Gaining insight in the solution space of the MPE problem when changing evidence or parameter values

Master Thesis Computing Science

Author
Rik van Rooij (3685381)
R.P.vanRooij@uu.nl

Primary Supervisor:
dr. S. Renooij
S.Renooij@uu.nl

Secondary Supervisor:
prof. dr. ir. L. C. van der Gaag
L.C.vandergaag@uu.nl

Departement Informatica
Universiteit Utrecht
Thesisnumber: ICA-3685381
November 12, 2015

Gaining insight in the solution space of the MPE problem when changing evidence or parameter values

Rik van Rooij
Universiteit Utrecht
3685381

Abstract

For Bayesian networks, the MPE problem is the problem of finding a configuration of all unobserved variables such that this configuration has the highest posterior probability given the evidence. In this paper, we aim to gain more insight in the solution space of the MPE problem when evidence or a parameter value is changed and the previous MPE solution is known. Gaining more insight is a requisite for developing better algorithms for solving the problem. We do so by introducing a new representation of the probabilities of configurations in the Bayesian network.

I. INTRODUCTION

To represent probabilistic interactions between variables in a probabilistic model, Bayesian networks are often used. For example, the probabilistic interaction between symptoms and diseases can be represented by a Bayesian network. The power of Bayesian networks lies in the fact that conditional probabilities over the variables can easily be calculated even when evidence, a group of variables of which the value is observed in the real world, is involved. The network can be used for example to calculate probabilities of diseases given an evidence set of symptoms. Moreover, the space required for storing a Bayesian network is much smaller than conventional notation of the problem [6].

A Bayesian network consists of a Directed Acyclic Graph (DAG) and probability tables. The nodes in the graph represent the variables, whereas the probability tables give the probabilities of how often certain values appear. A configuration of a Bayesian network is a value assignment to all the variables in the network.

Evidence variables are variables of which the value can be observed in the real world. In the Bayesian network, the value of a such variable can then be set to the observed value. A Most Probable Explanation (MPE) is then a configuration of all non-evidence variables, which has the highest posterior probability out of all possible configurations. Informally, the MPE can be seen as the most probable explanation of the found evidence. For example, in the example with diseases and symptoms the MPE can be used to find the most likely disease or set of diseases a patient has, based on the observed symptoms.

To find an MPE, simply taking the value with the highest (marginal) probability of each variable often yields a wrong result. As an illustration, take playing in a lottery where over half the tickets have been sold. Individually, each player has a very low chance of winning, and a very high chance of losing. By simply giving each player the value with the highest probability, the resulting answer will be that every player loses, which is untrue since the lottery is more likely to have a winner than not. When a player has bought two lottery tickets, he has a higher chance of winning than a player who has bought only one. An MPE in this case would be that every player loses except the player who bought two tickets, since that player is the most likely to win. Finding an MPE solution is thus more complex than it initially seems. In fact, it is NP-complete [3]. Currently, a few algorithms exist which are linear in the tree width of the underlying graph [7].

The solution space of the MPE problem, the space containing all possible configurations eligible to be the MPE solution, has had little attention so far. Knowledge of this solution space will lead to a better understanding of the MPE problem in general, and might lead to better algorithms which solve the MPE problem. In this paper the question is addressed whether in the solution space some configurations can be excluded from becoming the MPE solution, and thus from the solution space, after evidence or a parameter value has changed. In other words, is it possible to reduce the size of the solution space of the MPE problem?

Initially, an experimental approach was used. An idea was inspired by the work by Pastink and Van der Gaag [5]. If evidence or parameter values would change, could change in the solution space be restricted such that some cliques where excluded from change? For a random network with evidence the MPE solution space was calculated. After evidence was changed, the new solution space was calculated for the network. This procedure was repeated for different networks. The goal was to compare the new solution spaces, and see if a common rule based on the cliques could be found which all solution spaces agree to. This would indicate a way the cliques influence the solution space.

Due to the fact that we were interested in cliques, the networks had to be of a minimum size to obtain non trivial cliques. The exponential nature of the solution size however resulted in very large solution spaces, which could not easily be compared to one another. (Some solutions spaces had over 500.000 solutions.) However, within some solution spaces we found seemingly random configurations having the same probability. Looking into this, we found that this could be explained by the decomposition of the Bayesian network. This prompted the basis for our theoretical work. Instead of using cliques, the theoretical work uses Markov blankets in combination with the decomposition of the joint probability given by the Bayesian network to research the solution space of the MPE problem.

That the probability of two configurations that differ in value in only one variable can be related by a simple multiplication was already noted by Park and Darwiche [4]. *"When changing the state of variable X from x to x', the only values in the product that change are those from the CPTs of X and its children. If none of the CPT entries are 0, Pr(s-X,x',e) can be computed by dividing Pr(s,e) by the old and multiplying by the new entry for the CPTs for X and its children. "*

Based on this observation, we define a new representation which encodes the probability of configurations. This so called base-neighbour representation consists of a part which is shared across all probabilities, and a part which is a multiplier, which in itself depends on the encoded probability.

We propose a way to exclude several configurations from being the new MPE solution once evidence or a parameter value has changed. After defining the base-neighbour representation we use this representation to show how several configurations can be excluded. Moreover, the results are compared to related work.

In Section II some preliminaries are given. Then, the base-neighbour representation is explained in Section III, after which we explain how to exclude configurations when changing evidence in Section IV, and exclude configurations when a parameter value is changed in Section V. Finally, in Section VI, we present the discussion and conclusions.

II. Preliminaries

In the preliminaries, formal notation regarding Bayesian networks and explanations are introduced. Furthermore, the notation of variables and values is explained.

A. Bayesian network

A Bayesian network, formally defined below, is a way to represent a probability distribution on a set of random variables, and their dependencies, in a Directed Acyclic Graph (DAG) and Conditional Probability Tables (CPTs) [6].

Definition 1. Bayesian network A Bayesian network consists of a tuple with three values: $(\boldsymbol{V}, \boldsymbol{E}, \boldsymbol{T})$. In this tuple, $(\boldsymbol{V}, \boldsymbol{E})$ is a DAG with n nodes $\boldsymbol{V}$, which represent the random variables. The set of arcs in the DAG is given by $\boldsymbol{E}$. $\boldsymbol{T}$ is a set of conditional probability distributions: $\boldsymbol{T} = \{\Pr(V_1|\Pi(V_1)), ..., \Pr(V_n|\Pi(V_n))\}$. $\Pr(V_i|\Pi(V_i))$ denotes the set of conditional probability distributions on $V_i \in \boldsymbol{V}$. The notation $\Pi(V_i)$ stands for the parent variables of V_i. To denote a configuration of the parents of V_i, the notation $\pi(V_i)$ is used. The graph $(\boldsymbol{V}, \boldsymbol{E})$ encodes independency assumptions which imply the following joint probability:

$$\Pr(\boldsymbol{v}) = \prod_{i=1}^{n=|V|} \Pr(v_i|\pi(V_i)) \tag{1}$$

where $\Pr(\boldsymbol{v}) = \Pr(V_1 = v_1, ..., V_n = v_n)$

Furthermore, the variable set M_V^+ is defined to include the variables in the Markov Blanket of a node V *and* node V itself. The definition of the Markov Blanket is given in Definition 2 [6]. A configuration of the variables of M_V^+ is denoted by m_V^+.

Definition 2. Markov Blanket A Markov Blanket of a variable V is the variable set $\boldsymbol{M_V}$ such that the following holds: $\Pr(V|\boldsymbol{M_v}, \boldsymbol{W}) = \Pr(V|\boldsymbol{M_v})$ for all $\boldsymbol{W} \notin \boldsymbol{M_v}$. In other words, the Markov Blanket includes all variables which block influence from the other variables in the Bayesian network.

B. Configurations and explanations

Each variable $V_i \in \boldsymbol{V}$ has a set of values associated with it. A configuration c on a Bayesian network is a value assignment to a set of variables $\boldsymbol{V_c} \subset \boldsymbol{V}$. It is represented by a set of values, where for each variable $V_i \in \boldsymbol{V_c}$ a value out of their associated value set is picked.

Throughout this work it is assumed the variables can only take on two different values. Therefore, only a distinction needs to be made between these two values. For variable V, this means that we use v and $\overline{v}$ to denote one of the possible values each.

In a Bayesian network, evidence may be entered into the network. This is accomplished by locking some nodes to the observed value. A configuration on the nodes without evidence is called an explanation:

Definition 3. Explanations Consider a Bayesian network $B = (\boldsymbol{V}, \boldsymbol{E}, \boldsymbol{T})$, where the following evidence e has been observed: $e = \{v_1 v_2 \ldots 0 v_m\} = \{V_1 = v1, V_2 = v_2, \ldots V_m = v_m\}$. An explanation based on evidence e is now defined as a configuration $c = (v_{m+1} v_{m+2} \ldots v_n)$. This configuration denotes the following value assignment to the variables: $V_{m+1} = v_{m+1}$,$V_{m+2} = v_{m+2}$,$\ldots V_n = v_n$

TABLE I
A REVERSE TOPOLOGICAL ORDENING

Variable name	Ordering number
A	4
B	3
C	2
D	1

If no evidence at all is recorded, any configuration on all variables of the Bayesian network can be labelled as explanation.

The Most Probable Explanation for a recorded set of evidence e is defined as the explanation c_{MPE} which has the highest probability of all explanations. [8]

Definition 4. Most Probable Explanation (MPE) The MPE for a set of recorded evidence e is an explanation c_{MPE} on the non-evidence variables, such that $\Pr(c_{MPE}|e) \geq \Pr(x|e)$ where x is any other explanation for the evidence e in the Bayesian network.

It is possible that several explanations exist with the same probability. Therefore, multiple MPEs can exist. In this work, if *an* MPE is mentioned, the concept of MPE is meant, when *the* MPE is mentioned, a specific MPE out of the pool of MPEs is meant.

Instead of trying to find a configuration c on the non-evidence variables such that $\Pr(c|e)$ is maximal, we try to find a configuration c such that $\Pr(c, e)$ is maximal. This yields the same configuration as per the definition of conditional probability, $\Pr(c|e) = \frac{\Pr(c,e)}{\Pr(e)}$ and $\Pr(e)$ is constant for all configurations c.

III. A NEW REPRESENTATION OF THE PROBABILITIES IN A BAYESIAN NETWORK

In this section we propose a new representation of the probabilities of configurations in a Bayesian network. This *base-neighbour* representation will help us gain new insights in the solution space of the MPE problem.

In order to explain the base-neighbour representation, the concepts of base, neighbour, multiplier and fundamental are defined. Notation regarding these concepts is introduced first. Initially, evidence is not considered when explaining the base-neighbour representation. After the new concepts are defined, it is shown how the base-neighbour representation works with evidence.

A. Assumptions on notation

First, some notation is introduced regarding the concepts. An example Bayesian network, as shown in Figure 1, is used as an illustration.

- We consider a Bayesian network B as defined in Section II. We assume that the nodes in $\boldsymbol{V}$ are indexed based on a reverse topological ordering. In other words, the index of each node is higher than the index of its children. Node V_i, for example, can be a child of node V_{i+1}, but not its parent. Furthermore, let $n = |\boldsymbol{V}|$. Thus the nodes are labelled V_1 to V_n. Due to the indexing, V_1 has to be a leaf node and V_n has to be a root node. See Table I for an example reverse topological ordering on our Bayesian network. The reverse topological ordering will prove useful later on.
- Without loss of generality, a value variable V_i can take on is transcribed as v_i if the variable takes on this value in the MPE configuration. Otherwise, the value is transcribed as $\overline{v_i}$.
- $V \sim c$ denotes a configuration of variable V such that the value of V is equal to the value of V in configuration c. For example, $B \sim (abcd)$ entails that $B = b$. A similar reasoning holds for groups of variables.

Furthermore, the notion of a *neighbour in variable* V is defined below.

Definition 5. Neighbour in variable V A configuration c_x of all variables is a *neighbour in variable* V of configuration c_y of all variables if and only if configuration c_x differs from configuration c_y in variable V and is the same in the other variables. The set of neighbours of configuration c_x in variable V is $N^V_{c_x}$. An element of this set is a configuration and is denoted by $n^V_{c_x}$.

For example, in our example Bayesian network the set N^A_{abcd} would be $\{(\overline{a}bcd)\}$.

B. Multipliers between neighbours

As Park and Darwiche have noted, the probabilities of two different configurations can be related by a simple multiplication [4]. Without loss of generality, pick $\Pr(c_x)$ and $\Pr(c_y)$ as two probabilities of configurations. There now is a multiplier, λ, such that $\Pr(c_x) = \Pr(c_y) \cdot \lambda$. This multiplier is the core of the new representation.

In this section a way is shown how the multipliers as described above can be found. We pick a fixed base configuration c_b. Then, we can determine $\Pr(c_x)$ where $c_x \in N^{V_i}_b$ is based on a multiplier which can be deduced from the CPT's of V_i and its children (The so called *donna con bambini*). Such a multiplier is called a *Neighbour Multiplier*.

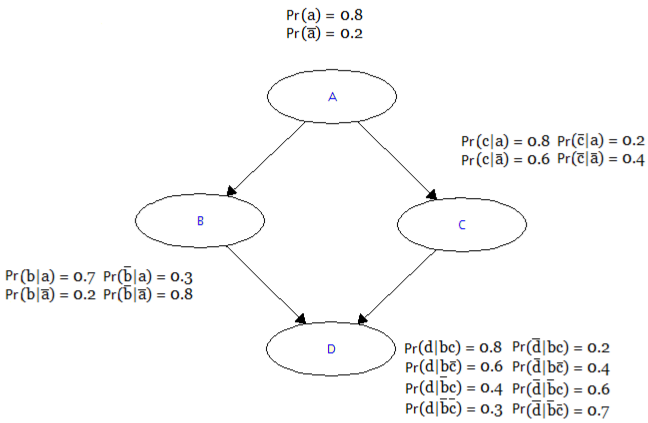

Fig. 1. An example Bayesian network

Definition 6. Neighbour multiplier A neighbour multiplier λ_j^i between a base configuration c_b of the variables in $\boldsymbol{V}$ and a goal configuration c_x of the variables in $\boldsymbol{V}$, where $c_x \in N_{c_b}^{V_i}$ is a value such that $\Pr(c_x) = \Pr(c_b) \cdot \lambda_j^i$. i denotes the *order of the multiplier* (See Definition 7), and j denotes an index to distinguish between multipliers of the same order.

Definition 7. Order of the multiplier A multiplier of order i works on configurations which are neighbours in variable V_i, the ith variable in the ordering on the variables.

Proposition 1 describes how a neighbour multiplier can be calculated.

Proposition 1. Consider a base configuration $c_b = (v_1 v_2 \dots v_n)$ and a configuration $c_x \in N_{c_b}^{V_i}$ Without loss of generality, pick $c_x = (v_1 v_2 \dots \overline{v_i} \dots v_n)$. Then, $\Pr(c_x) = Pr(c_b) \cdot \lambda_j^i$ where λ_j^i is equal to:

$$\lambda_j^i = \frac{\Pr(\overline{v_i}|\pi(V_i) \sim c_x)}{\Pr(v_i|\pi(V_i) \sim c_b)} \cdot \prod_{X \in \sigma(V_i)} \frac{\Pr(X \sim c_x|\pi(X) \sim c_x)}{\Pr(X \sim c_x|\pi(X) \sim c_b)} \tag{2}$$

Note that all variables appearing in the conditional probabilities in Theorem 1 are in the set $M_{V_i}^+$, and indeed, in the *donna con bambini*.

Proof. Theorem 1 can be easily proven by simply substituting the given value for λ_j^i and verifying it is indeed correct.

$$\lambda_j^i \cdot \Pr(c_b) = \frac{\Pr(\overline{v_i}|\pi(V_i) \sim c_x)}{\Pr(v_i|\pi(V_i) \sim c_b)} \cdot \left(\prod_{X \in \sigma(V_i)} \frac{\Pr(X \sim c_x|\pi(X) \sim c_x)}{\Pr(X \sim c_x|\pi(X) \sim c_b)} \right) \cdot \Pr(v_1 v_2 \dots v_i \dots v_n)$$

Using the factorisation of the joint probability distribution defined by a Bayesian network it follows that:

$$\frac{\Pr(\overline{v_i}|\pi(V_i) \sim c_x)}{\Pr(v_i|\pi(V_i) \sim c_b)} \cdot \left(\prod_{X \in \sigma(V_i)} \frac{\Pr(X \sim c_x|\pi(X) \sim c_x)}{\Pr(X \sim c_x|\pi(X) \sim c_b)} \right) \cdot \Pr(v_1 v_2 \dots v_i \dots v_n) =$$
$$\frac{\Pr(\overline{v_i}|\pi(V_i) \sim c_x)}{\Pr(v_i|\pi(V_i) \sim c_b)} \cdot \left(\prod_{X \in \sigma(V_i)} \frac{\Pr(X \sim c_x|\pi(X) \sim c_x)}{\Pr(X \sim c_x|\pi(X) \sim c_b)} \right) \cdot$$
$$\Pr(v_1|\pi(V_1) \sim c_b) \cdot \Pr(v_2|\pi(V_2) \sim c_b) \cdot \dots \cdot \Pr(v_i|\pi(V_i) \sim c_b) \cdot \dots \cdot \Pr(v_n|\pi(V_n) \sim c_b)$$

The factors containing variable V_i in the denominator of the fraction agree with configuration c_b. These factors however, appear also in the factorisation of $\Pr(c_b)$. Therefore, these factors cancel each other out. What remains are the factors which do not contain variable V_i corresponding with c_b, and the factors containing variable V_i corresponding with c_x.

$$\frac{\Pr(\overline{v_i}|\pi(V_i) \sim c_x)}{\Pr(v_i|\pi(V_i) \sim c_b)} \cdot \prod_{X \in Ch(v_i)} \frac{\Pr(X \sim c_x|\pi(X) \sim c_x)}{\Pr(X \sim c_b|\pi(X) \sim c_b)} \cdot$$

$$\Pr(v_1|\pi(V_1) \sim c_b) \cdot \Pr(v_2|\pi(V_2) \sim c_b) \cdot \dots \cdot \Pr(v_i|\pi(V_i) \sim c_b) \cdot \dots \cdot \Pr(v_n|\pi(V_n) \sim c_b) =$$

$$\Pr(v_1|\pi(V_1) \sim c_b) \cdot \Pr(v_2|\pi(V_2) \sim c_b) \cdot \dots \cdot \Pr(\overline{v_1}|\pi(V_i) \sim c_x) \cdot \dots \cdot \Pr(v_n|\pi(V_n) \sim c_b)$$

The configurations c_x and c_b are neighbours in V_i and thus differ only in variable V_i. Therefore, any probability from a CPT not containing V_i will have the same values irrespective of which instantiation is used.

$$\Pr(v_1|\pi(V_1) \sim c_b) \cdot \Pr(v_2|\pi(V_2) \sim c_b) \cdot \ldots \cdot \Pr(\overline{v_1}|\pi(V_i) \sim c_x) \cdot \ldots \cdot \Pr(v_n|\pi(V_n) \sim c_b) =$$

$$\Pr(v_1|\pi(V_1) \sim c_x) \cdot \Pr(v_2|\pi(V_2) \sim c_x) \cdot \ldots \cdot \Pr(\overline{v_1}|\pi(V_i) \sim c_x) \cdot \ldots \cdot \Pr(v_n|\pi(V_n) \sim c_x) = \Pr(c_x)$$

□

To illustrate Theorem 1, neighbour multiplier λ_j^i is calculated for the Bayesian network depicted in Figure 1. Arbitrarily, configuration $(abc\overline{d})$ is picked as our goal configuration. The base configuration will be $(abcd)$, which is a neighbour of $(abc\overline{d})$ in variable D.

Calculating the probability of $(abcd)$ can be done by the factorisation (See 3) and results in $\Pr(abcd) = 0.3584$

$$\begin{aligned} \Pr(abcd) = \Pr(a) \cdot \Pr(b|a) \cdot \Pr(c|a) \cdot \Pr(d|bc) = \\ 0.8 \cdot 0.7 \cdot 0.8 \cdot 0.8 = 0.3584 \end{aligned} \tag{3}$$

Now, λ is calculated as per Theorem 1. Note that $c_x = abc\overline{d}$

$$\lambda = \frac{\Pr(\overline{d}|\pi(D) \sim c_v)}{\Pr(d|\pi(D) \sim c_v)}$$

Since variable D has no children, no extra probabilities are needed. The next step is filling in the values for variables in the set $\pi(D)$ according to c_v.

$$\lambda = \frac{\Pr(\overline{d}|bc)}{\Pr(d|bc)} = \frac{0.2}{0.8} = 0.25$$

Using λ, it is possible to calculate the value of $\Pr(abc\overline{d})$:

$$\Pr(abc\overline{d}) = \Pr(abcd) \cdot \lambda = 0.3584 \cdot 0.25 = 0.0896$$

For completeness, $\Pr(abc\overline{d})$ is calculated using the factorisation.

$$\begin{aligned} \Pr(abc\overline{d}) = \Pr(a) \cdot \Pr(b|a) \cdot \Pr(c|a) \cdot \Pr(\overline{d}|bc) = \\ 0.8 \cdot 0.7 \cdot 0.8 \cdot 0.2 = 0.0896 \end{aligned}$$

As expected, the value calculated using the multiplier matches the value calculated in the traditional way.

C. The fundamental configuration

Thus far, a multiplier has been defined, and proven to be correct, for changing the probability of one configuration on the variables $\boldsymbol{V}$ to another, where the configurations are neighbours in one variable. In this section, the multipliers are generalised such that we can establish each probability of configurations from a single base configuration. This is done such that each multiplier has the same basis for comparison, and in the base-neighbour representation probabilities of configurations can be ordered by ordering the neighbour multipliers instead.

Definition 8. fundamental configuration The fundamental configuration is defined as a configuration used as base for all the goal configurations.

We begin by explaining how multipliers can be found which, combined with the probability of the fundamental configuration, can generate the probability of all other configurations. Then, a suitable fundamental configuration is chosen after which an example is given for our example Bayesian network for clarification purposes.

1) Combining the multipliers: It is not possible to pick a fundamental configuration such that each configuration is a neighbour of the fundamental configuration. This is due to the fact that not every configuration differs in only one variable from the some given configuration. Thanks to the associative property of multiplication however, two multipliers can be multiplied to create a new multiplier.

Take three configurations c_v, c_w and c_z of all variables in $\boldsymbol{V}$. Assume that the configurations c_v and c_w are neighbours in a variable V. c_z is also a neighbour of c_v, but in variable W. Now both the probabilities of c_z and c_v are expressed using a multiplier and the probability on a neighbour as base.

$$Pr(c_v) = Pr(c_w) \cdot \lambda_1$$

$$Pr(c_z) = Pr(c_v) \cdot \lambda_2$$

Substituting $Pr(c_v)$ returns

$$Pr(c_z) = Pr(c_w) \cdot \lambda_1 \cdot \lambda_2$$

TABLE II
NEIGHBOUR ROUTES FROM FUNDAMENTAL CONFIGURATION $(abcd)$ TO A GOAL CONFIGURATION USED IN OUR EXAMPLE

Goal configuration	Neighbour route
$(abc\bar{d})$	$(abcd) \rightarrow (abc\bar{d})$
$(ab\bar{c}d)$	$(abcd) \rightarrow (ab\bar{c}d)$
$(ab\bar{c}\bar{d})$	$(abcd) \rightarrow (abc\bar{d}) \rightarrow (ab\bar{c}\bar{d})$
$(a\bar{b}cd)$	$(abcd) \rightarrow (a\bar{b}cd)$
$(a\bar{b}c\bar{d})$	$(abcd) \rightarrow (abc\bar{d}) \rightarrow (a\bar{b}c\bar{d})$
$(a\bar{b}\bar{c}d)$	$(abcd) \rightarrow (ab\bar{c}d) \rightarrow (a\bar{b}\bar{c}d)$
$(a\bar{b}\bar{c}\bar{d})$	$(abcd) \rightarrow (abc\bar{d}) \rightarrow (ab\bar{c}\bar{d}) \rightarrow (a\bar{b}\bar{c}\bar{d})$
$(\bar{a}bcd)$	$(abcd) \rightarrow (\bar{a}bcd)$
$(\bar{a}bc\bar{d})$	$(abcd) \rightarrow (abc\bar{d}) \rightarrow (\bar{a}bc\bar{d})$
$(\bar{a}b\bar{c}d)$	$(abcd) \rightarrow (ab\bar{c}d) \rightarrow (\bar{a}b\bar{c}d)$
$(\bar{a}b\bar{c}\bar{d})$	$(abcd) \rightarrow (abc\bar{d}) \rightarrow (ab\bar{c}\bar{d}) \rightarrow (\bar{a}b\bar{c}\bar{d})$
$(\bar{a}\bar{b}cd)$	$(abcd) \rightarrow (a\bar{b}cd) \rightarrow (\bar{a}\bar{b}cd)$
$(\bar{a}\bar{b}c\bar{d})$	$(abcd) \rightarrow (abc\bar{d}) \rightarrow (a\bar{b}c\bar{d}) \rightarrow (\bar{a}\bar{b}c\bar{d})$
$(\bar{a}\bar{b}\bar{c}d)$	$(abcd) \rightarrow (ab\bar{c}d) \rightarrow (a\bar{b}\bar{c}d) \rightarrow (\bar{a}\bar{b}\bar{c}d)$
$(\bar{a}\bar{b}\bar{c}\bar{d})$	$(abcd) \rightarrow (abc\bar{d}) \rightarrow (ab\bar{c}\bar{d}) \rightarrow (a\bar{b}\bar{c}\bar{d}) \rightarrow (\bar{a}\bar{b}\bar{c}\bar{d})$

By combining multipliers, it now is possible to express the probability of c_z as a multiplier and the probability of base configuration c_w, even though c_w and c_z are not neighbours in a variable. The *fundamental multiplier* for c_z is now defined as $\lambda_{c_z} = \lambda_1 \cdot \lambda_2$.

Definition 9. Fundamental multiplier The multiplier between the probability of the fundamental configuration and the probability of a goal configuration is called the fundamental multiplier.

2) A suitable fundamental: Theoretically the fundamental multipliers can take on any positive real value, since probabilities take on values between 0 and 1 and the multipliers are fractions of products of probabilities. To avoid having to deal with large numbers, we pick a fundamental configuration such that the values of the fundamental multipliers are restricted. More specifically, we take the fundamental configuration to be the current c_{MPE}. Proposition 2 states that the fundamental multipliers are indeed restricted.

Proposition 2. When the MPE configuration is picked as the fundamental configuration, the fundamental multipliers lie between 0 and 1.

Proof. $Pr(c) = Pr(c_{MPE}) \cdot \lambda \Rightarrow \lambda = \frac{Pr(c)}{Pr(c_{MPE})}$. Since $Pr(c_{MPE}) \geq Pr(c)$, the fraction $\frac{Pr(c)}{Pr(c_{MPE})} \leq 1$, thus $\lambda \leq 1$. Since $Pr(c_{MPE}) \neq 0$, division by 0 can not happen.

Since probability values always lie between 0 and 1, the multipliers cannot be negative. □

Now that a suitable fundamental has been picked, it is possible to construct the fundamental multipliers. In order to do so, we first define the concept of neighbour route in Definition 10. In Proposition 3 is stated how the fundamental multipliers can be found.

Definition 10. Neighbour route A neighbour route NR_c for a configuration c on variables $\boldsymbol{V}$ is defined as a sequence of configurations. The sequence has length k, where k is the number of variables having a different value in c and c_{MPE}. The first element of the sequence is the fundamental configuration, and the last element of the sequence is c. Furthermore, for each two consecutive configurations nr_i and nr_{i+1} in NR_c it holds that $nr_{i+1} \in N_{nr_i}^{V_j}$ for a variable $V_j \in V$.

Proposition 3. A fundamental multiplier λ_c for a goal configuration c can be found as follows: $\prod_i \lambda_{nr_i}$ where λ_{nr_i} is the neighbour multiplier between element nr_i and nr_{i+1} in a neighbour route NR_c of c.

Proof. For a goal configuration $c_i \neq c_{MPE}$, there is a multiplier λ_{nr_k} such that $Pr(c_{k-1}) * \lambda_{nr_k} = Pr(c_k)$, where $c_{k-1} \in N_{c_k}^V$ for a variable V. Note that k is the amount of variables differing in value between the goal configuration and the fundamental configuration. Then, for $\Pr(c_{k-1})$ a similar reasoning holds. This reasoning can be repeated until the fundamental c_{MPE} is reached. Combining these results leads to $Pr(c_k) = Pr(c_{k-1}) \cdot \lambda_{nr_k} = Pr(c_{k-2}) * \lambda_{nr_k} \cdot \lambda_{nr_{k-1}} = \ldots = Pr(c_{MPE}) \cdot \lambda_{nr_k} \cdot \lambda_{nr_{k-1}} .. \cdot \lambda_{nr_1}$ which proves our proposition. □

We show an example to illustrate Proposition 3. We will calculate the fundamental multipliers for the configurations in the Bayesian network shown in Figure 1. The neighbour routes we use in the example are depicted in Table II. The neighbour routes are based on the topological ordering we defined earlier, this way an unique neighbour route exists between every two configuration.

Using the neighbour routes, we first calculate the necessary neighbour multipliers. The results are depicted in Table III.

TABLE III
MULTIPLIERS BETWEEN THE PROBABILITY OF A BASE CONFIGURATION AND THAT OF A GOAL CONFIGURATION. NOTE THAT HORIZONTAL BARS INDICATE A DIFFERENT VARIABLE GETTING A NEW VALUE.

base	multiplier	fraction	value	goal
$Pr(abcd)$	λ_1^1	$\frac{Pr(\overline{d}\|bc)}{Pr(d\|bc)}$	0.25	$Pr(abc\overline{d})$
$Pr(abcd)$	λ_1^2	$\frac{Pr(\overline{c}\|a)\cdot Pr(d\|b\overline{c})}{Pr(c\|a)\cdot Pr(d\|bc)}$	0.1875	$Pr(ab\overline{c}d)$
$Pr(abc\overline{d})$	λ_2^2	$\frac{Pr(\overline{c}\|a)\cdot Pr(\overline{d}\|b\overline{c})}{Pr(c\|a)\cdot Pr(\overline{d}\|bc)}$	0.5	$Pr(ab\overline{c}\overline{d})$
$Pr(abcd)$	λ_1^3	$\frac{Pr(\overline{b}\|a)\cdot Pr(d\|\overline{b}c)}{Pr(b\|a)\cdot Pr(d\|bc)}$	0.214	$Pr(a\overline{b}cd)$
$Pr(abc\overline{d})$	λ_2^3	$\frac{Pr(\overline{b}\|a)\cdot Pr(\overline{d}\|\overline{b}c)}{Pr(b\|a)\cdot Pr(\overline{d}\|bc)}$	1.285	$Pr(a\overline{b}c\overline{d})$
$Pr(ab\overline{c}d)$	λ_3^3	$\frac{Pr(\overline{b}\|a)\cdot Pr(d\|\overline{b}\overline{c})}{Pr(b\|a)\cdot Pr(d\|b\overline{c})}$	0.214	$Pr(a\overline{b}\overline{c}d)$
$Pr(ab\overline{c}\overline{d})$	λ_4^3	$\frac{Pr(\overline{b}\|a)\cdot Pr(\overline{d}\|\overline{b}\overline{c})}{Pr(b\|a)\cdot Pr(\overline{d}\|b\overline{c})}$	0.75	$Pr(a\overline{b}\overline{c}\overline{d})$
$Pr(abcd)$	λ_1^4	$\frac{Pr(\overline{a})\cdot Pr(b\|\overline{a})\cdot Pr(c\|\overline{a})}{Pr(a)\cdot Pr(b\|a)\cdot Pr(c\|a)}$	0.0536	$Pr(\overline{a}bcd)$
$Pr(abc\overline{d})$	λ_2^4	$\frac{Pr(\overline{a})\cdot Pr(b\|\overline{a})\cdot Pr(c\|\overline{a})}{Pr(a)\cdot Pr(b\|a)\cdot Pr(c\|a)}$	0.0536	$Pr(\overline{a}bc\overline{d})$
$Pr(ab\overline{c}d)$	λ_3^4	$\frac{Pr(\overline{a})\cdot Pr(b\|\overline{a})\cdot Pr(\overline{c}\|\overline{a})}{Pr(a)\cdot Pr(b\|a)\cdot Pr(\overline{c}\|a)}$	0.143	$Pr(\overline{a}b\overline{c}d)$
$Pr(ab\overline{c}\overline{d})$	λ_4^4	$\frac{Pr(\overline{a})\cdot Pr(b\|\overline{a})\cdot Pr(\overline{c}\|\overline{a})}{Pr(a)\cdot Pr(b\|a)\cdot Pr(\overline{c}\|a)}$	0.143	$Pr(\overline{a}b\overline{c}\overline{d})$
$Pr(a\overline{b}cd)$	λ_5^4	$\frac{Pr(\overline{a})\cdot Pr(\overline{b}\|\overline{a})\cdot Pr(c\|\overline{a})}{Pr(a)\cdot Pr(\overline{b}\|a)\cdot Pr(c\|a)}$	0.5	$Pr(\overline{a}\overline{b}cd)$
$Pr(a\overline{b}c\overline{d})$	λ_6^4	$\frac{Pr(\overline{a})\cdot Pr(\overline{b}\|\overline{a})\cdot Pr(c\|\overline{a})}{Pr(a)\cdot Pr(\overline{b}\|a)\cdot Pr(c\|a)}$	0.5	$Pr(\overline{a}\overline{b}c\overline{d})$
$Pr(a\overline{b}\overline{c}d)$	λ_7^4	$\frac{Pr(\overline{a})\cdot Pr(\overline{b}\|\overline{a})\cdot Pr(\overline{c}\|\overline{a})}{Pr(a)\cdot Pr(\overline{b}\|a)\cdot Pr(\overline{c}\|a)}$	1.333	$Pr(\overline{a}\overline{b}\overline{c}d)$
$Pr(a\overline{b}\overline{c}\overline{d})$	λ_8^4	$\frac{Pr(\overline{a})\cdot Pr(\overline{b}\|\overline{a})\cdot Pr(\overline{c}\|\overline{a})}{Pr(a)\cdot Pr(\overline{b}\|a)\cdot Pr(\overline{c}\|a)}$	1.333	$Pr(\overline{a}\overline{b}\overline{c}\overline{d})$

Next, the neighbour routes as described above are used to find the fundamental multipliers. These are contained in the set Λ. An example is worked out in Table IV. As expected, all values in Λ are between 0 and 1.

D. *Example of calculating a probability using the fundamental*

In this section, a fundamental multiplier is used to calculate a probability. First, a goal configuration is picked. For this example, configuration $(a\overline{b}\overline{c}d)$ is picked. The reader is reminded that the fundamental configuration is $c_{MPE} = (abcd)$. For configuration $(a\overline{b}\overline{c}d)$, multiplier $\lambda_6 \in \Lambda$ is needed. As calculated in Table IV, this is 0.09375. c_{MPE} has a probability of 0.3584 as calculated in the previous example. Now: $Pr(a\overline{b}\overline{c}d) = \lambda_6 \cdot Pr(c_{MPE}) = 0.0402 \cdot 0.3584 = 0.0144$

Calculating $Pr(a\overline{b}\overline{c}d)$ by using the factorisation of the Bayesian network, yields the following: $Pr(a\overline{b}\overline{c}d) = Pr(a) \cdot Pr(\overline{b}|a) \cdot Pr(\overline{c}|a) \cdot Pr(d|\overline{b}\overline{c}) = 0.8 \cdot 0.3 \cdot 0.2 \cdot 0.3 = 0.0144$. This indeed matches the value found using the fundamental.

E. *Evidence in the new representation*

When evidence is present in the network, the values of the multipliers do not change. The neighbour multiplier and fundamental multipliers are a property of the Bayesian network. However, not all possible configurations over the variables in the network agree with the given evidence. To handle evidence in the new representation, we introduce the notion of legality in configurations and multipliers.

Definition 11. Legal configuration A configuration is called legal if the values of the variables do not conflict with the given evidence. A configuration is called illegal if the values do conflict with the given evidence.

Definition 12. Legal multiplier A multiplier is called legal if applying the multiplier to the probability of the corresponding base configuration does not yield the probability of an illegal configuration. An illegal multiplier is a multiplier which results in the probability of an illegal configuration.

For example, assume that in the example Bayesian network evidence $A = a$ is observed. In c_{MPE} it holds that $A = a$. The goal configurations of multipliers of the fourth order (λ_j^4), see Table III, all contain the value $A = \overline{a}$. Therefore the fourth order multipliers are all illegal, and henceforth the fundamental multipliers $\lambda_8 - \lambda_1 5$ containing multipliers of the fourth order are illegal as well.

Now that the solution space of the MPE problem can be described using the new representation, we use this new representation to look into how change in evidence affects the MPE solution.

TABLE IV
MULTIPLIERS FROM THE PROBABILITY OF THE FUNDAMENTAL CONFIGURATION $(abcd)$ AND RESULTING CONFIGURATIONS.

Multipliers	name	value	goal configuration
1	λ_0	1	$Pr(abcd)$
λ_1^1	λ_1	0.25	$Pr(abc\overline{d})$
λ_1^2	λ_2	0.1875	$Pr(ab\overline{c}d)$
$\lambda_1^1 \cdot \lambda_2^2$	λ_3	0.175	$Pr(ab\overline{c}\overline{d})$
λ_1^3	λ_4	0.214	$Pr(a\overline{b}cd)$
$\lambda_1^1 \cdot \lambda_2^3$	λ_5	0.322	$Pr(a\overline{b}c\overline{d})$
$\lambda_1^2 \cdot \lambda_3^3$	λ_6	0.040125	$Pr(a\overline{b}\overline{c}d)$
$\lambda_1^1 \cdot \lambda_2^2 \cdot \lambda_4^3$	λ_7	0.0938	$Pr(a\overline{b}\overline{c}\overline{d})$
λ_1^4	λ_8	0.0536	$Pr(\overline{a}bcd)$
$\lambda_1^1 \cdot \lambda_2^4$	λ_9	0.01345	$Pr(\overline{a}bc\overline{d})$
$\lambda_1^2 \cdot \lambda_3^4$	λ_{10}	0.02681	$Pr(\overline{a}b\overline{c}d)$
$\lambda_1^1 \cdot \lambda_2^2 \cdot \lambda_4^4$	λ_{11}	0.0179	$Pr(\overline{a}b\overline{c}\overline{d})$
$\lambda_5^4 \cdot \lambda_1^3$	λ_{12}	0.107	$Pr(\overline{a}\overline{b}cd)$
$\lambda_1^1 \cdot \lambda_2^3 \cdot \lambda_6^4$	λ_{13}	0.160625	$Pr(\overline{a}\overline{b}c\overline{d})$
$\lambda_1^2 \cdot \lambda_3^3 \cdot \lambda_7^4$	λ_{14}	0.0535	$Pr(\overline{a}\overline{b}\overline{c}d)$
$\lambda_1^1 \cdot \lambda_2^2 \cdot \lambda_4^3 \cdot \lambda_8^4$	λ_{15}	0.175	$Pr(\overline{a}\overline{b}\overline{c}\overline{d})$

IV. EXCLUDING POTENTIAL MPE CANDIDATES WHEN CHANGING EVIDENCE

Consider a Bayesian network for which evidence e is given and the MPE solution, c_{MPE} is known. In this section we use the multipliers to research what happens to the MPE solution, when one variable in the evidence changes value.

We show that when one variable in the evidence set changes value in a Bayesian network, some configurations are excluded from becoming the new MPE solution.

Once evidence is changed, the legality of configurations changes. First we show that the set of configurations legal under the changed evidence can be linked to the set of configurations legal under the original evidence by using multipliers. Secondly, we show that the ordering of the probabilities of the configurations under original evidence is partly preserved when changing evidence.

Assume we are given a Bayesian network B where for a set of variables E evidence e has been observed. Furthermore, the MPE solution given evidence e, c_{MPE}, is known. Assume that the evidence for a variable $V_c \in E$ is changed. In the proposition, the set $M_{V_c}^+$ containing both the variables in the Markov blanket of V_c and V_c itself is used.

Proposition 4. In Bayesian network B with observed evidence e, where variable $V_c \in E$ changes value, and the current MPE solution is known, it is possible to identify $\frac{2^{|\boldsymbol{V}|-|E|}}{2^{|M_{V_c}|}} - 1$ legal configurations that can not possibly be the new MPE solution.

To prove the proposition we will first prove two lemmas. From the first lemma we conclude that it is indeed possible to exclude some legal configurations from being the new MPE solution. The second lemma shows that we can identify $\frac{2^{|\boldsymbol{V}|-|E|}}{2^{|M_{V_c}|}} - 1$ of those legal configurations. We first establish some properties necessary to prove the first lemma.

When evidence in variable V_c changes, the legality of some of the fundamental multipliers changes. Assuming the value of variable V_c in the evidence was v_c, the fundamental multipliers containing neighbour multipliers in V_c were illegal (since they change the value of a variable to a value which does not agree with the original evidence). These now become precisely the legal fundamental multipliers.

We propose a way to relate the probabilities of configurations legal under original evidence to the probabilities of configurations legal under changed evidence. See Proposition 5.

Proposition 5. Consider a configuration c_{old} of the variables $\boldsymbol{V}$ legal under evidence e. Assume that in evidence e variable V_c changes value. The probability of a configuration $c_{new} \in N_{c_{old}}^{V_c}$ legal under the changed evidence can now be found by multiplying the fundamental multiplier of c_{old} by the neighbour multiplier in V_c between c_{old} and c_{new}.

Proof. Let λ_{old} be the fundamental multiplier of c_{old}. Thus: $\Pr(c_{old}) = \Pr(c_{MPE}) \cdot \lambda_{old}$. Furthermore, we have that $\Pr(c_{new}) = \Pr(c_{old}) \cdot \lambda_{n+1}^{m+1}$, where λ_{n+1}^{m+1} is the neighbour multiplier in V_c between c_{old} and c_{new}. Then, $\Pr(c_{new}) = \Pr(c_{MPE}) \cdot \lambda_{old} \cdot \lambda_{n+1}^{m+1} = \Pr(c_{MPE}) \cdot \lambda_{new}$ where λ_{new} is the fundamental multiplier of c_{new}. $\square$

We now show that between probabilities of certain configurations the neighbour multipliers in V_c sometimes have the same value. The configurations legal under the original evidence are split into groups, such that all the neighbour multipliers in V_c of configurations in this group have the same value. The groups are defined, after which we prove that indeed the neighbour multipliers of V_c in this group have the same value. Moreover, the resulting probabilities are probabilities of configurations legal under the new evidence.

Definition 13. Configuration siblings in m_V Consider a set $m_V^+ \in M_V^+$ and two configurations on all variables (V) c_x and c_y. c_x and c_y are called siblings in m_V^+ if the following holds: $m_V^+ \sim c_x \Leftrightarrow m_V^+ \sim c_y$. The set of all configurations which are siblings in m_V^+ is called the sibling set of m_V^+, denoted by $S_{m_V^+}$. The collection of all different sibling sets of M_V^+ is called the sibling set of M_V^+.

Lemma 1. *For each configuration $m_{V_c}^+$ of variables in $M_{V_c}^+$, all the neighbour multipliers in V_c for configurations in the sibling set $S_{m_{V_c}^+}$ have the same value.*

Proof. The variables in the conditional probabilities used to calculate neighbour multipliers in V_c all are an element of $M_{V_c}^+$ as per the definition of the multiplier. Since the configurations agree on values of variables in $M_{V_c}^+$, the conditional probabilities used will be the same, and therefore the multipliers' values will be the same. □

For an example of Lemma 1, we again consider Table III. For the sibling set $S_{m_A^+}$ where $m_A^+ = (abc)$, which includes configurations $abcd$ and $abc\overline{d}$, the neighbour multipliers in A, λ_1^4 and λ_2^4 have the same value.

Under the original evidence, the probability of configurations yielded an ordering. Lemma 1 proved that in sibling sets on m_V^+, neighbour multipliers in V have the same values. Multiplying two ordered probabilities by the same value does not change the ordering. This leads to Corollary 1.

Corollary 1. For any configuration $m_{V_c}^+$ on variables in $M_{V_c}^+$ and any configuration $c_{old} \in S_{m_{V_c}^+}$ on the variables in V where $\Pr(c_{old}) \geq \Pr(c_i)$, for any $c_i \in S_{m_{V_c}^+}$, $c_i \neq c_{old}$. Let $c_{new} \in N_{c_{old}}^{V_c}$. Then, $\Pr(c_{new}) \geq \Pr(c_j)$ where $c_j \in N_{c_i}^{V_c}$.

As an example we again take the sibling set of $m_A^+ = (abc)$. The evidence $e = (a)$ is observed. We have that $S_{m_A^+} = \{(abcd), (abc\overline{d})\}$. In this sibling set, we have that $\Pr(abcd) > \Pr(abc\overline{d})$ since $(abcd)$ is the MPE solution. We now assume that in the evidence, instead of a the value $\overline{a}$ is observed. Using Corollary 1, we can now immediately see that $\Pr(\overline{a}bcd) > \Pr(\overline{a}bc\overline{d})$, since $(\overline{a}bcd) \in N_{abcd}^A$ and $(\overline{a}bc\overline{d}) \in N_{abc\overline{d}}^A$.

Left to prove is the number of candidates which can be excluded based on knowing the original MPE solution. As shown above, most likely configurations in a sibling set can be a candidate for the new MPE solution. Therefore, we only need to consider on configuration per sibling set for the MPE solution. However, we know for sure only one most likely configuration in a sibling set: the MPE solution. The other most likely configurations are unknown. Left to show is thus that the size of a sibling set is $\frac{2^{|\boldsymbol{V}|-|E|}}{2^{|M_{V_c}|}}$. We state this formally in a lemma.

Lemma 2. *The size of a sibling set of $m_{V_c}^+$ for any configuration $m_{V_c}^+$ on the variables in $M_{V_c}^+$ is equal to* $\frac{2^{|\boldsymbol{V}|-|E|}}{2^{|M_{V_c}|}}$.

Proof. The total amount of configurations possible on a Bayesian network where each variable can take on two values is $2^{|V|}$. Since the size of the evidence set remains constant, the configurations not agreeing with the evidence need be removed: $2^{|V|-|E|}$. There are $2^{|M_{V_c}|}$ configurations on M_{V_c}, and keeping these constant as well yields a total of $2^{|V|-|E|-|M_{V_c}|}$ configurations. □

Knowing the size of a sibling set, we now can deduce that the number of configurations which can be excluded from being the new MPE solution is equal to the size of the sibling set minus one, since the MPE solution itself is still a candidate. The proof of Proposition 4 now follows from this together with Corollary 1. We continue with an example on the example Bayesian network.

A. Example of using the sibling set

To illustrate Proposition 4 an example is given. We assume the evidence $e = a$ was previously observed. Now, the observed value in variable A changes. The configurations legal under evidence $e = a$ are divided in sibling sets in $S_{M_A^+}$. In Figure 2 these are sorted by probability. The shaded probabilities have the highest value in each sibling set.

From Corollary 1 it now follows that when evidence in A is changed, only the neighbours in A of the configurations with the highest probability (shaded in the figure) are eligible to be the new MPE configuration. In Figure 3 the neighbours configurations in A of configurations in the sibling sets are depicted. Note that these are sibling sets of their own. In each group, the configuration of the shaded probability is a neighbour of the configuration of the shaded probability in Figure 2. As expected, these are again the most likely probability. Moreover, knowing the MPE configuration $(abcd)$ we can immediately see that $(\overline{a}bc\overline{d})$ is not a possible MPE candidate. We see that indeed $\frac{2^{|V|-|E|}}{2^{|M_{V_c}|}} = \frac{2^2}{2^2} = 1$ configurations can be removed from consideration.

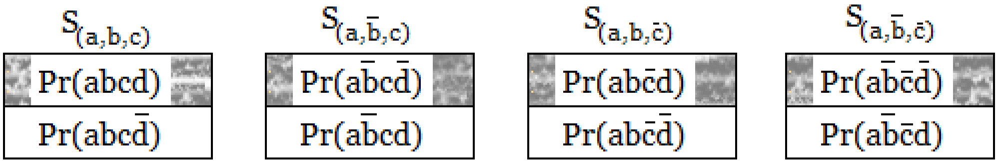

Fig. 2. The probabilities of legal configurations in the sibling sets in $S_{M_A^+}$ for evidence $e = a$

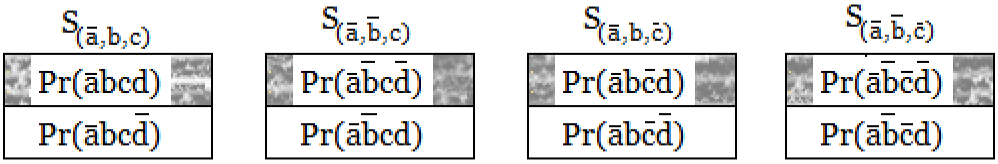

Fig. 3. The probabilities of legal configurations under evidence $e = \overline{a}$, which are neighbours of the configurations in the sibling sets in $S_{M_A^+}$.

B. Adding a variable to the evidence set

When instead of changing the value of a variable a variable is added to the evidence set, a modified version of Proposition 4 still holds. Indeed, it is still possible to select configurations which are eligible to be the new MPE configuration. The proof of Proposition 6 is similar to the proof of Proposition 4.

Proposition 6. For a Bayesian network B for which the current MPE solution is known, it is possible to identify $\frac{2^{|V|-|E|}}{2^{|M_{V_c}|}} - 1$ legal configurations that can not possibly be the new MPE solution when new evidence in variable V_c is observed.

Proposition 6 and 4 are based on the fact that only the MPE solution is known. However, knowing the k-th MPE, the top k most likely configurations, would lead to potentially more legal configurations being excluded from becoming the MPE solution. If k is larger than the size of one sibling set, more legal configurations can be excluded. This is due to the fact that all the legal configuration being excluded by knowing the actual MPE solution could already be known in the k-th MPE solution. For example, assume the the $3-th$ MPE solution is known in Figure 2. Then, at least two most likely configurations of different sibling sets are known, and two legal configurations instead of one legal configurations can be excluded.

V. Analysing the effect of parameter changes using the new representation

In addition to studying the solution space of the MPE problem when evidence has changed, we also use the new representation to study the MPE problem when a parameter in one of the CPTs of the Bayesian network changes value. The notion of sibling sets will again provide valuable insights in the solution space.

Previous work by Van der Gaag and Renooij researched methods on finding parameters which under parameter change lead to a change in the MPE solution [9]. Regarding MPE solutions, Chan and Darwiche studied the amount a parameter value can vary before the MPE solution is changed. They have created an algorithm which calculates this variation for all parameters in $O(n\ e^w)$ time, where n is the number of network variables and w is its treewidth [2]. This algorithm uses the fact that some parts of the robustness calculations are independent of the parameter value. We focus on the question what happens to the solution space of the MPE problem if a parameter value is changed.

The neighbour multipliers are made up of parameters in the Bayesian network. Changing a parameter will thus also change the neighbour multipliers this parameter is part of, and by doing so will also change the fundamental multipliers. In the previous section it was shown that under certain circumstances the neighbour multipliers between configurations have the same value. Similarly, we show in this section that under parameter change, groups of fundamental multipliers increase (or decrease) by the same factor. The ordering on the probabilities of the goal configurations of these fundamental multipliers is thus kept intact. Again, some configurations can then be excluded from being the new MPE solution.

We first show that when a parameter of the form $\Pr(V_c|\Pi(V_c))$ changes value, the probabilities of configurations in the sibling set $S_{m_{V_c}^+} \in S_{M_{V_c}^+}$ change by the same factor. This is formally stated in Proposition 7, after which we show an example before giving the proof.

Proposition 7. Consider a node V_c and a sibling set $S_{m_{V_c}^+}$ of a configuration of $M_{V_c}^+$. When a parameter of the form $\Pr(V_c|\Pi(V_c))$ changes value, the ordering on the probabilities of the configurations in the sibling set $S_{m_{V_c}^+}$ remains constant.

Assume that in our example Bayesian network parameter $\Pr(a)$ changes value from $\Pr(a) = 0.8$ to $\Pr(a) = 0.9$. In this example, we only work out the sibling set $S_{(abc)} = \{(abcd), (abc\overline{d})\}$. Calculating the probabilities of these configurations when $\Pr(a) = 0.8$ yields the following results: $\Pr(abcd) = 0.3584$ and $\Pr(abc\overline{d}) = 0.0896$. This means that $(abcd)$ has a higher probability than $(abc\overline{d})$ and thus is ordered higher.

Now we calculate the probabilities of the same configurations but the parameter value has changed: $\Pr(a) = 0.9$. The following results are obtained: $\Pr(abcd) = 0.4032$ and $\Pr(abc\overline{d}) = 0.1008$. As evident, configuration $(abcd)$ still has a higher probability than configuration $(abc\overline{d})$.

Proof. We aim to prove Proposition 7 by showing that whenever a parameter changes value, fundamental multipliers of configurations in a sibling set either remain at their current value, or all change with the same factor. In order to do so, the fundamental multipliers are broken down in the neighbour multiplier representation. It is then shown, that the subset of the neighbour multipliers containing parameters of the form $\Pr(V_c|\Pi(V_c))$ are invariant in the different fundamental multipliers in a sibling set in $S_{M^+_{V_c}}$.

First, we notice that only neighbour multipliers in V_c and neighbour multipliers in W where $W \in \Pi(V_c)$ can contain parameters of the form $\Pr(V_c|\Pi(V_c))$. This can be seen from the definition of the neighbour multiplier, see Equation 2.

Furthermore, for each fundamental multiplier λ_i it holds that a neighbour multiplier λ_i^k in V_c is one of the factors of λ_i if $V_c \sim m^+_{V_c}$, where k is the order of the multiplier. Since all the configurations in the sibling set agree on value in $m^+_{V_c}$, and thus $V_c \sim m^+_{V_c}$ for either all configurations $m^+_{V_c}$ or none, λ_m^k is either present in all fundamental multipliers of configurations in $S_{m^+_{V_c}}$ or absent in all of them.

Thus far, we have established that parameters of the form $\Pr(V_c|\pi(V_c))$ only appear in neighbour multipliers in V_c and in neighbour multipliers in W where $W \in \Pi(V_c)$ for all V_c. Moreover, a multiplier in V_c or W either is a factor of all fundamental multipliers of configurations in $S_{m^+_{V_c}}$ or none. Left to prove is that each neighbour multiplier in V_c or W is influenced in the same way by the parameter changes.

Since the values of variables in $m^+_{V_c}$ are the same in each configuration in sibling set $S_{m^+_{V_c}}$, the parameters used in the calculation of a neighbour in variable V_c all are the same for each configuration. Therefore, the neighbour multipliers in V_c have the same value for each configuration in the sibling set $S_{m^+_{V_c}}$.

Moreover, a neighbour multiplier in W, $W \in \Pi(V_c)$ can be refactored in two parts. One part P_1 containing parameters of the form $\Pr(V_c|\Pi(V_c))$, and one part P_2 containing the parameters which do not have the form $\Pr(V_c|\Pi(V_c))$. Since the variables in $M^+_{V_c}$ have the same value in each configuration in $S_{m^+_{V_c}}$, P_1 has the same value for each neighbour multiplier in W. Moreover, P_2 does not contain a parameter of the form $\Pr(V_c|\Pi(V_c))$ and is thus constant when changing the value of a parameter of the form $\Pr(V_c|\Pi(V_c))$. Therefore, when changing the value of parameter of the form $\Pr(V_c|\Pi(V_c))$ P_2 will remain constant for each probability of a configuration, and P_1 has the same value in each probability of a configuration. Therefore the ordering of the neighbour multipliers does not change. □

Note that Proposition 7 does not explicitly mention co-varying parameters. However, as the co-varying parameters of a parameter of the from $\Pr(V_c|\Pi(V_c))$ are also of the form $\Pr(V_c|\Pi(V_c))$, Proposition 7 can also be applied to the co-varying parameters. This means the effect of the co-varying parameters can studied more effectively.

Since the ordering of the configurations in a sibling set $S_{m^+_V} \in S_{M^+_V}$ is constant when a parameter of the form $\Pr(V|\Pi(V))$ changes, the MPE solution is restricted to a few configurations. In particular, these are the most probable configurations of the sibling sets in $S_{M^+_V}$. This is formally stated in a corollary.

Corollary 2. When the value of a parameter of the form $\Pr(V_c|\Pi(V_c))$ is changed, the MPE solution is restricted to a set of configurations $\{c_1, \dots, c_n\}$ where c_i is the most likely configuration in a sibling set $S_{m^+_{V_c}} \in S_{M^+_{V_c}}$ and $n = |S_{M^+_{V_c}}|$. Moreover, the MPE solution is equal to $\max_{S_{m^+_{V_c}} \in S_{M^+_{V_c}}} (\max_{c_i \in S_{m^+_{V_c}}} (\Pr(c_i)))$.

Corollary 2 shows that when a parameter changes, within each sibling set the most likely configuration does not change. However, between the sibling sets the most likely configuration can change. In our working Bayesian network example, Corollary 2 implies that the MPE solution is equal to:

$\max(\Pr(abcd), \Pr(a\overline{b}c\overline{d}), \Pr(ab\overline{c}d), \Pr(a\overline{b}\overline{c}\overline{d}), \Pr(\overline{a}bcd), \Pr(\overline{a}\overline{b}c\overline{d}), \Pr(\overline{a}b\overline{c}\overline{d}), \Pr(\overline{a}\overline{b}\overline{c}\overline{d}))$

as these are the most likely probabilities in the sibling sets, see Figures 2 and 3.

We have shown that when a parameter of the form $\Pr(V_c|\Pi(V_c))$ changes value, the MPE solution can be restricted to the most probable values in the sibling sets on $M^+_{V_c}$. However, the set of possible values can be restricted even more.

Without loss of generality assume that a parameter goes up in value. When a parameter goes up in value, co-varying parameters go down in value. Due to the laws of probability theory, the probabilities that constitute a total probability distribution need to sum to one. Therefore, when changing one parameter at least one other parameter needs to co-vary. For the case of binary valued variables, equation 4 is always true.

$$\Pr(v|\pi(V)) + \Pr(\overline{v}|\pi(V)) = 1 \quad (4)$$

Therefore, for the most likely configurations in sibling sets then, three different possibilities for the probabilities of these configurations are possible when changing the value of a parameter:

- The probability of the configuration goes up in value
- The probability of the configuration goes down in value
- The probability of the configurations stays the same value

The sibling sets can be divided in three different groups representing these different outcomes on the most likely probability. The configurations within one such group all change with the same factor, and the ordering thus never changes. See Proposition 8 which describes what these different groups are, and that indeed the probabilities of configurations in these groups change by the same factor.

Proposition 8. Consider a parameter $P = \Pr(v_c|\pi(V_c))$ and the set of sibling sets $S^{v_c} = \{S_{m^+_{V_c}} | m^+_{V_c} \sim (v_c \pi(V_c))\}$. If P goes up in value, the probabilities of configurations in S^{v_c} go up in value. The probabilities of configuration in the set of sibling sets $S^{\overline{v_c}} = \{S_{m^+_{V_c}} | m^+_{V_c} \sim (\overline{v_c}, \pi(V_c))\}$ go down in value. The configurations not belonging to either S^{v_c} or $S^{\overline{v_c}}$ belong in the third group, S, where the probabilities of configuration in S do not change. Moreover, the MPE solution is the configuration with the highest probability in one of the three groups.

As an example, assume that in our example Bayesian network the parameter $\Pr(a) = 0.8$ is changed to $\Pr(a) = 0.9$. Then, the sibling sets in figure 2 are in the set S^a, and the sibling sets in figure 3 are in the set $S^{\overline{a}}$. The third set is empty in this case because there are no configurations left.

A proposition analogous to Proposition 8 is also stated in the work by Chan and Darwiche [2]. For the proof of Proposition 8 we therefore refer to their work. Moreover, in their paper a way is shown how to find the maximum change a parameter will allow before the MPE solution changes. This maximum change can be found for each parameter in $O(n\ e^w)$ time, where n is the number of network variables and w is its treewidth.

Whereas we used the base-neighbour representation of the configurations in a Bayesian network, Chan and Darwiche used the factorisation of the Bayesian network itself. They noted that the factorisation of a joint probability of a configuration can be split in a part containing the changed parameter and a constant, which leads to the same result as Proposition 8.

By using the multipliers and sibling sets, the problem of finding the new best MPE in the entire solution space is broken down into finding multiple MPEs in smaller solution spaces. By doing so, the individual elements of the problem can be analysed independent of each other, leading to more precise insights in the solution space.

Where Chan and Darwiche also divide the problem in smaller problems, we feel that the division they make is a very crude division. All the co-varying parameters are considered in one group, whereas using the base-neighbour representation each co-varying parameter has its own group. This means that the effect of the co-varying parameter can be studied more effectively.

VI. Conclusion and discussion

For Bayesian networks, the MPE problem is the problem of finding a configuration of the variables such that given the evidence the posterior probability is maximal. However, when an MPE solution is known and the evidence changes, it is possible the MPE solution changes as well. Moreover, a similar statement can be made when a parameter in the set of parameters changes value. This paper was motivated by the aim to gain a better understanding of the solution space where the new MPE solution can be found. The results might lead to a faster algorithm to calculate the MPE solution in the future when evidence or a parameter value changes. In this paper we have shown that not every configuration in the solution space needs to be considered for the new MPE solution.

Initially, the problem was approached by an experimental approach. By comparing the solution spaces of the problem on examples networks, we hoped to find a link between cliques and the solution space. When this approach failed, we redefined the problem and used a theoretical approach using the decomposition of the probability function.

The problem now is approached by first defining a new representation of the probabilities of configurations of the Bayesian network. The representation consists of a global fundamental base configuration, and a fundamental multiplier for each configuration. These fundamental multipliers are constructed by multiplying neighbour multipliers, which can be calculated by using the conditional probabilities found in the CPTs of the Bayesian network.

Once evidence changes in one variable, the legality of configurations changes as well. Dividing the configurations in groups called sibling sets, the ordering between the new legal configurations can be related to the ordering of the old legal configurations. We showed that within a sibling set, only the configuration which has the highest probability is eligible to be an MPE solution. Therefore, all the other configurations in the sibling set need not be considered for the new MPE solution. Moreover, the sibling sets can be used when a new variable enters the evidence set as well.

The solution space under influence of parameter change yields a similar result. Not all possible configurations need be considered for the new MPE solution. Only the most likely probabilities in each sibling set are eligible for the new MPE solution. The amount of configurations eligible to be the new MPE solution can be further reduced to three by creating three groups based on the parameter changed, and then again picking the most likely configurations. The latter result was also found by Chan and Darwiche [2].

Whereas we approached the problem by breaking the factor between the probabilities of two different configurations down in smaller factors, Chan and Darwiche approached the problem by focusing on the difference between two configurations as a whole. Although the final results are the same, we believe using the sibling sets yields more insight in the MPE problem. By breaking the problem down in smaller problems, individual elements can be analysed rather than the problem as a whole.

The paper " *Preprocessing the MAP problem*" by Bolt and Van der Gaag discusses a way to preprocess the Maximum A-posteriori Probability (MAP) problem [1]. The MAP problem is a variant of the MPE problem. Whereas in the MPE problem the variables are either evidence variables or variables used in the explanation, in the MAP problem there is a third group of variables. This group of variables is neither evidence nor a explanation variable but a variable of no interest. In the work by Bolt and van der Gaag, the concept of the Markov blanket is used to ease the calculations for finding a MAP solution.

Since the Markov blanket can be used in preprocessing the MAP problem, and the sibling sets are based on the Markov blankets as well, there is an indication sibling sets can be useful for the MAP problem as well. Again, this would enable us to analyse individual aspects of the problem rather than the problem as a whole.

The new representation opens up more possibilities for future research. The sibling sets can be used when evidence changes in only one variable. The question now remains if equivalent sets can be found when more than one variable in the evidence changes value. If such sets exists, would it also be possible to exclude configurations from being the new MPE solution? A similar question can be asked for the parameter change: if multiple parameters are changed at once, can the sibling sets again help to exclude configurations?

Furthermore, we have shown that knowing the MPE solution allows for configurations of only one sibling set being removed. The question left open is whether there are methods which allow configurations of other sibling sets to be prevented from consideration for being an MPE solution as well.

REFERENCES

[1] Janneke H Bolt and Linda C Van Der Gaag. Preprocessing the map problem. In *Probabilistic Graphical Models*, pages 51–58, 2006.

[2] Hei Chan and Adnan Darwiche. On the robustness of most probable explanations. In *In Proceedings of the Twenty Second Conference on Uncertainty in Artificial Intelligence*, 2012.

[3] Michael L Littman. Initial experiments in stochastic satisfiability. In *Proceedings of the sixteenth national conference on Artificial intelligence and the eleventh Innovative applications of artificial intelligence conference innovative applications of artificial intelligence*, pages 667–672. American Association for Artificial Intelligence, 1999.

[4] James D Park and Adnan Darwiche. Complexity results and approximation strategies for map explanations. *Journal of Artificial Intelligence Research*, pages 101–133, 2006.

[5] Arnoud Pastink, Linda van der Gaag, et al. The persistence of most probable explanations in bayesian networks. *ECAI 2014*, 263:693–698, 2014.

[6] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Representation and Reasoning Series. Morgan Kaufmann Publishers, 1988.

[7] Glenn R Shafer and Prakash P Shenoy. Probability propagation. *Annals of Mathematics and Artificial Intelligence*, 2(1-4):327–351, 1990.

[8] Bon K Sy. Reasoning mpe to multiply connected belief networks using message passing. In *Proceedings of the tenth national conference on Artificial intelligence*, pages 570–576, 1992.

[9] Linda C van der Gaag and Silja Renooij. Analysing sensitivity data from probabilistic networks. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pages 530–537. Morgan Kaufmann Publishers Inc., 2001.