

Graphical Explanation in Belief Networks

David Madigan¹

University of Washington and Fred Hutchinson Cancer Research Center

Krzysztof Mosurski

Trinity College Dublin, Ireland

Russell G. Almond

Educational Testing Services

May 8, 1996

Abstract

Belief networks provide an important bridge between statistical modeling and expert systems. In this paper we present methods for visualizing probabilistic “evidence flows” in belief networks, thereby enabling belief networks to explain their behavior. Building on earlier research on explanation in expert systems, we present a hierarchy of explanations, ranging from simple colorings to detailed displays. Our approach complements parallel work on textual explanations in belief networks. GRAPHICAL-BELIEF, Mathsoft Inc.’s belief network software, implements the methods.

1 Introduction

A fundamental reason for building a mathematical or statistical model is to foster deeper understanding of complex, real-world systems. Consequently, explanations—descriptions of the mechanisms which comprise such models—form an important part of model validation, exploration, and use.

Early tests of rule-based expert system models indicated the critical need for detailed explanations in that setting (Buchanan and Shortliffe, 1984; Barr and Feigenbaum, 1982; Coyne, 1990; Chandrasakaran, *et al.*, 1989). Users of these early systems found that understanding *why* the system had reached a particular conclusion or decision was as important as reaching the decision. Swartout (1983) comments that “trust in a system is developed not only by the quality of the results but also by clear description of how they were derived... In addition to providing diagnoses or prescriptions, a consultant program must be able to explain what it is doing and why it is doing it.” Explanation facilities help the user validate the model and provide the user with a better understanding of the underlying reality the model is trying to represent.

¹*Address for correspondence:* Department of Statistics Box 354322, University of Washington, Seattle WA 98195-4322. E-mail: madigan@stat.washington.edu. WWW: <http://bayes.stat.washington.edu>.

Belief networks (Lauritzen and Spiegelhalter, 1988; Pearl, 1988; Almond, 1995a, Spiegelhalter, *et al.*, 1993) provide an important bridge between statistical modeling and expert system modeling. Formally, a belief network is a set of probability measures defined by distributional assumptions (such as multivariate normality) together with a set of conditional independencies that can be represented by an acyclic directed graph (in a directed graph, all the edges connecting nodes are directed; a directed graph is acyclic if it contains no cycles.) The graph, with nodes corresponding to random variables and edges connecting directly related variables, provides a precise representation of the independencies. The graph serves as both a visual representation of the model (a sort of informal explanation) and a guide to efficient probability computation algorithms (see, for example, Dawid, 1992). In what follows, we will refer to the graph as the “belief network”.

In recent years, researchers in Statistics and in Computer Science have directed considerable attention at belief networks for discrete-valued random variables (see, for example, Spiegelhalter, *et al.*, 1993, or Poole, 1994). Viewed from the Bayesian perspective, these belief networks generate predictive distributions through elicitation of informative prior opinion and subsequent combination with data. For this reason, as well as the emergence of efficient computation algorithms, expert system builders now commonly use belief networks. Important applications include forecasting, risk assessment, classification, and decision making.

Using belief network models in expert system applications requires appropriate explanation facilities. Chamberlain and Nordahl (1988), Cooper (1989), Druzdzel (1996), Henrion and Druzdzel (1990), Lauritzen and Spiegelhalter (1988), Pearl (1987), Suermondt (1991), and Suermondt and Cooper (1993) and have all made suggestions as to how such explanations might be generated. These approaches, however, fail to fully exploit the natural visual metaphor of the belief network.

This paper describes an approach to explanation in belief networks based primarily on visualizing the propagation of evidence through the belief network. We use Good’s *weight of evidence* (Good, 1977) as our basic metric of explanatory power and the graph itself to provide context for the “evidence flows”. We envision these explanations as part of a graphical display of the model in which the explanation is dynamically generated in response to user queries. This should have similar strengths to other analysis methods built on dynamic graphics, such as scatterplot brushing (Becker and Chambers, 1987).

We have implemented the methods of this paper in GRAPHICAL-BELIEF, Mathsoft, Inc.’s belief network modeling software (Almond, 1995b). All the color figures are taken directly from that software. While the medical context motivates our examples and terminology, the methods are equally applicable to domains such as risk analysis and forecasting.

Section 2 provides the mathematical basis for our approach to explanation; in particular, it reviews weights of evidence and belief networks and discusses techniques for calculating evidence flows in belief networks. Section 3 develops selected graphical displays for visualizing explanations. Section 4 contrasts our methods with other approaches to explanation in belief networks.

2 Background

Consider the following scenario:

A urologist has an expert system for selection of an optimal treatment technique for renal calculi (kidney stones). The clinician enters a patient's findings and the system presents a ranked list of possible treatments. The clinician then wants to know *why* the expert system is recommending (or not recommending) a given procedure. For example, the expert system might assign low success probability to extracorporeal shockwave lithotripsy (ESWL), a relatively low cost therapy. When asked why, the expert system would explain that while ESWL is successful in the majority of cases, this particular patient is likely to have scarred kidneys and the stone is likely to be difficult to image both of which reduce the chance of successful ESWL (i.e., stone clearance). The expert system can further explain that the scarring is likely because of previous renal surgery and the imaging is likely to be difficult because the stone is small.

This example reveals the basic components of an explanation. First, the explanation pertains to a *target variable*, in this case, stone clearance. Second, along with the prior information about the target variable (“ESWL is successful in the majority of cases”), the explanation lists the key findings which influence the target variable (scarred kidneys and image quality). An obvious refinement here would be to quantify the relative importance of these findings. Third, the findings which “explain” the conclusion about the target variable, are in turn explained by further findings (scarred kidneys by previous renal surgery and image quality by stone size). This introduces the idea of *evidence chains* which form an important part of our analysis.

Section 2.1 explores the lithotripsy example in the context of belief networks, conditional independence models that support the reasoning through intermediate variables implicit in the example. Section 2.2 introduces a basic metric for explanatory importance: the *weight of evidence*, and Section 2.3 shows how to calculate evidence in chains. For simplicity, we focus primarily on binary random variables. However, the concepts we describe have immediate extensions to non-binary discrete random variables.

2.1 Lithotripsy and Belief Networks

Kidney, ureteric, and bladder calculi are abnormal concretions occurring within the body, usually composed of mineral salts (Herbut, 1952). They occur more commonly in males than in females, seldom occur in black skinned people and are particularly common in certain geographical regions with a dry hot climate. Calculi have been found weighing as much as 286 grams (Mayers, 1940). Historically, stones have been treated by inducing the spontaneous passage of the stone with various drugs and failing that, operative removal. There are many dangers associated with the operative procedure including hemorrhage, stricture, kinks, infections, and dislodgment of the calculus with inability to find it at the time of operation. In recent years, Extracorporeal Shock Wave Lithotripsy (ESWL) has become established as the most popular treatment for urinary tract calculi (Kiely, *et al.*, 1989, Lingeman, *et al.*, 1987, Marberger, *et al.*, 1988). ESWL focuses

hundreds of high frequency shockwaves on the stone, creating high energy at the point of focus. This disintegrates the stone, which is then cleared by the normal functioning of the kidneys. ESWL is non-invasive and typically requires between one and five, 30-minute treatment sessions. However, ESWL fails to disintegrate about 20% of all stones and these subsequently require more traditional treatments. The annual cost associated with these failed ESWL treatments in the United States alone is in excess of \$50 million.

In order to flag potentially problematic cases *before* treatment, we constructed a belief network to predict the outcome of ESWL treatment. From an exploratory analysis of the available data (see Kiely, *et al.*, 1989) and in consultation with urologists, we selected a set of 6 indicants to predict the two outcome variables, Disintegration (*D*) and Clearance (*C*). Table 1 shows the indicants.

Table 1. Indicants for the renal calculi example. The numbers in parentheses are a shorthand notation for the possible values of the indicants.

	<i>Indicant</i>	<i>Possible Values</i>
<i>P</i>	Pain	Mild or None (0) / Moderate or Severe (1)
<i>T</i>	Stone Site	Upper Kidney (0) / Elsewhere (1)
<i>Z</i>	Stone Size	≤ 2 cm (0) / > 2 cm (1)
<i>S</i>	Scarred Kidneys	Yes (1) / No (0)
<i>R</i>	Previous Renal Surgery	Yes (1) / No (0)
<i>I</i>	Ultrasound Image Quality	Excellent or Good (1) / Vague or None (0)

The final indicant concerns the appearance of the stone on an ultrasound examination. This is of predictive relevance because lithotripter operators image calculi via ultrasound during ESWL treatment.

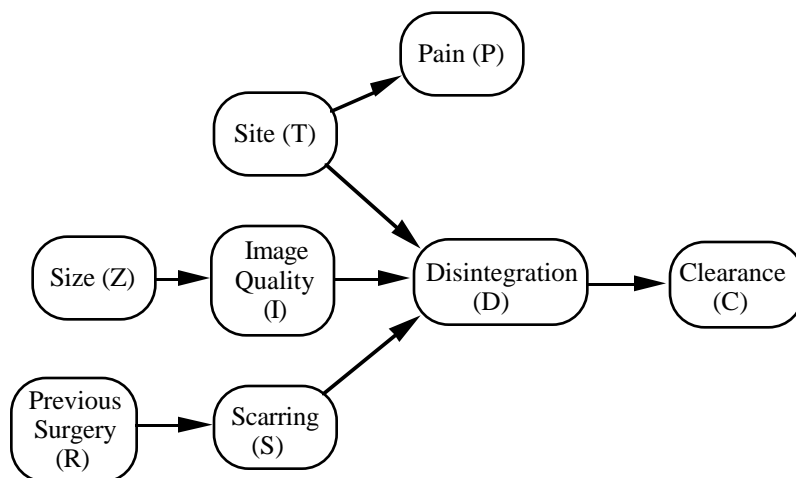


Figure 1. Belief network for the renal calculi example

We present a belief network model for this application in Figure 1. This belief network implies a factorization of the joint distribution of the eight variables as follows:

$$\Pr(P, T, Z, S, R, I, D, C) = \Pr(C|D) \Pr(D|T, I, S) \Pr(S|R) \Pr(R) \Pr(I|Z) \Pr(Z) \Pr(T) \Pr(P|T).$$

This factorization embodies a number of conditional independencies such as:

$$C \perp (P, T, I, S, R, Z) \mid D, \text{ and}$$

$$D \perp R \mid S$$

where \perp denotes conditional independence. Lauritzen, *et al.* (1990) provide a definitive description of the Markov properties implied by a belief network.

A complete specification of the belief network model requires values for all the conditional probabilities listed in the factorization. We elicited these probabilities from an experienced urologist. Madigan (1989) provides details of the elicitation procedure. Table 2 presents the probabilities.

Table 2. Conditional probabilities for the renal calculi example.

$\Pr(S=1 R=0)=0.30$	$\Pr(P=1 T=0)=0.6$
$\Pr(S=1 R=1)=0.95$	$\Pr(P=1 T=1)=0.8$
$\Pr(I=1 Z=0)=0.50$	$\Pr(C=1 D=0)=0.10$
$\Pr(I=1 Z=1)=0.80$	$\Pr(C=1 D=1)=0.95$
$\Pr(T=1)=0.50$	$\Pr(Z=1)=0.50$
$\Pr(R=1)=0.10$	
$\Pr(D=1 T=0, I=0, S=0)=0.70$	$\Pr(D=1 T=0, I=0, S=1)=0.55$
$\Pr(D=1 T=0, I=1, S=0)=0.90$	$\Pr(D=1 T=0, I=1, S=1)=0.75$
$\Pr(D=1 T=1, I=0, S=0)=0.65$	$\Pr(D=1 T=1, I=0, S=1)=0.50$
$\Pr(D=1 T=1, I=1, S=0)=0.80$	$\Pr(D=1 T=1, I=1, S=1)=0.60$

An important purpose of belief network models such as this is to facilitate calculation of arbitrary conditional probabilities. For example, the clinician might want to calculate:

$$\Pr(C|P=\text{Moderate or Severe}, T=\text{Elsewhere}, I=\text{Vague or None}),$$

that is, the probability of clearance in a patient with a painful bladder stone and poor image quality. In general, such a calculation would involve calculating the full joint distribution of the variables in the model, and then calculating the appropriate marginal distributions leading to the required probability. However, in applications of even moderate size (say, 50 variables or more), these calculations become impractical. A breakthrough occurred in the 1980s when Spiegelhalter (1987), Pearl (1988), and Lauritzen and Spiegelhalter (1988) presented algorithms for efficiently calculating conditional probabilities through a series of local calculations by exploiting the conditional independencies in the model. Dawid (1992) and Shenoy and Shafer (1990) present very general forms of these algorithms. These are based on the idea of passing messages between the cliques of the model (cliques are the maximal subgraphs where every node is connected to every other node by an edge.) The messages in fact provide primitive explanations. For instance, the facilities for examining those messages contained in the computer program GRAPHICAL-BELIEF have proved useful as model diagnostics (Almond, 1995a). Unfortunately, directly

interpreting the messages requires a sophisticated understanding of the message passing algorithm.

Note that all the edges in the belief network of Figure 1 are directed. While such a directed view is convenient for model construction and elicitation, the probability calculation algorithms of Dawid (1992) and Shenoy and Shafer (1990) exploit an undirected view of the belief network model. We show such an undirected view in Figure 2. All the conditional independencies implied by this undirected model are also implied by the directed model.

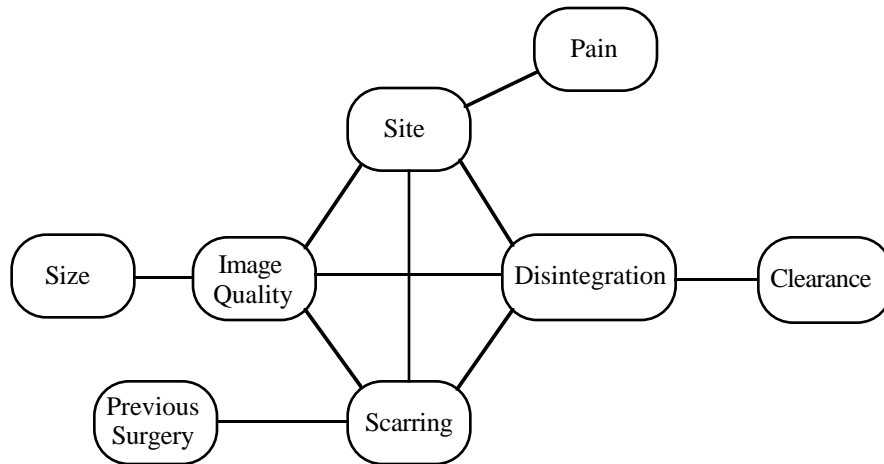


Figure 2. Belief network (undirected view) for the renal calculi example

It is possible to present explanations in the context of either the directed or undirected view. However, the directed view can lead to counter-intuitive displays, especially when all the arrows in the model flow in a causal direction. In this case, while the model points from disease to finding, the evidence (and hence the explanation) flows from finding to disease. Consequently, we mostly use the undirected view for explanation.

2.2 Metric for Explanation: Weight of Evidence

Here we introduce a measure for the explanatory importance of a particular finding, E (such as an observation or a test result), for a target hypothesis, H (such as a disease state). Let H be the hypothesis and $\neg H$ its negation. The weight of evidence (Good, 1977, 1985, van Fraassen, 1980, Schum, 1988) for H provided by E is:

$$W(H:E) = \log \frac{\Pr(E|H)}{\Pr(E|\neg H)}$$

Good (1985) provides a comprehensive review of the properties of weights of evidence and presents a detailed justification for using weights of evidence to measure explanatory importance. For ease of presentation, Good (1985) suggests taking the logarithm to the base 10 and multiplying the weight of evidence by 100, calling the resulting units *centibans*.

Note that the probability calculation algorithms provide an efficient method for calculating the weight of evidence for each node. First, we temporarily set the target hypothesis H to “true”, and “propagate” this to all variables, E (thus calculating $p(E|H)$ for each one). Second, we temporarily set the target variable H to “false”, and propagate this to all variables (thus calculating $p(E|\neg H)$ for each one). In this manner, we can easily calculate weights of evidence using standard software.

Finally in this section, we define potential and expected weights of evidence. Suppose we have a test possible outcomes t_1, \dots, t_n . Then, we can calculate the *potential weights of evidence*, $W(H:t_1), \dots, W(H:t_n)$, for each test outcome. This is, $W(H:t_i)$ provides the weight of evidence that *would* be provided for H , if the test result was known to be t_i . The *expected weight of evidence* provided by a test for a target variable is the average weight of evidence of the possible test results when the hypothesis is true:

$$EW(H:E) = \sum_{i=1}^n W(H:t_i) \Pr(t_i|H).$$

This provides a measure of the information content of a future finding. Madigan and Almond (1995) and Heckerman, *et al.* (1993) illustrate some properties and uses of the expected weight of evidence in belief network models. In particular they suggest using expected weight of evidence as a “test selection” measure; the basic idea is that a decision maker acquiring evidence sequentially, should, at each step, acquire the piece of evidence that provides that largest expected weight of evidence.

Although we believe that the weight of evidence is the best metric for explanation, many of the suggestions for visualizing weights of evidence presented in this paper would work with other measures as well. For example, Suermondt (1991) uses the Kulback-Leibler distance, which, like the weight of evidence, is another entropy based measure of explanatory power.

2.3 Evidence Flows

Recall that in the example, the computer reasoned that ESWL was unlikely to be successful because the kidneys were likely to be scarred because of previous renal surgery. This is an example of an *evidence flow* from a finding (previous renal surgery) through an intermediate variable (scarring) to the target variable (disintegration). Understanding and visualizing evidence flows is a key part of the explanation process in belief networks. We start with a simple example.

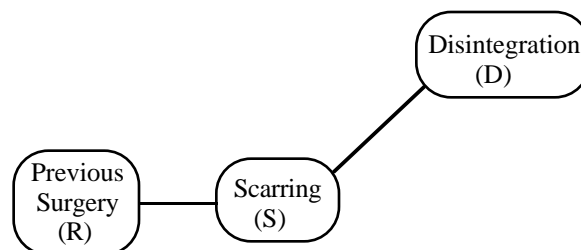


Figure 3: Belief network fragment (undirected view)

Figure 3 shows an undirected view of a fragment of the ESWL belief network model of Figure 2 corresponding to evidence flow from Previous Renal Surgery (R) through Scarring (S) to Stone Disintegration (D). Since this model fragment implies that R and D are conditionally independent given S , the following four probabilities (plus the probability of disintegration) fully define the probability distribution for this model fragment:

$$p_S = \Pr(S = 1|D = 1), \quad q_S = \Pr(S = 1|D = 0); \quad p_R = \Pr(R = 1|S = 1), \quad q_R = \Pr(R = 1|S = 0);$$

where $p_S, q_S, p_R,$ and $q_R \neq 0$. Here $R, S,$ and D are each binary variables.

Assume now that we know that the patient has had renal surgery (i.e., $R=1$) and are trying to establish whether or not ESWL will disintegrate the stone (i.e., $D=1$). Since $R \perp D | S$, the weight of evidence $R=1$ provides for $D=1$ is:

$$W(D = 1:R = 1) = \log \frac{\Pr(R = 1|D = 1)}{\Pr(R = 1|D = 0)} = \log \frac{p_R p_S + q_R (1 - p_S)}{p_R q_S + q_R (1 - q_S)}.$$

We are interested in the “flow” through the intermediate node S . The “incoming” evidence at S is:

$$W(S = 1:R = 1) = \log \frac{p_R}{q_R}.$$

There are *two* “outgoing” potential weights of evidence, corresponding to the two possible states of S . These are:

$$W(D = 1:S = 1) = \log \frac{p_S}{q_S} \quad \text{and} \quad W(D = 1:S = 0) = \log \frac{1 - p_S}{1 - q_S}$$

The following properties of weights of evidence provide useful insights into the evidence flows through a single node:

N1. $\text{sign } W(D=1:R=1) = \text{sign } W(D=1:S=1) \times \text{sign } W(S=1:R=1).$

For instance, if $R=1$ provides positive (negative) evidence for $S=1$, and $S=1$ provides positive (negative) evidence for $D=1$, then $R=1$ provides positive evidence for $D=1$.

N2. $|W(D=1:R=1)| \leq |W(S=1:R=1)|$

The evidence flow from R to D is constrained by the evidence on the incoming edge (i.e., the edge connecting R and S).

We define the *relevant outgoing weight of evidence*:

$$W_{rel: R=1}(D=1:S) = \begin{cases} W(D=1:S=1) & \text{if } W(S=1:R=1) > 0 \\ W(D=1:S=0) & \text{if } W(S=1:R=1) \leq 0 \end{cases}$$

So, if $R=1$ provides positive evidence for $S=1$, then $W(D=1:S=1)$ is the relevant weight of evidence, otherwise $W(D=1:S=0)$ is the relevant weight.

$$\mathbf{N3.} \quad |W(D=1:R=1)| \leq |W_{rel: R=1}(D=1:S)|$$

The evidence flow from $R=1$ to $D=1$ is constrained by the *relevant* evidence on the outgoing edge (i.e., the relevant evidence on the edge connecting S and D).

Thus, the total weight of evidence $W(D=1:R=1)$ is constrained both by the magnitude of the weight of evidence coming into S , $W(S=1:R=1)$, and by the *relevant* outgoing weight of evidence out of S , $W_{rel:R=1}(D=1:S)$. In effect, the evidence from R passes through S and into a channel provided by the relevant outgoing weight of evidence flowing from S to D .

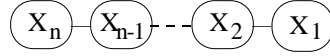


Figure 4. Simple evidence chain

Now consider the more complex belief network of Figure 4. This is an example of an *evidence chain*. Assume that we observe $X_n=1$ and want to calculate the evidence it provides for $X_1=1$. The weight of evidence our observation provides for $X_1=1$ is then:

$$\begin{aligned} W(X_i=1:X_n=1) &= \log \frac{\Pr(X_n=1|X_i=1)}{\Pr(X_n=1|X_i=0)} \\ &= \log \frac{\Pr(X_{i+1}=1|X_i=1)\Pr(X_n=1|X_{i+1}=1) + \Pr(X_{i+1}=0|X_i=1)\Pr(X_n=1|X_{i+1}=0)}{\Pr(X_{i+1}=1|X_i=0)\Pr(X_n=1|X_{i+1}=1) + \Pr(X_{i+1}=0|X_i=0)\Pr(X_n=1|X_{i+1}=0)} \end{aligned}$$

In this manner, $W(X_1=1|X_n=1)$ can be calculated recursively starting with $\Pr(X_n=1|X_n=1)=1$ and $\Pr(X_n=1|X_n=0)=0$.

Recall that in the three node chain, the total weight of evidence was limited by both the weight of evidence coming into the intermediate node and by the relevant weight of evidence going out of the intermediate node. Thus, the three-node evidence chain is “no stronger than its weakest link.” A simple induction argument extends this property to evidence chains of arbitrary length (as in Figure 4). The causal chains of Good (1961) and the evidential chains of Schum and Martin (1982) also have this property.

3 Visualizing Explanations

Throughout this section, we assume that we have a number of different findings about a particular patient and that we are interested in determining the truth (or falsehood) of a target variable such as Clearance. There are two questions the user might ask for which the model must provide explanations: “What was the relative importance of each of the findings in determining

the probability distribution of the target variable?” and “Why is a particular finding influential/not influential?” The first involves looking at the weights of evidence provided by the findings for the target variable (Section 3.1). The second involves looking at evidence flows. Section 3.2 visualizes evidence flows for trees with binary variables. Section 3.3 extends the visualization techniques to more complex belief networks and Section 3.4 extends the techniques to non-binary variables.

3.1 Evidence Balance Sheets and Node Coloring

To answer the question “What was the relative importance of each of the findings in determining the probability distribution of the target variable?” the computer can simply display the weight of evidence each finding provides for the target variable. Spiegelhalter and Knill-Jones (1984) suggest using an “evidence balance sheet”, separately listing positive and negative findings. This approach allows the user to see which findings were most important. Contrasting the positive and negative findings also explicates conflicts in the evidence. John Tukey, in the discussion of Spiegelhalter and Knill-Jones (1984) suggested displaying the evidence balance sheet graphically. Figure 5 presents the approach we have implemented in GRAPHICAL-BELIEF.

To initiate an evidence balance sheet, the user uses the mouse to attach a “probe” to the target variable in the GRAPHICAL-BELIEF window displaying the model (see, for example, Figure 7). From a pop-up menu, the user then selects the “desired” or “positive” state of the target variable, and generates the evidence balance sheet. In the balance sheet in Figure 5, the column labeled “Target Probability” shows the probability of successful clearance, i.e., the probability associated with the “desired” or “positive” state of this particular target variable. Here, the target probability is 0.72 *a priori*, and then 0.71 conditional on Pain=Yes. That is, Pain provides a small negative weight of evidence for Clearance, as indicated by the small red bar in the column labeled “WOE”. The next row shows the effect of also conditioning on Previous-Surgery=Yes. This provides a larger negative weight of evidence for Clearance, and the probability of clearance drops accordingly. Finally, the stone is greater than 2 cm in diameter providing a small positive weight of evidence (shown in blue) for Clearance, and the final probability of clearance, conditional on these three findings is 0.65. Note that the “16” in the icon atop the third column, shows the half-width in centibans of the three rectangles beneath it. By default, GRAPHICAL-BELIEF sets this value to maximum absolute weight of evidence (in centibans) provided by the findings in the evidence balance sheet. However, using the left and right arrows either side of the “16”, the user can scale the rectangles to any value bigger than 16. This feature allows a user to compare different balance sheets with different maximum absolute weights of evidence.

Evidence Balance Sheet [Clearance = YES]




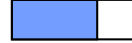




Indicant	State	WOE	Target Probability
Initial			 0.72
Pain	<input type="checkbox"/> Yes		 0.71
Previous-Surgery	<input type="checkbox"/> Yes		 0.63
Size	<input type="checkbox"/> >2		 0.65

Figure 5: Graphical Evidence Balance Sheet. Pain provides a small negative weight of evidence for Clearance, Previous-Surgery provides a larger negative weight of evidence, while Size provides a small positive weight of evidence. The prior probability of Clearance was 0.72; a posteriori, the probability is 0.65. The “16” under “WOE” shows the half-width in centibans of the boxes directly below.

In general, the weight of evidence provided by a single variable will change according to its location in the balance sheet. In our implementation, the balance sheet initially presents the variables in the order in which the user set their values. That is, if D is the target variable, and GRAPHICAL-BELIEF receives findings E_1, \dots, E_n in that order, then GRAPHICAL-BELIEF calculates the evidence provided by E_i for D , conditioned on the findings E_1, \dots, E_{i-1} . Thus, the entries in Figure 5 represent incremental weights of evidence. Other approaches are also possible, such as averaging over all orderings—see Kruskal (1987) and Theil and Chang (1988).

By selecting and dragging a variable in the balance sheet from one location to another, the user can manually change the order and thereby informally assess interactions between the variables. Furthermore, by clicking on a variable, the user can temporarily retract or change findings. These tools allow the user to better understand the effect of various findings on the probability of the target variable. In Figure 6, we show the effect of reversing the order of Previous Surgery and Size. In the absence of Previous Surgery, Size has a considerably larger positive weight of evidence for Clearance. Similarly, conditional on Size > 2 cm, Previous Surgery has a larger negative weight of evidence for Clearance. In either case, the evidence balance sheet highlights the fact that Previous Renal Surgery is the key finding and it has a negative influence on the probability of successful clearance. The other two findings, Pain=Yes and Size>2, provide conflicting evidence but have little influence.

Evidence Balance Sheet [Clearance = YES]







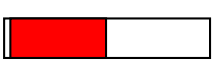
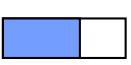
Indicant	State	WOE	Target Probability
			
Initial			 0.72
Pain	<input type="checkbox"/> Yes		 0.71
Size	<input type="checkbox"/> >2		 0.75
Previous-Surgery	<input type="checkbox"/> Yes		 0.65

Figure 6: Graphical Evidence Balance Sheets showing the effect of re-ordering the indicants. Pain still provides a small negative weight of evidence for Clearance, Size provides a small positive weight of evidence (but larger than in Figure 5), and Previous-Surgery provides a large negative weight of evidence, while . The prior and posterior probabilities of Clearance remain 0.72 and 0.65 respectively. The “21” under “WOE” shows the half-width in centibans of the boxes directly below.

Graphical evidence balance sheets can also supplement the actual weights of evidence for observed findings with the expected weights of evidence for future findings. Thus, the same display can assist with selecting a test to perform which is most likely to provide the most information about the target variable. Madigan and Almond (1995) discuss this problem of “test selection” at length.

We can display some of the information in the evidence balance sheet directly on the belief network via node coloring. A temperature scale provides the colors for the nodes. Nodes with high weight of evidence for the designated target variable are deep blue, nodes with low positive weight of evidence are pale blue. Nodes with negative evidence are red with intensity varying according to the magnitude of the appropriate weight of evidence. Nodes corresponding to unobserved variables are colored from light green to dark green according to their expected weight of evidence (expected weights of evidence are non-negative). The bars on either side of each node show the initial probability of the node's variable on the left hand side and the current probability on the right hand side.

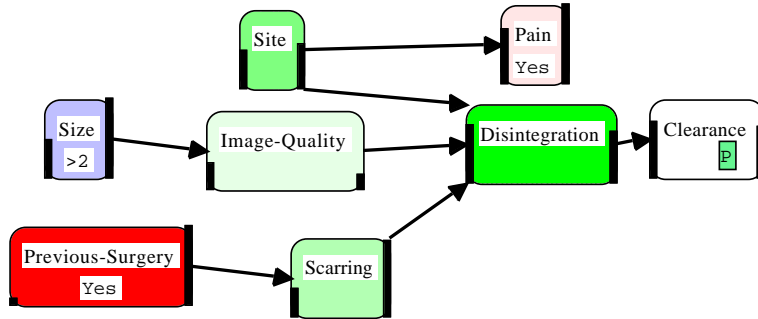


Figure 7: Graphical belief network with nodes colored according to weights of evidence and expected weights of evidence. As in Figure 5, we see that Pain provides a small negative weight of evidence for clearance, Size provides a small positive weight of evidence, while Previous-Surgery provides a larger negative weight of evidence. The green-colored nodes show expected weight of evidence for clearance. On average, Disintegration would provide the largest weight of evidence, followed by Site, Scarring, and Image Quality. Since Disintegration is an outcome variable, this would suggest that the physician should first establish the Site of the stone. Note that the Clearance node is tagged with “P”. In GRAPHICAL-BELIEF this indicates that the user has attached a *probe* to the node.

We note in passing that nodes can also be colored according to their current probability conditional on all findings. We show such a display in Figure 8. Because all the nodes are binary, we can label the states of each node as either *positive* or *negative*. The colors range from blue to red as the probability of the positive state ranges from one to zero.

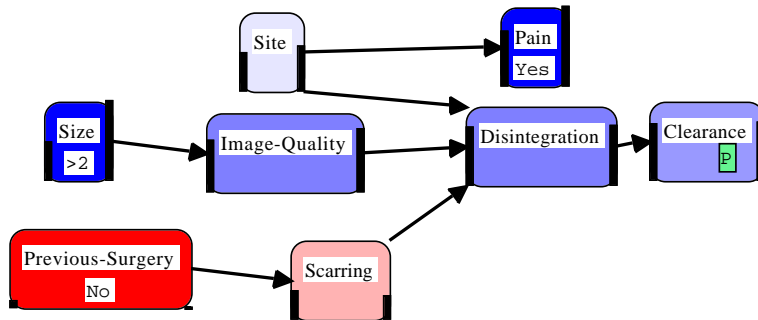


Figure 8: Belief network with nodes colored according to probabilities. Pain and Size are known to be in their “positive” states. Image-Quality, Disintegration, and Clearance have a high probability of being in their positive states. It is more likely that Site is in its positive state than not, but not by a large margin. Previous-Surgery is known to be in its negative state, so that Scarring not very likely.

3.2 Evidence Flows in Trees with Binary Variables

We now turn to the second question, “*Why* does a particular finding have such a large (or small) impact on the target variable?” To see this, we will want to examine the flow of evidence along an appropriate evidence chain from the finding to the target variable. In order to simplify the problem, for the moment we will assume that the belief network model (undirected view) is a tree, and we continue to assume that all variables are binary. Sections 3.3 and 3.4 show what happens when we relax those assumptions.

Since we are assuming that the belief network is a tree, there is a unique path between the finding and the target variable. This is the relevant *evidence chain*. In order to visualize the evidence flow along the chain, we calculate the actual and relevant potential weights of evidence at each step in the chain. Consider, for example, the evidence chain in Figure 9. Here, $X_5=1$ is the finding and X_1 is the target variable. Examine, for instance, the edge between X_4 and X_3 in Figure 9. The actual weight of evidence (the width of the inner bar) is the actual amount of evidence provided by $X_5=1$ for X_3 . In Section 2.3 we saw that evidence flowing along a chain is non-increasing. Therefore, this actual weight of evidence provides an upper bound on the amount of evidence that can eventually flow to X_1 . We use the outer width of the edge to display the *relevant* outgoing potential weight of evidence. This shows the weight of evidence that $X_4=1$ would provide for X_3 if it was observed, and provides an upper bound on the amount of evidence that $X_4=1$ could provide for X_1 .

As in Figure 8 above, we can label the states of each node as either positive or negative. In each step along the chain, we show evidence for the positive state as blue and evidence for the negative state as red.

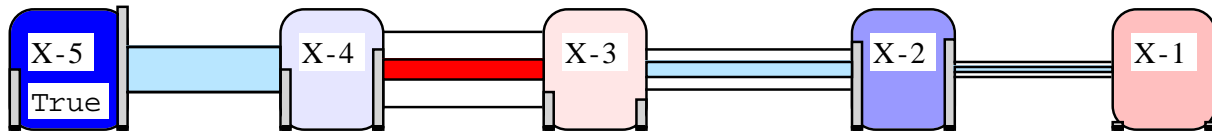


Figure 9: Simple Evidence Chain. The width of the channel represents the potential weight of evidence available (if the variable’s value was known exactly). The width of the interior bar represents the actual weight of evidence for the next node in the chain. The color is blue for evidence supporting the “positive” state and red for evidence supporting the “negative” state. This figure corresponds to the probabilities in Table 3 below.

Table 3. Simple example for Figures 9 and 10

$\Pr(X_5=1 X_4=1)=0.80$	$\Pr(X_4=1 X_3=1)=0.35$	$\Pr(X_3=1 X_2=1)=0.25$	$\Pr(X_2=1 X_1=1)=0.08$
$\Pr(X_5=1 X_4=0)=0.20$	$\Pr(X_4=1 X_3=0)=0.50$	$\Pr(X_3=1 X_2=0)=0.90$	$\Pr(X_2=1 X_1=0)=0.25$

Figure 9 shows the evidence flow corresponding the conditional probabilities given in Table 3. Note that:

1. From left to right, $X_5=1$ provides positive evidence for $X_4=1$, which provides negative evidence for $X_3=1$, which provides positive evidence for $X_2=1$, which provides positive evidence for $X_1=1$ (the positive state of the target variable);
2. X_4 “blocks” much of the evidence flow from X_5 ; and
3. Establishing a value for X_4 could “unblock” the chain, since it has a high potential weight of evidence for X_3 . However, little of this evidence will flow through to the target variable X_1 ; X_3 could provide considerable support for X_2 , but X_2 has a small potential weight of evidence for X_1 (as indicated by the width of the outer bar on the rightmost link) and will block incoming evidence.

As an alternative to Figure 9, we can use the idea of “Petri tokens” (Reisig, 1985) to dynamically show the flow of evidence along the chain. Figure 10 shows such an evidence flow. The width of the balls encodes the actual weight of evidence and the width of the channel encodes the relevant outgoing potential weight of evidence. The direction of motion of the balls shows the direction of evidence flow.

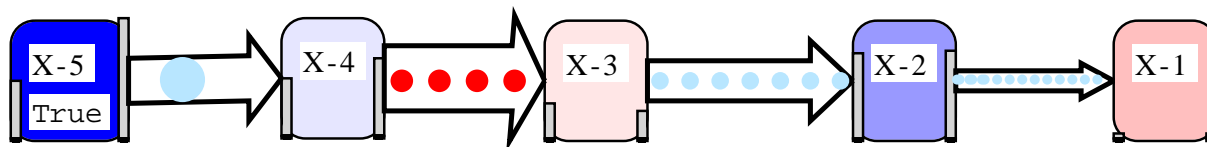


Figure 10: Evidence Chain with Petri Display. The width of the channel represents the potential strength of evidence available (if the variables value was known exactly). The width of the balls represents the actual strength of evidence for the next node in the chain. The color is blue for evidence supporting the “positive” state and red for evidence supporting the “negative” state. The balls move to follow the direction of evidence. This figure corresponds to the numbers in Table 3.

In the event that there are multiple findings, resulting in multiple chains of evidence leading from the findings to the target variable, all the evidence chains are displayed simultaneously, using either the Petri tokens or the static display (see Figure 13 below).

3.3 Beyond Trees: Berge Networks

The methods of the previous section rely on the existence of a unique evidence chain connecting every pair of nodes in the belief network. The following example demonstrates the difficulties that can arise in more general belief networks. Figure 11 depicts a simple belief network (undirected view) that is not a tree.

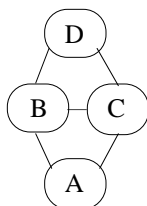


Figure 11. Belief network that is not a tree

Suppose we have the following probability distribution for this belief network:

		C=1		C=0	
		D=1	D=0	D=1	D=0
A=1	B=1	0.063	0.148	0.015	0.118
	B=0	0.010	0.064	0.000	0.001
A=0	B=1	0.007	0.016	0.015	0.118
	B=0	0.041	0.256	0.049	0.079

From these we can compute:

$$\begin{array}{ll}
\Pr(B=1|D=1)=0.500 & \Pr(B=1|D=0)=0.500 \\
\Pr(C=1|D=1)=0.605 & \Pr(C=1|D=0)=0.605 \\
\Pr(A=1|D=1)=0.443 & \Pr(A=1|D=0)=0.413
\end{array}$$

Thus, both of the potential weights of evidence provided by B for D are zero, as are the potential weights of evidence provided by C for D —that is, the potential weights of evidence on the two edges in the top half of the diamond are zero. However, the weight of evidence provided by $A=1$ for D is 3 centibans! It is not possible to use evidence chains to explain the evidence flows in this example.

Fortunately, there does exist a class of belief networks that is more general than trees and facilitates the graphical explanations introduced in Section 3.2. *Berge networks* (Berge, 1976) are chordal graphs with clique intersections of size one (an undirected graph is *chordal* if all cycles of length greater than three in the graph have a connection between two intermediate nodes of the cycle.) For example, in Figure 12(a), the cycle (B, C, D, E) has the chord (B, E) so the graph is chordal. However, this graph is not a Berge network since the cliques (B, C, E) and (B, D, E) have two nodes in common. Figure 12(b) is a Berge network since it is chordal and both its clique intersections are of size one.

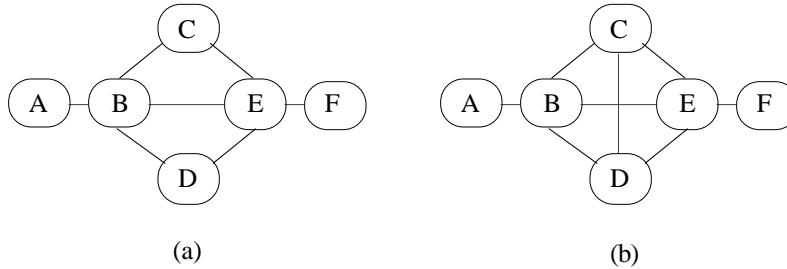


Figure 12: The graph on the left is chordal but not Berge. The graph on the right is both chordal and Berge.

Theorem 1 introduces a key property of Berge networks:

Theorem 1. For any pair of nodes X_i and X_j in a Berge network, B , the model defined by B is collapsible onto a unique evidence chain connecting X_i and X_j .

Proof. Follows from Madigan and Mosurski (1990).

A belief network, G , with corresponding nodes, V , is *collapsible* onto $S \subseteq V$, if the process of marginalizing onto S neither introduces any new conditional independencies not implied by G , nor removes any conditional independencies implied by G , concerning the variables in S . For example, the chordal graph of Figure 3 is *not* collapsible onto (R, D) , since the induced subgraph on (R, D) has no edges, implying, incorrectly, that R and D are marginally independent. On the other hand, the graph of Figure 3 *is* collapsible onto (S, D) , since the induced subgraph on (S, D) contains an edge and does not imply any new independencies.

For a given chordal graph, G , and a set of nodes, $S \subseteq V$, the SAHR algorithm (Madigan and Mosurski, 1990) finds $S' \subseteq V$, the (unique) smallest set of nodes containing S such that the model

defined by G is collapsible onto S' . The SAHR algorithm recursively removes simplicial vertices from $G \setminus S$ (a node $v \in V$ is simplicial if all its neighbors, i.e., the nodes connected to v by an edge, are themselves connected with edges.) The SAHR algorithm applied to Berge networks provides the following result:

Theorem 2. Let (X_i, X_j) be a pair of nodes in a Berge network, B . Then the SAHR algorithm applied to B with $S=(X_i, X_j)$ terminates in the (unique) evidence chain joining X_i and X_j .

Proof. Follows from Madigan and Mosurski (1990).

Thus, for any pair of nodes X_i and X_j in a Berge network, B , the model defined by B is collapsible onto a unique evidence chain connecting X_i and X_j . The SAHR algorithm finds the “relevant” variables for an explanation. For example, in assessing the evidence flow from A to F in the Berge network of Figure 12(b), the variables C and D are irrelevant. The SAHR algorithm, collapses these variables out of the graph, leaving the evidence chain (A, B, E, F) . In Figure 12(a), no such chain exists between, for instance, D and C (although other evidence chains do exist.)

Note that in Figure 12(a), both B and E are relevant when studying the evidence flow from C to D and the SAHR algorithm cannot collapse over these variables. We could convert Figure 12(a) into a Berge network by adding a edge between C and D (as in Figure 12(b)). In general, however, adding edges in this fashion can result in a highly connected graph and complex evidence flows. An alternative approach is to replace the nodes B and E by a single node (B, E) with four possible values. This violates our restriction to binary variables; the next section looks at the impact of lifting this restriction.

The lithotripsy example of Figure 2 is a Berge network. Figure 13 shows evidence flows along unique evidence chains from Pain, Size, and Previous Surgery to Clearance. The Figure shows that Disintegration and subsequent Clearance were both probable *a priori*. In the light of the three findings, they are both slightly more probable. That the patient is experiencing pain makes it more likely that the stone is not in the upper kidney, which in turn reduces the chances of successful disintegration. However, the stone is large, which improves the chances of getting good image quality, and, the patient has not had previous renal surgery, which makes scarring of the kidneys less likely. Both of these improve the chances of successful disintegration. The net effect is an increased probability of disintegration, which in turns makes successful clearance more likely.

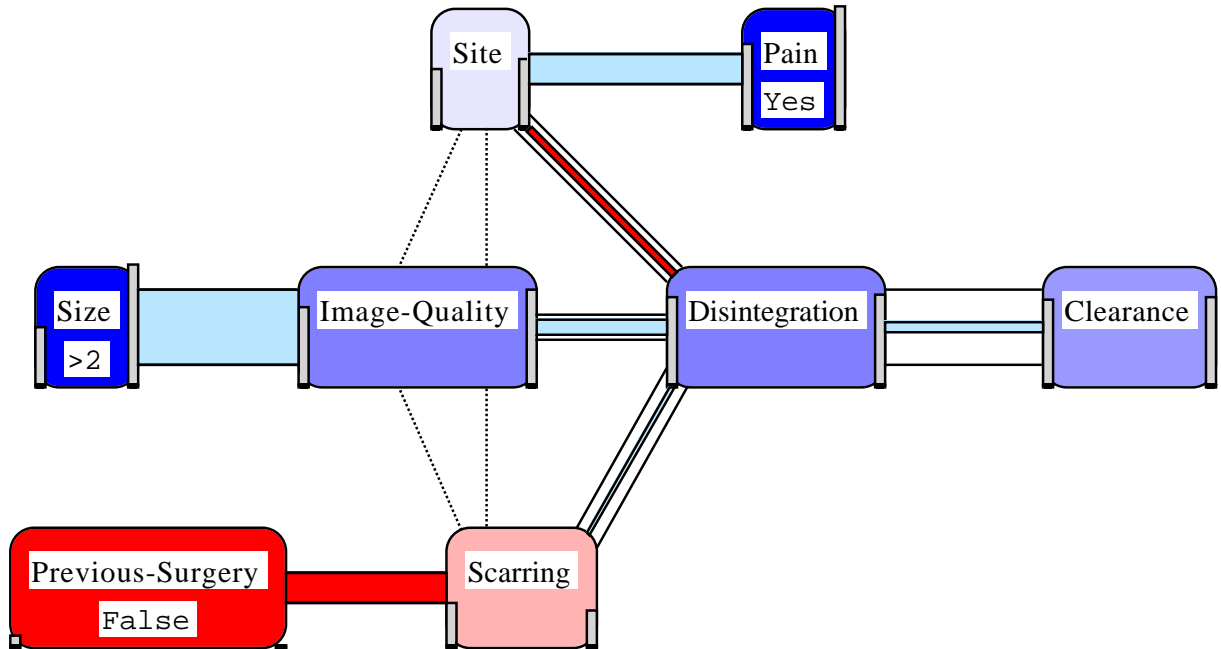


Figure 13: Evidence Chains in a Berge Network. There are three evidence chains converging on Disintegration. No Previous Renal Surgery makes scarring less likely which increases the probability of successful disintegration; Size > 2 makes good Image-Quality more likely, which also increases the probability of successful disintegration; Pain makes the site less likely to be the upper kidney, which reduces the probability of successful disintegration. Looking at the potential weights of evidence, we see that establishing values for Site, Image-Quality, or Scarring, would probably not change the probability of disintegration substantially, although amongst the three of them, Scarring has the largest potential weight of evidence. Not surprisingly, if we confirm successful Disintegration, successful Clearance will be much more likely.^β

3.4 General Belief Networks: Non-binary Variables and Clustering

For non-Berge networks, we first use the SAHR algorithm to find the smallest sets of nodes containing each finding and the target variable. For any of the sets that are *not* evidence chains, we combine nodes to create a unique evidence chain. Consider, for example, the non-Berge network of Figure 12(a). If A represents the finding and F represents the target variable, then the SAHR algorithm will collapse over C and D and onto the evidence chain (A, B, E, F) . For the non-Berge network of Figure 14, again with A representing the finding and F representing the target variable, the SAHR algorithm fails to remove any node.

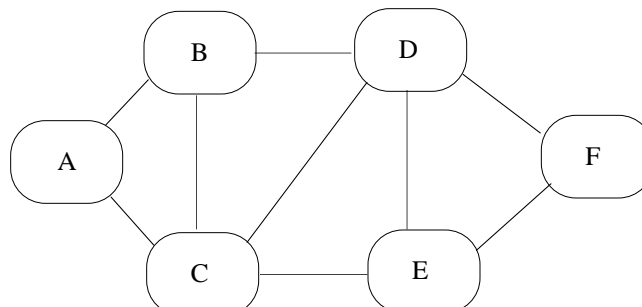


Figure 14: A non-Berge Network.

However, by combining, for instance, C and D into a single node “ CD ” we get the belief network of Figure 15. This is a Berge network, and applying the SAHR algorithm we get the evidence chain (A, CD, F) .

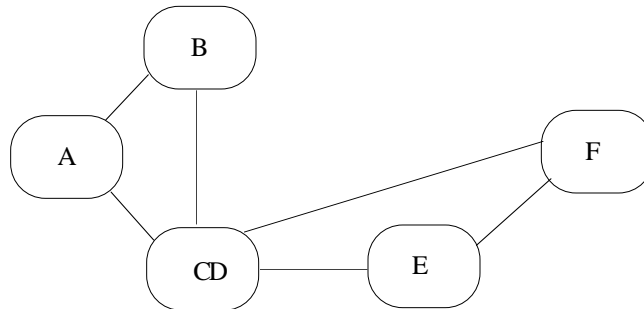


Figure 15: Merging C and D in Figure 14 produces a Berge Network.

In future versions of GRAPHICAL-BELIEF, the user will use the mouse to select nodes to combine, and GRAPHICAL-BELIEF will indicate at each step whether or not the graph is a Berge network. Berge (1976, p.32) provides a simple method for checking if a chordal graph is a Berge network.

The complexity of the explanations increases rapidly with both node size and the size of clique intersections. Our approach is to develop a hierarchy of explanations ranging from simple node coloring to extremely detailed displays, with the user delving only as deep as each situation demands. Patil, *et al.* (1984) described a similar philosophy.

Level 1: Node Coloring

Section 3.1 described node coloring. This can display either the current node probabilities or the weight of evidence provided by each node for a target variable.

Level 2: Node Marginals

Figure 7 augments the Level 1 explanation by displaying the prior and posterior marginal probabilities for each node. The MUNIN system of Andreassen, *et al.* (1987) adopts a similar approach.

Level 3: Evidence Balance Sheet

Section 3.1 introduced the evidence balance sheet. This level of explanation shows the sources of evidence for the target variable and their magnitude. It allows the user to carry out a limited sensitivity analysis by temporarily choosing values for nodes, assessing interactions by re-ordering nodes, and selecting optimal tests using expected weights of evidence.

Level 4: Relevant Potential Weights of Evidence

This level augments Level 1 or Level 2 by using the widths of the edges (see, for example, Figure 9) to display relevant potential weights of evidence on *all* chains connecting observed variables with the target variable.

Level 5: Dichotomized Weights of Evidence

For non-binary valued variables, our approach is to label one state as the “positive” state and group the other states together as the “negative” state. For Berge networks, evidence flows can then be displayed in the usual manner. The graphical evidence flows will depend on which states are labeled “positive”. However, a pop-up menu at each non-binary valued node can allow the user to choose any state as the “positive” state and observe the effect on the evidence flow.

4 Discussion

In this section we briefly review previous work on explanation in belief networks and discuss future directions.

4.1 Explanation in Belief Networks: Other Approaches

Pearl (1987) suggests that “verbal” explanations might be generated from a belief network. For example: *Disease A provides a fairly strong explanation for indicant E but provides only a partial explanation for indicant C. Indicant B, if it were present, would provide almost conclusive evidence for A.* Cooper (1989) and Druzdzel (1996) discuss similar approaches. Such explanations would be desirable in many situations. We believe that they could be derived from the quantitative/graphical approach we propose, combined with pre-stored text.

Henrion and Druzdzel (1990) consider explanation in the context of directed graphs and emphasize “linguistic probabilities”, i.e. the conversion of probabilities into phrases such as “very unlikely” or “fairly likely”. They consider singly connected graphs (Pearl, 1988) with binary nodes and provide textual explanations of dynamic behavior. Further developments of this approach and also their scenario-based explanations may be quite complementary to the more graphical approach we adopt. However, while the graphical explanations presented in this paper require more training to use than verbal explanations, we think they will be easier to use in complex situations.

Suermondt (1991) and Suermondt and Cooper (1993) also consider directed graphs and provide a useful discussion of metrics for influential findings and conflicts of evidence. They present an approach to the identification of “chains of evidence” which is similar in spirit to our approach. However, their method for comparing the strengths of competing chains involves temporarily severing the chains which can have unpredictable side-effects. As with the work of Henrion and Druzdzel, Suermondt and Henrion present explanations in a textual form.

In the legal context, Schum and Martin (1982) have defined fifteen generic types of evidence which they term “inference structures.” They claim that legal cases or collections of evidence can be thought of as collections of inference structures. It was suggested by Phillips in the discussion of Lauritzen and Spiegelhalter (1988) that these generic structures could be combined to model complex belief networks. In their response, the authors suggested that such networks “would seem to be good tests of any proposed automatic explanation facilities.” However, the

complexity of linking multiple explanations from different inference structures would seem to render such an approach impossible.

4.2 Future Directions

GRAPHICAL-BELIEF users have reacted positively to the hierarchy of explanations presented above. Node coloring proves adequate for many situations, but the facility to delve deeper into the inner workings of the model is especially useful for model debugging.

We have not addressed automated graph layout facilities in our work. Many belief networks will be too large to display in their entirety on a computer screen. In providing a particular explanation, it then becomes important to simultaneously display as many as possible of the nodes and edges that are participating in the explanation. For reasons of clarity, a secondary objective is to minimize edge crossings. There is a considerable literature on automatic graph layout—see Tamassia, *et al.* (1988) for a review.

A number of researchers have suggested that it is frequently desirable to consider several belief network models simultaneously. These models may be provided by a number of experts or may arise from a model selection procedure (Edwards and Havranek, 1985, Madigan and Raftery, 1994). Explanations could be averaged over the different models under consideration. Ideally, these would be integrated with support for making comparisons between models, such as proposed in Almond (1994).

Various researchers have proposed methods for the sequential updating of belief networks as data become available. It would be desirable to explain the impact of prior distributions on the behavior of the network and to alert the user to possible conflicts between prior opinions and subsequent data (see Spiegelhalter, *et al.*, 1993).

Acknowledgments

This work was supported in part by the Irish Stone Foundation, the National Science Foundation (DMS-92111629) and MathSoft's GRAPHICAL-BELIEF project, NASA SBIR Phase I NAS 9-18699 and Phase II, NASA SBIR NAS 9-18908.

The authors are grateful to Thien Nguyen who helped implement some of the explanation displays in GRAPHICAL-BELIEF. We are also grateful to the Associate Editor, the referees, and especially the former Associate Editor for detailed constructive advice.

References

Almond, R.G. (1995a). *Graphical Belief Modeling*. Chapman and Hall, London.

Almond, R.G. (1995b). *GRAPHICAL-BELIEF Overview*. World Wide Web: <http://bayes.stat.washington.edu/almond/gb/graphical-belief.html>.

Almond, R.G. (1994). Brushing Histories to Compare Models. StatSci Research Report 17. StatSci division of MathSoft, Inc. 1700 Westlake Ave, N, Seattle, WA 98109. (Submitted for publication).

Andreassen, S., Woldbye, M., Falck, B., and Anderson, S. (1987). MUNIN—A causal probabilistic network for interpretation of electromyographic findings. In: *Proceedings of the 10th International Joint Conference on Artificial Intelligence*. Milan, Italy.

Barr, A. and Feigenbaum, E. (1982). *Handbook of Artificial Intelligence*, Volume 2. Kaufmann, Los Altos.

Becker, R.A. and Chambers, J.M. (1987) Brushing Scatterplots. *Technometrics*, **29**, 127–142.

Berge, C. (1976). *Graphs and Hypergraphs*. Amsterdam: North Holland.

Buchanan, B.G. and Shortliffe, E.H. (1984). *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison-Wesley, Reading, Massachusetts.

Chamberlain, B. and Nordahl, T. (1988). Explanation in causal networks. Technical Report HF 88-19, Institute for Electronic Systems, Aalborg University.

Chandrasakaran, B., Tanner, M., and Josephson, J. (1989). Explaining control strategies in problem solving. *IEEE Expert*, **4**, 9–24.

Cleal, D.M. and Heaton, N.O. (1988). *Knowledge Based Systems : The User Interface*. Chichester: Ellis Horwood.

Coyne, R. (1990) Design reasoning without explanations. *AI Magazine*, **11**, 72–80.

Cooper, G. (1989). Current research directions in the development of expert systems based on belief networks. *Applied Stochastic Models and Data Analysis*, **5**, 39–52.

Dawid, A.P. (1992). Applications of a general propagation algorithm for probabilistic expert systems. *Statistics and Computing*, **2**, 25–36.

Druzdzal, M.K. (1996). Qualitative verbal explanations in Bayesian belief networks. *Artificial Intelligence and Simulation of Behaviour Quarterly*, to appear.

Edwards, D. and Havranek, T. (1985). A fast procedure for model search in multidimensional contingency tables. *Biometrika*, **72**, 339–351.

Good, I.J. (1961). A causal calculus. *Brit.J.Philos.Sci* **11**, 305–318; **12**, 43–51; **13** (1962) 88.

- Good, I.J. (1977). Explicativity: A mathematical theory of explanation with statistical applications. *Proceedings of the Royal Society (London) A*, **354**, 303–330.
- Good, I.J. (1985). Weight of evidence: a brief survey, in *Bayesian Statistics 2 : Proceedings of the Second Valencia International Meeting*, September, 1983, J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith (eds.). New York: North Holland, 249–269 (including discussion).
- Heckerman, D., Horvitz, E., and Middleton, B. (1993). An approximate nonmyopic computation for value of information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15**, 292–298.
- Henrion, M. and Druzdzel, M. (1990). Qualitative propagation and scenario-based approaches to explanation of probabilistic reasoning. In: *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence*, pp.10–20, Cambridge, MA.
- Herbut, P.A. (1952). *Urological Pathology*, Vol I. London: Henry Kimpton.
- Jensen, F.V., Lauritzen, S.L., and Olesen, K.G. (1989). Bayesian Updating in Recursive Graphical Models by Local Computations. *Computational Statistics Quarterly*, **4**, 269–282.
- Kiely, E.A., Madigan, D., Ryan, P.C., Grainger, R., McDermott, T.E.D., Lane, V. and Butler, M.R. (1989). Significant factors in ultrasound-imaged piezo-electric extracorporeal shockwave lithotripsy. *British Journal of Urology*, **66**, 127–131.
- Kruskal, W. (1987). Relative importance by averaging over orderings. *The American Statistician*, **41**, 6–10.
- Lauritzen, S.L. and Spiegelhalter, D.J. (1988). Local Computation with Probabilities on Graphical Structures and their Application to Expert Systems (with discussion). *Journal of the Royal Statistical Society (Series B)*, **50**, 205–247.
- Lauritzen, S.L., Dawid, A.P., Larsen, B.N., and Leimer, H-G. (1990). Independence properties of directed Markov fields. *Networks*, **20**, 491–505.
- Lingeman, J.E., Shirrell, W.L., Newmam, D.M., Mosbaugh, P.G., Steele, R.E. & Woods, J.R. (1987). Management of upper ureteral calculi with extracorporeal shockwave lithotripsy. *Journal of Urology*, **138**, 720–723.
- Madigan, D. (1989). *An investigation of weights of evidence in the context of probabilistic expert systems*. PhD Dissertation, Department of Statistics, Trinity College, Dublin.
- Madigan, D. and Mosurski, K. (1990). An extension of the results of Asmussen and Edwards on collapsibility in contingency tables. *Biometrika*, **77**, 315–319.

- Madigan, D. and Raftery, A.E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of the American Statistical Association*, **89**, 1535-1546.
- Madigan, D. and Almond, R.G. (1995). On test selection measures for belief networks. *Proceedings of the Fifth International Workshop on Artificial Intelligence and Statistics*. To appear.
- Marberger, M., Turk, C, and Steinkogher, I. (1988). Painless piezoelectric extracorporeal lithotripsy. *Journal of Urology*, **139**, 695–699.
- Mayers, M.M. (1940). Giant ureteral calculus. *Journal of Urology*, **44**, 47.
- Patil, R., Szolovits, P., and Buchanan, B. (1984). Causal understanding of patient illness in medical diagnosis. In: Clancey, W. and Shortliffe, E. (Eds.) *Readings in Medical Artificial Intelligence*, pp.339–360, Addison-Wesley, reading, Mass.
- Pearl, J. (1987). Distributed revision of composite beliefs. *Artificial Intelligence*, **33**, 173–215.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, California.
- Poole, D. (1994). *Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence*, D.Poole (Ed.), Morgan Kaufmann.
- Reisig, W. (1985). *Petri Nets*. Springer-Verlag.
- Salmon, W.C. (1984). *Scientific Explanation and the Causal Structure of the World*. New Jersey: Princeton University Press.
- Schum, D.A. (1988). Probability and the processes of discovery, proof, and choice, In *Probability and Inference in the Law of Evidence*, P. Tillers and E.D. Green (eds.). Kluwer Academic Publishers, 213–270.
- Schum, D.A. and Martin, A.W. (1982). Formal and empirical research on cascaded inference in jurisprudence. *Law Society Review*, **17**, 105–151.
- Shenoy, P.P. and Shafer, G. (1990). Axioms for Probability and Belief-Function Propagation. In: *Uncertainty in Artificial Intelligence*, **4**, 169–198. Reprinted in Shafer and Pearl (1990).
- Smith, J.Q. (1990). Statistical principles on graphs. In: Oliver, R. and Smith, J.Q. (eds.) *Influence Diagrams, Belief Nets, and Decision Analysis*, pp.89–120, Wiley.
- Spiegelhalter, D.J. and Knill-Jones, R.P. (1984). Statistical and knowledge based approaches to clinical decision support systems, with an application in gastroenterology (with discussion). *Journal of the Royal Statistical Society (Series A)*, **147**, 35–77.

- Spiegelhalter, D.J. (1986). Probabilistic reasoning in predictive expert systems. In: *Uncertainty in Artificial Intelligence*, Kanal, L.N. and Lemmer, J. (eds.). Amsterdam: North-Holland.
- Spiegelhalter, D.J. (1987). Coherent evidence propagation in expert systems. *The Statistician* **36**, 201-210.
- Spiegelhalter, D.J., Dawid, A.P., Lauritzen, S.L., and Cowell, R.G. (1993). Bayesian analysis in expert systems. *Statistical Science*, **8**, 219–283.
- Suermondt, H. (1991) Explanation of probabilistic inference in Bayesian belief networks. Technical Report KSL-91-39, Knowledge Systems Laboratory, Stanford University.
- Suermondt, H. and Cooper, G.F. (1993). AN evaluation of explanations of probabilistic inference. *Computers and Biomedical Research*, **26**, 242–254.
- Swartout, W. (1983). XPLAIN: A System for Creating and Explaining Expert Consulting Programs. *Artificial Intelligence*, **21**, 285–325.
- Tamassia, R., Di Battista, G., and Batini, C. (1988). Automatic graph drawing and readability of diagrams. *IEEE Transactions on Systems, Man, and Cybernetics*, **18**, 61–78.
- Theil, H. and Chang, C. (1988). Information theoretic measures of fit for univariate and multivariate linear regressions. *The American Statistician*, **42**, 249–252.
- van Fraassen, B.C. (1980). *The Scientific Image*. Oxford: Clarendon Press.