Relevance of Evidence in Bayesian Networks

Michelle Meekes, Silja Renooij, and Linda C. van der Gaag

Utrecht University, The Netherlands
Email: meekes.michelle@gmail.com; {s.renooij,l.c.vandergaag}@uu.nl

Abstract. For many inference tasks in Bayesian networks, computational efforts can be restricted to a relevant part of the network. Researchers have studied the relevance of a network's variables and parameter probabilities for such tasks as sensitivity analysis and probabilistic inference in general, and identified relevant sets of variables by graphical considerations. In this paper we study relevance of the evidence variables of a network for such tasks as evidence sensitivity analysis and diagnostic test selection, and identify sets of variables on which computational efforts can focus. We relate the newly identified sets of relevant variables to previously established relevance sets and address their computation compared to these sets. We thereby paint an overall picture of the relevance of various variable sets for answering questions concerning inference and analysis in Bayesian network applications.

1 Introduction

Bayesian networks have become increasingly popular for decision support in a range of application domains. Capturing general domain knowledge, Bayesian networks owe much of their strength to their ability to derive probability distributions for individual problem instances, given the evidence available from that instance. In view of practical applications however, decision makers should have insight not just in the established probability distributions themselves but in their robustness as well. This observation has motivated researchers to develop techniques for this purpose. The sensitivity of a network's output probabilities to inaccuracies in its parameters can be studied using a parameter sensitivity analysis [2]. A sensitivity-to-evidence analysis allows studying the contribution of specific observations to the output of interest and investigating the effects of changing or removing a particular observation [4]. Algorithms developed for these types of analysis typically rely on (multiple) propagations throughout a network, and hence incur high computational costs.

To relieve the computational burden of probabilistic inference with a Bayesian network, the runtime efforts of computing an output probability of interest can be focused on a relevant part of the network, which depends on the target variables and the specific set of observed variables at hand [3]. This relevant part can to a large extent be identified from graphical considerations only. For example, d-separated nodes and barren nodes are readily identified from a network's graph (we refer to [3] for an overview of available methods) and subsequently pruned without affecting the computed output distribution [1].

Although the concept of relevance has been studied for probabilistic inference in general, it has hardly been addressed in the context of the analyses mentioned above. An exception is the concept of parameter sensitivity set which was introduced to describe the set of variables to which a parameter sensitivity analysis can be restricted [2]. In this paper we will study the relevance of various sets of network variables for answering questions related to evidence, such as sensitivity-to-evidence analyses and test-selection procedures. Where previous relevance studies often focused on a single output variable, we consider in this paper the more general case of a set of target variables; the insights developed will therefore be relevant to MAP and MPE studies as well [7]. We will define three new sets of relevant nodes and show how these relate to existing relevance sets; we further show that these sets can be efficiently determined from a network by graphical considerations only. We thereby provide an overall view of the relevance of both known and newly defined sets of nodes, for answering various types of question related to practical applications of Bayesian networks.

The paper is organised as follows. In Sect. 2 we present some preliminaries. Section 3 introduces our new sets of relevant and irrelevant nodes. In Sect. 4 we show how to efficiently establish these sets, and illustrate their possible application. The paper ends with our concluding remarks in Sect. 5.

2 Preliminaries

A Bayesian network is a concise representation of a joint probability distribution Pr over a set of random variables [5]. It consists of a directed acyclic graph $G = (\mathbf{V}_G, \mathbf{A}_G)$, which captures the random variables as nodes and their interdependencies through arcs; in the sequel we will use the term node to refer to nodes and variables alike. The network further includes a set of conditional probabilities for its parameters, which jointly define the distribution Pr through:

$$\Pr(\mathbf{V}_G) = \prod_{V_i \in \mathbf{V}_G} \Pr(V_i \mid \pi(V_i))$$

where $\pi(V_i)$ denotes the parent set of V_i in the graph. The factorisation of the distribution Pr derives from the well-known concept of d-separation which provides a semantics for the network's graph [8]. For any three disjoint sets of nodes $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subset \mathbf{V}_G$, the set $\mathbf{Z}$ is said to d-separate the sets $\mathbf{X}$ and $\mathbf{Y}$ in G, written $\langle \mathbf{X} \mid \mathbf{Z} \mid \mathbf{Y} \rangle^d_G$, if there do not exist any active chains between $\mathbf{X}$ and $\mathbf{Y}$ given evidence for $\mathbf{Z}$. A chain between two nodes is active if each of its head-to-head nodes is either observed or has an observed descendant, and none of its other nodes are observed. The variables captured by d-separated nodes are considered probabilistically independent.

For computing probabilities of interest from a Bayesian network, general inference algorithms have been designed which derive their efficiency from the d-separation properties of a network's graphical structure. In view of these properties, researchers have studied the computation of an output distribution $\Pr(\mathbf{T} \mid \mathbf{e})$ for a set $\mathbf{T}$ of target nodes given evidence $\mathbf{e}$, and identified sets of nodes whose parameter probabilities are not involved in establishing this distribution. Two well-known examples of such sets are the set of nodes d-separated from $\mathbf{T}$ given $\mathbf{E}$, denoted $DSep(\mathbf{T}, \mathbf{E})$, and the set $Barren(\mathbf{T}, \mathbf{E})$ of barren nodes, where a barren node is a node in $\mathbf{V}_G \backslash (\mathbf{T} \cup \mathbf{E})$ without descendants, or with barren descendants only. These sets of nodes are efficiently established through the Bayes-ball algorithm [9], which runs on the network's graph and does not require probabilistic inference. After pruning these nodes from the graph, a minimal computationally equivalent subgraph results from which the output distribution can be established [1].

3 Defining Sets of (Ir)relevant Nodes

Inspired by the well-known concept of parameter sensitivity set and its role in reducing the computational burden of a parameter sensitivity analysis [2], we develop the concept of evidence sensitivity set as the set of nodes for which a change in observed value, or a change in observational status, may affect a posterior probability distribution of interest.

3.1 Parameter and Evidence Sensitivity Sets

Parameter sensitivity analysis is a well-known technique for studying the possible effects of inaccuracies in the parameter probabilities of a Bayesian network [2]. To reduce the computational burden involved, such an analysis is typically restricted to the parameters of a network which, based upon graphical considerations, cannot be discarded as uninfluential. The concept of parameter sensitivity set was introduced to identify the nodes to which these possibly influential parameter probabilities apply [2]. We briefly review this concept, generalising it to marginal output distributions $\Pr(\mathbf{T} \mid \mathbf{e})$ for sets of target nodes $\mathbf{T}$.

Definition 1. *Let $G = (\mathbf{V}_G, \mathbf{A}_G)$ be the digraph of a Bayesian network. Let $\mathbf{T} \subset \mathbf{V}_G$, $\mathbf{T} \neq \varnothing$, be a set of target nodes and let $\mathbf{E} \subset \mathbf{V}_G \setminus \mathbf{T}$ be a set of evidence nodes in G. The parameter sensitivity set for $\mathbf{T}$ given $\mathbf{E}$ is the set*

$$ParSens(\mathbf{T}, \mathbf{E}) = \{X \in \mathbf{V}_G \mid \neg \langle \{P_X\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle^d_{G^*}\}$$

where G^* is the parented graph of G in which each node X has an additional auxiliary parent P_X.

As described by Coupé and Van der Gaag [2], the parent nodes P_X used for defining the parameter sensitivity set can be viewed as capturing the uncertainty in the parameters for the node X. If this uncertainty is not d-separated from the target nodes, it may affect their probability distribution. The authors proved that a sensitivity analysis can be restricted to this parameter sensitivity set.

While a parameter sensitivity analysis addresses the effects of inaccuracies in a network's parameters, a sensitivity-to-evidence analysis focuses on the effects of changes in the observation entered for a node or in a node's observational status [4]. Similar to the parameter sensitivity set, we now develop the concept of evidence sensitivity set as the set of nodes to which an evidence sensitivity analysis can be restricted. We begin by distinguishing between two types of evidence sensitivity set. The given-evidence sensitivity set consists of all observed nodes for which a change in value or in observational status may affect the probability distribution of interest. While the given-evidence sensitivity set includes observed nodes only, the potential-evidence sensitivity set comprises all yet *un*observed evidence nodes for which obtaining evidence may affect the output distribution. We define the evidence sensitivity sets more formally.

Definition 2. *Let $G, \mathbf{T}, \mathbf{E}$ be as before. Then,*

- *the given-evidence sensitivity set for $\mathbf{T}$ given $\mathbf{E}$ is the set*
$$\mathit{GivEvSens}(\mathbf{T},\mathbf{E}) = \{X \in \mathbf{E} \mid \neg \langle \{X\} \,|\, \mathbf{E}\backslash\{X\} \,|\, \mathbf{T}\rangle_G^d\}$$
- *the potential-evidence sensitivity set for $\mathbf{T}$ given $\mathbf{E}$ is the set*
$$\mathit{PotEvSens}(\mathbf{T},\mathbf{E}) = \mathbf{T} \cup \{X \in \mathbf{V}_G \setminus \mathbf{E} \mid \neg \langle \{X\} \,|\, \mathbf{E} \,|\, \mathbf{T}\rangle_G^d\}$$
- *the evidence sensitivity set for $\mathbf{T}$ given $\mathbf{E}$ is the set*
$$\mathit{EvSens}(\mathbf{T},\mathbf{E}) = \mathit{GivEvSens}(\mathbf{T},\mathbf{E}) \ \cup \ \mathit{PotEvSens}(\mathbf{T},\mathbf{E})$$

We note that the given-evidence sensitivity set contains only nodes from the set $\mathbf{E}$ of observed nodes. If such a node $X \in \mathbf{E}$ is not d-separated from a target node given the remaining evidence, then X and $\mathbf{T}$ may be conditionally dependent, and any change in or removal of the observation for X may affect the output probabilities $\Pr(\mathbf{T} \mid \mathbf{e})$. The potential-evidence sensitivity set on the other hand, contains only nodes which are yet *un*observed. The given-evidence sensitivity set and the potential-evidence sensitivity set thus are disjoint. If an unobserved node $X \notin \mathbf{E}$ is not d-separated from a target node given the available evidence, then X and $\mathbf{T}$ may be conditionally dependent, and entering an observation for X may affect the probabilities $\Pr(\mathbf{T} \mid \mathbf{e})$. Although we could assume that nodes in the target set will never be observed, we include $\mathbf{T}$ in the potential-evidence sensitivity set since observations for nodes in $\mathbf{T}$ most likely affect the probability distribution over the set of target nodes. We further note that all sensitivity sets are defined for a specific $\mathbf{T}$ and $\mathbf{E}$ and may therefore change upon adding or removing an observation. The dynamics involved may then be more complex than just moving nodes between the various sensitivity sets.

Our new concept of evidence sensitivity set is closely related to the concept of parameter sensitivity set, yet is different. The following proposition shows in fact that the parameter sensitivity set is a subset of the evidence sensitivity set.

Proposition 1. *Let $G, \mathbf{T}, \mathbf{E}$ be as before. Then,*

(i) $\mathbf{T} \subseteq \mathit{ParSens}(\mathbf{T},\mathbf{E})$*;*
(ii) $\mathit{ParSens}(\mathbf{T},\mathbf{E}) \cap \mathbf{E} \subseteq \mathit{GivEvSens}(\mathbf{T},\mathbf{E})$*;*
(iii) $\mathit{ParSens}(\mathbf{T},\mathbf{E})\backslash\mathbf{E} \subseteq \mathit{PotEvSens}(\mathbf{T},\mathbf{E})$*.*

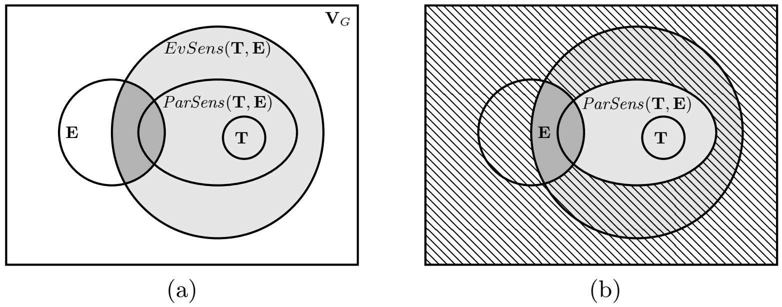

Fig. 1: The relations between the various sets. (a) The light grey area indicates the potential-evidence sensitivity set, whereas the dark grey constitutes the given-evidence sensitivity set; (b) the hatched area represents all nodes that can be pruned, i.e. d-separated nodes, barren nodes and irrelevant evidence nodes

Proof. (i) We consider adding an auxiliary parent P_T to a target node $T \in \mathbf{T}$. Since P_T and T are directly connected, we have that $\neg \langle \{P_T\} \mid \mathbf{E} \mid \mathbf{T} \rangle^d_{G^*}$. Therefore, $T \in ParSens(\mathbf{T},\mathbf{E})$ by definition.

(ii) We consider a node $X \in ParSens(\mathbf{T},\mathbf{E}) \cap \mathbf{E}$. Since $X \in \mathbf{E}$, the auxiliary parent P_X has active chains only to other parents of X. Since $X \in ParSens(\mathbf{T},\mathbf{E})$, at least one such parent Y must have an active chain to a target node. As a result, X cannot be d-separated from $\mathbf{T}$ given $\mathbf{E}\backslash\{X\}$.

(iii) We consider a node $X \in ParSens(\mathbf{T},\mathbf{E})\backslash\mathbf{E}$. Since P_X is not d-separated from $\mathbf{T}$ given $\mathbf{E}$ in G^*, there must be an active chain from X to $\mathbf{T}$ in G. □

A schematic summary of the above properties is given in Fig. 1(a). The dark grey area represents the intersection of the sensitivity sets with the set of observed nodes $\mathbf{E}$; this area thus coincides with the given-evidence sensitivity set. The light grey area constitutes the potential-evidence sensitivity set; the diagram shows that the set of target nodes is a subset of this set.

3.2 Ignoring Irrelevant Evidence Nodes

The concept of computationally equivalent subgraph was introduced to describe a subgraph of a Bayesian network, with its associated parameters, from which the correct output distribution over the network's target variables can be established. Although the minimal computationally equivalent subgraph identified by Baker and Boult [1], contains no nodes $X \notin \mathbf{T} \cup \mathbf{E}$ that can be pruned, it may contain evidence nodes that are d-separated from the target nodes given the remaining evidence, that is, it may contain evidence nodes outside $GivEvSens(\mathbf{T},\mathbf{E})$. The identified subgraph thus is not minimal in the sense that no proper subgraph exists from which the output distribution can be correctly established. For the sake of completeness, we define the set of irrelevant evidence nodes and explicitly state the property that these nodes cannot affect the output distribution.

Definition 3. *Let $G, \mathbf{T}, \mathbf{E}$ be as before. Then, the irrelevant evidence set for $\mathbf{T}$ given $\mathbf{E}$ equals*

$$IrrEv(\mathbf{T}, \mathbf{E}) = \{E \in \mathbf{E} \mid \langle \{E\} \,|\, \mathbf{E} \backslash \{E\} \,|\, \mathbf{T} \rangle_G^d \}$$

The irrelevant evidence nodes together constitute a set of nodes whose parameter probabilities indeed are not required for computing $\Pr(\mathbf{T} \mid \mathbf{E})$.

Proposition 2. *Let G,$\mathbf{T}$,$\mathbf{E}$ be as before. Then,* $\Pr(\mathbf{T} \,|\, \mathbf{E}) = \Pr(\mathbf{T} \,|\, \mathbf{E} \backslash IrrEv(\mathbf{T}, \mathbf{E}))$.

Proof. Assuming that $\Pr(\mathbf{E})$ is strictly positive, the proposition is proven by repeated application of the intersection property of independence relations. $\square$

From the above property we conclude that the minimal computationally equivalent subgraph can be further pruned by removing all nodes from $IrrEv(\mathbf{T}, \mathbf{E})$. Moreover, since $IrrEv(\mathbf{T}, \mathbf{E}) = \mathbf{E} \setminus GivEvSens(\mathbf{T}, \mathbf{E})$, the proposition also shows that $\Pr(\mathbf{T} \mid \mathbf{E})$ can indeed be correctly computed by restricting the set of evidence nodes to $GivEvSens(\mathbf{T}, \mathbf{E})$.

3.3 Relating the Different Sets

We now establish the relationship between the various sensitivity sets and well-known types of irrelevant node.

Proposition 3. *Let $G, \mathbf{T}, \mathbf{E}$ be as before. Then,*

(i) $DSep(\mathbf{T}, \mathbf{E}) = V_G \setminus (EvSens(\mathbf{T}, \mathbf{E}) \cup \mathbf{E})$;
(ii) $Barren(\mathbf{T}, \mathbf{E}) = PotEvSens(\mathbf{T}, \mathbf{E}) \setminus ParSens(\mathbf{T}, \mathbf{E})$;

Proof. (i) The set $DSep(\mathbf{T}, \mathbf{E})$ contains all nodes $X \in V_G \backslash (\mathbf{T} \cup \mathbf{E})$ such that $\langle \{X\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle_G^d$. By definition, this set equals $\mathbf{V}_G \setminus (PotEvSens(\mathbf{T}, \mathbf{E}) \cup \mathbf{E})$, and corresponds with the white area *outside* the circles in Fig. 1(a).

(ii) Barren nodes are unobserved nodes that are not d-separated from the target nodes given the evidence, yet are not involved in the computation of the output distribution over these nodes; once observed however (directly or indirectly), barren nodes can become computationally relevant. $\square$

For computing the output distribution $\Pr(\mathbf{T} \mid \mathbf{E})$ over the target nodes $\mathbf{T}$ of a Bayesian network, we can safely prune all nodes from $DSep(\mathbf{T}, \mathbf{E}) \cup Barren(\mathbf{T}, \mathbf{E}) \cup IrrEv(\mathbf{T}, \mathbf{E})$. From the above proposition we find that this set equals:

$$\mathbf{V}_G \setminus (ParSens(\mathbf{T}, \mathbf{E}) \cup (\mathbf{E} \backslash IrrEv(\mathbf{T}, \mathbf{E}))),$$

where $\mathbf{E} \backslash IrrEv(\mathbf{T}, \mathbf{E})$ equals $GivEvSens(\mathbf{T}, \mathbf{E})$. We can therefore prune all nodes from $\mathbf{V}_G$ except those in $ParSens(\mathbf{T}, \mathbf{E}) \cup GivEvSens(\mathbf{T}, \mathbf{E})$; this set of nodes is illustrated in Fig. 1(b).

4 Identifying (Ir)relevant Nodes

Many Bayesian network properties can be recognised just by inspecting the network's graph. Core d-separation statements, for example, can be verified in time linear to the size of the graph. A well-known algorithm for this purpose is the *Bayes-ball algorithm* [9]. Although this algorithm was not designed for establishing sensitivity sets as defined in the previous section, we will demonstrate that the information maintained by the algorithm suffices for identifying these sets. We will subsequently illustrate all concepts introduced in this paper, as well as their potential use, by means of an example.

4.1 Bayes-ball for Sensitivity Sets

The Bayes-ball algorithm was designed to identify various sets of relevant and irrelevant nodes. The algorithm explores the graph of a Bayesian network in view of an output distribution $\Pr(\mathbf{T} \mid \mathbf{E})$ over its target nodes. It starts from these target nodes and "bounces a ball" through the graph, respecting d-separation properties. Visited nodes are marked as such, and in addition get a top or bottom mark when their parents or children, respectively, are scheduled for a visit. Initially, all target nodes are marked on top and at the bottom; evidence nodes, if visited, can receive a top mark only. After exploring the graph, the algorithm establishes the following sets of nodes, based on the marks received:

- $N_i(\mathbf{T} \mid \mathbf{E}) = \{X \in \mathbf{V}_G \mid X \text{ is not marked at the bottom}\}$;
- $N_p(\mathbf{T} \mid \mathbf{E}) = \{X \in \mathbf{V}_G \mid X \text{ is marked on top}\}$;
- $N_e(\mathbf{T} \mid \mathbf{E}) = \{X \in \mathbf{E} \mid X \text{ is marked as visited}\}$.

The set $N_i(\mathbf{T} \mid \mathbf{E})$, termed the set of *irrelevant nodes*, includes all d-separated nodes [9]. We note that the algorithm includes all evidence nodes in the set of irrelevant nodes as well. Evidence nodes in general are not irrelevant to the computation at hand, however, with the exception of nodes in $\mathit{IrrEv}(\mathbf{T}, \mathbf{E})$. The set $N_p(\mathbf{T} \mid \mathbf{E})$, called the set of *requisite probability nodes*, includes the nodes whose parameters are needed for the computation of the output probability; we note that the adjective 'requisite' refers to the node's parameter probabilities. The set $N_e(\mathbf{T} \mid \mathbf{E})$, coined the set of *requisite observation nodes*, includes the evidence nodes whose value is required for the computations. A computationally equivalent subgraph for the computations can now be obtained from the original Bayesian network by pruning all nodes outside the set $N_p(\mathbf{T} \mid \mathbf{E}) \cup N_e(\mathbf{T} \mid \mathbf{E})$.

The sensitivity sets defined and reviewed in the previous section can be readily identified from the information recorded by the Bayes-ball algorithm as stated in the following proposition; for a formal proof of the proposition, we refer to [6].

Proposition 4. *Let* $G, \mathbf{T}, \mathbf{E}$ *be as before. Consider running Bayes-ball on* G *with respect to* $\Pr(\mathbf{T} \mid \mathbf{E})$. *Then,*

- $\mathit{ParSens}(\mathbf{T}, \mathbf{E}) = \{X \in \mathbf{V}_G \mid X \textit{ is marked on top}\}$;

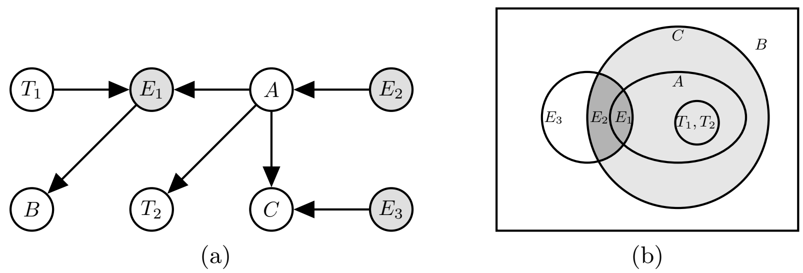

Fig. 2: (a) The digraph of the example Bayesian network, and (b) the sensitivity sets for $\Pr(\mathbf{T} \mid \mathbf{E})$

- $\mathit{GivEvSens}(\mathbf{T}, \mathbf{E}) = \{X \in \mathbf{E} \mid X$ *is marked as visited*$\}$;
- $\mathit{PotEvSens}(\mathbf{T}, \mathbf{E}) = \{X \in \mathbf{V}_G \mid X$ *is marked at the bottom*$\}$;
- $\mathit{IrrEv}(\mathbf{T}, \mathbf{E}) = \{X \in \mathbf{E} \mid X$ *is not marked as visited*$\}$.

The proposition reveals that the different sensitivity sets identified in the previous section actually provide alternative semantics for the three Bayes-ball sets:

- $N_i(\mathbf{T} \mid \mathbf{E}) = \mathbf{V}_G \setminus \mathit{PotEvSens}(\mathbf{T}, \mathbf{E})$;
- $N_p(\mathbf{T} \mid \mathbf{E}) = \mathit{ParSens}(\mathbf{T}, \mathbf{E})$;
- $N_e(\mathbf{T} \mid \mathbf{E}) = \mathit{GivEvSens}(\mathbf{T}, \mathbf{E})$.

4.2 An Example

To illustrate the use of the various relevance sets introduced in Section 3, we consider an example Bayesian network defining a joint probability distribution over eight nodes; the graph of the network is shown in Fig. 2(a). For the network, we consider output probability distributions $\Pr(\mathbf{T} \mid \mathbf{E})$ for the target nodes $\mathbf{T} = \{T_1, T_2\}$ given observations for the evidence nodes $\mathbf{E} = \{E_1, E_2, E_3\}$. Using Bayes-ball, we find the following sensitivity sets, summarised in Fig. 2(b):

- $\mathit{ParSens}(\mathbf{T}, \mathbf{E}) = \{A, E_1, T_1, T_2\}$;
- $\mathit{PotEvSens}(\mathbf{T}, \mathbf{E}) = \{A, C, T_1, T_2\}$;
- $\mathit{GivEvSens}(\mathbf{T}, \mathbf{E}) = \{E_1, E_2\}$.

We recall that the parameter sensitivity set was designed to describe the nodes in a Bayesian network whose parameter inaccuracies may affect the output probabilities from the network. For the example network, we conclude that only changes in the parameter probabilities of the nodes A, E_1, T_1 and T_2 may influence the probabilities $\Pr(\mathbf{T} \mid \mathbf{E})$. A parameter sensitivity analysis may thus be restricted to the parameters for these nodes and forego variation of the parameters of the nodes B, C, E_2 and E_3, that is, the parameters of only half of the nodes need be investigated upon the analysis.

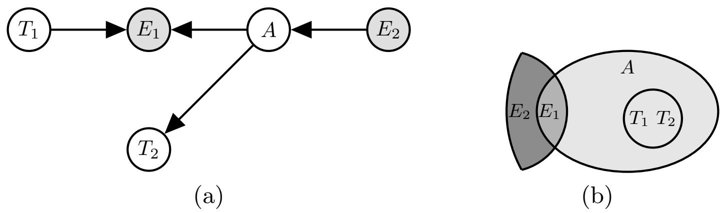

Fig. 3: (a) Graph of the example network after pruning all nodes outside $\mathit{ParSens}(\mathbf{T}, \mathbf{E}) \cup (\mathbf{E} \backslash \mathit{IrrEv}(\mathbf{T}, \mathbf{E}))$, and (b) a summary of the pruning result

The evidence sensitivity set captures the nodes for which a change of value or in observational status may affect the output probabilities established from the network. In the example network, only the evidence nodes E_1 and E_2 are contained in the given-evidence sensitivity set for $\mathbf{T}$ given $\mathbf{E}$. Changing their observed value, or removing their observations, may therefore change the output probabilities $\Pr(\mathbf{T} \mid \mathbf{E})$. Since node E_3 is not included in the given-evidence sensitivity set, we can change or remove its observation, without affecting the output probabilities. The network's output therefore is robust against an inaccurate observation for E_3. The potential-evidence sensitivity set for $\mathbf{T}$ given $\mathbf{E}$ provides information about the effects of additional evidence. The potential-evidence sensitivity set established from the example network shows that obtaining additional evidence for one of the nodes A, C, T_1 and T_2 may change the output. Since node B is not in the set, gathering an observation for this node cannot change the current distribution over $\mathbf{T}$. We note that such a finding may be exploited by a test-selection procedure. In fact, test selection can focus on collecting evidence for the nodes A and C, if appropriate.

The resulting computationally equivalent subgraph is shown in Fig. 3(a), along with a schematic summary of the roles of the remaining nodes (b). We would like to emphasize that as a result of the dynamics of the various sets upon changes in the observational status of nodes, a change in $\mathbf{E}$ may require a differently pruned graph. For example, if an observation would be entered for node C, the parameters of node E_3 would no longer be immaterial for the output distribution over the target nodes. We further note that the various sets of relevant nodes identified above can be instrumental in focusing the efforts of a large variety of inference tasks. We note for example that the above conclusions also pertain to MAP computations, that is, for establishing $\mathrm{MAP}(\mathbf{T}, \mathbf{e}) = \mathrm{argmax}_{\mathbf{t}} \Pr(\mathbf{t} \mid \mathbf{e})$ for a specific assignment $\mathbf{e}$ to $\mathbf{E}$: taking the output of the Bayes-ball algorithm in fact, we know that we can safely prune the nodes $\{B, C, E_3\}$ from the network without affecting the established MAP.

5 Conclusions and Future Research

Relevance of the nodes of a Bayesian network had so far been studied primarily in the context of probabilistic inference. In this paper we focused on a network's evidence nodes and addressed their relevance for such tasks as sensitivity-to-evidence analysis and diagnostic test selection. To this end, we defined two types of evidence sensitivity set and studied the relationships between these sets and with previously known sets of (ir)relevant nodes. We thereby presented a more complete picture of the relevance of various node sets for answering questions concerning inference and analysis in Bayesian network applications. By demonstrating that our evidence sensitivity sets can be determined from the well-known Bayes-ball algorithm, moreover, we provided an efficient way of establishing these sets from graphical considerations only.

The various relevance sets discussed in this paper are not static in a Bayesian network application, but will change dynamically as the set of observed nodes changes. More extensive sensitivity-to-evidence analyses and test-selection procedures therefore entail re-establishing the relevance sets after each change in observational status. In the near future we would like to study the dynamics involved and investigate whether we can predict, at least partly, how these sets will change without having to re-invoke the Bayes-ball algorithm. We would further like to extend our investigations and study the concept of relevance for yet other computational tasks in Bayesian network applications.

References

1. M. Baker, T. Boult. Pruning Bayesian networks for efficient computation. In: P.P. Bonissone, M. Henrion, L.N. Kanal, and J.F. Lemmer (eds) *Uncertainty in Artificial Intelligence 6*, pp. 225-232. Elsevier Science, Amsterdam, 1991.
2. V.M.H. Coupé, L.C. van der Gaag. Properties of sensitivity analysis of Bayesian belief networks. *Annals of Mathematics and Artificial Intelligence*, vol. 36(4), pp. 323-356, 2002.
3. M. Druzdzel, H. Suermondt. Relevance in probabilistic models: "backyards" in a "small world". In: *Working Notes of the AAAI 1994 Fall Symposium Series: Relevance*, pp. 60-63, 1994.
4. F.V. Jensen, S.A. Aldenryd, K.B. Jensen. Sensitivity analysis in Bayesian networks. In: C. Froidevaux, J. Kohlas (eds) *Proceedings of the European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, Lecture Notes in Computer Science 946, pp. 243-250, Springer, 1995.
5. F.V. Jensen, T.D. Nielsen. *Bayesian Networks and Decision Graphs*. Springer, 2007.
6. M. Meekes. *Sensitivity to Evidence in Probabilistic Networks*. Master's thesis, Faculty of Science, Utrecht University, 2013.
7. J. Park, A. Darwiche. Complexity results and approximation strategies for MAP explanations. *Journal of Artificial Intelligence Research*, vol. 21, pp. 101-133, 2004.
8. J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers, 1988.
9. R. Shachter. Bayes-ball: The rational pastime (for determining irrelevance and requisite information in belief networks and influence diagrams). In: G. Cooper, S. Moral (eds) *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, pp. 480-487, Morgan Kaufmann, San Francisco, 1998.