Two new notions of abduction in Bayesian networks

Johan Kwisthout

Radboud University Nijmegen, Institute for Computing and Information Sciences,
P.O. Box 9010, 6500GL Nijmegen, The Netherlands
johank@science.ru.nl

Abstract

Most Probable Explanation and (Partial) MAP are well-known problems in Bayesian networks that correspond to Bayesian or probabilistic inference of the most probable explanation of observed phenomena given full or partial evidence. These problems have been studied extensively, both from a knowledge-engineering starting point (see [10] for an overview) as well as a complexity-theoretic point of view (see [9] for an overview). Algorithms, both exact and approximate, are studied in e.g. [14, 17, 12, 20]. In this paper, we introduce two new notions of abduction-like problems in Bayesian networks, motivated from cognitive science, namely the problem of finding the most *simple* and the most *informative* explanation for a set of variables, given evidence. We define and motivate these problems, show that these problems are computationally intractable in general, but become tractable when some particular constraints are met.

1 Introduction

Abduction (the process of finding a suitable explanation of observed phenomena) is, according to the well-known philosopher Charles Sanders Pierce[1] perhaps the most essential mechanism through which we acquire knowledge or information. While since long obvious in fields like medicine ("which disease is causing these symptoms"), history ("which circumstances led to this uprising"), and the sciences ("what theory explains this behavior"), the problem becomes also more and more important in fields as cognitive science, e.g., in computational models of goal inference [1], action understanding [3], or vision [21].

In Bayesian networks, the abduction problem is often encoded as the problem of determining the joint value assignment to a set of variables in the network which has the highest posterior probability given the observed values of other variables in the network. This problem is known as MPE or (PARTIAL) MAP in the literature. The abduction problem in Bayesian networks (be it formalized as MPE or PARTIAL MAP) is intractable in general, yet can be constrained to make computation tractable [9]. For example, the MPE problem, where the network is bi-partitioned into a set of variables for which the most probable joint value assignment is sought (the *explanation set*) and a set of variables whose values are observed (the *evidence*), can be solved fast if the most probable explanation is very likely (and thus, correspondingly, competing explanations are very improbable) [2]. However, the bi-partition constraint is often not satisfied in practice or may lead to unexpected results, forcing us to solve a PARTIAL MAP problem (where only partial evidence is available, i.e., not all variables outside the explanation set are observed).

The merit of giving results that intuitively better match our notion of abduction (i.e., solving a PARTIAL MAP problem rather than a MPE problem) comes at the price of higher computational demands [15, 13, 9]. Moreover, when used as a computational framework of cognitive tasks in which abduction plays a central role, it is questionable whether PARTIAL MAP is actually the most plausible underlying computational problem in such cognitive tasks. We will give two examples to illustrate this stance.

Example 1.1. *Mr. Jones typically comes to work by train. Today Mr. Jones is late while he has been seen to leave his house at the usual time. One explanation can be that the train is delayed. However, it might also*

[1] See [5] for an overview of Pierce's work on abduction.

be the case that Mr. Jones was the unlucky individual who walked through 11th Street at 8.03 AM and was shot during an armed bank robbery. When trying to explain why Mr. Jones is not at his desk on 8.30 AM, there are a number of variables we might take into account, for example whether he has to change trains. A whole lot of variables are typically not taken into account because they are normally not relevant in most of the cases, for example whether Mr. Jones walked on the left or right pavement in 11th Street. Only in the awkward coincidence that Mr. Jones was in the wrong place at the wrong time they become relevant to explain why he is not at work.

Is it plausible that, in order to determine the most likely cause of Mr. Jones' lateness, we really marginalize over all these unknown–and usually irrelevant–variables in order to infer the most likely explanation? Or do we actually think a train delay is a likely explanation because it explains his absence for the overwhelming majority of values for these variables, save that particular situation where he was walking on the left pavement and passed the bank at 8.03 AM while there was a shooting at that very moment? We will argue that the latter may be a more plausible cognitive model, congruent with recent findings in cognitive science.

An other potential problem when using PARTIAL MAP as the underlying computational model is that the choice of the explanation set is crucial: we seek to find the most probable joint value assignment to that set of variables, i.e., the explanation set is given in the input. The question *what constitutes a good explanation* will be illustrated in the following example.

Example 1.2. *Dr. Brown is examining Mr. Smith. She notices an increased body temperature and Mr. Smith complains of shortness of breath, coughing with phlegm, and pain while breathing. A correct, but hardly informative, explanation of these signs is 'Mr. Smith is ill.' This explanation has a higher probability (say* 0.998*) than the much more informative explanation 'Mr. Smith has pneumonia.' which may have a probability of* 0.96*. The latter explanation of course has more explanatory power at the cost of little probability mass, and is thus preferred over the former although this explanation has a higher probability.*

Note that there are many situation-specific circumstances that may determine whether a more specific explanation is needed. While a general practitioner will need an explanation that is specific enough to successfully describe medication, Mr. Smith's project manager needs only a general explanation why he won't be at his desk for some time. Sometimes it might be costly to determine more specific explanations. Typically there is a trade-off between specificity and probability and the impact of making the *wrong* decision is crucial in determining the probability threshold.

In this paper we introduce two new notions of abduction in Bayesian networks, namely finding the explanation that *assumes as little as possible of the values of other variables that are not observed* (which we will denote as the MOST SIMPLE EXPLANATION problem), and finding the explanation that *carries the most information given some probability threshold* (which we will denote as the MOST INFORMATIVE EXPLANATION problem). After the introduction of some preliminaries in Section 2, the problem variants will be formalized in Section 3 and 4, respectively. We will give complexity results for the problems in general and show under which constraints these problems are feasible. We conclude the paper in Section 5.

2 Preliminaries

A Bayesian or probabilistic network $\mathcal{B}$ is a graphical structure that models a set of stochastic variables, the (in-)dependencies among these variables, and a joint probability distribution over these variables. $\mathcal{B}$ includes a directed acyclic graph $\mathbf{G} = (\mathbf{V}, \mathbf{A})$, modeling the variables and (in-) dependencies in the network, and a set of parameter probabilities Γ in the form of conditional probability tables (CPTs), capturing the strengths of the relationships between the variables. The network models a joint probability distribution $\Pr(\mathbf{V}) = \prod_{i=1}^{n} \Pr(V_i \mid \pi(V_i))$ over its variables, where $\pi(V_i)$ denotes the parents of V_i in $\mathbf{G}$. We will use upper case letters to denote individual nodes in the network, upper case bold letters to denote sets of nodes, lower case letters to denote value assignments to nodes, and lower case bold letters to denote joint value assignments to sets of nodes. Every (posterior) probability of interest in Bayesian networks can be computed using well known lemmas in probability theory, like Bayes' theorem ($\Pr(H \mid E) = \frac{\Pr(E|H)\Pr(H)}{\Pr(E)}$), marginalization ($\Pr(H) = \sum_{g_i} \Pr(H \wedge G = g_i)$), and the factorization property of Bayesian networks ($\Pr(\mathbf{V}) = \prod_{i=1}^{n} \Pr(V_i \mid \pi(V_i))$). More background can be found in textbooks like [14] and [8].

In the remainder, we assume that the reader is familiar with basic concepts of computational complexity theory, such as Turing Machines, the complexity classes P and NP, and NP-completeness proofs. For more background we refer to classical textbooks like [6]. In addition to these basic concepts, to describe

the complexity of various problems we will use the *probabilistic* classes PP and BPP, oracles, and fixed-parameter tractability.

The class PP contains languages L accepted in polynomial time by a *Probabilistic Turing Machine*. Such a machine augments the more traditional non-deterministic Turing Machine with a probability distribution associated with each state transition, e.g., by providing the machine with a tape, randomly filled with symbols [7]. If all choice points are binary and the probability of each transition is $\frac{1}{2}$, then the *majority* of the computation paths accept a string s if and only if $s \in L$. This majority, however, is not fixed and may (exponentially) depend on the input, e.g., a problem in PP may accept 'yes'-instances with size n with probability $\frac{1}{2} + \frac{1}{2^n}$. This makes problems in PP intractable in general, in contrast to the related complexity class BPP which is associated with problems which allow for efficient randomized computation[2]. BPP accepts 'yes'-inputs with a *bounded*[3] majority (say $\frac{3}{4}$). This means we can amplify the probability of a correct answer arbitrary close to one by running the algorithm a polynomial amount of times and taking a majority vote on the outcome. This approach fails for unbounded majorities as $\frac{1}{2} + \frac{1}{2^n}$ as allowed by the class PP: here an exponential number of simulations (with respect to the input size) is needed to meet a constant threshold on the probability of answering correctly.

The canonical PP-complete problem is MAJSAT: given a Boolean formula ϕ, does the majority of the truth instantiations satisfy ϕ? Indeed it is easily shown that MAJSAT encodes the NP-complete SATISFIABILITY problem: take a formula ϕ with n variables and construct $\psi = \phi \vee x_{n+1}$. Now, the majority of truth assignments satisfy ψ if and only if ϕ is satisfiable, thus $\mathsf{NP} \subseteq \mathsf{PP}$. In probabilistic networks, the INFERENCE problem of determining whether the probability $\Pr(\mathbf{X} = \mathbf{x}) > q$ is PP-complete.

A Turing Machine $\mathcal{M}$ has *oracle access* to languages in the class A, denoted as $\mathcal{M}^{\mathsf{A}}$, if it can "query the oracle" in one state transition, i.e., in $\mathcal{O}(1)$. We can regard the oracle as a 'black box' that can answer membership queries in constant time. For example, $\mathsf{NP}^{\mathsf{PP}}$ is defined as the class of languages which are decidable in polynomial time on a non-deterministic Turing Machine with access to an oracle deciding problems in PP. Informally, computational problems related to probabilistic networks that are in $\mathsf{NP}^{\mathsf{PP}}$ typically combine some sort of *selecting* with *probabilistic inference*. The canonical $\mathsf{NP}^{\mathsf{PP}}$-complete satisfiability variant is E-MAJSAT: given a formula ϕ with variable sets $x_1 \ldots x_k$ and $x_{k+1} \ldots x_n$, is there an instantiation to $x_1 \ldots x_k$ such that the majority of the instantiations to $x_{k+1} \ldots x_n$ satisfy ϕ?

A problem is called *fixed parameter tractable* [4] for a parameter l if it can be solved in time, exponential *only* in l and polynomial in the input size n, i.e., when the running time is $\mathcal{O}(f(l) \cdot n^c)$ for an arbitrary function f and a constant c, independent of n. In practice, this means that problem instances can be solved efficiently, even when the problem is NP-hard in general, if l is known to be small. If an NP-hard problem Π is *fixed parameter tractable* for a parameter l then l is denoted a *source of complexity* [18] of Π: bounding l renders the problem tractable, whereas leaving l unbounded ensures intractability under usual complexity-theoretic assumptions like $\mathsf{P} \neq \mathsf{NP}$.

To conclude this preliminaries section, we define decision variants of the MPE and PARTIAL MAP problems as follows[4]:

MPE
Instance: A probabilistic network $\mathcal{B} = (\mathbf{G}, \Gamma)$, where $\mathbf{V}$ is partitioned into a set of evidence nodes $\mathbf{E}$ with a joint value assignment $\mathbf{e}$, and an explanation set $\mathbf{M}$; a rational number $0 \leq q < 1$.
Question: Is there a joint value assignment $\mathbf{m}$ to the nodes in $\mathbf{M}$ with evidence $\mathbf{e}$ with probability $\Pr(\mathbf{m}, \mathbf{e}) > q$?

PARTIAL MAP
Instance: A probabilistic network $\mathcal{B} = (\mathbf{G}, \Gamma)$, where $\mathbf{V}$ is partitioned into a set of evidence nodes $\mathbf{E}$ with a joint value assignment $\mathbf{e}$, an explanation set $\mathbf{M}$, and a set of intermediate variables $\mathbf{I}$; a rational number $0 \leq q < 1$.
Question: Is there a joint value assignment $\mathbf{m}$ to the nodes in $\mathbf{M}$ with evidence $\mathbf{e}$ with probability $\Pr(\mathbf{m}, \mathbf{e}) > q$?

MPE has been proven NP-complete by Shimony [15] and is fixed parameter tractable in graphs with bounded treewidth and when the most probable explanation has a high probability [17, 2]. PARTIAL MAP

[2] Indeed BPP is assumed to be equal to P by many complexity theorists.

[3] To be precise: polynomially bounded in the input size.

[4] Observe that we consistently use marginal probabilities $\Pr(\mathbf{m}, \mathbf{e})$ here rather than conditional probabilities $\Pr(\mathbf{m} \mid \mathbf{e})$. This is due to the different complexity results for MPE defined with marginal and conditional probabilities [9].

has been proven to be $\mathsf{NP}^{\mathsf{PP}}$-complete by Park and Darwich [13]. In contrast with MPE, PARTIAL MAP remains intractable when the network has bounded treewidth or when the most probable explanation has a high probability, yet is fixed parameter tractable when *both* conditions are met [9].

3 Most Simple Explanations

In this section we will introduce Most Simple Explanations as an alternative notion of Bayesian abduction with partial evidence. Our approach is partially inspired by the *Decision by Sampling* model of Stewart et al. [16], where the authors *"...do not assume that people have stable, long-term internal scales along which they represent value, probability, temporal duration, or any other magnitudes."*, but, *"...assume that people can only sample items from memory and then judge whether a target value is larger or smaller than these items."* [16, p. 2]. This was recently confirmed by Vul et al. [19]: few samples from a probability distribution are often adequate to make reasonably well decisions.

Based on the intuitive observation that we often do not take information into account (for example, whether Mr. Jones was walking on the left or right pavement) unless we have reason to believe that this might influence the most probable explanation, we suggest that the PARTIAL MAP problem, in which marginalization of the intermediate variables takes place, might not be the formalism that describes human behavior best in such circumstances. Instead, we suggest that one does not choose the *most probable* explanation—based on a marginalization of potentially many intermediate variables—but one rather picks the *most simple* explanation, i.e., the explanation that assumes as little as possible of the values of the intermediate variables. This can be seen as an implementation of *Occam's razor* within the context of Bayesian abduction. We formalize this as follows: select the joint value assignment to the explanation set that is the *most probable explanation* for the *majority* of joint value assignments to the intermediate variables.

To stick with the *Mr. Jones*-example: for almost every joint value assignment to the intermediate variables Time-Of-Passing, Shooting-at-Bank, Pavement-walking-on, Bus-17-riding-by, Color-of-Mr.Jones'-suit etcetera, the explanation "Train-is-delayed = TRUE" would be the most probable explanation of Mr. Jones' absence at work. That would be the most *simple* explanation assuming in general no particular value assignments to the color of Mr. Jones' suit or whether bus 17 or 29 was riding by. Only for a particular combination of circumstances, the most probable explanation would shift to "Mr.Jones-got-shot-on his-way-to-work = TRUE". Thus, when picking a few arbitrary *samples* out of this vast domain space, chances are high that all of them have "Train-is-delayed = TRUE" as the most probable explanation.

Typically, such a decision might be difficult to make when there are multiple competing alternatives. In such cases, apart from solving MPE problems, we may have a difficult task in deciding upon the majority. This is reflected in the computational complexity result of MOST SIMPLE EXPLANATION.

3.1 Computational Complexity

We will prove that MOST SIMPLE EXPLANATION is $\mathsf{NP}^{\mathsf{PP}}$-complete, and thus resides in the same complexity class as PARTIAL MAP. First we give a decision variant of the MOST SIMPLE EXPLANATION problem.

MOST SIMPLE EXPLANATION
Instance: A probabilistic network $\mathcal{B} = (\mathbf{G}, \Gamma)$, where $\mathbf{V}$ is partitioned into a set of evidence nodes $\mathbf{E}$ with a joint value assignment $\mathbf{e}$, an explanation set $\mathbf{M}$, and intermediate variables $\mathbf{I}$; a rational number $0 \leq q < 1$.
Question: Is there a joint value assignment $\mathbf{m}$ to the nodes in $\mathbf{M}$ such that for the majority of instantiations $\mathbf{i}$ to $\mathbf{I}$, $\Pr(\mathbf{m}, \mathbf{i}, \mathbf{e}) > q$?

We construct a probabilistic network $\mathcal{B}_\phi$ from an E-MAJSAT instance $(\phi, \mathbf{E}, \mathbf{M})$, where ϕ is a Boolean formula with n variables, partitioned into sets $\mathbf{E} = x_1 \ldots x_k$ and $\mathbf{M} = x_{k+1} \ldots x_n$ for some number $0 \leq k \leq n$. For each propositional variable x_i in ϕ, a binary stochastic variable X_i is added to $\mathcal{B}_\phi$, with possible values TRUE and FALSE and a uniform probability distribution. These stochastic variables in $\mathcal{B}_\phi$ are bi-partitioned into sets $\mathbf{X_E}$ and $\mathbf{X_M}$ according to the partition of ϕ. For each logical operator in ϕ, an additional binary variable in $\mathcal{B}_\phi$ is introduced, whose parents are the variables that correspond to the input of the operator, and whose conditional probability table is equal to the truth table of that operator. For example, the value TRUE of a stochastic variable mimicking an *and*-operator would have a conditional probability of 1 if and only if both its parents have the value TRUE, and 0 otherwise. The variable associated with the

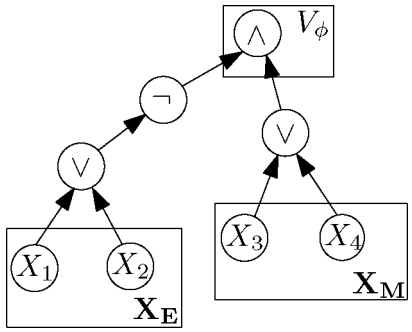

Figure 1: Example of construction of a probabilistic network $\mathcal{B}_{\phi_{\text{ex}}}$ for the Boolean formula $\phi_{\text{ex}} = \neg(x_1 \vee x_2) \wedge (x_3 \vee x_4)$ with $\mathbf{E} = \{x_1, x_2\}$ and $\mathbf{M} = \{x_3, x_4\}$

top-level operator in ϕ is denoted as V_ϕ. Figure 1 shows the graphical structure of the probabilistic network constructed for the E-MAJSAT instance $(\phi_{\text{ex}}, \mathbf{E}, \mathbf{M})$, where $\phi_{\text{ex}} = \neg(x_1 \vee x_2) \wedge (x_3 \vee x_4)$, $\mathbf{E} = \{x_1, x_2\}$, and $\mathbf{M} = \{x_3, x_4\}$.

Theorem 3.1. MOST SIMPLE EXPLANATION *is* $\mathsf{NP}^{\mathsf{PP}}$*-complete.*

Proof. Membership in $\mathsf{NP}^{\mathsf{PP}}$ follows from the following algorithm: non-deterministically guess a value assignment $\mathbf{m}$, count the joint value assignments $\mathbf{i}$ to $\mathbf{I}$ such that $\Pr(\mathbf{m}, \mathbf{i}, \mathbf{e}) > q$ and decide whether this is a majority. This can be done using a non-deterministic algorithm with an oracle for problems in $\#\mathsf{P}$, since deciding whether there *exists* a joint value assignment $\mathbf{i}$ such that $\Pr(\mathbf{m}, \mathbf{i}, \mathbf{e}) > q$ is in NP, and thus the counting problem is in $\#\mathsf{P}$ by definition. Hence the problem is in $\mathsf{NP}^{\#\mathsf{P}} = \mathsf{NP}^{\mathsf{PP}}$ since $\#\mathsf{P}$ is Turing reducible to PP.

To prove $\mathsf{NP}^{\mathsf{PP}}$-hardness, we reduce MOST SIMPLE EXPLANATION from E-MAJSAT. We fix $q = \frac{1}{2}$. Let $(\phi, \mathbf{E}, \mathbf{M})$ be an instance of E-MAJSAT and let $\mathcal{B}_\phi$ be the network constructed from that instance as shown above. If there exists a satisfying solution to $(\phi, \mathbf{E}, \mathbf{M})$, then there is a truth instantiation to $\mathbf{E}$ such that for the majority of truth instantiations to $\mathbf{M}$, ϕ is satisfied. For any particular corresponding joint value assignment $\mathbf{x_E}$ to $\mathbf{X_E}$ in $\mathcal{B}_\phi$, $\Pr(V_\phi = \text{TRUE}, \mathbf{x_E}, \mathbf{x_M}) = 1$ for a particular joint value assignment $\mathbf{x_M}$ to $\mathbf{X_M}$, if and only if the corresponding truth instantiation to $\mathbf{M}$ satisfies ϕ, and 0 otherwise. Correspondingly, if $(\phi, \mathbf{E}, \mathbf{M})$ is satisfiable, then $\Pr(V_\phi = \text{TRUE}, \mathbf{x_E}, \mathbf{x_M}) = 1$ for the majority of joint value assignments $\mathbf{x_M}$ to $\mathbf{X_M}$, and thus the MOST SIMPLE EXPLANATION instance $(\mathcal{B}_\phi, V_\phi = \text{TRUE}, \mathbf{X_E})$ is satisfied.

Now assume $(\mathcal{B}_\phi, V_\phi = \text{TRUE}, \mathbf{X_E}, q = \frac{1}{2})$ is satisfied, i.e., there exists a joint value assignment $\mathbf{x_E}$ to $\mathbf{X_E}$ such that for the majority of joint value assignments $\mathbf{x_M}$ to $\mathbf{X_M}$, $\Pr(V_\phi = \text{TRUE}, \mathbf{x_E}, \mathbf{x_M}) > \frac{1}{2}$. Then there exists a truth instantiation to $\mathbf{E}$ (namely the truth instantiation that corresponds to the joint value assignment $\mathbf{x_E}$) such that the majority of truth instantiations to $\mathbf{M}$ satisfies ϕ, and thus the E-MAJSAT instance $(\phi, \mathbf{E}, \mathbf{M})$ is satisfied. Since the reduction can obviously be done in polynomial time, this proves that MOST SIMPLE EXPLANATION is $\mathsf{NP}^{\mathsf{PP}}$-complete. □

Note that MOST SIMPLE EXPLANATION remains $\mathsf{NP}^{\mathsf{PP}}$-complete, even if all variables are binary, each node has at most two incoming arcs, and the probability distribution of each variable is either uniform or deterministic. If $\mathbf{I} = \emptyset$ then MOST SIMPLE EXPLANATION degenerates to MPE and hence is NP-complete. If $\mathbf{M} = \emptyset$ then MOST SIMPLE EXPLANATION remains $\#\mathsf{P}$-complete via a reduction of #SAT, using a similar construction as above, the proof of which is omitted for reasons of space.

3.2 A Randomized Algorithm

A deterministic algorithm for MOST SIMPLE EXPLANATION might iterate over $\mathbf{m}$ and count the joint value assignments $\mathbf{i}$ to $\mathbf{I}$ for which $\text{argmax}_{\mathbf{M}} \Pr(\mathbf{M}, \mathbf{i}, \mathbf{e}) = \mathbf{m}$, i.e., for which $\mathbf{m}$ is the most probable explanation to $\mathbf{M}$, and decide upon the majority. However, this algorithm trivially runs in exponential time, even if the treewidth of the network is bounded: there are exponentially many joint value assignments to $\mathbf{M} \cup \mathbf{I}$. We present a randomized algorithm that performs much better in many circumstances.

for $n = 1$ to N **do**
  Choose $\mathbf{i}$ at random
  Determine $\mathbf{m} = \text{argmax}_{\mathbf{M}} \Pr(\mathbf{M}, \mathbf{i}, \mathbf{e})$
  Count the joint value assignments $(\mathbf{m}, \mathbf{i})$
end for
Decide upon the majority and output $\mathbf{m_{maj}}$

This randomized algorithm repeatedly picks a joint value assignment to $\mathbf{I}$ at random, determines the most probable explanation, and at the end decides upon the majority. Due to its stochastic nature, this algorithm is not guaranteed to give correct answers all the time. However, the error margin ϵ can be made sufficiently low by choosing N large enough, where the threshold value of N can be computed using the *Chernoff bound*: $N \geq \frac{1}{(p-\frac{1}{2})^2} \ln \frac{1}{\sqrt{\epsilon}}$. Assume we require an error margin of less than 0.1, then the number of repeats depends on the probability of picking a joint value assignment $\mathbf{i}$ for which $\mathbf{m_{maj}}$ is the most probable explanation[5]. If this probability is high (say $p = 0.85$), then N can be fairly low ($N \geq 10$), however, if the probability distribution is almost uniform then an exponential number of repetitions is needed.

If the majority is bounded (i.e., larger than a particular fixed threshold) we thus need only polynomially many repetitions to obtain any constant error rate, and thus the MOST SIMPLE EXPLANATION problem is in $\mathsf{NP}^{\mathsf{BPP}}$. When determining the most probable explanation is easy—in particular, when the treewidth of $\mathcal{B}$ is low—the algorithm thus runs in polynomial time. Since the treewidth of $\mathcal{B}$ is independent of the choice of $\mathbf{m}$ and $\mathbf{i}$, MOST SIMPLE EXPLANATION can be decided, with a small possibility of error, in polynomial time when the treewidth is low and the majority is sufficiently large.

This sampling approach is conceptually different from Monte-Carlo-like approximations of probabilistic inference (as might be used in an approximation algorithm for PARTIAL MAP), where the marginalization process is approximated by taking samples of the intermediate variables according to their prior probability distribution. We argue, in line with the *Decision by Sampling* model as discussed in the beginning of this section, that it may be plausible that one *does not marginalize at all* to obtain most probable explanations, but often bases ones judgement on what appears to hold for the vast majority of intermediate variables.

4 Most Informative Explanations

In this section we introduce Most Informative Explanations that enhance the traditional interpretation of Bayesian abduction with a notion of *information-richness*. This alternative interpretation is based on the intuitive idea that the explanation set (and thus the specificity of the most probable explanation) is not always fixed, but can be flexibly adjusted to be more general or more specific, depending on the probability of the most probable explanation for a particular size of the explanation set. Since we marginalize over variables outside both the evidence and the explanation set, the probability of the most probable explanation is by definition larger when the explanation set becomes smaller, however the information carried in that explanation is less specific, as was illustrated by the *Dr. Brown*-example in the introduction. In this section, we formalise this problem, analyse its computational complexity and show under what circumstances the problem becomes tractable.

4.1 Computational Complexity

The decision version of MOST INFORMATIVE EXPLANATION will be formulated as follows.

MOST INFORMATIVE EXPLANATION
Instance: A probabilistic network $\mathcal{B} = (\mathbf{G}, \Gamma)$, where $\mathbf{V}$ is partitioned into a set of evidence nodes $\mathbf{E}$ with a joint value assignment $\mathbf{e}$, an explanation set $\mathbf{M}$, and intermediate variables $\mathbf{I}$; a rational number $0 \leq q < 1$ and a natural number $0 \leq l \leq |\mathbf{M}|$.
Question: Is there a joint value assignment $\mathbf{m_l}$ to a subset of length l of the nodes in $\mathbf{M}$ such that $\Pr(\mathbf{m_l}, \mathbf{e}) > q$?

Theorem 4.1. MOST INFORMATIVE EXPLANATION *is* $\mathsf{NP}^{\mathsf{PP}}$*-complete.*

[5]For ease of exposure, we assume here that there are only two competing joint value assignments, and consequently, this probability is by definition larger than $\frac{1}{2}$. If this assumption is violated, the standard Chernoff bound can not be used to compute N for a given error margin.

Proof. Membership can be shown by non-deterministically guessing $\mathbf{m_1}$ and deciding, using the PP oracle, whether $\Pr(\mathbf{m_1}, \mathbf{e}) > q$. $\mathsf{NP}^{\mathsf{PP}}$-hardness follows since MOST INFORMATIVE EXPLANATION has PARTIAL MAP as a special case: take $l = |\mathbf{M}|$. □

If $l = 0$ then MOST INFORMATIVE EXPLANATION degenerates to INFERENCE. If $l = |\mathbf{M}|$ then MOST INFORMATIVE EXPLANATION degenerates to PARTIAL MAP, and if in addition $\mathbf{I} = \emptyset$ then MOST INFORMATIVE EXPLANATION degenerates to MPE. Furthermore, MOST INFORMATIVE EXPLANATION inherits the constraints and inapproximability results of PARTIAL MAP [9].

While PARTIAL MAP is fixed parameter tractable for $(1 - p, \mathrm{tw})$, i.e., PARTIAL MAP can be solved fast when the treewidth of the network is bounded *and* the most probable explanation has a high probability, this may not hold for MOST INFORMATIVE EXPLANATION since we need to choose $\mathbf{m_1}$ which by itself is a source of complexity. However, Bodlaenders algorithm for PARTIAL MAP, which is fixed parameter tractable for $(1-p, \mathrm{tw})$ [9], can be adjusted by branching on each l-sized subset of the explanation set. Since there are $\binom{l}{|\mathbf{M}|}$ many such subsets, the number of subsets to consider is low when either l is low (the subset is small) or $|M| - l$ is low (the subset is large), but increases with l approaching $\frac{|M|}{2}$. Therefore, MOST INFORMATIVE EXPLANATION is fixed parameter tractable for both $(1 - p, \mathrm{tw}, l)$ and $(1 - p, \mathrm{tw}, |M| - l)$. Since MOST INFORMATIVE EXPLANATION is a generalization of both PARTIAL MAP (for $l = |\mathbf{M}|$) and INFERENCE (for $l = 0$) it follows that MOST INFORMATIVE EXPLANATION remains intractable for the sets of parameters $(1 - p, l)$, $(1 - p, |M| - l)$ and $(\mathrm{tw}, |M| - l)$, but is fixed parameter tractable for the set of parameters (tw, l). We do not know whether MOST INFORMATIVE EXPLANATION is fixed parameter tractable for the subset $(1 - p, tw)$.

5 Conclusion

In this paper we proposed two alternative notions of abduction in probabilistic networks, based on observations and intuitions in cognitive science. MOST SIMPLE EXPLANATION focuses on the problem of finding a *parsimonious* explanation, MOST SIMPLE EXPLANATION focuses on the problem of finding an *informative* explanation.

We proved $\mathsf{NP}^{\mathsf{PP}}$-hardness of both problem variants, placing them in the same complexity class as PARTIAL MAP, however, with different fixed parameter tractability results. In particular, MOST SIMPLE EXPLANATION is fixed parameter tractable when the treewidth of the network is bounded and there is a bounded majority of the joint value assignments to $\mathbf{I}$ for which a particular explanation $\mathbf{m}$ is the most probable explanation[6]. Informally, one can envisage some sort of sampling: if $\mathbf{m}$ is the most probable explanation for almost all joint value asignments to the set of intermediate varables $\mathbf{I}$, then only a few samples will suffice to say so with a relatively healthy error margin[7]. We see MOST INFORMATIVE EXPLANATION as an extension, rather than an alternative, to PARTIAL MAP. The formulation of the problem variant with a flexible size of the explanation set may serve as a starting point for more thorough investigations regarding the computational model of the cognitive task of deciding *what* to explain.

We do not put forward claims on whether MOST SIMPLE EXPLANATION or PARTIAL MAP (or yet an other notion of abduction) better represent the cognitive processes that are modeled by abductive Bayesian reasoning. One highly speculative hypothesis might be that a 'sampling' approach (i.e., MOST SIMPLE EXPLANATION) is used when the set of intermediate variables $\mathbf{I}$ does not contain variables whose values may have a non-trivial impact on the most probable explanation, and a 'marginalization' approach (i.e., PARTIAL MAP) is used when they do. Further elaboration, however, is beyond the scope of this paper.

Acknowledgements

The author is supported by the OCTOPUS project under the responsibility of the Embedded Systems Institute. This project is partially supported by the Netherlands Ministry of Economic Affairs under the Embedded Systems Institute program. The author wishes to thank Iris van Rooij, Pim Haselager, Todd Wareham, Moritz Müller and Hans Bodlaender for valuable discussion regarding various ideas put forward in this paper and comments on earlier drafts.

[6] This is a slight abuse of the definitions of fixed parameter tractability, as the boundedness of the majority is strictly speaken not a parameter; however, the reader is assumed to understand the intuition behind this claim.

[7] Note that this is an explanation at Marr's [11] *algorithmic*, rather than *computational*, level since there may be alternative fast algorithms that do not use sampling at all.

References

[1] C. L. Baker, R. Saxe, and J. B. Tenenbaum. Action understanding as inverse planning. *Cognition*, 113(3):329–349, 2009.

[2] H. L. Bodlaender, F. van den Eijkhof, and L. C. van der Gaag. On the complexity of the MPA problem in probabilistic networks. In F. van Harmelen, editor, *Proceedings of the 15th European Conference on Artificial Intelligence*, pages 675–679, 2002.

[3] R. H. Cuijpers, H. T. van Schie, M. Koppen, W. Erlhagen, and H. Bekkering. Neural networks. *Goals and means in action observation: A computational approach*, 19:311–322, 2006.

[4] R. G. Downey and M. R. Fellows. *Parameterized Complexity*. Springer Verlag, Berlin, 1999.

[5] K. T. Fann. *Peirce's theory of abduction*. Springer, Berlin, 1970.

[6] M. R. Garey and D. S. Johnson. *Computers and Intractability. A Guide to the Theory of NP-Completeness*. W. H. Freeman and Co., San Francisco, CA, 1979.

[7] J. T. Gill. Computational complexity of Probabilistic Turing Machines. *SIAM Journal of Computing*, 6(4):675–695, 1977.

[8] F. V. Jensen and T. D Nielsen. *Bayesian Networks and Decision Graphs*. Springer Verlag, New York, NY, second edition, 2007.

[9] J. Kwisthout. Most Probable Explanations in Bayesian networks: complexity and tractability. Technical Report ICIS–R10001, Radboud University Nijmegen, 2010.

[10] C. Lacave and F. J. Díez. A review of explanation methods for Bayesian networks. *The Knowledge Engineering Review*, 17(2):107–127, 2002.

[11] D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Freeman, New York, NY, 1982.

[12] D. Nilsson. An efficient algorithm for finding the M most probable configurations in probabilistic expert systems. *Statistics and Computing*, 8:159–173, 1998.

[13] J. D. Park and A. Darwiche. Complexity results and approximation settings for MAP explanations. *Journal of Artificial Intelligence Research*, 21:101–133, 2004.

[14] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, Palo Alto, CA, 1988.

[15] S. E. Shimony. Finding MAPs for belief networks is NP-hard. *Artificial Intelligence*, 68(2):399–410, 1994.

[16] N. Stewart, N. Chater, and G. D. A. Brown. Decision by sampling. *Cognitive Psychology*, 53:1–26, 2006.

[17] B. K. Sy. Reasoning MPE to multiply connected belief networks using message-passing. In P. Rosenbloom and P. Szolovits, editors, *Proceedings of the Tenth National Conference on Artificial Intelligence*, pages 570–576. AAAI Press, Arlington, Va, 1992.

[18] I. van Rooij and T. Wareham. Parameterized complexity in cognitive modeling: Foundations, applications and opportunities. *The Computer Journal*, 51(3):385–404, 2008.

[19] E. Vul, N. D. Goodman, T. L. Griffiths, and J. B. Tenenbaum. One and done? optimal decisions from very few samples. In *31st Annual Meeting of the Cognitive Science Society*, 2009.

[20] C. Yuan, T. Lu, and M. Druzdzel. Annealed MAP. In J. Halpern, editor, *Proceedings of the Twentieth Conference Annual Conference on Uncertainty in Artificial Intelligence*, pages 628–635. AUAI Press, Arlington, Va, 2004.

[21] A. Yuille and D. Kersten. Vision as Bayesian inference: analysis by synthesis? *TRENDS in Cognitive Sciences*, 10(7):301–308, 2006.