Universiteit Utrecht

Department of Information and Computing Sciences

Sensitivity to Evidence in Probabilistic Networks

Master's thesis
ICA-3151182

Author:
Michelle Meekes
meekes.michelle@gmail.com

Supervisors:
prof. dr. ir. Linda C. van der Gaag
dr. Silja Renooij

October 15, 2013

Contents

Chapter 1

Introduction

The probabilistic network framework is an approach to apply probability theory to reasoning with uncertainty in knowledge-based systems. It is characterized by a powerful formalism for representing a joint probability distribution on a set of statistical variables. Such a network consists of a qualitative part, a directed acyclic graph, and a quantitative part, conditional probability distributions for every node given its parents'. The conditional probabilities are also called parameters.

After the construction of a Bayesian network, several analysis techniques can be used to measure the robustness and the reliability of the network. A probabilistic network is constructed to solve problems of reasoning under uncertainty, and we would like the network to represent the domain as precise as possible, and that it behaves the way we think it should. It is therefore important to gain insight in *how* a certain probability of interest is affected by changes in, for example, certain parameter values or in the observations. The analysis that can be done to this end, is called *sensitivity analysis*. In a mathematical model, this basically means that one or more parameters are being varied stepwise in order to study the influence on the output of the model in each step. In probabilistic networks, we can distinguish between different types of sensitivity analysis.

The parameters associated with the nodes in the probabilistic network can be inaccurate, which may influence the reliability of the probability of interest. By applying *parameter sensitivity analysis* to the network, this reliability can be investigated. As mentioned in [4], the straightforward way of parameter sensitivity analysis, where every parameter is varied to study its influence on the output, is highly time-consuming. Namely, the number of different parameters is exponential in the number of variables in the network. Luckily, not every parameter has to be varied. By only taking graphical properties of the probabilistic network into account, the qualitative part, enough information is available to determine the *(parameter) sensitivity set*. This set of nodes consists of all nodes whose parameters upon variation may influence the probability of interest. For a certain probability of interest, the *(parameter) sensitivity function* can be studied, for example in a one-way (see for example [4]) or two-way (see [2]) analysis. One way to compute the parameter sensitivity function can be found in [4].

Besides performing parameter sensitivity analysis, it can also be useful to perform *evidence sensitivity analysis.* Evidence sensitivity analysis (or SE Analysis) is the analysis of how sensitive the probability of interest is to variations in the set of evidence [6]. There exist some measures and ways to describe the sensitivity to evidence in a probabilistic network. An introduction in this field is given in [5] and [6]. Applications of evidence sensitivity analysis can be found in, for example, [14], where the sensitivity of human fatigue in the marine industry to observations, such as the weather conditions, hours awake and alcohol consumption, is analyzed, and in [3], where the sensitivity of a fruit fly outbreak to different observations, such as amount of trapped fruit flies and location properties, is analyzed.

Whereas researchers have studied the properties of parameter sensitivity analysis for probabilistic networks to quite some extent [1], evidence sensitivity analysis has received far less attention. The aim of this thesis is to present new, fundamental insights on sensitivity to evidence in probabilistic networks. We will not only identify *which* nodes may influence the probability of interest upon a change in the observation, by defining the *evidence sensitivity set*, but also *how*, by defining the *evidence sensitivity function.*

The thesis is organized as follows. In Chapter 2, we will formally define a probabilistic network, and introduce the notations and terminology needed in the remainder of this thesis. In Chapter 3, we will discuss the parameter sensitivity set, and we will introduce the evidence sensitivity set, which consists of all nodes for which a change in observed value, or a change in status of being observed or not, may influence the probability of interest. We will also propose an algorithm to identify the sensitivity sets. In Chapter 4, we use results from previous studies of probabilistic network pruning to look for simplifications in the probability calculation. In Chapter 5, we combine the results of Chapter 4 in order to extend our results of Chapter 3 by the definition and computation of the evidence sensitivity function, which describes *how* a single observation upon variation influences the probability of interest. Finally, in Chapter 6, we outline our results and conclusions as well as options for further research.

[1] An overview of research results of parameter sensitivity analysis can be found in [10].

Chapter 2

Preliminaries

We will present in this chapter the notations and terminology which we will use in the remainder of this thesis.

2.1 Probabilistic network definition

A probabilistic network consists of a qualitative part, which is a directed acyclic graph G, and a quantitative part, which is a set of conditional probability distributions Γ associated with G.

The set of nodes in the graph G is denoted $V(G)$, and its set of arcs $A(G)$. In view of probabilistic networks, we use the terms "variable" and "node" interchangeably. Every node $X \in V(G)$ stands for a discrete random variable. The set of possible values for X, its domain, will also be denoted X. An instantiation of X will be denoted by a small letter, x. A set of nodes will be denoted by a bold capital, for example $\mathbf{X} \subseteq V(G)$. $\mathbf{X}$ and $\mathbf{x}$ are the domain and a joint value assignment, or instantiation, to all variables in $\mathbf{X}$, respectively.

By $\rho(\mathbf{X})$ we will mean the set of direct ancestors (or parents) of the set of nodes $\mathbf{X} \subseteq V(G)$, i.e.

$$\rho(\mathbf{X}) = \{Y \mid \exists X \in \mathbf{X} : (Y \to X) \in A(G)\}.$$

The set of direct ancestors of a single node X, $\rho(\{X\})$, will be denoted $\rho(X)$ for short. Furthermore, the set of all ancestors of $\mathbf{X}$ will be denoted $\rho^*(\mathbf{X})$.

By $\delta(\mathbf{X})$ we will mean the set of direct descendants (or children) of the set of nodes $\mathbf{X} \subseteq V(G)$, i.e.

$$\delta(\mathbf{X}) = \{Y \mid \exists X \in \mathbf{X} : (X \to Y) \in A(G)\}.$$

The set of direct descendants of a single node X, $\delta(\{X\})$, will be denoted $\delta(X)$ for short. Furthermore, the set of all descendants of $\mathbf{X}$ will be denoted $\delta^*(\mathbf{X})$.

We define a set of nodes $\mathbf{X} \subseteq V(G)$ together with their children $\delta(\mathbf{X})$, as

$$\mathcal{D}(\mathbf{X}) = \mathbf{X} \cup \delta(\mathbf{X}),$$

the *'donna con bambini'* of $\mathbf{X}$. The donna con bambini of a single node X, $\mathcal{D}(\{X\})$, will be denoted $\mathcal{D}(X)$ for short. We define a set of nodes $\mathbf{X} \subseteq V(G)$ together with their children $\delta(\mathbf{X})$ and all their parents $\rho(\mathbf{X} \cup \delta(\mathbf{X}))$ as

$$\mathcal{MB}(\mathbf{X}) = \mathcal{D}(\mathbf{X}) \cup \rho(\mathcal{D}(\mathbf{X})),$$

the well-known *Markov blanket* of $\mathbf{X}$. The Markov blanket of a single node X, $\mathcal{MB}(\{X\})$, will be denoted $\mathcal{MB}(X)$ for short.

We now define a probabilistic (or Bayesian) network as follows.

Definition 2.1.1 (Probabilistic network)**.** *A probabilistic (or Bayesian) network consists of a tuple $B = (G, \Gamma)$, where*

- $G = (V(G), A(G))$ *is a directed acyclic graph which consists of nodes $V(G)$ and arcs $A(G)$, and*
- $\Gamma = \{\Pr(X \mid \rho(X)) \mid X \in V(G)\}$ *is a set of conditional probability distributions associated with each node $X \in V(G)$.*

The graph G of a probabilistic network represents the independences in the joint probability distribution on a set of random variables. The conditional probabilities associated with each node are also called its *parameters*. These parameters describe the strength of the probabilistic relationship of X with all its parents in the directed acyclic graph.

The *probabilistic network framework* was first introduced in 1988 by Pearl [11], as an approach to apply probability theory for reasoning with uncertainty in knowledge-based systems. Different inference algorithms can be used to make probabilistic statements concerning the variables that are represented in the network: any prior or posterior probability of interest over these variables can essentially be computed. In its most general form, with no further restrictions on the probabilistic network, probabilistic inference costs exponential time. This thesis will not go into the details of available inference algorithms, such as Pearl's algorithm [11] and the algorithm by Lauritzen and Spiegelhalter [8].

The prior joint probability of a probabilistic network $B = (G, \Gamma)$ is defined by

$$\Pr(V(G)) = \prod_{X \in V(G)} \Pr(X \mid \rho(X)). \tag{2.1}$$

For any subset $\mathbf{Y} \subseteq V(G)$, the marginal probability $\Pr(\mathbf{Y})$ is defined by

$$\Pr(\mathbf{Y}) = \sum_{V(G) \setminus \mathbf{Y}} \Pr(V(G)). \tag{2.2}$$

In this thesis, we will be interested in computing from a probabilistic network a probability distribution $\Pr(\mathbf{T} \mid \mathbf{E})$ of interest, for the target node set $\mathbf{T}$ and the evidence node set $\mathbf{E}$. We will assume no overlap between the node sets, i.e. $\mathbf{T} \cap \mathbf{E} = \emptyset$. For calculating a *probability of interest* of the form $\Pr(\mathbf{t} \mid \mathbf{e})$, the values $\mathbf{t}$ of the target node set $\mathbf{T}$ and, in case of a posterior probability of interest, the values $\mathbf{e}$ of nodes in the evidence node set $\mathbf{E}$ are given. The values $\mathbf{e}$ of the evidence nodes are also called "observations". In this thesis $\mathbf{T}$ stands for the most general case, where it is a set of nodes instead of a single node.

Probabilistic independence can be read from the probabilistic network's digraph, by means of the d-separation criterion. This criterion distinguishes two special node sequences; the following definitions are taken from [15].

Definition 2.1.2 (chain)**.** *Let $G = (V(G), A(G))$ be a directed acyclic digraph. A chain from node $X_0 \in V(G)$ to $X_k \in V(G)$ is a sequence of nodes $X_0, ..., X_k, k \geq 0$, with distinct arcs $(X_{i-1} \rightarrow X_i)$ or $(X_{i-1} \leftarrow X_i) \in A(G)$, $i = 1, ..., k$, between them.*

Definition 2.1.3 (active chain)**.** *Let $G = (V(G), A(G))$ be an acyclic digraph. Let s be a chain between $X \in V(G)$ and $Y \in V(G)$. The chain s is active given $\mathbf{Z} \subseteq V(G)$, if*

- *every node with two incoming arcs on s is or has a descendant in $\mathbf{Z}$, and*
- *all other nodes on the chain are not in $\mathbf{Z}$.*

If a chain is *not* active, it is said to be *blocked*. In a probabilistic network, we have that if there exists an active chain given $\mathbf{Z} \subseteq V(G)$ between nodes $X \in V(G)$ and $Y \in V(G)$, they *may* be dependent given $\mathbf{Z}$. On the other hand, if there does not exist such a chain, X and Y are conditionally independent given $\mathbf{Z}$. In this case, the two nodes are said to be *d-separated*. The following definition formalizes this notion in general, for sets of nodes, using the notation introduced by Pearl [11].

Definition 2.1.4 (d-separation)**.** *Let $G = (V(G), A(G))$ be a directed acyclic graph. Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq V(G)$ be mutually disjoint sets of nodes in G. The set of nodes $\mathbf{Z}$ is said to d-separate the sets of nodes $\mathbf{X}$ and $\mathbf{Y}$ in G, denoted as $\langle \mathbf{X} \,|\, \mathbf{Z} \,|\, \mathbf{Y} \rangle^d_G$, if for each node $X \in \mathbf{X}$ and each node $Y \in \mathbf{Y}$ there exists no active chain from X to Y in G given $\mathbf{Z}$.*

If the set of nodes $\mathbf{Z}$ d-separates the sets of nodes $\mathbf{X}$ and $\mathbf{Y}$, then $\mathbf{X}$ and $\mathbf{Y}$ are conditionally independent given $\mathbf{Z}$, i.e.

$$\Pr(\mathbf{X} \wedge \mathbf{Y} \mid \mathbf{Z}) = \Pr(\mathbf{X} \mid \mathbf{Z}) \cdot \Pr(\mathbf{Y} \mid \mathbf{Z}).$$

Or, equivalently,

$$\Pr(\mathbf{X} \mid \mathbf{Y} \wedge \mathbf{Z}) = \Pr(\mathbf{X} \mid \mathbf{Z}).$$

If two sets of nodes are *not* d-separated, they are said to be *d-connected*. We want to emphasize that Pearl only defined d-separation for mutually disjoint sets of nodes. Therefore, we will use the definition only in mutually disjoint cases. Figure 2.1 illustrates our visual representation of a graph, observations and d-separation.

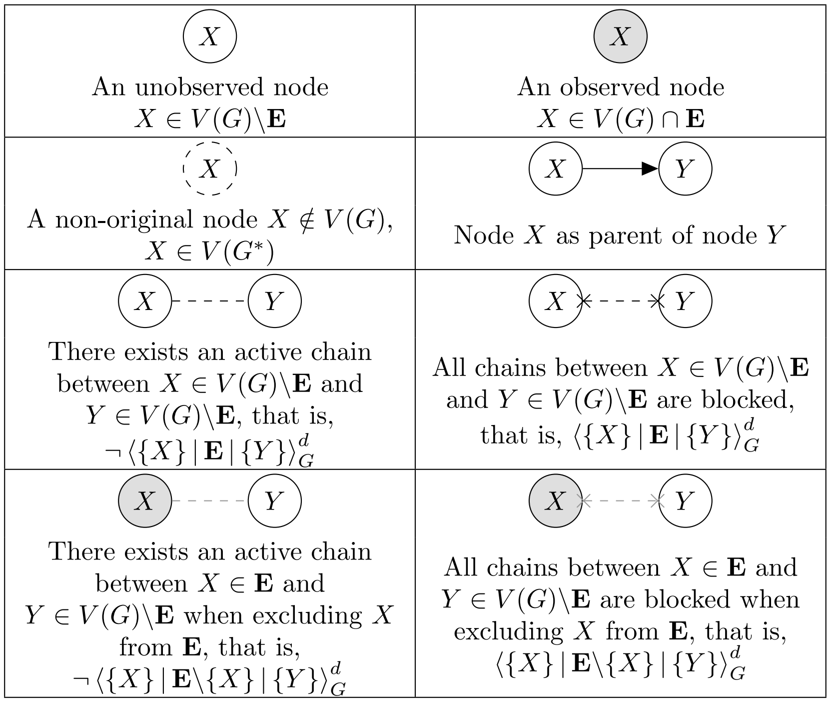


Figure 2.1: Explanation of the depiction of nodes and arcs of $G = (V(G), A(G))$.

Chapter 3

Sensitivity Sets

As mentioned in Chapter 1, upon performing a parameter sensitivity analysis, it is not necessary to subject the whole probabilistic network to a numerical analysis. By investigating just the qualitative part of the network, it is possible to determine the sets of nodes to which the probability of interest may be sensitive and to which it is not.

In the first section a recap of the *parameter sensitivity set* will be given. The parameter sensitivity set contains all variables whose parameters may, upon variation, affect the posterior probability distribution of interest for target variable set $\mathbf{T}$ given evidence for the nodes $\mathbf{E}$.

The second section will introduce the *evidence sensitivity set*. This set contains all nodes for which a change in observed value, or a change in status of being observed or not, may influence the probability of interest. These two sets are closely related, as will become clear at the end of the second section.

The third section presents an algorithm to identify efficiently the two types of sensitivity set in the network.

3.1 The parameter sensitivity set

The parameter sensitivity set contains all variables whose parameters may upon variation affect the posterior probability distribution of interest for target variable set $\mathbf{T}$ given evidence node set $\mathbf{E}$. To formalize this concept, we will use an extended version of the original graph G, namely the graph with an extra parent node added to every node in G, as in [4].

Definition 3.1.1 (Parented graph)**.** *Let $G = (V(G), A(G))$, $\mathbf{T}$ and $\mathbf{E}$ be as defined before. Let $G^* = (V(G^*), A(G^*))$ be the digraph that is constructed by* *adding a parent node P_X to every node $X \in V(G)$, such that*

$$V(G^*) = V(G) \cup \{P_X \mid X \in V(G)\},$$
$$A(G^*) = A(G) \cup \{(P_X \to X) \mid X \in V(G)\}.$$

G^* is called the parented graph of G.

Here follows the definition of the parameter sensitivity set.

Definition 3.1.2 (Parameter sensitivity set)**.** *Let $G = (V(G), A(G))$, $\mathbf{T}$ and $\mathbf{E}$ be as defined before. Let G^* be the parented graph of G. The parameter sensitivity set $ParSens^{\mathbf{E}}(\mathbf{T})$ for $\mathbf{T}$ given $\mathbf{E}$ contains all variables $X \in V(G)$ for which we have that*

$$\neg \langle \{P_X\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle^d_{G^*}$$

To show that the parameters of the nodes contained in the parameter sensitivity set may indeed influence the probability of interest upon variation, we state the following proposition, from [4].

Proposition 3.1.1. *Let $G = (V(G), A(G))$, $\mathbf{T}$ and $\mathbf{E}$ be as defined before. Let $ParSens^{\mathbf{E}}(\mathbf{T})$ be the parameter sensitivity set for $\mathbf{T}$ given $\mathbf{E}$ as defined in Definition 3.1.2. Then, for every node $X \notin ParSens^{\mathbf{E}}(\mathbf{T})$, we have that $\Pr(\mathbf{T} \mid \mathbf{E})$ is insensitive to changes in the parameter values $\Pr(X \mid \rho(X))$.*

A proof of Proposition 3.1.1 can be found in [4]. Intuitively, the stated property can be understood as follows. An added parent node P_X can be seen as a representation of the possible inaccuracy in the conditional probability distributions associated with node X. If this inaccuracy is not d-separated from the target variable set, then varying the parameter values of these node may influence the probabilities of interest $\Pr(\mathbf{T} \mid \mathbf{E})$.

The following lemma now states that the property $\neg \langle \{P_T\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle^d_{G^*}$ always holds for all $T \in \mathbf{T}$. In other words, variations in the parameter values of nodes in $\mathbf{T}$ may influence the probabilities of interest $\Pr(\mathbf{T} \mid \mathbf{E})$.

Lemma 3.1.1. *Let $G = (V(G), A(G))$, $\mathbf{T}$ and $\mathbf{E}$ be as defined before. Let $ParSens^{\mathbf{E}}(\mathbf{T})$ be the parameter sensitivity set for $\mathbf{T}$ given $\mathbf{E}$. Then, we have that*

$$\mathbf{T} \subseteq ParSens^{\mathbf{E}}(\mathbf{T}).$$

Proof. Adding a parent to a target node implies directly a d-connection, therefore it holds that $\mathbf{T} \subseteq ParSens^{\mathbf{E}}(\mathbf{T})$. □

3.2 The evidence sensitivity set

We will now define the *evidence sensitivity set*, the set that contains all variables for which a change in observed value, or a change in status of being observed or not, may affect the probability of interest. Our concept is inspired by the parameter sensitivity set and its role in parameter sensitivity analysis. Identifying all evidence nodes that upon change in observation may influence the posterior probability of interest is a fist step towards studying *how* certain observations influence the posterior probability of interest.

We begin by distinguishing between two types of evidence sensitivity set. The *given evidence sensitivity set* consists of all nodes $X \in \mathbf{E}$ for which a change in the observed value, or removal of this value, may influence the probability distribution of interest.

Definition 3.2.1 (Given evidence sensitivity set)**.** *Let $G = (V(G), A(G))$, $\mathbf{T}$ and $\mathbf{E}$ be as defined before. The given evidence sensitivity set $GivEvSens^{\mathbf{E}}(\mathbf{T})$ for $\mathbf{T}$ given $\mathbf{E}$ contains all variables $X \in \mathbf{E}$ for which we have that*

$$\neg \langle \{X\} \,|\, \mathbf{E} \backslash \{X\} \,|\, \mathbf{T} \rangle^d_G$$

To show that changing the observed value or removing this value for a variable from the set $GivEvSens^{\mathbf{E}}(\mathbf{T})$ may indeed influence the probability of interest, we state the following proposition.

Proposition 3.2.1. *Let $G = (V(G), A(G))$, $\mathbf{T}$ and $\mathbf{E}$ be as defined before. Let $GivEvSens^{\mathbf{E}}(\mathbf{T})$ be the given evidence sensitivity set for $\mathbf{T}$ given $\mathbf{E}$ as defined in Definition 3.2.1. Then,*

1. *for every node $X \notin GivEvSens^{\mathbf{E}}(\mathbf{T})$, we have that $\Pr(\mathbf{T} \mid \mathbf{E})$ is insensitive to changes in or removal of the observation of node X.*
2. *for every node $X \in GivEvSens^{\mathbf{E}}(\mathbf{T})$, we have that $\Pr(\mathbf{T} \mid \mathbf{E})$ may be sensitive to changes in or removal of the observation of node X.*

Proof. Given is $X \in \mathbf{E}$.

(1) Suppose that $X \notin GivEvSens^{\mathbf{E}}(\mathbf{T})$. Then, by definition we have that

$$\langle \{X\} \,|\, \mathbf{E} \backslash \{X\} \,|\, \mathbf{T} \rangle^d_G$$

i.e. X and $\mathbf{T}$ are d-separated given $\mathbf{E}\backslash\{X\}$. This means that there does not exist an active chain between X and any node in $\mathbf{T}$ given $\mathbf{E}\backslash\{X\}$, or equivalently, X and $\mathbf{T}$ are independent given $\mathbf{E}\backslash\{X\}$. Then,

$$\Pr(\mathbf{T} \mid \mathbf{E}\backslash\{X\}) = \Pr(\mathbf{T} \mid \mathbf{E}).$$

Because of this equality, for every $x_1, x_2 \in X$ it holds that $\Pr(\mathbf{T} \mid \mathbf{E}\backslash\{X\}, x_1) = \Pr(\mathbf{T} \mid \mathbf{E}\backslash\{X\}, x_2)$, which proves the insensitivity of $\Pr(\mathbf{T} \mid \mathbf{E})$ to changes in the observed value of a node $X \notin GivEvSens^{\mathbf{E}}(\mathbf{T})$.

(2) Now, suppose that $X \in GivEvSens^{\mathbf{E}}(\mathbf{T})$, that is, by definition we have that

$$\neg \langle \{X\} \,|\, \mathbf{E} \backslash \{X\} \,|\, \mathbf{T} \rangle^d_G$$

i.e. X and $\mathbf{T}$ are d-connected given $\mathbf{E}\backslash\{X\}$. This means that there exists an active chain between X and a node in $\mathbf{T}$ given $\mathbf{E}\backslash\{X\}$, or equivalently, X and $\mathbf{T}$ *may* be dependent given $\mathbf{E}\backslash\{X\}$. It may then hold that

$$\Pr(\mathbf{T} \mid \mathbf{E}\backslash\{X\}) \neq \Pr(\mathbf{T} \mid \mathbf{E})$$

from which we have that $x_1, x_2 \in X$ could exist, such that $\Pr(\mathbf{T} \mid \mathbf{E}\backslash\{X\}) \neq \Pr(\mathbf{T} \mid \mathbf{E}\backslash\{X\}, x_1) \neq \Pr(\mathbf{T} \mid \mathbf{E}\backslash\{X\}, x_2)$, which proves the possible sensitivity of $\Pr(\mathbf{T} \mid \mathbf{E})$ to changes in, or removal of the observed value of a node $X \in GivEvSens^{\mathbf{E}}(\mathbf{T})$. □

Note that the given evidence sensitivity set contains only nodes from $\mathbf{E}$. Now, we will introduce the *potential evidence sensitivity set*. This set consists of all nodes $X \in V(G), X \notin \mathbf{E}$, for which adding an observation may influence the posterior probability of interest.

Definition 3.2.2 (Potential evidence sensitivity set)**.** *Let $G = (V(G), A(G))$, $\mathbf{T}$ and $\mathbf{E}$ be as defined before. The potential evidence sensitivity set $PotEvSens^{\mathbf{E}}(\mathbf{T})$ for $\mathbf{T}$ given $\mathbf{E}$ contains $\mathbf{T}$ and all variables $X \notin \mathbf{E}$ for which we have that*

$$\neg \langle \{X\} \mid \mathbf{E} \mid \mathbf{T} \rangle_G^d .$$

To show that observing the value of a specific variable X from the potential evidence sensitivity set may indeed influence the probability of interest, we state the following proposition.

Proposition 3.2.2. *Let $G = (V(G), A(G))$, $\mathbf{T}$ and $\mathbf{E}$ be as defined before. Let $PotEvSens^{\mathbf{E}}(\mathbf{T})$ be the potential evidence sensitivity set for $\mathbf{T}$ given $\mathbf{E}$ as defined in Definition 3.2.2. Then,*

1. *for every node $X \notin PotEvSens^{\mathbf{E}}(\mathbf{T})$, we have that $\Pr(\mathbf{T} \mid \mathbf{E})$ is insensitive to observing the value of node X.*
2. *for every node $X \in PotEvSens^{\mathbf{E}}(\mathbf{T})$, we have that $\Pr(\mathbf{T} \mid \mathbf{E})$ may be sensitive to observing the value of node X.*

Proof. Given is $X \notin \mathbf{E}$.

(1) Suppose that $X \notin PotEvSens^{\mathbf{E}}(\mathbf{T})$. Then, by definition we have that

$$\langle \{X\} \mid \mathbf{E} \mid \mathbf{T} \rangle_G^d$$

i.e. X and $\mathbf{T}$ are d-separated given $\mathbf{E}$. This means that there does not exist an active chain between X and any node in $\mathbf{T}$ given $\mathbf{E}$, or equivalently, X and $\mathbf{T}$ are independent given $\mathbf{E}$. Then,

$$\Pr(\mathbf{T} \mid \mathbf{E}) = \Pr(\mathbf{T} \mid \mathbf{E} \cup \{X\}).$$

Because of this equality, for every $x \in X$ it holds that $\Pr(\mathbf{T} \mid \mathbf{E}) = \Pr(\mathbf{T} \mid \mathbf{E}, x)$, which proves the insensitivity of $\Pr(\mathbf{T} \mid \mathbf{E})$ to observing the value of a node $X \notin PotEvSens^{\mathbf{E}}(\mathbf{T})$.

(2) First, suppose that $X \in PotEvSens^{\mathbf{E}}(\mathbf{T})$ and $X \notin \mathbf{T}$, that is, by definition we have that

$$\neg \langle \{X\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle^d_G$$

i.e. X and $\mathbf{T}$ are d-connected given $\mathbf{E}$. This means that there exists an active chain between X and a node in $\mathbf{T}$ given $\mathbf{E}$, or equivalently, X and $\mathbf{T}$ *may* be dependent given $\mathbf{E}$. It may then hold that

$$\Pr(\mathbf{T} \mid \mathbf{E}) \neq \Pr(\mathbf{T} \mid \mathbf{E} \cup \{X\})$$

from which we have that a value $x \in X$ could exist, such that $\Pr(\mathbf{T} \mid \mathbf{E}) \neq \Pr(\mathbf{T} \mid \mathbf{E}, x)$.

Second, suppose that $X \in \mathbf{T}$. Then, since it always holds that $\mathbf{T} \cap \mathbf{E} = \emptyset$, X has to be removed from $\mathbf{T}$ after observing. It may then hold that

$$\Pr(\mathbf{T} \mid \mathbf{E}) \neq \Pr(\mathbf{T} \backslash \{X\} \mid \mathbf{E} \cup \{X\})$$

from which we have that a value $x \in X$ could exist, such that $\Pr(\mathbf{T} \mid \mathbf{E}) \neq \Pr(\mathbf{T} \backslash \{X\} \mid \mathbf{E}, x)$. Both cases together prove the possible sensitivity of $\Pr(\mathbf{T} \mid \mathbf{E})$ to observing the value of a node $X \in PotEvSens^{\mathbf{E}}(\mathbf{T})$. □

The two previously defined sets together can be seen as containing all variables for which any kind of change in the observation could influence the posterior probability distribution $\Pr(\mathbf{T} \mid \mathbf{E})$ of interest. The network may thus be sensitive to changes in the observations of these nodes. We will define the union of the two previously defined sets as follows.

Definition 3.2.3 (Evidence sensitivity set)**.** *Let $G = (V(G), A(G))$, $\mathbf{T}$ and $\mathbf{E}$ be as defined before. The set*

$$EvSens^{\mathbf{E}}(\mathbf{T}) = GivEvSens^{\mathbf{E}}(\mathbf{T}) \cup PotEvSens^{\mathbf{E}}(\mathbf{T})$$

is called the evidence sensitivity set *for $\mathbf{T}$ given $\mathbf{E}$.*

Note that the given evidence sensitivity set and the potential evidence sensitivity set are *disjoint*, and that the evidence sensitivity set is just the union of those two sets. This way of defining the sensitivity sets relates them to each other, as demonstrated in the following proposition. For a schematic summary of all statements of the proposition by means of an Euler diagram, see Figure 3.1. The evidence sensitivity set consists of a given (dark grey) and a potential (light grey) part, as defined before.

Proposition 3.2.3. *Let $G = (V(G), A(G))$, $\mathbf{T}$ and $\mathbf{E}$ be as defined before. Then,*

1. $ParSens^{\mathbf{E}}(\mathbf{T}) \backslash \mathbf{E} \subseteq PotEvSens^{\mathbf{E}}(\mathbf{T})$
2. $ParSens^{\mathbf{E}}(\mathbf{T}) \cap \mathbf{E} \subseteq GivEvSens^{\mathbf{E}}(\mathbf{T})$

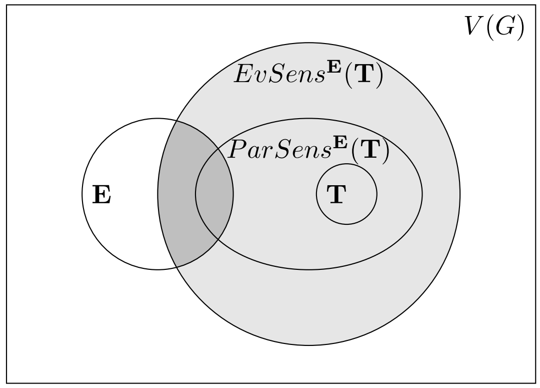


Figure 3.1: All statements of Proposition 3.2.3 summarized in an Euler diagram.

Proof. We recall that $\mathbf{T} \cap \mathbf{E} = \emptyset$ and that we always have that $\mathbf{T} \subseteq ParSens^{\mathbf{E}}(\mathbf{T})$ and, by definition, $\mathbf{T} \subseteq PotEvSens^{\mathbf{E}}(\mathbf{T})$.

Suppose that $X \in ParSens^{\mathbf{E}}(\mathbf{T})$, that is, by definition we have that

$$\neg \langle \{P_X\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle^d_{G^*}$$

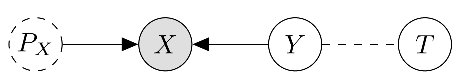
i.e. P_X and $\mathbf{T}$ are d-connected in the parented graph G^*. We want to prove that X is an element of the evidence sensitivity set. We will distinguish between two cases, where X is or is not in $\mathbf{E}$. Recall that the given evidence sensitivity set and the potential evidence sensitivity set together define the evidence sensitivity set, and that they are disjoint because of the way they are defined.

(1) For the first case, we assume that $X \notin \mathbf{E}$. We want to show that X is in the potential evidence sensitivity set. That is, we need to prove that $\neg \langle \{X\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle^d_G$ holds:

$$\neg \langle \{P_X\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle^d_{G^*} \implies \neg \langle \{X\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle^d_{G^*} \implies \neg \langle \{X\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle^d_G$$

Because P_X has no other arcs than the one to X, the d-connection between P_X and $\mathbf{T}$ has to be a chain including X, which implies that $\neg \langle \{X\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle^d_{G^*}$, and therefore also $\neg \langle \{X\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle^d_G$, holds. We conclude that $X \in PotEvSens^{\mathbf{E}}(\mathbf{T})$.

(2) Now we assume that $X \in \mathbf{E}$. We want to show that X is in the given evidence sensitivity set, i.e. $\neg \langle \{X\} \,|\, \mathbf{E} \backslash \{X\} \,|\, \mathbf{T} \rangle^d_G$ holds. Because P_X has no other arcs than the one to X, and X is in the evidence set, P_X and $\mathbf{T}$ can only be d-connected if X has a head-to-head connection on an active chain with some $T \in \mathbf{T}$ as an endpoint. This situation is illustrated in the following graph:



On this chain, parent Y has to be d-connected to T. By excluding X from the evidence set $\mathbf{E}$, X and T are d-connected, and therefore $\neg \langle\{X\} \mid \mathbf{E}\backslash\{X\} \mid \mathbf{T}\rangle_G^d$ holds. □

The above proposition gives us insights about nodes whose parameter values may influence the posterior probability of interest. If such a node is in $\mathbf{E}$, its observed value may always influence the posterior probability of interest, because $ParSens^{\mathbf{E}}(\mathbf{T}) \cap \mathbf{E} \subseteq GivEvSens^{\mathbf{E}}(\mathbf{T})$. In the case that the node is not in $\mathbf{E}$, it will always in the potential evidence sensitivity set, because $ParSens^{E}(T)\backslash\mathbf{E} \subseteq PotEvSens^{\mathbf{E}}(\mathbf{T})$.

Note that deleting an observation of a node from the given evidence sensitivity set or adding an observation for a node from the potential evidence sensitivity set may change the dependences and therefore the sensitivity sets of the network.

3.3 Identifying the sensitivity sets

One of the benefits of probabilistic networks, is that much knowledge is captured in the graphical structure. Many properties can be recognized just by looking at the topology of the underlying graph. Statements of conditional independence for example, can be verified in time linear to the size of the graph. An algorithm to this end is the *Bayes-Ball Algorithm* [13], which determines irrelevant sets of nodes and requisite information in time linear to the size of the graph. Bayes-Ball determines all d-connected nodes, all nodes with relevant parameters and all relevant observations.

Some of the sensitivity sets as defined in the previous section, are easily recognizable in the definitions of Shachter in [13]. In fact, all sensitivity sets can be easily determined using a simplified version of the Bayes-Ball Algorithm, given by Algorithm 1. The algorithm is simplified in the way that the original algorithm contains also a distinction for deterministic nodes, something that is unnecessary to include for our purpose.

Informally, the algorithm sends balls from all target nodes on chains through the network. The way a ball behaves corresponds to the different situations on a chain, as illustrated in Table 3.1. Each situation results in certain marks on the node, which can be translated to sensitivity sets in the end. The algorithm works as follows.

- Initially, all nodes are unvisited and unmarked, and all target nodes are scheduled to be visited from a child (line 1-5).
- While there is still a node in the schedule, this node will be removed and actually becomes visited by a bouncing ball (line 6-8).

Algorithm 1 Simplified Bayes-Ball Algorithm

Input: Graph $G = (V(G), A(G))$, evidence nodes $\mathbf{E}$ and target nodes $\mathbf{T}$.
Output: The sets $Visited, Top, Bottom \subseteq V(G)$.

```
for each node X ∈ V(G) do
    visited(X) = FALSE
    top(X) = UNMARKED
    bottom(X) = UNMARKED
S = {(N, CHILD) | N ∈ T}
while there exists an (X, from) ∈ S do
    S = S\{(X, from)}
    visited(X) = TRUE
    if X ∉ E and from == CHILD then
        if top(X) == UNMARKED then
            top(X) = MARKED
            S = S ∪ {(N, CHILD) | N ∈ ρ(X)}
        if bottom(X) == UNMARKED then
            bottom(X) = MARKED
            S = S ∪ {(N, PARENT) | N ∈ δ(X)}
    if from == PARENT then
        if X ∈ E and top(X) == UNMARKED then
            top(X) = MARKED
            S = S ∪ {(N, CHILD) | N ∈ ρ(X)}
        if X ∉ E and bottom(X) == UNMARKED then
            bottom(X) = MARKED
            S = S ∪ {(N, PARENT) | N ∈ δ(X)}
Visited = {X ∈ V(G) | visited(X) == TRUE}
    Top = {X ∈ V(G) | top(X) == MARKED}
Bottom = {X ∈ V(G) | bottom(X) == MARKED}
return Visited, Top, Bottom
```

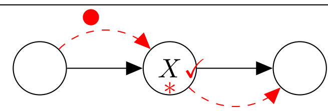
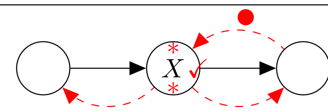
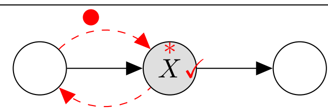
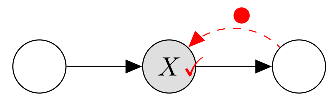
 Unobserved node X gets a ball from a parent and bounces balls to all children.	 Unobserved node X gets a ball from a child and bounces balls to all parents and children.
 Observed node X gets a ball from a parent and bounces balls to all parents.	 Observed node X gets a ball from a child and blocks bouncing.

Table 3.1: Behavior of the balls in the Bayes-Ball Algorithm. The mark $*$ at node X stands marked on top and/or on bottom by the algorithm, respectively, for X bouncing balls to its parents and/or children. The mark ✓ at node X stands for visited by the algorithm.

- The if-statements in the algorithm handle the different cases of how the balls will bounce through the network, with respect to the d-separation criterion. (line 9-22).
  - If a node bounces balls to its children, its bottom will become marked (line 14,21).
  - If a node bounces balls to its parents, its top will become marked (line 11,18).
- The algorithm terminates when the schedule is empty.

Note that a node becomes visited if there is at least one incoming ball. After running Algorithm 1, the sensitivity sets are easily recognizable by looking at the marked and visited nodes. This is summarized in the following theorem:

Theorem 3.3.1. *Let $G = (V(G), A(G))$, $\mathbf{T}$ and $\mathbf{E}$ be as defined before. Run the Bayes-Ball Algorithm on G, $\mathbf{T}$ and $\mathbf{E}$. Let $Visited, Top, Bottom \subseteq V(G)$ be as determined by the Bayes-Ball Algorithm. Then,*

1. *$PotEvSens^{\mathbf{E}}(\mathbf{T}) = Bottom$.*
2. *$GivEvSens^{\mathbf{E}}(\mathbf{T}) = Visited \cap \mathbf{E}$.*
3. *$ParSens^{\mathbf{E}}(\mathbf{T}) = Top$.*

To prove Theorem 3.3.1, we prove the following lemma first, to make the connection between the bouncing of the balls and the d-separation criterion. [1]

[1] In [13], Shachter mentions in his complexity proof that the statement in Lemma 3.3.1 is true, but the proof is absent.

Lemma 3.3.1. *Let $G = (V(G), A(G))$, $\mathbf{T}$ and $\mathbf{E}$ be as defined before. Then, for every $\mathbf{X} \in V(G)$,*

$$\neg \langle \{X\} \mid \mathbf{E} \backslash \{X\} \mid \mathbf{T} \rangle_G^d \iff X \in Visited.$$

Proof of Lemma 3.3.1. We will prove the lemma in two steps:

1. $\neg \langle \{X\} \mid \mathbf{E} \backslash \{X\} \mid \mathbf{T} \rangle_G^d \implies X \in Visited.$
2. $\langle \{X\} \mid \mathbf{E} \backslash \{X\} \mid \mathbf{T} \rangle_G^d \implies X \notin Visited.$

(1) Suppose that $\neg \langle \{X\} \mid \mathbf{E} \backslash \{X\} \mid \mathbf{T} \rangle_G^d$, that is, by the definition we have that there exists a $T \in \mathbf{T}$ such that $\neg \langle \{X\} \mid \mathbf{E} \backslash \{X\} \mid \{T\} \rangle_G^d$. In other words, there exists an active chain s between X and T given $\mathbf{E} \backslash \{X\}$. This means, by the definition of an active chain, that

- any node N on s with two incoming arcs is in or has a descendant in $\mathbf{E} \backslash \{X\}$, and
- any node N on s *not* having two incoming arcs is not in $\mathbf{E} \backslash \{X\}$.

Now, we will show that in both situations it holds that if a ball visits node N, then a ball will passed further along the chain:

- If a node N with two incoming arcs on s is or has a descendant in $\mathbf{E}$, and N receives a ball from a parent, it will always bounce the ball along the chain to the other parent. That is because the ball will "bump" on the evidence node, which either is a descendant of N or is N itself. This makes the ball bounce back to all its parents, and the ball will continue on s. This corresponds with lines 17-19 of Algorithm 1.
- If a node on s without being in $\mathbf{E}$ receives a ball, it will pass the ball along s, if N has no two incoming arcs. This corresponds with lines 9-15 (the "from a child"-case) and lines 20-22 (the "from a parent"-case without an evidence node) of Algorithm 1.

Therefore, $\neg \langle \{X\} \mid \mathbf{E} \backslash \{X\} \mid \mathbf{T} \rangle_G^d \implies X \in Visited.$

(2) Now, suppose that $\langle \{X\} \mid \mathbf{E} \backslash \{X\} \mid \mathbf{T} \rangle_G^d$, that is, by the definition we have that for all $T \in \mathbf{T}$ it holds that $\langle \{X\} \mid \mathbf{E} \backslash \{X\} \mid \{T\} \rangle_G^d$. In other words, any chain s between X and T given $\mathbf{E} \backslash \{X\}$ is blocked. This means, by the definition of a blocked chain, that

- there exists a node N on s with two incoming arcs which is not in and does not have a descendant in $\mathbf{E} \backslash \{X\}$, or
- there exists a node N on s *not* having two incoming arcs which is in $\mathbf{E} \backslash \{X\}$.

Now, we will show that in both situations it holds that if a ball visits node N, then no ball will passed further along the chain:

- If a node N with two incoming arcs on s is or has not a descendant in $\mathbf{E}$, and N receives a ball from a parent, it will never bounce a ball along the chain to the other parent. That is because the ball will only be passed to the children of N. This corresponds with a block on this chain.
- A node not having two incoming arcs which is in $\mathbf{E}$, corresponds too with a block on the chain.

Therefore, $\langle \{X\} \mid \mathbf{E} \backslash \{X\} \mid \mathbf{T} \rangle_G^d \implies X \notin Visited$, and together with the result from (1), it proves the lemma. □

By using the previous lemma, it is quite straightforward to prove the earlier stated theorem:

Proof of Theorem 3.3.1. We will distinguish between the three cases stated in the theorem:

(1) We want to prove that $X \in Bottom$ if and only if $X \in PotEvSens^{\mathbf{E}}(\mathbf{T})$, or equivalently,

$$\neg \langle \{X\} \mid \mathbf{E} \mid \mathbf{T} \rangle_G^d .$$

Note that $X \notin \mathbf{E}$. Then, by applying Lemma 3.3.1,

$$\neg \langle \{X\} \mid \mathbf{E} \mid \mathbf{T} \rangle_G^d \iff X \in Visited.$$

By the algorithm, $X \notin \mathbf{E}$ will become marked on bottom if and only if X will be visited, if and only if $X \in PotEvSens^{\mathbf{E}}(\mathbf{T})$.

(2) Now, we want to prove that $X \in \mathbf{E} \cap Visited$ if and only if $X \in GivEvSens^{\mathbf{E}}(\mathbf{T})$, or equivalently,

$$\neg \langle \{X\} \mid \mathbf{E} \backslash \{X\} \mid \mathbf{T} \rangle_G^d .$$

Then, by applying Lemma 3.3.1,

$$\neg \langle \{X\} \mid \mathbf{E} \backslash \{X\} \mid \mathbf{T} \rangle_G^d \iff X \in Visited.$$

By the algorithm, $X \in \mathbf{E}$ will be visited if and only if $X \in GivEvSens^{\mathbf{E}}(\mathbf{T})$.

(3) Now, we want to prove that $X \in Top$ if and only if $X \in ParSens^{\mathbf{E}}(\mathbf{T})$, or equivalently,

$$\neg \langle \{P_X\} \mid \mathbf{E} \mid \mathbf{T} \rangle_{G^*}^d .$$

Note that always holds that $P_X \notin \mathbf{E}$. Suppose we run the algorithm on the parented graph G^*. Then, by Lemma 3.3.1,

$$\neg \langle \{P_X\} \mid \mathbf{E} \mid \mathbf{T} \rangle_G^d \iff P_X \in Visited.$$

Thus, during the algorithm, $X \in ParSens^{\mathbf{E}}(\mathbf{T})$ if and only if P_X will be visited, if and only if X sends the ball to its auxiliary parent P_X, if and only if the top of node X will become marked. □

This gives us an algorithm to identify exactly the sensitivity sets of a network in an efficient way. Shachter used the terms relevant and requisite sets of nodes, but (as far as we know), the direct link to the parameter sensitivity set was not made before. Moreover, the evidence sensitivity set introduced in this thesis can also be identified by the Bayes-Ball algorithm.

3.4 Example

We will end this chapter with an example that demonstrates all concepts from this chapter. The graph of the probabilistic network that we consider is depicted in Figure 3.2. Here, $\mathbf{T} = \{T_1, T_2\}$ and $\mathbf{E} = \{E_1, E_2, E_3\}$. We will determine all sensitivity sets for the probabilistic network by the Bayes-Ball Algorithm, Algorithm 1, and we will explain the implications for a sensitivity analysis being performed.

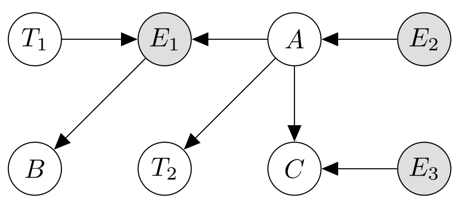


Figure 3.2: Graph of the example probabilistic network.

After the termination of Algorithm 1 on the example graph of Figure 3.2, the nodes are marked as in Figure 3.3, that is,

$$\begin{aligned} Visited &= \{A, C, E_1, E_2, T_1, T_2\}, \\ Top &= \{A, E_1, T_1, T_2\}, \\ Bottom &= \{A, C, T_1, T_2\}. \end{aligned}$$

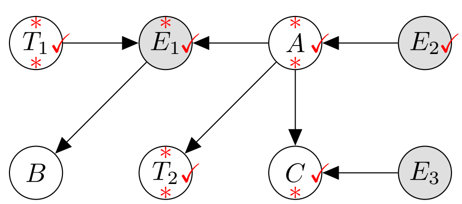


Figure 3.3: The example graph with all marks at the top and bottom and visited (✓) nodes, after running the Bayes-Ball Algorithm.

Now, we will use Theorem 3.3.1 to determine the sensitivity sets, which results in:

$$PotEvSens^{\mathbf{E}}(\mathbf{T}) = Bottom = \{A, C, T_1, T_2\},$$
$$GivEvSens^{\mathbf{E}}(\mathbf{T}) = Visited \cap \mathbf{E} = \{E_1, E_2\},$$
$$ParSens^{\mathbf{E}}(\mathbf{T}) = Top = \{A, E_1, T_1, T_2\}.$$

In Figure 3.4, the sensitivity sets are depicted in the style of Figure 3.1, where all implications by Proposition 3.2.3 are summarized in an Euler diagram.

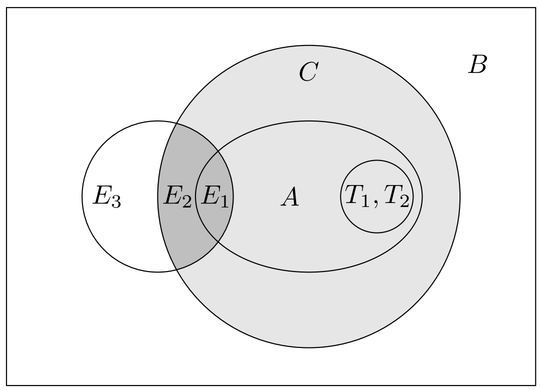


Figure 3.4: Euler diagram corresponding to the example graph of Figure 3.2.

The parameter sensitivity set can be used as a basis for doing a full parameter sensitivity analysis. For the example probabilistic network, the following can be concluded. The nodes A, E_1, T_1 and T_2 are contained in the parameter sensitivity set for $\mathbf{T}$ given $\mathbf{E}$. Therefore, changing the parameter values of these nodes may influence the outcome of the probability of interest $\Pr(\mathbf{t} \mid \mathbf{e})$. On the other hand, nodes B, C, E_2 and E_3 are not contained in the parameter sensitivity set for $\mathbf{T}$ given $\mathbf{E}$. These four nodes can therefore be excluded from a parameter sensitivity analysis, since by Proposition 3.1.1 we have that the probability of interest is insensitive to changes in the parameter values of these nodes.

Note that by Proposition 3.2.3, we know that at least all nodes in the parameter sensitivity set for $\mathbf{T}$ given $\mathbf{E}$ have to be included in an evidence sensitivity analysis. But by looking at the given evidence sensitivity set and the potential evidence sensitivity set, we can determine *all* nodes that have to be included.

First, evidence nodes E_1 and E_2 are contained in the given evidence sensitivity set for $\mathbf{T}$ given $\mathbf{E}$. Therefore, changing or removing the observations of these nodes may influence the outcome of the probability of interest $\Pr(\mathbf{t} \mid \mathbf{e})$. On the other hand, evidence node E_3 is not contained in the given evidence sensitivity set for $\mathbf{T}$ given $\mathbf{E}$. The nodes for which their observations have to be varied in order to investigate their influence on the probability of interest are therefore E_1 and E_2.

Second, nodes A, C, T_1 and T_2 are contained in the potential evidence sensitivity set for $\mathbf{T}$ given $\mathbf{E}$. Therefore, adding observations for these variables may influence the outcome of the probability of interest $\Pr(\mathbf{t} \mid \mathbf{e})$. On the other hand, Nodes B and E_3 are not in the potential evidence sensitivity set for $\mathbf{T}$ given $\mathbf{E}$. The nodes for which observations can be added and varied in order to investigate their influence on the probability of interest are therefore A, C, T_1 and T_2. However, it is possible that nodes are not observable at all. Note that the sensitivity sets may change after adding an observation for nodes in the potential evidence sensitivity set for $\mathbf{T}$ given $\mathbf{E}$.

3.5 Summary

In this chapter, we introduced the evidence sensitivity set for $\mathbf{T}$ given $\mathbf{E}$. This set captures all nodes for which a change in observed value, or change in status of being observed or not, affects the probability of interest. We proved that a node is in the evidence sensitivity set if and only if it might affect the probability of interest upon undergoing abovementioned changes. Subsequently, we formalized the relation between the parameter sensitivity set for $\mathbf{T}$ given $\mathbf{E}$ and the evidence sensitivity set for $\mathbf{T}$ given $\mathbf{E}$. We proved that the parameter sensitivity set is always a subset of the evidence sensitivity set.

Finally, we reintroduced the already existing Bayes-Ball Algorithm [13] in order to identify all sensitivity sets of a probabilistic network. We analyzed how to interpret the output of the algorithm in order to identify the sensitivity sets, and we proved the correctness of this interpretation.

The evidence sensitivity set can be used as a basis for doing a full evidence sensitivity analysis. The evidence nodes contained in the given evidence sensitivity set for $\mathbf{T}$ given $\mathbf{E}$ have to be varied in their observations in order to investigate their influence on the probability of interest.

In addition, the nodes contained in the potential evidence sensitivity set for $\mathbf{T}$ given $\mathbf{E}$ can be added to $\mathbf{E}$ and varied in their observations in order to investigate their influence on the probability of interest. Note that the sensitivity sets may change after adding an observation for nodes in the potential evidence sensitivity set for $\mathbf{T}$ given $\mathbf{E}$.

Chapter 4

The effects of pruning on the probability calculation

In this chapter, we will show how *probabilistic network pruning* can be translated to simplifications in the calculation of the probability of interest. The most naive approach to computing $\Pr(\mathbf{T} \mid \mathbf{E}) = \Pr(\mathbf{T} \wedge \mathbf{E}) / \Pr(\mathbf{E})$ is by summing over all variables $V(G) \backslash (\mathbf{T} \cup \mathbf{E})$ and $V(G) \backslash \mathbf{E}$ in the numerator and denominator, respectively:

$$\Pr(\mathbf{T} \wedge \mathbf{E}) = \sum_{V(G) \backslash (\mathbf{T} \cup \mathbf{E})} \prod_{X \in V(G)} \Pr(X \mid \rho(X)) \tag{4.1}$$

$$\Pr(\mathbf{E}) = \sum_{V(G) \backslash \mathbf{E}} \prod_{X \in V(G)} \Pr(X \mid \rho(X)) \tag{4.2}$$

However, not all variables are relevant for the calculation of the probability of interest at hand. These variables can be *pruned* from the network, as demonstrated in, for example, [1] and [9]. In this chapter we will show *why* certain nodes can be pruned, from the perspective of the probability calculation. Our approach differs from [1], where Pearl's equations regarding message-passing [11] are used in explaining why certain nodes can be pruned safely from the network.

Our approach will provide us with new insights in the structure of the probability calculation. In addition, an additional subset of nodes will be identified, that can safely be pruned from the network. We will apply these insights in the next chapter to evidence sensitivity analysis.

4.1 Probability calculation simplifications

Nodes can be pruned from the network if their parameters as well as their children's parameters are not needed in the calculation for a probability of interest. To demonstrate that these parameters are not needed, we will present a basic toolkit of rewriting rules given by Lemmas 4.1.1- 4.1.3.

In calculating a specific marginal probability in a given probabilistic network, it is often possible to simplify this calculation by looking closely at the dependences in the network, and at specific properties of probabilities. Recall from Equation 2.2, that for any subset $\mathbf{Y} \subseteq V(G)$, the marginal probability $\Pr(\mathbf{Y})$ is defined by

$$\Pr(\mathbf{Y}) = \sum_{V(G)\backslash\mathbf{Y}} \prod_{X \in V(G)} \Pr(X \mid \rho(X)) \tag{4.3}$$

The following lemma shows that summing over *all* variables contained in the conditional probability distributions for a certain subset $\mathbf{X} \subseteq V(G)$, results in a probability of 1.

Lemma 4.1.1. *Let $G = (V(G), A(G))$ and $\Gamma = \{\Pr(X \mid \rho(X)) \mid X \in V(G)\}$ be as defined before. Let $\mathbf{X} \subseteq V(G)$. Then,*

$$\sum_{\mathbf{X}} \prod_{X \in \mathbf{X}} \Pr(X \mid \rho(X)) = 1.$$

Proof. It always holds that $\prod_{X \in \mathbf{X}} \Pr(X \mid \rho(X)) = \Pr(\mathbf{X} \mid \rho(\mathbf{X})\backslash\mathbf{X})$, since G is a probabilistic network, where the chain rule can be applied. Then,

$$\sum_{\mathbf{X}} \prod_{X \in \mathbf{X}} \Pr(X \mid \rho(X)) = \sum_{\mathbf{X}} \Pr(\mathbf{X} \mid \rho(\mathbf{X})\backslash\mathbf{X}) = 1$$

always holds. □

A probability calculation computationally can often be made more efficient, by rearranging terms. Consider the following expression,

$$\sum_{\mathbf{Q} \cup \mathbf{R}} \prod_{X \in \mathbf{X} \cup \mathbf{Y}} \Pr(X \mid \rho(X))$$

where $\mathbf{Q}, \mathbf{R}, \mathbf{X}, \mathbf{Y} \subseteq V(G)$, such that $\mathbf{Q} \cap \mathbf{R} = \emptyset$ and $\mathbf{X} \cap \mathbf{Y} = \emptyset$. This expression can be rephrased as

$$\sum_{\mathbf{Q}} \sum_{\mathbf{R}} \prod_{X \in \mathbf{X}} \Pr(X \mid \rho(X)) \cdot \prod_{Y \in \mathbf{Y}} \Pr(Y \mid \rho(Y)).$$

Now, if subset $\mathbf{R}$ meets the conditions stated in the following lemma, then the product over $\mathbf{X}$ is constant with respect to all variables in $\mathbf{R}$. The above expression can then be rewritten in a way that is computationally more efficient.

Lemma 4.1.2. *Let $G = (V(G), A(G))$ and $\Gamma = \{\Pr(X \mid \rho(X)) \mid X \in V(G)\}$ be as defined before. Let $\mathbf{Q}, \mathbf{R}, \mathbf{X}, \mathbf{Y} \subseteq V(G)$. If $\mathbf{Q} \cap \mathbf{R} = \emptyset$, $\mathbf{X} \cap \mathbf{Y} = \emptyset$ and $(\mathbf{X} \cup \rho(\mathbf{X})) \cap \mathbf{R} = \emptyset$, then*

$$\sum_{\mathbf{Q}\cup\mathbf{R}} \prod_{X\in\mathbf{X}\cup\mathbf{Y}} \Pr(X \mid \rho(X)) = \left[\sum_{\mathbf{Q}} \prod_{X\in\mathbf{X}} \Pr(X \mid \rho(X)) \cdot \left[\sum_{\mathbf{R}} \prod_{Y\in\mathbf{Y}} \Pr(Y \mid \rho(Y))\right]\right]$$

Proof. If $\mathbf{X}$ and all parents of the nodes in $\mathbf{X}$ do not overlap with $\mathbf{R}$, then the product over the parameters of $\mathbf{X}$ is constant with respect to all variables in $\mathbf{R}$. The distributive law now tells us that the product over the parameters of $\mathbf{X}$ can be taken out of the summation over $\mathbf{R}$. □

In other words, the summation over all variables in $\mathbf{R}$ can be *pulled in* by the use of the distributive law, which is used by most of the inference algorithms. An even stronger form of this proposition can be formalized, where the distributive law is applied twice, if subsets $\mathbf{Q}$ and $\mathbf{R}$ meet the conditions as are stated in the following lemma.

Lemma 4.1.3. *Let $G = (V(G), A(G))$ and $\Gamma = \{\Pr(X \mid \rho(X)) \mid X \in V(G)\}$ be as defined before. Let $\mathbf{Q}, \mathbf{R}, \mathbf{X}, \mathbf{Y} \subseteq V(G)$. If $\mathbf{Q} \cap \mathbf{R} = \emptyset$, $\mathbf{X} \cap \mathbf{Y} = \emptyset$, $(\mathbf{X} \cup \rho(\mathbf{X})) \cap \mathbf{R} = \emptyset$ and $(\mathbf{Y} \cup \rho(\mathbf{Y})) \cap \mathbf{Q} = \emptyset$, then*

$$\sum_{\mathbf{Q}\cup\mathbf{R}} \prod_{X\in\mathbf{X}\cup\mathbf{Y}} \Pr(X \mid \rho(X)) = \left[\sum_{\mathbf{Q}} \prod_{X\in\mathbf{X}} \Pr(X \mid \rho(X))\right] \cdot \left[\sum_{\mathbf{R}} \prod_{Y\in\mathbf{Y}} \Pr(Y \mid \rho(Y))\right]$$

Proof. Following the proof of Lemma 4.1.2, we have that from the distributive law that both the product over the parameters of $\mathbf{X}$ can be taken out of the summation over $\mathbf{R}$ and the parameters of $\mathbf{Y}$ can be taken out of the summation over $\mathbf{Q}$. □

In other words, under the constraints of Lemma 4.1.3 the summations over the variables in $\mathbf{Q}$ and $\mathbf{R}$ can be completely *separated.* We will now use the tools provided by Lemmas 4.1.1-4.1.3 to explain network pruning by means of the probability calculation.

4.2 Probabilistic network pruning

For each probabilistic network, a *minimal computationally equivalent subgraph* can be constructed by pruning computationally irrelevant nodes, as described in [1]. The goal is to find the smallest possible subgraph of the original graph of the probabilistic network, such that the probability of interest can still be calculated in the right way. This means that each node is pruned for which its own parameters as well as its children's parameters are not needed in the calculation.

The subgraph resulting after pruning two sets of nodes as described in [1] is called the *minimal computationally equivalent subgraph.* The following theorem is from [1] and is proven by using Pearl's equations.

Theorem 4.2.1 (Minimal computationally equivalent subgraph)**.** *Let* $G = (V(G), A(G))$*,* $\mathbf{T}$ *and* $\mathbf{E}$ *be as defined before. A subgraph* G^M *is* minimal computationally equivalent *to* G *with respect to* $\mathbf{T}$ *and* $\mathbf{E}$*, if it is constructed by*

- *removing all nodes d-separated from* $\mathbf{T}$ *by* $\mathbf{E}$*, and*
- *removing all barren nodes with respect to* $\mathbf{T}$ *and* $\mathbf{E}$*.*

Of course, all arcs that are not incident on two nodes in $V(G^M)$ *will be removed.*

Minimal in this case means, that there are no nodes in $V(G^M)\backslash(\mathbf{T} \cup \mathbf{E})$ for which we need neither its own parameters, nor its children's parameters in the calculation. Note that this is by taking only the graphical structure of the probabilistic network into consideration.

In the next subsections, we will explain why all nodes d-separated from $\mathbf{T}$ given $\mathbf{E}$ and all barren nodes with respect to $\mathbf{T}$ and $\mathbf{E}$ can be pruned safely, by means of the probability calculation. In addition, we present another set of nodes which can be pruned.

4.2.1 Pruning d-separated nodes

The first set of nodes which can be pruned in order to make the graph smaller, but computationally equivalent, is the set of *d-separated* nodes from $\mathbf{T}$ given $\mathbf{E}$, as defined in Definition 2.1.4. Let $\mathbf{S} \subseteq V(G)$ be such that $\mathbf{S}$ consists of all nodes d-separated from $\mathbf{T}$ given $\mathbf{E}$, that is,

$$\mathbf{S} = \{X \mid X \in V(G)\colon \langle \{X\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle_G^d\}.$$

Note that $\mathbf{S} \subseteq V(G)\backslash(\mathbf{T} \cup \mathbf{E})$, since d-separation is only defined for non-overlapping sets. We will now prove, by means of probability calculations, that nodes d-separated from $\mathbf{T}$ given $\mathbf{E}$ can be safely pruned.

Proposition 4.2.1. *Let* $G = (V(G), A(G))$ *and* $\Gamma = \{\Pr(X \mid \rho(X)) \mid X \in V(G)\}$ *be as defined before. Let* $\mathbf{S}$ *be as defined above and let* $\mathcal{D}(\mathbf{S})$ *be the 'donna con bambini' of* $\mathbf{S}$*. Then,*

$$\Pr(\mathbf{T} \mid \mathbf{E}) = \frac{\displaystyle\sum_{\substack{V(G)\backslash \\ (\mathbf{T} \cup \mathbf{E} \cup \mathcal{D}(\mathbf{S}))}} \prod_{X \in V(G)\backslash \mathcal{D}(\mathbf{S})} \Pr(X \mid \rho(X))}{\displaystyle\sum_{\substack{V(G)\backslash \\ (\mathbf{E} \cup \mathcal{D}(\mathbf{S}))}} \prod_{X \in V(G)\backslash \mathcal{D}(\mathbf{S})} \Pr(X \mid \rho(X))}$$

Proof. Since all nodes in $\mathbf{S}$ are d-separated from $\mathbf{T}$ given $\mathbf{E}$, the following holds:

$$\Pr(\mathbf{T} \wedge \mathbf{S} \mid \mathbf{E}) = \Pr(\mathbf{T} \mid \mathbf{E}) \cdot \Pr(\mathbf{S} \mid \mathbf{E})$$

or, equivalently,

$$\Pr(\mathbf{T} \mid \mathbf{E}) = \frac{\Pr(\mathbf{T} \wedge \mathbf{S} \mid \mathbf{E})}{\Pr(\mathbf{S} \mid \mathbf{E})} = \frac{\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E})}{\Pr(\mathbf{S} \wedge \mathbf{E})} \tag{4.4}$$

$$= \frac{\displaystyle\sum_{V(G)\backslash(\mathbf{T}\cup\mathbf{E}\cup\mathbf{S})} \prod_{X\in V(G)} \Pr(X \mid \rho(X))}{\displaystyle\sum_{V(G)\backslash(\mathbf{E}\cup\mathbf{S})} \prod_{X\in V(G)} \Pr(X \mid \rho(X))}. \tag{4.5}$$

Now, we will prove for the numerator, $\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E})$, that the parameters of all nodes in $\mathcal{D}(\mathbf{S})$ are not needed in the calculation. Since

- $(V(G)\backslash\mathbf{S}) \cup \mathbf{S} = V(G)$, and
- $(V(G)\backslash\mathbf{S}) \cap \mathbf{S} = \emptyset$

hold, we can rephrase $\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E})$ as:

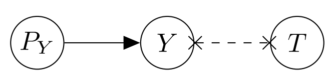
$$\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E}) = \sum_{\substack{V(G)\backslash \\ (\mathbf{T}\cup\mathbf{E}\cup\mathbf{S})}} \prod_{X\in V(G)\backslash\mathbf{S}} \Pr(X \mid \rho(X)) \cdot \prod_{Y\in\mathbf{S}} \Pr(Y \mid \rho(Y)).$$

We want to take the product over the parameters of the nodes in $\mathbf{S}$ out of the summation over $V(G)\backslash(\mathbf{T} \cup \mathbf{E} \cup \mathbf{S})$. In order to apply Lemma 4.1.2 to the calculation, we have to show that $(\mathbf{X} \cup \rho(\mathbf{X})) \cap \mathbf{R} = (\mathbf{S} \cup \rho(\mathbf{S})) \cap (V(G)\backslash(\mathbf{T} \cup \mathbf{E} \cup \mathbf{S})) = \emptyset$, i.e.

1. $\mathbf{S} \cap (V(G)\backslash(\mathbf{T} \cup \mathbf{E} \cup \mathbf{S})) = \emptyset$, and
2. for all $Y \in \mathbf{S}$, it holds that $\rho(Y) \cap (V(G)\backslash(\mathbf{T} \cup \mathbf{E} \cup \mathbf{S})) = \emptyset$.

Property 1 holds trivially.

Proof of property 2.
We assume that $Y \in \mathbf{S}$. We will now show by contradiction that $\rho(Y) \subseteq \mathbf{S} \cup \mathbf{E}$ always holds in order to prove the property. Suppose on the contrary, that $\rho(Y) \not\subseteq \mathbf{S} \cup \mathbf{E}$, in other words, there exists a $P_Y \in \rho(Y)$, such that $P_Y \notin \mathbf{S}$ and $P_Y \notin \mathbf{E}$. This situation is illustrated in the following graph:



Then, since $Y \in \mathbf{S}$, $P_Y \notin \mathbf{T}$ always holds. Since $P_Y \notin \mathbf{S}$, $P_Y \notin \mathbf{T}$ and $P_Y \notin \mathbf{E}$, $\neg \langle\{P_Y\} \mid \mathbf{E} \mid \mathbf{T}\rangle_G^d$ always holds. Then, also $\neg \langle\{Y\} \mid \mathbf{E} \mid \mathbf{T}\rangle_G^d$ always holds, which contradicts the supposition that $Y \in \mathbf{S}$. Hence, $\rho(Y) \subseteq \mathbf{S} \cup \mathbf{E}$ always holds.

Now, we can apply Lemma 4.1.2 to the calculation of $\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E})$:

$$\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E}) = \left[\sum_{\substack{V(G)\backslash \\ (\mathbf{T}\cup\mathbf{E}\cup\mathbf{S})}} \prod_{X \in V(G)\backslash \mathbf{S}} \Pr(X \mid \rho(X)) \right] \cdot \prod_{Y \in \mathbf{S}} \Pr(Y \mid \rho(Y)).$$

Since

- $(V(G)\backslash(\mathbf{T} \cup \mathbf{E} \cup \mathcal{D}(\mathbf{S}))) \cup (\delta(\mathbf{S})\backslash(\mathbf{E} \cup \mathbf{S})) = V(G)\backslash(\mathbf{T} \cup \mathbf{E} \cup \mathbf{S})$, and
- $(V(G)\backslash(\mathbf{T} \cup \mathbf{E} \cup \mathcal{D}(\mathbf{S}))) \cap (\delta(\mathbf{S})\backslash(\mathbf{E} \cup \mathbf{S})) = \emptyset$, and
- $(V(G)\backslash\mathcal{D}(\mathbf{S})) \cup (\delta(\mathbf{S})\backslash\mathbf{S}) = V(G)\backslash\mathbf{S}$, and
- $(V(G)\backslash\mathcal{D}(\mathbf{S})) \cap (\delta(\mathbf{S})\backslash\mathbf{S}) = \emptyset$

hold, we can subsequently rephrase $\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E})$ as:

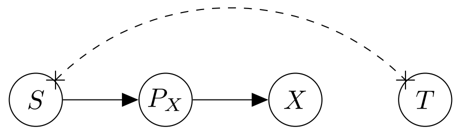
$$\begin{aligned} &\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E}) \\ &= \left[\sum_{\substack{V(G)\backslash \\ (\mathbf{T}\cup\mathbf{E}\cup\mathcal{D}(\mathbf{S}))}} \sum_{\substack{\delta(\mathbf{S})\backslash \\ (\mathbf{E}\cup\mathbf{S})}} \prod_{X \in V(G)\backslash\mathcal{D}(\mathbf{S})} \Pr(X \mid \rho(X)) \cdot \prod_{Y \in \delta(\mathbf{S})\backslash\mathbf{S}} \Pr(Y \mid \rho(Y)) \right] \\ &\quad \cdot \prod_{Z \in \mathbf{S}} \Pr(Z \mid \rho(Z)). \end{aligned}$$

Note that we now rephrased the product in such a way, that for all $X \in V(G)\backslash\mathcal{D}(\mathbf{S})$ we have that $\rho(X) \cap \mathbf{S} = \emptyset$. Our next step is to take the product over the parameters of nodes in $V(G)\backslash\mathcal{D}(\mathbf{S})$ out of the summation over $\delta(\mathbf{S})\backslash(\mathbf{E}\cup\mathbf{S})$. In order to apply Lemma 4.1.2 to the calculation of $\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E})$, we have to show that $(\mathbf{X}\cup\rho(\mathbf{X}))\cap\mathbf{R} = ((V(G)\backslash\mathcal{D}(\mathbf{S}))\cup\rho(V(G)\backslash\mathcal{D}(\mathbf{S})))\cap(\delta(\mathbf{S})\backslash(\mathbf{E}\cup\mathbf{S})) = \emptyset$, i.e.

3. $(V(G)\backslash\mathcal{D}(\mathbf{S})) \cap (\delta(\mathbf{S})\backslash(\mathbf{E} \cup \mathbf{S})) = \emptyset$, and
4. for all $X \in V(G)\backslash\mathcal{D}(\mathbf{S})$, it holds that $\rho(X) \cap (\delta(\mathbf{S})\backslash(\mathbf{E} \cup \mathbf{S})) = \emptyset$.

Property 3 holds trivially.

Proof of property 4.
We assume that $X \in V(G)\backslash\mathcal{D}(\mathbf{S})$, in other words, $X \notin \mathbf{S}$ and $X \notin \delta(\mathbf{S})$. We will now show by contradiction that $\rho(X) \cap (\delta(\mathbf{S})\backslash(\mathbf{E} \cup \mathbf{S})) = \emptyset$. Suppose on the contrary, that $\rho(X) \cap (\delta(\mathbf{S})\backslash(\mathbf{E} \cup \mathbf{S})) \neq \emptyset$, in other words, there exists a $P_X \in \rho(X)$, such that $P_X \in \delta(\mathbf{S})$, $P_X \notin \mathbf{S}$ and $P_X \notin \mathbf{E}$. This situation is illustrated in the following graph:



where $T \in \mathbf{T}$ and $S \in \mathbf{S}$, i.e. $\langle \{S\} \mid \mathbf{E} \mid \mathbf{T} \rangle_G^d$ by definition. Now, either $P_X \in \mathbf{T}$ or $\neg \langle \{P_X\} \mid \mathbf{E} \mid \mathbf{T} \rangle_G^d$ must hold. If $P_X \in \mathbf{T}$, then $\neg \langle \{S\} \mid \mathbf{E} \mid \mathbf{T} \rangle_G^d$, which contradicts that $S \in \mathbf{S}$. If $\neg \langle \{P_X\} \mid \mathbf{E} \mid \mathbf{T} \rangle_G^d$, then $\neg \langle \{S\} \mid \mathbf{E} \mid \mathbf{T} \rangle_G^d$, which contradicts that $S \in \mathbf{S}$. This contradicts the supposition that $\rho(X) \cap (\delta(\mathbf{S}) \backslash (\mathbf{E} \cup \mathbf{S})) \neq \emptyset$. Hence, $\rho(X) \cap (\delta(\mathbf{S}) \backslash (\mathbf{E} \cup \mathbf{S})) = \emptyset$ always holds.

Now, we can apply Lemma 4.1.2 to the calculation of $\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E})$:

$$\begin{aligned}
&\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E}) \\
&\quad = \left[\sum_{\substack{V(G)\backslash \\ (\mathbf{T} \cup \mathbf{E} \cup \mathcal{D}(\mathbf{S}))}} \prod_{X \in V(G) \backslash \mathcal{D}(\mathbf{S})} \Pr(X \mid \rho(X)) \cdot \left[\sum_{\substack{\delta(\mathbf{S})\backslash \\ (\mathbf{E} \cup \mathbf{S})}} \prod_{Y \in \delta(\mathbf{S}) \backslash \mathbf{S}} \Pr(Y \mid \rho(Y)) \right] \right] \\
&\qquad \cdot \prod_{Z \in \mathbf{S}} \Pr(Z \mid \rho(Z)).
\end{aligned}$$

Now, we will look closer at the innermost factor, where we want to split out the parameters of nodes in $\mathbf{E}$. Since it holds that

- $(\delta(\mathbf{S}) \backslash (\mathbf{S} \cup \mathbf{E})) \cup (\delta(\mathbf{S}) \cap \mathbf{E}) = \delta(\mathbf{S}) \backslash \mathbf{S}$, and
- $(\delta(\mathbf{S}) \backslash (\mathbf{S} \cup \mathbf{E})) \cap (\delta(\mathbf{S}) \cap \mathbf{E}) = \emptyset$,

we can rephrase the innermost factor as:

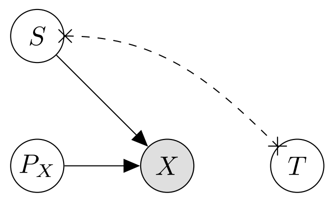
$$\begin{aligned}
&\sum_{\substack{\delta(\mathbf{S})\backslash \\ (\mathbf{E} \cup \mathbf{S})}} \prod_{Y \in \delta(\mathbf{S}) \backslash \mathbf{S}} \Pr(Y \mid \rho(Y)) \\
&\qquad = \sum_{\substack{\delta(\mathbf{S})\backslash \\ (\mathbf{E} \cup \mathbf{S})}} \prod_{X \in \delta(\mathbf{S}) \cap \mathbf{E}} \Pr(X \mid \rho(X)) \cdot \prod_{Y \in \delta(\mathbf{S}) \backslash (\mathbf{S} \cup \mathbf{E})} \Pr(Y \mid \rho(Y))
\end{aligned}$$

We want to apply Lemma 4.1.2, to take the product over the parameters of the nodes in $\delta(\mathbf{S}) \cap \mathbf{E}$ out of *both* of the summations, i.e. the summations over $(V(G) \backslash (\mathbf{T} \cup \mathbf{E} \cup \mathcal{D}(\mathbf{S}))) \cup (\delta(\mathbf{S}) \backslash (\mathbf{E} \cup \mathbf{S})) = V(G) \backslash (\mathbf{T} \cup \mathbf{E} \cup \mathbf{S})$. Therefore, we have to show that $(X \cup \rho(X))) \cap \mathbf{R} = ((\delta(\mathbf{S}) \cap \mathbf{E}) \cup \rho(\delta(\mathbf{S}) \cap \mathbf{E})) \cap (V(G) \backslash (\mathbf{T} \cup \mathbf{E} \cup \mathbf{S})) = \emptyset$, i.e.

5. $(\delta(\mathbf{S}) \cap \mathbf{E}) \cap (V(G) \backslash (\mathbf{T} \cup \mathbf{E} \cup \mathbf{S})) = \emptyset$, and
6. for all $X \in (\delta(\mathbf{S}) \cap \mathbf{E})$, it holds that $\rho(X) \cap (V(G) \backslash (\mathbf{T} \cup \mathbf{E} \cup \mathbf{S})) = \emptyset$.

Property 5 holds trivially.

Proof of property 6.
We assume that $X \in \delta(\mathbf{S}) \cap \mathbf{E}$. We will now show by contradiction that $\rho(X) \cap (V(G) \backslash (\mathbf{T} \cup \mathbf{E} \cup \mathbf{S})) = \emptyset$. Suppose on the contrary, that $\rho(X) \cap (V(G) \backslash (\mathbf{T} \cup \mathbf{E} \cup \mathbf{S})) \neq \emptyset$, in other words, there exists a $P_X \in \rho(X)$, such that $P_X \notin \mathbf{T}$, $P_X \notin \mathbf{E}$ and $P_X \notin \mathbf{S}$. This situation is illustrated in the following graph:



where $T \in \mathbf{T}$ and $S \in \mathbf{S}$, i.e. $\langle \{S\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle_G^d$ by definition. Then, since $X \in \mathbf{E}$ and $P_X \notin \mathbf{E}$, we have that $\langle \{P_X\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle_G^d$, i.e. $P_X \in \mathbf{S}$. This contradicts the supposition that $\rho(X) \cap (V(G) \backslash (\mathbf{T} \cup \mathbf{E} \cup \mathbf{S})) \neq \emptyset$. Hence, $\rho(X) \cap (V(G) \backslash (\mathbf{T} \cup \mathbf{E} \cup \mathbf{S})) = \emptyset$ always holds.

Now, we can apply Lemma 4.1.2 to take the product over the parameters of the nodes in $\delta(\mathbf{S}) \cap \mathbf{E}$ out of the summations over $V(G) \backslash (\mathbf{T} \cup \mathbf{E} \cup \mathbf{S})$:

$$\begin{aligned}
&\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E}) \\
&\quad = \left[\sum_{\substack{V(G)\backslash \\ (\mathbf{T} \cup \mathbf{E} \cup \mathcal{D}(\mathbf{S}))}} \prod_{X \in V(G) \backslash \mathcal{D}(\mathbf{S})} \Pr(X \mid \rho(X)) \cdot \left[\sum_{\substack{\delta(\mathbf{S})\backslash \\ (\mathbf{E} \cup \mathbf{S})}} \prod_{Y \in \delta(\mathbf{S}) \backslash (\mathbf{E} \cup \mathbf{S})} \Pr(Y \mid \rho(Y)) \right] \right] \\
&\qquad \cdot \prod_{W \in \delta(\mathbf{S}) \cap \mathbf{E}} \Pr(W \mid \rho(W)) \cdot \prod_{Z \in \mathbf{S}} \Pr(Z \mid \rho(Z)).
\end{aligned}$$

By applying Lemma 4.1.1, the term over $\delta(\mathbf{S}) \backslash (\mathbf{E} \cup \mathbf{S})$ sums to 1. This gives us

$$\begin{aligned}
&\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E}) \\
&\quad = \left[\sum_{\substack{V(G)\backslash \\ (\mathbf{T} \cup \mathbf{E} \cup \mathcal{D}(\mathbf{S}))}} \prod_{X \in V(G) \backslash \mathcal{D}(\mathbf{S})} \Pr(X \mid \rho(X)) \right] \cdot \prod_{Y \in \mathbf{S} \cup (\delta(\mathbf{S}) \cap \mathbf{E})} \Pr(Y \mid \rho(Y))
\end{aligned}$$

And analogous to $\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E})$, $\Pr(\mathbf{S} \wedge \mathbf{E})$ is:

$$\begin{aligned}
&\Pr(\mathbf{S} \wedge \mathbf{E}) \\
&\quad = \left[\sum_{\substack{V(G)\setminus \\ (\mathbf{E} \cup \mathcal{D}(\mathbf{S}))}} \prod_{X \in V(G)\setminus \mathcal{D}(\mathbf{S})} \Pr(X \mid \rho(X)) \right] \cdot \prod_{Y \in \mathbf{S} \cup (\delta(\mathbf{S}) \cap \mathbf{E})} \Pr(Y \mid \rho(Y))
\end{aligned}$$

This results in

$$\begin{aligned}
\Pr(\mathbf{T} \mid \mathbf{E}) &= \frac{\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E})}{\Pr(\mathbf{S} \wedge \mathbf{E})} \\
&= \frac{\displaystyle\sum_{\substack{V(G)\setminus \\ (\mathbf{T} \cup \mathbf{E} \cup \mathcal{D}(\mathbf{S}))}} \prod_{X \in V(G)\setminus \mathcal{D}(\mathbf{S})} \Pr(X \mid \rho(X))}{\displaystyle\sum_{\substack{V(G)\setminus \\ (\mathbf{E} \cup \mathcal{D}(\mathbf{S}))}} \prod_{X \in V(G)\setminus \mathcal{D}(\mathbf{S})} \Pr(X \mid \rho(X))}
\end{aligned}$$

□

From the above proposition, we have that for all nodes in $\mathbf{S}$ it holds that neither its own parameters nor its children's parameters are needed in the calculation, which allows us to prune all nodes in $\mathbf{S}$. Note that the Bayes-Ball Algorithm, Algorithm 1, can be used to identify all nodes d-separated from $\mathbf{T}$ given $\mathbf{E}$, since $\mathbf{S} = V(G)\setminus(PotEvSens^{\mathbf{E}}(\mathbf{T}) \cup \mathbf{E})$ by the definitions of $EvSens^{\mathbf{E}}(\mathbf{T})$ and $\mathbf{S}$. After pruning this first set of nodes, a subgraph G' of G results where

- $V(G') = \{X \in V(G) \mid X \notin \mathbf{S}\}$
- $A(G') = \{(X \to Y) \in A(G) \mid X \in V(G') \wedge Y \in V(G')\}$

For a schematic summary of pruning all nodes d-separated from $\mathbf{T}$ given $\mathbf{E}$, see Figure 4.1. It is clear that $\mathbf{S} \cap \mathbf{E} = \emptyset$, $EvSens^{\mathbf{E}}(\mathbf{T}) \cap \mathbf{S} = \emptyset$ and $\mathbf{S} \cup EvSens^{\mathbf{E}}(\mathbf{T}) \cup \mathbf{E} = V(G)$ by the definitions of $EvSens^{\mathbf{E}}(\mathbf{T})$ and $\mathbf{S}$.

4.2.2 Pruning barren nodes

Another set of nodes which can be pruned in order to make the graph smaller, but computationally equivalent, is the set of *barren nodes* with respect to $\mathbf{T}$ and $\mathbf{E}$. A barren node is computationally irrelevant to the probability of interest, because no other non-barren nodes are conditioned on a barren node. Barren nodes were first introduced in [12]; the definition used here is from [6].

Definition 4.2.1 (Barren node)**.** *Let $G = (V(G), A(G))$, $\mathbf{T}$ and $\mathbf{E}$ be as defined before. A node $X \in V(G)$ is called* barren *if $X \notin \mathbf{T}$, $X \notin \mathbf{E}$, and all its descendants $\delta^*(X)$ are barren.*

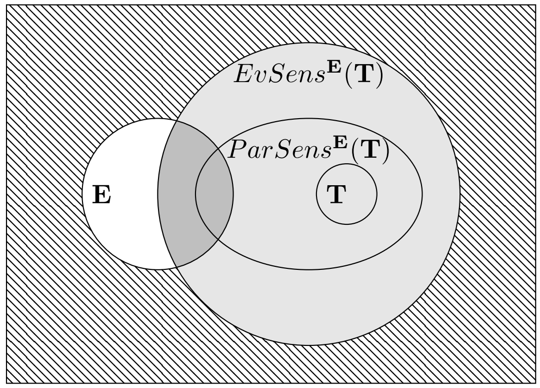


Figure 4.1: An Euler diagram representing all nodes in $V(G') \subseteq V(G)$, after pruning all nodes d-separated from $\mathbf{T}$ given $\mathbf{E}$; the hatched area represents the set of pruned nodes.

According to the definition of barren nodes, it is not necessary to make the distinction between nodes in $\mathbf{T}$ and $\mathbf{E}$. In other words, if a node of interest later becomes observed, this does not change the set of barren nodes. This gives us the possibility to define a *barren set* in the following way:

Definition 4.2.2 (Barren set)**.** *Let $G = (V(G), A(G))$, $\mathbf{T}$ and $\mathbf{E}$ be as defined before. $BarrenSet(\mathbf{X})$ contains all nodes $X \in V(G)$, such that $X \notin \mathbf{X}$ and all its descendants $\delta^*(X)$ are also in $BarrenSet(\mathbf{X})$.*

Now, let $\mathbf{B} \subseteq V(G)$ consist of all barren nodes with respect to $\mathbf{T}$ and $\mathbf{E}$, that is,

$$\mathbf{B} = BarrenSet(\mathbf{T} \cup \mathbf{E}).$$

and let $\mathbf{B}^\Delta \subseteq V(G)$ consist of all barren nodes with respect to $\mathbf{E}$ only, that is,

$$\mathbf{B}^\Delta = BarrenSet(\mathbf{E}).$$

We will call these nodes *semi-barren* with respect to $\mathbf{T}$ and $\mathbf{E}$. Note that $\mathbf{B} \cap (\mathbf{T} \cup \mathbf{E}) = \emptyset$, $\mathbf{B}^\Delta \cap \mathbf{E} = \emptyset$, $\mathbf{B} \subseteq \mathbf{B}^\Delta$, $\delta(\mathbf{B}) \subseteq \mathbf{B}$ and $\delta(\mathbf{B}^\Delta) \subseteq \mathbf{B}^\Delta$ hold by definition.

We will now prove, by means of probability calculations, that barren nodes with respect to $\mathbf{T}$ and $\mathbf{E}$ can safely be pruned. Moreover, we will show that the semi-barren nodes with respect to $\mathbf{T}$ and $\mathbf{E}$ can be disregarded in the calculation of the denominator as well.

Proposition 4.2.2. *Let $G = (V(G), A(G))$ and $\Gamma = \{\Pr(X \mid \rho(X)) \mid X \in$*

$V(G)\}$ be as defined before. Let $\mathbf{B}$ and $\mathbf{B}^{\Delta}$ be as defined above. Then,

$$Pr(\mathbf{T} \mid \mathbf{E}) = \frac{\sum\limits_{\substack{V(G)\setminus \\ (\mathbf{T}\cup\mathbf{E}\cup\mathbf{B})}} \prod\limits_{X \in V(G)\setminus\mathbf{B}} \Pr(X \mid \rho(X))}{\sum\limits_{\substack{V(G)\setminus \\ (\mathbf{E}\cup\mathbf{B}^{\Delta})}} \prod\limits_{X \in V(G)\setminus\mathbf{B}^{\Delta}} \Pr(X \mid \rho(X))}.$$

Proof. We recall, that

$$\begin{aligned}\Pr(\mathbf{T} \mid \mathbf{E}) &= \frac{\Pr(\mathbf{T} \wedge \mathbf{E})}{\Pr(\mathbf{E})} \\ &= \frac{\sum\limits_{V(G)\setminus(\mathbf{T}\cup\mathbf{E})} \prod\limits_{X \in V(G)} \Pr(X \mid \rho(X))}{\sum\limits_{V(G)\setminus\mathbf{E}} \prod\limits_{X \in V(G)} \Pr(X \mid \rho(X))}.\end{aligned}$$

First, we will prove for the numerator, $\Pr(\mathbf{T} \wedge \mathbf{E})$, that the parameters of all nodes in $\mathbf{B}$ are not needed in the calculation. Since

- $(V(G)\setminus(\mathbf{T} \cup \mathbf{E} \cup \mathbf{B})) \cup \mathbf{B} = V(G)\setminus(\mathbf{T} \cup \mathbf{E})$, and
- $(V(G)\setminus(\mathbf{T} \cup \mathbf{E} \cup \mathbf{B})) \cap \mathbf{B} = \emptyset$, and
- $(V(G)\setminus\mathbf{B}) \cup \mathbf{B} = V(G)$, and
- $(V(G)\setminus\mathbf{B}) \cap \mathbf{B} = \emptyset$

hold, we can rephrase $\Pr(\mathbf{T} \wedge \mathbf{E})$ as:

$$\Pr(\mathbf{T} \wedge \mathbf{E}) = \sum_{\substack{V(G)\setminus \\ (\mathbf{T}\cup\mathbf{E}\cup\mathbf{B})}} \sum_{\mathbf{B}} \prod_{X \in V(G)\setminus\mathbf{B}} \Pr(X \mid \rho(X)) \cdot \prod_{Y \in \mathbf{B}} \Pr(Y \mid \rho(Y)).$$

We want to take the product over the parameters of nodes in $V(G)\setminus\mathbf{B}$ out of the summation over $\mathbf{B}$. Note that since $\delta(\mathbf{B}) \subseteq \mathbf{B}$ by definition, $\mathcal{D}(\mathbf{B}) = \mathbf{B}$. In order to apply Lemma 4.1.2 to the calculation, we have to show that

1. $(V(G)\setminus\mathbf{B}) \cap \mathbf{B} = \emptyset$, and
2. for all $X \in V(G)\setminus\mathbf{B}$, it holds that $\rho(X) \cap \mathbf{B} = \emptyset$.

Property 1 holds trivially.

Proof of property 2. We assume that $X \in V(G)\setminus\mathbf{B}$. Then, since $\delta(\mathbf{B}) \subseteq \mathbf{B}$ by definition, $X \notin \delta(\mathbf{B})$ and therefore $\rho(X) \cap \mathbf{B} = \emptyset$.

Now, we can apply Lemma 4.1.2 to the calculation of $\Pr(\mathbf{T} \wedge \mathbf{E})$, and Lemma 4.1.1 to sum the inner term to 1:

$$\begin{aligned}\Pr(\mathbf{T} \wedge \mathbf{E}) &= \left[\sum_{\substack{V(G)\setminus\\(\mathbf{T}\cup\mathbf{E}\cup\mathbf{B})}} \prod_{X \in V(G)\setminus\mathbf{B}} \Pr(X \mid \rho(X)) \cdot \left[\sum_{\mathbf{B}} \prod_{Y \in \mathbf{B}} \Pr(Y \mid \rho(Y))\right]\right] \\ &= \sum_{\substack{V(G)\setminus\\(\mathbf{T}\cup\mathbf{E}\cup\mathbf{B})}} \prod_{X \in V(G)\setminus\mathbf{B}} \Pr(X \mid \rho(X)) \cdot 1.\end{aligned}$$

Analogous to $\Pr(\mathbf{T} \wedge \mathbf{E})$, where $\mathbf{B}$ is replaced by $\mathbf{B}^{\triangle}$, it follows that

$$\Pr(\mathbf{E}) = \sum_{\substack{V(G)\setminus\\(\mathbf{T}\cup\mathbf{E}\cup\mathbf{B}^{\triangle})}} \prod_{X \in V(G)\setminus\mathbf{B}^{\triangle}} \Pr(X \mid \rho(X)) \cdot 1.$$

□

From the above proposition, we have that it holds for all nodes in $\mathbf{B}$ that its own parameters are not needed in the calculation. Since, by the definition of a barren set $\mathbf{B}$, $\delta(\mathbf{B}) \subseteq \mathbf{B}$, we have that the parameters of all children of nodes in $\mathbf{B}$ are not needed in the calculation either. This allows us to prune all nodes in $\mathbf{B}$.

Note that the parameters of nodes in $\mathbf{B}^{\triangle}\setminus\mathbf{B}$ are needed in the calculation of the numerator, but they are not needed in the calculation of the denominator. Therefore, pruning all nodes in $\mathbf{B}^{\triangle}$ is not allowed, but it gives more information about the structure of the calculation, which we will use in Chapter 5.

A very nice corollary, implied by the definitions of barren set in this chapter and the sensitivity sets in the previous chapter, is the following. After pruning all d-separated nodes of $\mathbf{T}$ given $\mathbf{E}$, the barren set with respect to $\mathbf{T}$ and $\mathbf{E}$ contains all nodes in the potential evidence sensitivity set not contained in the parameter sensitivity set for $\mathbf{T}$ given $\mathbf{E}$, more formally,

Corollary 4.2.1. *Let $G = (V(G), A(G))$, $\mathbf{T}$ and $\mathbf{E}$ be as defined before. Let $\mathbf{S}$ denote all nodes d-separated from $\mathbf{T}$ given $\mathbf{E}$. Then,*

$$\mathit{BarrenSet}(\mathbf{T} \cup \mathbf{E})\setminus\mathbf{S} = \mathit{PotEvSens}^{\mathbf{E}}(\mathbf{T})\setminus\mathit{ParSens}^{\mathbf{E}}(\mathbf{T})$$

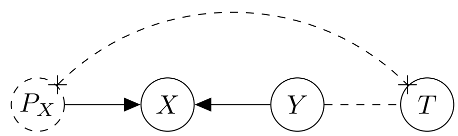
Proof. We will prove subsequently,

1. $\mathit{BarrenSet}(\mathbf{T} \cup \mathbf{E})\setminus\mathbf{S} \subseteq \mathit{PotEvSens}^{\mathbf{E}}(\mathbf{T})\setminus\mathit{ParSens}^{\mathbf{E}}(\mathbf{T})$,
2. $\mathit{BarrenSet}(\mathbf{T} \cup \mathbf{E})\setminus\mathbf{S} \supseteq \mathit{PotEvSens}^{\mathbf{E}}(\mathbf{T})\setminus\mathit{ParSens}^{\mathbf{E}}(\mathbf{T})$.

Proof of property 1.
Assume that $X \in \mathit{BarrenSet}(\mathbf{T} \cup \mathbf{E})\setminus\mathbf{S}$, i.e. $X \notin \mathbf{T}$, $X \notin \mathbf{E}$, $X \notin \mathbf{S}$, and

$\delta^*(X) \cap (\mathbf{T} \cup \mathbf{E}) = \emptyset$. It then holds trivially that $\neg \langle \{X\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle^d_G$. Therefore, $X \in PotEvSens^{\mathbf{E}}(\mathbf{T})$, by definition.

Now, we want to show that $X \notin ParSens^{\mathbf{E}}(\mathbf{T})$, that is, $\langle \{P_X\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle^d_{G^*}$ holds in the parented graph G^* of G. This situation is illustrated in the following graph:



where P_X is the parent added to X in G^*. Since X is d-connected to a $T \in \mathbf{T}$ given $\mathbf{E}$, and $\delta^*(X) \cap (\mathbf{T} \cup \mathbf{E}) = \emptyset$, we know that X has a parent $Y \in V(G)$ d-connected to T. Therefore, G^* includes a chain between P_X and T in which X is a head-to-head node. Since $\delta^*(X) \cap \mathbf{E} = \emptyset$, we have that $\langle \{P_X\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle_{G^*}$ holds and therefore $X \notin ParSens^{\mathbf{E}}(\mathbf{T})$.

Proof of property 2.
Assume that $X \in PotEvSens^{\mathbf{E}}(\mathbf{T})$, i.e. $\neg \langle \{X\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle^d_G$, and that $X \notin ParSens^{\mathbf{E}}(\mathbf{T})$, i.e. $\langle \{P_X\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle^d_{G^*}$ holds. Now, we want to show that $X \in BarrenSet(\mathbf{T} \cup \mathbf{E}) \backslash \mathbf{S}$.

Since X is d-connected to a $T \in \mathbf{T}$ given $\mathbf{E}$, we have that $X \notin \mathbf{S}$. Since the added P_X in the parented graph is d-separated from $\mathbf{T}$ given $\mathbf{E}$, we know that node X has to be a head-to-head node on a blocked chain. In other words, $\delta^*(X) \cap (\mathbf{E} \cup \mathbf{T}) = \emptyset$. Therefore, $X \in BarrenSet(\mathbf{T} \cup \mathbf{E}) \backslash \mathbf{S}$. $\square$

Note that by Corollary 4.2.1 the Bayes-Ball Algorithm, Algorithm 1, can be used to identify all d-connected barren nodes with respect to $\mathbf{T}$ and $\mathbf{E}$. After pruning all nodes d-separated from $\mathbf{T}$ given $\mathbf{E}$ and all barren nodes with respect to $\mathbf{T}$ and $\mathbf{E}$, the subgraph G^M of G results where

- $V(G^M) = \{X \in V(G) \mid X \notin \mathbf{B} \wedge X \notin \mathbf{S}\}$
- $A(G^M) = \{(X \to Y) \in A(G) \mid X \in V(G^M) \wedge Y \in V(G^M)\}$

From Theorem 4.2.1 we have that G^M is the minimal computationally equivalent subgraph of our original graph G.

For a schematic summary of pruning all nodes d-separated from $\mathbf{T}$ given $\mathbf{E}$ and all barren nodes with respect to $\mathbf{T}$ and $\mathbf{E}$, see Figure 4.2. Note that $\mathbf{B}^{\Delta} \backslash \mathbf{B} \subseteq ParSens^{\mathbf{E}}(\mathbf{T})$, since these nodes are needed in the calculation of the numerator of the probability of interest.

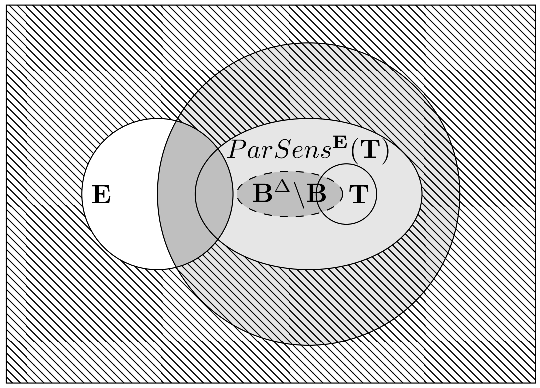


Figure 4.2: An Euler diagram representing all nodes in $V(G^M) \subseteq V(G)$, after pruning all nodes d-separated from $\mathbf{T}$ given $\mathbf{E}$ and all barren nodes with respect to $\mathbf{T}$ and $\mathbf{E}$.

4.3 Investigating evidence nodes

By Theorem 4.2.1, the minimal computationally equivalent subgraph G^M contains no more nodes $X \notin \mathbf{T} \cup \mathbf{E}$ that can be pruned [1]. By taking a close look at the probability calculation, however, we can identify two subsets of $\mathbf{E}$ that capture information not required for the probability calculation. These subsets consist of nodes with, respectively, irrelevant parameters and irrelevant observations.

4.3.1 Evidence nodes with irrelevant parameters

An evidence node with irrelevant parameters is a node $E \in \mathbf{E}$ for which its own parameters are not needed in the probability calculation. In other words, the parameters of this set of nodes are irrelevant, but their observations may be relevant.

Definition 4.3.1 (Evidence nodes with irrelevant parameters)**.** *Let $G = (V(G), A(G))$, $\mathbf{T}$ and $\mathbf{E}$ be as defined before. Let G^* be the parented graph of G. The set of evidence nodes with irrelevant parameters $\mathbf{E}^{\neg\gamma} \subseteq \mathbf{E}$ contains all $E \in \mathbf{E}$ for which it holds that*

$$\langle \{P_E\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle^d_{G^*}$$

Note that $\mathbf{E}^{\neg\gamma}$ contains all $E \in \mathbf{E}$ which are *not* contained in the parameter sensitivity set for $\mathbf{T}$ given $\mathbf{E}$. By Proposition 3.1.1, we know that $\Pr(\mathbf{T} \mid \mathbf{E})$ is insensitive to changes in the parameter values of a node $E \in \mathbf{E}^{\neg\gamma}$. We will define the set of evidence nodes with relevant parameters as $\overline{\mathbf{E}^{\neg\gamma}} = \mathbf{E}\backslash\mathbf{E}^{\neg\gamma}$. Although the nodes in $\mathbf{E}^{\neg\gamma}$ cannot be pruned because of their possible relevancy for the parameters of their children, that is, there may exist $X \in V(G)$ such that

$\rho(X) \cap \mathbf{E}^{\neg\gamma} \neq \emptyset$, identifying these nodes will give us more information about the structure of the calculation, which we will exploit in Chapter 5.

Proposition 4.3.1. *Let $G = (V(G), A(G))$ and $\Gamma = \{\Pr(X \mid \rho(X)) \mid X \in V(G)\}$ be as defined before. Then,*

$$\Pr(\mathbf{T} \mid \mathbf{E}) = \frac{\displaystyle\sum_{\substack{V(G)\setminus\\(\mathbf{T}\cup\mathbf{S}\cup\mathbf{E})}} \prod_{X\in V(G)\setminus\mathbf{E}^{\neg\gamma}} \Pr(X \mid \rho(X))}{\displaystyle\sum_{\substack{V(G)\setminus\\(\mathbf{S}\cup\mathbf{E})}} \prod_{X\in V(G)\setminus\mathbf{E}^{\neg\gamma}} \Pr(X \mid \rho(X))}.$$

Proof. We recall from Equation 4.4, that

$$\begin{aligned}\Pr(\mathbf{T} \mid \mathbf{E}) &= \frac{\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E})}{\Pr(\mathbf{S} \wedge \mathbf{E})} \\ &= \frac{\displaystyle\sum_{\substack{V(G)\setminus\\(\mathbf{T}\cup\mathbf{S}\cup\mathbf{E})}} \prod_{X\in V(G)} \Pr(X \mid \rho(X))}{\displaystyle\sum_{\substack{V(G)\setminus\\(\mathbf{S}\cup\mathbf{E})}} \prod_{X\in V(G)} \Pr(X \mid \rho(X))}.\end{aligned}$$

First, we will prove for the numerator, $\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E})$, that the parameters of all nodes in $\mathbf{E}^{\neg\gamma}$ are not needed in the calculation. Since

- $(V(G)\setminus\mathbf{E}^{\neg\gamma}) \cup \mathbf{E}^{\neg\gamma} = V(G)$, and
- $(V(G)\setminus\mathbf{E}^{\neg\gamma}) \cap \mathbf{E}^{\neg\gamma} = \emptyset$

hold, we can rephrase $\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E})$ as:

$$\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E}) = \sum_{\substack{V(G)\setminus\\(\mathbf{T}\cup\mathbf{S}\cup\mathbf{E})}} \prod_{X\in V(G)\setminus\mathbf{E}^{\neg\gamma}} \Pr(X \mid \rho(X)) \cdot \prod_{Y\in\mathbf{E}^{\neg\gamma}} \Pr(Y \mid \rho(Y)).$$

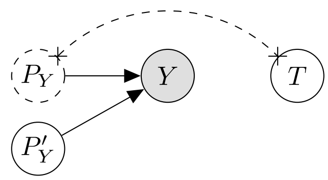
We want to take the product over the parameters of nodes in $\mathbf{E}^{\neg\gamma}$ out of the summation over $V(G)\setminus(\mathbf{T} \cup \mathbf{S} \cup \mathbf{E})$. In order to apply Lemma 4.1.2 to the calculation, we have to show that $(\mathbf{X}\cup\rho(\mathbf{X}))\cap\mathbf{R} = (\mathbf{E}^{\neg\gamma}\cup\rho(\mathbf{E}^{\neg\gamma}))\cap(V(G)\setminus(\mathbf{T}\cup \mathbf{S} \cup \mathbf{E})) = \emptyset$, i.e.

1. $\mathbf{E}^{\neg\gamma} \cap (V(G)\setminus(\mathbf{T} \cup \mathbf{S} \cup \mathbf{E})) = \emptyset$, and
2. for all $Y \in \mathbf{E}^{\neg\gamma}$, it holds that $\rho(Y) \cap (V(G)\setminus(\mathbf{T} \cup \mathbf{S} \cup \mathbf{E})) = \emptyset$.

Property 1 holds trivially.

Proof of property 2.
We assume that $Y \in \mathbf{E}^{\neg\gamma}$, that is, $\langle\{P_Y\} \,|\, \mathbf{E} \,|\, \mathbf{T}\rangle^d_{G^*}$ in the parented graph G^* of G. We will now show by contradiction that $\rho(Y) \subseteq \mathbf{S} \cup \mathbf{E}$ in order to prove the

property. Suppose on the contrary, that $\rho(Y) \not\subseteq \mathbf{S} \cup \mathbf{E}$, in other words, there exists a $P'_Y \in \rho(Y)$, such that $P'_Y \notin \mathbf{S}$ and $P'_Y \notin \mathbf{E}$. This situation is illustrated in the following graph:



Then, given $Y \in \mathbf{E}^{\neg\gamma}$, and the head-to-head connection between P_Y and P'_Y, we must have that $P'_Y \notin \mathbf{T}$ holds. Since $P'_Y \notin \mathbf{S}$, $P'_Y \notin \mathbf{E}$ and $P'_Y \notin \mathbf{T}$, $\neg \langle \{P'_Y\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle^d_G$. But, then, $\neg \langle \{Y\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle^d_G$ always holds, which contradicts the supposition that $Y \in \mathbf{E}^{\neg\gamma}$. Hence, $\rho(Y) \subseteq \mathbf{S} \cup \mathbf{E}$ always holds.

Now, we can apply Lemma 4.1.2 to the calculations of $\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E})$,

$$\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E}) = \left[\sum_{\substack{V(G)\setminus \\ (\mathbf{T}\cup\mathbf{S}\cup\mathbf{E})}} \prod_{X \in V(G)\setminus \mathbf{E}^{\neg\gamma}} \Pr(X \mid \rho(X)) \right] \cdot \prod_{Y \in \mathbf{E}^{\neg\gamma}} \Pr(Y \mid \rho(Y)).$$

And analogous to $\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E})$, $\Pr(\mathbf{S} \wedge \mathbf{E})$ is:

$$\Pr(\mathbf{S} \wedge \mathbf{E}) = \left[\sum_{\substack{V(G)\setminus \\ (\mathbf{S}\cup\mathbf{E})}} \prod_{X \in V(G)\setminus \mathbf{E}^{\neg\gamma}} \Pr(X \mid \rho(X)) \right] \cdot \prod_{Y \in \mathbf{E}^{\neg\gamma}} \Pr(Y \mid \rho(Y)).$$

This results in

$$Pr(\mathbf{T} \mid \mathbf{E}) = \frac{\displaystyle\sum_{\substack{V(G)\setminus \\ (\mathbf{T}\cup\mathbf{S}\cup\mathbf{E})}} \prod_{X \in V(G)\setminus \mathbf{E}^{\neg\gamma}} \Pr(X \mid \rho(X))}{\displaystyle\sum_{\substack{V(G)\setminus \\ (\mathbf{S}\cup\mathbf{E})}} \prod_{X \in V(G)\setminus \mathbf{E}^{\neg\gamma}} \Pr(X \mid \rho(X))}$$

□

From the above proposition, we have that for all nodes in $\mathbf{E}^{\neg\gamma}$ it holds that its own parameters are not needed in the calculation. Because the parameters of its children may be needed, we are not allowed to prune the nodes in $\mathbf{E}^{\neg\gamma}$.

Note that the Bayes-Ball Algorithm, Algorithm 1, can be used to identify all evidence nodes with irrelevant parameters, since $\mathbf{E}^{\neg\gamma} = \mathbf{E} \setminus ParEvSens^{\mathbf{E}}(\mathbf{T})$.

4.3.2 Irrelevant evidence nodes

An irrelevant evidence node is a node $E \in \mathbf{E}$ for which neither its own parameters nor its children's parameters are needed in the probability calculation. In other words, besides the parameters of E, its observation is also irrelevant.

Definition 4.3.2 (Irrelevant evidence nodes)**.** *Let* $G = (V(G), A(G))$*,* $\mathbf{T}$ *and* $\mathbf{E}$ *be as defined before. The set of irrelevant evidence nodes* $\mathbf{E}^{\sigma} \subseteq \mathbf{E}$ *contains all* $E \in \mathbf{E}$ *for which we have that*

$$\langle \{E\} \,|\, \mathbf{E}\backslash\{E\} \,|\, \mathbf{T}\rangle^{d}_{G}$$

Note that $\mathbf{E}^{\sigma}$ contains all $E \in \mathbf{E}$ which are *not* contained in the given evidence sensitivity set for $\mathbf{T}$ given $\mathbf{E}$, as defined in Definition 3.2.1. By Proposition 3.2.1, we know that $\Pr(\mathbf{T} \mid \mathbf{E})$ is insensitive to changes in the observation of a node $E \in \mathbf{E}^{\sigma}$. We will define the set of relevant evidence nodes as the set of evidence nodes which are not irrelevant, that is, $\overline{\mathbf{E}^{\sigma}} = \mathbf{E}\backslash\mathbf{E}^{\sigma}$.

Note that $\mathbf{E}^{\sigma} \subseteq \mathbf{E}^{\neg\gamma}$, as stated in the following corollary.

Corollary 4.3.1. *Let* $G = (V(G), A(G))$*,* $\mathbf{T}$ *and* $\mathbf{E}$ *be as defined before. Then,* $\mathbf{E}^{\sigma} \subseteq \mathbf{E}^{\neg\gamma} \subseteq \mathbf{E}$*.*

Proof. By Proposition 3.2.3 and Definitions 3.2.1, 4.3.1 and 4.3.2:

$$\begin{aligned} & ParSens^{\mathbf{E}}(\mathbf{T}) \cap \mathbf{E} \subseteq GivEvSens^{\mathbf{E}}(\mathbf{T}) \subseteq \mathbf{E} \\ \iff & \mathbf{E}\backslash GivEvSens^{\mathbf{E}}(\mathbf{T}) \subseteq \mathbf{E}\backslash ParSens^{\mathbf{E}}(\mathbf{T}) \subseteq \mathbf{E} \\ \iff & \mathbf{E}^{\sigma} \subseteq \mathbf{E}^{\neg\gamma} \subseteq \mathbf{E}. \end{aligned}$$

□

We will now prove, by means of probability calculations, that all irrelevant evidence nodes can be safely pruned.

Proposition 4.3.2. *Let* $G = (V(G), A(G))$ *and* $\Gamma = \{\Pr(X \mid \rho(X)) \mid X \in V(G)\}$ *be as defined before. Then,*

$$Pr(\mathbf{T} \mid \mathbf{E}) = \frac{\displaystyle\sum_{\substack{V(G)\backslash \\ (\mathbf{T}\cup\mathbf{S}\cup\mathbf{E}\cup\delta(\mathbf{E}^{\sigma}))}} \prod_{X \in V(G)\backslash\mathcal{D}(\mathbf{E}^{\sigma})} \Pr(X \mid \rho(X))}{\displaystyle\sum_{\substack{V(G)\backslash \\ (\mathbf{S}\cup\mathbf{E}\cup\delta(\mathbf{E}^{\sigma}))}} \prod_{X \in V(G)\backslash\mathcal{D}(\mathbf{E}^{\sigma})} \Pr(X \mid \rho(X))}$$

Proof. We recall from Equation 4.4, that

$$\Pr(\mathbf{T} \mid \mathbf{E}) = \frac{\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E})}{\Pr(\mathbf{S} \wedge \mathbf{E})} = \frac{\displaystyle\sum_{\substack{V(G)\backslash \\ (\mathbf{T}\cup\mathbf{S}\cup\mathbf{E})}} \prod_{X \in V(G)} \Pr(X \mid \rho(X))}{\displaystyle\sum_{\substack{V(G)\backslash \\ (\mathbf{S}\cup\mathbf{E})}} \prod_{X \in V(G)} \Pr(X \mid \rho(X))}.$$

First, we will prove for the numerator, $\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E})$, that the parameters of all nodes in $\mathcal{D}(\mathbf{E}^\sigma)$ are not needed in the calculation.

Since $\mathbf{E}^\sigma \subseteq \mathbf{E}^{\neg\gamma}$ by Corollary 4.3.1, we can apply Lemma 4.1.2 to $\mathbf{E}^\sigma$ in the calculation of $\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E})$. The proof is analogous to the proof of applying Lemma 4.1.2 to $\mathbf{E}^{\neg\gamma}$. The resulting equation equals:

$$\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E}) = \left[\sum_{\substack{V(G)\backslash \\ (\mathbf{T}\cup\mathbf{S}\cup\mathbf{E})}} \prod_{X \in V(G)\backslash\mathbf{E}^\sigma} \Pr(X \mid \rho(X))\right] \cdot \prod_{Y \in \mathbf{E}^\sigma} \Pr(Y \mid \rho(Y)).$$

Since

- $(V(G)\backslash(\mathbf{T}\cup\mathbf{S}\cup\mathbf{E}\cup\delta(\mathbf{E}^\sigma))) \cup (\delta(\mathbf{E}^\sigma)\backslash(\mathbf{E}\cup\mathbf{S})) = V(G)\backslash(\mathbf{T}\cup\mathbf{S}\cup\mathbf{E})$ since $\delta(\mathbf{E}^\sigma) \cap \mathbf{T} = \emptyset$, and
- $(V(G)\backslash(\mathbf{T}\cup\mathbf{S}\cup\mathbf{E}\cup\delta(\mathbf{E}^\sigma))) \cap (\delta(\mathbf{E}^\sigma)\backslash(\mathbf{E}\cup\mathbf{S})) = \emptyset$, and
- $(V(G)\backslash\mathcal{D}(\mathbf{E}^\sigma)) \cup (\delta(\mathbf{E}^\sigma)\backslash\mathbf{E}^\sigma) = V(G)\backslash\mathbf{E}^\sigma$, and
- $(V(G)\backslash\mathcal{D}(\mathbf{E}^\sigma)) \cap (\delta(\mathbf{E}^\sigma)\backslash\mathbf{E}^\sigma) = \emptyset$

hold, we can subsequently rephrase $\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E})$ as:

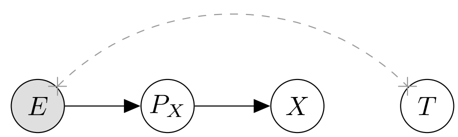
$$\begin{aligned}
&\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E}) \\
&= \left[\sum_{\substack{V(G)\backslash \\ (\mathbf{T}\cup\mathbf{S}\cup\mathbf{E}\cup\delta(\mathbf{E}^\sigma))}} \sum_{\substack{\delta(\mathbf{E}^\sigma)\backslash \\ (\mathbf{E}\cup\mathbf{S})}} \prod_{X \in V(G)\backslash\mathcal{D}(\mathbf{E}^\sigma)} \Pr(X \mid \rho(X)) \cdot \prod_{Y \in \delta(\mathbf{E}^\sigma)\backslash\mathbf{E}^\sigma} \Pr(Y \mid \rho(Y))\right] \\
&\quad \cdot \prod_{Z \in \mathbf{E}^\sigma} \Pr(Z \mid \rho(Z)).
\end{aligned}$$

Note that we now separate the product in such a way, that for all $X \in V(G)\backslash\mathcal{D}(\mathbf{E}^\sigma)$ we have that $\rho(X) \cap \mathbf{E}^\sigma = \emptyset$. Our next step is to take the product over the parameters of nodes in $V(G)\backslash\mathcal{D}(\mathbf{E}^\sigma)$ out of the summation over $\delta(\mathbf{E}^\sigma)\backslash(\mathbf{E}\cup\mathbf{S})$. In order to apply Lemma 4.1.2 to the calculation of $\Pr(\mathbf{T}\wedge\mathbf{S}\wedge\mathbf{E})$, we have to show that $(\mathbf{X}\cup\rho(\mathbf{X}))\cap\mathbf{R} = ((V(G)\backslash\mathcal{D}(\mathbf{E}^\sigma))\cup\rho(V(G)\backslash\mathcal{D}(\mathbf{E}^\sigma)))\cap(\delta(\mathbf{E}^\sigma)\backslash(\mathbf{E}\cup\mathbf{S})) = \emptyset$, i.e.

1. $(V(G)\backslash\mathcal{D}(\mathbf{E}^\sigma)) \cap (\delta(\mathbf{E}^\sigma)\backslash(\mathbf{E}\cup\mathbf{S})) = \emptyset$, and
2. for all $X \in V(G)\backslash\mathcal{D}(\mathbf{E}^\sigma)$, it holds that $\rho(X) \cap (\delta(\mathbf{E}^\sigma)\backslash(\mathbf{E}\cup\mathbf{S})) = \emptyset$.

Property 1 holds trivially.

Proof of property 2.
We assume that $X \in V(G)\backslash\mathcal{D}(\mathbf{E}^\sigma)$, in other words, $X \notin \mathbf{E}^\sigma$ and $X \notin \delta(\mathbf{E}^\sigma)$. We will now show by contradiction that $\rho(X) \cap (\delta(\mathbf{E}^\sigma)\backslash(\mathbf{E}\cup\mathbf{S})) = \emptyset$. Suppose on the contrary, that $\rho(X) \cap (\delta(\mathbf{E}^\sigma)\backslash(\mathbf{E}\cup\mathbf{S})) \neq \emptyset$, in other words, there exists a $P_X \in \rho(X)$, such that $P_X \in \delta(\mathbf{E}^\sigma)$, $P_X \notin \mathbf{E}$ and $P_X \notin \mathbf{S}$. This situation is illustrated in the following graph:



where $T \in \mathbf{T}$ and $E \in \mathbf{E}^\sigma$, i.e. $\langle\{E\}\,|\,\mathbf{E}\backslash\{E\}\,|\,\mathbf{T}\rangle_G^d$ by definition. Now, either $P_X \in \mathbf{T}$ or $\neg\langle\{P_X\}\,|\,\mathbf{E}\,|\,\mathbf{T}\rangle_G^d$ must hold. If $P_X \in \mathbf{T}$, then $\neg\langle\{E\}\,|\,\mathbf{E}\backslash\{E\}\,|\,\mathbf{T}\rangle_G^d$, which contradicts that $E \in \mathbf{E}^\sigma$. If $\neg\langle\{P_X\}\,|\,\mathbf{E}\,|\,\mathbf{T}\rangle_G^d$, then $\neg\langle\{E\}\,|\,\mathbf{E}\backslash\{E\}\,|\,\mathbf{T}\rangle_G^d$, which contradicts that $E \in \mathbf{E}^\sigma$. This contradicts the supposition that $\rho(X) \cap (\delta(\mathbf{E}^\sigma)\backslash(\mathbf{E}\cup\mathbf{S})) \neq \emptyset$. Hence, $\rho(X) \cap (\delta(\mathbf{E}^\sigma)\backslash(\mathbf{E}\cup\mathbf{S})) = \emptyset$ always holds.

Now, we can apply Lemma 4.1.2 to the calculation of $\Pr(\mathbf{T}\wedge\mathbf{S}\wedge\mathbf{E})$:

$$\begin{aligned}
&\Pr(\mathbf{T}\wedge\mathbf{S}\wedge\mathbf{E})\\
&= \left[\sum_{\substack{V(G)\backslash\\(\mathbf{T}\cup\mathbf{S}\cup\mathbf{E}\cup\delta(\mathbf{E}^\sigma))}} \prod_{X\in V(G)\backslash\mathcal{D}(\mathbf{E}^\sigma)} \Pr(X \mid \rho(X)) \cdot \left[\sum_{\substack{\delta(\mathbf{E}^\sigma)\backslash\\(\mathbf{E}\cup\mathbf{S})}} \prod_{Y\in\delta(\mathbf{E}^\sigma)\backslash\mathbf{E}^\sigma} \Pr(Y \mid \rho(Y))\right]\right]\\
&\quad \cdot \prod_{Z\in\mathbf{E}^\sigma} \Pr(Z \mid \rho(Z)).
\end{aligned}$$

Now, we will look closer at the innermost factor, where we want to split out the parameters of nodes in $(\mathbf{S}\cup\overline{\mathbf{E}^\sigma})$. Since it holds that

- $(\delta(\mathbf{E}^\sigma) \cap (\mathbf{S}\cup\overline{\mathbf{E}^\sigma})) \cup (\delta(\mathbf{E}^\sigma)\backslash(\mathbf{E}\cup\mathbf{S})) = \delta(\mathbf{E}^\sigma)\backslash\mathbf{E}^\sigma$, and
- $(\delta(\mathbf{E}^\sigma) \cap (\mathbf{S}\cup\overline{\mathbf{E}^\sigma})) \cap (\delta(\mathbf{E}^\sigma)\backslash(\mathbf{E}\cup\mathbf{S})) = \emptyset$,

we can rephrase the innermost factor as:

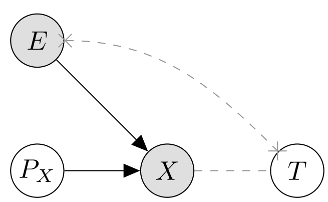
$$\sum_{\substack{\delta(\mathbf{E}^\sigma)\setminus\\(\mathbf{E}\cup\mathbf{S})}} \prod_{Y\in\delta(\mathbf{E}^\sigma)\setminus\mathbf{E}^\sigma} \Pr(Y \mid \rho(Y))$$
$$= \sum_{\substack{\delta(\mathbf{E}^\sigma)\setminus\\(\mathbf{E}\cup\mathbf{S})}} \prod_{X\in\delta(\mathbf{E}^\sigma)\cap(\mathbf{S}\cup\overline{\mathbf{E}^\sigma})} \Pr(X \mid \rho(X)) \cdot \prod_{Y\in\delta(\mathbf{E}^\sigma)\setminus(\mathbf{S}\cup\mathbf{E})} \Pr(Y \mid \rho(Y))$$

We want to apply Lemma 4.1.2, to take the product over the parameters of the nodes in $\delta(\mathbf{E}^\sigma)\cap(\mathbf{S}\cup\overline{\mathbf{E}^\sigma})$ out of *both* of the summations, i.e. the summations over $(V(G)\setminus(\mathbf{T}\cup\mathbf{S}\cup\mathbf{E}\cup\delta(\mathbf{E}^\sigma)))\cup(\delta(\mathbf{E}^\sigma)\setminus(\mathbf{E}\cup\mathbf{S})) = V(G)\setminus(\mathbf{T}\cup\mathbf{E}\cup\mathbf{S})$. Therefore, we have to show that $(X\cup\rho(X)))\cap\mathbf{R} = ((\delta(\mathbf{E}^\sigma)\cap(\mathbf{S}\cup\overline{\mathbf{E}^\sigma}))\cup\rho(\delta(\mathbf{E}^\sigma)\cap(\mathbf{S}\cup\overline{\mathbf{E}^\sigma}))\cap(V(G)\setminus(\mathbf{T}\cup\mathbf{E}\cup\mathbf{S})) = \emptyset$, i.e.

3. $(\delta(\mathbf{E}^\sigma)\cap(\mathbf{S}\cup\overline{\mathbf{E}^\sigma}))\cap(V(G)\setminus(\mathbf{T}\cup\mathbf{E}\cup\mathbf{S})) = \emptyset$, and
4. for all $X\in\delta(\mathbf{E}^\sigma)\cap(\mathbf{S}\cup\overline{\mathbf{E}^\sigma})$, it holds that $\rho(X)\cap(V(G)\setminus(\mathbf{T}\cup\mathbf{E}\cup\mathbf{S})) = \emptyset$.

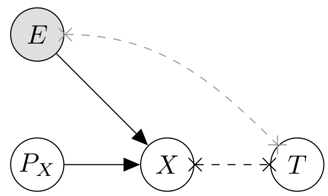
Property 3 holds trivially.

Proof of property 4.
We assume that $X\in\delta(\mathbf{E}^\sigma)\cap(\mathbf{S}\cup\overline{\mathbf{E}^\sigma})$, then either $X\in\mathbf{S}$ or $X\in\overline{\mathbf{E}^\sigma}$. We will now show by contradiction that $\rho(X)\cap(V(G)\setminus(\mathbf{T}\cup\mathbf{E}\cup\mathbf{S})) = \emptyset$. Suppose on the contrary, that $\rho(X)\cap(V(G)\setminus(\mathbf{T}\cup\mathbf{E}\cup\mathbf{S})) \neq \emptyset$, in other words, there exists a $P_X\in\rho(X)$, such that $P_X\notin\mathbf{T}$, $P_X\notin\mathbf{E}$ and $P_X\notin\mathbf{S}$. First, consider the case where $X\in\delta(\mathbf{E}^\sigma)\cap\overline{\mathbf{E}^\sigma}$. This situation is illustrated in the following graph:



where $T\in\mathbf{T}$ and $E\in\mathbf{E}^\sigma$, i.e. $\langle\{E\}\,|\,\mathbf{E}\setminus\{E\}\,|\,\mathbf{T}\rangle^d_G$ by definition. Then, since $X\in\overline{\mathbf{E}^\sigma}$, i.e. $\neg\langle\{X\}\,|\,\mathbf{E}\setminus\{X\}\,|\,\mathbf{T}\rangle^d_G$, and $\langle\{E\}\,|\,\mathbf{E}\setminus\{E\}\,|\,\mathbf{T}\rangle^d_G$, we have that there cannot exist an active chain from P_X to T through a parent of X. Since $P_X\notin\mathbf{E}$, we then have that $\langle\{P_X\}\,|\,\mathbf{E}\,|\,\mathbf{T}\rangle^d_G$ holds, i.e. $P_X\in\mathbf{S}$. This contradicts the supposition that $\rho(X)\cap(V(G)\setminus(\mathbf{T}\cup\mathbf{E}\cup\mathbf{S})) \neq \emptyset$.

Second, we consider the case where $X\in\delta(\mathbf{E}^\sigma)\cap\mathbf{S}$, illustrated in the following graph:



where again $T \in \mathbf{T}$ and $E \in \mathbf{E}^\sigma$. Then, since $X \in \mathbf{S}$ and $P_X \notin \mathbf{E}$, we have that $\langle \{P_X\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle^d_G$, i.e. $P_X \in \mathbf{S}$. This contradicts the supposition that $\rho(X) \cap (V(G) \backslash (\mathbf{T} \cup \mathbf{E} \cup \mathbf{S})) \neq \emptyset$. Hence, $\rho(X) \cap (V(G) \backslash (\mathbf{T} \cup \mathbf{E} \cup \mathbf{S})) = \emptyset$ always holds.

Now, we can apply Lemma 4.1.2 to take the product over the parameters of the nodes in $\delta(\mathbf{E}^\sigma) \cap (\mathbf{S} \cup \overline{\mathbf{E}^\sigma})$ out of the summations over $V(G) \backslash (\mathbf{T} \cup \mathbf{E} \cup \mathbf{S})$:

$$\begin{aligned}
&\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E}) \\
&= \left[\sum_{\substack{V(G)\backslash \\ (\mathbf{T} \cup \mathbf{S} \cup \mathbf{E} \cup \delta(\mathbf{E}^\sigma))}} \prod_{X \in V(G) \backslash \mathcal{D}(\mathbf{E}^\sigma)} \Pr(X \mid \rho(X)) \cdot \left[\sum_{\substack{\delta(\mathbf{E}^\sigma)\backslash \\ (\mathbf{E} \cup \mathbf{S})}} \prod_{Y \in \delta(\mathbf{E}^\sigma) \backslash (\mathbf{E} \cup \mathbf{S})} \Pr(Y \mid \rho(Y)) \right] \right] \\
&\quad \cdot \prod_{W \in \delta(\mathbf{E}^\sigma) \cap (\mathbf{S} \cup \overline{\mathbf{E}^\sigma})} \Pr(W \mid \rho(W)) \cdot \prod_{Z \in \mathbf{E}^\sigma} \Pr(Z \mid \rho(Z)).
\end{aligned}$$

By applying Lemma 4.1.1, the term over $\delta(\mathbf{E}^\sigma) \backslash (\mathbf{E} \cup \mathbf{S})$ sums to 1. This gives us

$$\begin{aligned}
&\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E}) \\
&= \left[\sum_{\substack{V(G)\backslash \\ (\mathbf{T} \cup \mathbf{S} \cup \mathbf{E} \cup \delta(\mathbf{E}^\sigma))}} \prod_{X \in V(G) \backslash \mathcal{D}(\mathbf{E}^\sigma)} \Pr(X \mid \rho(X)) \right] \cdot \prod_{Y \in \mathcal{D}(\mathbf{E}^\sigma) \cap (\mathbf{S} \cup \mathbf{E})} \Pr(Y \mid \rho(Y))
\end{aligned}$$

And analogous to $\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E})$, $\Pr(\mathbf{S} \wedge \mathbf{E})$ is:

$$\begin{aligned}
&\Pr(\mathbf{S} \wedge \mathbf{E}) \\
&= \left[\sum_{\substack{V(G)\backslash \\ (\mathbf{S} \cup \mathbf{E} \cup \delta(\mathbf{E}^\sigma))}} \prod_{X \in V(G) \backslash \mathcal{D}(\mathbf{E}^\sigma)} \Pr(X \mid \rho(X)) \right] \cdot \prod_{Y \in \mathcal{D}(\mathbf{E}^\sigma) \cap (\mathbf{S} \cup \mathbf{E})} \Pr(Y \mid \rho(Y))
\end{aligned}$$

This results in

$$\Pr(\mathbf{T} \mid \mathbf{E}) = \frac{\Pr(\mathbf{T} \wedge \mathbf{S} \wedge \mathbf{E})}{\Pr(\mathbf{S} \wedge \mathbf{E})}$$

$$= \frac{\displaystyle\sum_{\substack{V(G)\setminus \\ (\mathbf{T}\cup\mathbf{S}\cup\mathbf{E}\cup\delta(\mathbf{E}^\sigma))}} \prod_{X\in V(G)\setminus\mathcal{D}(\mathbf{E}^\sigma)} \Pr(X \mid \rho(X))}{\displaystyle\sum_{\substack{V(G)\setminus \\ (\mathbf{S}\cup\mathbf{E}\cup\delta(\mathbf{E}^\sigma))}} \prod_{X\in V(G)\setminus\mathcal{D}(\mathbf{E}^\sigma)} \Pr(X \mid \rho(X))}$$

□

From the above proposition, we have that for all nodes in $\mathbf{E}^\sigma$ it holds that neither its own parameters nor its children's parameters are needed in the calculation, which allows us to prune all nodes in $\mathbf{E}^\sigma$. Note that by Corollary 4.3.1 the Bayes-Ball Algorithm, Algorithm 1, can be used to identify all irrelevant evidence nodes, since $\mathbf{E}^\sigma = \mathbf{E}\setminus GivEvSens^{\mathbf{E}}(\mathbf{T})$.

After pruning all nodes d-separated from $\mathbf{T}$ given $\mathbf{E}$, all barren nodes with respect to $\mathbf{T}$ and $\mathbf{E}$ and all irrelevant evidence nodes, the subgraph G'^M of G results where

- $V(G'^M) = \{X \in V(G) \mid X \notin \mathbf{S} \wedge X \notin \mathbf{B} \wedge X \notin \mathbf{E}^\sigma\}$
- $A(G'^M) = \{(X \to Y) \in A(G) \mid X \in V(G'^M) \wedge Y \in V(G'^M)\}$

For a schematic summary of pruning all nodes d-separated from $\mathbf{T}$ given $\mathbf{E}$, all barren nodes with respect to $\mathbf{T}$ and $\mathbf{E}$ and all irrelevant evidence nodes, see Figure 4.3.

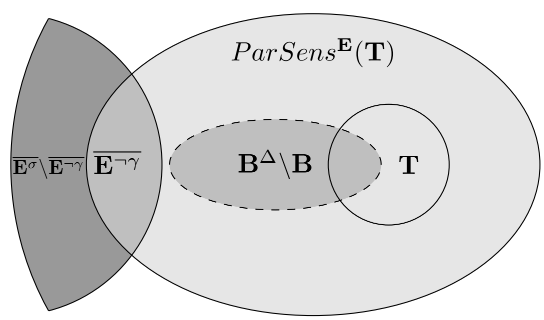


Figure 4.3: An Euler diagram representing all nodes in $V(G'^M) \subseteq V(G)$, after pruning all nodes d-separated from $\mathbf{T}$ given $\mathbf{E}$, all barren nodes with respect to $\mathbf{T}$ and $\mathbf{E}$ and all irrelevant evidence nodes.

4.4 Combining the results

After pruning all nodes as described in this chapter, and identifying other node sets with parameters (partly) irrelevant to the calculation of the probability of interest, we can conclude with the following theorem.

Theorem 4.4.1. *Let $G = (V(G), A(G))$ and $\Gamma = \{\Pr(X \mid \rho(X)) \mid X \in V(G)\}$ be as defined before. Then,*

$$\begin{aligned}\Pr(\mathbf{T} \mid \mathbf{E}) &= \frac{\Pr(\mathbf{T} \wedge \mathbf{E})}{\Pr(\mathbf{E})} \\ &= \frac{\displaystyle\sum_{ParSens^{\mathbf{E}}(\mathbf{T})\backslash(\mathbf{T}\cup\mathbf{E})} \prod_{X \in ParSens^{\mathbf{E}}(\mathbf{T})} \Pr(X \mid \rho(X))}{\displaystyle\sum_{ParSens^{\mathbf{E}}(\mathbf{T})\backslash(\mathbf{E}\cup\mathbf{B}^{\triangle})} \prod_{X \in ParSens^{\mathbf{E}}(\mathbf{T})\backslash\mathbf{B}^{\triangle}} \Pr(X \mid \rho(X))}. \end{aligned} \tag{4.6}$$

Proof. Proposition 4.2.1 implies that the parameters of all nodes in $\mathcal{D}(\mathbf{S})$ can be excluded in the probability of interest calculation. Furthermore, Proposition 4.2.2 implies that the parameters of all nodes in $\mathbf{B}$ and $\mathbf{B}^{\triangle}$ can be excluded in the numerator and denominator of the probability of interest calculation, respectively. Finally, Proposition 4.3.1 implies that the parameters of all nodes in $\mathbf{E}^{\neg\gamma}$ can be excluded in the probability of interest calculation.

Now, we have to show that

1. $V(G)\backslash(\mathcal{D}(\mathbf{S}) \cup \mathbf{B} \cup \mathbf{E}^{\neg\gamma}) = ParSens^{\mathbf{E}}(\mathbf{T})$
2. $V(G)\backslash(\mathcal{D}(\mathbf{S}) \cup \mathbf{B}^{\triangle} \cup \mathbf{E}^{\neg\gamma}) = ParSens^{\mathbf{E}}(\mathbf{T})\backslash\mathbf{B}^{\triangle}$

Note that property 2 always holds when property 1 holds, since $\mathbf{B} \subseteq \mathbf{B}^{\triangle}$.

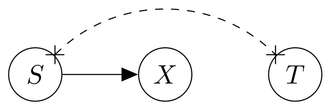
Proof of property 1.
We know by Corollary 4.2.1, and by Definitions 2.1.4, 3.1.2, 3.2.2 and 4.3.1, that:

$$\begin{aligned}\mathbf{S} &= V(G)\backslash(PotEvSens^{\mathbf{E}}(\mathbf{T}) \cup \mathbf{E}), \\ \mathbf{B}\backslash\mathbf{S} &= PotEvSens^{\mathbf{E}}(\mathbf{T})\backslash ParSens^{\mathbf{E}}(\mathbf{T}), \\ \mathbf{E}^{\neg\gamma} &= \mathbf{E}\backslash ParSens^{\mathbf{E}}(\mathbf{T}).\end{aligned}$$

Therefore, we can verify that $V(G)\backslash(\mathbf{S} \cup \mathbf{B} \cup \mathbf{E}^{\neg\gamma}) = ParSens^{\mathbf{E}}(\mathbf{T})$, since

$$\begin{aligned}\mathbf{S} \cup \mathbf{B} \cup \mathbf{E}^{\neg\gamma} &= \mathbf{S} \cup \mathbf{B}\backslash\mathbf{S} \cup \mathbf{E}^{\neg\gamma} \\ &= (V(G)\backslash(PotEvSens^{\mathbf{E}}(\mathbf{T}) \cup \mathbf{E})) \\ &\quad \cup (PotEvSens^{\mathbf{E}}(\mathbf{T})\backslash ParSens^{\mathbf{E}}(\mathbf{T})) \cup (\mathbf{E}\backslash ParSens^{\mathbf{E}}(\mathbf{T})) \\ &= (V(G)\backslash(\mathbf{E} \cup ParSens^{\mathbf{E}}(\mathbf{T}))) \cup (\mathbf{E}\backslash ParSens^{\mathbf{E}}(\mathbf{T})) \\ &= V(G)\backslash ParSens^{\mathbf{E}}(\mathbf{T}).\end{aligned}$$

Now, we have to prove that $\delta(\mathbf{S}) \subseteq \mathbf{S} \cup \mathbf{B} \cup \mathbf{E}^{\neg\gamma}$ in order to proof the proposition. We assume that $\mathbf{X} \in \delta(\mathbf{S})$. This situation is illustrated in the following graph:



where $S \in \mathbf{S}$ and $T \in \mathbf{T}$. Note that $X \notin \mathbf{T}$ holds trivially. Then, there are three possible explanations for the d-separation of node S. First, if $X \in \mathbf{S}$, then S is d-separated. Second, if $X \notin \mathbf{S}$ but $X \in \mathbf{E}$, then $X \notin \overline{\mathbf{E}^{\neg\gamma}}$, since then $\langle \{P_X\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle^d_{G^*}$ must hold. This implies that it may hold that $\neg \langle \{S\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle^d_G$, which contradicts the supposition that $S \in \mathbf{S}$, and therefore $X \in \mathbf{E}^{\neg\gamma}$ must hold. Finally, if $X \notin \mathbf{S}$, $X \notin \mathbf{E}$ and $X \notin \mathbf{T}$, then $\neg \langle \{X\} \,|\, \mathbf{E} \,|\, \mathbf{T} \rangle^d_G$ must hold. Then, S can only be d-separated from T if $X \in \mathbf{B}$. □

4.5 Example

We will end this chapter with an example that demonstrates all concepts introduced in this chapter. We will revisit the graph in Figure 3.2, which is depicted once again in Figure 4.4. We will determine which nodes can be safely pruned. After pruning these nodes, we will give the implications of the pruning steps, and identify the other node sets with (partly) irrelevant parameters to the calculation of the probability of interest.

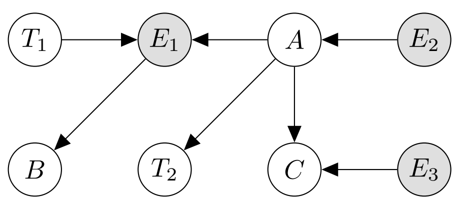


Figure 4.4: Graph of the example probabilistic network.

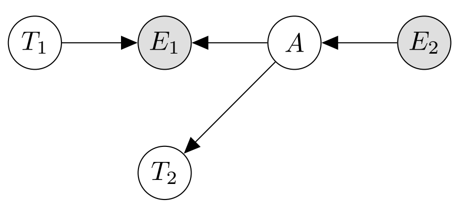


Figure 4.5: Graph of the example probabilistic network after pruning all nodes in $\mathbf{S} \cup \mathbf{B} \cup \mathbf{E}^{\sigma} = \{B, C, E_3\}$.

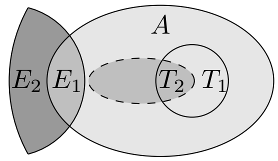


Figure 4.6: An Euler diagram representing the nodes of the example probabilistic network after pruning all nodes in $\mathbf{S} \cup \mathbf{B} \cup \mathbf{E}^{\sigma}$.

First, we will determine all sets of nodes which can be safely pruned from the network, that is, by Figure 4.3 we know that all nodes *not* in $ParSens^{\mathbf{E}}(\mathbf{T}) \cup \overline{\mathbf{E}^{\sigma}}$ can be safely pruned. Recall that, by the Bayes-Ball Algorithm, Algorithm 1, we know that

$$\begin{aligned} PotEvSens^{\mathbf{E}}(\mathbf{T}) &= \{A, C, T_1, T_2\}, \\ GivEvSens^{\mathbf{E}}(\mathbf{T}) &= \{E_1, E_2\}, \\ ParSens^{\mathbf{E}}(\mathbf{T}) &= \{A, E_1, T_1, T_2\}. \end{aligned}$$

We can use these sensitivity sets in order to determine all nodes which can be pruned, since we know from Definition 4.3.2 that $\overline{\mathbf{E}^{\sigma}} = GivEvSens^{\mathbf{E}}(\mathbf{T})$. Then,

$$\begin{aligned} ParSens^{\mathbf{E}}(\mathbf{T}) \cup \overline{\mathbf{E}^{\sigma}} &= ParSens^{\mathbf{E}}(\mathbf{T}) \cup GivEvSens^{\mathbf{E}}(\mathbf{T}) \\ &= \{A, E_1, T_1, T_2\} \cup \{E_1, E_2\} \\ &= \{A, E_1, T_1, T_2, E_1, E_2\}, \\ V(G) \backslash (ParSens^{\mathbf{E}}(\mathbf{T}) \cup \overline{\mathbf{E}^{\sigma}}) &= \{B, C, E_3\}. \end{aligned}$$

That is, nodes B, C and E_3 can be safely pruned from the network. Furthermore, by Definition 4.3.1 and the definition of $\mathbf{B}^{\Delta}$, we have that

$$\begin{aligned} \mathbf{E}^{\neg\gamma} &= \mathbf{E} \backslash ParSens^{\mathbf{E}}(\mathbf{T}) \\ &= \{E_2, E_3\}, \\ \mathbf{E}^{\neg\gamma} \backslash \mathbf{E}^{\sigma} &= \{E_2\}, \\ \mathbf{B}^{\Delta} &= \{B, C, T_2\}, \\ \mathbf{B}^{\Delta} \backslash \mathbf{B} &= \{T_2\}. \end{aligned}$$

That is, besides the parameters of node E_3, the parameters of node E_2 are also irrelevant in the probability of interest calculation. In addition, besides the parameters of nodes B and C, also the parameters of node T_2 are not needed in the denominator of the probability of interest calculation. In Figure 4.6, the remaining nodes after pruning are depicted in the style of Figure 3.1, where all implications of Theorem 4.4.1 are summarized in an Euler diagram.

Now, we will apply our findings to the probability calculation of $\Pr(\mathbf{t} \mid \mathbf{e})$, that is,

$$\begin{aligned}\Pr(\mathbf{t} \mid \mathbf{e}) &= \frac{\Pr(\mathbf{t} \wedge \mathbf{e})}{\Pr(\mathbf{e})} \\ &= \frac{\displaystyle\sum_{ParSens^{\mathbf{E}}(\mathbf{T})\setminus(\mathbf{T}\cup\mathbf{E})} \prod_{X\in ParSens^{\mathbf{E}}(\mathbf{T})} \Pr(X \mid \rho(X))}{\displaystyle\sum_{ParSens^{\mathbf{E}}(\mathbf{T})\setminus(\mathbf{E}\cup\mathbf{B}^{\triangle})} \prod_{X\in ParSens^{\mathbf{E}}(\mathbf{T})\setminus\mathbf{B}^{\triangle}} \Pr(X \mid \rho(X))}.\end{aligned}$$

In the numerator we have to take the product over all nodes in

$$ParSens^{\mathbf{E}}(\mathbf{T}) = \{A, E_1, T_1, T_2\},$$

and we have to sum over all possible value combinations of the nodes in

$$\begin{aligned}ParSens^{\mathbf{E}}(\mathbf{T})\setminus(\mathbf{T}\cup\mathbf{E}) &= \{A, E_1, T_1, T_2\}\setminus\{E_1, E_2, E_3, T_1, T_2\} \\ &= \{A\}.\end{aligned}$$

In the denominator we have to take the product over all nodes in

$$\begin{aligned}ParSens^{\mathbf{E}}(\mathbf{T})\setminus\mathbf{B}^{\triangle} &= \{A, E_1, T_1, T_2\}\setminus\{B, C, T_2\} \\ &= \{A, E_1, T_1\},\end{aligned}$$

and we have to sum over all possible value combinations of the nodes in

$$\begin{aligned}ParSens^{\mathbf{E}}(\mathbf{T})\setminus(\mathbf{E}\cup\mathbf{B}^{\triangle}) &= \{A, E_1, T_1, T_2\}\setminus\{E_1, E_2, E_3, B, C, T_2\} \\ &= \{A, T_1\}.\end{aligned}$$

That is,

$$\begin{aligned}\Pr(\mathbf{t} \mid \mathbf{e}) &= \frac{\displaystyle\sum_{A} \prod_{X\in\{A,E_1,T_1,T_2\}} \Pr(X \mid \rho(X))}{\displaystyle\sum_{A\times T_1} \prod_{X\in\{A,E_1,T_1\}} \Pr(X \mid \rho(X))} \\ &= \frac{\displaystyle\sum_{A} \Pr(A \mid e_2)\Pr(e_1 \mid t_1 \wedge A)\Pr(t_1)\Pr(t_2 \mid A)}{\displaystyle\sum_{A\times T_1} \Pr(A \mid e_2)\Pr(e_1 \mid T_1 \wedge A)\Pr(T_1)}. \qquad (4.7)\end{aligned}$$

Now, only parameters of nodes in the parameter sensitivity set of $\mathbf{T}$ given $\mathbf{E}$ are used in the probability calculation, and in the denominator even less.

4.6 Summary

In this chapter, we explained by means of the probability calculation why the parameters of certain nodes in a probabilistic network are not required when interested in computing a specific probability of interest, and when nodes can even be pruned entirely from the network. This resulted in a simplified equation to calculate a probability of interest, where only parameters of of nodes in the parameter sensitivity set for $\mathbf{T}$ given $\mathbf{E}$ are needed for the numerator, and even less parameters are required in the denominator.

Chapter 5

Evidence sensitivity analysis

In Chapter 3, we used the graphical representation of a probabilistic network to determine *which* nodes may influence the probability of interest upon variation in their observations.

In this chapter, we will use the results of Chapter 4 to locate all terms in the probability of interest calculation that are related to a single evidence node under investigation, in order to define *how* the nodes upon variation in the observations influence the probability of interest. To this end, we will define an evidence sensitivity function, which expresses the probability of interest in terms of the parameters of nodes related to the evidence node under investigation.

5.1 Defining the **evidence sensitivity function**

In this section we assume that there is a single evidence node $E_i \in \mathbf{E}$ which we would like to investigate. The currently observed value for E_i is present in the probability of interest calculation in all conditional probability distributions $\Pr(X \mid \rho(X)) \in \Gamma$ for which either $X = E_i$ or $E_i \in \rho(X)$. Upon changing the observed value for E_i, therefore, different parameters from the distributions for nodes in $\mathcal{D}(E_i)$ will be selected. These distributions involve not only nodes in $\mathcal{D}(E_i)$, but their parents $\rho(\mathcal{D}(E_i))$ also. For changing the observed value for E_i, therefore, we need to consider the entire Markov blanket $\mathcal{MB}(E_i)$ of E_i.

The following lemma shows how to isolate $\mathcal{MB}(E_i)$ in a probability calculation. More specifically, it shows that we can pull in the summation over all variables $\mathbf{Q} \backslash \mathcal{MB}(E_i)$ for any $\mathbf{Q} \subseteq V(G)$, in the probability calculation, which captures the fact that the product over the parameters of all $X \in \mathcal{D}(E_i) \cap \mathbf{X}$ is constant with respect to all variables in $\mathbf{Q} \backslash \mathcal{MB}(E_i)$.

Lemma 5.1.1. *Let $G = (V(G), A(G))$ and $\Gamma = \{\Pr(X \mid \rho(X)) \mid X \in V(G)\}$ be as defined before. Let $\mathbf{Q}, \mathbf{X} \subseteq V(G)$ and $E_i \in V(G)$. Then,*

$$\begin{aligned}
&\sum_{\mathbf{Q}} \prod_{X \in \mathbf{X}} \Pr(X \mid \rho(X)) \\
&\quad = \left[\sum_{\mathcal{MB}(E_i) \cap \mathbf{Q}} \prod_{X \in \mathcal{D}(E_i) \cap \mathbf{X}} \Pr(X \mid \rho(X)) \cdot \left[\sum_{\mathbf{Q} \backslash \mathcal{MB}(E_i)} \prod_{Y \in \mathbf{X} \backslash \mathcal{D}(E_i)} \Pr(Y \mid \rho(Y)) \right] \right]
\end{aligned}$$

Proof. In order to pull in the summation over all variables $\mathbf{Q} \backslash \mathcal{MB}(E_i)$, we want to apply Lemma 4.1.2 to the equation. In order to do that, we have to show that

1. $(\mathcal{MB}(E_i) \cap \mathbf{Q}) \cap (\mathbf{Q} \backslash \mathcal{MB}(E_i)) = \emptyset$,
2. $(\mathcal{D}(E_i) \cap \mathbf{X}) \cap (\mathbf{X} \backslash \mathcal{D}(E_i)) = \emptyset$,
3. $((\mathcal{D}(E_i) \cap \mathbf{X}) \cup \rho(\mathcal{D}(E_i) \cap \mathbf{X})) \cap (\mathbf{Q} \backslash \mathcal{MB}(E_i)) = \emptyset$.

Property 1 and 2 hold trivially.

Proof of property 3.

Suppose $X \in \mathcal{D}(E_i) \cap \mathbf{X}$. Then, by definition,

$$X \in \mathcal{D}(E_i) \implies X \in \mathcal{MB}(E_i) \implies X \notin \mathbf{Q} \backslash \mathcal{MB}(E_i) \text{ for any } \mathbf{Q} \subseteq V(G).$$

Now, suppose $X \in \rho(\mathcal{D}(E_i) \cap \mathbf{X})$. Then, by definition,

$$X \in \rho(\mathcal{D}(E_i)) \implies X \in \mathcal{MB}(E_i) \implies X \notin \mathbf{Q} \backslash \mathcal{MB}(E_i) \text{ for any } \mathbf{Q} \subseteq V(G).$$

Therefore, property 3 holds. □

We will apply Lemma 5.1.1 to Equation 4.6 to pull in all terms which are constant with respect to the so-called evidence sensitivity function parameters. Recall Equation 4.6, the result of the previous chapter, where only parameters of the parameter sensitivity set of $\mathbf{T}$ given $\mathbf{E}$ are used in the probability of interest calculation:

$$\Pr(\mathbf{T} \mid \mathbf{E}) = \frac{\displaystyle\sum_{ParSens^{\mathbf{E}}(\mathbf{T}) \backslash (\mathbf{T} \cup \mathbf{E})} \prod_{X \in ParSens^{\mathbf{E}}(\mathbf{T})} \Pr(X \mid \rho(X))}{\displaystyle\sum_{ParSens^{\mathbf{E}}(\mathbf{T}) \backslash (\mathbf{E} \cup \mathbf{B}^{\triangle})} \prod_{X \in ParSens^{\mathbf{E}}(\mathbf{T}) \backslash \mathbf{B}^{\triangle}} \Pr(X \mid \rho(X))}$$

We will locate all terms in the probability calculation, which contain parameters related to E_i, the node under investigation. Then, all other terms in the probability calculation can be considered as constants.

In order to apply Lemma 5.1.1 to the above equation , we will define the evidence sensitivity function parameters in the following way. First, we define the set $\mathfrak{V}$ of all possible value combinations, the Cartesian product, of the nodes associated with the parameters in $\mathcal{D}(E_i)$ in the numerator:

$$\mathfrak{V} = \bigtimes \mathcal{MB}(E_i) \cap ParSens^{\mathbf{E}}(\mathbf{T}) \backslash (\mathbf{T} \cup \mathbf{E}).$$

Using $\mathfrak{V}$, we can now define the evidence sensitivity function parameters in the numerator as follows.

Definition 5.1.1 (evidence sensitivity function parameters in the numerator)**.** *Let $G = (V(G), A(G))$, $\mathbf{T}$ and $\mathbf{E}$ be as defined before. Let $ParSens^{\mathbf{E}}(\mathbf{T})$ be the parameter sensitivity set for $\mathbf{T}$ given $\mathbf{E}$, let E_i be the evidence node under investigation, and let $\mathfrak{V}$ be defined as before. Then, $x_{\mathbf{v}}^{(\mathbf{t},\mathbf{e}\backslash\{e_i\})}(e_i)$ is a parameter in the evidence sensitivity function's numerator, defined as*

$$x_{\mathbf{v}}^{(\mathbf{t},\mathbf{e}\backslash\{e_i\})}(e_i) = \prod_{X \in \mathcal{D}(E_i) \cap ParSens^{\mathbf{E}}(\mathbf{T})} \Pr(X \mid \rho(X))$$

for any $\mathbf{v} \in \mathfrak{V}$.

Note that $x_{\mathbf{v}}^{(\mathbf{t},\mathbf{e}\backslash\{e_i\})}(e_i)$ depends on the specific value combination $\mathbf{t}$ and $\mathbf{e}$ under consideration for $\mathbf{T}$ and $\mathbf{E}$, respectively. We will, however, write $x_{\mathbf{v}}(e_i)$ for short, if it will not cause confusion.

Second, we define the set $\mathfrak{W}$ of all possible value combinations, the Cartesian product, of the nodes relevant for the parameters in $\mathcal{D}(E_i)$ in the denominator:

$$\mathfrak{W} = \bigtimes \mathcal{MB}(E_i) \cap ParSens^{\mathbf{E}}(\mathbf{T}) \backslash (\mathbf{E} \cup \mathbf{B}^{\Delta}).$$

Using $\mathfrak{W}$, we can now define the evidence sensitivity function parameters in the denominator as follows.

Definition 5.1.2 (evidence sensitivity function parameters in the denominator)**.** *Let $G = (V(G), A(G))$, $\mathbf{T}$ and $\mathbf{E}$ be as defined before. Let $ParSens^{\mathbf{E}}(\mathbf{T})$ be the parameter sensitivity set for $\mathbf{T}$ given $\mathbf{E}$, let $\mathbf{B}^{\Delta}$ be the barren nodes with respect to $\mathbf{E}$, let E_i be the evidence node under investigation, and let $\mathfrak{W}$ defined as before. Then, $y_{\mathbf{w}}^{(\mathbf{e}\backslash\{e_i\})}(e_i)$ is a parameter in the evidence sensitivity function's denominator, defined as*

$$y_{\mathbf{w}}^{(\mathbf{e}\backslash\{e_i\})}(e_i) = \prod_{X \in \mathcal{D}(E_i) \cap ParSens^{\mathbf{E}}(\mathbf{T}) \backslash \mathbf{B}^{\Delta}} \Pr(X \mid \rho(X))$$

for any $\mathbf{w} \in \mathfrak{W}$.

Note that $y_{\mathbf{w}}^{(\mathbf{e}\backslash\{e_i\})}(e_i)$ depends on the specific value combination $\mathbf{e}$ under consideration for $\mathbf{E}$. We will, however, write $y_{\mathbf{w}}(e_i)$ for short, if it will not cause confusion.

Note, that the set of evidence sensitivity function parameters corresponds to a certain value for node E_i. By changing the observation for E_i from, for example, e_i to $\overline{e_i}$, the parameters change from $x_{\mathbf{v}}(e_i)$ and $y_{\mathbf{w}}(e_i)$ to $x_{\mathbf{v}}(\overline{e_i})$ and $y_{\mathbf{w}}(\overline{e_i})$, for all $\mathbf{v} \in \mathfrak{V}$, $\mathbf{w} \in \mathfrak{W}$.

Now, we can define the evidence sensitivity function.

Theorem 5.1.1 (The evidence sensitivity function)**.** *Let $G = (V(G), A(G))$, $\mathbf{T}$ and $\mathbf{E}$ be as defined before. Let E_i be the evidence node under investigation. Then,*

$$\Pr(\mathbf{t} \mid \mathbf{e})(e_i) = \frac{\sum_{\mathbf{v} \in \mathfrak{V}} x_{\mathbf{v}}(e_i) \cdot c_{\mathbf{v}}}{\sum_{\mathbf{w} \in \mathfrak{W}} y_{\mathbf{w}}(e_i) \cdot c'_{\mathbf{w}}}$$

where $c_{\mathbf{v}}, c'_{\mathbf{w}}$ are constants with respect to E_i, and $x_{\mathbf{v}}(e_i), y_{\mathbf{w}}(e_i)$ are the evidence sensitivity function parameters for all $\mathbf{v} \in \mathfrak{V}$, $\mathbf{w} \in \mathfrak{W}$.

Proof. Using Equation 4.6, we can compute a certain probability of interest $\Pr(\mathbf{t} \mid \mathbf{e})$ as follows:

$$\Pr(\mathbf{t} \mid \mathbf{e}) = \frac{\sum_{ParSens^{\mathbf{E}}(\mathbf{T}) \setminus (\mathbf{T} \cup \mathbf{E})} \prod_{X \in ParSens^{\mathbf{E}}(\mathbf{T})} \Pr(X \mid \rho(X))}{\sum_{ParSens^{\mathbf{E}}(\mathbf{T}) \setminus (\mathbf{E} \cup \mathbf{B}^{\Delta})} \prod_{X \in ParSens^{\mathbf{E}}(\mathbf{T}) \setminus \mathbf{B}^{\Delta}} \Pr(X \mid \rho(X))}$$

By applying Lemma 5.1.1 to this equation, and replacing the products representing the evidence sensitivity function parameters by $x_{\mathbf{v}}(e_i)$ and $y_{\mathbf{w}}(e_i)$ as defined in Definitions 5.1.1 and 5.1.2, the resulting equation is:

$$\begin{aligned}
&\Pr(\mathbf{t} \mid \mathbf{e})(e_i) \\
&= \frac{\left[\sum_{\mathfrak{V}} x_{\mathbf{v}}(e_i) \cdot \left[\sum_{\substack{ParSens^{\mathbf{E}}(\mathbf{T}) \setminus \\ (\mathbf{T} \cup \mathbf{E} \cup \mathcal{MB}(E_i))}} \prod_{\substack{X \in ParSens^{\mathbf{E}}(\mathbf{T}) \setminus \\ \mathcal{D}(E_i)}} \Pr(X \mid \rho(X))\right]\right]}{\left[\sum_{\mathfrak{W}} y_{\mathbf{w}}(e_i) \cdot \left[\sum_{\substack{ParSens^{\mathbf{E}}(\mathbf{T}) \setminus \\ (\mathbf{E} \cup \mathbf{B}^{\Delta} \cup \mathcal{MB}(E_i))}} \prod_{\substack{X \in ParSens^{\mathbf{E}}(\mathbf{T}) \setminus \\ (\mathbf{B}^{\Delta} \cup \mathcal{D}(E_i))}} \Pr(X \mid \rho(X))\right]\right]} \\
&= \frac{\sum_{\mathbf{v} \in \mathfrak{V}} x_{\mathbf{v}}(e_i) \cdot c_{\mathbf{v}}}{\sum_{\mathbf{w} \in \mathfrak{W}} y_{\mathbf{w}}(e_i) \cdot c'_{\mathbf{w}}}.
\end{aligned}$$

Since the inner summations in the numerator and denominator are not related to node E_i, due to using Lemma 5.1.1, we can consider these factors as constants with respect to E_i, for all $\mathbf{v} \in \mathfrak{V}$, $\mathbf{w} \in \mathfrak{W}$. $\square$

Therefore, an evidence sensitivity function is always a fraction of two *linear functions* with multiple variables.

5.2 Computing the evidence sensitivity function

To compute the constants of an evidence sensitivity function, previous research results for the parameter sensitivity function can be used. For example, you can compute from the network the probability of interest for as many different observations for node E_i as there are evidence sensitivity function constants minus 1, and subsequently solve the resulting system of multilinear equations to determine the constants. This method is used for computing the parameter sensitivity function in [4]. However, this method is often not applicable to evidence, since it requires a node to have at least as many possible observations as there are evidence sensitivity function constants, that is, the size of the relevant part for calculating the numerator and the denominator of the Markov blanket $\mathcal{MB}(E_i)$, respectively. Note that in most practical applications of probabilistic networks, the Markov blanket of an evidence node is relatively small, since evidence nodes are often root nodes or leaf nodes.

In the next section, we give an example where this approach does work. For the general case, alternative algorithms have to be designed which integrate the computation of constants with standard inference. Such algorithms already exist for computing the parameter sensitivity function in an efficient way [7].

5.3 Example

We will end this chapter with an example that demonstrates all concepts introduced in this chapter. We will revisit the graph in Figure 3.2, which is depicted once again in Figure 5.1. We will investigate the sensitivity of the probability of interest to changes in the observation of nodes E_1, E_2 and E_3, respectively. For each evidence node, we will define the evidence sensitivity function. For evidence node E_2, we will also compute the evidence sensitivity function.

5.3.1 Defining the evidence sensitivity function

We will define the evidence sensitivity function for each evidence node. Recall from Equation 4.7, that

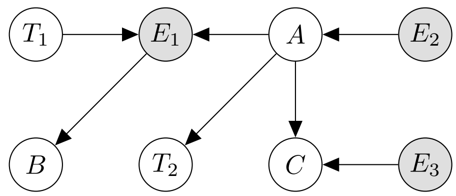


Figure 5.1: Graph of the example probabilistic network.

$$\begin{aligned}
\Pr(\mathbf{t} \mid \mathbf{e}) &= \frac{\Pr(\mathbf{t} \wedge \mathbf{e})}{\Pr(\mathbf{e})} \\
&= \frac{\displaystyle\sum_{ParSens^{\mathbf{E}}(\mathbf{T})\setminus(\mathbf{T}\cup\mathbf{E})} \prod_{X \in ParSens^{\mathbf{E}}(\mathbf{T})} \Pr(X \mid \rho(X))}{\displaystyle\sum_{ParSens^{\mathbf{E}}(\mathbf{T})\setminus(\mathbf{E}\cup\mathbf{B}^{\triangle})} \prod_{X \in ParSens^{\mathbf{E}}(\mathbf{T})\setminus\mathbf{B}^{\triangle}} \Pr(X \mid \rho(X))} \\
&= \frac{\displaystyle\sum_{A} \Pr(A \mid e_2) \Pr(e_1 \mid t_1 \wedge A) \Pr(t_1) \Pr(t_2 \mid A)}{\displaystyle\sum_{A \times T_1} \Pr(A \mid e_2) \Pr(e_1 \mid T_1 \wedge A) \Pr(T_1)}.
\end{aligned}$$

We will apply Lemma 5.1.1 to the above equation to pull in all terms which are constant with respect to the evidence sensitivity function parameters for a single evidence node under investigation, and we will use Theorem 5.1.1 to determine their evidence sensitivity function.

First, we will determine the evidence sensitivity function for node E_1 under investigation. Note that

$$\begin{aligned}
\mathcal{D}(E_1) &= \{B, E_1\}, \\
\mathcal{MB}(E_1) &= \{A, B, E_1, T_1\}.
\end{aligned}$$

Then, the set $\mathfrak{V}$ of all possible value combinations of the nodes associated with the parameters in $\mathcal{D}(E_1)$ in the numerator is:

$$\begin{aligned}
\mathfrak{V} &= \bigtimes \mathcal{MB}(E_1) \cap (ParSens^{\mathbf{E}}(\mathbf{T})\setminus(\mathbf{T} \cup \mathbf{E})) \\
&= \bigtimes \{A, B, E_1, T_1\} \cap (\{A, E_1, T_1, T_2\}\setminus\{E_1, E_2, E_3, T_1, T_2\}) \\
&= \bigtimes \{A, B, E_1, T_1\} \cap \{A\} \\
&= \bigtimes \{A\} \\
&= A.
\end{aligned}$$

The set $\mathfrak{W}$ of all possible value combinations of the nodes relevant for the parameters in $\mathcal{D}(E_1)$ in the denominator is:

$$\begin{aligned}\mathfrak{W} &= \bigtimes \mathcal{MB}(E_1) \cap (ParSens^{\mathbf{E}}(\mathbf{T}) \backslash (\mathbf{E} \cup \mathbf{B}^{\Delta}))\\ &= \bigtimes \{A, B, E_1, T_1\} \cap (\{A, E_1, T_1, T_2\} \backslash \{E_1, E_2, E_3, B, C, T_2\}\\ &= \bigtimes \{A, B, E_1, T_1\} \cap \{A, T_1\}\\ &= \bigtimes \{A, T_1\}\\ &= A \times T_1.\end{aligned}$$

Then, the evidence sensitivity function for node E_1 under investigation equals for $E_1 = e_1$,

$$\begin{aligned}\Pr(\mathbf{t} \mid \mathbf{e})(e_1) &= \frac{\sum_{\mathbf{v} \in \mathfrak{V}} x_{\mathbf{v}}(e_1) \cdot c_{\mathbf{v}}}{\sum_{\mathbf{w} \in \mathfrak{W}} y_{\mathbf{w}}(e_1) \cdot c'_{\mathbf{w}}}\\ &= \frac{\sum_{A} x_{\mathbf{v}}(e_1) \cdot c_{\mathbf{v}}}{\sum_{A \times T_1} y_{\mathbf{w}}(e_1) \cdot c'_{\mathbf{w}}}\end{aligned}$$

where $c_{\mathbf{v}}, c'_{\mathbf{w}}$ are constants with respect to E_1, and

$$\begin{aligned}x_{\mathbf{v}}(e_1) &= \prod_{X \in \mathcal{D}(E_1) \cap ParSens^{\mathbf{E}}(\mathbf{T})} \Pr(X \mid \rho(X)) = \Pr(e_1 \mid t_1 \wedge A)\\ y_{\mathbf{w}}(e_1) &= \prod_{X \in \mathcal{D}(E_1) \cap ParSens^{\mathbf{E}}(\mathbf{T}) \backslash \mathbf{B}^{\Delta}} \Pr(X \mid \rho(X)) = \Pr(e_1 \mid T_1 \wedge A)\end{aligned}$$

are the evidence sensitivity function parameters for all $\mathbf{v} \in A$, $\mathbf{w} \in A \times T_1$. We already knew that, by the Bayes-Ball Algorithm, Algorithm 1, $E_1 \in GivEvSens^{\mathbf{E}}(\mathbf{T})$, and therefore we knew that by changing the observation for E_1, the probability of interest may change too. We now know how the evidence sensitivity function parameters change, and therefore we know how the probability of interest changes.

Second, we will determine the evidence sensitivity function for node E_2 under investigation. Note that

$$\begin{aligned}\mathcal{D}(E_2) &= \{A, E_2\},\\ \mathcal{MB}(E_2) &= \{A, E_2\}.\end{aligned}$$

Then, the set $\mathfrak{V}$ of all possible value combinations of the nodes associated with the parameters in $\mathcal{D}(E_2)$ in the numerator is:

$$\begin{aligned}\mathfrak{V} &= \bigtimes \mathcal{MB}(E_2) \cap (ParSens^{\mathbf{E}}(\mathbf{T})\backslash(\mathbf{T} \cup \mathbf{E}))\\ &= \bigtimes \{A, E_2\} \cap (\{A, E_1, T_1, T_2\}\backslash\{E_1, E_2, E_3, T_1, T_2\})\\ &= \bigtimes \{A\}\\ &= A.\end{aligned}$$

The set $\mathfrak{W}$ of all possible value combinations of the nodes relevant for the parameters in $\mathcal{D}(E_2)$ in the denominator is:

$$\begin{aligned}\mathfrak{W} &= \bigtimes \mathcal{MB}(E_2) \cap (ParSens^{\mathbf{E}}(\mathbf{T})\backslash(\mathbf{E} \cup \mathbf{B}^{\Delta}))\\ &= \bigtimes \{A, E_2\} \cap (\{A, E_1, T_1, T_2\}\backslash\{E_1, E_2, E_3, B, C, T_2\}\\ &= \bigtimes \{A\}\\ &= A.\end{aligned}$$

Then, the evidence sensitivity function for node E_2 under investigation equals for $E_2 = e_2$,

$$\Pr(\mathbf{t} \mid \mathbf{e})(e_2) = \frac{\sum_{A} x_{\mathbf{v}}(e_2) \cdot c_{\mathbf{v}}}{\sum_{A} y_{\mathbf{w}}(e_2) \cdot c'_{\mathbf{w}}} \tag{5.1}$$

where $c_{\mathbf{v}}, c'_{\mathbf{w}}$ are constants with respect to E_2, and

$$\begin{aligned}x_{\mathbf{v}}(e_2) &= \prod_{X \in \mathcal{D}(E_2) \cap ParSens^{\mathbf{E}}(\mathbf{T})} \Pr(X \mid \rho(X)) = \Pr(A \mid e_2)\\ y_{\mathbf{w}}(e_2) &= \prod_{X \in \mathcal{D}(E_2) \cap ParSens^{\mathbf{E}}(\mathbf{T})\backslash\mathbf{B}^{\Delta}} \Pr(X \mid \rho(X)) = \Pr(A \mid e_2)\end{aligned} \tag{5.2}$$

are the evidence sensitivity function parameters for all $\mathbf{v} \in A$, $\mathbf{w} \in A$. We knew already that, by the BayesBall Algorithm, Algorithm 1, $E_2 \in GivEvSens^{\mathbf{E}}(\mathbf{T})$, and therefore we knew that by changing the observation for E_2, the probability of interest may change too. We now know the evidence sensitivity function parameters that change, and therefore we know how the probability of interest changes.

Finally, we will determine the evidence sensitivity function for node E_3 under investigation. Note that

$$\begin{aligned}\mathcal{D}(E_3) &= \{C, E_3\},\\ \mathcal{MB}(E_3) &= \{A, C, E_3\}.\end{aligned}$$

Then, the set $\mathfrak{V}$ of all possible value combinations of the nodes associated with the parameters in $\mathcal{D}(E_3)$ in the numerator is:

$$\begin{aligned}
\mathfrak{V} &= \bigtimes \mathcal{MB}(E_3) \cap (ParSens^{\mathbf{E}}(\mathbf{T}) \backslash (\mathbf{T} \cup \mathbf{E})) \\
&= \bigtimes \{A, C, E_3\} \cap (\{A, E_1, T_1, T_2\} \backslash \{E_1, E_2, E_3, T_1, T_2\}) \\
&= \bigtimes \{A, C, E_3\} \cap \{A\} \\
&= \bigtimes \{A\} \\
&= A.
\end{aligned}$$

The set $\mathfrak{W}$ of all possible value combinations of the nodes relevant for the parameters in $\mathcal{D}(E_3)$ in the denominator is:

$$\begin{aligned}
\mathfrak{W} &= \bigtimes \mathcal{MB}(E_3) \cap (ParSens^{\mathbf{E}}(\mathbf{T}) \backslash (\mathbf{E} \cup \mathbf{B}^{\Delta})) \\
&= \bigtimes \{A, C, E_3\} \cap (\{A, E_1, T_1, T_2\} \backslash \{E_1, E_2, E_3, B, C, T_2\} \\
&= \bigtimes \{A, C, E_3\} \cap \{A, T_1\} \\
&= \bigtimes \{A\} \\
&= A.
\end{aligned}$$

Then, the evidence sensitivity function for node E_3 under investigation is

$$\Pr(\mathbf{t} \mid \mathbf{e})(e_3) = \frac{\sum_{A} c_{\mathbf{v}}}{\sum_{A} c'_{\mathbf{w}}}$$

where $c_{\mathbf{v}}, c'_{\mathbf{w}}$ are constants with respect to E_3, and all evidence sensitivity function parameters are equal to 1, since $\mathcal{D}(E_3) \cap ParSens^{\mathbf{E}}(\mathbf{T}) = \emptyset$ and $\mathcal{D}(E_3) \cap ParSens^{\mathbf{E}}(\mathbf{T}) \backslash \mathbf{B}^{\Delta} = \emptyset$. By changing the observation for E_3, there are no evidence sensitivity function parameters changing, and therefore also the probability of interest stays the same. We actually knew this already, since we knew that, by the BayesBall Algorithm, Algorithm 1, that $E_3 \notin GivEvSens^{\mathbf{E}}(\mathbf{T})$.

5.3.2 Computing the evidence sensitivity function

This subsection will demonstrate how we can compute the evidence sensitivity function for evidence node E_2. The conditional probability distributions which will be used for this example, can be found in Table 5.1. Recall from Equations

$\Pr(A \mid E_2)$	E_2	e_2^0	e_2^1	e_2^2	e_2^3	e_2^4	e_2^5	e_2^6
A	a	1	0.95	0.75	0.55	0.25	0.15	0.05
	$\overline{a}$	0	0.05	0.25	0.45	0.75	0.85	0.95

$\Pr(E_1 \mid T_1 \wedge A)$	T_1	t_1	t_1	$\overline{t_1}$	$\overline{t_1}$
	A	a	$\overline{a}$	a	$\overline{a}$
E_1	e_1	0.2	0.1	0.3	0.4
	$\overline{e_1}$	0.8	0.9	0.7	0.6

$\Pr(T_1)$		
T_1	t_1	0.75
	$\overline{t_1}$	0.25

$\Pr(T_2 \mid A)$	A	a	$\overline{a}$
T_2	t_2	0.99	0.8
	$\overline{t_2}$	0.01	0.2

Table 5.1: The conditional probability distributions of all nodes in the parameter sensitivity set for $\mathbf{T}$ given $\mathbf{E}$ of the example probabilistic network.

5.1 and 5.2, that the evidence sensitivity function for node E_2 under investigation is

$$\Pr(\mathbf{t} \mid \mathbf{e})(e_2) = \frac{\sum_A x_{\mathbf{v}}(e_2) \cdot c_{\mathbf{v}}}{\sum_A y_{\mathbf{w}}(e_2) \cdot c'_{\mathbf{w}}}$$

where $c_{\mathbf{v}}, c'_{\mathbf{w}}$ are constants with respect to E_2, and

$$x_{\mathbf{v}}(e_2) = \Pr(A \mid e_2)$$
$$y_{\mathbf{w}}(e_2) = \Pr(A \mid e_2)$$

are the evidence sensitivity function parameters for all $\mathbf{v} \in A$, $\mathbf{w} \in A$.

More specifically, in our example probabilistic network, we have that

$$\Pr(\{t_1, t_2\} \mid \{e_1, e_2, e_3\})(e_2) = \frac{\Pr(a \mid e_2) \cdot c_1 + \Pr(\overline{a} \mid e_2) \cdot c_2}{\Pr(a \mid e_2) \cdot c_3 + \Pr(\overline{a} \mid e_2) \cdot c_4}$$

for any value e_2 of E_2.

Since $\Pr(\overline{a} \mid e_2) = 1 - \Pr(a \mid e_2)$, we can rewrite the function as

$$\begin{aligned}\Pr(\{t_1, t_2\} \mid \{e_1, e_2, e_3\})(e_2) &= \frac{\Pr(a \mid e_2) \cdot (c_1 - c_2) + c_2}{\Pr(a \mid e_2) \cdot (c_3 - c_4) + c_4} \\ &= \frac{\Pr(a \mid e_2) \cdot c'_1 + c_2}{\Pr(a \mid e_2) \cdot c'_3 + c_4}\end{aligned}$$

Moreover, we can reduce the number of required constants to three:

$$\begin{aligned}\Pr(\{t_1,t_2\} \mid \{e_1,e_2,e_3\})(e_2) &= \frac{\Pr(a \mid e_2)\cdot\frac{c_1'}{c_1'}+\frac{c_2}{c_1'}}{\Pr(a \mid e_2)\cdot\frac{c_3'}{c_1'}+\frac{c_4}{c_1'}}\\ &= \frac{\Pr(a \mid e_2)+c_2}{\Pr(a \mid e_2)\cdot c_3+c_4}\end{aligned}$$

for new constants c_2, c_3 and c_4.

To compute these constants we use any standard inference algorithm on our example network to find three posteriors for three different values of e_2:

$$\begin{aligned}\Pr(\mathbf{t} \mid e_1, e_2^0, e_3) &= 0.81\\ \Pr(\mathbf{t} \mid e_1, e_2^2, e_3) &= 0.78978261\\ \Pr(\mathbf{t} \mid e_1, e_2^4, e_3) &= 0.7542\end{aligned}$$

where $\mathbf{t} = \{t_1, t_2\}$. This gives us a system of linear equations:

$$\begin{aligned}\Pr(a \mid e_2^0)+c_2 &= 0.81\cdot(\Pr(a \mid e_2^0)\cdot c_3+c_4)\\ \iff 1+c_2 &= 0.81\cdot(c_3+c_4)\end{aligned}$$

and similar equations for e_2^2 and e_2^4.

Solving this system gives us $c_2 = -13.913, c_3 = 2.899$ and $c_4 = -18.841$. Now, we can calculate easily all other probabilities of interest by changing the observation for node E_2 to, for example, e_2^6. Then, with $\Pr(a \mid e_2^6) = 0.05$:

$$\begin{aligned}\Pr(\mathbf{t} \mid \mathbf{e})(e_2^6) &= \frac{\Pr(a \mid e_2^6)-13.913}{2.899\cdot\Pr(a \mid e_2^6)-18.841}\\ &= \frac{0.05-13.913}{2.899\cdot 0.05-18.841}\\ &= 0.7415.\end{aligned}$$

Note that in this specific example, where we can write the evidence sensitivity function as a function with only one evidence sensitivity function parameter. Due to the binary variable A and the small Markov blanket of node E_2, the evidence sensitivity function has the same form as a parameter sensitivity function. This evidence sensitivity function can be analyzed in the same way as a parameter sensitivity function, and previous results as in [10] can be used as an inspiration to analyze the effects of changing the observation for node E_2.

5.4 Summary

In this chapter, we introduced the evidence sensitivity function for a single evidence node under investigation, E_i. This function captures *how* a change in observation will influence the probability of interest. We showed how to isolate $\mathcal{MB}(E_i)$ in a probability calculation, and we applied this to the probability calculation where only parameters of the nodes in the parameter sensitivity set for $\mathbf{T}$ given $\mathbf{E}$ are used. We defined the evidence sensitivity function parameters in such a way, that they capture all nodes for which different parameters of the conditional probability distributions have to be selected, upon changing an observation.

The evidence sensitivity function can be described as a fraction of two linear functions with multiple variables. The form of this function depends only on the properties of all nodes in the Markov blanket $\mathcal{MB}(E_i)$ of E_i. The evidence sensitivity function can be used to perform an evidence sensitivity analysis, by investigating the function's properties.

Chapter 6

Conclusion and further research

In this thesis, we studied the sensitivity of a probability of interest to changes in the evidence in a probabilistic network. We defined the evidence sensitivity set for $\mathbf{T}$ given $\mathbf{E}$, which consists of all nodes for which a change in the observed value, or a change in status of being observed or not, may influence the probability of interest. We proved that a node is in the evidence sensitivity set if and only if it might affect the probability of interest upon undergoing abovementioned changes, based on the graphical properties of a probabilistic network alone. The evidence sensitivity set, as well as the parameter sensitivity set. for $\mathbf{T}$ given $\mathbf{E}$ can be efficiently identified by the Bayes-Ball Algorithm. Although the algorithm was already available, its ability to identify the parameter sensitivity set was, as far as we know, not mentioned before. We proved in addition that it also correctly identifies the by us introduced given evidence sensitivity set and potential evidence sensitivity set. In addition, we proved that the parameter sensitivity set is always a subset of the evidence sensitivity set, which gives us a good understanding of the relations between the sensitivity sets in the probabilistic network. The results can be applied directly in experimental research where probabilistic networks are used, to narrow down the amount of evidence nodes that have to be analyzed in order to perform a full evidence sensitivity analysis.

Furthermore, we explained by means of the probability calculation why the parameters of certain nodes in a probabilistic network are not needed, and when nodes can even be pruned entirely from the network. This resulted in a simple equation to calculate a probability of interest, where only parameters of nodes in the parameter sensitivity set for $\mathbf{T}$ given $\mathbf{E}$ are needed. In the denominator, probably even less parameters are needed, since the by us defined semi-barren nodes can be excluded here. Finally, we introduced the evidence sensitivity function for a single evidence node under investigation. The evidence sensitivity function parameters depend only on parameters of nodes in the 'donna con bambini' of this evidence node, which are also in the parameter sensitivity set for $\mathbf{T}$ given $\mathbf{E}$.

More research can be done on pruning in specific situations, for example when the probabilistic network contains deterministic nodes. Moreover, for the evidence sensitivity function, we limited ourselves to investigating how the change of a single observation influences the probability of interest. It would be interesting to investigate how the change of more than one observation influences the probability of interest. Moreover, we presented one way of computing the evidence sensitivity function constants by solving a system of linear equations. However, a more sophisticated and practical way may be to compute the constants directly during propagation in an inference algorithm using a junction tree, which already exists for computing the parameter sensitivity function [7]. Furthermore, more research on how we have to interpret the evidence sensitivity function to measure the robustness and reliability of the probabilistic network is necessary. We expect that the existing methods from the parameter sensitivity analysis as in [10] are a good inspiration. Finally, since we limited ourselves to investigating hard evidence nodes only, it can be even more interesting to investigate the evidence sensitivity function of an evidence node in a probabilistic network where soft evidence is allowed.

Bibliography

[1] Michelle Baker and Terrance Boult. Pruning bayesian networks for efficient computation. In *Uncertainty in Artificial Intelligence 6 Annual Conference on Uncertainty in Artificial Intelligence (UAI-90)*, pages 225–232. Elsevier Science, 1990.

[2] Hei Chan and Adnan Darwiche. Sensitivity analysis in bayesian networks: From single to multiple parameters. In *UAI*, pages 67–75, 2004.

[3] Alan Clift and Alfie Meats. Use of a bayesian belief network to identify situations that favour fruit fly incursions in inland SE Australia. *MODSIM 2005 International Congress on Modelling and Simulation*, 2006.

[4] Veerle M. H. Coupé and Linda C. van der Gaag. Properties of sensitivity analysis of bayesian belief networks. *Annals of Mathematics and Artificial Intelligence*, 36(4):323–356, 2002.

[5] Finn V. Jensen and Thomas D. Nielsen. *Bayesian Networks and Decision Graphs*. Information Science and Statistics. Springer, 2007.

[6] Uffe B. Kjærulff and Anders L. Madsen. *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis*. Information Science and Statistics. Springer-Verlag New York, 2008.

[7] Uffe B. Kjærulff and Linda C. van der Gaag. Making sensitivity analysis computationally efficient. *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, pages 317–325, 2000.

[8] Steffen L. Lauritzen and David J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 50(2):157–224, 1988.

[9] Yan Lin and Marek Druzdzel. Computational advantages of relevance reasoning in bayesian belief networks. In *Proceedings of the Thirteenth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-97)*, pages 342–350. Morgan Kaufmann, 1997.

[10] Silja Renooij Linda C. van der Gaag and Veerle M.H. Coupé. *Sensitivity Analysis of Probabilistic Networks*, volume 213. Springer Verlag, 2007.

[11] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann Publishers Inc., 1988.

[12] Ross D. Shachter. Evaluating influence diagrams. *Operations Research*, 33(6), 1986.

[13] Ross D. Shachter. *Bayes-ball: The rational pastime (for determining irrelevance and requisite information in belief networks and influence diagrams)*. Morgan Kaufmann, 1998.

[14] Nikolaos A. Vagias. *A Bayesian Network Application for the Prediction of Human Fatigue in the Marine Industry*. National Technical University of Athens, 2010.

[15] Linda C. van der Gaag and Silja Renooij. *Probabilistic Reasoning Syllabus*. Utrecht University, 2009.