Most Inforbable Explanations: Finding Explanations in Bayesian Networks that are both Probable *and* Informative

Johan Kwisthout

Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, Montessorilaan 3, 6525 HR Nijmegen, The Netherlands, j.kwisthout@donders.ru.nl

Abstract. The problems of generating candidate hypotheses and inferring the best hypothesis out of this set are typically seen as two distinct aspects of the more general problem of non-demonstrative inference or abduction. In the context of Bayesian networks the latter problem (computing most probable explanations) is well understood, while the former problem is typically left as an exercise to the modeler. In other words, the candidate hypotheses are pre-selected and hard-coded. In reality, however, non-demonstrative inference is rather an interactive process, switching between hypothesis generation, inference to the best explanation, evidence gathering and deciding which information is relevant. In this paper we will discuss a possible computational formalization of finding an explanation which is both probable and as informative as possible, thereby combining (at least some aspects of) both the 'hypotheses-generating' and 'inference' steps of the abduction process. The computational complexity of this formal problem, denoted Most Inforbable Explanation, is then established and some problem parameters are investigated in order to get a deeper understanding of what makes this problem intractable in general, and under which circumstances the problem becomes tractable.

1 Introduction

Inference to the best explanation is a well-known and well-studied computational problem in Bayesian networks. When "best" is operationalized as "most probable" (as is typically the case in the Bayesian network community, but see, e.g., [11] for alternative notions) it is commonly known as MAP[1]: given a partition of a Bayesian network into an *evidence* set with observed variables, a set of *explanation* variables which together constitute candidate hypotheses, and a set of *intermediate* variables that fall in neither category, compute the most probable joint value assignment to the explanation variables. This computational problem

[1] Also *Partial* or *Marginal* MAP to distinguish the problem from the more constrained MPE problem, in which the variables of the graph are bi-partitioned in evidence variables and hypothesis variables and no marginalization over other variables is needed.

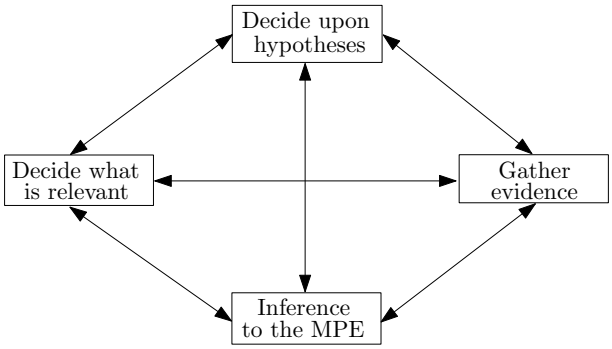

Fig. 1. In everyday problem solving the selection of hypotheses, determining upon relevant information, gathering evidence, and inference to the most probable explanation are concurrent (rather than sequential) and highly connected sub-tasks of the broader *abduction* problem

has been studied from an engineering [17] and computational complexity [5, 14, 22] point of view, and exact and approximate algorithms for MAP are available in abundance [3, 6, 7, 20–22, 25]. However, the *abduction* or *non-demonstrative inference problem* is broader and more complex than 'merely' solving a MAP problem. It is a heavily intertwined combination of deciding which are the relevant variables, deciding upon candidate hypotheses, evidence gathering, and inference to the most probable explanation (Fig. 1).

Clinical examination (i.e., diagnosing the patient) is an excellent example of such an abduction process, consisting of hypothesis generation, obtaining evidence, evaluating hypotheses, and determining throughout this process what of all the available information is relevant to diagnosing (and preferably curing) the patient; see, e.g., [19] and in particular the highly illustrative case study on page 26-27. Some observations and findings may not be relevant to the diagnosis. The clinician needs to decide which are to be taken into account and which are not. Often, symptoms and signs come in patterns; for example, polyuria, polydipsia, and polyphagia are well known symptoms for diabetes mellitus. Clustering or lumping such observations may benefit hypothesis generation towards a diagnosis. On the other hand, the clinician may miss important aspects in doing so: There is a high probability that orthostatic hypotension is caused by vomiting and diarrhea. Thus, they could be lumped together as cause and effect. In so doing, however, the clinician is at risk of excluding a completely separate and important problem, namely, extracellular volume depletion.

During this process, initial hypotheses are generated and evidence is gathered and judged. Based on the evidence and the posterior probabilities of these initial hypotheses, additional evidence may be gathered and the hypotheses may be further refined, eventually leading to a diagnosis and possibly a treatment procedure. These "real world" aspects of abduction problems, as illustrated in

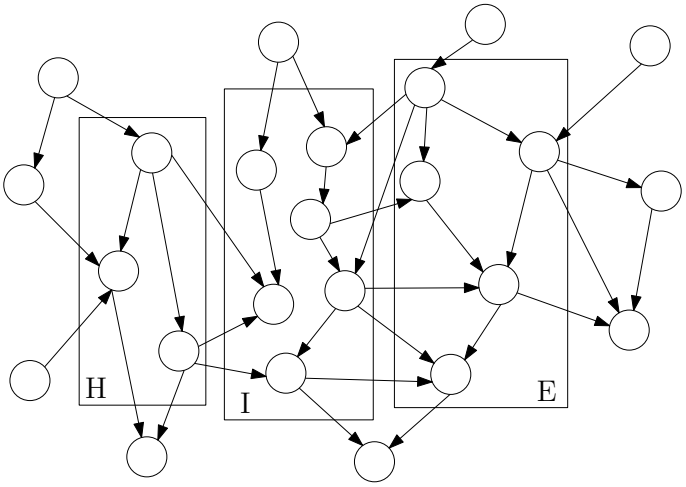

Fig. 2. Partitioning the domain model into hypotheses variables **H**, evidence variables **E**, intermediate variables **I**, and irrelevant or "outside" variables that are not part of the model can be graphically depicted as *establishing boundaries* ("drawing boxes") within a knowledge structure

the above example, are typically not part of the computational problem: they are 'left to the modeler'. Instinctively, this modeling process can be seen as *establishing boundaries* in a knowledge structure such as a Bayesian network (Fig. 2). In this process, numerous decisions need to be made, such as which nodes in the knowledge structure can be dismissed as being irrelevant to the goal or how detailed the explanation should be. These choices are driven by the goal of the abduction process: what counts as a candidate hypotheses or as a relevant variables is determined by what we seek to explain; see for example [8] and [18, Ch. 3, and the references therein].

1.1 Granularity of Explanations

A correct, but hardly informative, explanation of the signs "shortness of breath, coughing with phlegm, and pain while breathing" will be "patient-X is ill". This explanation has (by definition) a higher probability (say 0.95) than the much more informative explanation "patient-X has pneumonia" (say 0.8). The latter explanation of course has more explanatory power at the cost of little probability mass, and thus will, in general, be preferred over the former although this explanation has a higher probability.

This trade off between information and probability is known as the Inverse Relationship Principe [1]: the more *specific* an explanation is, the lower its probability will be. From a mathematical point of view, this may be trivial: surely, $\Pr(A) \leq \Pr(B)$ if $A \subseteq B$. However, in practical situations, there can be 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, a project manager needs only a general explanation why one of her team members won't be at his desk for some time. Sometimes it might be costly or impractical to determine more specific explanations. The impact of making the *wrong* decision may be crucial in determining the probability threshold; what risks are we willing to accept?

In this paper, we seek to combine two aspects of the abduction problem into one computational formalism: choosing what to explain (and at which granularity) and inference to the most probable explanation. This computational problem of seeking an explanation which is both *informative* enough for our means and has a high enough *probability* is denoted as the MOST INFORBABLE EXPLANATION problem to emphasize the trade off between informativeness and probability. The remainder of this paper is structured as follows. In the next section we will offer some needed preliminaries on Bayesian networks and computational complexity theory. In Section 3 we formally define MOST INFORBABLE EXPLANATION. We discuss the computational complexity of a decision variant of MOST INFORBABLE EXPLANATION in Section 4. In Section 5 we conclude the paper.

2 Preliminaries

In this section, we give a short overview of a number of concepts from Bayesian networks, graph theory, and complexity theory, in particular definitions of probabilistic networks and treewidth, some background on complexity classes defined by Probabilistic Turing Machines and oracles, and fixed-parameter tractability. For a more thorough discussion of these concepts, the reader is referred to textbooks like [9, 10, 12, 23].

2.1 Bayesian Networks

A Bayesian or probabilistic network $\mathcal{B} = (\mathbf{G}_\mathcal{B}, \Pr)$ is a graphical structure that models a set of stochastic variables, the conditional independences among these variables, and a joint probability distribution over these variables. $\mathcal{B}$ includes a directed acyclic graph $\mathbf{G}_\mathcal{B} = (\mathbf{V}, \mathbf{A})$, modeling the variables and conditional independences in the network, and a set of parameter probabilities Pr 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 $\mathbf{G}_\mathcal{B}$. 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.

One of the key computational problems in Bayesian networks is the problem to find the most probable explanation for a set of observations, i.e., the joint value assignment to a designated set of variables that has highest posterior probability

given the observed variables in the network. If the network is bi-partitioned into explanation variables and evidence variables this problem is known as Most Probable Explanation, however, in practice there will often be variables that are neither observed nor to be explained; for example, variables that influence the posterior probability distribution but whose value is impractical or even impossible to observe. In that case, the problem is denoted (Partial) MAP (or Marginal MAP, to emphasize that we need to marginalize over the unobserved variables); the decision variant of this problem is defined as follows:

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

MAP is NP-hard under a wide range of constraints, both to compute exact and to approximate [22, 14, 5, 15].

An important structural property of a probabilistic network is its *treewidth*. Treewidth is a graph-theoretical concept, which can be loosely described as a measure on the *'localness'* of the dependencies in the network: when the variables tend to be clustered in small groups with few connections between groups, treewidth is typically low, whereas treewidth tends to be high if the connections between variables are scattered all over the network. Formally, the treewidth of a Bayesian network $\mathcal{B}$ is defined as the minimum width over all tree-decompositions of triangulations of the moralization $\mathbf{G}_{\mathcal{B}}^{\mathbf{M}}$ of the network [24]. Treewidth plays an important role in the complexity analysis of Bayesian networks, as many otherwise intractable computational problems become tractable when the treewidth of the network is bounded.

2.2 Computational Complexity Theory

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. In addition to these basic concepts, to describe the complexity of various problems we will use the *probabilistic* class PP, oracles, and some aspects from parameterized complexity theory.

The class PP contains languages L accepted in polynomial time by a *Probabilistic Turing Machine*. Such a machine augments the more traditional nondeterministic Turing Machine with a probability distribution associated with each state transition. Acceptance of an input x is defined as follows: the probability of arriving in an *accept state* is strictly larger than $\frac{1}{2}$ if and only if $x \in L$. This probability of acceptance, however, is not fixed and may (exponentially) depend on the input, e.g., a problem in PP may accept 'yes'-instances with size $|x|$ with probability $\frac{1}{2} + \frac{1}{2^{|x|}}$. PP-complete problems are considered to be intractable. The canonical PP-complete problem is MajSAT: given a Boolean formula ϕ, does the majority of the truth assignments satisfy ϕ? In Bayesian networks, the canonical

problem of determining whether the probability $\Pr(\mathbf{H} = \mathbf{h} \mid \mathbf{E} = \mathbf{e}) > q$ for a given rational q (known as the INFERENCE problem) is PP-complete [4, 13].

A Turing Machine $\mathcal{M}$ has *oracle access* to languages in the class C, denoted as $\mathcal{M}^{\mathsf{C}}$, 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.

Sometimes problems are intractable (i.e., NP-hard) in general, but become tractable if some *parameters* of the problem can be assumed to be small. Informally, a problem is called fixed-parameter tractable for a parameter k (or a set $\{k_1, \ldots, k_n\}$ of parameters) if it can be solved in time, exponential *only* in k and polynomial in the input size $|x|$, i.e., in time $\mathcal{O}(f(k) \cdot |x|^c)$ for a constant c and an arbitrary function f. In practice, this means that problem instances can be solved efficiently, even when the problem is NP-hard in general, if k is known to be small.

3 Most Inforbable Explanations

In the MAP problem, one seeks to find the joint value assignment to a set of variables that has maximum posterior probability. Here the candidate solutions consist of joint value assignments to exactly that set of variables, i.e., a conjunction of value assignments $\{(H_1 = h_1) \wedge \ldots \wedge (H_n = h_n)\}$ to the individual variables of the explanation set. This assumes that both the candidate hypotheses and the granularity of the explanation are set beforehand.

In real life, however, candidate hypotheses are formed and considered during the inference process, and the granularity of the explanation varies. Let us assume there is evidence that a patient suffers from a lung disease. On examination, when further evidence becomes available, the diagnosis may be refined to an obstructive lung disease, and later on, even further refined to the more specfic COPD and finally chronic bronchitis (Fig. 3). Preferably, we would like to find an explanation that has high probability and is specific, like $\{(\text{CB} = \text{TRUE}) \wedge (\text{EM} = \text{FALSE}) \wedge \ldots \wedge (\text{LP} = \text{FALSE})\}$, denoting that the patient has chronic bronchitis and no other lung disease is present. But what if there is not enough evidence to clearly distuinguish between chronic bronchitis and emphysema? Would it be wise to ignore the possibility of other lung diseases being present (maybe altering the advised medication) if the probability of their presense is maybe not convincing, but still non-neglectable?

Let us consider the three cases as presented in Table 1. In case a), the explanation is as specific as possible and has a high probability: the patient suffers from chronic bronchitis and no other lung disease is present. Case b) reflects that no clear distinction between chronic bronchitis and emphysema could be made. Note, however, that the probability of the three joint value assignments that correspond with $\{((\text{CB} = \text{TRUE}) \vee (\text{EM} = \text{TRUE})) \wedge \ldots \wedge (\text{LP} = \text{FALSE})\}$ is high. Here, it seems best to restrict the diagnosis to "COPD", rather than to refine it

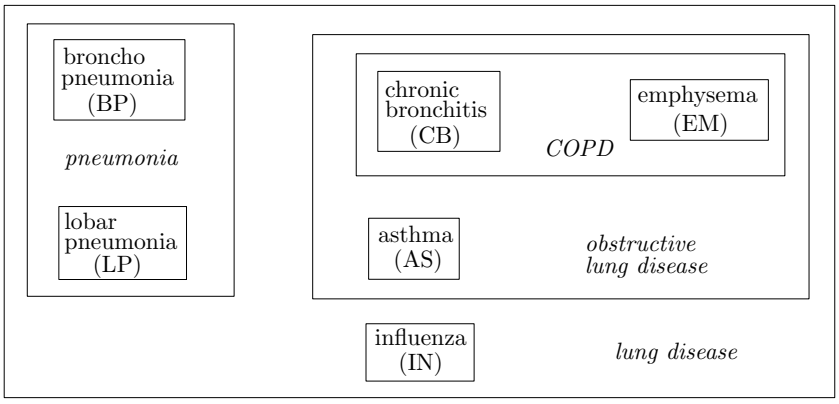

Fig. 3. Example of part of a classification of lung diseases

further. In case c) the patient definately suffers from chronic bronchitis, but in addition, some form of pneumonia may be present. Here, it would be wise (if no further evidence can be gathered) to settle for the diagnosis "chronic bronchitis, and maybe also pneumonia" and describe medication that covers both.

case	BP	LP	CB	EM	AS	IN	prob.
a	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	0.87
	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	0.04
	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	0.02
	other						0.07
b	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	0.48
	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	0.37
	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	0.10
	other						0.05
c	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	0.48
	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	0.21
	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	0.17
	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	0.08
	other						0.06

Table 1. Joint value assignments and their probabilities in the *lung disease* example

What we did in case a) corresponds to 'plain' MAP. In case b) and c), however, we choose as explanation a *set of joint value assignments* rather than a singleton joint value assignment, namely the set that corresponds to the (informal) diagnoses "COPD" (case b), respectively "chronic bronchitis, and maybe also pneumonia" (case c). Or to put it more formally, the sets of joint

value assignments that correspond to the sentences $\{((\text{CB} = \text{TRUE}) \vee (\text{EM} = \text{TRUE})) \wedge (\text{AS} = \text{FALSE}) \wedge (\text{IN} = \text{FALSE}) \wedge (\text{BP} = \text{FALSE}) \wedge (\text{LP} = \text{FALSE})\}$, respectively $\{(\text{CB} = \text{TRUE}) \wedge (\text{EM} = \text{FALSE}) \wedge (\text{AS} = \text{FALSE}) \wedge (\text{IN} = \text{FALSE})\}$. Thus, we extended MAP to deal with sets of joint value assignments, each consisting of a conjunction of value assignments to the variables in the explanation set[2].

We can also use a possible world semantics to describe these explanations. In case a) the explanation corresponds to the world where CB is TRUE and all other variables are FALSE. In case b) the explanation corresponds to the worlds where either CB or EM or both are set to TRUE, and all other variables are FALSE. In case c) the set of possible worlds are those where CB is TRUE, BP and LP are either TRUE or FALSE, and the other variables are FALSE. If we count these worlds in the three cases, we see that there is a single world in case a) with probability 0.87, there are three worlds in case b) whose probabilities add up to 0.95, and there are four worlds in case c) with total probability 0.94. Thus, in order to gain probabibility mass, in case b) and c) we needed to trade off informativeness, where we define explanation H to be more informative than H' if H corresponds to fewer possible worlds than H'.

3.1 Succinct encodings

We saw that the formal definition of "chronic bronchitis, and maybe also pneumonia", which corresponds to four possible worlds in the *lung disease* example, can be quite succinctly described as $\{(\text{CB} = \text{TRUE}) \wedge (\text{EM} = \text{FALSE}) \wedge (\text{AS} = \text{FALSE}) \wedge (\text{IN} = \text{FALSE})\}$ because the values of BP and LP are "don't cares". Surely, not every combination of four possible worlds can be described so easily, and we may need to resort to a full enumeration of four joint value assignments to describe that explanation.

That feels quite unnatural and unsatisfactory, in the sense that such an explanation (that consists of an arbitrarily complex sentence over the values of the variables) does not appear to be very informative at first sight. The sentence $(\text{AS} = \text{TRUE})$ corresponds to 32 possible worlds in which the patient has asthma (without committing to a particular value of the other variables), yet this is far more comprehensible and informative than a plain enumeration of, say, 11 possible worlds, so despite being "less informative" given the possible worlds semantics, we would like to enforce some reasonable encoding that makes the explanation easy to understand and to reason with. But there are also complexity-theoretic reason to constrain how the explanation should be encoded: if we allow the explanation to be encoded as an arbitrary set of w possible worlds, where w is given as a binary number, we may need an exponential (in w) number of bits to describe that explanation. Therefore, apart from high probability and a low number of possible worlds, we also require that the sentence describing these possible worlds is short, i.e., we also demand succinct encodings. To be

[2] Observe that we assume binary variables here for ease of exposition, but we might also include variables with a higher cardinality, like TEMP with values $\{\text{low}, \text{normal}, \text{high}\}$, stating, e.g., $\{((\text{TEMP} = \text{low}) \vee (\text{TEMP} = \text{normal})) \wedge \ldots\}$.

precise, we require that the explanation can be encoded by the addition of at most $\hat{w} = \mathcal{O}(\lceil \log_2(w+1) \rceil)$ *partial* joint value assignments to subsets of the explanation set.

We finish this section with an informal problem definition of Most Inforbable Explanation, combining these three requirements:

Most Inforbable Explanation (informal)
Instance: A Bayesian network, partitioned into evidence nodes, explanation nodes, and intermediate variables.
Output: An explanation that has high probability, corresponds to few possible worlds, and is succinctly encodable.

4 Computational Complexity

To investigate the computational complexity of Most Inforbable Explanation, we will formally define a decision variant of this problem as follows.

Most Inforbable Explanation
Instance: A Bayesian network $\mathcal{B} = (\mathbf{G}, \Pr)$, where $\mathbf{V}$ is partitioned into a set of evidence nodes $\mathbf{E}$ with a joint value assignment $\mathbf{e}$, an explanation set $\mathbf{H}$, and intermediate variables $\mathbf{I}$; a rational number $0 \leq q < 1$ and a natural number w.
Question: Is there a set $\{\mathbf{h_1}, \ldots, \mathbf{h_w}\}$ of w distinct joint value assignments $\mathbf{h_1}, \ldots, \mathbf{h_w}$ to $\mathbf{H}$, encodable by the addition of at most $\hat{w} = \mathcal{O}(\lceil \log_2(w+1) \rceil)$ joint value assignments $\mathbf{h'}$ to subsets of $\mathbf{H}$, such that $\sum_{i=1}^{w} \Pr(\mathbf{h_i}, \mathbf{e}) = \sum_{j=1}^{\hat{w}} \Pr(\mathbf{h'_j}, \mathbf{e}) > q$?

Theorem 1. Most Inforbable Explanation *is* $\mathsf{NP}^{\mathsf{PP}}$*-complete.*

Proof. We prove membership in $\mathsf{NP}^{\#\mathsf{P}}$, membership in $\mathsf{NP}^{\mathsf{PP}}$ follows as $\mathsf{P}^{\#\mathsf{P}} = \mathsf{P}^{\mathsf{PP}}$. Membership can be shown by non-deterministically guessing a certificate, consisting of a set of at most $\hat{w}$ joint value assignments $\mathbf{h'}$ to subsets of $\mathbf{H}$; checking that this certificate yields at most w distinct joint value assignments to $\mathbf{H}$; computing, using the $\#\mathsf{P}$ oracle, $\sum_{j=1}^{\hat{w}} \Pr(\mathbf{h'_j}, \mathbf{e})$ (note that $\#\mathsf{P}$ is closed under addition), and finally deciding whether $\sum_{j=1}^{\hat{w}} \Pr(\mathbf{h'_j}, \mathbf{e}) > q$. Note that the number of joint value assignments w may grow exponentially in the input size, as w is encoded in binary notation, but that all three steps of the verification algorithm can be done in polynomial time given the constraint that $\{\mathbf{h_1}, \ldots, \mathbf{h_w}\}$ must be succinctly (i.e., logarithmically in w) encodable. Note that $\mathsf{NP}^{\mathsf{PP}}$-hardness follows since Most Inforbable Explanation has MAP as a special case: take $w = 1$. □

If $w = 0$ then Most Inforbable Explanation degenerates to Inference. If $w = 1$ then Most Inforbable Explanation degenerates to MAP. Furthermore, Most Inforbable Explanation inherits the inapproximability results of MAP [22]. MAP is fixed parameter tractable (fp-tractable) for $\{c, 1-p, \mathrm{tw}\}$, i.e., MAP can be solved fast when the treewidth tw of the restricted junction

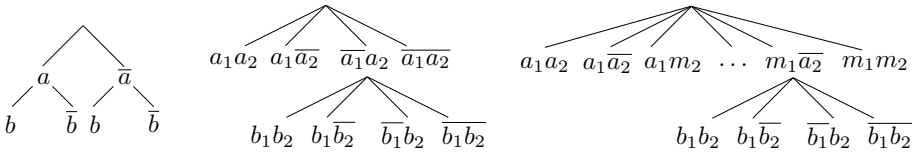

Fig. 4. The fp-tractable MAP algorithm branches on each value assignment to the variables in $\mathbf{H}$, computing marginal distributions over the variables not in $\mathbf{H}$; in the left subfigure the example $\mathbf{H} = \{A, B\}$, $c = 2$ is illustrated. The size of the branching tree is bounded by the probability of the most probable joint value assignment. Here, we extend this algorithm by branching over all possible value assignments in all possible worlds: part of the branching tree for $w = 2$ is drawn in the middle subtree. In the right subtree part of the branching tree for $\hat{w} = 2$ is drawn. Here, for each choice of $\hat{w}$, a variable can take any of its values, or it can take no value at all, denoted with m to illustrate that we marginalize over that variable rather than assign it a value

tree and cardinality c of the variables are small *and* the most probable explanation has a high probability[3] ($1 - p$ is low) [2, 14]. However, this may not hold for Most Inforbable Explanation since we need to choose w joint value assignments out of maximally $c^{|\mathbf{H}|}$ which by itself is a source of complexity. However, the $\{c, 1 - p, \mathrm{tw}\}$-fixed-parameter tractable algorithm[4] for MAP can be adjusted by branching on each of the (at most) c^w combinations of values for each variable, rather than on each of the (at most) c values (see Fig. 4). Therefore, Most Inforbable Explanation is fp-tractable for $\{c, 1-p, \mathrm{tw}, w\}$. Since Most Inforbable Explanation is a generalization of both MAP (for $w = 1$) and Inference (for $w = 0$) it follows that Most Inforbable Explanation remains intractable for the set of parameters $\{c, 1 - p, w\}$ and $\{c, \mathrm{tw}, w\}$ [16, 5].

It can be shown that Most Inforbable Explanation is also fp-tractable for $\{c, 1 - p, \mathrm{tw}, \hat{w}\}$, i.e., instead of bounding the number of possible worlds, we bound the size of the encoding. This can be done by further augmenting the above-mentioned algorithm, allowing it to branch on the (at most) $c^{\hat{w}} + c^{\hat{w}-1} + \ldots + 1$ combinations of values and non-assigned variables (that are marginalized over) — see again Fig. 4. Thus, from a computational point of view, Most Inforbable Explanation is not harder than 'plain' MAP, as both are $\mathsf{NP}^{\mathsf{PP}}$-complete. However, to render Most Inforbable Explanation fixed-parameter tractable, an additional constraint needs to be imposed on wither the number of possible worlds w or the number of (partial) joint value assignments $\hat{w}$ encoding these worlds.

[3] Technically speaking, $1-p$ is not a parameter as it is not a natural number; however, it can be mapped one-to-one to a suitable natural parameter [14].

[4] See [2] for the original algorithm for Most Probable Explanation, and [14] for the augmented algorithm for MAP.

5 Conclusion

In this paper, we introduced Most Inforbable Explanation as an extention to MAP, in order to combine both inference to the best explanation and (some aspects of) selecting candidate hypotheses and determining the granularity of the explanations. In human reasoning, the sets h_i are not likely to be arbitrarily chosen, but may correspond to common phrases as "either A or B, or both", "maybe A, but definitely not B", or "likely A, and possibly also B"; simple heuristics may exist that favor such phrases in practice and penalizing more complex structures, thus enforcing the formal logarithmic bound introduced in the formal definition and the fpt-result for $\hat{w}$. A succint encoding of "Asthma, but also at least one other disease" (spanning 31 possible worlds in the example) may be $\{(\text{AS} = \textsc{true})\} \setminus \{(\text{BP} = \textsc{false}) \wedge (\text{LP} = \textsc{false}) \wedge (\text{CB} = \textsc{false}) \wedge (\text{EM} = \textsc{false}) \wedge (\text{IN} = \textsc{false})\}$. We did not include such encodings (allowing for substraction, as well as addition, of partial joint value assignments) as it is not obvious that the above mentioned algorithm is fp-tractable in this case.

A particularly interesting aspect of informativeness of explanations lies in the often *contrasting* nature of explanations: often, we do not simply want to explain: *'Why this?'*, but *'Why this, rather than that?'* [18]. For example, to explain why Alice got tenure, referring to her quality teaching is unsufficient when Bob is an excellent teacher as well, but happend to be denied tenure: a better explanation would (also) refer to her many high-rated publications that Bob lacked. We leave a formal study of how such aspects may be implemented in a computational problem for future work.

Acknowledgments The author wishes to thank Iris van Rooij, Linda van der Gaag, and Pim Haselager for helpful discussions and literature suggestions.

References

1. J. Barwise. Information and impossibilities. *The Notre Dame Journal of Formal Logic*, 38(4):488–515, 1997.
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 *Proceedings of the 15th European Conference on Artificial Intelligence*, pages 675–679, 2002.
3. E. Charniak and S. E. Shimony. Cost-based abduction and MAP explanation. *Artificial Intelligence*, 66(2):345–374, 1994.
4. A. Darwiche. *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press, 2009.
5. C. P. De Campos. New complexity results for MAP in Bayesian networks. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, pages 2100–2106, 2011.
6. L. de Campos, J. Gamez, and S. Moral. Partial abductive inference in Bayesian belief networks using a genetic algorithm. *Pattern Recognition Letters*, 20(11-13):1211–1217, 1999.

7. L. de Campos, J. Gamez, and S. Moral. Partial abductive inference in Bayesian belief networks by simulated annealing. *International Journal of Approximate Reasoning*, 27(3):263–283, 2001.
8. L. de Campos, J. Gámez, and S. Moral. Simplifying explanations in Bayesian belief networks. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(4):461–489, 2001.
9. R. G. Downey and M. R. Fellows. *Parameterized complexity*. Springer Verlag, Berlin, 1999.
10. 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.
11. D. H. Glass. Inference to the best explanation: a comparison of approaches. In *Second Symposium on Computing and Philosophy*, 2009.
12. F. V. Jensen and T. D Nielsen. *Bayesian Networks and Decision Graphs*. Springer Verlag, New York, NY, second edition, 2007.
13. J. Kwisthout. *The Computational Complexity of Probabilistic Networks*. PhD thesis, Faculty of Science, Utrecht University, The Netherlands, 2009.
14. J. Kwisthout. Most probable explanations in Bayesian networks: Complexity and tractability. *International Journal of Approximate Reasoning*, 52(9):1452 – 1469, 2011.
15. J. Kwisthout. Structure approximation of most probable explanations in Bayesian networks. In *Proceedings of the 24th Benelux Conference on Artificial Intelligence (BNAIC'12)*, 2012.
16. Johan Kwisthout. The computational complexity of probabilistic inference. Technical Report ICIS–R11003, Radboud University Nijmegen, 2011.
17. C. Lacave and F. J. Díez. A review of explanation methods for Bayesian networks. *The Knowledge Engineering Review*, 17(2):107–127, 2002.
18. P. Lipton. *Inference to the best explanation*. Routledge, 2004.
19. D.A. Nardone. Collecting and analyzing data: Doing and thinking. In H.K. Walker, W.D. Hall, and J.W. Hurst, editors, *Clinical Methods: The History, Physical, and Laboratory Examinations*, chapter 2. Boston: Butterworths, 3rd edition, 1990.
20. J. D. Park and A. Darwiche. Approximating MAP using local search. In *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence*, pages 403–410. Morgan Kaufmann Publishers, San Francisco, California, 2001., 2001.
21. J. D. Park and A. Darwiche. Solving MAP exactly using systematic search. In *Proceedings of the 19th Annual Conference on Uncertainty in Artificial Intelligence (UAI-03)*, pages 459–46. Morgan Kaufmann, 2003.
22. J. D. Park and A. Darwiche. Complexity results and approximation settings for MAP explanations. *Journal of Artificial Intelligence Research*, 21:101–133, 2004.
23. J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, Palo Alto, CA, 1988.
24. N. Robertson and P.D. Seymour. Graph minors II: Algorithmic aspects of tree-width. *Journal of Algorithms*, 7:309–322, 1986.
25. C. Yuan, T. Lu, and M. J. Druzdzel. Annealed MAP. In *Proceedings of the Twentieth Conference in Uncertainty in Artificial Intelligence*, pages 628–635. AUA, 2004.