond the scope of this paper, the proposed algorithm will merely include these alternative explanations in the synthesised ADs. Their effectiveness as potential counterarguments can be assessed again by analysing the CPTs. Such alternative explanations may be mutually independent, or to a larger or smaller extent mutually exclusive or synergetic [24]. If they are mutually independent, the likelihood of one explanation does not affect the probability the other. If they are partially mutually exclusive, evidence supporting one explanation undermines the support for the other. Conversely, if they are partially mutually synergetic, evidence supporting one also enhances support for the other. How this idea can be formalised into a criteria for the content of a CPT is to be examined in future work.

3.2.3 Convergent and linked arguments

Argument diagrams and Bayesian networks possess fundamentally different ways of denoting the distinct ways in which two sets of antecedents affect a consequent. Argument diagramming techniques often seek to distinguish between so-called *linked* and *convergent* arguments [21]. Two propositions are said to be convergent arguments for a conclusion if both support the conclusion independently, whereas they are said to be linked arguments if the strength of support of one depends on the truth of the other.

It can be quite difficult to classify real-world arguments into the convergent or linked category [6, 23]. As explained in 2.1, a BN models the support of one or more premises for a conclusion by means of CPTs. These CPTs, therefore, are to inform the degree to which the support of an argument for a conclusion is independent from other arguments. Let a_1, a_2 and c be propositions in an AD derived from a BN, such that a_1 supports c and a_2 supports c. The support of a_1 for c in isolation of the a_2 argument equals $P(c|a_1, \overline{a_2})$. Similarly, the support of a_2 for c in isolation of a_1 is $P(c|\overline{a_1}, a_2)$. If these effects are independent from one another, then the combined effect (i.e. $P(c|a_1, a_2)$) must equal:

$$P(c|a_1, \overline{a_2}) + P(c|\overline{a_1}, a_2) - P(c|a_1, \overline{a_2}) \times P(c|\overline{a_1}, a_2) \quad (3)$$

This criterion is similar to the one obtained by Yanal [26]. However, Yanal represents the support of a_1 for c as $P(c|a_1)$, which does consider the effect of a_2 since $P(c|a_1) = P(c|a_1, \overline{a_2})P(\overline{a_2}) + P(c|a_1, a_2)P(a_2)$.

As explained in 3.2 and 4.2, the propositions a_1, a_2 and c stem from variables A_1, A_2 and C respectively, of a BN in which C is a parent variable to A_1 or A_2 or both. Therefore, the conditional probabilities of the criterion specified by (3) must be derived from CPTs expressing $P(A_1|A_2, C)$, $P(A_2|A_1, C)$ or $P(A_1|C)$ and $P(A_2|C)$, using Bayes' law. The result is, therefore, affected by prior probabilities, in addition to the aforementioned CPTs[1]. This raises two questions, beyond the scope of this paper, but interesting as further research. Firstly, does the classification of arguments into linked and convergent ones, based on identification of (near) matches of criterion (3), meet with the expectations of an expert designing the BN? Secondly, how should the identification of linked arguments, and therefore the applicability of (3), constrain CPTs and/or priors?

3.3 Qualifiers

BNs and ADs both possess a means to represent the strength of links between vertices. BNs employ a precise numeric calculus to describe how the probability of a node is affected by knowledge of its parent variables. ADs employ qualitative schemes to describe the probative force of inferences. As such, the proposed argument extraction algorithm must convert the information contained in the CPTs into a qualitative representation that categorises the information contained in the CPTs into a notion of probative force.

Let A and N be variables of a BN with an influence $A \rightarrow N$, a and n be propositions derived from A and N respectively, and $n \rightarrow a$ be an inference of the AD derived from the BN. According to (2), the strength of support of n for a, as opposed to $\neg q$, is expressed by:

$$LR = \frac{P(n|a)}{P(n|\neg a)}$$

Let $A_1, \ldots, A_j$ be the antecedents of a second-order influence to $A \rightarrow N$ in the BN. The conditional probabilities in the likelihood ratio need to be constrained by the propositions $a_1, \ldots, a_j$ to be derived from $A_1, \ldots, A_j$. That is:

$$LR = \frac{P(n|a, a_1, \ldots, a_j)}{P(n|\neg a, a_1, \ldots, a_j)} \quad (4)$$

If the AD incorporates the second-order influence, $a_1, \ldots, a_j$ are chosen such that it affects its first-order influence. If the AD does not contain the second-order influence, $a_1, \ldots, a_j$ are chosen such that it is inactive.

Let $A_{j+1}, \ldots, A_k$ be further antecedents to N in the BN. Then, the outcome of the calculation of the likelihood ratio associated with an inference in the AD will depend on the values that are assigned to $A_{j+1}, \ldots, A_k$. A number of different approaches can be employed to deal with this. For example, (4) could be calculated as

$$\frac{\sum_{a_{j+1},\ldots,a_k} P(n|a, a_1, \ldots, a_j, a_{j+1}, \ldots, a_k)P(a_{j+1}) \ldots P(a_k)}{\sum_{a_{j+1},\ldots,a_k} P(n|\neg a, a_1, \ldots, a_j, a_{j+1}, \ldots, a_k)P(a_{j+1}) \ldots P(a_k)}$$

where $a_{j+1} \in \mathbf{D}_{A_{j+1}}, \ldots, a_k \in \mathbf{D}_{A_k}$ and the $P(a_i)$ can be computed by the BN. There are two problems with this approach. Firstly, $P(a_i)$ is affected by prior probabilities. Secondly, the calculation averages values conditional probabilities under different circumstances. This defeats the purpose of using the AD to scrutinise the individual inferences for potential weaknesses. Therefore, the approach proposed herein will be to calculate a range of likelihood ratios:

$$[\underline{LR}, \overline{LR}] \quad (5)$$

where

$$\underline{LR} = \min_{a_{j+1},\ldots,a_k} \frac{P(n|a, a_1, \ldots, a_j, a_{j+1}, \ldots, a_k)}{P(n|\neg a, a_1, \ldots, a_j, a_{j+1}, \ldots, a_k)}$$

$$\overline{LR} = \max_{a_{j+1},\ldots,a_k} \frac{P(n|a, a_1, \ldots, a_j, a_{j+1}, \ldots, a_k)}{P(n|\neg a, a_1, \ldots, a_j, a_{j+1}, \ldots, a_k)}$$

where $a_{j+1} \in \mathbf{D}_{A_{j+1}}, \ldots, a_k \in \mathbf{D}_{A_k}$.

[1] For example, to derive $P(c|a_1, a_2)$ from CPTs expressing $P(A_1|C)$ and $P(A_2|C)$, Bayes' law is applied as follows:

$$P(c|a_1, a_2) = \frac{P(a_1|c)P(a_2|c)P(c)}{P(a_1|c)P(a_2|c)P(c) + P(a_1|\overline{c})P(a_2|\overline{c})P(\overline{c})}$$

The values for $P(a_i|c)$ and $P(a_i|\overline{c})$ are given by the CPTs for $P(A_i|C)$. However, the calculation of $P(c)$ and $P(\overline{c})$ relies on prior probabilities.

LR	Description
<1	not impossible
1	plausible
>1 to <2	tenuous
2 to <5	weak
5 to <10	limited
10 to <100	moderate
100 to <1000	moderately strong
1000 to <10000	strong
10000 and <∞	very strong
∞	certainly

Table 2: Qualitative descriptions of probative force derived from likelihood ratios (based on [10])

Finally, the likelihood ratios need to be translated into verbal descriptions of probative force. A, relatively conservative, conversion table derived from [10] (and extended to refine qualifiers assigned to low likelihood ratios) is shown in Table 2.

4. ARGUMENT EXTRACTION

4.1 Inputs, output and assumptions

The basic argument extraction algorithm takes as input a Bayesian network, a hypothesis corresponding and a set of pieces of evidence. Let us denote the hypothesis variable H, a value assignment for H corresponding to the ultimate probandum, the set of evidence variables $\mathbf{O}$ (for observations) and their value assignments, and a value for ϵ (i.e. what constitutes a low probability). The output of the algorithm is an argumentation diagram that explains how the evidence supports the chosen hypothesis, including indications of probative force for the inferences.

As explained in Section 3, the algorithm assumes that the variables are boolean. It is also assumed that the influences in the BN are either causal or definitions in nature. Without loss of generality, a value of $\epsilon = 0.01$ is assumed and the likelihood ratio translation scheme of Table 2 is adopted.

4.2 Outline algorithm

Step 1: Initialisation of the graph

The algorithm begins by constructing an initial structure for the AD. This initial construct will consist of the minimum elements that are required to capture the ways in which the evidence support the hypothesis according to the knowledge contained in the BN $\langle \mathbf{V}_b, \mathbf{E}_b, P \rangle$. Put simply, this initial construct describes the way the evidence supports the hypothesis in an evidence/observation to hypothesis/probandum direction rather than a causal direction. It also excludes vertices and edges that are not on a path between hypothesis and evidence, as well as edges that are implied through transitivity. The next steps will add the vertices and edges that were ignored at this stage if they are deemed to possess a qualitatively distinct meaning in the BN rather than merely alter conditional probability distributions in subtle ways.

More formally, the initial construct is a directed acyclic graph $AD = (\mathbf{V}_a, \mathbf{E}_a)$, in which the vertices in $\mathbf{V}_a$ correspond to propositions and the edges in $\mathbf{E}_a$ to arguments of the AD. Each vertex in the AD will eventually contain a variable from the BN and a value from that variable's domain. In the initial construct, the vertices that correspond to the hypothesis and to the evidence nodes are assigned both a variable and a domain value. All the other vertices will be assigned a value in a later step. The initialisation algorithm is as follows:

- Given that the hypothesis is $H : v$, where H is the hypothesis variable in the BN and v the value it is assigned (e.g. true or false), add a node with variable H and value v to $\mathbf{V}_a$.
- For each piece of evidence $E : v$, where E is the evidence variable in the BN and v the value it is assigned (e.g. true or false), add a node with variable E and value v to $\mathbf{V}_a$.
- For each variable $V \in \mathbf{V}_b$ that is on a path $H \rightarrow \ldots \rightarrow V \rightarrow \ldots \rightarrow E_i$ in the DAG $(\mathbf{V}_b, \mathbf{E}_b)$ defined by the BN, where E_i is one of the pieces of evidence, add a node containing variable V to $\mathbf{V}_a$.
- For each edge $V_1 \rightarrow V_2 \in \mathbf{E}_b$, such that $(\mathbf{V}_b, \mathbf{E}_b)$ does not define a longer path $V_1 \rightarrow \ldots \rightarrow V_3 \rightarrow \ldots \rightarrow V_2$, with $V_3 \neq V_1, V_2$, add an edge $V_2 \rightarrow V_1$ to $\mathbf{E}_a$.

Step 2: Initialisation of the propositions

Because an AD is concerned with propositions rather than variables, values must be associated with each of the nodes in the emerging AD. More precisely, at least the variables currently in the set $\mathbf{V}_a - (\{H\} \cup \mathbf{O})$ should be assigned values. A number of different schemes can be devised to accomplish this assignment and software designed to support the developer of BNs with arguments generated from BNs should allow the user to overrule any set of automatically generated value assignments. The only constraint the initial value assignments is the requirement that the combination of value assignments, including the hypothesis and observations, are possible according the probability distribution defined by the BN.

Arguably, in the absence of any further requirements, the most suitable set of value assignments is the most probable combination of value assignments given the hypothesis to be argued for and the evidence. This is an instance of the problem of finding the *most probable explanation* (MPE) of a BN [15]. Let $\mathbf{X} = \mathbf{V}_a - (\{H\} \cup \mathbf{O}) = \{X_1, \ldots, X_n\}$ and let $\mathbf{D}_i$ denote the domain of X_i, then the MPE we require equals:

$$\max_{x_1 \in \mathbf{D}_1, \ldots, x_n \in \mathbf{D}_n} P(X_1 : x_1, \ldots, X_n : x_n | \{H\} \cup \mathbf{O}) \qquad (6)$$

This problem can be solved by a number of standard algorithms developed by Shimony [19] and Suermondt [20].

Step 3: Extend

Let a precedence-ordered queue $\mathbf{Q}_\mathbf{E}$ be an ordered set of vertices, such that for any pair of nodes N_1 and N_2 where $\mathbf{E}$ defines a path $N_1 \rightarrow \cdots \rightarrow N_2$, N_1 precedes N_2 in $\mathbf{Q}_\mathbf{E}$. Let *enqueue*$(\mathbf{Q}_\mathbf{E}, N)$ adds the node N to the precedence-ordered queue $\mathbf{Q}_\mathbf{E}$. Let *dequeue*$(\mathbf{Q}_\mathbf{E})$ be an operation that removes and returns the first node N from the queue (i.e. such that there is no node in $\mathbf{Q}_\mathbf{E}$ that precedes N). The algorithm below employs a queue $\mathbf{Q}_{\mathbf{E}_a}$ that contains vertices taken from $\mathbf{V}_a$. Because $(\mathbf{V}_a, \mathbf{E}_a)$ is a DAG, a precedence ordered queue that defines a partial order over its elements exists.

- Add all the nodes of $\mathbf{V}_a$ to the precedence-ordered queue. That is, for each $N \in \mathbf{V}_a - \{H\}$, *enqueue*$(\mathbf{Q}_{\mathbf{E}_a}, N)$.
- While $\mathbf{Q}_{\mathbf{E}_a}$ is not empty:
    - Let $N =$ *dequeue*$(\mathbf{Q}_\mathbf{E})$.

  - Apply *Step 3*: *process*(N).

Step 4: *process*(N)

- Let $\mathbf{A} = \{A_1, \ldots, A_i, A_{i+1}, \ldots, A_j\}$ be the set of nodes of nodes such that for each $A \in \mathbf{A}$, there exists an edge $A \rightarrow N \in \mathbf{E}_b$. Without loss of generality, it is assumed that $\mathbf{A}$ can be partitioned into the set $\{A_1, \ldots, A_i\}$ of nodes of $\mathbf{A}$ already in $\mathbf{V}_a$ and the set $\{A_{i+1}, \ldots, A_j\}$ of nodes of $\mathbf{A}$ currently not in $\mathbf{V}_a$.
- For each $A_k \in \{A_{i+1}, \ldots, A_j\}$:
  - If N is assigned the value true and a subset of variables of $\{A_{i+1}, \ldots, A_j\}$ constitute an inhibitor or a disabler with regards to the effect of $A_1, \ldots, A_i$ on N, then ignore that subset of variables. If N is assigned the value false and a subset of variables of $\{A_{i+1}, \ldots, A_j\}$ constitute an amplifier or an enabler with regards to the effect of $A_1, \ldots, A_i$ on N, then ignore that subset of variables.
  - If N is assigned the value false and a subset of variables of $\{A_{i+1}, \ldots, A_j\}$ constitute an inhibitor or a disabler with regards to the effect of $A_1, \ldots, A_i$ on N, or if N is assigned the value true and a subset of variables of $\{A_{i+1}, \ldots, A_j\}$ constitute an amplifier or an enabler with regards to the effect of $A_1, \ldots, A_i$ on N, then:
    - Let $\mathbf{A}$ be a set of variables containing $\{A_{i+1}, \ldots, A_j\}$ and all their ancestor variables in the BN. Assign the variables in $\mathbf{A}$ their MPE, as defined in (6), given the assignments already made. Add each variable and corresponding assignment to $\mathbf{V}_a$.
    - For each edge $A_1 \rightarrow A_2 \in \mathbf{E}_b$ such that $A_1, A_2 \in \mathbf{A}$ add an edge $A_1 \rightarrow A_2$ to $\mathbf{E}_a$.
    - For each $A \in \mathbf{A}$, *enqueue*$(\mathbf{Q}_{\mathbf{E}_a}, A)$.
  - For each remaining $A_k \in \{A_{i+1}, \ldots, A_j\}$ that was not considered previously:
    - Let $\mathbf{A}$ be a set of variables A_k and its their descendent variables in the BN. Assign the variables in $\mathbf{A}$ their MPE, as defined in (6), given the assignments already made. Add each variable and corresponding assignment to $\mathbf{V}_a$.
    - For each edge $A_1 \rightarrow A_2 \in \mathbf{E}_b$ such that $A_1, A_2 \in \mathbf{A}$ add an edge $A_2 \rightarrow A_1$ to $\mathbf{E}_a$.
    - For each $A \in \mathbf{A}$, *enqueue*$(\mathbf{Q}_{\mathbf{E}_a}, A)$.
  - Calculate the qualifiers for the inferences departing from N in the emerging AD by means of (5).

This algorithm is best explained further by means of some examples. This is covered by the next Section.

5. RESULTS

5.1 Two-way transfer evidence

5.1.1 The Bayesian network

Aitken et. al. have developed a BN for the analysis of two-way transfer evidence. The scenario to which this BN is applicable is one where there a violent exchange between two people that leads to the death of one of them, the victim. A suspect is arrested later and blood splatter is retrieved from both the victim and the suspect, and analysed with a view to determine whether the suspect carries blood traces matching the victim's (RM_s) and the victim traces matching the suspect's (RM_v). This evidence is to be related to a hypothesis indicating whether the suspect is guilty of killing the victim (G) and a related proposition that there was violent contact between suspect and victim (C). The DAG of this BN is shown in Figure 4, the meaning of the symbols is explained in Table 3 and the conditional probability tables of the BN are shown in Table 4.

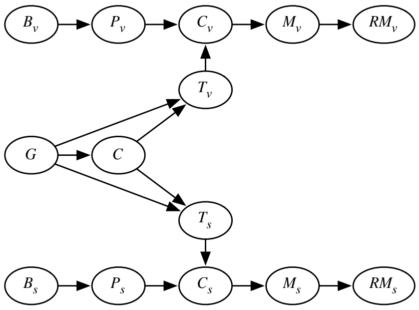

Figure 4: DAG of the two way transfer BN [1]

Variable	Meaning
G	The suspect is guilty
C	The suspect and victim had violent contact
T_v	Blood traces transferred from suspect to victim
B_v	Victim comes into contact with blood
P_v	Blood traces exist on victim unrelated to crime
C_v	Investigator selects a blood trace from the victim that is related to the crime
M_v	Blood trace retrieved from victim matches suspect
RM_v	Reported match of blood trace retrieved from victim
T_s	Blood traces transferred from victim to suspect
B_s	Suspect comes into contact with blood
P_s	Blood traces exist on suspect unrelated to crime
C_s	Investigator selects a blood trace from the suspect that is related to the crime
M_s	Blood trace retrieved from suspect matches victim
RM_s	Reported match of blood trace retrieved from suspect

Table 3: Variables of the two way transfer BN [1]

In the BN developed by Aitken et. al., transfer of trace material from suspect to victim (T_v) and from victim to suspect (T_s) is much more likely under the hypothesis that the suspect is guilty than under the hypothesis that the suspect is not guilty, and these transfers are only possible if there was violent contact between suspect and victim. If there was a transfer of trace material between suspect and victim, then it is possible that traces related to the incident are found on the victim (C_v) and on the suspect (C_s). If, however, there is another potential source of the same type of trace material for the victim/suspect (P_v/P_s), then it becomes harder to find such trace material. For that reason, $P(c_i|t_i, p_i) = 0.3$ whereas $P(c_i|t_i, \overline{p_i}) = 1$. The presence of trace material from another source on the victim/suspect (P_v/P_s) depends on the victim's/suspect's background (B_v/B_s). If transfer related trace material is retrieved from the victim/suspect, it is likely to match the suspect's/victim's (M_v/M_s), though a Type I error (false positive) of 0.0001 is assumed. As $P(\overline{m_i}|c_i) = 0$, a Type II error is deemed

	g	$\overline{g}$
$P(c\|G)$	1	0.01
$P(\overline{c}\|G)$	0	0.99

	g		$\overline{g}$	
	c	$\overline{c}$	c	$\overline{c}$
$P(t_i\|G,C)$	0.95	0	0.095	0
$P(\overline{t_i}\|G,C)$	0.05	1	0.905	1

	b_i	$\overline{b_i}$
$P(p_i\|B_i)$	1	0
$P(\overline{p_i}\|B_i)$	0	1

	t_i		$\overline{t_i}$	
	p_i	$\overline{p_i}$	p_i	$\overline{p_i}$
$P(c_i\|T_i,P_i)$	0.3	1	0	0
$P(\overline{c_i}\|T_i,P_i)$	0.7	0	1	1

	c_i	$\overline{c_i}$
$P(m_i\|C_i)$	1	0.0001
$P(\overline{m_i}\|C_i)$	0	0.9999

	m_i	$\overline{m_i}$
$P(rm_i\|M_i)$	1	0.001
$P(\overline{rm_i}\|M_i)$	0	0.999

Table 4: Conditional probability tables of the two-way transfer network (initial values proposed by Aitken et. al. [1])

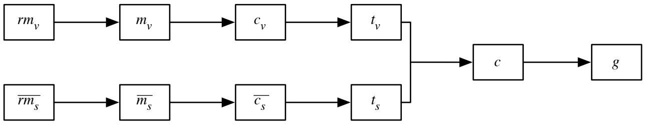

Figure 5: Intermediate argumentation diagram of a two-way transfer case

to be impossible. In case of such a match, evidence is likely to be reported accordingly (RM_v/RM_s), with a Type I error of 0.001 and Type II error of 0.

5.1.2 *Argumentation diagram*

To propose an AD based on this BN, hypothesis and evidence propositions must be given. To illustrate the algorithm, consider a case where blood splatter found on the victim has been matched to the suspect (rm_v) and blood splatter found on the suspect has not been matched to the victim ($\overline{rm_s}$). Based on this, the algorithm is employed to produce an AD that argues in favour of the guilt of the suspect. As such, the hypothesis under consideration is g.

In step 1 of the algorithm, the AD is initialised with the evidence and hypothesis propositions as well as all the nodes on paths between evidence and hypothesis. The direction of the directed paths from hypothesis to evidence in the BN are inverted in the AD. As such, step 1 of the algorithm results in a graph containing two directed paths: $rm_v \rightarrow M_v \rightarrow C_v \rightarrow T_v \rightarrow C \rightarrow g$ and $\overline{rm_s} \rightarrow M_s \rightarrow C_s \rightarrow T_s \rightarrow C \rightarrow g$.

Step 2 of the algorithm generates propositions for the variables M_v, M_s, C_v, C_s, T_v, T_s and C. The most probable explanation of these variables, given g, rm_v and $\overline{rm_s}$ is m_v, $\overline{m_s}$, c_v, $\overline{c_s}$, t_v, t_s and c. More precisely, the most probable explanation for the evidence rm_v in support of the hypothesis is the argument that the trace evidence found on the victim (rm_v) is the result of blood splatter matched to the suspect (m_v) that was found on the victim (c_v) and the result from transfer of blood from suspect to victim (t_s) as a consequence of violent contact (c). The most probable explanation for the evidence $\overline{rm_s}$ in support of the hypothesis is the argument that the trace evidence could not be matched to the victim ($\overline{m_s}$) and was not crime related ($\overline{c_s}$). The latter leaves open the possibility that blood was transferred from victim to suspect (t_s) as a consequence of violent contact (c). Figure 5 shows the resulting emerging AD.

In the BN, $P_v \rightarrow C_v$ is an inhibitor to $T_v \rightarrow C_v$ and $P_s \rightarrow C_s$ is an inhibitor to $T_s \rightarrow C_s$. Because C_v is assigned true, $P_v \rightarrow C_v$ is irrelevant to the argument. Because C_s is assigned false, $P_s \rightarrow C_s$ is relevant to the argument. Thus, following step 4 of the algorithm, a directed path $B_s \rightarrow P_s \rightarrow \overline{c_s}$ is added to the emerging AD. The most probable explanations for B_s and P_s are b_s and p_s respectively (i.e. the inhibitor is active). This version of the emerging AD is depicted in Figure 6.

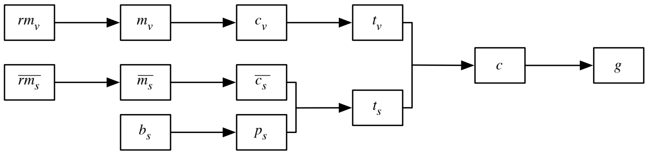

Figure 6: Intermediate argumentation diagram of a two-way transfer case

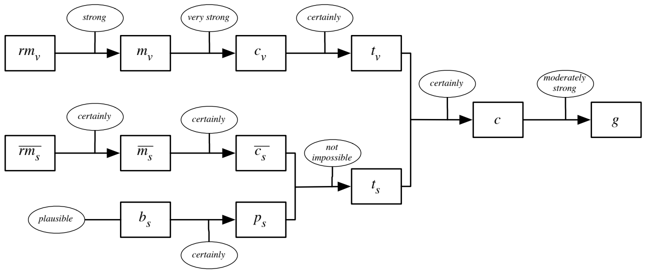

Figure 7: Outline argumentation diagram of a two-way transfer case

In the process of generating the graph, qualifiers expressing probative force are added. For example, because

$$\frac{P(rm_v|m_v)}{P(rm_v|\overline{m_v})} = \frac{1}{0.001} = 1000 \simeq \text{strong}$$

the inference $rm_v \rightarrow m_v$ is annotated with the qualifier "strong".

The AD generated by following the algorithm is shown in Figure 7. Arguably, the explanation provided by the AD is largely uncontroversial. The diagram contains one unconvincing inference: $\overline{c_s} \wedge p_s \rightarrow t_s$ or choosing a non-crime related bloodstain on the suspect and the presence of bloodstains on the suspect from sources other than the crime imply a transfer of blood from victim to suspect. In this case, the qualifier provides a helpful elaboration. Because

$$\frac{P(\overline{c_s}|t_s,p_s)}{P(\overline{c_s}|\overline{t_s},p_s)} = \frac{0.7}{1} = 0.7 \simeq \text{not impossible}$$

As such, the AD states that because bloodstains were present on the suspect from sources other than the crime and non-crime related bloodstain on the suspect were chosen for analysis, it is "not impossible" that blood was transferred from victim to suspect. In this way, the qualifier identifies the tenuous nature of this problematic inference.

5.2 Terpenes traces in fire incidents

5.2.1 *The Bayesian network*

Biedermann et. al. have developed a set of BNs for the analysis of traces of certain flammable materials in the forensic investigation of fire incidents [3]. One representative sample for evaluating the argument extraction algorithm: a BN designed to analyse traces of

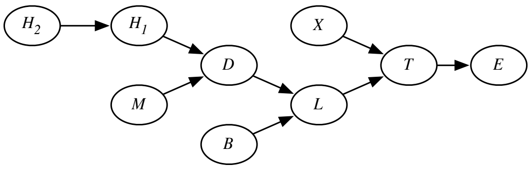

Figure 8: DAG of the fire incident BN [3]

Variable	Meaning
H_2	The fire was started by human action
H_1	The fire was started by human action, using an terpenes containing accelerant
M	The sampling point was located at origin of the fire
D	Combustable liquid with terpenes was spilled
B	Background presence of terpenes
L	Presence of terpenes from source other than building itself
X	Building constructed from materials with terpenes
T	Presence of terpenes at sampling point
E	Detection of terpenes in fire debris sample

Table 5: Variables of the fire-incident BN [3]

terpenes in fire debris. The DAG of the BN is shown in Figure 8, the meaning of the symbols is explained in Table 5 and the conditional probability tables of the BN are shown in Table 6.

The BN consists of two hypothesis variables, indicating whether the fire incident was the result of human action, and one evidence variable (E) indicating the identification of terpenes traces in fire debris. The first hypothesis (H_1) is that the fire was started by human action by means of an accelerant containing terpenes. If this hypothesis is true, and the origin of the fire was sampled (M), then that explains that a liquid containing terpenes was spilled (D). The second hypothesis (H_2) is that the fire was started by human action. If H_2 is true, there is a small probability of 0.02 that the fire was started by human action using a terpenes based accelerant ($P(h_1|h_2) = 0.02$). $P(h_1|h_2)$, which is rather low, is based on the composition and distribution of different types of accelerant.

Variable D is one possible source of terpenes other than the building itself (L). Another is storage of materials that have been contaminated with terpenes (B). Terpenes may also be contained in the building itself as certain types of wood contain the substance (X). Both L and X can account for the presence of terpenes at the sampling point (T). Generally speaking, there the presence of terpenes will lead to the identification of terpenes in the fire debris, but the model suggests a 0.001 chance of Type I error (false positive) and a 0.01 chance of Type II error (false negative).

5.2.2 *Arguing for the specific hypothesis (h_1)*

Consider a case where terpenes traces has been discovered near the source of the fire (e) and the specific hypothesis that a fire was started by human action using a terpenes containing accelerant (h_1). An AD can be generated by means of the algorithm in the same was as in the two-way transfer example. The result is shown in Figure 9. This diagram has one distinguishing feature compared to the two-way transfer case. Here, two alternative explanations are provided that are equally plausible to elements of the central argument. The evidence may also be the result of the use of terpenes containing wood in the construction of the building (x) and of a background presence of terpenes (b).

	h_2	$\overline{h_2}$
$P(h_1\|H_2)$	0.02	0
$P(\overline{h_1}\|H_2)$	0.98	1

	h		$\overline{h}$	
	m	$\overline{m}$	m	$\overline{m}$
$P(d\|H,M)$	1	0	0	0
$P(\overline{d}\|H,M)$	0	1	1	1

	t	$\overline{t}$
$P(e\|T)$	0.99	0.001
$P(\overline{e}\|T)$	0.01	0.999

	d		$\overline{d}$	
	b	$\overline{b}$	b	$\overline{b}$
$P(l\|D,B)$	1	1	1	0
$P(\overline{l}\|D,B)$	0	0	0	1

	l		$\overline{l}$	
	x	$\overline{x}$	x	$\overline{x}$
$P(t\|L,X)$	1	1	1	0
$P(\overline{t}\|L,X)$	0	0	0	1

Table 6: Conditional probability tables of the fire incident network (initial values proposed by Biedermann et. al. [3])

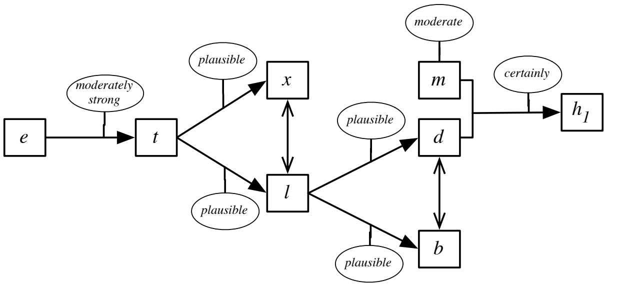

Figure 9: Outline argumentation diagram supporting the specific hypothesis

5.2.3 *Arguing for the general hypothesis (h_2)*

Consider a case where terpenes traces has been discovered near the source of the fire (e) and the generic hypothesis that a fire was started by human action (h_2). The result is shown in Figure 10.

In this case, there are three key differences compared to the AD of Figure 9. Firstly, some of the propositions are different because the MPE need not commit to hypothesis h_1. As a result, the more likely explanation that includes $\overline{l}$, $\overline{d}$ and $\overline{h_1}$ is chosen. Secondly, because this AD commits to proposition $\overline{h_1}$, the enabler $M \rightarrow D$ to $H_1 \rightarrow D$ is not relevant. Neither of these these differences is controversial. Thirdly, the inference $\overline{h_1} \rightarrow h_2$ is added. This inference is dubious, as is recognised by the automatically chosen qualifier "not impossible". The problem with this case is the lack of a complete set of partitioning hypothesis for h_2. In other words, there is not a complete set of specific hypothesis to justify h_2. Instead,

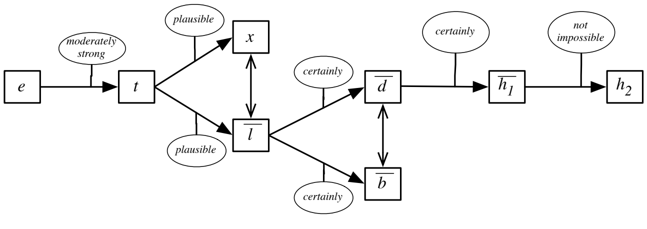

Figure 10: Outline argumentation diagram supporting the generic hypothesis

there is only one specific hypothesis h_1 supporting the general hypothesis h_2 and it is relatively unlikely (i.e. $P(h_1|h_2) = 0.02$. This constitutes a possible fault model for the BN that an AD can explain. Future work should examine this issue further.

6. CONCLUSIONS AND FUTURE WORK

This paper has introduced an approach to propose argument diagrams (ADs) based on Bayesian networks (BNs). Based on an analysis of similarities and differences of the representation formalisms of ADs and BNs, the component translations that such an approach requires have been identified, and initial solutions to these challenges have been proposed. These include identification of the scope of the AD (i.e. the part of BN to be converted to an AD), extraction of AD propositions from BN variables and domains, proposing an appropriate direction of inferences between propositions, distinguishing between linked and convergent arguments and assessing the strength of inferential support between inferences. The limits on the kinds of information that are normally a part of ADs and that can be extract from BNs (not enhanced with additional information have been explored. The approach has been evaluated by means of BNs from the Forensic Science literature.

The approach has been developed with a view to support the development of BNs for evidential reasoning. Therefore, a key objective of future research is to incorporate into software for evidential reasoning BN design. In such software, the algorithm would provide the basis of an interactive tool to scrutinise BN under development by testing it against test cases, for which argument diagrams are to synthesised and evaluated. The algorithm relies on a number of simplifying assumptions that are to be relaxed in future work. In particular, the present version of the approach relies on boolean variables, which is a significant limitation. In general, BNs may employ significantly larger nominal or ordered domains. To relax the assumption of boolean variables, the approach needs to be extended with a means to produce suitable propositions. The work also assumes that the influences in the BN are causal of definitional in nature. Although this does not appear to be a significant limitation as influences in BNs then to be causal or definitional, future work may seek to address this. While the paper has presented a method to synthesise arguments, it has not produced a corresponding means to produce counterarguments. The latter would likely provide a suitable addition to an evidential reasoning BN design tool as it would facilitate more elaborate case-based scrutiny.

7. REFERENCES

[1] C. Aitken, F. Taroni, and P. Garbolino. A graphical model for the evaluation of cross-transfer evidence in dna profiles. *Theoretical Population Biology*, 63:179–190, 2003.

[2] F. Bex, P. van Koppen, H. Prakken, and B. Verheij. A hybrid formal theory of arguments, stories and criminal evidence. *Artificial Intelligence and Law*, 18(2):123–152, 2010.

[3] A. Biedermann, F. Taroni, O. Delemont, C. Semadeni, and A. Davison. The evaluation of evidence in the forensic investigation of fire incidents. part ii. practical examples of the use of bayesian networks. *Forensic Science International*, 147:59–69, 2005.

[4] J. Buckleton, C. Triggs, and C. Champod. An extended likelihood ratio framework for interpreting evidence. *Science and Justice*, 46(2):69–78, 2006.

[5] P. Condliffe, B. Abrahams, and J. Zeleznikow. An OWL ontology and bayesian network to suport legal reasoning in the owners corporation domain. In *Proceedings of the 6th International Workshop on Online Dispute Resolution*, pages 51–62, 2010.

[6] D. Conway. On the distinction between convergent and linked arguments. *Informal Logic*, 13(3):145–158, 1991.

[7] R. Cook, I. Evett, G. Jackson, P. Jones, and J. Lambert. A model for case assessment and interpretation. *Science and Justice*, 38(6):151–156, 1998.

[8] D. Corfield and J. Williamson. *Foundations of Bayesianism*. Springer, 2001.

[9] G. Davis. Bayesian reconstruction of traffic accidents. *Law, Probability and Risk*, 2:69–89, 2003.

[10] I. Evett, G. Jackson, J. Lambert, and S. McCrossan. The impact of the principles of evidence interpretation on the structure and content of statements. *Science and Justice*, 40(4):233–239, 2000.

[11] M. Grabmair, T. Gordon, and D. Walton. Probabilistic semantics for the carneades argument model using bayesian networks. In *Proceedings of the International Conference on Computational Models of Argument*, pages 255–266. IOS Press, 2010.

[12] A. Hepler, P. Dawid, and V. Leucari. Object-oriented graphical representations of complex patterns of evidence. *Law, Probability and Risk*, 6(1–4):275–293, 2007.

[13] J. Keppens. Towards qualitative approaches to bayesian evidential reasoning. In *Proceedings of the 11th International Conference on Artificial Intelligence and Law*, pages 17–25, 2007.

[14] J. Keppens, Q. Shen, and C. Price. Compositional bayesian modelling for computation of evidence collection strategies. *Applied Intelligence*, pages DOI: 10.1007/s10489–009–0208–5, 2010.

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

[16] J. Mortera, A. Dawid, and S. Lauritzen. Probabilistic expert systems for dna mixture profiling. *Theoretical Population Biology*, 63:191–205, 2003.

[17] C. Reed, D. Walton, and F. Macagno. Argument diagramming in logic, law and artificial intelligence. *Knowledge Engineering Review*, 22:87–109, 2007.

[18] D. Schum. *The Evidential Foundations of Probabilistic Reasoning*. Northwestern University Press, 1994.

[19] S. Shimony. *A probabilistic framework for explanation*. PhD thesis, Brown University, Department of Computer Science, 1991.

[20] H. Suermondt. *Explanation in Bayesian belief networks*. PhD thesis, Stanford University, Department of Computer Science, 1992.

[21] S. Thomas. *Practical Reasoning in Natural Language*. Prentice-Hall, 1986.

[22] S. Toulmin. *The Uses of Argument*. Cambridge University Press, 1958.

[23] D. Walton. *Argumentation methods for artificial intelligence in law*. Springer, 2005.

[24] M. Wellman and M. Henrion. Explaining "explaining away". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15:287–291, 1993.

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

[26] R. Yanal. Dependent and independent reasons. *Informal Logic*, 13(3):137–144, 1991.