Hierarchical explanation of inference in Bayesian networks that represent a population of independent agents

Peter Šutovský and Gregory F. Cooper[1]

Abstract. This paper describes a novel method for explaining Bayesian network (BN) inference when the network is modeling a population of conditionally independent agents, each of which is modeled as a subnetwork. For example, consider disease-outbreak detection, in which the agents are patients who are modeled as independent, conditioned on the factors that cause disease spread. Given evidence about these patients, such as their symptoms, suppose that the BN system infers that a respiratory anthrax outbreak is highly likely. A public-health official who received such a report would generally want to know *why* anthrax is being given a high posterior probability. This paper describes the design of a system that explains such inferences. The explanation approach is applicable in general to inference in BNs that model conditionally independent agents; it complements previous approaches for explaining inference on BNs that model a single agent (e.g., explaining the diagnostic inference for a single patient using a BN that models just that patient).

1 Introduction

The importance of an explanation facility in intelligent systems was recognized early on. There are several studies that experimentally confirmed the positive impact of explanation on learning [2, 22, 11], on user perception of the system [33], and on the accuracy of decision making [2, 31]. Only one experimental study has evaluated the impact of explanation of inference in Bayesian networks on decision making [31]. It showed that appropriate explanation can improve decision making.

Developments in Bayesian network (BN) research in the 1980s and 1990s [5, 18] made BNs one of the most powerful tools for modeling of uncertainty in AI. Today, there are applications of BNs in various domains: tutoring systems [27], user interfaces [17], information retrieval [12], locomotive diagnostics [26], financial operational risk assessment [23], ecology [34], genetics [3], biosurveillance [6] and medical diagnosis [20].

There are methods for explanation of inference that were designed for general BNs [7, 31, 15, 32] or for special BNs [29, 28, 21, 4]. Some explanation methods have been successfully applied to BNs for real world applications [19, 31]. However, current explanation methods may not be feasible for very large networks.

BNs can become huge in size when they model a large population of agents by representing each agent with an individual subnetwork. We refer to these networks as *Bayesian Networks with a Population of Agents* (BNPA). These networks are useful in situations in which we want to learn something about a population based on information about the agents in the population. In disease-outbreak detection (aka biosurveillance), for example, the agents are often people who are reporting their symptoms when admitted to the hospital. Another type of agent could be a sensor that periodically measures and reports information about air quality at a given location in a city. An example of an agent-based BN in the military domain is the collection of intelligence from soldiers engaged in combat about the size and nature of an enemy force in order to derive an estimate of the enemy's military capabilities.

Agents in a population will be independent of one another, if we condition on all those factors that make them dependent, which are shown as interface nodes (I) in Figure 1. We refer to such networks as *Bayesian Networks with a Population of Independent Agents* (BNPIA). In detecting non-contagious infectious diseases, for example, we can model and condition on all the significant factors that cause a person to acquire the disease, such as the amount, location, and timing of the source of the disease. This paper describes a new approach, the *hierarchical explanation method* (HEM), for explaining inference in a BNPIA. Unlike previous explanation methods, the HEM exploits the modular character of a BNPIA.

2 Background

A BN is a framework for efficient representation of joint probability distribution over a set of random variables. A Bayesian network has two main components: a graph and local probability distributions [25]. The graph is the qualitative component of a BN, locally representing the relationships among domain variables. The graph of a BN is a directed acyclic graph (DAG), which means that it cannot contain cycles, that is, closed loops of directed links. It consists of nodes that represent random variables and directed arcs connecting nodes. We use the term node and variable interchangeably in the text. The arcs are directed from parents to children. An arc expresses the dependency of a child node on a parent node. Every node is associated with a *conditional probability distribution* (CPD). In general, a BN network can contain continuous variables, discrete variables or both.

An outcome of inference in a BN is a posterior probability distribution of variables as an effect of observed evidence. Hence the explanation of inference in Bayesian network is focused on the posterior distribution of a node of interest. Methods that explain BN inference try to clarify why and how a certain posterior probability was obtained given observed evidence.

The posterior probability that we want to explain results from a combination of several factors: the evidence, the BN structure (vari-

[1] University of Pittsburgh , Pittsburgh, email: {pesst17, gfc}@pitt.edu

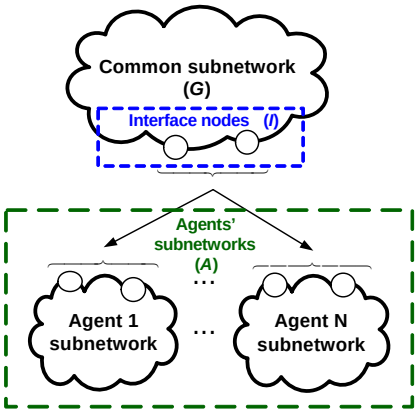

Figure 1. General representation of a Bayesian Networks with a Population of Independent Agents. The closed regions represent Bayesian subnetworks. The circles on the edge of the networks denote nodes that are connected by arcs that bridge subnetworks. Only two such "I/O" nodes are shown per subnetwork, but in general there could be any number. The arrows between subnetworks show the direction in which the Bayesian-network arcs are oriented between the subnetworks. The braces show which nodes can possibly be connected by arcs. Interface nodes *I* are the "I/O" nodes in the common subnetwork that connect common subnetworks with the agent subnetworks.

ables and arcs), the BN parameters (local conditional probabilities), and inference. Complete information about these factors could be included in the explanation. The calculation of posterior probability, however, involves many calculations with many numbers. Moreover, the evidence can consist of many findings. An explanation that simply lists all this information is unlikely to be very useful. Gregor and Benbasat [13] concluded that explanations which require less cognitive effort to access and understand will be used more often and will have larger positive effect on the performance, learning, or user perception.

Therefore, many explanation methods start construction of explanation with the selection of a subset of the most important findings from the set of all findings [31, 14, 4]. These methods use some quality of the explanation measure to evaluate the set of selected findings, which is usually based on a distance measure that measures the distance between the desired posterior probability obtained using complete evidence and the posterior probability obtained using the selected subset of findings. The smaller the distance, the better the selected subset represents the complete evidence. In the next step, explanation methods select paths between the selected subset of findings and the *node of interest* (NOI). Further simplification of explanation is achieved by examining which components of network structure (nodes and arcs) are important for propagation of evidence from the selected evidence nodes to the node of interest. Basically, if the removal of a node or arc does not appreciably change the posterior probability distribution of some node of interest, the node or arc is removed for the purpose of providing an explanation. INSITE, the explanation method proposed by Suermondt [30, 31], was the first comprehensive work based on this approach.

There are several variations of explanation methods. Madigan et al. [21] proposed an explanation method that provided explanation of inference in graphical form. Henrion and Druzdzel [16, 10] explained inference in Bayesian networks by means of qualitative explanations. Sember, Zukerman, and Wiegerink [28, 32] proposed methods that explain inference in BNs in terms of messages defined by Pearl's [24] belief propagation algorithm. Scenario-based explanation [11] provides explanation in the form of a list containing the most probable scenarios that are consistent with hypothesis and evidence. Some of these methods can be combined with the two-step method sketched in the previous paragraph. The proposed HEM introduced in this paper is closely related to INSITE [31] and to an explanation method based on analysis of evidence in naive Bayes networks, where weight of evidence is used to determine the influence of each piece of evidence (a finding) on the selected diagnosis [9].

3 Agent-based Bayesian networks

An agent-based BN (Figure 1) consists of several parts: a subnetwork that represents the whole population (the common part) and subnetworks that represent agents in the population individually (the agent part). The common part of the BN is connected with the agent parts of the BNPIA via the common part nodes that we call *interface nodes* (*I*). The size of the population can be very large. For example, a Bayesian model for biosurveillance called PANDA-CDCA [8] was tested using a population of 423,000 agents. PANDA-CDCA (Figure 2) is a BNPIA for diagnosing outbreaks of CDC (Center for Disease Control and Prevention) Category A diseases [1], which include anthrax, tularemia, plague, and several other serious infectious diseases. In this paper we use PANDA-CDCA as an example of an agent-based BN with independent agents. Agents in PANDA-CDCA are modeled as not interacting directly with each other in contracting disease, and hence, there are no directed arcs connecting variables in different agent subnetworks. Therefore, the agents are conditionally independent if we condition on the nodes *Outbreak Disease in Population* and *Fraction of Population with Outbreak Disease*. This independence assumption seems reasonable for non-contagious disease outbreaks. In case of non-contagious diseases due to bioterrorism, for example, the main increase in infected individuals in the population is due to a release of some biological agent for which we can assume non-transmission of the disease among individuals (agents) in the population.

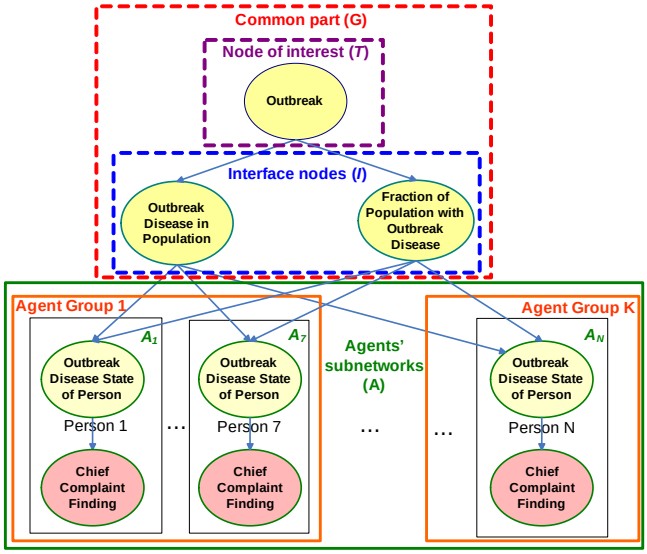

Figure 2. Example of Bayesian Networks with a Population of Independent Agents: *PANDA-CDCA* [8].

PANDA-CDCA takes as input chief complaints observed at an emergency department (ED) during the previous 24-hour period, and it outputs the posterior probability of CDC Category A diseases (plus a few other diseases, including influenza). The common part (**G**) for the PANDA-CDCA BN model is made up of the nodes that represent features common to the whole population, and an agent part (**A**) that consists of subnetworks $A = \{A_1, \ldots, A_n\}$ of all *n* individuals in the population, where A_i is subnetwork of i^{th} agent. Nodes in the common subnetwork are *Outbreak*, *Outbreak Disease in Population,* and *Fraction of Population with Outbreak Disease*. The *Outbreak* node represents the presence or absence of an outbreak. The node *Outbreak Disease* represents 12 explicitly modeled diseases. *Fraction of Population* represents the population that has an outbreak disease and has come to the ED within the previous 24 hours. The *Fraction of Population* and *Outbreak Disease* create an interface between the other global node (Outbreak node) and the agent subnetworks. In PANDA-CDCA all individuals (agents) are represented by identical subnetworks, although in general they could differ. An agent's subnetwork consists of the nodes *Outbreak Disease State of Patient* and *Chief Complaint Finding*. The node *Outbreak Disease State of Patient* represents diseases that each person can have according to the model. The CDC category A diseases and influenza, cryptosporidiosis, hepatitis A and asthma are modeled explicitly; any other disease which the patient may have is represented by the state "other", meaning some other disease. PANDA-CDCA was tested using semi-synthetic and real data with encouraging results [8].

4 Hierarchical explanation method

Unlike previously existing explanation methods for Bayesian networks, the hierarchical explanation method takes advantage of the modularity in BNPIAs. The structure of a BNPIA allows us to identify those agents that are most important for obtaining inference results (Figure 2). When HEM selects evidence for explanation it selects all evidence of the agent and, hence, the agent's subnetwork. Figure 3 shows how HEM is applied using PANDA-CDCA. Since in PANDA-CDCA all agents have the same subnetwork, it does not represent the most general example of BNPIA. However it provides an apropriately complex model for demonstrating HEM. HEM builds up explanation hierarchically using three levels. The information collected is represented by the schema in Figure 3. The tree structure in the figure represents variables (ellipses) at each level of the explanation and the instantiation of these variables (rectangles) are ranked and sorted from most important to least important for use in explanation. The instantiations of the nodes are ranked with respect to instantiations selected on the previous level and the evidence. HEM starts explanation with the *node of interest* (NOI), whose probability we want to explain. NOI constitutes the top explanation level in Figure 3. HEM assumes that the NOI is in the common part of BNPIA. For PANDA-CDCA it is the Outbreak (O) node. The lowest level represents patient models (subnetworks), each with its respective set of evidence. People modeled in PANDA-CDCA correspond to the agents of a BNPIA. Explanation can be simplified by creating groups of patients with the same model and evidence (model-evidence equivalence groups). As PANDA-CDCA uses the same subnetwork for all agents, we can simplify explanation by grouping agents with the same evidence. Therefore the explanation in PANDA-CDCA is constructed based on groups of agents with identical evidence rather than on individual agents. We will refer to groups of agents as evidence equivalence groups. Although, the HEM does not require agents to have identical subnetworks, it makes the grouping of the agents more efficient. The middle level consists of interface nodes; conditioning on them renders the agents independent of each other. In PANDA-CDCA, there are two interface nodes: *Outbreak Disease in Population* and *Fraction of Population with Outbreak Disease.* In order to keep the example simple, we assume in Figure 3 that there is only one interface node, namely, *Outbreak Disease in Population.*

Having introduced the three levels, we now summarize how explanation is performed using them. At the first level, all possible states $t_1, t_2, \ldots, t_j, \ldots, t_{N_T}$ of the NOI, T, are ranked using the posterior probability $P(t_j|\boldsymbol{e})$, where $\boldsymbol{e}$ is the evidence for all individuals in the population. Since there is no higher level, this posterior probability is conditioned only on evidence observed for the individuals. In this way we select the most probable instantiation of the NOI for the subsequent analysis. Alternatively, the user can select the state of NOI that he is interested in having explained. Suppose the selected instantiation of the NOI is t_3. At the middle level of the explanation tree in Figure 3, the instantiations of the states $i_1, i_2, \ldots, i_j, \ldots, i_{N_\mathcal{I}}$ of nodes $\mathcal{I}$ are ranked using the posterior probability $P(i_j \mid t_3, \boldsymbol{e})$. Suppose that i_7 is the instantiation with the highest posterior probability. At the lowest level of the explanation tree, we identify those individuals that most contributed to a high posterior probability of i_7, when compared to other possible instantiations of $\mathcal{I}$: $\neg i_7 = i_1 \vee \ldots \vee i_6 \vee i_8 \vee i_{N_\mathcal{I}}$. Recall that individuals with identical evidence are considered as one evidence equivalence group. Let $\boldsymbol{e}_j$ be the total evidence of the j^{th} group. The relative support of i_7 by the j^{th} evidence equivalence group is measured using the conditional *likelihood ratio* (LR) given by Equation 1.

$$L(i_7 : \boldsymbol{e}_j | t_3, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_{j-1}) = \\ = \frac{p(\boldsymbol{e}_j \mid i_7, t_3, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_{j-1})}{p(\boldsymbol{e}_j \mid \neg i_7, t_3, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_{j-1})}, \quad (1)$$

The denominator in Equation 1 is derived as follows:

$$p(\boldsymbol{e}_j \mid \neg i_7, t_3, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_{j-1}) = \frac{\sum_{k \neq 7}^{N_I} p(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_j \mid i_k)\, p(i_k \mid t_3)}{\sum_{k \neq 7}^{N_I} p(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{j-1} \mid i_k)\, p(i_k \mid t_3)}.$$

We have to use a conditional LR since in general

$$p(e_1 \mid \neg i_7, t_3,)\, p(e_2 \mid \neg i_7, t_3) \neq p(e_1, e_2 \mid \neg i_7, t_3)\ .$$

The likelihood ratio in Equation 1 allows us to decompose the posterior odds of i_7 into contribution of each evidence equivalence group $\boldsymbol{e}_j$ (Equation 2).

$$\frac{p(i_7|\boldsymbol{e}, t_3)}{p(\neg i_7|\boldsymbol{e}, t_3)} = \\ \frac{p(i_7|t_3)}{p(\neg i_7|t_3)} \prod_{j=1}^{N_G} L(i_7 : \boldsymbol{e}_j \mid t_3, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_{j-1}), \quad (2)$$

where N_g is number of evidence groups. Let $\mathcal{B}_{j-1}$ represent background information consisting of $T = t_3$ and $j-1$ already selected equivalence evidence groups $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{j-1}\}$. The LR allows us to determine which evidence equivalence group is supporting and which evidence equivalence group is contradicting the instantiation i_7, shown in Equation 3.

$$L(i_7 : \boldsymbol{e_j} \mid \mathcal{B}_{j-1}) = \begin{cases} > 1 & \text{evidence } \boldsymbol{e_j} \\ & \text{supports instantiantion } i_7 \\ & \text{over instantiation } \neg i_7 \\ = 1 & \text{evidence } \boldsymbol{e_j} \\ & \text{supports instantiantion } i_7 \\ & \text{equally to instantiation } \neg i_7 \\ < 1 & \text{evidence } \boldsymbol{e_j} \\ & \text{supports instantiantion } \neg i_7 \\ & \text{over instantiation } i_7 \ . \end{cases} \quad (3)$$

Moreover, the LR in Equation 1 allows us say that instantiation i_7 is $LR(i_7 : \boldsymbol{e_j}|\mathcal{B}_{j-1})$ times more (or alternatively, less) supported by e_j than $\neg i_7$ given the background information $\mathcal{B}_{j-1}$ if $LR(i_7 : \boldsymbol{e_j}|\mathcal{B}_{j-1}) > 0$ (or alternatively, $LR(i_7 : \boldsymbol{e_j}|\mathcal{B}_{j-1}) < 0$). The likelihood ratio given by Equation 1 depends on the order of selected group evidence and only in the case of a binary $\mathcal{I}$ can the LR be replaced by unconditional LRs, $L\,(i_7 : \boldsymbol{e_j})$, in Equation 2. A simple heuristic for ordering evidence in the case of an $\mathcal{I}$ with more than two states is to initially select the evidence equivalence group with the highest likelihood ratio. In this case, we select first the high-ranked evidence equivalence group. The remaining evidence equivalence groups are then sorted and applied similarly. This approach selects evidence that is most supportive of the instantiation of i_7 itself, regardless of the interaction of the selected evidence with the rest of the observed evidence. Once instantiations and evidence equivalence groups are selected, HEM will select information for explanation using the scores calculated for each instantiation and evidence equivalence group. Explanation presented to the user includes the score for the selected information.

5 Example of hierarchical explanation

This section provides an example of applying the HEM methodology to produce an explanation of inference for PANDA-CDCA. In particular, Figure 3 shows the scheme of HEM as applied to PANDA-CDCA. The observed evidence for PANDA-CDCA, $\boldsymbol{e}$, consists of chief-complaint findings extracted from chief-complaint strings recorded for each patient who comes to the emergency department. Chief-complaint findings included in the model are fever, cough, and headache, for example. Since *Outbreak* is the NOI, the posterior probability of each possible instantiation of that node is derived given the observed evidence on all individuals in the population. The posterior probability of $Outbreak = true$ is 0.9999 and posterior probability of $Outbreak = false$ is 0.0001. Assume we would like to know why the posterior of $Outbreak = true$ is so high. HEM next identifies the instantiations of those intermediate nodes that most support $Outbreak = true$. In this section, we will refer to the node *Outbreak Disease in Population* simply as to *Outbreak Disease*. For simplicity of exposition, we will not include the node *Fraction of population with outbreak disease* in the explanation described here. All possible instantiations of the variable *Outbreak Disease* are scored using the posterior of *Outbreak Disease* given evidence $\boldsymbol{e}$ and $Oubreak = true$. Suppose that the top scoring instantiation of *Outbreak Disease* is $Outbreak\,Disease = botulism$ $(score = 0.998)$ and the second most highly scored instantiation is $Outbreak\,Disease = plague$ (0.001). Explanation focuses on the most important (highest scoring) instantiation, that is, in $oubreak\,disease = botulism$ $(score = 0.998)$. As a final stage, HEM searches for groups of individuals that provide the highest evidential support for instantiation $Outbreak\,Disease = botulism$. Equation 1 is used to quantify such support. The highest support for $Outbreak\,Disease = botulism$ given $Outbreak = true$ is provided by a group of 36 patients with the chief complaint of *difficulty swallowing*. The second highest support is provided by a group of patients with the chief complaint of *slurred speech*. Using the information derived above, a simple verbal explanation can be constructed, such as: "PANDA-CDCA detected an outbreak (*Outbreak=true*) with probability 0.9999. The most probable outbreak disease is *botulism* with probability 0.998. Evidence that supports *botulism* as the outbreak disease is a group of 36 patients with a chief complaint of *difficulty of swallowing*. When 36 such patients come to the emergency department, the probability of botulism increases with respect to alternative outbreak diseases by a factor of 22".

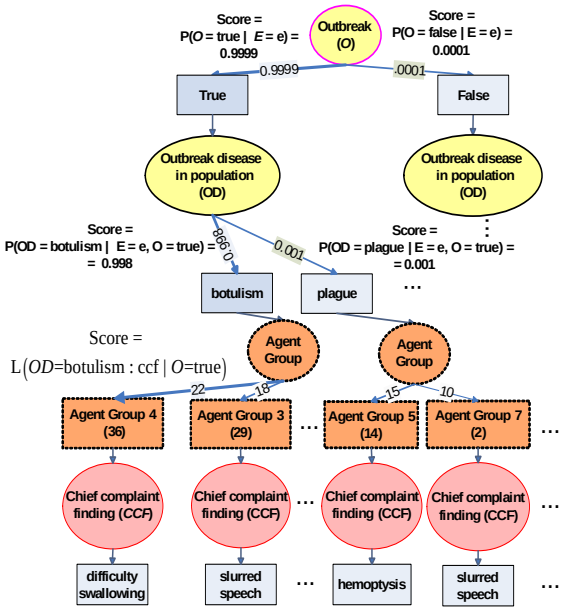

Figure 3. Schema of hierarchical explanation for PANDA-CDCA. The ellipses represent variables and the rectangles represent values. The numbers in parentheses represent the number of patients in the corresponding evidence group and the numbers on the edges represent the scores.

6 Summary and future research

This paper describes a novel method called HEM for explanation of inference in BNs that model populations of conditionally independent agents. HEM complements previous explanation methods that focus on explaining inference for BN models of single individuals. HEM exploits the modularity of BNPIA models to structure its explanations.

We currently are completing the implementation of the HEM explanation system. In the near future we plan to evaluate how effectively the system provides human users with explanations of inference.

ACKNOWLEDGEMENTS

This research was supported by the U.S. National Science Foundation grant IIS-0325581.

References

[1] http://www.bt.cdc.gov/agent/agentlist-category.asp.

[2] D. C. Berry and D. E. Broadbent. Explanation and verbalization in a computer-assisted search task. *The Quarterly Journal of Experimental Psychology*, 39a(4):585 – 609, November 1987.

[3] S. Bulashevska, O. Szakacs, B. Brors, R. Eils, and G. Kovacs. Pathways of urothelial cancer progression suggested by bayesian network analysis of allelotyping data. *International Journal of Cancer*, 110(6):850–856, 2004.

[4] U. Chajewska and D. L. Draper. Explaining predictions in Bayesian networks and influence diagrams. In *Interactive and Mixed-Initiative Decision-Theoretic Systems*, pages 23–31. AAAI Spring Symposium, 1998.

[5] E. Charniak. The Bayesian basis of common sense medical diagnosis. In *AAAI*, pages 70–73, 1983.

[6] G. F. Cooper, D. H. Dash, J. D. Levander, W. K. Wong, W. R. Hogan, and M. M. Wagner. Bayesian biosurveillance of disease outbreaks. *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, pages 94–103, 2004.

[7] Gregory F. Cooper. *NESTOR: A Computer-based Medical Diagnostic Aid that Integrates Causal and Probabilistic Knowledge*. PhD thesis, Stanford University, 1984.

[8] Gregory F. Cooper, John N. Dowling, John D. Levander, and Peter Sutovsky. A bayesian algorithm for detecting cdc category a outbreak diseases from emergency department chief complaints. In *Advances in Disease Surveillance*. International Society for Disease Surveillance, 2006.

[9] Robert G. Cowell, Steffen L. Lauritzen, A. Philip David, David J. Spiegelhalter, and David J. Spiegelhater. Probabilistic networks and expert systems. 1999.

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

[11] Marek J. Druzdzel and Max Henrion. Using scenarios to explain probabilistic inference. In *Working notes of the AAAI–90 Workshop on Explanation*, pages 133–141, Boston, MA, 1990. AAAI.

[12] Robert Fung and Brendan del Favero. Applying Bayesian networks to information retrieval. In *Communications of the ACM*, volume 38, pages 42–48, 1995.

[13] Shirley Gregor and Izak Benbasat. Explanations from intelligent systems: Theoretical foundations and implications for practice. *MIS Quarterly*, 23(4):497–530, 1999.

[14] P. Haddawy, J. Jacobson, and C. E. Kahn. Generating explanations and tutorial problems from Bayesian networks. *JAMIA*, pages 770–774, 1994. Suppl. S.

[15] Peter Haddawy, Joel Jacobson, and Charles E. Kahn. Banter: a bayesian network tutoring shell. *Artificial Intelligence in Medicine*, 10(2):177–200, June 1997.

[16] Max Henrion and Marek J. Druzdzel. Qualitative and linguistic explanation of probabilistic reasoning in belief networks. In *Proceedings of the Third International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, pages 225–227, Paris, France, July 1990.

[17] Eric Horvitz, John Breese, David Heckerman, David Hovel, and Koos Rommelse. The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users. In *Proceedings of the Fourteenth Annual Conference on UAI (UAI-98)*, pages 256–265, San Francisco, CA, 1998. Morgan Kaufmann Publishers.

[18] Jin H. Kim and Judea Pearl. A computational model for causal and diagnostic reasoning in inference systems. In *IJCAI*, pages 190–193, 1983.

[19] C. Lacave, A. Onisko, and F. J. Díez. Debugging medical Bayesian networks with Elvira's explanation capability. In *Workshop onBayesian Models in Medicine, Eighth European Conference on Artificial Intelligence in Medicine (AIME-2001)*, Cascais, Portugal, 2001.

[20] Carmen Lacave and Francisco Javier Díez. Knowledge acquisition in prostanet - a bayesian network for diagnosing prostate cancer. In Vasile Palade, Robert J. Howlett, and Lakhmi C. Jain, editors, *Knowledge-Based Intelligent Information and Engineering Systems*, volume 2774 of *Lecture Notes in Computer Science*, pages 1345–1350. Springer, 2003.

[21] David Madigan, Krzysztof Mosurski, and G. Almond, Russell. Graphical explanation in belief networks. *Journal of Computational and Graphical Statistics*, 6(2):160–181, June 1997.

[22] Kathleen Ellen Moffitt. *An empirical test of expert system explanation facility effects on incidental learning and decision-making*. PhD thesis, ARIZONA STATE UNIVERSITY, 1989. Chairperson-James Hershauer.

[23] M. Neil, N. E. Fenton, and M. Tailor. Using bayesian networks to model expected and unexpected operational losses. *Risk Analysis: An International Journal*, 25(4):963–972, 2005.

[24] Judea Pearl. Reverend Bayes on inference engines: A distributed hierarchical approach. In *AAAI*, pages 133–136, 1982.

[25] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. 1988.

[26] K. Wojtek Przytula and Don Thompson. Construction of Bayesian networks for diagnostics. In *Proceedings of 2000 IEEE Aerospace Conference*, 2000.

[27] K. G. Schulze, R. N. Shelby, D. J. Treacy, and M. C. Wintersgill. Andes: A coached learning environment for classical Newtonian physics. In *Proceedings of the 11th International Conference on College Teaching and Learning*, Jacksonville, FL, April 2000.

[28] Peter Sember and Ingrid Zukerman. Strategies for generating micro explanations for bayesian belief networks. In *Proceedings of the 5th Workshop on Uncertainty in Artificial Intelligence*, pages 295–302, Windsor, Ontario, 1989.

[29] D.J. Spiegelhalter and R.P. Knill-Jones. Statistical and Knowledge-Based Approaches to Clinical Decision-Support Systems, with an Application in Gastroenterology. *Journal of the Royal Statistical Society. Series A (General)*, 147(1):35–77, 1984.

[30] Henri J. Suermondt. *Explanation in Bayesian belief networks*. PhD thesis, Stanford, CA, USA, 1992.

[31] Henri J. Suermondt and Gregory F. Cooper. An evaluation of explanations of probabilistic inference. *Comput. Biomed. Res.*, 26(3):242–254, 1993.

[32] Wim Wiegerinck. Approximate explanation of reasoning in Bayesian networks. In *Workshop Probabilistic Graphical Models (PGM '04)*, volume 2004, pages 209–216, 2004.

[33] L. Richard Ye and Paul E. Johnson. The impact of explanation facilities on user acceptance of expert systems advice. *MIS Quarterly*, 19(2):157–172, 1995.

[34] J. Y. Zhu and A. Deshmukh. Application of bayesian decision networks to life cycle engineering in green design and manufacturing. *Engineering Applications of AI*, 16(2):91–103, March 2003.