UTRECHT UNIVERSITY

MASTER THESIS

Analysis of the behaviour of the MPE in a Bayesian network that is subject to changes

---

Author:
Bob REIMINK
ICA-5814588

Supervisor:
prof. dr. ir. L. C. VAN DER GAAG
Second examiner:
dr. ir. D. THIERENS

June 29, 2018

Abstract

An important computational problem in Bayesian networks is to find the most likely set of states of all unobserved variables in the network given the states of the observed variables, the evidence. This problem is known as the *MPE* problem. Especially for applications with real-time requirements, re-computations of the MPE after changes in the evidence will not always be manageable. In this thesis, we will investigate how the MPE changes as a result of a change in the evidence. The research is performed on basis of a junction tree that can be constructed from a Bayesian network. We present theoretical results about how changes in the evidence affect the probabilities in the junction tree and introduce a way to visualize this. We further on carry out several experiments on well-known Bayesian networks. In these experiments, the consequences of a single change in the evidence are studied. We investigate how the change is propagated through the junction tree and how the MPE variables are affected. We extend these experiments by considering the more general problem - the MAP problem. The results of these experiments indicate that changes caused by a change in evidence decreases quickly as we propagate it through the junction tree. Furthermore, in general only a small number of MPE and MAP variables change of state and these variables are close to the variable that changed in the evidence.

Contents

List of Figures

List of Tables

1 Introduction

A Bayesian network is a model to reason about probabilities of variables in a specific domain. It consists of a directed graph in which the nodes represent the variables. This graph represents independencies between the variables. Take for example a network with the variables *grass*, *rain* and *cloud type*. The grass is either wet or dry. The probability that the grass is wet may depend on the type of clouds. It will be more likely to rain when it is cloudy than partial cloudy. And if there are no clouds, the probability of rain is zero. If it rains however, we know that the grass is wet and therefore, it does not matter whether we know it is cloudy or not. This can be represented in a Bayesian network by having an arc from *cloud type* to *rain* and from *rain* to *grass*. *cloud type* is called the parent of *rain* and likewise *rain* the parent of *grass*. While this is one kind of independency, other kinds can be represented in the Bayesian network.
Besides the graphical structure, a Bayesian network consists of a conditional probability table for each variable. The table contains the probability distribution of the states the variable can be in. When the variable has parents, a probability distribution is stated for each possible combination of the states of the parents.

A Bayesian network may include variables of which the state is known, *observed* variables. There will however be variables of which the state is unobserved. In order to reason about the states of these variables, we have to take other variables in the Bayesian network into account. Suppose we want to know whether the *grass* is wet and we do not know anything about *rain* and *cloud type*. In order to compute the probability, we can determine the probability in each situation: partially cloudy and raining, no clouds and dry etc. Adding these probabilities up gives us the probability over all possible combinations. Suppose there are three states for *cloud type* and two states for *rain*, then the number of combinations is $2 \times 3 = 6$. Suppose now that we extend the Bayesian network by adding a new variable *season*. This variable becomes the parent of *cloud type* and has four states. If we again want to know the probability of whether the grass is wet, we have to include the new variable as well. Because there are four seasons, this results in $6 \times 4 = 24$ combinations. As we can see, the number of combinations growths exponentially with the number of unobserved variables. We can however also exploit the structure of the Bayesian network and look first at the probabilities of each state of *cloud type* given its parent, then look at the probabilities of rain or dry based on *cloud type* and finally look at the probability of wet grass given the probabilities of *rain*. This gives us: $3 \times 4 + 2 \times 3 + 1 \times 2 = 22$ combinations. The number of combinations is a bit smaller but grows linearly instead of exponentially with the number of variables and will therefore be much smaller with a larger number of variables.

Computing a probability is called inference. We could infer the probability of each state the variable can be in. This gives us information about what the most likely state of the variable is. While computing the most likely state of one variable may be interesting, it can also be the case that we are interested in the most likely states of a combination of variables. This problem is called the *Maximum a posteriori estimation* (MAP). The problem is defined as: Given a set of variables, which set of states for these variables will give us the highest probability. We cannot simply look at the most likely state of each variable separately and combine them together. It is even the case that this problem is NP-hard [2]. A more specific instance of this problem is the Most Probable Explanation (MPE) problem. Here, we look at the set of all variables that are unobserved. Although this problem is NP-hard in general as well [2], it is easier to compute under specific conditions. We will see that there exist algorithms to compute the MPE quite fast under these conditions - polynomial time.

While these algorithms are sufficient when our network is small or we do not have any real-time requirements, it will become a problem when this is not the case. For example in monitoring applications, where we have a Bayesian network in which a set of variables is monitored. We get an update of these variables on a regular basis. When no variable changes, we do not have to do anything. It could however occur that in two consecutive updates, the state of a monitored variable changes. We have to recompute the MPE in both updates which gives us a limited amount of time for the first re-computation. This may not be manageable when, for example, the Bayesian network is big, or the update interval is too small.

1.1 Research Questions

In this thesis, our goal is to analyze how the Most Probable Explanation of a set variables in a Bayesian network changes when information in the Bayesian network changes and use this to get insight in how the re-computations of the Most Probable Explanation can be reduced. While the main focus of the analysis is the case where the set of variables of interest consists of all unknown variables, the MPE problem, we will also address the more general MAP problem. We define two types of changes that can occur in a Bayesian network. The first is an evidence change. With evidence, we mean the states of the observed variables. A change occurs when the state of one of the variables changes, an unobserved variable becomes observed or an observed variable becomes unobserved. The second type of change occurs when a probability in a conditional probability table changes. In this thesis, we only consider the first type of change.

During our analysis we will consider three research questions. The first research question that will be addressed is:

- How does the MPE change when evidence in the Bayesian network changes?

This question will be addressed by considering how the complete distribution of probabilities of possible explanations of the Bayesian network are affected. When a change in the Bayesian network occurs, the MPE along with the other explanations may change. One of these other explanations may become the new MPE. It is therefore important to consider the complete distribution and not only the MPE in order to gain more insight in how the MPE changes. Furthermore, we will not directly use the Bayesian network to investigate the MPE. Instead, a junction tree will be used in the analysis. We will later on see what a junction tree exactly is. A lot of current algorithms translate the Bayesian network into a junction tree before performing inference on probabilities [3, 4, 5]. A junction tree has sets of variables as nodes and contains potentials instead of conditional probability tables. This structure makes it easier to reason about properties and can also be used to compute the MPE. We will first investigate how the potentials can be used to look at changes in the distribution. After this, we can look how changes within potentials can be related to changes in the MPE.

The second research question that will be addressed is:

- How does the MAP change when probabilities in the Bayesian network change?

In the MAP problem, we want to find the configuration of a subset of unobserved variables that has the highest joint probability. The unobserved variables that are not part of this subset have to be marginalized out. This makes the MAP problem more difficult than the MPE problem. However, the approach used for analyzing the MPE may lend itself for the MAP problem as well. Therefore, we will first investigate whether we can analyze the MAP with the same method and if this is the case, we can look how the changes in potentials of the junction tree can be related to changes in the MAP.

The final research questions is:

- How can we use the relations between the changes in the MPE and the changes in the potentials in order to reduce computations?

When relations have been found, it is interesting to investigate how they can be used. Important here is how to deal with consecutive changes as in the monitoring explanation. We should not only look at the reduction of computations after one change, but also how this can be combined with the computations of the next change.

1.2 Organization

The thesis is organized as follows: In section 2, the Bayesian network, junction tree and a way to infer probabilities are introduced. In section 3, we will define the MPE and MAP problem and review previous research that have been performed around the MPE and MAP problem. Section 4 contains theoretical results about processing changes in the junction tree and a visualization of processing the changes is described in section 5. In section 6, theoretical results concerning persistence of MPE variables is discussed. We will discuss experiments concerning the MPE in section 7 and experiments about the MAP in section 8. We will end with a conclusion in section 9.

2 Preliminaries

In this section, we will first review the concept of a Bayesian network and describe some of its properties. After that, we will address how a Bayesian network can be converted to another representation called a junction tree. This representation allows more efficient computation of probabilities. We will also show how the potentials in the tree are updated when evidence changes. Throughout this section, we will use an example Bayesian network. The graphical structure of this network is shown in Figure 1.

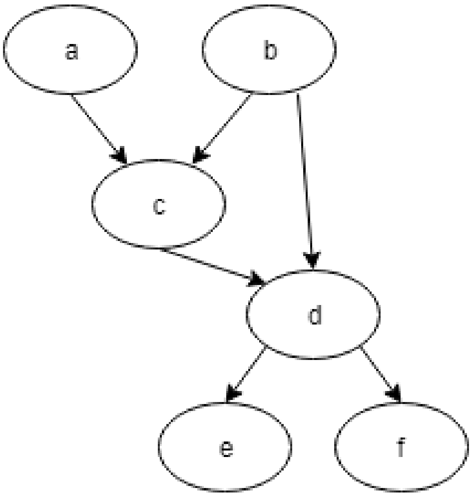
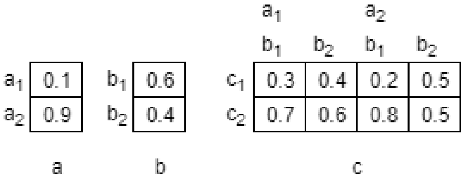


Figure 1: Graphical structure of the example Bayesian network

2.1 Bayesian network

Bayesian networks are used to reason about probabilities. A Bayesian network is defined to include a graphical structure composed of a set of nodes representing variables and a set of arcs between these nodes (See Figure 1). The arcs indicate relations between the variables and are not allowed to form a directed cycle within the graph. We will use an uppercase letter to indicate a variable set, and a lowercase letter to indicate a variable. We can use, for example, $V = \{a, b, c, d, e, f\}$ as the variable set of our Bayesian network. Each variable has a state space which represents the states in which it can be. The state space for variable b will for example be indicated with Ω_b. A state of a variable will be indicated with the same letter as the variable but with a subscript indicating the state. For example, the state space of variable b consists of the states: $\Omega_b = \{b_1, b_2\}$. The state space of a variable should have a size of at least two. We call a variable v a parent of w if there is an arc pointing from v to w. The set of parents of a variable v is indicated with $Pa(v)$. In our example network, we have that $Pa(c) = \{a, b\}$.

In addition to the graphical structure, a Bayesian network contains for each variable a conditional probability table, or CPT for short. Probabilities, or parameters, are defined in the CPT of a variable for each of its states. The probabilities form the probability distribution of the variable. When a variable has parents, a conditional probability distribution is defined given each possible combination of states of its parent variables. Example CPTs for the variables a, b and c are shown in Figure 2.



a_1	0.1
a_2	0.9

a

b_1	0.6
b_2	0.4

b

	a_1		a_2	
	b_1	b_2	b_1	b_2
c_1	0.3	0.4	0.2	0.5
c_2	0.7	0.6	0.8	0.5

c

Figure 2: Conditional probability tables of variables a, b and c. Each variable has two states.

Variable c has parents and has therefore a conditional probability distribution for each combination of states of the parents. Since the variable can only be in one of the states in its state space, all probabilities for a variable with the same parent configuration together sum up to 1. We will call a combination of states of variables a configuration and indicate it with $\mathbf{c}$. A configuration of the parents of c is indicated with $\mathbf{c}_{Pa(c)}$ or $\mathbf{c}_{\{a,b\}}$ and a configuration of a variable set V is indicated with $\mathbf{c}_V$. We will use an additional type of notation for subset configurations. Suppose we have the variable sets V and W where $W \subseteq V$ and a configuration $\mathbf{c}_V$. The configuration of W in which the same states are chosen as in $\mathbf{c}_V$ is indicated with $\mathbf{c}_W^{\downarrow V}$. We will call the set of all possible configurations of V the configuration space of V and indicate it with Ω_V. It can be constructed by taking the Cartesian product of the state spaces of its variables: $\times_{v \in V} \Omega_v$. Note that the same symbol is used for the state space, they can be distinguished by their subscript. If the subscript is a single variable, it indicates a state space, otherwise a configuration space.
For some variables in a Bayesian network, the state is known. We call these variables *observed* variables. The set of observed variables is called the evidence set and will be indicated with E. The configuration of this set - the evidence - is indicated with $\mathbf{c}_E$.

The graphical structure of a Bayesian network represents how variables are independent of each other and not necessarily how they are dependent. In the graphical structure, a path between two variables is defined as a set of incident arcs that connect the variables. The direction of the arcs does not matter. We speak of a blocked path when either a variable in E lies on the path and one of the two adjacent arcs on the path is an outgoing arc or when there is a variable on the path of which both adjacent arcs on the path are incoming arcs and this variable along with its descendants are not in E. A variable v is a descendant of w when there is a *directed* path from w to v. When all paths between two variable sets V and W in the graph are blocked by a third variable set E, we say that E d-separates V and W. This means that the variables of V are independent of W given that all variables in the variable set E are observed. In our example network, we have that d d-separates a of e. This is the case because all paths between the two variables go through d and on both paths, the two arcs adjacent to d are not both incoming arcs. Due to this d-separation, we can say, for example, that the probabilities of the states of e will not change when there is evidence introduced for variable a in the network while d is observed. On the other hand, variable c does not d-separate variable a and b because of the two incoming arcs for c. So, when variable c and its descendants are not observed, variable a and b are independent of each other. But when evidence for one of these variables is found, the path will be unblocked.
An important concept in d-separation is the **Markov blanket** of a variable, which is defined as the set of variables that d-separate it from the rest of the variables in the Bayesian network.

Now that we can read from a given Bayesian network how variables are independent of each other, we can use this information for factorizing the joint probability distribution. The joint probability distribution of a variable set V is indicated with Pr and is a function that, for a given configuration of the variables in V, returns the probability of this configuration. While the function can be used to compute the probability of a configuration of all variables in the Bayesian network, we can also use it to compute the joint probability of a configuration of a smaller set of variables. In order to do this, we have to marginalize out the other variables. In case of our example Bayesian network, if we want to compute the joint probability distribution of b and c, we have to marginalize out a, d, e and f. This comes down to:

$$\Pr(\mathbf{c}_{\{b,c\}}) = \sum_{\mathbf{c}_{\{a,d,e,f\}} \in \Omega_{\{a,d,e,f\}}} \Pr(\mathbf{c}_{\{b,c\}}, \mathbf{c}_{\{a,d,e,f\}})$$

Suppose each variable has a state space of size 2, then the number of configurations over which we need to sum probabilities is equal to $2^4 = 16$. While this is doable, imagine you want to know the joint probability distribution of two variables and need to marginalize out ten other variables with state spaces of 2. This requires us to sum 2^{10} probabilities. As we can see, this will quickly become infeasible. Fortunately, the independencies represented by the structure of the Bayesian network can be used to factorize the joint probability distribution. The joint probability distribution can be written as the product of the conditional probability tables:

$$\Pr(\mathbf{c}_V) = \prod_{v \in V} \Pr(\mathbf{c}_v | \mathbf{c}_{Pa(v)})$$

Due to this property, we can marginalize locally and take the product of the separate tasks to get our final marginalized probability.

2.2 Computing probabilities

The method for actually computing probabilities from a Bayesian network used in this thesis, is based on a junction-tree representation of the network. The junction tree was introduced by *Lauritzen* and *Spiegelhalter* [6]. We will first describe what a junction tree is and how it can be constructed from a Bayesian network. Next, a method called propagation is introduced comes. This method comes down to sending and processing messages between the nodes in the junction tree. We will explain this method and see how this can be used again when the Bayesian network changes after, for example, a change in the evidence.

2.2.1 From a Bayesian network to a junction tree

A junction tree is an undirected tree representation of a Bayesian network in which a node represents a set of variables in the Bayesian network. A node representing variable set X will be indicated with Cl_X. The junction tree does not contain cycles and has the *Running Intersection Property* (RIP): For each node Cl_X, Cl_Y and Cl_Z where Cl_Y lies between Cl_X and Cl_Z, if X and Z contain the variable v, then Y contains v as well.

In order to convert a Bayesian network to a junction tree, a couple of steps have to be taken. First, we will make the Bayesian network undirected. Important here is that we want to preserve as much independency relations as possible while we do not want to introduce new independencies that did not exist in the original Bayesian network. In undirected graphs, separation comes down to: If each path between two nodes contains a node with evidence, then all paths are blocked and the two nodes are independent. Separation thus differs from d-separation. In Bayesian networks, there are nodes with multiple incoming arcs. For example, if we have a Bayesian network with node a, b and c and a has two incoming arcs of b and c, then b is independent of c if there is no evidence for a. Preserving this independency in the undirected representation would mean that we have to remove one of the arcs.

This however introduces a new independency: b and c will also be independent when there is evidence for a. Keeping both arcs will result in a new independency as well: a and b are independent when a is observed. Therefore, an edge needs to be added between b and c. Removing the directions of the arcs will then give us no new independencies but we will lose information about the independency of b and c. This process is called *moralization*. For each two unconnected parents, an edge will be added between them. Afterwards, all arcs will be made undirected resulting in an undirected graph. All independency relations that can be read from this graph were also present in the Bayesian network.
Next, we want to get rid of the cycles by using specific node sets in the graph as nodes in the junction tree. However, in order to make sure that the joint probability distribution factorizes over the junction tree, we need to deal with cycles having more than three nodes in which two non-consecutive nodes do not have an edge in between them. By adding edges between such two nodes until these cycles do not exist anymore, called *triangulation*, we get a triangulated network. Skipping this step would result in a junction tree that violates the RIP. Take for example a cycle of four nodes. There are four cliques with each two nodes. If we enforce a tree structure, there will be two cliques that share a variable while they are not connected in the junction tree. Their shared node will cause the RIP to be violated. We would like to keep the cliques as small as possible when triangulating. Unfortunately, optimal triangulation is proved to be NP-hard [7].
Finally, the cliques in the triangulated graph can be converted to nodes. Connecting these nodes while making sure that the resulting graph is a tree and has the running intersection property can be done in polynomial time which gives us the junction tree. In Figure 3, a junction tree of the example Bayesian network can be seen. Note that a junction tree is not unique. In this example, the node containing variables d and e could be connected to the node containing d and f instead of the node containing b, c and d. Throughout the thesis, we will the word *clique* to refer to a node in the junction tree.

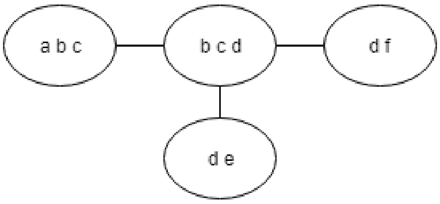


Figure 3: A junction tree of the example Bayesian network

2.2.2 Propagation

Propagation in a junction tree is performed by passing messages between cliques. Multiple algorithms have been developed for propagation. The complexity of the algorithms involving the junction tree are the same and depend on the maximum clique size in the network. A comparison between them is described by *Lepar* and *Shenoy* [8].
In this thesis, we will use the propagation algorithm introduced by *Jensen et al.* [3]. In this algorithm, the junction tree is extended with so-called *separators*. We will indicate a separator with variable set K with Se_K. For each edge in the junction tree, a separator is added. Given two adjacent cliques Cl_V and Cl_W, the separator Se_K on this edge contains the variable set K where $K = V \cap W$. The variable set of a separator will always contain at least one variable. An example is shown in Figure 4.
For each clique and separator, a potential is defined. The potential of a clique Cl_R or separator Se_R will be indicated with ϕ_R. A potential ϕ_R is a function that, given a configuration $\mathbf{c}_R \in \Omega_R$, returns a non-negative number. As we will see, a potential is actually the marginal probability distribution of

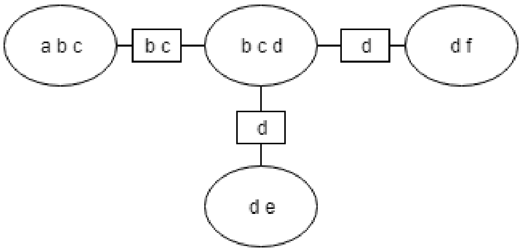


Figure 4: A junction tree for the example Bayesian network with separators

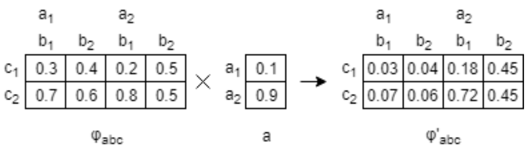
the variables in the corresponding clique or separator. These potentials are computed from the CPTs of the original Bayesian network and will be used to propagate evidence and compute probabilities correctly within the junction tree.
The joint probability distribution of the Bayesian network can also be expressed in terms of the clique and separator potentials in the junction tree. Suppose a junction tree with variable set V, then the joint probability distribution of V is:

$$\Pr(\mathbf{c}_V) = \frac{\prod_{Cl_S \in \mathbf{Cl}} \phi_S(\mathbf{c}_S^{\downarrow V})}{\prod_{Se_L \in \mathbf{Se}} \phi_L(\mathbf{c}_L^{\downarrow V})}$$

Here, **Cl** and **Se** represent the set of cliques and separators respectively. The potentials actually represent the marginal probability distribution of the variables in those potentials. If we would replace the potentials by Pr, the joint probability distribution can be rewritten to the joint probability distribution as the Bayesian network stated it. More information is stated in the paper of *Lauritzen et al.* [6]. Because each potential contains correct probabilities for its variables, we can take for a variable v an arbitrary clique or separator of which the variable set contains v to compute its probability. This can be achieved by marginalizing out the other variables in the separator or clique.
In order to compute the correct values in the potentials, we first initialize each potential to give as output 1 for each given configuration of its variables. Next, we go over each variable x in the Bayesian network and do the following: Pick a clique Cl_X such that $v \in X$ and $Pa(v) \subseteq X$. Suppose W presents the variables in the CPT of v and f returns, given a configuration, the probability in the CPT, then the processing of the CPT into Cl_X goes as follows:

$$\phi_X(\mathbf{c}_X) = \phi_X(\mathbf{c}_X) f(\mathbf{c}_W^{\downarrow X}) \ \ \forall \mathbf{c}_X \in \Omega_X$$

An example is shown in Figure 5.



	a_1		a_2	
	b_1	b_2	b_1	b_2
c_1	0.3	0.4	0.2	0.5
c_2	0.7	0.6	0.8	0.5

×

a_1	0.1
a_2	0.9

→

	a_1		a_2	
	b_1	b_2	b_1	b_2
c_1	0.03	0.04	0.18	0.45
c_2	0.07	0.06	0.72	0.45

Figure 5: Processing CPT of variable a into the potential with variable set $\{a, b, c\}$. The CPT of variable c has already been processed into the potential

After all CPTs have been processed into the tree, we want to make sure that the values in the potentials of both cliques and separators are the actual marginal probabilities of the corresponding configurations. Since variables can be present in multiple cliques, these marginal probabilities ensure that computing the probability for a variable in one clique will match the computations for the same variable in another clique. In order to realize this property, propagation is performed in two phases: *Collection* and *Distribution* phase.

Collection
In the *Collection* phase, the idea is to choose a clique, say Cl_R, and send messages from all cliques toward Cl_R. A clique can only send messages to the adjacent separators. Separators in turn can only send messages to their adjacent cliques. The message from a clique to a separator contains a potential that will function as the new potential for the separator. Suppose clique Cl_S and adjacent separator Se_L. The message from Cl_S to Se_L is defined as follows:

$$\phi_L(\mathbf{c}_L) = \sum_{\mathbf{c}_R \in \Omega'_R} \phi_R(\mathbf{c}_R) \; where \; \Omega'_R = \{\mathbf{c}_R | \mathbf{c}_R \in \Omega_R \wedge \mathbf{c}_L^{\downarrow R} = \mathbf{c}_L\} \quad \forall \mathbf{c}_L \in \Omega_L$$

What we do here is marginalizing out all variables that are present in the clique but not in the separator and use the resulting sums as values for the new potential in the separator. The separator then sends a message to the other adjacent clique. This message is the division of the values of its new potential by the values of the previous potential. The clique updates its potential by multiplying its current values by the values in the message. Suppose a separator Se_L and Cl_S, we compute the new potential of Cl_S as follows:

$$\phi_S(\mathbf{c}_S) = \phi_S(\mathbf{c}_S) \frac{\phi'_L(\mathbf{c}_L^{\downarrow S})}{\phi_L(\mathbf{c}_L^{\downarrow S})} \quad \forall \mathbf{c}_S \in \Omega_S$$

Here, ϕ' indicates the new potential. Note that we initialized the potentials to output 1. Therefore, the message from a separator in the initial *Collection* phase comes down to the values of the new potential.
In order to reduce the computations as much as possible, we start at the cliques that are leafs in the junction tree. Subsequently, we let a clique compute its message for the separator that lies between the clique and Cl_R only if it has received and processed the messages of its other adjacent separators. In this way, we send exactly one message over each edge in the junction tree.

Distribution
In the *Distribution* phase, a second round of message passing starts. Cl_R starts by sending messages to its separators. The separators send messages to their other adjacent clique. These cliques will send in turn messages to all their separators except for the separator from which the message came. This process continues until we have reached the leaves of the junction tree. The generating and processing of messages is the same as in the *Collection* phase. When this phase ends, messages have been sent across each edge in the junction tree in both ways.

Finally, we have to normalize the values in each potential to let them sum up to 1. After this normalization, the potentials show the correct marginal probabilities for the configurations. If we would send again messages across the junction tree, there will be no changes because the values in the new and old potentials of each separator will be the same resulting in a division by the value 1 and thus no change in the next clique. When there is new or changed evidence, we can select the clique that contains the variables for which new or adjusted evidence is found and change the probabilities. After normalizing the cliques, we can execute the two phases again to obtain the new probabilities for the potentials.

Because the values in a potential of a clique or separator represent probabilities of the configurations, we will use the word **probability** to refer to a value in a potential. The values sent in a message from a separator is a division of old and new probabilities. Since these values are only used to be multiplied with the probabilities in a clique to adjust them correctly, we will indicate a value in this message with **factor** in the remaining part of the thesis.

3 MPE

In this section, we review the MPE problem. The problem will first be formally defined after which we look into the complexity of the problem, and compare it with the MAP problem and its complexity. Subsequently, an algorithm for computing the MPE problem is described. A related problem called the k-MPE will be discussed next. Finally, we will review some research as to how the computation of the MPE can be reduced.

3.1 Definition

From a Bayesian network, a *Maximum a posteriori* (MAP) configuration can be extracted. The term *posterior* indicates that evidence is taken into account when computing the probabilities of configurations of the variables. Given a Bayesian network with the variable set V, an evidence set $E \subseteq V$ with configuration $\mathbf{c}_E$ and a variable set $U \subseteq V \backslash E$, the MAP of U given $\mathbf{c}_E$ is:

$$\arg\max_{\mathbf{c}_U \in \Omega_U} \Pr(\mathbf{c}_U | \mathbf{c}_E) = \arg\max_{\mathbf{c}_U \in \Omega_U} \sum_{\mathbf{c}_V \in \Omega_{V \backslash (U \cup E)}} \Pr(\mathbf{c}_U, \mathbf{c}_V | \mathbf{c}_E)$$

For computing a *Maximum a posteriori* configuration, it will not suffice to just look at each variable separately, pick the state that has the highest posterior probability and combine them into a configuration $\mathbf{c}_U$ in order to get the MAP. This would only work when the variables are independent of each other which is most of the time not the case.

In this thesis, the focus will mainly be on the case that $U = V \backslash E$. This is called the *Most Probable Explanation* (MPE).

3.2 Complexity

In this section, the difficulty of both problems will be described. The input size that is considered for both problems depends on the number of variables and the size of the conditional probability tables of these variables.

In order to solve the problems, we could for the MPE just go over every configuration and keep the configuration with the highest probability, while for the MAP problem, we could first marginalize out every variable that is not part of U and then go over the remaining configurations and keep the configuration with the highest probability. Unfortunately, the number of configurations growths exponentially with the number of variables present in the Bayesian network. This makes the computations quickly infeasible as the number of variables increases. For the general case, no polynomial algorithms have been found for the problems. However, while both the MAP and MPE problem are hard to solve, the MAP problem is more difficult than the MPE problem.

In order to explain this difference, we need to introduce some complexity classes. First of all the **NP** complexity class. This class contains the set of decision problems of which the solution can be verified in polynomial time. A decision problem is a problem of which the outcome is a *yes* or a *no*. The complexity class contains both problems for which we know a polynomial time algorithm as well as problems for which only exponential time algorithms are known. The class of **NP**-complete problems is a subclass of **NP** and contains the problems whose solutions suffice for solving any other problem in **NP** in polynomial time. **PP** is another complexity class and contains the decision variants of the problems in yet another class called the **#P** class. While **NP** is concerned about whether there is a solution to the given problem, the **#P** class contains counting problems. These problems are concerned about the number of solutions.

Bayesian networks are concerned with probabilities and in order to get some intuition in how probabilities are related with counting, let us consider the following decision problem: *Can we find more than p solutions that satisfy a specific problem?* Suppose we have a total of n solutions and each solution is equally likely to pick, then we can reformulate the problem as follows: *Is the probability that a solution suffices higher than* $\frac{p}{n}$*?*
We will not go into further detail but in the end, *Littman et al.* proved that the problem of determining whether the joint probability of a configuration of a set variables is greater than a given threshold probability belongs to the **PP** class [9].

Now, in order to use these classes to explain the difference in difficulty between MAP and MPE, we will look at the decision variants of these problems: Given the evidence and a set of unobserved variables - for the MPE case all variables except the evidence variables - is there an assignment to these variables that has a probability greater than p for a given p that lies between 0 and 1? For the MPE, this problem is **NP**-complete [2]. For MAP, *Park* and *Darwiche* proved that it is $\mathbf{NP^{PP}}$-complete [10]. It can be intuitively seen that the MAP is in this complexity class by dividing it into two problems: Find the best assignment to the variables in the MAP set like the MPE, but find it on marginalizing over the variables that are neither in the evidence set nor in the MAP set like the problem described by *Littman et al.* The complexity classes for the MPE and MAP problem are for the general case. Research in complexity for Bayesian networks with a bounded induced width has been performed as well. If we look at junction trees, then the clique containing the most variables can be seen as an indicator of the induced width of the Bayesian network. The complexity of the MPE problem is then polynomially solvable when the induced width is bounded [11]. While the complexity of the MAP problem reduces as well, the problem is still **NP-complete** [10]; this bound also applies to Bayesian networks of which the graph is a singly connected graph since they have a bounded induced width by definition.

3.3 Computation of the MPE

In order to compute the MPE, the junction-tree representation explained in section 2.2 can be used. *Dawid* describes an algorithm, called max-propagation, that uses the junction tree [4]. The propagation algorithm is adjusted by replacing marginalization with maximization. When a clique Cl_V passes a message to the separator Se_L, we consider for each configuration $\mathbf{c}_L \in \Omega_L$ the set: $\{\mathbf{c}_V | \mathbf{c}_V \in \Omega_V \wedge \mathbf{c}_L^{\downarrow V} = \mathbf{c}_L\}$. Instead of performing a summation of the probabilities of these configurations, we only take the maximum. From the separator's perspective, we get from each partition the probability of the best configuration. The other adjacent clique gets the best option for each partition and multiplies it with the values in its own potential. After the *Collection* and *Distribution* phase, the highest probability in each potential has the same value and represents the probability of the MPE. We can easily look up the configuration by joining the configurations for which the potential outputs the highest value. If multiple configurations have the highest value, we have multiple MPEs. Updating the tree, for example after new evidence, follows the same procedure as standard propagation but again with maximization instead of marginalization.

3.4 k-MPE

A related problem to the MPE problem is the k-MPE problem. In this problem, we try to find the k most probable configurations where k is a positive integer. Having multiple probable configurations could be interesting. You have alternative configurations that you can take into account as well. It also

shows how probable the best configuration is relative to other configurations. If more configurations have a probability that is almost equal to the probability of the MPE, the MPE is not that probable compared to the other configurations.
Solving the k-MPE problem is obviously harder than the MPE problem. However, *Nilsson* came up with an algorithm that is quite efficient [12]. The idea is as follows: Suppose we have a Bayesian network with the variable set V with $|V| = n$. We start by computing the MPE according to the max-propagation algorithm introduced in the previous section. Suppose that the MPE is $\mathbf{c}_V$. The second-best configuration should then be different in at least one place. In order to enforce this, we can partition the configuration space of the complete variable set into n parts. The configurations in the first part are not allowed to have the same state as in $\mathbf{c}_V$ for the first variable. The state of the other variables can be the same or different. The second set contains the configurations in which the first variable is the same as in $\mathbf{c}_V$ while the second should be different and the rest does not matter. We proceed in the same way for the other sets and end with the set in which the configurations should only be different from $\mathbf{c}_V$ in the last variable while the other variables stay the same. These sets partition the configuration space $V \backslash \{\mathbf{c}_V\}$. The restrictions for each of the sets can be filled in as evidence. By introducing the n evidences separately in the junction tree and performing max-propagation after each one, we get an MPE of each partition. The MPE configuration with the maximum value is the second MPE. For our third MPE, we have to partition the set that contained the second MPE. The idea is here the same as for the second MPE. This will give us n new sets on which we have to perform max-propagation. This will be repeated k times. Summing up all max-propagations that have been performed, we come to the order of $\mathcal{O}(kn)$ max-propagations. So, the complexity of finding the best k MPEs is a factor n larger than the complexity of the MPE problem.

3.5 Pre-processing

One way to tackle the difficulty of the MAP and MPE problems is by trying to pre-process the input in order to reduce the computations later on. *Bolt* and *Van der Gaag* describe for MAP a way to exclude specific states of a variable or even determine the exact state of the variable in the MAP by looking at local properties [13]. Two propositions are introduced. The first proposition involves a variable and its Markov blanket. Suppose a variable v and its Markov blanket containing the variable set M. A proof is given that the state of v will be in the MAP solution if the probability of the state given any configuration $\mathbf{c}_M \in \Omega_M$ is always higher than the probability of the other states of the variable based on the same configuration. So, given a state $\mathbf{c}_v$:

$$\Pr(\mathbf{c}_v|\mathbf{c}_M) \geqslant max_{\mathbf{c}_l \in \Omega_v \backslash \{\mathbf{c}_v\}} \Pr(\mathbf{c}_l|\mathbf{c}_M) \quad \forall \mathbf{c}_M \in \Omega_M$$

The other way around is true as well: if the probability of the state is lower given any configuration, then the state can be excluded as candidate for the MAP solution.

For the second proposition, they use the so-called *MAP blanket*. Given the set of MAP variables V and a variable v, the MAP blanket is a set $W \subset V$ such that v is independent of $V \backslash W$ given W. A similar property is proven for v as in proposition 1 but now given the MAP blanket and possible evidence.

While determining whether the state of a variable can be fixed or excluded is exponential in the size of the Markov blanket and MAP blanket, the blanket is local. The size of the MAP blanket and the computations involved are negligibly small if the network is big enough. Moreover, the outcome that a state of a variable should be included or excluded may hold for multiple changes of the network when the changes in evidence affect only variables outside the Markov blanket or MAP blanket. This is a useful property in monitoring applications.

3.6 Robustness of the MPE

Other studies were conducted in which the MPE was investigated under conditions where the Bayesian network changed. We will first review a research in which the robustness of the MPE is examined when parameters of conditional probability tables change. The next research considers the solution space of the MPE after a change in evidence. The last research proves a couple of properties regarding persistence of the MPE in the junction tree when evidence changes.

Chan and *Darwiche* investigated how much a single parameter θ can be changed while the most probable explanation stays the same [14]. Given a Bayesian network with the variable set V, a variable $x \in V$ and a parameter $\theta = Pr(x_1|\mathbf{c}_P)$ in the conditional probability table of x in which x_1 is the state of x and $\mathbf{c}_P$ the configuration of $Pa(x)$. Now, let Ω_- be defined as follows:

$$\Omega_- = \{\mathbf{c}_V | \mathbf{c}_V \in \Omega_V \wedge \mathbf{c}_P^{\downarrow V} = \mathbf{c}_P \wedge x_1 \in \mathbf{c}_V\}$$

The authors now observe that the probabilities of the configurations in Ω_- are linear functions of θ. The MPE of Ω_- therefore as well. In essence, the probabilities of the other configurations not in Ω_- remain constant when θ changes and therefore, the MPE of these configurations remains constant as well. However, changing θ means that the parameter for at least one configuration $\mathbf{c}_Q$ that does not contain x_1 but is compatible with $\mathbf{c}_P$ ($\mathbf{c}_P = \mathbf{c}_P^{\downarrow Q}$) should change, as the configurations compatible with $\mathbf{c}_P$ should after all sum up to 1. Therefore, the configurations that are not compatible with $\mathbf{c}_P$ or do not contain x_1 are divided into groups. The first group is not compatible with $\mathbf{c}_P$. The other groups are compatible with $\mathbf{c}_P$ but are divided according to the states in $\Omega_x \backslash \{x_1\}$. The probabilities of the configurations in the first group that is not compatible with $\mathbf{c}_P$ remain constant when θ changes, while the others are a linear function of the other parameters.
The authors then describe on how to compute the constant representing the slope of the linear function for each parameter that is compatible with $\mathbf{c}_P$ and the constant of the group that is not compatible with $\mathbf{c}_P$. With this data, it can be computed how much θ can be adjusted before the MPE in one of the groups will be bigger. This can be used to look at the parameters and see which ones are important for the MPE and which ones have small influence.

Van Rooij looked at the set of configurations that could be candidate for the new MPE after a change in the evidence [15]. He proved that given the set of variables of a Bayesian network N, an evidence set E and the Markov blanket M_E of E, $\frac{2^{|V|-|E|}}{2^{|M_E|}}$ configurations cannot become the new MPE under the assumption that the state space of each variable has size two and M_E does not contain any variable in E.

Pastink and *Van der Gaag* researched the persistence of the MPE under evidence change [16] in a junction tree, in the context of max-propagation. Persistence can occur in cliques and separators and means that the configuration in the clique or separator that is part of the MPE remains part of the MPE after a change in evidence occurs.
Suppose a junction tree that contains a clique Cl_V with two separators Se_K and Se_L which both contain one binary variable. We could then divide the configuration space Ω_V such that for each configuration $\mathbf{c}_{KL} \in \Omega_{K \cup L}$ we have: $\{\mathbf{c}_V | \mathbf{c}_V \in \Omega_V \wedge \mathbf{c}_{KL} = \mathbf{c}_{KL}^{\downarrow V}\}$. This gives us four groups. When we determine of each group the MPE, then one of these MPE's is in accordance with the MPE of the

complete network. And the MPE's of these four groups will persist when changes occur outside Cl_V, since all configurations within a group will be altered with the same factor. Using these concepts, two terms are defined: Strong and weak persistence. When a change in a clique Cl_V occurs, we have strong persistence if:

- The configurations of the separators adjacent to Cl_V in the previous MPE of the network stay the same in the new MPE.
- The ratio with which the probability of the configuration is changed is bigger than or equal to the ratios with which the probabilities of the other configurations are changed.

Weak persistence occurs when the second condition does not hold. If strong persistence occurs, it is not necessary to continue max-propagation because the other values after the separator will persist into the new MPE as well. When weak persistence occurs, we do not have to propagate further as well, but when after another evidence change again a weak persistence occurs somewhere in the network, we cannot guarantee persistence anymore and should therefore do a full propagation.

Another interesting property of a changed MPE stated in the paper is as follows: Suppose we have a clique Cl_V in which a change occurs and two separators Se_K, and Se_L, such that Se_L is adjacent to Cl_V and lies between Cl_V and Se_K. If the new MPE contains a different configuration for Se_K as the previous one, Se_L will have a different configuration as its previous one as well. This property limits the number of configurations that can become the new MPE after a change in evidence. Configurations that have, for example, the same state for the variables in K but different states for the variables in L compared to the MPE, cannot become the new MPE after a change in evidence.

4 Monotone decrease

In this section, it will be proved that given three cliques - Cl_V, Cl_W and Cl_X - where Cl_W lies between Cl_V and Cl_X in the junction tree, if a change occurs in the potential of Cl_V, then the effects on the potential of Cl_X will be the same as or smaller than the effect on the potential of Cl_W. We will call this monotone decrease.

When propagation is performed, new probabilities for the potential of a clique are computed by multiplying the old probabilities with the incoming messages from the adjacent separators. Since a factor in such a message represents a division of the probabilities in the new potential by the probabilities of the old potential in the separator, we can look at the ratio between the probabilities of the new and old potential of a separator in order to gain insight into how the potential of the clique changes. Suppose we have a clique Cl_V and adjacent separator Se_L. The separator sends a message to the clique so that the clique can process the changes correctly. This message contains for each configuration in the separator a factor. Each factor corresponding to a configuration $\mathbf{c}_L$ will be multiplied with the probabilities of the following configurations in the clique:

$$\{\mathbf{c}_V | \mathbf{c}_V \in \Omega_V \wedge \mathbf{c}_L^{\downarrow V} = \mathbf{c}_L\}$$

The configurations in the separator will partition the configurations in the clique. Suppose now that one factor is equal to 2 and an other factor equal to 0.5. Then, the probabilities in the corresponding partitions in the clique will be multiplied with respectively 2 and 0.5. This causes the ratios between the probabilities of any configuration from one partition and any configuration of the other partition is to be increased by 4. This ratio in the clique is what the proof is about: in subsequent cliques, we will know that the ratio will never be more than 4. In order to prove this, we will first take a look at the two phases in propagation.
We further on make the following assumptions:

- Only changes of probabilities in the potential of a single clique may occur at the same time.
- Before the change will be propagated through the junction tree, we assume that there is no potential that has a probability of zero for a configuration. Later on, we will discuss how zero probabilities can be interpreted.

4.1 Only the distribution phase

When there are changes introduced in a single clique, we can skip the collection phase and only perform the distribution phase in order to process the evidence correctly in the junction tree. This was already mentioned by *Dawid* [4]. The requirement is that we start the distribution phase at the clique in which changes are introduced. We will state this property in a lemma and provide a proof.

Lemma 1. When the probabilities in the potential of a single clique change, the distribution phase starting from that clique is the only phase needed in order to update the potentials of the other cliques correctly.

Proof. Before the introduction of evidence in the junction tree, the potential of each clique and separator represents the marginal distribution over the variables that are present in the clique or separator. Suppose we have a separator Se_L and clique Cl_R where $L \subset R$, then we know that computing the marginal probabilities of the variables in L by marginalizing out the variables $R \backslash L$ in

ϕ_R will give us the same probabilities as ϕ_L.
Suppose now that evidence is introduced in clique Cl_O. By choosing Cl_O as the source clique, all separators will receive a message in the collection phase from a clique of which the potential is the same as before the introduction of evidence. Clique Cl_O will only receive messages in this phase. Recall that the message sent by a clique Cl_R to an adjacent separator Se_L is defined as follows:

$$\phi_L(\mathbf{c}_L) = \sum_{\mathbf{c}_R \in \Omega'_R} \phi_R(\mathbf{c}_R) \textit{ where } \Omega'_R = \{\mathbf{c}_R | \mathbf{c}_R \in \Omega_R \wedge \mathbf{c}_L^{\downarrow R} = \mathbf{c}_L\} \quad \forall \mathbf{c}_L \in \Omega_L$$

We already stated that this marginalized distribution computed by Cl_R is the same as the current potential of separator Se_L because potential ϕ_R has not changed. Replacing the potential with the new potential in Se_L will therefore not cause any changes. The processing of a message sent by a separator Se_L to a clique Cl_S is defined as follows:

$$\phi_S(\mathbf{c}_S) = \phi_S(\mathbf{c}_S)\frac{\phi'_L(\mathbf{c}_L^{\downarrow S})}{\phi_L(\mathbf{c}_L^{\downarrow S})} \quad \forall \mathbf{c}_S \in \Omega_S$$

Because the old and new potential of the separator is the same, the probabilities in the clique will be multiplied with a factor equal to one. Therefore, the new potential of each clique including the source clique Cl_O remain the same after the collection phase. The potentials of the separators did not change as well which makes the collection phase redundant. □

4.2 Proof of monotone decrease

Now that we have proven that the distribution phase suffices to propagate a change throughout the junction tree, we can prove the following theorem.

Theorem 2. Given a clique Cl_V in which the probabilities in the potential changed. Then, for any two separators Se_K and Se_L where Se_K lies on the path between Cl_V and Se_L in the junction tree, it holds that:

$$\max_{\mathbf{c}_K \in \Omega_K} \frac{\phi'_K(\mathbf{c}_K)}{\phi_K(\mathbf{c}_K)} \geqslant \frac{\phi'_L(\mathbf{c}_L)}{\phi_L(\mathbf{c}_L)} \quad \textit{and} \quad \min_{\mathbf{c}_K \in \Omega_K} \frac{\phi'_K(\mathbf{c}_K)}{\phi_K(\mathbf{c}_K)} \leqslant \frac{\phi'_L(\mathbf{c}_L)}{\phi_L(\mathbf{c}_L)} \quad \forall\ \mathbf{c}_L \in \Omega_L$$

where $\mathbf{c}$ ϕ indicates a potential with probabilities before the update from Cl_V and ϕ' a potential that contains updated probabilities as a result of the changes in Cl_V.

Proof. Suppose Se_K and Se_L are connected to the same clique Cl_V and Se_K is closer to the source clique Cl_S in which the potential changed. And suppose that potential $\phi'_K(\mathbf{c}_K)$ is updated to the new potential $\phi'_K(\mathbf{c}_K)$. New probabilities for the potential of Cl_V can be computed by:

$$\phi'_V(\mathbf{c}_V) = \phi_V(\mathbf{c}_V) \cdot \frac{\phi'_K(\mathbf{c}_K^{\downarrow V})}{\phi_K(\mathbf{c}_K^{\downarrow V})} \quad \forall\ \mathbf{c}_V \in \Omega_V$$

By lemma 1, these will be the final probabilities for the potential of Cl_V, as we do not need to collect the results from the other separators connected to Cl_V since the factors will be equal to 1.
New probabilities for the potential of Se_L can be computed by:

$$\phi'_L(\mathbf{c}_L) = \sum_{\mathbf{c}_V \in \Omega_{V \setminus L}} \phi'_V(\mathbf{c}_V, \mathbf{c}_L) \quad \forall \mathbf{c}_L \in \Omega_L$$

Let us take a specific configuration $\mathbf{c}_L$ of Se_L. The new probability for the potential of Se_L can be rewritten as:

$$\phi'_L(\mathbf{c}_L) = \sum_{\mathbf{c}_V \in \Omega_{V \setminus L}} \phi'_V(\mathbf{c}_V, \mathbf{c}_L) = \sum_{\mathbf{c}_V \in \Omega_{V \setminus L}} \phi_V(\mathbf{c}_V, \mathbf{c}_L) \cdot \frac{\phi'_K(\mathbf{c}_K^{\downarrow V})}{\phi_K(\mathbf{c}_K^{\downarrow V})}$$

The new factor of Se_L for configuration $\mathbf{c}_L$ is equal to: $\frac{\phi'_L(\mathbf{c}_L)}{\phi_L(\mathbf{c}_L)}$. Now, it can be easily shown that the largest factor of Se_K is always larger than or equal to the factors in Se_L:

$$\begin{aligned}
\frac{\phi'_L(\mathbf{c}_L)}{\phi_L(\mathbf{c}_L)} &= \frac{\sum_{\mathbf{c}_V \in \Omega_{V \setminus L}} \phi_V(\mathbf{c}_V, \mathbf{c}_L) \cdot \frac{\phi'_K(\mathbf{c}_K^{\downarrow V})}{\phi_K(\mathbf{c}_K^{\downarrow V})}}{\sum_{\mathbf{c}_V \in \Omega_{V \setminus L}} \phi_V(\mathbf{c}_V, \mathbf{c}_L)} \\
&\leqslant \frac{\sum_{\mathbf{c}_V \in \Omega_{V \setminus L}} \phi_V(\mathbf{c}_V, \mathbf{c}_L) \cdot \max_{\mathbf{c}_m \in \Omega_K} \frac{\phi'_K(\mathbf{c}_m)}{\phi_K(\mathbf{c}_m)}}{\sum_{\mathbf{c}_V \in \Omega_{V \setminus L}} \phi_V(\mathbf{c}_V, \mathbf{c}_L)} \\
&\leqslant \max_{\mathbf{c}_m \in \Omega_K} \frac{\phi'_K(\mathbf{c}_m)}{\phi_K(\mathbf{c}_m)} \cdot \frac{\sum_{\mathbf{c}_V \in \Omega_{V \setminus L}} \phi_V(\mathbf{c}_V, \mathbf{c}_L)}{\sum_{\mathbf{c}_V \in \Omega_{V \setminus L}} \phi_V(\mathbf{c}_V, \mathbf{c}_L)} \\
&\leqslant \max_{\mathbf{c}_m \in \Omega_K} \frac{\phi'_K(\mathbf{c}_m)}{\phi_K(\mathbf{c}_m)}.
\end{aligned}$$

The same procedure can be used to show that the smallest factor of Se_K is always smaller than or equal to the factors in Se_L:

$$\begin{aligned}
\frac{\phi'_L(\mathbf{c}_L)}{\phi_L(\mathbf{c}_L)} &= \frac{\sum_{\mathbf{c}_V \in \Omega_{V \setminus L}} \phi_V(\mathbf{c}_V, \mathbf{c}_L) \cdot \frac{\phi'_K(\mathbf{c}_K^{\downarrow V})}{\phi_K(\mathbf{c}_K^{\downarrow V})}}{\sum_{\mathbf{c}_V \in \Omega_{V \setminus L}} \phi_V(\mathbf{c}_V, \mathbf{c}_L)} \\
&\geqslant \frac{\sum_{\mathbf{c}_V \in \Omega_{V \setminus L}} \phi_V(\mathbf{c}_V, \mathbf{c}_L) \cdot \min_{\mathbf{c}_m \in \Omega_K} \frac{\phi'_K(\mathbf{c}_m)}{\phi_K(\mathbf{c}_m)}}{\sum_{\mathbf{c}_V \in \Omega_{V \setminus L}} \phi_V(\mathbf{c}_V, \mathbf{c}_L)} \\
&\geqslant \min_{\mathbf{c}_m \in \Omega_K} \frac{\phi'_K(\mathbf{c}_m)}{\phi_K(\mathbf{c}_m)} \cdot \frac{\sum_{\mathbf{c}_V \in \Omega_{V \setminus L}} \phi_V(\mathbf{c}_V, \mathbf{c}_L)}{\sum_{\mathbf{c}_V \in \Omega_{V \setminus L}} \phi_V(\mathbf{c}_V, \mathbf{c}_L)} \\
&\geqslant \min_{\mathbf{c}_m \in \Omega_K} \frac{\phi'_K(\mathbf{c}_m)}{\phi_K(\mathbf{c}_m)}.
\end{aligned}$$

Note that we made an assumption about Se_K and Se_L being two separators adjacent to the same clique. But because of the transitivity property of $\leqslant$ and $\geqslant$, Se_K can just be any separator between the source clique Cl_S and separator Se_L. □

4.3 Assumptions

Let us now consider the two assumptions we have made. First of all, we allowed only changes in a single clique. Multiple changes can be handled by propagating them one-by-one. So, introduce the first changes, perform a full propagation, and continue with the next change after that until all changes have been processed.

The second assumption said that the separators may not contain zero probabilities before propagation. Zero probabilities are however likely to be present in a junction tree. Either a CPT contains zero probabilities resulting in zero probabilities in a potential or evidence that a variable has a specific state is introduced and propagated. Due to this, potentials may contain probabilities of zero.

When we apply the theorem while zero probabilities exist in the junction tree, we will come across division by zero. This is not possible and gives us ratios of infinity. One way to deal with these zero probabilities is by simply ignoring these configurations. Theorem 2 will still hold for the remaining configurations.

4.4 Concluding observations

We have now proven that given:

- two separators Se_K and Se_L and a clique Cl_V;
- such that Se_K lies on the path between Cl_V and Se_L;
- a change of the probabilities in the potential of Cl_V,

the largest factor of Se_L will be smaller than or equal to the largest factor of Se_K and likewise, the smallest factor of Se_L is larger than or equal to the smallest factor of Se_K.

This theorem helps us in getting insight in the changes of the MPE after a change in evidence. The factors in a message sent by a separator tell us by how much the probabilities in the adjacent clique will increase or decrease and therefore give information about how the joint probability distribution in a clique changes. We know that the distribution in a clique, that receives a message from the separator Se_L during the propagation, will be subject to changes that are at most the same as the changes in the clique that received a message from the separator Se_K. Let us consider the joint probability distribution written as the product of potentials of cliques divided by the product of potentials of separators. Filling in the MPE will gives us the highest possible probability. Probably not all potentials will return the configuration of the highest probability since some have to compromise with their neighbouring cliques. However, we will see maximum probabilities or small groups of potentials that combined will give the highest probability. If these single cliques or groups of cliques are far away from the source of change in evidence, the influence of this change on their joint probability distributions will likely be small, causing the best configuration of this set of cliques to retain its highest probability and is therefore likely to be part of the MPE again. While for single cliques or group of cliques closer to the source of change in evidence, the probabilities in the joint probability distribution will be changed more rigorously due to which other configurations may become the best configuration and therefore cause changes in the MPE.
However, although we know that the changes do not increase, we do not know how strongly the changes reduce in size and therefore do not know to what extent the change in evidence influences the cliques. Theorem 2 is however useful as basis on how to investigate the changes in the MPE. We will later on perform experiments in which we use the theorem as a basis and see how the changes reduces in size.

5 Visualizing propagation of evidence

While the monotone decrease of changes in cliques during propagation of evidence through the junction tree has been proven, the strength of the monotone decrease is not known. In this section, we will introduce a visualization to gain insight in this monotone decrease. The visualization should meet various requirements. First of all, the effects of propagation after each clique should be visible. Secondly, the components in the visualization should show the monotone decrease through an appropriate ordering of cliques. With respect to the second requirement, we note that propagation is performed in a tree of cliques. When we continue propagation from a clique via multiple paths, the monotone decrease in one path will have no relation with the monotone decrease in another path. Therefore, decreases along different paths should be visualized separately. Before we introduce our visualization, we review another visualization proposed for the MPE problem.

5.1 Visualization of Joint Probability Distribution

Haipeng Guo used a visualization of the joint probability distribution [17] for his research on the MPE problem. He used the visualization to demonstrate, what he called, the *multifractal*-property of joint probability distributions. The *multifractal*-property was subsequently used by him as a heuristic for an approximation algorithm for the MPE problem. We will first address the visualization after which we will briefly state the *multifractal*-property. Further information about the visualization and the *multifractal*-property can be found in *Haipeng Guo*'s dissertation [17].

The visualization for a joint probability distribution of a Bayesian network consists of a barchart in which each configuration of the variables in the Bayesian network is represented by a bar. The height of the bar indicates the probability of the configuration. The MPE is the configuration with the highest bar. An example of such a visualization is depicted in Figure 6. The order in which the bars are arranged is a systematic order of all configurations. Take for example variable a and b with $\Omega_a = \{a_1, a_2\}$ and $\Omega_b = \{b_1, b_2\}$. A resulting order could then be: $\{a_1, b_1\}, \{a_2, b_1\}, \{a_1, b_2\}, \{a_2, b_2\}$.

With this visualization, the *multifractal*-property comes to the fore. In order to understand the meaning of this property, let us look at Figure 6. The distribution is characterized by three clusters of configurations with high probabilities compared to their neighbourhood. The structure of these clusters is very similar. When we analyze each cluster from left to right, we can make the following observations: It starts with configurations that have the highest probabilities in the cluster. Thereafter, the probability decreases as we shift to the configurations in the middle. Continuing to the right side of the cluster, the probabilities increase again. The three clusters themselves can be found in other parts of the distribution as well. For example, on the left side of the middle cluster, we see three small clusters. These three clusters are relatively very similar. The *multifractal*-property is about these re-occurrences of distributions in the joint probability distribution.

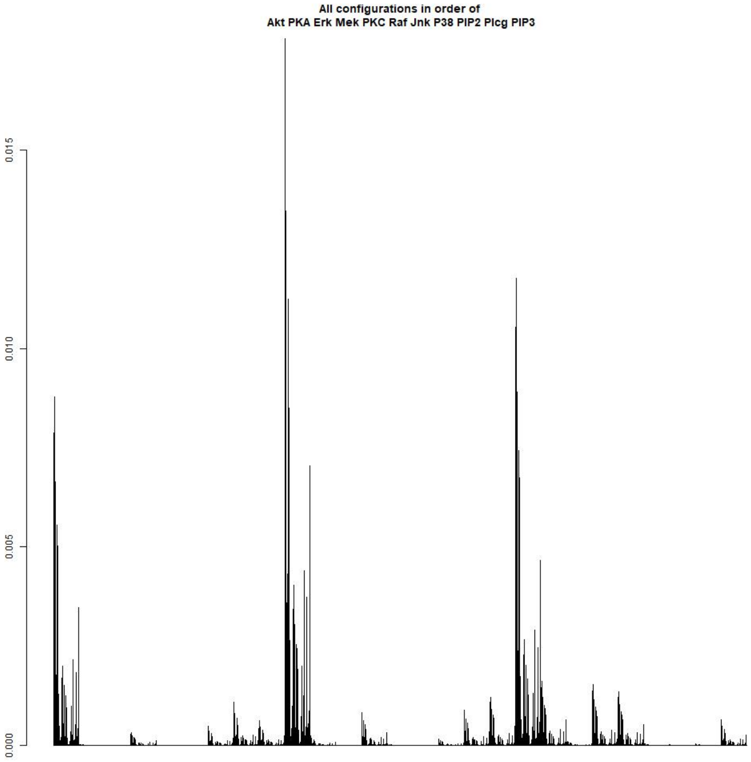


Figure 6: The probability distribution of all joint configurations of the *Sachs*-network [1]

5.2 The visualization

In contrast with *Haipeng Guo*'s work, we are not so much interested in the probabilities of all joint configurations, but rather in the ratios between the new and old probability of a configuration after a change of evidence. Therefore, this ratio will be used as the height of the bar instead of the probability. In our visualization, we distinguish between five types of ratio for a given joint configuration:

- The ratio is greater than 1, which means that the probability of this configuration increases after the change in evidence. The bar representing the ratio is colored red.
- The ratio is between 1 and 0, which means that the probability of the configuration decreases. The bar representing the ratio is colored blue. The bar will be displayed on the negative y-axis in order to indicate the decrease of probability. The size of the decrease is visualized by the reciprocal of the ratio. This ensures that the meaning of the size of a bar is equal for both increase and decrease.
- The ratio is equal to 1, which means that the probability of the configuration does not change. This will be indicated with a gray bar across the whole y-axis.
- The ratio is equal to 0, which indicates that the probability after the change in evidence is equal to zero. These bars will be displayed on the negative y-axis as well but with the color yellow. In order to display the bar, the height is equal to the height of a bar with the most increase

instead of infinity.

- The ratio is undefined, which happens when the probability before the change in evidence is zero. This can happen when the value of one of the variables in the evidence changes. These configurations will be indicated with a yellow bar as well but then on the positive y-axis. The height of the bar is equal to the height of a bar with the most decrease instead of an infinite height.

The different types of bar are displayed in Figure 7.

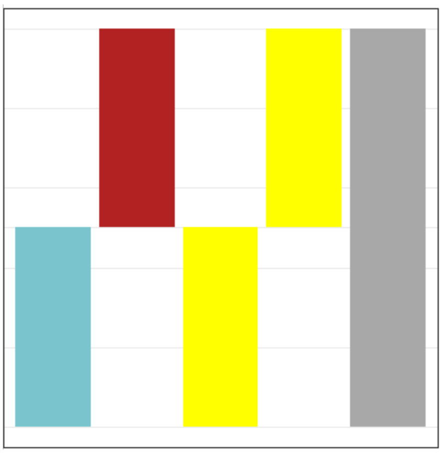


Figure 7: The types of bar used in the visualization

Now, we want to visualize the monotone decrease after propagation through the junction tree. In order to show the progress of the decrease through the tree, we visualize parts of configurations instead of complete configurations. Suppose we have currently propagated the evidence from Cl_V to Se_L and have to update the adjacent clique Cl_S next. As we have seen, the formula for updating the clique is as follows:

$$\phi_S(\mathbf{c}_S) = \phi_S(\mathbf{c}_S)\frac{\phi'_L(\mathbf{c}_L^{\downarrow S})}{\phi_L(\mathbf{c}_L^{\downarrow S})} \quad \forall \mathbf{c}_S \in \Omega_S$$

Since the ratios between the new and old potential values for the various configurations of the separator Se_L fully characterizes the ratios by which the potential of clique Cl_S changes, it suffices to visualize the ratios for just the separators. There is however one clique that is not preceded by a separator, the source clique in which the evidence changed. In order to handle this situation, a dummy clique is used. This dummy clique consists of the variable that caused the change in evidence and is connected to the source clique by a separator. The variables in the separator and the dummy clique are the same. Instead of introducing the evidence in the source clique, it will be introduced in the dummy clique. The separator can then be used in the visualization to show how the probabilities of the configurations in the source clique change.

We visualize configurations per separator in the junction tree instead of using configurations involving all variables. A major advantage of this visualization is that the number of bars does not grow exponentially in the number of variables, as in the visualization of *Haipeng Guo*. For the potential of each separator in the junction tree, a barchart is visualized in which each configuration is represented by a bar that tells us how the joint probability of this configuration is changed. The order

on configurations of a set variables is the same as in the visualization of *Haipend Guo*. We further arrange the barcharts per separator such that the paths in the junction tree are clearly visible. To this end, the charts on a path that is not split into multiple subpaths are visualized consecutively. When there is a split, we will continue the charts for one subtree on the same row, while for every other subpath, a new row of charts is constructed, connected by a line to indicate the charts that are prior to the subpath. This procedure is repeated when new splits occur.
We will now make the above description of the visualization more clear with an example.

Example 1. In Figure 8, a visualization can be seen that is the result of a propagation through our example network after evidence of variable f is introduced. Let us say that $\Omega_f = \{f_1, f_2\}$. The probability of $\mathbf{c}_f = f_2$ has become zero. The probability that $\mathbf{c}_f = f_1$ is increased because we have to normalize the probabilities. Since there are no other possible values for f then f_1, the probability will be normalized to one. This is depicted in the first chart. The two bars shows us how the clique - that contains f - will be altered. The probabilities of all configurations of the variables in the clique containing f_2 will be put to zero, while the probabilities of the other configuration will be multiplied by a factor roughly equal to 4.5. The clique has one other separator containing the variable d. The second chart shows us how the probability for the first configuration increases by a factor two, while the probability of the other configuration decreases by more than two. The probabilities of the configurations in the adjacent clique will be updated with these factors. Next, the path splits because the clique has two more separators. One separator with variables $\{b, c\}$ and the other with

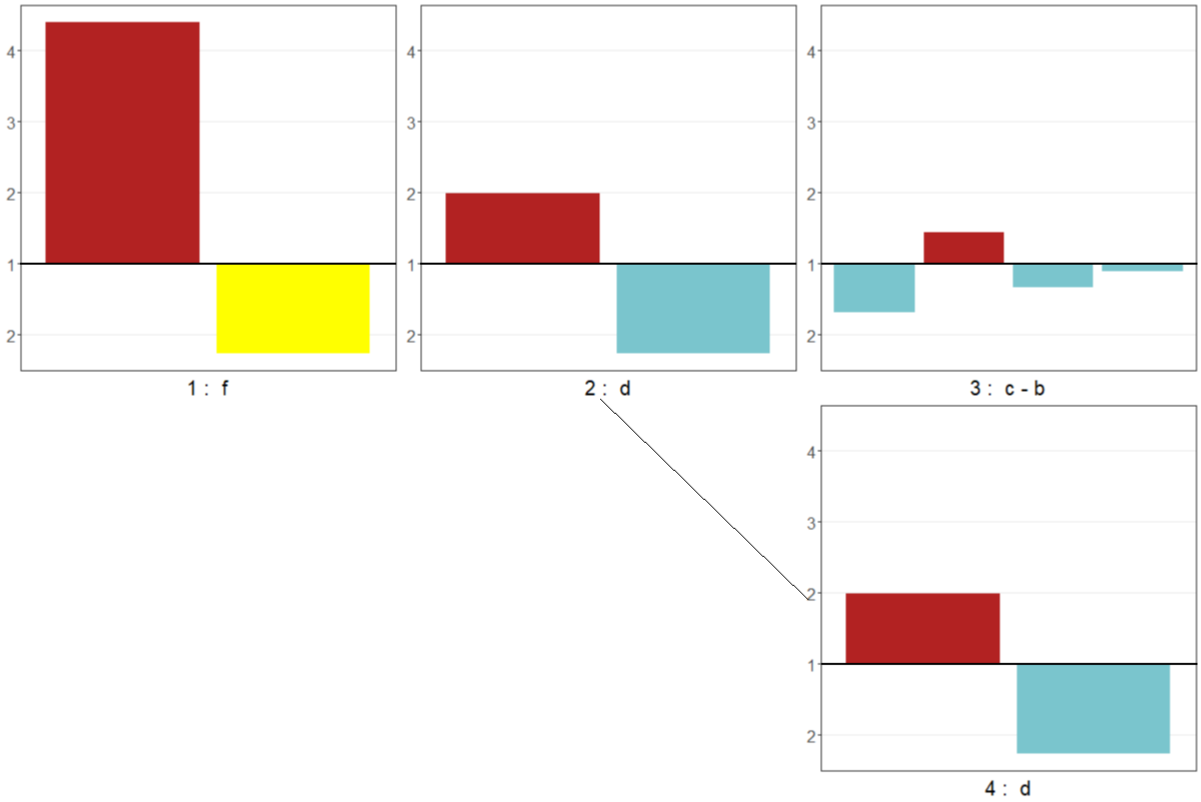


Figure 8: Propagation of evidence for f in the example Bayesian network

d again. The chart of the last named separator will be put on a new row and since it contains the same variable as the previous separator, the chart will be the same. For the other separator, we see four configurations. The probability of one configuration increases and the probabilities of the other three decrease. $\square$

5.3 Interpretation of visualization

The visualization shows how the potentials of the cliques are changed by presenting the change in potentials of the separators. The advantage of using the potentials of the separators is that we just proved that the change decreases monotonically as we propagate down the junction tree. This means that given a chart G, all charts that come after G on the right will have rates of increase and decrease that are smaller than the maximum rates of increase and decrease in G.

Furthermore, the change of probabilities in the potentials has an additional interpretation. Given a separator Se_L and clique Cl_A. The changes in Se_L do not only show how the marginal probability distribution of the variables in Cl_A changes, but it shows us also how the probability distribution of the variables in Cl_A plus the variables in the cliques behind Cl_A is changed. Suppose we have a smaller junction tree JT with variable set N, cliqueset $\mathbf{Cl} = \{Cl_A, Cl_B, Cl_C\}$ and separator set $\mathbf{Se} = \{Se_K, Se_L\}$ where $K = A \cap B$ and $L = B \cap C$. The joint probability distribution can be written as:

$$\Pr(\mathbf{c}_N) = \frac{\phi_A(\mathbf{c}_A^{\downarrow N})\phi_B(\mathbf{c}_B^{\downarrow N})\phi_C(\mathbf{c}_C^{\downarrow N})}{\phi_K(\mathbf{c}_K^{\downarrow N})\phi_L(\mathbf{c}_L^{\downarrow N})}$$

Suppose there is new evidence in clique Cl_A. After propagation of the evidence, we get a junction tree with new potentials, say JT' with the joint probability distribution:

$$\Pr(\mathbf{c}_N) = \frac{\phi'_A(\mathbf{c}_A^{\downarrow N})\phi'_B(\mathbf{c}_B^{\downarrow N})\phi'_C(\mathbf{c}_C^{\downarrow N})}{\phi'_K(\mathbf{c}_K^{\downarrow N})\phi'_L(\mathbf{c}_L^{\downarrow N})}$$

If we now take a close look at $\phi'_C(\mathbf{c}_C^{\downarrow N})$, then according to the rules in propagation, it is equal to:

$$\frac{\phi'_L(\mathbf{c}_L^{\downarrow N})}{\phi_L(\mathbf{c}_L^{\downarrow N})}\phi_C(\mathbf{c}_C^{\downarrow N})$$

Substituting this into the joint probability distribution of JT' and factoring out $\phi'_L(c_L)$ gives us:

$$\Pr(\mathbf{c}_N) = \frac{\phi'_A(\mathbf{c}_A^{\downarrow N})\phi'_B(\mathbf{c}_B^{\downarrow N})\phi_C(\mathbf{c}_C^{\downarrow N})}{\phi'_K(\mathbf{c}_K^{\downarrow N})\phi_L(\mathbf{c}_L^{\downarrow N})}$$

After doing the same for Cl_B, we see that the difference between the distributions of JT and JT' comes down to the ratio between $\phi_A(\mathbf{c}_A^{\downarrow N})$ and $\phi'_A(\mathbf{c}_A^{\downarrow N})$. These changes result from the change in probability of the configurations of the variables in the evidence which are presented in the first chart of our visualization. Therefore, the first chart does not only tell us how the joint probability distribution of the variables in the first cliques change, but it tells us exactly how the joint probability distribution containing all variables change.

Now let us remove clique ϕ_A and separator ϕ_K from JT and JT'. The remaining potentials form the joint probability distribution over the variables in $M = N \backslash (A \backslash K)$. This is true because of the following: In order to have a joint probability distribution over the variables in M, we have

to marginalize out the variables in $A\backslash K$. Since Cl_A contains the marginal distribution over the variables in A, we only have to perform marginalization in Cl_A over the variables $A\backslash K$ to get the joint probability distribution over the variables in M. However, marginalizing out the variables $A\backslash K$ results in a potential that is equal to the potential of Se_K and therefore cancel each other out.
Let us look at the ratio between these partial joint probability distribution of JT and JT'. For the new potentials, we can again rewrite $\phi'_L(\mathbf{c}_L^{\downarrow N})$ and $\phi'_K(\mathbf{c}_K^{\downarrow N})$ which gives us:

$$\frac{\frac{\phi'_K(\mathbf{c}_K^{\downarrow N})}{\phi_K(\mathbf{c}_K^{\downarrow N})}\phi_B(\mathbf{c}_B^{\downarrow N})\phi_C(\mathbf{c}_C^{\downarrow N})}{\phi_L(\mathbf{c}_L^{\downarrow N})}$$

As we can see, the ratio between the partial joint probability distribution of JT and JT' comes down to the ratio between the old and new potential of separator Se_K. Hence, each chart in the visualizations also shows how the joint probability distribution changes of all variables of the cliques in the subtree.

6 The effect of changes on persistence of the MPE

In section 4, we have seen that standard propagation of evidence has a property of monotone decrease in the junction tree. Previous research showed that persistence properties could be derived when performing max-propagation. These two findings lead us to investigate whether persistence properties could be defined for the MPE when performing standard propagation after a change in evidence.

6.1 Persistence

As we have seen in chapter 3.6, *Pastink* and *Van der Gaag* came up with two types of persistence: Weak and strong persistence [16]. While these two types of persistence were defined in the context of the max-propagation algorithm for computing the MPE, we can ask ourselves whether we can derive similar types of persistence for standard propagation. For weak persistence the results are negative. An example is shown in Figures 9 and 10, where Figure 10 results from Figure 9 after updating the prior distribution of the variable k in clique Cl_R with the multipliers: 1.25 for k_1, 0.901 for k_2 and 0.905 for k_3 and subsequently propagating changes through Se_L to Cl_S. The MPE of the original distribution from Figure 9 is the configuration $\{k_2, b_2, c_2\}$; the MPE from Figure 10 is the configuration $\{k_1, b_1, c_1\}$. Weak persistence is defined for separators and occurs when the probability of the configuration in the separator that is part of the MPE remains the highest probability after propagation of the changed evidence. As we can see, the configuration $\{b_2\}$ in separator Se_L remains to have the highest probability, thus satisfying the criterion for weak persistence. However, the state for the variable b in the MPE changes. Therefore, there is no actual persistence upon standard propagation. So, we can conclude that weak persistence is only valid in the context of max-propagation.

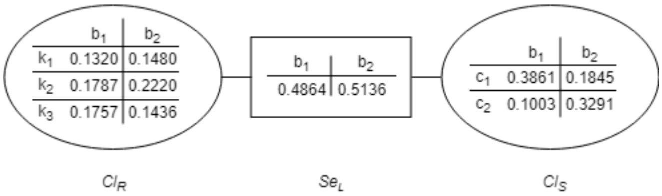


Figure 9: A junction tree with its potentials for Cl_R, Cl_S and Se_L. The MPE is the configuration $\{k_2, b_2, c_2\}$

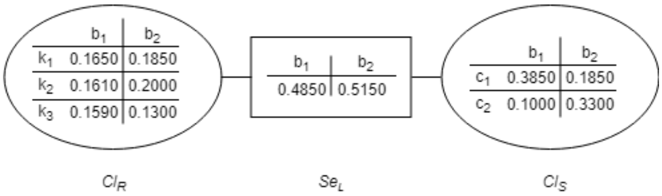


Figure 10: Potentials for Cl_R, Cl_S and Se_L after adjusting the prior distribution of the variable k and updating the junction tree. The MPE changes to the configuration $\{k_1, b_1, c_1\}$

The example also shows that strong persistence does not hold for standard propagation. Recall that strong persistence occurs in a separator when the following conditions hold for the configuration that is part of the MPE: Its probability remains the highest in the separator and the increase of its probability is at least as high as that of any other probability in the separator. We can use the visualization from section 5 to look at the changes in the junction tree of our example. These changes are shown in Figure 11. The second graph shows that the configuration with the highest increase is b_2. However, we have just seen that b_1 is the configuration for variable b in the new MPE and does therefore not persist. This means that strong persistence does not hold either in standard propagation.

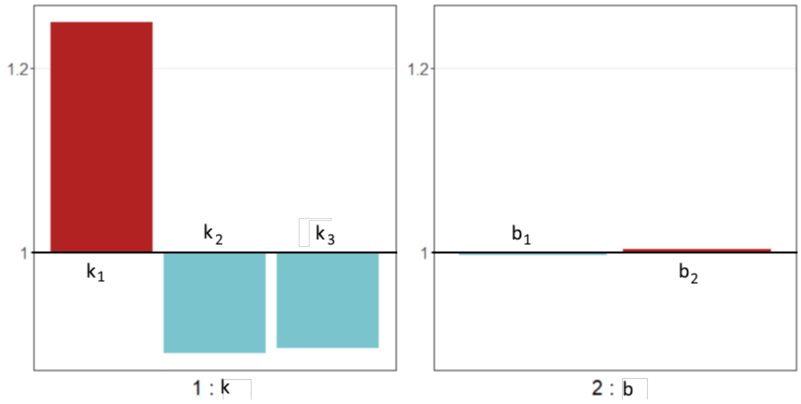
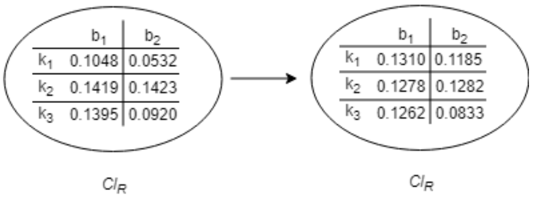


Figure 11: Propagation of changes in prior distribution of variable k

We take a look at why these two properties hold for max-propagation and not for standard propagation. In order to do this, we should consider the knowledge we have available about the MPE in max-propagation and in standard propagation, respectively. The knowledge can be divided in global and local knowledge. Global knowledge is the knowledge that is available at each clique and separator, while local knowledge is only available at a specific clique or separator. Let V be the set of all variables in a junction tree. When max-propagation has taken place, we have the following knowledge:

- *Global knowledge*: The configuration of the MPE $\mathbf{c}_{MPE}$ and its probability $\Pr(\mathbf{c}_{MPE})$
- *Local knowledge*: For a clique or separator and its potential ϕ_X, we know for each of its configurations $\mathbf{c}_X$ that $\phi_X(\mathbf{c}_X) = max_{\mathbf{c}_V \in \Omega_V} \Pr(\mathbf{c}_V)$ under the condition that $\mathbf{c}_X = \mathbf{c}_X^{\downarrow V}$.

Suppose that max-propagation is performed in our example instead of standard propagation and we want to process the changes to the prior distribution of the variable k into clique Cl_R. In this case, the process of updating the local knowledge in max-propagation is the same as for standard propagation. Figure 12 displays the probabilities in clique Cl_R before and after the update.



	b1	b2
k1	0.1048	0.0532
k2	0.1419	0.1423
k3	0.1395	0.0920

	b1	b2
k1	0.1310	0.1185
k2	0.1278	0.1282
k3	0.1262	0.0833

Figure 12: The potential of Cl_R on the left contains probabilities resulting from max-propagation. The potential on the right shows the probabilities in Cl_R after an update of the local knowledge.

From this local knowledge, we know that the highest probability is the probability of the MPE. In order to find the configuration $\mathbf{c}_{MPE}$, we have to continue max-propagation through the junction tree. When the highest probability in Se_L remains the highest after max-propagation has taken place in the separator, we know that for the other cliques and separators after Se_L the highest probability will remain the highest as well. We therefore have persistence since the configurations that are part of the MPE in all cliques and separators after Se_L (and Se_L itself) will not change.

In standard propagation however, we are missing some crucial information. Before the change in evidence, we have the same global knowledge as in max-propagation. However, the local knowledge is different. We only have the joint probabilities of the variables in the specific clique or separator at our disposal. Suppose we are at clique Cl_R and collapse clique Cl_S and separator Se_L into a single clique Cl_X with potential ϕ_X. The joint probability distribution can then be written as: $\Pr(\mathbf{c}_V) = \phi_R(\mathbf{c}_R^{\downarrow V})\phi_X(\mathbf{c}_X^{\downarrow V})$. Since we have the probability of the MPE and the probabilities ϕ_R available, we can find the probability $\phi_X(\mathbf{c}_X^{\downarrow MPE})$. From this, we can also compute the probability for any other configuration $\mathbf{c}_V$ for which it holds that $\mathbf{c}_X^{\downarrow V} = \mathbf{c}_X^{\downarrow MPE}$. For any other configuration however, we will not be able to compute the probability. The only thing we can say for each of these configurations, say $\mathbf{c'}_V$, is that $\phi_R(\mathbf{c'}_R^{\downarrow V})\phi_X(\mathbf{c'}_X^{\downarrow V}) < \phi_R(\mathbf{c}_R^{\downarrow MPE})\phi_X(\mathbf{c}_X^{\downarrow MPE})$. When evidence is introduced in clique Cl_R, we update our local knowledge. This gives us the new joint probabilities of the variables in R. From section 5.3, we know that updating Cl_R is enough to have the correct joint probability distribution again. This means that we know the new probabilities for $\Pr(\mathbf{c}_{MPE})$ and the other configurations that are consistent with $\mathbf{c}_X^{\downarrow MPE}$. We will however not know whether the MPE remains the MPE, since we have too little information about $\phi_X(\mathbf{c}_X^{\downarrow V'})$. The only thing we know is how the probabilities of the configurations will change, but we don't know how much the probability of one of these configurations has to change in order to become the MPE.

6.2 Persistence in standard propagation

Interestingly, another insight can be used to say something about persistence. While strong persistence concerns the increase and decrease in probabilities of the configurations in the separator, we will look at the increase and decrease in probabilities of the configurations in the clique after the separator. We will see that in some cases, it can be proven that the configuration of the variables, present in the clique but absent in the separator, will persist. We will first describe the situation in which such persistence can occur. This situation will then be used in a lemma and theorem to prove the persistence.

We consider a separator Se_L and its adjacent clique Cl_R. Se_L has the single variable k with

$\Omega_k = \{k_1, k_2\}$; later on, the theorem will be extended to separators with multiple variables and larger state spaces. The separator Se_L splits the junction tree into two parts in the sense that these parts are independent given Se_L. The part that does not contain Cl_R will be indicated with X and the other part with Y'; Y indicates the part of the junction tree of Y' without Cl_R. X, Y and Y' will also be used to indicate the sets of variables that are present in the cliques and separators in that part. The cliques and separators in a part can be collapsed into a single big clique. Doing this for $X$and Y gives us two cliques with the potentials ϕ_X and ϕ_Y. In our theorem, the joint probability distribution will be used. Since we will not perform any changes to individual cliques in X or Y, we can use ϕ_X and ϕ_Y in the joint probability distribution to represent the potentials of the cliques and separators in X and Y. The situation is depicted in Figure 13. In this figure, clique Cl_R contains two other variables in addition to the variable k, both with a state space of size two. For our theorem, it will not matter what the number of variables and the size of their state spaces are in Cl_R as long as they are not present in X. For ease of exposition, we assume that any other separator adjacent to Cl_R does not contain the variable k. Let $\{k_1, b_1, c_1\}$ be part of the MPE before any change in evidence.

Figure 13: Example junction tree

Now suppose that a change in evidence occurs in one of the variables in X. Propagating the changed evidence gives a new potential ϕ'_L for Se_L. The probabilities of the configurations in ϕ_R containing k_1 are subsequently multiplied with $\mathbf{x} = \frac{\phi'_L(k_1)}{\phi_L(k_1)}$ and the ones containing k_2 with $\mathbf{y} = \frac{\phi'_L(k_2)}{\phi_L(k_2)}$, giving the new potential ϕ'_R for Cl_R. We now take the potential $\phi_{R\backslash L}$ as the potential that is obtained from ϕ_R by summing up the probabilities of all configurations in Ω_R that are similar except for the state of the variable k. So, we actually marginalize out variable k. We will now show that if the following is true:

- $\mathbf{y} > \mathbf{x}$, and
- $\frac{\phi'_{R\backslash L}(b_1,c_1)}{\phi_{R\backslash L}(b_1,c_1)} > \frac{\phi'_{R\backslash L}(\mathbf{c}')}{\phi_{R\backslash L}(\mathbf{c}')} \quad \forall \mathbf{c}' \in \Omega_{\{b,c\}} \backslash \{b_1 c_1\}$,

then the configuration of Y will persist after the evidence change. In order to proof this property, a lemma is introduced first.

Lemma 3. Let the situation be as described above. Then, $\frac{\phi_R(\{k_1\},\mathbf{c}_{R_1})}{\phi_R(\{k_2\},\mathbf{c}_{R_1})} < \frac{\phi_R(\{k_1\},\mathbf{c}_{R_2})}{\phi_R(\{k_2\},\mathbf{c}_{R_2})}$, $\forall \mathbf{c}_{R_1}, \mathbf{c}_{R_2} \in \Omega_{R\backslash L}$, if the following conditions hold:

- $\frac{\phi'_{R\backslash L}(\mathbf{c}_{R_1})}{\phi_{R\backslash L}(\mathbf{c}_{R_1})} > \frac{\phi'_{R\backslash L}(\mathbf{c}_{R_2})}{\phi_{R\backslash L}(\mathbf{c}_{R_2})}$
- $\mathbf{y} > \mathbf{x}$

Proof. Suppose that the two conditions mentioned above hold. Since $\frac{\phi'_{R\backslash L}(\mathbf{c}_{R_1})}{\phi_{R\backslash L}(\mathbf{c}_{R_1})} > \frac{\phi'_{R\backslash L}(\mathbf{c}_{R_2})}{\phi_{R\backslash L}(\mathbf{c}_{R_2})}$, there exists a $z \in \mathbb{R}$ such that $\frac{\phi'_{R\backslash L}(\mathbf{c}_{R_1})}{\phi_{R\backslash L}(\mathbf{c}_{R_1})} > z$ and $\frac{\phi'_{R\backslash L}(\mathbf{c}_{R_2})}{\phi_{R\backslash L}(\mathbf{c}_{R_2})} < z$. We first consider $\frac{\phi'_{R\backslash L}(\mathbf{c}_{R_1})}{\phi_{R\backslash L}(\mathbf{c}_{R_1})} > z$, which can be rewritten as follows:

$$\begin{aligned}
&\frac{\phi'_{R\setminus L}(\mathbf{c}_{R_1})}{\phi_{R\setminus L}(\mathbf{c}_{R_1})} > z\\
\Leftrightarrow\ &\frac{\phi'_R(\{k_1\},\mathbf{c}_{R_1}) + \phi'_R(\{k_2\},\mathbf{c}_{R_1})}{\phi_R(\{k_1\},\mathbf{c}_{R_1}) + \phi_R(\{k_2\},\mathbf{c}_{R_1})} > z\\
\Leftrightarrow\ &\frac{\mathbf{x}\cdot\phi_R(\{k_1\},\mathbf{c}_{R_1}) + \mathbf{y}\cdot\phi_R(\{k_2\},\mathbf{c}_{R_1})}{\phi_R(\{k_1\},\mathbf{c}_{R_1}) + \phi_R(\{k_2\},\mathbf{c}_{R_1})} > z\\
\Leftrightarrow\ &\mathbf{x}\cdot\phi_R(\{k_1\},\mathbf{c}_{R_1}) + \mathbf{y}\cdot\phi_R(\{k_2\},\mathbf{c}_{R_1}) > z\cdot\phi_R(\{k_1\},\mathbf{c}_{R_1}) + z\cdot\phi_R(\{k_2\},\mathbf{c}_{R_1})\\
\Leftrightarrow\ &(\mathbf{y}-z)\cdot\phi_R(\{k_2\},\mathbf{c}_{R_1}) > (z-\mathbf{x})\cdot\phi_R(\{k_1\},\mathbf{c}_{R_1})
\end{aligned}$$

$\frac{\phi'_{R\setminus L}(\mathbf{c}_R)}{\phi_{R\setminus L}(\mathbf{c}_R)} < z$ can be rewritten in the same way:

$$\begin{aligned}
&\frac{\phi'_{R\setminus L}(\mathbf{c}_{R_2})}{\phi_{R\setminus L}(\mathbf{c}_{R_2})} < z\\
\Leftrightarrow\ &\frac{\phi'_R(\{k_1\},\mathbf{c}_{R_2}) + \phi'_R(\{k_2\},\mathbf{c}_{R_2})}{\phi_R(\{k_1\},\mathbf{c}_{R_2}) + \phi_R(\{k_2\},\mathbf{c}_{R_2})} < z\\
\Leftrightarrow\ &\frac{\mathbf{x}\cdot\phi_R(\{k_1\},\mathbf{c}_{R_2}) + \mathbf{y}\cdot\phi_R(\{k_2\},\mathbf{c}_{R_2})}{\phi_R(\{k_1\},\mathbf{c}_{R_2}) + \phi_R(\{k_2\},\mathbf{c}_{R_2})} < z\\
\Leftrightarrow\ &\mathbf{x}\cdot\phi_R(\{k_1\},\mathbf{c}_{R_2}) + \mathbf{y}\cdot\phi_R(\{k_2\},\mathbf{c}_{R_2}) < z\cdot\phi_R(\{k_1\},\mathbf{c}_{R_2}) + z\cdot\phi_R(\{k_2\},\mathbf{c}_{R_2})\\
\Leftrightarrow\ &(\mathbf{y}-z)\cdot\phi_R(\{k_2\},\mathbf{c}_{R_2}) < (z-\mathbf{x})\cdot\phi_R(\{k_1\},\mathbf{c}_{R_2})
\end{aligned}$$

We know that the probabilities in the potentials are positive. Therefore, $\frac{\phi'_{R\setminus L}(\mathbf{c}_{R_1})}{\phi_{R\setminus L}(\mathbf{c}_{R_1})} > z$ and $\frac{\phi'_{R\setminus L}(\mathbf{c}_{R_2})}{\phi_{R\setminus L}(\mathbf{c}_{R_2})} < z$ can only be both true if $\mathbf{y}-z$ and $z-\mathbf{x}$ are both positive or both negative. The second condition requires $\mathbf{y}$ to be larger than $\mathbf{x}$. This makes it impossible for $\mathbf{y}-z$ and $z-\mathbf{x}$ to be negative. The only possibility therefore is that both terms are positive. We can thus continue our rewriting as follows:

$$\begin{aligned}
&(\mathbf{y}-z)\cdot\phi_R(\{k_2\},\mathbf{c}_{R_1}) > (z-\mathbf{x})\cdot\phi_R(\{k_1\},\mathbf{c}_{R_1})\\
\Leftrightarrow\ &\frac{\mathbf{y}-z}{z-\mathbf{x}} > \frac{\phi_R(\{k_1\},\mathbf{c}_{R_1})}{\phi_R(\{k_2\},\mathbf{c}_{R_1})}
\end{aligned}$$

and

$$\begin{aligned}
&(\mathbf{y}-z)\cdot\phi_R(\{k_2\},\mathbf{c}_{R_2}) < (z-\mathbf{x})\cdot\phi_R(\{k_1\},\mathbf{c}_{R_2})\\
\Leftrightarrow\ &\frac{\mathbf{y}-z}{z-\mathbf{x}} < \frac{\phi_R(\{k_1\},\mathbf{c}_{R_2})}{\phi_R(\{k_2\},\mathbf{c}_{R_2})}
\end{aligned}$$

Combining the two results gives us:

$$\frac{\phi_R(\{k_1\},\mathbf{c}_{R_1})}{\phi_R(\{k_2\},\mathbf{c}_{R_1})} < \frac{\phi_R(\{k_1\},\mathbf{c}_{R_2})}{\phi_R(\{k_2\},\mathbf{c}_{R_2})}$$

Since given the two conditions, the configurations $\mathbf{c}_{R_1}$ and $\mathbf{c}_{R_2}$ were chosen arbitrarily, the results hold for all $\mathbf{c}_{R_1}, \mathbf{c}_{R_2} \in \Omega_{R\setminus L}$ □

We illustrate the lemma with an example.

Example 2. Take the junction tree in Figure 14 and suppose there is a change in evidence in X and we propagate this change through separator Se_L to clique Cl_R, resulting in the junction tree presented in Figure 15.

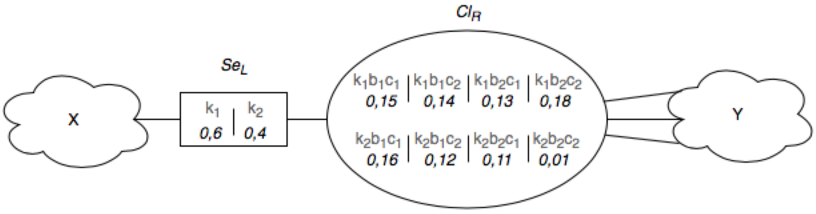


Figure 14: Potentials of cliques and separators in the junction tree before a change in evidence in X

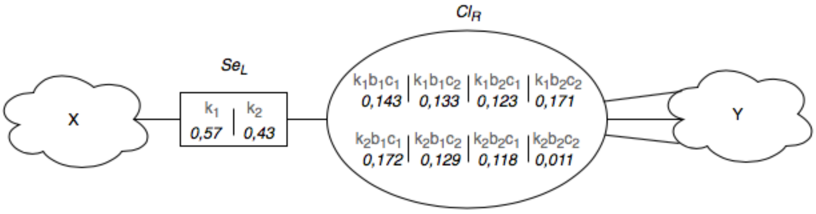


Figure 15: Potentials of cliques and separators in the junction tree after a change in evidence in X. The values in ϕ_R are rounded

As we can see, $\mathbf{y} = \frac{\phi_{L'}(\mathbf{c}_{k_2})}{\phi_L(\mathbf{c}_{k_2})} = \frac{0{,}43}{0{,}40} = 1,075$ and $\mathbf{x} = \frac{\phi_{L'}(\mathbf{c}_{k_1})}{\phi_L(\mathbf{c}_{k_1})} = \frac{0{,}57}{0{,}60} = 0,95$. So, the increase in the probability for $k = k_2$ is higher than the increase in the probability for $k = k_1$. The increase of the probability for the configuration b_1c_1 is equal to $\frac{0{,}143+0{,}172}{0{,}150+0{,}160} \approx 1,016$. This factor is larger than the increase in the probability for the configurations b_1c_2, b_2c_1 and b_2c_2 which are respectively $1,008$, $1,004$ and $0,958$. We can therefore conclude that before the change in evidence, the ratio between the probabilities of $k_1b_1c_1$ and $k_2b_1c_1$ is smaller than the ratios of the probabilities the other three pairs of configurations. And indeed, $\frac{\phi_R(k_1b_1c_1)}{\phi_R(k_2b_1c_1)} = \frac{0{,}150}{0{,}160} \approx 0,938$ while $\frac{\phi_R(k_1b_1c_2)}{\phi_R(k_2b_1c_2)} \approx 1,167$, $\frac{\phi_R(k_1b_2c_1)}{\phi_R(k_2b_2c_1)} \approx 1,182$ and $\frac{\phi_R(k_1b_2c_2)}{\phi_R(k_2b_2c_2)} = 18$. ☐

With this lemma, we can now proceed and formulate the theorem about persistence.

Theorem 4. Let the situation be as described above. Suppose $\mathbf{c}_{MPE}$ is the MPE before the change in evidence and $\mathbf{c}'_{MPE}$ is the MPE after the change in evidence. Then $\mathbf{c}_{Y'\setminus\{k\}}^{\downarrow MPE} = \mathbf{c}'^{\downarrow MPE}_{Y'\setminus\{k\}}$, if the following holds:

- $\frac{\phi'_{R\setminus L}(\mathbf{c}_{R_1})}{\phi_{R\setminus L}(\mathbf{c}_{R_1})} > \frac{\phi'_{R\setminus L}(\mathbf{c}_{R_2})}{\phi_{R\setminus L}(\mathbf{c}_{R_2})}$
- $\mathbf{y} > \mathbf{x}$

$\forall \mathbf{c}_{R_2} \in \Omega_{R\setminus L} \setminus \{\mathbf{c}_{R_1}\}$ where $\mathbf{c}_{R_1} = \mathbf{c}_{R\setminus L}^{\downarrow MPE}$.

Proof. We will prove this by contradiction. Let us indicate the variable set of the complete junction tree with V. We will represent the joint probability distribution as follows:

$$\Pr(\mathbf{c}_V) = \phi_X(\mathbf{c}_X^{\downarrow V}) \frac{\phi_R(\mathbf{c}_R^{\downarrow V})}{\phi_L(\mathbf{c}_L^{\downarrow V})} \phi_Y(\mathbf{c}_Y^{\downarrow V})$$

Next, the configuration space Ω_V can be partitioned into two groups:

- $\Omega_1 = \{\mathbf{c}_V | \mathbf{c}_L^{\downarrow V} = \{k_1\}\}$
- $\Omega_2 = \{\mathbf{c}_V | \mathbf{c}_L^{\downarrow V} = \{k_2\}\}$

Before the change in evidence, we assume that the $\mathbf{c}_{MPE}$ is in Ω_1. Let us pick another configuration $\mathbf{c}'_V$ from Ω_2. We will show that this configuration can not become the new MPE under the conditions mentioned in the theorem unless $\mathbf{c}'^{\downarrow V}_R \setminus \{k_2\} = \mathbf{c}_R^{\downarrow MPE} \setminus \{k_1\}$.

To contradict this, suppose that $\mathbf{c}'^{\downarrow V}_R \setminus \{k_2\} \neq \mathbf{c}_R^{\downarrow MPE} \setminus \{k_1\}$. Then, we can define two other configurations $\hat{\mathbf{c}}_V$ and $\hat{\mathbf{c}}'_V$ so that:

- $\hat{\mathbf{c}}_V$ is the configuration with $\hat{\mathbf{c}}'^{\downarrow V}_X = \mathbf{c}'^{\downarrow V}_X$ and $\mathbf{c}_{Y'}^{\downarrow V} \setminus \{k_2\} = \mathbf{c}_{Y'}^{\downarrow MPE} \setminus \{k_1\}$.
- $\hat{\mathbf{c}}'_V$ is the configuration with $\hat{\mathbf{c}}'^{\downarrow V}_X = \mathbf{c}_X^{\downarrow MPE}$ and $\hat{\mathbf{c}}'^{\downarrow V}_{Y'} \setminus \{k_1\} = \mathbf{c}'^{\downarrow V}_{Y'} \setminus \{k_2\}$.

Figure 16 depicts the situation for these four configurations. We have that both $\mathbf{c}_R^{\downarrow MPE}$ and $\hat{\mathbf{c}}_R^{\downarrow V}$ as well as $\hat{\mathbf{c}}'^{\downarrow V}_R$ and $\mathbf{c}'^{\downarrow V}_R$ marginalize out variable k. According to the conditions mentioned in the theorem, we know that:

$$\frac{\phi'_R(\mathbf{c}_R^{\downarrow MPE}) + \phi'_R(\hat{\mathbf{c}}_R^{\downarrow V})}{\phi_R(\mathbf{c}_R^{\downarrow MPE}) + \phi_R(\hat{\mathbf{c}}_R^{\downarrow V})} > \frac{\phi'_R(\mathbf{c}'^{\downarrow V}_R) + \phi'_R(\hat{\mathbf{c}}'^{\downarrow V}_R)}{\phi_R(\mathbf{c}'^{\downarrow V}_R) + \phi_R(\hat{\mathbf{c}}'^{\downarrow V}_R)}$$

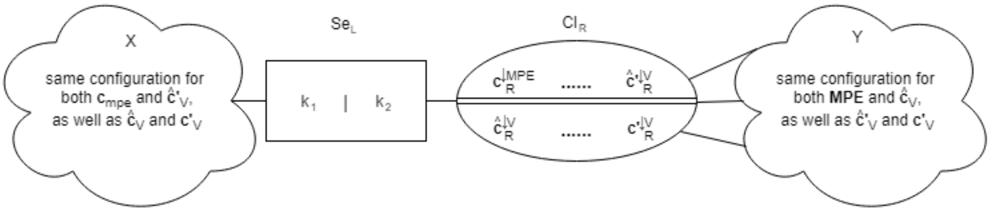


Figure 16: Situation of the four configurations in clique Cl_R. k_1 is part of $\mathbf{c}_{MPE}$ and $\hat{\mathbf{c}}'_V$. k_2 is part of $\hat{\mathbf{c}}_V$ and $\mathbf{c}'_V$. $\mathbf{c}_{MPE}$ and $\hat{\mathbf{c}}_V$ have the same configuration for the variables in Y. $\hat{\mathbf{c}}'_V$ and $\mathbf{c}'_V$ have the same configuration for the variables in Y.

Therefore, we can use lemma 3:

$$\frac{\phi_R(\mathbf{c}_R^{\downarrow MPE})}{\phi_R(\hat{\mathbf{c}}_R^{\downarrow V})} < \frac{\phi_R(\hat{\mathbf{c}}'^{\downarrow V}_R)}{\phi_R(\mathbf{c}'^{\downarrow V}_R)} \Leftrightarrow \frac{\phi_R(\mathbf{c}_R^{\downarrow MPE})\phi_R(\mathbf{c}'^{\downarrow V}_R)}{\phi_R(\hat{\mathbf{c}}'^{\downarrow V}_R)} < \phi_R(\hat{\mathbf{c}}_R^{\downarrow V})$$

We will now use this information to show that $\Pr(\mathbf{c}'_V)$ cannot be larger than $\Pr(\hat{\mathbf{c}}_V)$. Because both configurations have the same states for the variables in X and the change in evidence is in X as well, the probabilities of both configurations will be adjusted with the same factor. Therefore, the ratio between $\Pr(\hat{\mathbf{c}}_V)$ and $\Pr(\mathbf{c}'_V)$ after updating the junction tree does not change. This means that $\mathbf{c}'_V$ can never become the new MPE. Let us proof this by contradiction:

$$\begin{aligned}
& \Pr(\mathbf{c}'_V) > \Pr(\hat{\mathbf{c}}_V) \\
\Leftrightarrow\ & \phi_X(\mathbf{c}'^{\downarrow V}_X)\frac{\phi_R(\mathbf{c}'^{\downarrow V}_R)}{\phi_L(k_2)}\phi_Y(\mathbf{c}'^{\downarrow V}_Y) > \phi_X(\hat{\mathbf{c}}^{\downarrow V}_X)\frac{\phi_R(\hat{\mathbf{c}}^{\downarrow V}_R)}{\phi_L(k_2)}\phi_Y(\hat{\mathbf{c}}^{\downarrow V}_Y) \\
\Leftrightarrow\ & \phi_X(\hat{\mathbf{c}}^{\downarrow V}_X)\frac{\phi_R(\mathbf{c}'^{\downarrow V}_R)}{\phi_L(k_2)}\phi_Y(\mathbf{c}'^{\downarrow V}_Y) > \phi_X(\hat{\mathbf{c}}^{\downarrow V}_X)\frac{\phi_R(\hat{\mathbf{c}}^{\downarrow V}_R)}{\phi_L(k_2)}\phi_Y(\hat{\mathbf{c}}^{\downarrow V}_Y) \\
\Leftrightarrow\ & \phi_R(\mathbf{c}'^{\downarrow V}_R)\phi_Y(\mathbf{c}'^{\downarrow V}_Y) > \phi_R(\hat{\mathbf{c}}^{\downarrow V}_R)\phi_Y(\hat{\mathbf{c}}^{\downarrow V}_Y) \\
\Rightarrow\ & \phi_R(\mathbf{c}'^{\downarrow V}_R)\phi_Y(\mathbf{c}'^{\downarrow V}_Y) > \frac{\phi_R(\mathbf{c}_R^{\downarrow MPE})\phi_R(\mathbf{c}'^{\downarrow V}_R)}{\phi_R(\hat{\mathbf{c}}'^{\downarrow V}_R)}\phi_Y(\hat{\mathbf{c}}^{\downarrow V}_Y) \quad (\textit{lemma } 3) \\
\Leftrightarrow\ & \phi_R(\hat{\mathbf{c}}'^{\downarrow V}_R)\phi_Y(\mathbf{c}'^{\downarrow V}_Y) > \phi_R(\mathbf{c}_R^{\downarrow MPE})\phi_Y(\hat{\mathbf{c}}^{\downarrow V}_Y) \\
\Leftrightarrow\ & \phi_R(\hat{\mathbf{c}}'^{\downarrow V}_R)\phi_Y(\hat{\mathbf{c}}'^{\downarrow V}_Y) > \phi_R(\mathbf{c}_R^{\downarrow MPE})\phi_Y(\mathbf{c}_Y^{\downarrow MPE}) \\
\Leftrightarrow\ & \frac{\phi_R(\hat{\mathbf{c}}'^{\downarrow V}_R)}{\phi_L(k_1)}\phi_Y(\hat{\mathbf{c}}'^{\downarrow V}_Y) > \frac{\phi_R(\mathbf{c}_R^{\downarrow MPE})}{\phi_L(k_1)}\phi_Y(\mathbf{c}_Y^{\downarrow MPE}) \\
\Leftrightarrow\ & \phi_X(\hat{\mathbf{c}}'^{\downarrow V}_X)\frac{\phi_R(\hat{\mathbf{c}}'^{\downarrow V}_R)}{\phi_L(k_1)}\phi_Y(\hat{\mathbf{c}}'^{\downarrow V}_Y) > \phi_X(\hat{\mathbf{c}}'^{\downarrow V}_X)\frac{\phi_R(\mathbf{c}_R^{\downarrow MPE})}{\phi_L(k_1)}\phi_Y(\mathbf{c}_Y^{\downarrow MPE}) \\
\Leftrightarrow\ & \phi_X(\hat{\mathbf{c}}'^{\downarrow V}_X)\frac{\phi_R(\hat{\mathbf{c}}'^{\downarrow V}_R)}{\phi_L(k_1)}\phi_Y(\hat{\mathbf{c}}'^{\downarrow V}_Y) > \phi_X(\mathbf{c}_X^{\downarrow MPE})\frac{\phi_R(\mathbf{c}_R^{\downarrow MPE})}{\phi_L(k_1)}\phi_Y(\mathbf{c}_Y^{\downarrow MPE}) \\
\Leftrightarrow\ & \Pr(\hat{\mathbf{c}}'_V) > \Pr(\mathbf{c}_{MPE})
\end{aligned}$$

This is a contradiction since we supposed that $\mathbf{c}_{MPE}$ is the MPE. The only configuration to which $\mathbf{c}_{MPE}$ can change is $\hat{\mathbf{c}}_V$. This means that variable k may be the only variable in Y' for which the configuration changes. ☐

6.3 The assumptions

The situation that is used in our theorem is based on two assumptions. The first assumption is that the variable k should not occur in any other adjacent separator of Cl_R. In order to see what happens if the variable k is present in a separator located in Y, we divide the set of variables $Y'\backslash\{k\}$ into two parts: the variables in clique Cl_R: $R\backslash\{k\}$ and the variables not in Cl_R :$Y'\backslash R$. If the variable k is present in another separator, we still know that the MPE configuration of the variables in $R\backslash\{k\}$ will not change. For the variables in $Y'\backslash R$ however, this may not be the case. Because of the assumption that the variable k is not present in any of the other separators, we know that the MPE configuration of the variables in the separator will not change and that the MPE configuration of the variables in the junction tree after this separator will not change either. Removing the assumption means that we have to continue propagation for the separators that contain the variable k.

The second assumption is that the separator contains a single variable with a state space of size two. If separators contain a variable with a larger state space or more than one variable, the theorem holds. Since our lemma only works for two elements, we need to compare the MPE configuration in the separator with every other configuration in the separator separately. Suppose for example that the separator Se_L has more than two configurations. We pick the configuration that is part of the MPE, $\mathbf{c}_L^{\downarrow MPE}$, and compare it to every other configuration $\mathbf{c}'_L \in \Omega_L \backslash \{\mathbf{c}_L^{\downarrow MPE}\}$. For each such $\mathbf{c}'_L$, we consider only the configurations in $\Omega'_R = \{\mathbf{c}_R | (\mathbf{c}'_L = \mathbf{c}_L^{\downarrow R} \vee \mathbf{c}_L^{\downarrow MPE} = \mathbf{c}_L^{\downarrow R}) \wedge \mathbf{c}_R \in \Omega_R\}$. We can then use the theorem on configurations $\mathbf{c}_L^{\downarrow MPE}$, $\mathbf{c}'_L$ and the configuration space Ω'_R. If the theorem holds, we know that the MPE will not change to any of the configurations in Ω'_R that have other states for the variables $R \backslash L$. If the theorem holds for every configuration in $\Omega_L \backslash \{\mathbf{c}_L^{\downarrow MPE}\}$, we know that the variables in $R \backslash \{k\}$ will persist.

The advantage of our theorem is that we do not have to propagate further when we find persistence. The complexity of the procedure to check the persistence in a clique moreover is linear in the number of configurations in the clique. Another advantage is that the theorem does not require the MPE-specific max-propagation but only the standard propagation. There are however some disadvantages as well. We do not know whether and how the variables change state when persistence is not proved by this theorem. Furthermore, the theorem currently requires the junction tree to be fully propagated. Therefore, this theorem cannot be used when consecutive changes in the evidence are introduced unless full propagation between these changes is performed. Lastly, we haven't yet been able to find any situation that has persistence according to this theorem. Therefore, the theorem is likely not very useful in practice in our experience.

7 MPE Experiments

In this section, we will analyze the effects of changes in evidence on the MPE by performing multiple experiments on different junction trees. After a change in evidence is propagated through the junction tree, the MPE may change. Therefore, the configurations of some cliques that were part of the MPE, may change, while the MPE configurations of other cliques persist. In the experiments, we will look at how probability changes in cliques have effect on whether the configuration of the clique persists into the new MPE or not. We will also study the relation between the monotone decrease property and the persistence of configurations.

7.1 Set-up

Before we go to the results of the experiments, we will first describe the set-up we have used. We will discuss the Bayesian networks that were used, the way we performed runs in the experiments and finally define some concepts that we use in the experiments.

7.1.1 Bayesian networks

The experiments are performed on the junction trees of three well-known Bayesian networks: *asia*, *alarm* and *hepar2*.

asia
asia [18] is a small artificial network with eight binary variables. Its junction tree contains six cliques. The paths along which evidence will be propagated are short due to the small number of cliques. Because of this, the effects of changes in the potentials can be considered in more detail.

alarm
alarm [19] is a medium-sized network with 37 variables. The state spaces of the variables vary between two and four states. Its junction tree has 27 cliques. The tree includes long parts along which evidence will be propagated.

hepar2
Hepar2 [20] is a real-world network containing 70 variables with two to four states each. Its junction tree contains 58 cliques. The large number of cliques allows a large variety of evidence scenarios to be tested.

7.1.2 Runs

Each experiment consists of multiple runs and each run includes a set of scenarios. A scenario involves, for a given Bayesian network, two instances of the junction tree: T and T'. T' differs from T in that the evidence entered in T' is different from the evidence in T at one point. The difference can be in a variable that was not observed in T, a variable that is observed in T but changed state, or a variable that is removed from the evidence in T. We call the evidence entered in T the *pre-change evidence* and the evidence in T' the *post-change evidence*. The variables that will be part of the evidence are selected at random as are their states. Duplicate scenarios in a single run are not allowed. For the *alarm* and *hepar2* networks, a run involves a hundred scenarios. This number ensures that multiple types of scenarios are covered (e.g. a change of evidence in a leaf clique, a change in a clique of more than five variables). For *asia*, it was not possible in some cases to create a 100 distinct scenarios. For these cases, only 10 scenarios are used. We mark these runs with an asterisk in the tables.

Various parameters are set for a run:

- The number of variables involved in the *pre-change evidence*. More variables means that fewer variables remain for the MPE and therefore a smaller joint state space for the MPE. More variables in the evidence also can make specific variables independent due to blocking. But they can also cause independencies to be removed. 0, 1, 2 and 4 variables in the *pre-change evidence* are used for the *asia*-network, 0, 1, 2, 5 and 9 for the *alarm*-network, and 0, 1, 2, 5, 9, 15, 25 and 30 for the *hepar2*-network.
- Whether the change of evidence concerns a variable in the *pre-change evidence*, a variable that had no evidence before or a variable that is removed from the *pre-change evidence*. This is the difference between the three types of change in the evidence we stated before. The reason to make this distinction is as follows. When a variable is added to the evidence, we could say that its states changes from an 'optimal' state for the MPE to a random state. For the second type however, we just swap between two random states. And for the third type, the state goes from random to an optimal state. These three types may have different impacts on how the MPE changes.

For every setting of the parameters, ten runs are generated. This gives us an indication of the consistency of the results.

7.1.3 Characterization of the affected clique set

In the experiments, we will consider the part of the junction tree in which a change of the MPE to MPE′ occurred. The clique in which the change in evidence is introduced is indicated with Cl_{So} and we take $\mathbf{Cl}_{Affected}$ to be the set of all cliques in which the configuration of at least one variable changed in the MPE. Recall from section 3.6 that thea affected cliques form a contiguous part of the junction tree. We define the border of $\mathbf{Cl}_{Affected}$ to be the set $\mathbf{Cl}_{Border} \subseteq \mathbf{Cl}_{Affected}$ such that for each clique $Cl_B \in \mathbf{Cl}_{Border}$ one of the two properties hold:

- $L \subset B \wedge \mathbf{c}_L^{\downarrow MPE} = \mathbf{c}_L^{\downarrow MPE'} \quad \exists Se_L \in \mathbf{Se}$
- $\#\{Se_L | Se_L \in \mathbf{Se} \wedge L \subset B\} = 1 \wedge \mathbf{c}_B^{\downarrow MPE} \neq \mathbf{c}_B^{\downarrow MPE'}$

The border of $\mathbf{Cl}_{Affected}$ thus includes the set of cliques that have either a separator that is not in $\mathbf{Cl}_{Affected}$ or have a single separator. A clique in $\mathbf{Cl}_{Border}$ will be called a *border clique*.

Note that multiple configurations can be the most probable configuration in both T and T'. For T, an arbitrary configuration will be picked as MPE. For T', we have to be more careful. Suppose there are two MPEs $\mathbf{c}_1$ and $\mathbf{c}_2$ in T and that some variables have the same state in both MPEs. Next, a change in one of these variables occurs and only the common part of the two MPEs is affected. The two MPEs will change in the same way giving us two new MPEs $\mathbf{c'}_1$ and $\mathbf{c'}_2$. If we pick $\mathbf{c'}_2$ for T' while $\mathbf{c}_1$ was picked for T, then $\mathbf{Cl}_{Affected}$ may not be contiguous. We will therefore pick an MPE for T' that makes sure $\mathbf{Cl}_{Affected}$ is contiguous. During the experiments, we encountered only a small number of scenarios in which there was more than one MPE.

7.1.4 Distance

We will study the distance between a clique and the source of change in evidence in our analysis. This distance is measured as follows. Suppose we want the distance between Cl_R and the source of change in evidence. We will not simply pick Cl_{So} as the source. If multiple cliques contain the changed variable, it would not matter in which clique the evidence is introduced and therefore, the distance should not matter as well. We will therefore pick the clique that is closest to Cl_R. However, we cannot just take the length of the path between these two cliques in the junction as the distance. Take for example a junction tree containing three cliques Cl_R, Cl_S and Cl_T. Cl_S is connected to both Cl_R and Cl_T with a separator. If both separators have the same variables, we can alter the junction tree by removing the connection between Cl_R and Cl_S and add a connection between Cl_S and Cl_T which will yield a new correct junction tree. While the distance between Cl_R and Cl_T is two in the first junction tree, it is one in the other. We will pick the smallest possible distance for our distance measure.

7.2 Results: Affected clique set

The analysis of the results in this section will focus on two issues. We examine the situations at the border of the affected clique set and we will look into the size of the set of affected cliques.

7.2.1 The border

At the border, we distinguish between three cases. The first case concerns cliques in $\mathbf{Cl}_{Border}$ that constitute a leaf at the end of a branch in the junction tree. These *border cliques* do not have any separator with a clique outside $\mathbf{Cl}_{Affected}$. The other *border cliques* have one or more separators that are connected to another clique outside $\mathbf{Cl}_{Affected}$. We will consider each separator separately. Each such separator is either one of which the factors are all equal to one - the *second* case - or not - the *third* case. One clique may contain multiple separators of different cases. It is interesting to distinguish between the cases two and three. For case two, the probabilities in that clique and in the subsequent cliques are all algebraically independent of the change in Cl_{So}. Therefore, persistence of the MPE configuration of these cliques is likely. Changes in the MPE configuration can only be caused because the combination with configurations in previous cliques is better due to changes in probabilities of these previous cliques. This is not the case for type three cases in which the probabilities are still affected by the change of evidence.
The results for these three cases for the different networks are shown in Tables 1-9. The occurrences of each case is stated as a fraction of the total number of case occurences. It is the average over ten runs and the standard deviation is stated besides the average and gives an indication on how close the results of the separate runs are to each other.

Let us first analyze the results for the *asia*-network. The tables show common trends. The results for type one cases are similar across all three tables. Table 2 and 3 have similar results for the type two and three cases as well. The type two and three cases follow the same trend as the *pre-change evidence* increases, but the fractions in Table 1 seem to lay one step behind. The reason that Table 1 seems to lay one step behind Table 3 can be explained by the fact that the two types of evidence change are each others' opposite. Therefore, we should compare the numbers of the *pre-change evidence* from Table 1 with the number of the *post-change evidence* in Table 3. Table 1 and 2, should be compared by looking at the *post-change evidence* instead of the *pre-change evidence*.

Comparing the results for the same *post-change evidence* gives similar results. An explanation for this finding is that the number of variables in the *post-change evidence* indicate how likely paths are blocked or how many independencies have been removed by the evidence in the network. These removed or added independencies are important for the proportion of type two and type three cases.

In all three tables, about 50% of the cases are of type one. Three of the six cliques have only one adjacent separator, that is, they are at the end of a branch. The probability that an affected clique is one of these cliques is therefore high. Furthermore, the relatively high number of cliques with a single separator also indicates that the paths in the junction tree are short which increases the probability that changes reach the end of a branch. As the number of variables involved in the *pre-change evidence* increases, the fraction of type two cases increases, while the fraction of type three cases decreases. As a consequence of the small size of the network, entering evidence is likely to block a path, which increases the number of independencies between variables and therefore increases the fraction of type two cases while decreasing the fraction of type three cases.

pre-change evidence	post-change evidence	(1) End of branch	(2) Independence	(3) Change
0*	1	.505 ± .054	.029 ± .037	.466 ± .065
1	2	.523 ± .040	.193 ± .036	.284 ± .031
2	3	.526 ± .045	.245 ± .045	.229 ± .034
4	5	.524 ± .040	.388 ± .064	.089 ± .045

Table 1: Results for scenarios in the *asia*-network in which a new variable is introduced in the evidence. The fractions show how often each type of case occurs at the border of the set of affected cliques.

pre-change evidence	post-change evidence	(1) End of branch	(2) Independence	(3) Change
1*	1	.543 ± .071	.066 ± .020	.391 ± .076
2	2	.530 ± .011	.169 ± .032	.301 ± .033
4	4	.520 ± .024	.322 ± .048	.159 ± .035

Table 2: Results for scenarios in the *asia*-network in which the state of a variable in the evidence is changed. The fractions show how often each type of case occurs at the border of the set of affected cliques.

pre-change evidence	post-change evidence	(1) End of branch	(2) Independence	(3) Change
1*	0	.555 ± .055	.053 ± .037	.392 ± .067
2	1	.537 ± .035	.151 ± .019	.312 ± .042
4	3	.493 ± .040	.325 ± .057	.182 ± .041

Table 3: Results for scenarios in the *asia*-network in which a variable is removed from the evidence. The fractions show how often each type of case occurs at the border of the set of affected cliques.

Next, let us look at the results for *alarm* in Tables 4-6. The fraction of type one cases is smaller than for *asia*. In the junction tree of *alarm*, the branches are longer and only 9 of the 27 cliques have a single separator. The fraction of type two cases starts being higher in comparison with *asia*. However, the fraction does not change significantly when the number of variables in the *pre-change evidence* increases and becomes smaller than the fraction for *asia*. The unchanging fraction may be due to multiple issues. First of all, there are often multiple paths between two variables in the Bayesian network. The variables will only be independent when every path is blocked. Moreover, it can be that the number of new independencies is counterbalanced by the number of times an independency is removed. When evidence is introduced in one of the descendants of two independent variables, the two variables will not be independent anymore. In the *alarm*-network, there are multiple variables that are independent of each other and have a high number of common descendants, which makes introducing evidence quite likely to remove an independency.

pre-change evidence	post-change evidence	(1) End of branch	(2) Independence	(3) Change
0	1	.385 ± .005	.239 ± .005	.376 ± .004
1	2	.387 ± .025	.248 ± .030	.365 ± .039
2	3	.385 ± .034	.201 ± .027	.415 ± .038
5	6	.343 ± .031	.197 ± .038	.461 ± .040
9	10	.334 ± .032	.212 ± .046	.455 ± .044

Table 4: Results for scenarios in the *alarm*-network in which a new variable is introduced in the evidence. The fractions show how often each type of case occurs at the border of the set of affected cliques.

pre-change evidence	post-change evidence	(1) End of branch	(2) Independence	(3) Change
1	1	.387 ± .014	.239 ± .021	.374 ± .017
2	2	.393 ± .018	.239 ± .029	.368 ± .028
5	5	.369 ± .020	.195 ± .025	.435 ± .033
9	9	.348 ± .024	.221 ± .028	.431 ± .024

Table 5: Results for scenarios in the *alarm*-network in which the state of a variable in the evidence is changed. The fractions show how often each type of case occurs at the border of the set of affected cliques.

pre-change evidence	post-change evidence	(1) End of branch	(2) Independence	(3) Change
1	0	.388 ± .006	.236 ± .007	.376 ± .007
2	1	.381 ± .026	.231 ± .031	.388 ± .031
5	4	.369 ± .021	.190 ± .029	.441 ± .035
9	8	.351 ± .029	.195 ± .040	.455 ± .033

Table 6: Results for scenarios in the *alarm*-network in which a variable is removed from the evidence. The fractions show how often each type of case occurs at the border of the set of affected cliques.

A last finding to note are the different results for type one and type three cases. While their fractions are not significantly different with small numbers of variables in the *pre-change evidence*, they start to be different for five and nine variables. This could be due to the fact that more variables in the evidence means less possibilities for the MPE to change and therefore more persistence.

Lastly, let us look at the results for *hepar2* reported in table 7-9. The fraction of type one cases is smaller than in the other two networks. Since the paths in the junction tree are in general longer, the probability of cliques persisting before the end of the branch increases. It does not change significantly as the number of variables in the *pre-change evidence* increases. For cases two and three,

pre-change evidence	post-change evidence	(1) End of branch	(2) Independence	(3) Change
0	1	.251 ± .013	.131 ± .019	.618 ± .028
1	2	.270 ± .019	.111 ± .019	.619 ± .027
2	4	.269 ± .020	.077 ± .023	.654 ± .036
5	6	.279 ± .018	.047 ± .022	.675 ± .023
9	10	.290 ± .027	.058 ± .016	.652 ± .037
15	16	.283 ± .024	.087 ± .016	.630 ± .034
25	25	.290 ± .032	.161 ± .030	.548 ± .035
30	31	.270 ± .032	.179 ± .040	.551 ± .037

Table 7: Results for scenarios in the *hepar2*-network in which a new variable is introduced in the evidence. The fractions show how often each type of case occurs at the border of the set of affected cliques.

pre-change evidence	post-change evidence	(1) End of branch	(2) Independence	(3) Change
1	1	.258 ± .011	.136 ± .013	.606 ± .019
2	2	.253 ± .013	.098 ± .027	.649 ± .030
5	5	.267 ± .021	.058 ± .015	.675 ± .030
9	9	.272 ± .020	.052 ± .013	.676 ± .025
15	15	.282 ± .010	.082 ± .024	.636 ± .022
25	25	.281 ± .019	.145 ± .028	.574 ± .033
30	30	.289 ± .021	.187 ± .030	.524 ± .041

Table 8: Results for scenarios in the *hepar2*-network in which the state of a variable in the evidence is changed. The fractions show how often each type of case occurs at the border of the set of affected cliques.

pre-change evidence	post-change evidence	(1) End of branch	(2) Independence	(3) Change
1	0	.262 ± .012	.134 ± .017	.604 ± .024
2	1	.256 ± .025	.101 ± .030	.643 ± .045
5	4	.281 ± .023	.042 ± .022	.677 ± .023
9	8	.271 ± .027	.051 ± .014	.677 ± .017
15	14	.296 ± .029	.062 ± .018	.642 ± .032
25	24	.288 ± .040	.149 ± .018	.563 ± .031
30	29	.287 ± .027	.175 ± .023	.538 ± .039

Table 9: Results for scenarios in the *hepar2*-network in which a variable is removed from the evidence. The fractions show how often each type of case occurs at the border of the set of affected cliques.

there are two findings that are interesting. First of all, compared to *alarm*, the fraction of type two cases is in general smaller while the fraction of type three cases is much bigger. One reason could be that there are less independency structures in *hepar2* compared to *alarm*. Another possible reason, that cannot be inferred from these tables, is the number of affected cliques in *hepar2*. We will look more into this in the next section. Another striking finding is the way how the fraction of cases two and three changes as the variables in the *pre-change evidence* increases. As the number of variables in the *pre-change evidence* increases, the fraction of case two decreases first after which it increases. For case three, it is the other way around. While in the case of *alarm*, the number of new independencies appeared to be counterbalanced by the number of removed independencies, it seems that the probability of removing an independency in *hepar2* is more likely as the number of variables in the *pre-change evidence* increases. But at some point, the independencies increases again. This makes sense since an independency can only be removed once. Therefore, the probability to remove independencies decreases while more variables in the evidence increases the chance that paths between variables are blocked and therefore increasing the independencies again.

7.2.2 Number of affected cliques

The findings from the previous section show that a change in evidence does not typically affect the complete junction tree, but it does not provide much insight in the size of $\mathbf{Cl}_{Affected}$. In order to gain insight into the affected part of the junction tree, we consider the number of cliques and the number of variables involved for which the MPE configuration changes, including the variable that changed in the evidence. Besides the number of changed variables, we will also display the maximum number of variables that could have changed its MPE state. When a variable is able to change is defined as follows:

- The variable is located in an affected clique.
- The variable is not located in any non-affected clique; otherwise, the MPE configurations of cliques outside $\mathbf{Cl}_{Affected}$ could be changed as well.
- The variable may not be part of the *pre-change evidence* except for the variable that changed state.

The results are shown in Table 10, 11 and 12. As in the previous section, we compare the results for different numbers of variables in the evidence. We will however omit the cases in which a variable is added to or removed from the evidence and thus only look at the scenarios in which the state of a variable in the evidence changed. This because the mutual differences between these three types were small.

pre-change evidence	Avg. #Cliques	Avg. #Variables	Avg. Max #Variables
1*	$4.01 \pm .23$(67%)	$4.46 \pm .35$(56%)	$4.68 \pm .34$
2	$3.38 \pm .15$(56%)	$2.95 \pm .11$(37%)	$3.43 \pm .16$
4	$2.74 \pm .23$(46%)	$1.82 \pm .11$(23%)	$2.21 \pm .15$

Table 10: Results about the number of affected cliques and changed variables in the *asia*-network for different types of scenarios.

For the *asia*-network, the percentage of affected cliques and variables after a change in evidence is relatively high. Because the junction tree of the *asia*-network is small, all cliques are close to any change in the *pre-change evidence*. We know from the monotone decrease property that the effects of a change in the evidence is strongest at the neighbouring cliques. The finding is consistent with Table 1-3 in the previous section where the number of cases in which the end of a branch was reached is more than 50%.

For the *alarm*-network, the percentage of cliques and variables affected by a change of evidence is much smaller compared to *asia*. Interestingly, the number of affected variables is smaller than the number of cliques. This can be explained by the fact that cliques have overlapping variables. In

pre-change evidence	Avg. #Cliques	Avg. #Variables	Avg. Max #Variables
1	$7.30 \pm .30(27\%)$	$5.52 \pm .25(15\%)$	$8.09 \pm .40$
2	$6.93 \pm .35(26\%)$	$5.33 \pm .35(14\%)$	$7.64 \pm .43$
5	$5.84 \pm .44(22\%)$	$4.37 \pm .29(12\%)$	$6.02 \pm .48$
9	$5.23 \pm .27(19\%)$	$3.64 \pm .20(10\%)$	$4.90 \pm .26$

Table 11: Results about the number of affected cliques and changed variables in the *alarm*-network for different types of scenarios.

fact, the junction tree of *alarm* contains a variable that is contained in seven cliques. If this variable changes its state, a high clique count will result. The number of maximum possible variables puts the number of affected variables more into perspective. We see that around $50\% - 60\%$ of the variables that could have changed, did change their state in the new MPE.

In the *hepar2*-network, the number of cliques and variables that are affected by the change of evidence is surprisingly small. While the size of the network is two times that of *alarm*, the percentage of changed cliques is not higher than 30% of the total number of cliques while the percentage of changed variables is not higher than 20%. We suspect that a reason for the small size of the set of affected variables are the conditional probability tables of the *hepar2*-network. This will be investigated further in the next section. Another difference between the *hepar2*-network and the other two networks is that the numbers of affected cliques and variables do not change significantly when the number of variables in the *pre-change evidence* increases. This may be caused by two reasons. The size of the junction tree makes the distance between a variable in the *pre-change evidence* and an arbitrary clique relatively large. The effect of evidence is therefore often probably small. On top of that, the effect of a change in evidence is already small. Only four to five cliques are affected. Because of this small effect and the size of the network, future changes will have a small effect as well.

pre-change evidence	Avg. #Cliques	Avg. #Variables	Avg. Max #Variables
1	$4.38 \pm .42(8\%)$	$1.68 \pm .08(2\%)$	$4.10 \pm .37$
2	$4.34 \pm .43(7\%)$	$1.66 \pm .14(2\%)$	$4.03 \pm .36$
5	$4.34 \pm .34(7\%)$	$1.71 \pm .09(2\%)$	$3.95 \pm .30$
9	$4.83 \pm .38(8\%)$	$1.78 \pm .10(3\%)$	$4.17 \pm .34$
15	$5.25 \pm .77(9\%)$	$1.85 \pm .18(3\%)$	$4.19 \pm .56$
25	$4.75 \pm .54(8\%)$	$1.69 \pm .11(2\%)$	$3.35 \pm .31$
30	$5.03 \pm .79(9\%)$	$1.72 \pm .17(2\%)$	$3.25 \pm .36$

Table 12: Results about the number of affected cliques and changed variables in the *hepar2*-network for different types of scenarios.

7.2.3 Different conditional probability tables

To examine why the numbers of affected cliques and variables in the *hepar2*-network are small, we generated some new instances of the *hepar2*-network in which we changed the conditional probability tables. This was achieved as follows: For each CPT and for each state of the variable associated with the CPT, we generated probabilities by first sampling values from a normal distribution with a mean of zero, then use a normal cumulative distribution function with a mean of zero and a standard deviation of 0.5 to get a value between zero and one and finally normalize these values so that they sum to one. Different standard deviations for the normal distribution were used to get varying skewnesses of the probability distribution. The skewness we looked at represents the skewness of the Bayesian network and is computed according to a method introduced by *Jitnah* and *Nicholson* [21].

In Table 13, the results for the *hepar2*-networks are shown.

skewness	0.579 (original)	0.486	0.343	0.264	0.172
Avg. Cliques	4.38 ± .42(8%)	24.29 ± 1.26(42%)	18.73 ± .77(32%)	23.92 ± .51(41%)	13.01 ± .83(22%)
Avg. Variables	1.68 ± .08(2%)	17.00 ± .93(24%)	11.40 ± .47(16%)	16.14 ± .33(23%)	7.50 ± .43(11%)
Avg. Max Variables	4.10 ± .37	26.25 ± 1.41	19.25 ± .79	25.15 ± .58	12.84 ± .83

Table 13: Results for different initializations of the CPTs of the *hepar2*-network.

The average numbers of affected cliques and variables are much higher. The number of cliques is for two cases even more than 40% of the total number of cliques in the network. We can conclude from these results that the way the probabilities in a CPT are distributed is important for how evidence changes the MPE. Furthermore, the ratio between the numbers of affected cliques and variables is about the same as we saw in the *alarm*-network. At most a quarter of the variables changes on average. This is still considerably smaller than the maximum number of variables that could change state. Note further that from this table, there seems not to be any relation between the skewness and size of $\mathbf{Cl}_{Affected}$. One reason that skewness does not have a clear relation with the size is that on the one hand, if the skewness is low, the probabilities in potentials will be close to each other. When the change of evidence is propagated through the junction tree, the changes will attenuate fast. On the other hand, because the probabilities in the potentials are close to each other, the probabilities of configurations are close to each other and therefore, not a lot of change is required for a configuration to become the MPE.

For *alarm* en *asia*, we generated two different instances as well. The results are shown in Table 14 and 15.

skewness	0.834 (original)	0.342	0.042
Avg. Cliques	7.30 ± .30(27%)	8.59 ± 0.33(32%)	6.96 ± .35(25%)
Avg. Variables	5.52 ± .25(15%)	5.36 ± .21(14%)	5.28 ± .28(14%)
Avg. Max Variables	8.09 ± .40	9.66 ± 0.36	7.84 ± .45

Table 14: Results for different initializations of the CPTs of the *alarm*-network.

skewness	0.711 (original)	0.351	0.026
Avg. Cliques	4.01 ± .23(67%)	3.04 ± 0.32(51%)	2.80 ± .41(47%)
Avg. Variables	4.46 ± .35(56%)	2.02 ± .18(25%)	2.24 ± .32(28%)
Avg. Max Variables	4.68 ± .34	2.91 ± 0.31	3.05 ± .57

Table 15: Results for different initializations of the CPTs of the *asia*-network.

While the number of affected cliques and variables increase for *hepar2*-network, we did not see this increase for *asia* and *alarm*. For *alarm* the numbers are quite similar. It seems that the conditional probabilities in the original network does not have a special influence on how the MPE changes. For *asia*, it is even the case that the number of variables and cliques decreases. An explanation for this could be the conditional probability table for the variable *either* in the original *asia*-network. This CPT consists only of zeroes and ones. This has a great influence on the propagation of changes.

7.3 Results: Monotone decrease

In this section, the focus will be on the monotone decrease of the change in potentials as a result of the change in evidence. Two things will be addressed: Firstly, we will compare the distance between a clique in $\mathbf{Cl}_{Affected}$ and the source of the change in evidence and the change of probabilities in the potential of that clique. Secondly, we will zoom into the changes of cliques after the border of $\mathbf{Cl}_{Affected}$. The experiment will be performed on the *alarm*- and *hepar2*-network.

7.3.1 The monotone decrease in $\mathbf{Cl}_{Affected}$

When a change in evidence occurs, we propagate this through the junction tree. This will change the potentials of the cliques and we know that these changes will not increase. In order to get insight into the changes of cliques in $\mathbf{Cl}_{Affected}$, we first of all split up the increases of probabilities and the decreases of probabilities in potentials. Further on, we will only consider the maximum increase and the maximum decrease of a probability in a potential. This because the theorem about monotone decrease only concerns these two extremes. We will refer to the maximum increase or decrease of a potential as the change of a potential. From the context, it will be clear whether the maximum increase or decrease is meant.
We want to visualize multiple scenarios in order to get a more general idea about the monotone decrease. For this purpose, we will pick a *run* and consider all changes in the potentials of cliques that are part of $\mathbf{Cl}_{Affected}$. We will group the cliques by path. With path, we mean a set of cliques that form a path from the source of change in evidence to a *border clique*. Due to grouping cliques by path, a change in a potential could be displayed multiple times because some cliques will be part of multiple paths. Each change in a clique presents a point in the graph. The y-axis displays the size of the change and is logarithmic in scale. The x-axis displays the distance to the source of change in evidence. The colour of a point indicates the length of the path to which the point belongs. There will be areas in the graph with many points. In order to get an indication on how dense the areas are, a couple of measures have been taken. First of all, points with different colours are displayed next to each other. Secondly, the points are slightly transparent. This will give denser areas a more solid colour. Thirdly, a line representing the median of the points on the different lengths is drawn. For each length, we only consider the *border cliques*.

Let us first look at *alarm*-network with the randomly generated CPT-tables where the skewness of the Bayesian network is equal to 0.351. The decrease of change is clearly visible as the distance to the source becomes larger. Especially the blue points give a good indication on how the changes decrease along their paths. At the start of these paths, most probabilities decrease with a factor larger than four. Most paths end up at cliques where almost no changes occur.
Interestingly, while the influence of the change in evidence is almost negligible after a distance of six, there are still cliques at distances of seven, eight or even nine in which variables change of state in the new MPE. The reason for this could either be that a changed variable is present in all three or four cliques. Another reason could be that the probability of two different configurations of variables located in these three or four cliques were close to each other and as a result of the minor changes, the other configuration is more likely under the new evidence.

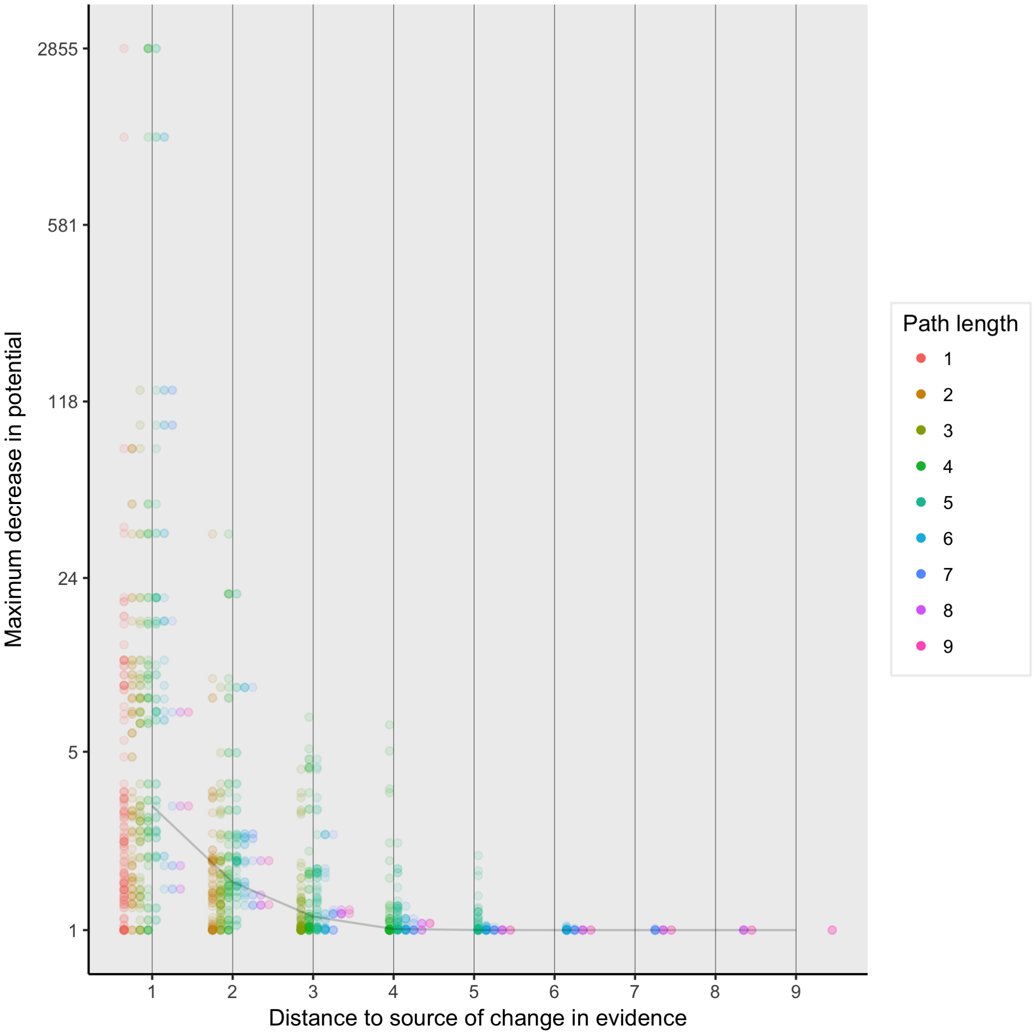


Figure 17: ***alarm*-network with skewness of 0.342. The graph contains 1289 points and shows the maximum decreases in the affected cliques. The data comes from a run in which the scenarios have one variable in the *pre-change evidence* and the state of this variable changes**

In Figure 18, results of a run for the *hepar2*-network are shown. The CPTs are again custom made and the Bayesian network has a skewness of 0.343. This graph is different compared to the graph of *alarm* on some points. The highest changes present in the graph are much smaller than the changes in *alarm*. This is however quite variable. There are scenarios in *hepar2* that start with changes higher than 1000 as well. Although the decrease in changes as the distance growths is visible here as well, it does not decrease as much as in the case of *alarm*. Another oddity in this graph is that the median for paths of length five is higher than the median for paths of length four. This phenomenon was visible in several other runs as well and also for different CPTs. The structure of the network seems therefore be responsible for this behaviour. In most cases, the median drops as the distance towards the source of change in evidence increases. The reason for this is that every step further away from the source is likely to cause a decrease of change and never an increase of change.

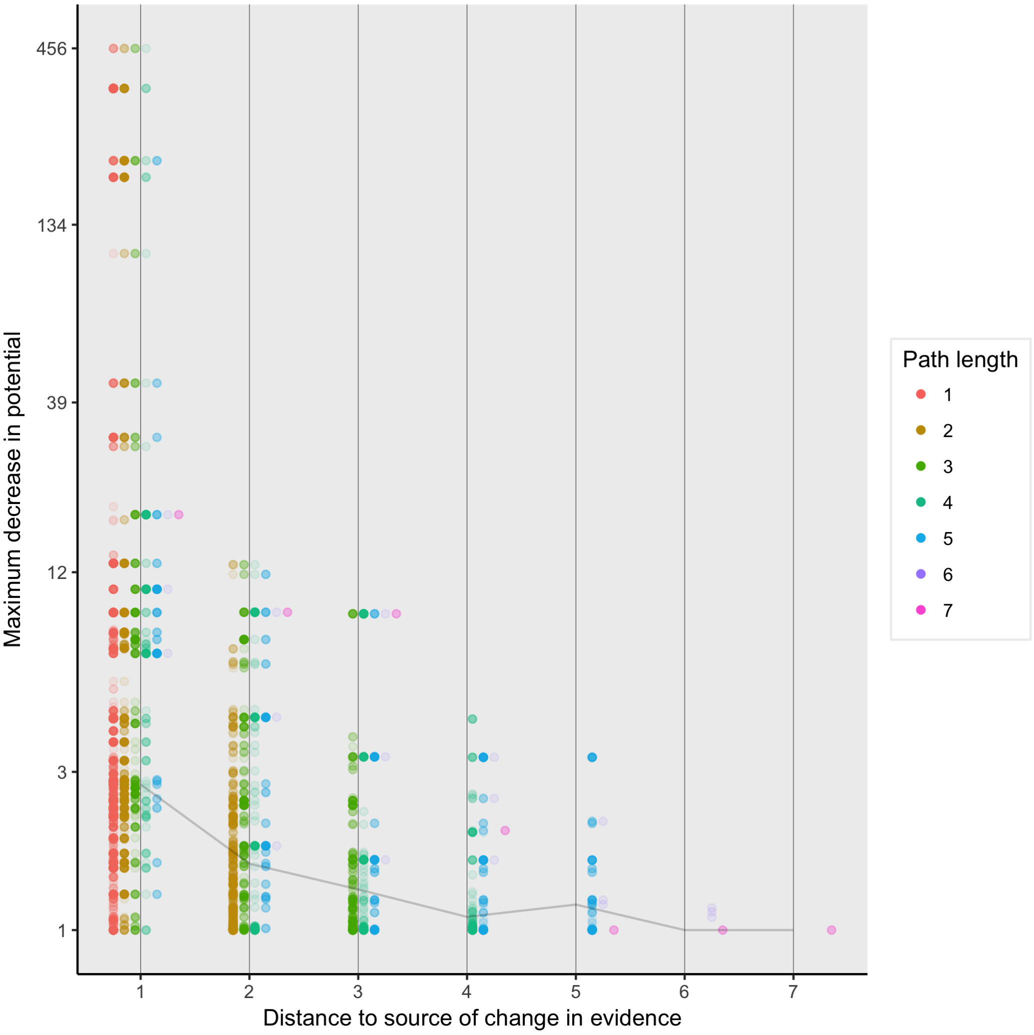


Figure 18: ***hepar2*-network with skewness of 0.343. The graph contains 5829 points and shows the maximum decreases in the affected cliques. The data comes from a run in which the scenarios have one variable in the *pre-change evidence* and the state of this variable changes**

However, you might expect that a change in evidence should in general be larger in order to change the state of a variable further away. Therefore, if you compare two cliques that have the same distance towards the source, then the clique that is part of the longest path will probably also have the largest changes to its probabilities. In the case of Figure 18, the distances to the source of change in evidence are not equal but since *hepar2* contains many situations in which the decrease of change is minimal, due to which the above reasoning can extend to not equal distances as well.

We will now look at *alarm* and *hepar2* with their original CPTs. The increases will be considered here as well. Figure 19 and 20 present relatively the decreases and increases in scenarios of the *alarm*-network. The first thing to note is the distribution of the points. While the changes converges quickly to one in the case of random CPTs, the changes in these two figures are less reluctant to

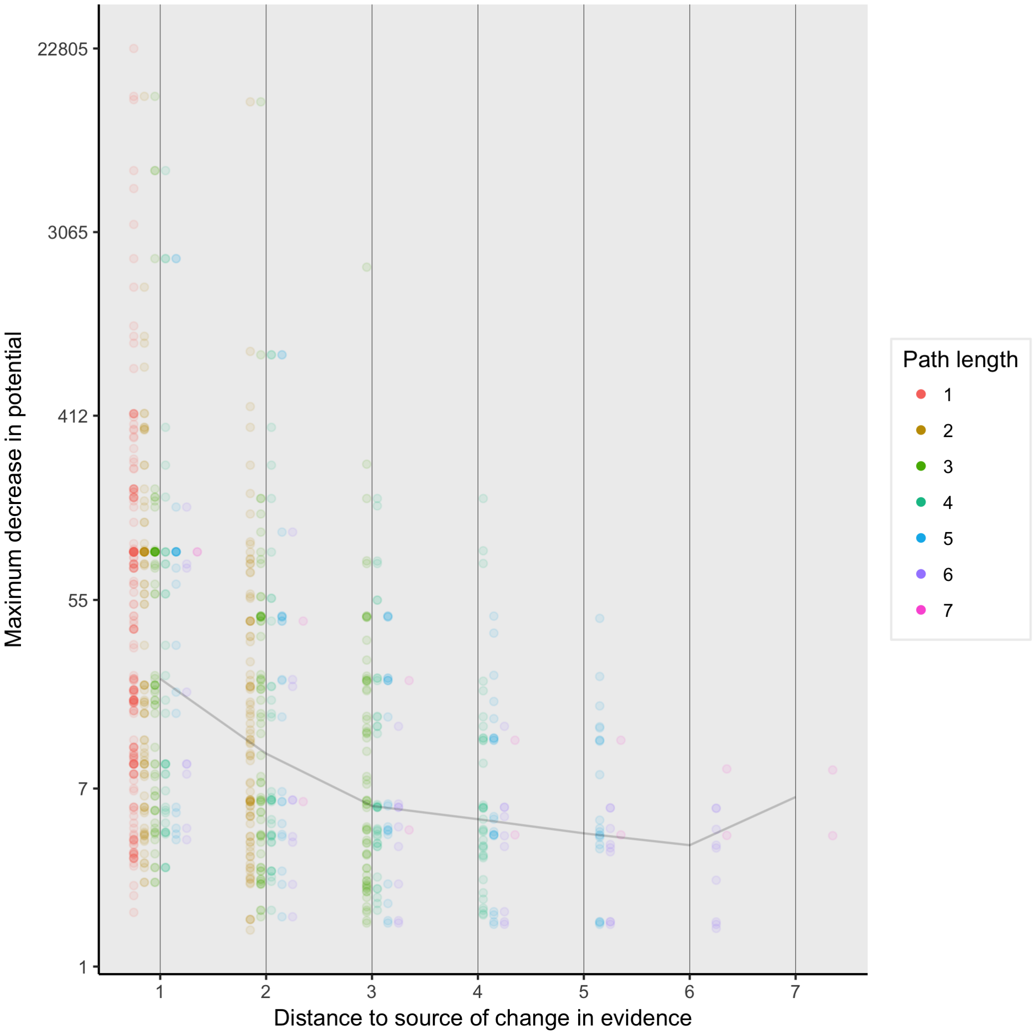


Figure 19: Original *alarm*-network. The graph contains 883 points and shows the maximum decreases in the affected cliques. The data comes from a run in which the scenarios have one variable in the *pre-change evidence* and the state of this variable changes

decrease and the cliques at the end of the paths are still often subject to change. It is further on the case that the probabilities in every affected clique are subject to change. This was not the case for the networks with random CPTs.
Despite the reluctant decrease, the relation between decrease of change and length of path is still visible. For the maximum increase, we see a drop from 65 at a distance of one to 5 at a distance of four. For the maximum decrease, we see a decrease from 23 to 5 over the same distance.

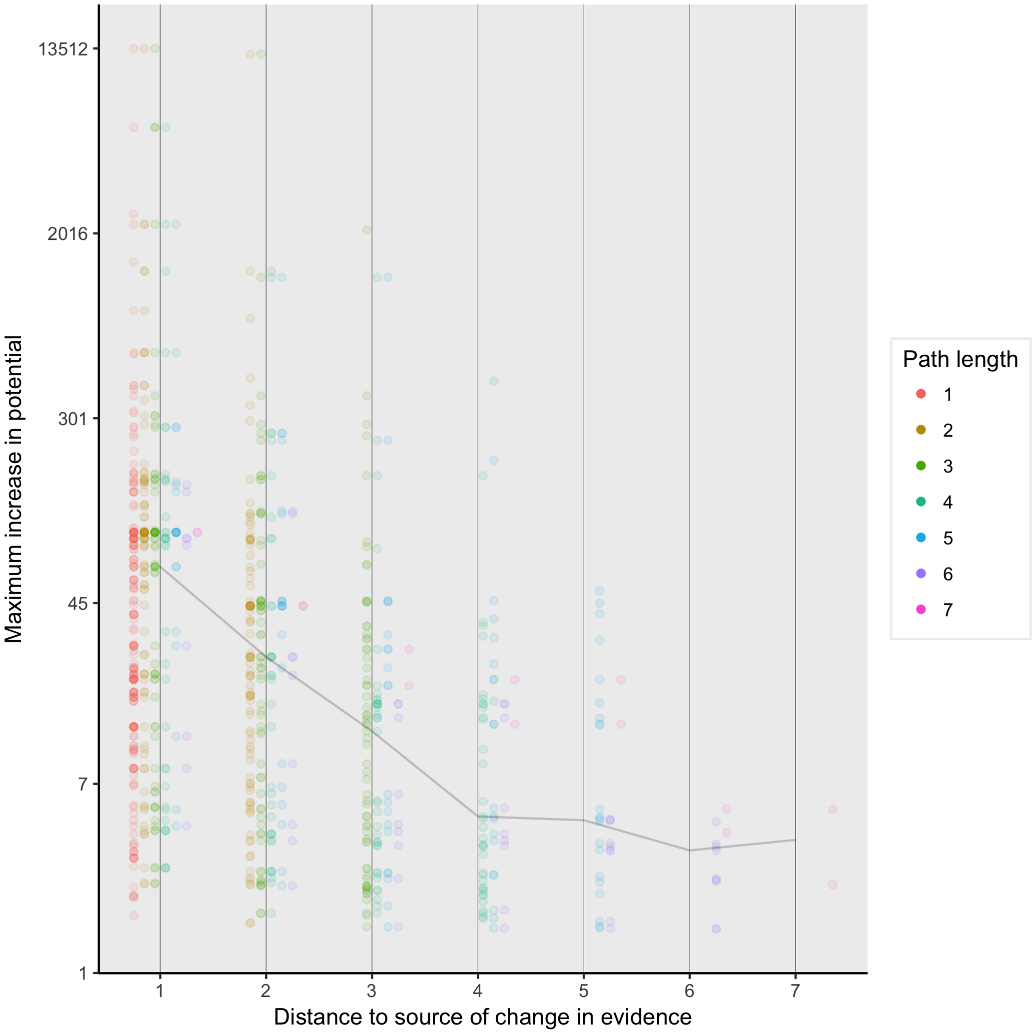


Figure 20: Original *alarm*-network. The graph contains 883 points and shows the maximum increases in the affected cliques. The data comes from a run in which the scenarios have one variable in the *pre-change evidence* and the state of this variable changes

In Figure 21 and 22 the results of changes in increase and decrease for the original *hepar2*-network are displayed.

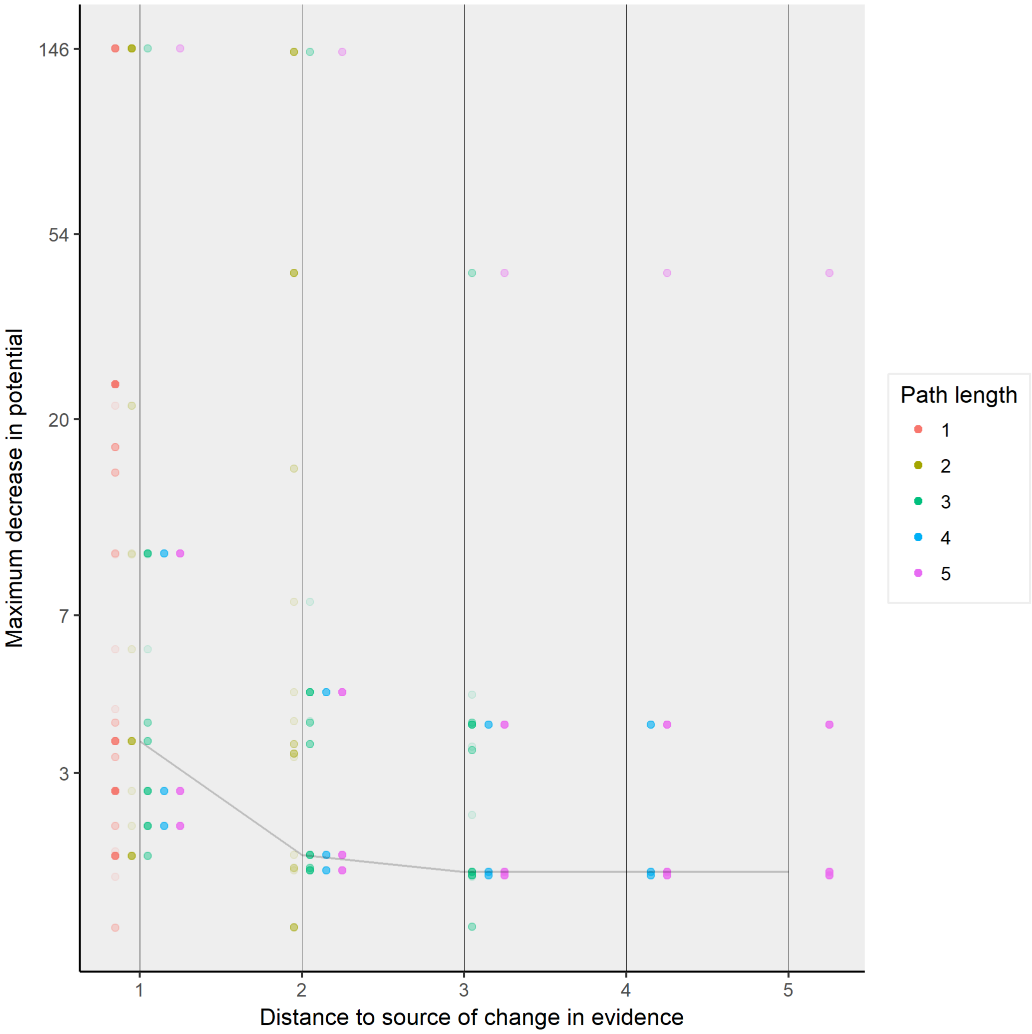


Figure 21: Original *hepar2*-network. The graph contains 760 points and shows the maximum decreases in the affected cliques. The data comes from a run in which the scenarios have one variable in the *pre-change evidence* that changes of state

Compared to *alarm*, these results contain much more cliques with similar changes. This is due to the fact that there are many separators which share most of their variables but are not completely the same. Due to this, we will have small changes as propagation continues along a path.

Further on, it seems that only a small number of points is displayed in the graph while the total number of points is 760 - which is just under the number of points displayed for *alarm*. To explain this, we have to look back at the tables of *alarm* and *hepar2* in section 7.2.1. The number of affected cliques for *alarm* is around 7.30 while the number of affected cliques for *hepar2* is 4.38. The number of cliques in $\mathbf{Cl}_{Border}$ for *alarm* is therefore likely to be higher than for *hepar2*. However, the proportion

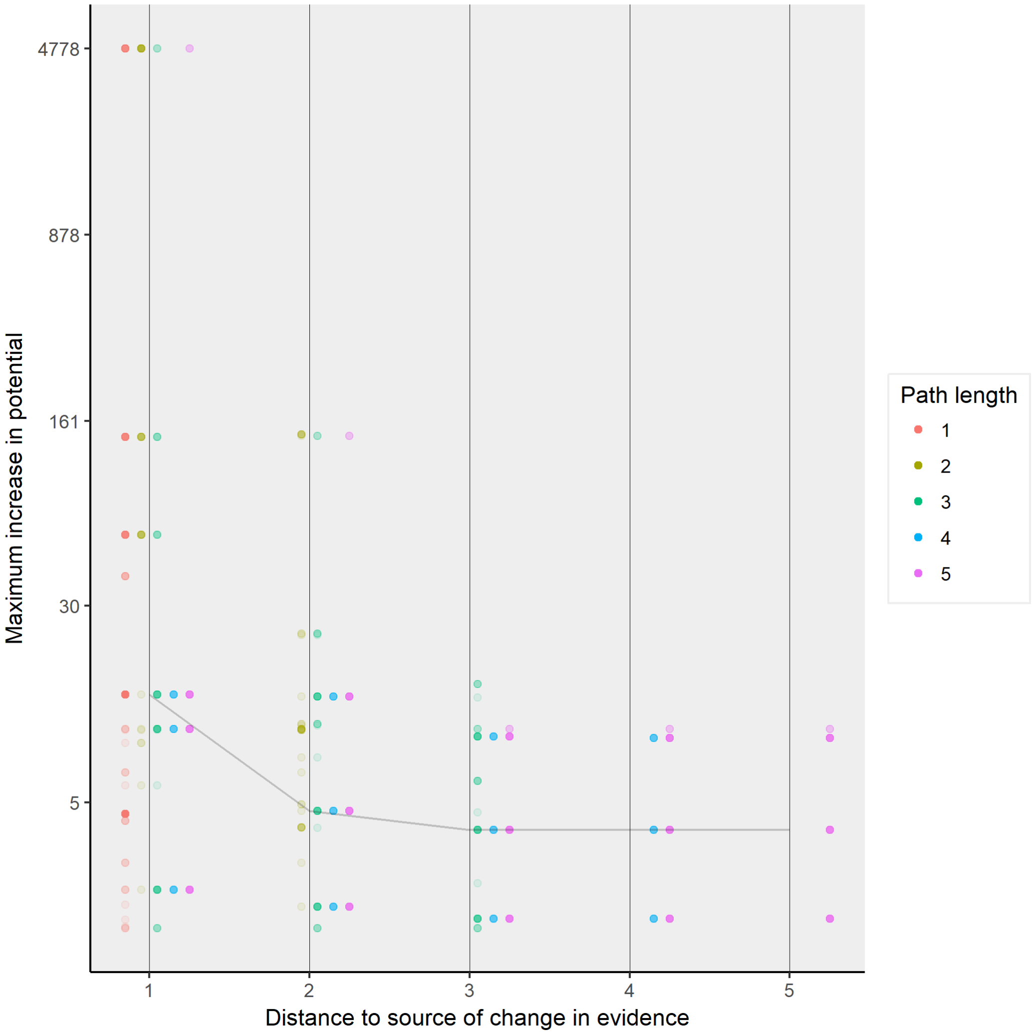


Figure 22: Original *hepar2*-network. The graph contains 760 points and shows the maximum increases in the affected cliques. The data comes from a run in which the scenarios have one variable in the *pre-change evidence* that changes of state

of cliques in $\mathbf{Cl}_{Border}$ that are at the end of a branch in the junction tree is higher - 0.387 against 0.258.

Another thing that can be noted from the figures is that despite the fact that we have often small decreases along a path, the monotone decrease is visible. Especially between cliques with a distance of one and two from the source of change in evidence. For a distance of two, the maximum decrease in half of the separators decrease to less than two while the maximum increase decrease to less than four.

7.3.2 Changes Outside $\mathbf{Cl}_{Affected}$

In the previous section, we have looked at the changes within $\mathbf{Cl}_{Affected}$ and although we saw the monotone decrease, it was also clearly visible that the changes in cliques at the end of the paths could still be large. However, these paths only concerned cliques in which the MPE configuration changed. What if we look at the cliques in which persistence occurs for the first time? We would expect that the changes of probabilities within these set of cliques is small. To gain insight into this set of cliques, we first of all gather separators that are adjacent to a clique in $\mathbf{Cl}_{Border}$ and a clique outside $\mathbf{Cl}_{Border}$. Each of these separators will be presented as a point in the graph. On the y-axis, it indicates the maximum increase or decrease that took place in the potential of the separator. The x-axis represents the distance between the source of change in evidence and the clique in $\mathbf{Cl}_{Border}$ that is adjacent to the concerned separator. In order to get insight in clusters of points, we add ± 0.2 to the distance at random and use transparent points like we did in the previous graphs. In order to compare the results with the previous section, we used the same runs for *alarm* and *hepar2* to show the increases and decreases in the separators.
Figure 23 and 24 contain the results for *alarm*. While we had max increases in cliques up to 13000

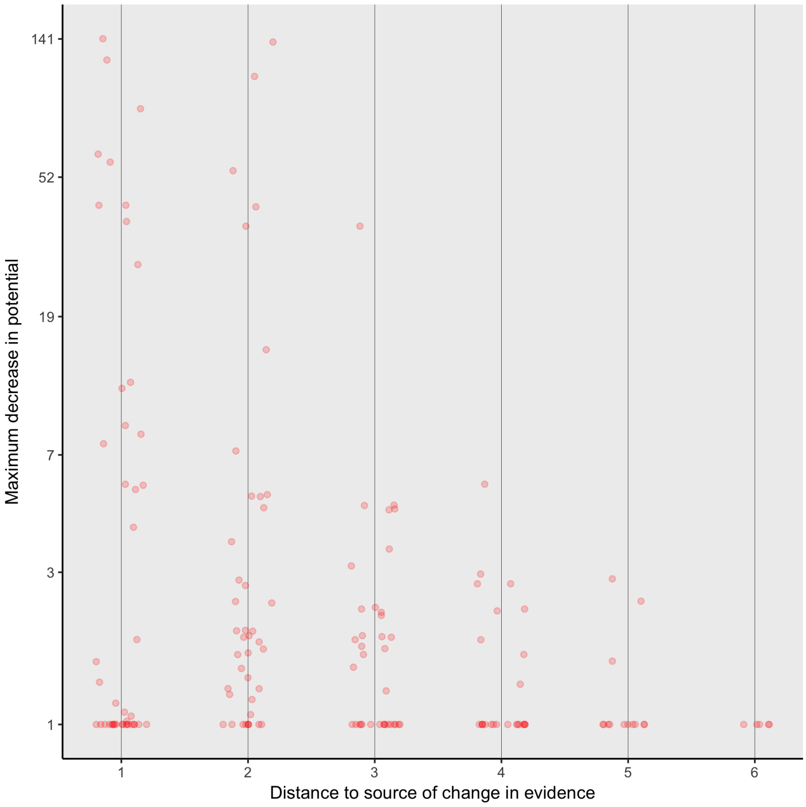


Figure 23: Original *alarm*-network. The graph contains 164 points and shows the maximum decreases of the cliques just outside the affected area of cliques. The data comes from a run in which the scenarios have one variable in the *pre-change evidence* that changes of state

and maximum decreases up to 22000, the probabilities in the potentials of cliques outside $\mathbf{Cl}_{Affected}$ are at most increased or decreased by 141. Only few potentials have a change of this magnitude while

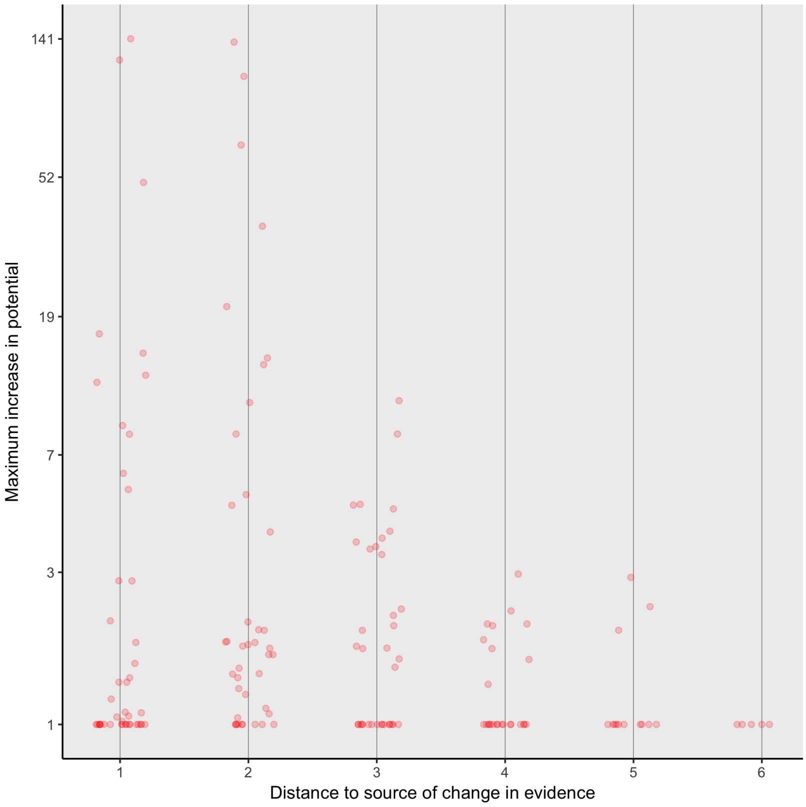


Figure 24: Original *alarm*-network. The graph contains 164 points and shows the maximum increases of the cliques just outside the affected area of cliques. The data comes from a run in which the scenarios have one variable in the *pre-change evidence* that changes of state

they persist. The probabilities in most cliques increase or decrease no more than seven times.
An other interesting thing to note concerns the independency. We already saw that many cliques are independent of the changes and is visible in this graph as well. In Figure 19 and 20 we didn't see any clique in $\mathbf{Cl}_{Affected}$ that was independent of the changes. However, we did see these cliques in the *alarm*-network in which randomly generated CPTs were used. So while it is possible, it is rare when the variables in the MPE change while they are independent of the change in evidence.

The results for *hepar2* are shown in Figure 25 and 26. and look similar to the results of *alarm*.

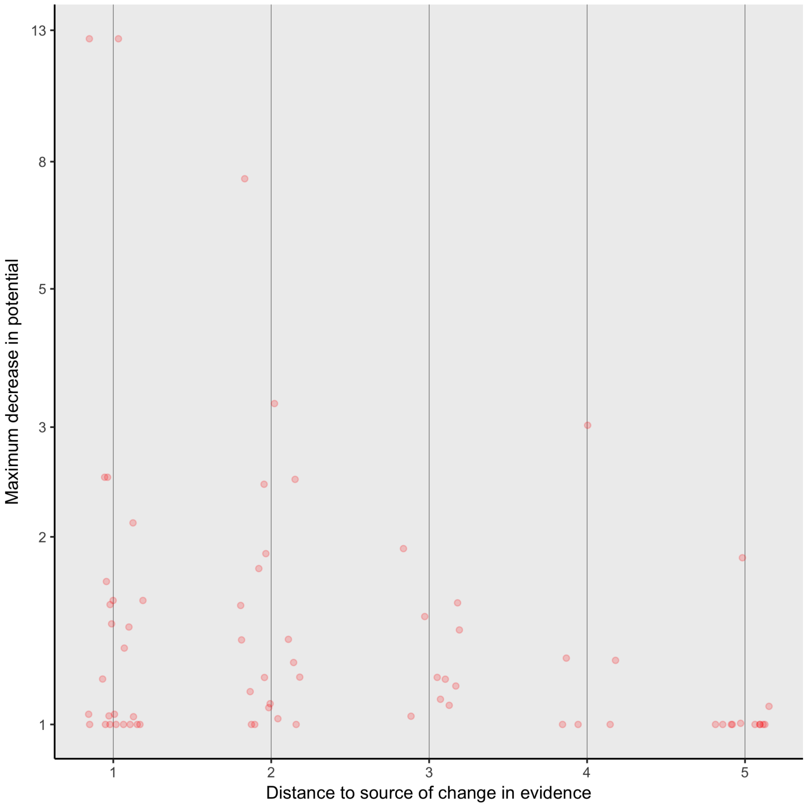


Figure 25: Original *hepar2*-network. The graph contains 72 points and shows the maximum decreases of the cliques just outside the affected area of cliques. The data comes from a run in which the scenarios have one variable in the *pre-change evidence* that changes of state

The changes in the first cliques that persist are significantly smaller than the changes in the cliques that do not persist. Almost all changes lie between a distance of one and three. A bit odd are the two cliques in Figure 26 that have a distance towards the change in evidence of two and a maximum increase in probability around 14 and 45. All cliques with a distance of one have a maximum increase of no more than three. This is however possible since the monotone decrease property only holds for cliques that are part of the same path. It could for example be that there is a scenario that has no endings at a distance of one and the separators with a distance of two are subject to changes that are coincidentally high.

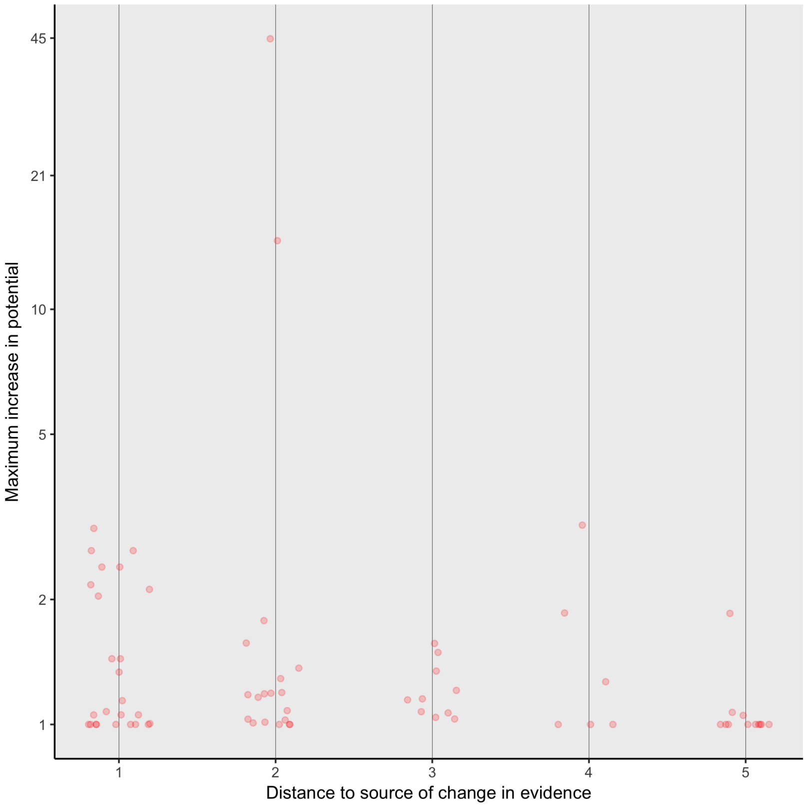


Figure 26: Original *hepar2*-network. The graph contains 72 points and shows the maximum increases of the cliques just outside the affected area of cliques. The data comes from a run in which the scenarios have one variable in the *pre-change evidence* and the state of this variable changes

7.4 Conclusion and Discussion

Summarizing the results, we have looked at characteristics of the affected clique set and its border. We first of all looked at the different cases that can occur at the border: The end of a branch, probabilities in the separator are independent of the changes in the evidence and the probabilities are affected by the change of evidence. We have seen that the fraction of type one cases decreases as we look at larger networks. For the type two and three cases, we have seen how they are affected by the number of variables in the *pre-change evidence*. For *alarm*, the effect was minimal, but the fractions did change for *hepar2* and *asia*. Possible causes for these changes are the probabilities of independencies being created or removed as the number of variables in the *pre-change evidence* increases. Further on, we have seen that the percentage of cases in which a clique is subject to change but does persist was in all three networks quite large.

Results about the size of the affected part of the junction tree shows us that the affected part is often only a fraction of the junction tree, especially for *hepar2*. The number of affected variables is often smaller than the number of affected cliques. We have also seen that the role of conditional probability tables can be important for the size of the affected clique set. While the results with different conditional probability tables for *alarm* did not give us very different results, the number of

affected cliques and variables for *asia* turned out to be lower while higher for *hepar2*.
Next, we considered the changes of probabilities in potentials. The monotone decrease is clearly visible in the potentials of cliques that are part of $\mathbf{Cl}_{Affected}$, especially for *alarm*. The difference between changes in the potentials of cliques inside and directly outside $\mathbf{Cl}_{Affected}$ was also evident.

In most cases, the number of variables that change of state in the MPE is small. Therefore, methods to find these persistences will save often many computations. Especially since persistence mostly occurs between two cliques and not just at the end of a branch. However, this will not always be the case. As we have seen for *hepar2*, different CPTs resulted in many more variables that changes of state in the MPE.
That being said, if the affected clique set is as small as was the case for the original *hepar2*-network, future evidence changes may not even influence previously affected cliques at all. This can be interesting in how to deal with partial propagation of consecutive changes in the evidence.
Another interesting thing is the relation between independency and persistence. It seems logical that variables independent of the change in evidence are not likely to change of state in the MPE, but we did see these cases when we used randomly generated CPTs. However, for the original networks of *alarm* and *hepar2*, these cases are rare. It would be interesting to find out how rare the cases are and under what circumstances the variable is likely to change of state.
Furthermore, the results for monotone decrease are quite varying. Although the decrease is often present, we also see cases in which the changes are not decreasing at all but remain the same. When separators persists varies widely as well. We have separators that are persistent while some of the probabilities in their potential become 40 times as high. However, it is clearly visible that the changes in cliques that are affected is significantly larger than the changes in cliques outside the affected clique set. Besides that the monotone decrease is varying, the course of it is not very smooth either. If you look at individual scenarios, one clique may have probabilities that becomes a 1000 times smaller while the next clique has a maximum decrease of 10. There are also situations in which a path starts with a maximum decrease of 2.5 in a clique and five cliques later, the maximum decrease dropped to 2.1. These abrupt changes are also clearly visible in the graphs for *hepar2*. In order to investigate the length of the path and the monotone decrease further, it would be useful to analyze more specific scenarios. Using random scenarios, we have evidence all across the junction tree. The evidence may cause the marginal probability distribution of neighbouring potentials to be more skewed which gives us the sudden decreases of the change that is propagated. In practise however, it is often the case that evidence in a network is located at a specific place like the roots or the leaves. The skewed probability distributions will as a result be more clustered at these places. This may result in more predictable courses of the monotone decrease.

8 MAP experiments

In this section, we will analyze how the configuration of a MAP is affected by a change of the evidence. As with the MPE experiments, we will use a junction tree as basis. We tried to perform similar experiments for the MAP as we have performed for the MPE experiments. But while we could adapt some parts, others had to be omitted. More details are stated in the set-up. After the set-up section, we will discuss the results of the experiments and end with a conclusion.

8.1 Set-up

Asia, *alarm* and *hepar2* will be used again as Bayesian networks in the experiments.
One concept that cannot be used for the MAP is $\mathbf{Cl}_{Affected}$. MAP variables may not belong to cliques that form a contiguous area. And even if the cliques are contiguous, we encountered cases in which we had two cliques Cl_R and Cl_S with both containing a variable from the MAP variables and while Cl_R lay between Cl_S and the source of change, the MAP variable located in Cl_S changed of state while the state of a MAP variable in Cl_R remained the same.

8.1.1 Runs

The run that is used here slightly deviates from the runs we used to analyze the MPE changes. We will only perform runs in which a variable in the *pre-change evidence* changes of state. Further on, while we considered in the previous experiments all variables that were not part of the evidence, we can now choose which variables should be part of the MAP. For most experiments, this number will be fixed to 25% of the variables in the Bayesian network. For *asia*, *alarm* and *hepar2* this means that respectively 2, 9 and 18 MAP variables will be used. The variables for the MAP will be chosen at random in each scenario like we do for the variables in the evidence. Because the number of distinct scenarios for MAP is much bigger than for MPE, the number of scenarios in one run will be set to 500.

8.1.2 Distance

We will make use of the distance measure described in section 7.1.4. However, while that measure represented the distance between two cliques, we are here more interested to measure the distance between a MAP variable and the source of change in evidence. Therefore, we first have to find a clique for the MAP variable in order to use the distance measure. We will pick the clique that contains the MAP variable and is closest to the source of change in evidence. We will refer to this type of clique with Cl_{Ma}. While the distance measure works for a specific junction tree, there is a problem with this measure. Recall that different sets of cliques can be constructed from a Bayesian network. This could result in a different distance between the MAP variable and source of change in evidence since they may be part of different cliques. Since we only consider one junction tree for each Bayesian network, we will use our distance measure as described. A paper detailing a solution for a distance measure that does not depend on the junction tree is under construction [22].

8.2 Results

Now that we have described the set-up, we can look at the results of the experiments. We will address three issues. First of all, we investigate the number of affected variables in the MAP after a change of evidence. Secondly, we will look into the distance between a MAP variable and the source of change in evidence and study how this relates to whether the variable changes state or not. Lastly, the changes in probabilities of cliques containing MAP variables will be studied.

8.2.1 Variable and clique count

In Tables 16-18, results for the three Bayesian networks are shown. The second column states the average number of affected cliques and the third column shows the average number of changed MAP variables. The percentage of scenarios in which no MAP variable changes is shown in the fourth column. The last two columns show again the average number of affected cliques and changed variables but they only consider the scenarios in which at least one MAP variable changes.

pre-change evidence	Avg. #Cliques	Avg. #Variables	No changes (%)	Avg. #Cliques with at least one change	Avg. #Variables with at least one change
1	$1.52 \pm .03$	$0.86 \pm .02$	37 ± 1	2.41 ± 0.04	1.37 ± 0.01
2	$1.25 \pm .07$	$0.67 \pm .03$	49 ± 2	2.45 ± 0.10	1.32 ± 0.03
4	$0.96 \pm .06$	$0.48 \pm .03$	63 ± 2	2.58 ± 0.08	1.31 ± 0.05

Table 16: Results about the number of affected cliques and changed variables in the *asia*-network for different types of scenarios. The set of MAP variables consists of two variables.

pre-change evidence	Avg. #Cliques	Avg. #Variables	No changes (%)	Avg. #Cliques with at least one change	Avg. #Variables with at least one change
1	$2.42 \pm .08$	$1.06 \pm .04$	44 ± 2	4.30 ± 0.09	1.88 ± 0.03
2	$2.37 \pm .08$	$1.03 \pm .03$	44 ± 2	4.25 ± 0.10	1.84 ± 0.04
5	$2.14 \pm .11$	$0.92 \pm .04$	49 ± 2	4.15 ± 0.16	1.79 ± 0.07
9	$1.74 \pm .14$	$0.77 \pm .06$	55 ± 3	3.84 ± 0.19	1.70 ± 0.04

Table 17: Results about the number of affected cliques and changed variables in the *alarm*-network for different types of scenarios. The set of MAP variables consists of nine variables.

pre-change evidence	Avg. #Cliques	Avg. #Variables	No changes (%)	Avg. #Cliques with at least one change	Avg. #Variables with at least one change
1	$0.82 \pm .18$	$0.25 \pm .02$	82 ± 1	4.63 ± 0.73	1.42 ± 0.05
2	$0.89 \pm .15$	$0.27 \pm .02$	81 ± 1	4.77 ± 0.66	1.44 ± 0.05
5	$1.09 \pm .17$	$0.30 \pm .04$	80 ± 2	5.56 ± 0.46	1.52 ± 0.09
9	$1.19 \pm .21$	$0.31 \pm .03$	80 ± 2	6.01 ± 0.72	1.57 ± 0.10
15	$1.36 \pm .22$	$0.32 \pm .05$	80 ± 2	6.87 ± 0.93	1.59 ± 0.10
25	$1.54 \pm .20$	$0.33 \pm .04$	80 ± 2	7.70 ± 0.89	1.62 ± 0.13

Table 18: Results about the number of affected cliques and changed variables in the *hepar2*-network for different types of scenarios. The set of MAP variables consists of 18 variables.

We see that the average number of cliques and variables that are affected by the change of evidence is small, especially for *hepar2*. In 80% of the cases, no change in the configuration of the MAP occurs. While the percentage of cases for the other networks starts low at around 40%, it increases to more than 50% as the number of variables in the *pre-change evidence* increases.
The cases in which at least one MAP variable changes does not give a larger number of changed variables than expected. In all three cases, the number of changed MAP variables stays below two on average. For *asia* and *alarm*, the numbers are quite stable in contrast to the overall average of changed variables as the number of variables in the *pre-change evidence* growths. This seems to indicate that the decrease of changed variables is mostly caused by the increasing number of cases in which no MAP variables changes. This is not the case for *hepar2*, the percentage of cases in which there is no change remains 80% and the number of changed variables increases slightly in both columns two and six.

Besides the different sizes of *pre-change evidence*, we also looked at different number of MAP variables for the *alarm*-network. Figure 27 shows how the number of variables and cliques growths as the number of MAP variables increases. The *pre-change evidence* consists of a single variable. We see a linear increase in the number of changed variables. The number of cliques seems to flatten at the end. This could be due to the fact that variables are likely to be present in multiple cliques. As the number of MAP variables increases it becomes less and less likely that newly introduced variables are present in cliques that do not already contain another MAP variable.

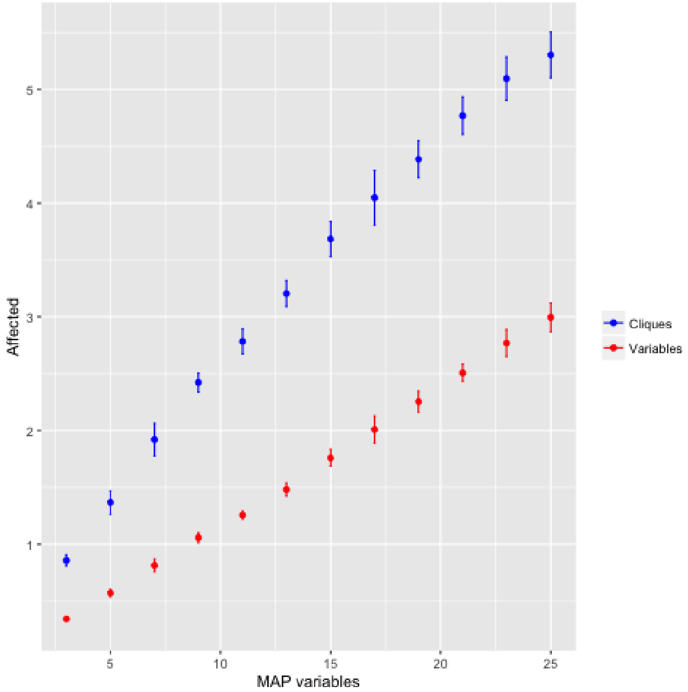


Figure 27: The average number of variables and cliques that are affected when we change the evidence for different numbers of MAP variables. It concerns the *alarm*-network with one variable in the *pre-change evidence*. The bars show the standard deviation over ten runs.

8.2.2 Distance of MAP variables

In the next part, we divide the MAP variables into two groups: The changed MAP variables and the persistent MAP variables. We want to compare the distance that changed MAP variables have to the the source of change in evidence versus the distance that persistent MAP variables have to the source of change in evidence. Table 19-21 contain the results for the three different Bayesian networks. The distance for persistent MAP variables is stated in the second column and the distance of changed MAP variables is shown in the last column. Different settings for the *pre-change evidence* are used.

pre-change evidence	Avg. Distance No Change	Avg. Distance Change
1	1.21 ± .02	0.65 ± .01
2	0.89 ± .04	0.80 ± .06
4	0.92 ± .04	0.81 ± .06

Table 19: Results about the distances between the MAP variables and the source of change in evidence in the *asia*-network.

pre-change evidence	Avg. Distance No Change	Avg. Distance Change
1	4.29 ± .06	1.27 ± .08
2	4.29 ± .05	1.25 ± .04
5	4.22 ± .08	1.14 ± .06
9	4.21 ± .04	0.96 ± .08

Table 20: Results about the distances between the MAP variables and the source of change in evidence in the *alarm*-network.

pre-change evidence	Avg. Distance No Change	Avg. Distance Change
1	2.36 ± .04	0.73 ± .13
2	2.36 ± .04	0.79 ± .12
5	2.37 ± .03	0.84 ± .11
9	2.38 ± .05	0.86 ± .10
15	2.38 ± .03	0.82 ± .12
25	2.39 ± .04	0.73 ± .13

Table 21: Results about the distances between the MAP variables and the source of change in evidence in the *hepar2*-network.

The distance for changed MAP variables is in all three tables smaller for the MAP variables that did notchage state. For *asia* however, the difference is small for the case of two and four variables in the *pre-change evidence*. For *alarm*, the distance for the persistent MAP variables shrinks slowly as the variables in the *pre-change evidence* increases. The distance for the changed MAP variables does not differ significantly for the different settings of the *pre-change evidence*. The difference between the distance of the two types of MAP variables is large. MAP variables that change are on average four times as close to the source of change in evidence than the other MAP variables.

The distances of the MAP variables that do not change in *hepar2* are even closer to the source. They are often in the same clique as the changed variable in the evidence. Note that although the distance for MAP variables that do not change is about three times as high, it is still quite close to the source and is about two times as small compared to the numbers for *alarm*. Despite that there are many more cliques in *hepar2* than in *alarm*. The reason for this is that there are a lot of chains in *hepar2* that consists of cliques that are connected with separators containing the same variables. Due to this, the distance from a certain clique is the same to all cliques in the chain. The maximum distance that we have seen between the source of change in evidence and a variable is eight while eleven was the maximum distance in the *alarm*-network.

8.3 Change of probabilities

In this part we will again consider the distance of changing and persistent MAP variables but, in addition, plot the maximum increases and decreases in the potential of Cl_{Ma} for each MAP variable as well. Not all MAP variables will be considered because some are located in the same clique as the variable that changed in the evidence, resulting in an infinite increase and decrease. These MAP variables will be discussed later on.

The max decreases in a run for the *alarm*-network are shown in Figure 28. We notice first of all that there is much more data for persistent MAP variables than for changed MAP variables. We already saw that on average, 1.06 of the *nine* MAP variables change of state. When ignoring the MAP variables with a maximum decrease of infinite, the average percentage of MAP variables that change is less than 9%. Like in the MPE experiments, the relation between the distance and the decrease of change in cliques is clearly visible. The maximum increases in Figure 29 show similar

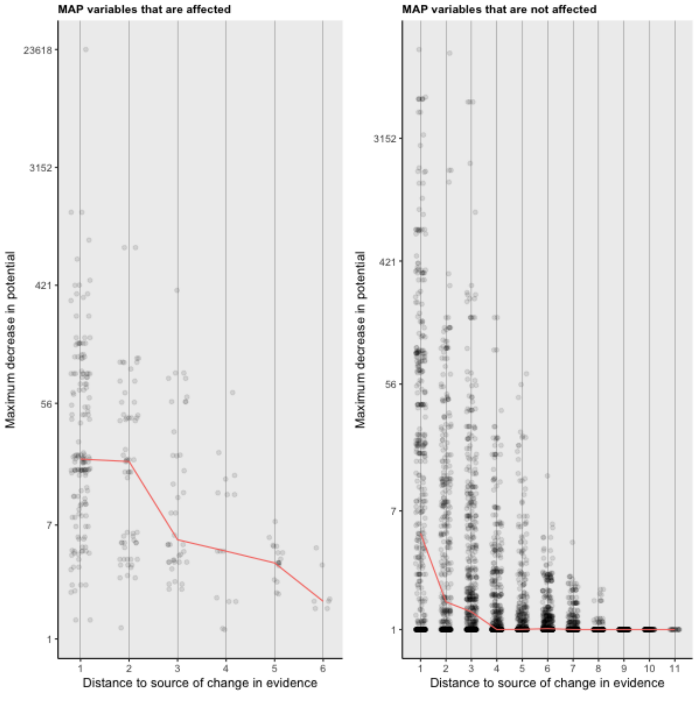


Figure 28: For each MAP variable, the maximum decrease in the potential of Cl_{Ma} is shown except for the MAP variables that have a distance of zero to the source of change in evidence. The figure on the left concerns MAP variables that changed. On the right are the MAP variables that persisted. The *pre-change evidence* for this run consists of one variable. The line represents the median.

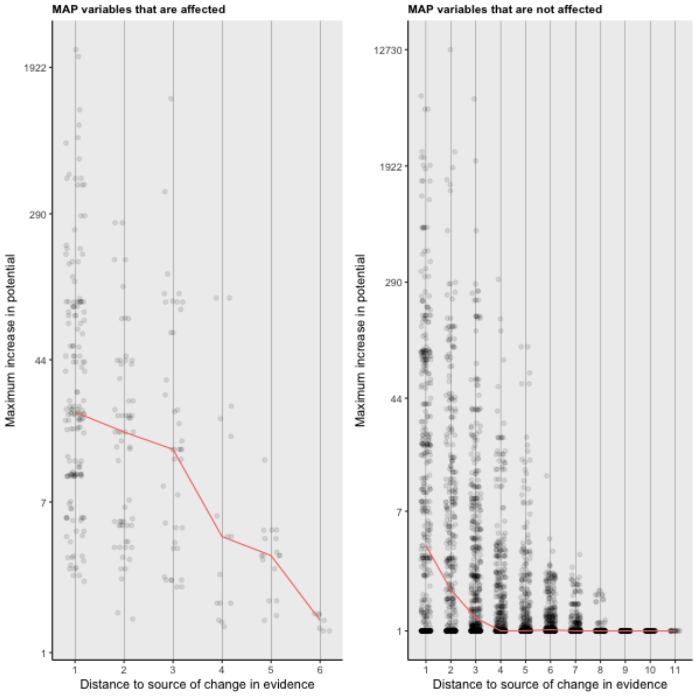


Figure 29: For each MAP variable, the maximum increase in the potential of Cl_{Ma} is shown except for the MAP variables that have a distance of zero to the source of change in evidence. The figure on the left concerns MAP variables that changed. On the right are the MAP variables that persisted. The *pre-change evidence* for this run consists of one variable. The line represents the median.

results. Interestingly, the affected cliques with the largest changes belong to MAP variables that are persistent while this is the other way around for the maximum decreases. The median shows that change in evidence has in most cases a small effect on the cliques of the persistent MAP variables. However, it seems also more likely that the MAP variables in cliques, that are subject to larger changes in their probabilities, will persist as well.

To further investigate this, we counted the cliques for different magnitudes of changes. Results are stated in Table 22. The table contains the numbers of cliques presented in the figures above. We looked at both the maximum increase and the maximum decrease. The first column in the table indicates the change that should be present in the potential of the clique in order to be counted. So, in the third row for example, we have that 133 cliques contain a changed MAP variable and at least one probability in their potential decreased more than 20 times. In the last row, the number of cliques with maximum increases and decreases of infinite are listed. A first thing to note is that in most cases, the MAP variable is more likely to persist than to change. The number of persistent MAP variables is very close to the number of changed MAP variables in the last row. This in contrast to the other rows.
For *asia* and *hepar2*, we have slightly different results. We picked for both a run as well and the results are shown in Table 23 and 24. *Asia* contains many more changed MAP variables and they form the majority when looking at larger changes to a clique. There are more changed MAP variables than persistent MAP variables in case of infinite as well.

Change more than	Changed MAP variables (Decrease)	Persistent MAP variables (Decrease)	Changed MAP variables (Increase)	Persistent MAP variables (Increase)
1	297	1908	297	1908
5	215	568	225	541
20	133	307	141	297
100	41	116	37	110
500	7	41	8	30
2000	1	17	2	10
Infinite	212	233	212	233

Table 22: Total numbers of the cliques that contain MAP variables based on how they are affected by the change in evidence. The numbers are gathered from a single run in the *alarm*-network.

For *hepar2*, we see the contrast between persistent and changed MAP variables even more clear compared to *alarm*. The same trend is visible however: For the infinite cases, we see that the number of changed MAP variables is again high, but still not close to the number of persistent MAP variables.

Change more than	Changed MAP variables (Decrease)	Persistent MAP variables (Decrease)	Changed MAP variables (Increase)	Persistent MAP variables (Increase)
1	103	297	151	307
5	59	88	113	84
20	29	24	62	23
100	29	24	42	23
Infinite	278	182	278	182

Table 23: Total numbers of the cliques that contain MAP variables based on how they are affected by the change in evidence. The numbers are gathered from a single run in the *asia*-network.

Change more than	Changed MAP variables (Decrease)	Persistent MAP variables (Decrease)	Changed MAP variables (Increase)	Persistent MAP variables (Increase)
1	52	6522	52	6522
5	6	497	13	622
20	2	113	9	158
100	1	32	2	34
2000	0	6	0	7
Infinite	73	582	73	582

Table 24: Total numbers of the cliques that contain MAP variables based on how they are affected by the change in evidence. The numbers are gathered from a single run in the *hepar2*-network.

8.4 Conclusion and Discussion

In this section, we looked at how the MAP configuration changes in a junction tree. We first focused on the number of variables that changes and in how many cliques these variables were present. Overall, a minority of the MAP variables changed. The least changes were found in the *hepar2*-network were the average number of changed MAP variables was around 0.3 out of 18. Different sizes for the set of MAP variables for the *alarm*-network were tested and we saw that the number of changed MAP variables growths linearly as the number of MAP variables increases.
We further on considered the distance of the MAP variables towards the source of change in evidence and saw that there is a significant difference between the distance of MAP variables and the distance of persistent MAP variables to the source of change in evidence. Lastly, we considered the maximum increases and decreases for MAP variables in the Bayesian network of *alarm*. At any rate of change of probabilities, MAP variables are more likely to persist than change except for the *asia*-network. Another interesting result for *alarm* and *hepar2* is that when the changes in a clique increases, it remains very likely that the MAP variables will persist. However, when MAP variables are in the same clique as the source change in evidence, we see for *hepar2* that the probability for changing increases again while it becomes for MAP variables in the *alarm* even more likely that it changes.

These experiments show us that a change in evidence has on average a small impact on the MAP configuration. When we consider the number of changed MAP variables relative to the total number of MAP variables and compare this with the relative number of changed MPE variables in our previous experiments, there is no big difference in how many variables changes of state. It is interesting to find methods that can detect persistence for the MAP variables, especially since the MAP problem seems to be more complex than the MPE problem and the MAP problem is more general. From our experiments, the changes in probabilities in potentials do not tell us much about whether a MAP variables persists or not. This could be explained by the fact that within cliques, most variables have to be marginalized out when computing the probability for the MAP. However, we currently only considered changes to the marginal distribution of all variables in the clique. It may gives us more insight when we marginalize further within the clique and only consider how these resulting distributions changes when a change in evidence is introduced.

Lastly, note that the results in section 8.3 are all based on single runs instead of an average over ten runs. The reason for this is that the highest maximum increases and decreases vary greatly over different runs. There are few datapoints in a run with a high increase or decrease. And because the number of scenarios is big, we have runs in which we have maximum increases of more than 40.000 while in other runs, the maximum increase is not more than 10.000. This makes it hard to combine the results of ten runs. However, we did look at separate runs and were able to make the same conclusions regarding relations between persistent and changed MAP variables and the distribution of the maximum increase and decrease over the distance.

9 Conclusion

9.1 Summary of results

In this thesis, we analyzed the MPE and MAP problems within Bayesian networks. We transformed a Bayesian network into a junction tree and used the method of standard junction tree propagation as the basis for our study. We performed our analysis by breaking it down into three research questions. The first research question concerns how the MPE changes in the Bayesian network after changes in the evidence. We approached this question by considering the changes to the probabilities of all configurations. We found a theorem that holds for standard propagation in a junction tree. It states that as propagation advances through the junction tree, the propagated changes can never become larger. We developed a way to visualize the course of standard propagation through a junction tree to show how the monotone decrease progresses. We used the monotone decrease in the design of the experiments. In these experiments, we considered three Bayesian networks: *asia*, *alarm* and *hepar2*. We generated many scenarios of a single change in the evidence. And despite the fact that the changes may stay the same instead of decreasing, we saw in general a clear decrease of changes as the distance to the source of change in evidence growths. At the border between cliques that are affected on one side and cliques that contain only persistent variables on the other side, we saw that the changes were much smaller compared to the changes in the affected cliques. We also looked at the number of variables that changed state in the MPE and the number of cliques that contain a changed variable. In general, these numbers were small. Especially for *hepar2* in which fewer than two of the 70 variables changed of state on average.

For our second research question, we focused on the more general form, the MAP configuration, and tried to find out using similar experiments how the MAP configuration differs when evidence changes. We used 25% of the variables as MAP variable. In all three Bayesian networks, we saw on average only small changes in the MAP configuration after a change in evidence. While the number of MAP variables is much smaller than the number of MPE variables, the relative number of MAP variables that changed was equal to or less than the relative number of MPE variables that changed. In addition, the changed MAP variables are on average close to the source of change in evidence. Most variables that changed were present in cliques in which the change in evidence took place as well. We could however not find a clear relation between how much a potential of a clique changes and whether a MAP variable within that clique changes or persists.

Our last research question addresses the way we can use the relations between changes in the MPE and the changes in the potentials of cliques in the junction tree to reduce computations. We found a theorem that can detect persistence of the MPE. The conditions can be checked without adding extra complexity to the standard propagation. When the theorem holds in a clique and the variable set, say P, in the separator in which propagation is performed is not included in any of the other adjacent separators, we know that all variables in the current clique and subsequent separators and cliques will persist, except for the variables in P. Unfortunately, we have not seen any situation yet that fulfills the conditions for the theorem. Therefore, the impact will probably be small. Furthermore, we do not know as yet how the variables in the non-persistent part of the junction tree will change. And the theorem cannot be used in the next iteration of standard propagation if the previous propagation was not fully completed. All things considered, the theorem does not seem to be useful in practice. We further have not yet been able to use the monotone decrease property in standard propagation in a direct way to detect persistence.

These answers to the research questions indicate that the impact of changes in the evidence are in general small on both the MPE and the MAP configuration. Finding persistence in order to prematurely stop propagations will therefore save many computations. The monotone decrease property may function as a basis for such a reduction of computations because the changes of the MPE are in a contiguous area of cliques; the distance seems to be instrumental for the MAP as well. Since the monotone decrease property is based on standard propagation, it can be used for both MPE and MAP.

9.2 Limitations

While we tried to look at the general behaviour of the MPE and MAP within the experiments, only three Bayesian networks were used. The networks were chosen so that we had networks representing small networks, medium networks and large networks. However, there are many more properties in a Bayesian network that may give different characteristics, such as the connectivity of a Bayesian network and the average size of state space of variables within the network. Due to the general approach taken, we may have missed information about the MPE or MAP. For example, we randomly picked variables to be part of the evidence, while in most applications, evidence is more biased towards the roots or the leaves of a Bayesian network. Generating more specific evidence could give us more distinctive results. The same observations apply to the MAP variables.

We further considered a single junction tree per Bayesian network. Multiple junction trees may however exist. These junction trees can for example give different values for the distance measure we used in the MAP experiments.

9.3 Future work

Besides the fact that variables can change state, we also have the situation in which the parameters in the CPTs of the variables in a network may change. We did not investigate this type of change. The changes will probably have even less impact on the changes in the MPE or MAP configurations than a change of the state of a variable. It is therefore interesting to investigate how we can process these changes in a junction tree and know when the MPE configurations of variables persist.

Although we have the monotone decrease property when performing standard propagation, we have not been able to use this property for determining the MPE after a change. It seems that this assurance of monotone decrease can be used as a bound on how much the probability of the MPE and other configurations of cliques can change. When propagation is performed, these bounds may lend themselves for a criterion to stop propagation prematurely.

Another point we have not included in our research are consecutive changes of evidence. Research should be performed in how much the effect of changes of one variable is on the affected area by another change and how we can partially update the junction tree over multiple updates.

Furthermore, more extended experiments for the MAP problem should be performed. For example, we only considered changes of the marginal distribution of all variables in a clique while most variables in a clique will not be a MAP variable. marginalizing further and investigating how this distribution changes may result in a relation with whether the MAP variables changes state or not.

Finally, while we only mentioned monitoring applications as our motivation for investigating the effects of changes in a Bayesian network on the MPE, we can think of other applications that will also benefit if we can find ways to find persistence in the MPE or MAP configuration in order to reduce computations. For example, applications in which we want to consider scenarios where we change the state of an observed variable in order to see what happens with the network and how the

MPE changes. The change is only temporarily. In finding methods for reducing the computations, such a application will give us the advantage that we do not have to think about how to deal with consecutive changes in the evidence.

References

[1] sachs network. `http://www.bnlearn.com/bnrepository/discrete-small.html#sachs`. Accessed: 2017-11-03.

[2] Solomon Eyal Shimony. Finding maps for belief networks is np-hard. *Artificial Intelligence*, 68(2):399–410, 1994.

[3] Finn Verner Jensen, Kristian G Olesen, and Stig Kjaer Andersen. An algebra of bayesian belief universes for knowledge-based systems. *Networks*, 20(5):637–659, 1990.

[4] A Philip Dawid. Applications of a general propagation algorithm for probabilistic expert systems. *Statistics and computing*, 2(1):25–36, 1992.

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

[6] 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)*, pages 157–224, 1988.

[7] Stefan Arnborg, Derek G Corneil, and Andrzej Proskurowski. Complexity of finding embeddings in ak-tree. *SIAM Journal on Algebraic Discrete Methods*, 8(2):277–284, 1987.

[8] Vasilica Lepar and Prakash P Shenoy. A comparison of lauritzen-spiegelhalter, hugin, and shenoy-shafer architectures for computing marginals of probability distributions. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pages 328–337. Morgan Kaufmann Publishers Inc., 1998.

[9] Michael L Littman, Stephen M Majercik, and Toniann Pitassi. Stochastic boolean satisfiability. *Journal of Automated Reasoning*, 27(3):251–296, 2001.

[10] James D Park and Adnan Darwiche. Complexity results and approximation strategies for map explanations. *Journal of Artificial Intelligence Research*, 21:101–133, 2006.

[11] Rina Dechter. Bucket elimination: A unifying framework for probabilistic inference. In *Learning in graphical models*, pages 75–104. Springer, 1998.

[12] Dennis Nilsson. An efficient algorithm for finding the m most probable configurations in probabilistic expert systems. *Statistics and computing*, 8(2):159–173, 1998.

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

[14] Hei Chan and Adnan Darwiche. On the robustness of most probable explanations. *arXiv preprint arXiv:1206.6819*, 2012.

[15] RP van Rooij. Gaining insight in the solution space of the mpe problem when changing evidence or parameter values. Master's thesis, Utrecht University, 2015. Retrieved from Utrecht University Repository.

[16] Arnoud Pastink and Linda C van der Gaag. The persistence of most probable explanations in bayesian networks. In *Proceedings of the Twenty-first European Conference on Artificial Intelligence*, pages 693–698. IOS Press, 2014.

[17] Haipeng Guo and William H Hsu. *Algorithm selection for sorting and probabilistic inference: a machine learning-based approach.* PhD thesis, Kansas State University, 2003.

[18] Asia network. `http://http://www.bnlearn.com/bnrepository/discrete-small.html#asia`. Accessed: 2018-03-29.

[19] Alarm network. `http://www.bnlearn.com/bnrepository/discrete-medium.html#alarm`. Accessed: 2018-02-07.

[20] Hepar2 network. `http://www.bnlearn.com/bnrepository/discrete-large.html#hepar2`. Accessed: 2018-03-29.

[21] Nathalie Jitnah and Ann E Nicholson. Belief network algorithms: A study of performance based on domain characterisation. In *Pacific Rim International Conference on Artificial Intelligence*, pages 168–187. Springer, 1996.

[22] Linda C van der Gaag, Hans L Bodlaender, and Bob Reimink. A distance measure for junction trees. Under construction.