Statistics and Computing (1998) **8**, 159–173

An efficient algorithm for finding the M most probable configurations in probabilistic expert systems

D. NILSSON

Aalborg University, Department of Mathematics and Computer Science, Institute for Electronic Systems, Fredrik Bajers Vej 7 E, 9220 Aalborg Øst, Denmark
E-mail: nilsson@iesd.auc.dk

Received January 1996 and accepted November 1997

A probabilistic expert system provides a graphical representation of a joint probability distribution which enables local computations of probabilities. Dawid (1992) provided a 'flow-propagation' algorithm for finding the most probable configuration of the joint distribution in such a system. This paper analyses that algorithm in detail, and shows how it can be combined with a clever partitioning scheme to formulate an efficient method for finding the M most probable configurations. The algorithm is a divide and conquer technique, that iteratively identifies the M most probable configurations.

Keywords: Bayesian network, belief revision, charge, conditional independence, divide-and-conquer, evidence, flow, junction tree, marginalization, maximization, most probable explanation, potential function, propagation

1. Introduction

A probabilistic expert system (PES) functions by providing an efficient representation of the joint probability distribution of a finite collection of random variables (Lauritzen and Spiegelhalter, 1988). Such a system is typically specified as a Bayesian network (Pearl, 1986; Jensen, 1996), a graphical model consisting of nodes joined by directed edges. Nodes in the graph represent variables and missing links in the graph represent conditional independencies among the variables. The problem of finding the M most probable configurations in a PES is an optimization problem that also is called belief revision (Pearl, 1988), and 'The most probable explanations problem' (Sy, 1993). Several approaches have been proposed for solving this problem. Pearl (1988) described an algorithm for finding the most probable configuration in singly connected Bayesian networks. On the basis of his work, Sy (1993) developed a computational method for finding the M most probable configurations in such networks. However, singly connected networks are rather limited in their power to represent probabilistic models, and recently researchers have focused on solving the problem for an arbitrary Bayesian network. A process of compilation of the Bayesian network, involving various manipulations, leaves a higher level graphical structure, the junction tree (Jensen, 1988).

Dawid (1992) presented an elegant and efficient algorithm to calculate the most probable configuration in a junction tree. His algorithm can be regarded as the propagation of 'max-flows' through the junction tree, each such flow only involving two adjacent vertices and the associated edge.

Li and D'Ambrosio (1993) presented a method that applies to any multiply connected network. It is based on the combination of partially instantiated distributions according to an optimal factoring. However, even though the M most probable configurations can be obtained with a linear time cost in the number of nodes in the network when the optimal factoring is given, finding the optimal factoring has polynomial complexity.

Seroussi and Golmard (1994) described a method, named fast belief revision (FBR) for finding the M most probable configurations in junction trees. Similar to the

Dawid algorithm, FBR propagate max-flows through the junction tree. But while the max-flows in the Dawid algorithm are vectors, the max-flows in FBR are matrices, and it leaves FBR with considerable computational cost.

Nilsson (1994) analysed the Dawid algorithm in detail and proved that the three most probable configurations can be identified directly from the propagation of max-flows, but in general the fourth cannot be found directly. The intuition behind this result is that the algorithm of Dawid provides the most probable configuration in various subsets of the configuration space. These 'max-configurations' can then be identified. The three most probable configurations belong to this set of max-configurations; however, in general the fourth most probable does not belong to the set. On the basis of this work, Nilsson combined the algorithm of Dawid with a partitioning scheme to formulate a method for finding the M most probable configurations.

In this paper we present a cleverer partitioning scheme and obtain an efficient algorithm for finding the M most probable configurations. The algorithm is a divide and conquer technique, that iteratively identifies the M most probable configurations. The algorithm is named max-flow propagation, since each of the M most probable configurations is identified by the propagation of 'simple' max-flows through the junction tree: the most probable configuration is identified by the propagation of vectors through the network (as in the Dawid algorithm); subsequently the algorithm uses the result of this initial 'vector-propagation', and identifies the next most probable configurations by the propagation of values (not vectors) through the network. This leaves the algorithm less complex than any other method known to the author. It has been implemented into the experimental shell XBAIES, which is an extension of BAIES (Cowell, 1992), and it is in the process of being implemented in the probabilistic expert system shell HUGIN (Andersen *et al.*, 1989).

The outline of the paper is as follows. Section 2 describes the basic structure and provides the notation that will be used throughout. Section 3 shows how to find the most probable configuration by the Dawid algorithm. Section 4 describes max-flow propagation and Section 5 discusses the computational complexity of the algorithm. Finally, Section 6 deals with related problems to that of finding the M most probable configurations. In particular we show how to find the M most probable configurations of any subset of the variables. Throughout we use the artificial ASIA example, reproduced from Lauritzen and Spiegelhalter (1988), to illustrate our points. We will borrow notations, definitions and results from the paper by Dawid (1992).

2. Basic structure

We are given a set of discrete variables X_u, $u \in U$, where U is a finite index-set. For any random variable X_u we use the symbol $\mathscr{H}_u$ to denote its possible values. Given a nonempty subset A of U, let $\mathscr{H}_A$ denote the Cartesian product of $\mathscr{H}_u$ for $u \in A$, that is $\mathscr{H}_A = \times_{u \in A} \mathscr{H}_u$, abbreviating $\mathscr{H}_U$ to $\mathscr{H}$. We refer to elements of $\mathscr{H}_A$ as *configurations* of $\mathscr{H}_A$. If the configuration $x = (x_u : u \in U) \in \mathscr{H}$ then x_A will denote $(x_u : u \in A)$.

Also for any $A, B \subseteq U$, we interpret the product (with similar conventions for ratios and sums) gh of two real functions g on $\mathscr{H}_A$ and h on $\mathscr{H}_B$ as the function on $\mathscr{H}_{A \cup B}$, where the value at $x_{A \cup B} \in \mathscr{H}_{A \cup B}$ is $g(x_A)h(x_B)$. If $A \subseteq B$ then $\max_{B \setminus A} h$ will denote the function on $\mathscr{H}_A$, where the value at $x_A \in \mathscr{H}_A$ is $\max_{z \in \mathscr{H}_B} \{h(z) : z_A = x_A\}$. Finally, we abbreviate $\max_{U \setminus A} h$ to $\hat{h}_A$.

2.1. *Junction tree representation*

We are given a tree $\mathscr{T}$ with vertex-set $\mathscr{C}$ and edge-set $\mathscr{S}$. Associated with any vertex C in $\mathscr{T}$ is a subset of U, which we likewise denote by C. Such subsets will be termed *cliques*. This is because the subsets $C \in \mathscr{C}$ can be identified with the cliques of an appropriate undirected graph constructed during compilation from the original Bayesian network (Lauritzen *et al.*, 1984). Any $u \in U$ belongs to at least one clique. For computational reasons we require that the tree $\mathscr{T}$ is a junction tree, with the property that, if $u \in U$ is contained in any two cliques C_i and C_j, then it is contained in all the cliques in the unique path from C_i to C_j. This property is termed the *junction tree property*, and is guaranteed from the compilation process.

Let C_1 and C_2 be two adjacent vertices of $\mathscr{T}$. We associate with the edge S joining them, the set $S = C_1 \cap C_2 \subset U$. This will be called a *separator*.

We suppose that $\mathscr{T}$ is connected, since the disconnected case can be handled by considering each component separately.

Finally we are given a probability function f of the discrete variables $X_u, u \in U$, defined as

$$f(x) = \frac{\prod_{C \in \mathscr{C}} a_C(x_C)}{\prod_{S \in \mathscr{S}} b_S(x_S)}, \tag{1}$$

where each a_C is a known non-negative real function on $\mathscr{H}_C$, and each b_S is a known non-negative real function on $\mathscr{H}_S$. Any non-negative function, which can be represented in the form of Equation 1 will be said to *factorize* on $\mathscr{T}$. The right-hand side of Equation 1 is interpreted as 0 if the denominator is 0. Throughout we define $0/0$ to be zero.

We shall term any collection of non-negative real functions $K = (\{a_C : C \in \mathscr{C}\}, \{b_S : S \in \mathscr{S}\})$ a *charge* on $\mathscr{T}$. The functions will be termed *potentials* (on the cliques and separators respectively).

Any charge that fulfils Equation 1 is termed a *representation* for f on $\mathscr{T}$. Note that f factorizes on $\mathscr{T}$ if and only if there exists a representation for f on $\mathscr{T}$.

The idea behind the building of a junction tree is that while the overall problem, i.e. the specification and calculation of the probabilities of f, is too large to handle in a joint table, the individual cliques are of manageable size.

2.2. *Passing flows in* $\mathscr{T}$

Let C_1 and C_2 be adjacent cliques in $\mathscr{T}$ and let S be the edge joining them. We will describe the effect of passing a *max-flow* from C_1 to C_2. Suppose that prior to this max-flow we have a charge $K = (\{a_C : C \in \mathscr{C}\}, \{b_S : S \in \mathscr{S}\})$. Then the effect of the max-flow is a new charge similar to K, except that it has a new potential b_S^* on S and a new potential $a_{C_2}^*$ on C_2 given by

$$b_S^* = \max_{C_1 \setminus S} a_{C_1} \tag{2}$$

and

$$a_{C_2}^* = a_{C_2} \cdot \lambda_S, \tag{3}$$

where the *update-factor* λ_S is given by $\lambda_S = b_S^*/b_S$. All other potentials are unaltered.

We can also pass a *sum-flow* from C_1 to C_2. The effect of the passage of a sum-flow is given by replacing 'max' with '$\sum$' in Equation 2. Note that the fraction in Equation 1 is invariant under max-flow passing as well as sum-flow passing.

Example 1. The ASIA-example is reproduced with minor modifications from Lauritzen and Spiegelhalter (1988).

There are eight binary variables α, δ, τ, ϵ, λ, ξ, β and σ, reflecting various attributes of a patient:

α	Recent visit to Asia?	σ	Smoker?
τ	Tuberculosis?	λ	Lung cancer?
ϵ	Either tuberculosis or lung cancer?	β	Bronchitis?
δ	Dyspnoea?	ξ	Positive X-ray?

The states of α are denoted by a for 'yes', $\bar{a}$ for 'no' with similar conventions for the other variables.

The initial qualitative structure is given in a Bayesian network not reproduced here. After compilation, we obtain the junction tree given in Fig. 1.

The charge on $\mathscr{T}$, derived from the original conditional probability specifications in the Bayesian network, is given in Table 1.

Suppose we now pass a max-flow from clique C_1 through separator S_1 to clique C_2. The new potential on S_1 becomes

b_{S_1}	t	$\bar{t}$
	0.0099	0.9801

and the update-factor is equal to the new potential on S_1. So, the new potential on C_2 will be

a_{C_2}			t	$\bar{t}$
	e	l	0.0099	0.9801
		$\bar{l}$	0.0099	0
	$\bar{e}$	l	0	0
		$\bar{l}$	0	0.9801

□

3. The most probable configuration

In this section we describe the algorithm of Dawid (1992) for finding the maximum of a function f, that factorizes on $\mathscr{T}$. Suppose that f is specified by the charge $(\{a_C : C \in \mathscr{C}\}, \{b_S : S \in \mathscr{S}\})$. We need not restrict f to be a probability function.

The algorithm proceeds by the propagation of max-flows through the junction tree. To achieve our purpose we need to order the flows appropriately. Dawid calls an ordered list of directed edges, a *schedule*. A schedule specifies which flows are to be passed, and in what order. Relative to a schedule, a flow is called *active* if, before it is passed, the sender has already received active flows from all its neighbouring cliques with the possible exception of the receiver. Note that the first active flow must originate from a leaf in $\mathscr{T}$. It is easy to show that for any tree $\mathscr{T}$, we can construct a schedule of flows so that eventually an active flow has been passed in both directions between each pair

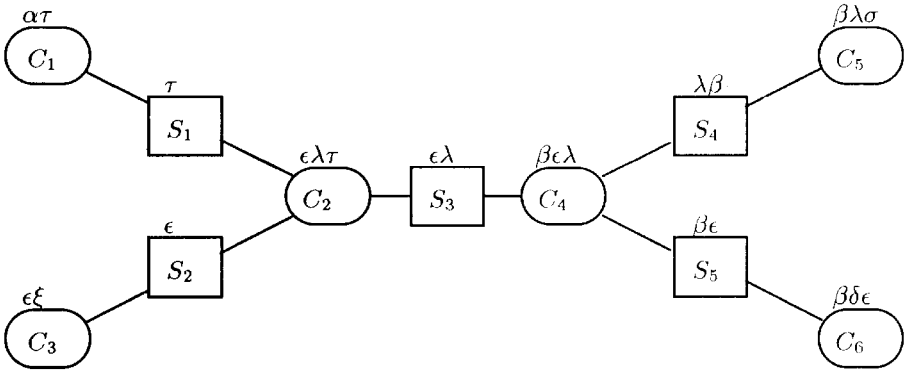

Fig. 1. *Junction tree for ASIA*

Table 1. *Initial charge for ASIA*

Clique-potentials:

a_{C_1}

	t	$\bar{t}$
a	.0005	.0095
$\bar{a}$	.0099	.9801

a_{C_2}

		t	$\bar{t}$
e	l	1	1
	$\bar{l}$	1	0
$\bar{e}$	l	0	0
	$\bar{l}$	0	1

a_{C_3}

	x	$\bar{x}$
e	.98	.02
$\bar{e}$	.05	.95

a_{C_4}

		t	$\bar{t}$
b	e	1	1
	$\bar{e}$	1	1
$\bar{b}$	e	1	1
	$\bar{e}$	1	1

a_{C_5}

		s	$\bar{s}$
b	l	.03	.0015
	$\bar{l}$	.27	.1485
$\bar{b}$	l	.02	.0035
	$\bar{l}$	.18	.3465

a_{C_6}

		s	$\bar{s}$
b	d	.7	.6
	$\bar{d}$	.3	.4
$\bar{b}$	d	.5	.1
	$\bar{d}$	.5	.9

Separator-potentials $b_{S_1} = b_{S_2} = b_{S_3} = b_{S_4} = b_{S_5} = 1$.

of neighbouring cliques. In this state, any other flow will leave the potentials unaltered, and we say that the system *has reached equilibrium*.

Jensen *et al.* (1990) propose a schedule consisting of two phases. They choose an arbitrary clique C_1 to be the 'root-clique'. An initial *collection* phase involving the passage of active flows only along edges towards C_1 is followed by a *distribution* phase, in which, starting from C_1, active flows are passed back towards the periphery.

3.1. *The max-marginal charge*

The following theorem is due to Dawid (1992). It states that when equilibrium is reached after the passage of max-flows in $\mathscr{T}$, the new charge on $\mathscr{T}$ is given by

$$(\{\hat{f}_C : C \in \mathscr{C}\}, \{\hat{f}_S : S \in \mathscr{S}\}).$$

Recall that for any $A \subseteq U$, $\hat{f}_A$ is a function on $\mathscr{H}_A$, where the value at $x_A \in \mathscr{H}_A$ is the maximum of f over the subset $\{z \in \mathscr{H} : z_A = x_A\}$.

Theorem 1. Suppose we start with a representation $K^0 = (\{a_C : C \in \mathscr{C}\}, \{b_S : S \in \mathscr{S}\})$ for a function f, which factorizes on $\mathscr{T}$; and progressively modify it by passing a sequence of active max-flows. Then when equilibrium has reached, the new charge on $\mathscr{T}$ is

$$(\{\hat{f}_C : C \in \mathscr{C}\}, \{\hat{f}_S : S \in \mathscr{S}\}).$$

Proof. See Dawid (1992) or Jensen (1996) □

The collection $(\{\hat{f}_C : C \in \mathscr{C}\}, \{\hat{f}_S : S \in \mathscr{S}\})$ is termed the *max-marginal charge* of f on $\mathscr{T}$.

If f is a probability function for $(X_u : u \in U)$, and $A \subseteq U$ then $\hat{f}_A(x_A)$ for $x_A \in \mathscr{H}_A$, can be interpreted as the probability of the most probable configuration in the subset $\{z \in \mathscr{H} : z_A = x_A\}$. Nilsson (1994) showed that the probability of each of the three most probable configurations is contained in the set $\{F_C(x_C) : C \in \mathscr{C},\ x_C \in \mathscr{H}_C\}$; however in general the probability of the fourth most probable configuration does not belong to the set.

We shall use that Equation 1 is invariant under max-flow passing. Therefore the max-marginal charge is a representation for f on $\mathscr{T}$, that is

$$f = \frac{\prod_{C \in \mathscr{C}} \hat{f}_C}{\prod_{S \in \mathscr{S}} \hat{f}_S}. \tag{4}$$

Note that for the max-marginal charge we have max-consistency, that is

$$\max_{C_1 \setminus S} \hat{f}_{C_1} = \hat{f}_S = \max_{C_2 \setminus S} \hat{f}_{C_2}$$

for any two cliques C_1 and C_2 with intersection S.

The following theorem shows how to find the maximum of f over $\mathscr{H}$ from the max-marginal charge.

Theorem 2. For any $C \in \mathscr{C}$ it holds that

$$\max_U f = \max_C \hat{f}_C.$$

Proof. The result follows immediately from the definition of $\hat{f}_C$ for $C \in \mathscr{C}$. □

Theorem 2 states that we can find the maximum of f from the max-marginal charge. To find where the maximum is achieved, say x^1, when this is uniquely determined, we note that $x^1_A = \arg\max \hat{f}_A(x_A)$ for any $A \subseteq U$. Thus x^1_A may be calculated directly for each $A \in \mathscr{C} \cup \mathscr{S}$, from where x^1 is determined.

To allow for the possibility that the maximizing configuration is not unique, we shall use the following routine, which we name *max-distribution in* $\mathscr{T}$.

The routine starts in any root-clique, say C_1, and chooses a configuration ξ_{C_1} that maximizes $\hat{f}_{C_1}$ over $\mathscr{H}_{C_1}$. Note that $\hat{f}_{C_1}(\xi_{C_1}) = \max f$, which follows from Theorem 2. Now proceed towards a neighbouring clique C_2, say, and let S_1 be the separator between C_1 and C_2. Define ξ_{S_1} by extracting the relevant coordinates from ξ_{C_1}, and note, that $\hat{f}_{S_1}(\xi_{S_1}) = \max f$ because of max-consistency. Now let ξ_{C_2} be a configuration that maximizes $\hat{f}_{C_2}$ over $\mathscr{H}_{C_2}$, and such that its coordinates for variables in S_1 agree with ξ_{S_1}. Note that $\hat{f}_{C_2}(\xi_{C_2}) = \hat{f}_{S_1}(\xi_{S_1}) = \max f$ because of max- consistency. Max-distribution in $\mathscr{T}$ proceeds in this manner with the propagation of max-flows through the network, and obtain a collection $(\{\xi_C : C \in \mathscr{C}\}, \{\xi_S : S \in \mathscr{S}\})$, where $\hat{f}_C(\xi_C) = \hat{f}_S(\xi_S) = \max f$ for all $C \in \mathscr{C}$ and all $S \in \mathscr{S}$. The junction tree guarantees that there exists a configuration v, say, with $v_C = \xi_C$ for $C \in \mathscr{C}$. Furthermore, v maximizes f, since the max-marginal charge is a representation for f, such that

$$f(v) = \frac{\prod_{C\in\mathscr{C}} \hat{f}_C(v_C)}{\prod_{S\in\mathscr{S}} \hat{f}_S(v_S)} = \frac{|\mathscr{C}| \max f}{|\mathscr{S}| \max f} = \max f.$$

3.2. *The sum-marginal charge*

Recall that the effect of passing a sum-flow is given by replacing 'max' with '$\sum$' in Equation 1. We have a similar result to that of Theorem 1 when sum-flows are passed rather than max-flows: after the passage of a sequence of sum-flows and equilibrium has reached, the new charge on $\mathscr{T}$ is given by

$$(\{f_C : C \in \mathscr{C}\}, \{f_S : S \in \mathscr{S}\}),$$

where functions of the form $f_A (A \subset U)$ is defined as

$$f_A = \sum_{U\setminus A} f.$$

When f is a probability function for $(X_u : u \in U)$, f_A is the implied marginal probability function for $(X_u : u \in A)$. The collection $(\{f_C : C \in \mathscr{C}\}, \{f_S : S \in \mathscr{S}\})$ is termed the *sum-marginal charge*.

3.3. *Incorporating of evidence*

Suppose that we observe 'evidence' $\mathscr{E}$ of the form

$$X_u \in \bar{\mathscr{H}}_u \subset \mathscr{H}_u \quad \text{for} \quad u \in V \subseteq U.$$

Define a new function $\bar{f}$ by

$$\bar{f}(x) = \begin{cases} f(x) & \text{if } x_u \in \bar{\mathscr{H}}_u \text{ for } u \in V \\ 0 & \text{otherwise.} \end{cases}$$

We will now find a representation for $\bar{f}$. The fact that $\bar{f}$ is proportional to, rather than equal to, the conditional probability function of $(X_u : u \in U)$ given the evidence $\mathscr{E}$ is of no consequence, and can be turned into an advantage.

Define functions l_u for $u \in V$ as

$$l_u(x_u) = \begin{cases} 1 & \text{if } x_u \in \bar{\mathscr{H}}_u \\ 0 & \text{otherwise.} \end{cases}$$

Thus l_u is the likelihood function based on the partial evidence $X_u \in \bar{\mathscr{H}}_u$. From the representation of f we obtain such a representation for $\bar{f}$ by associating each $u \in V$ with any one clique containing u, and replacing each a_C by

$$a_C(x_C) = a_C(x_C) \prod_{u\in V} \{l_u(x_u) : u \text{ is assigned to } C\},$$

where we define an empty product as 1.

We may be interested in finding the sum

$$\sum_{x\in\mathscr{H}} \{f(x) : x_u \in \bar{\mathscr{H}}_u \text{ for } u \in V\}. \tag{5}$$

Note that when f is a probability function, (5) is the probability of the evidence $\mathscr{E}$. In any case, (5) can be found as follows: starting from the representation of $\bar{f}$ we propagate sum-flows until the system has reached equilibrium. In this state, the potential for each $A \in \mathscr{C} \cup \mathscr{S}$ will thus be $\bar{f}_A = \sum_{U\setminus A} \bar{f}$, and (5) can now be found from

$$\sum_{x\in\mathscr{H}} \{f(x) : x_u \in \bar{\mathscr{H}}_u \text{ for } u \in V\} = \sum_{x\in\mathscr{H}} \bar{f}(x) = \sum_{x_A\in\mathscr{H}_A} \bar{f}_A(x_A).$$

Example 2. Suppose we observe that a patient has recently visited Asia ($\alpha = a$), he has positive X-ray ($\xi = x$) and he smokes ($\sigma = s$) and wish to find the most probable configuration consistent with this evidence $\mathscr{E}$. After incorporating the evidence and propagating active max-flows until equilibrium has reached, we obtain the max-marginal charge given in Table 2.

The order in which the active flows can be passed is not unique. One such ordering is

$$C_6 \to C_4, C_5 \to C_4, C_4 \to C_2, C_3 \to C_2, C_2 \to C_1$$
$$C_1 \to C_2, C_2 \to C_3, C_2 \to C_4, C_4 \to C_5, C_4 \to C_6.$$

From the max-marginal charge we see that the probability of the most probable configuration, x^1 say, consistent with the evidence $\mathscr{E}$ is $\max_{C_i} \hat{f}_{C_i} = 0.1955 \cdot 10^{-3}$.

The probability of the evidence $\mathscr{E}$ is $0.9242 \cdot 10^{-3}$, being obtained by prior or subsequent passage of sum-flows until equilibrium has reached.

Now we identify x^1 by max-distribution in $\mathscr{T}$ (see Fig. 2): we proceed from any root-clique, C_1, say, and choose a suitable configuration ξ_{C_1} which maximizes $\hat{f}_{C_1}$ over $\mathscr{H}_{C_1}$. So $\xi_{C_1} = (a, \bar{t})$ (see Table 2). Now we proceed outwards from C_1, along edge S_1 towards clique C_2: define $\xi_{S_1} = \bar{t}$ by extracting the relevant coordinates from ξ_{C_1}, and ξ_{C_2} is an arbitrary configuration that maximizes $\hat{f}_{C_2}$ over $\mathscr{H}_{C_2}$ and such that its variables in S_1 is ξ_{S_1}. The configuration $\xi_{C_2} = (e, l, \bar{t})$ fulfils this (see Table 2). Proceeding outwards in this way, starting from the root-clique C_1, we obtain a

Table 2. *The max-marginal charge (multiplied by 10^3) in ASIA, after incorporating the evidence $\alpha = a$, $\xi = x$, and $\sigma = s$*

Clique-potentials:

$\hat{f}_{C_1}$

	t	$\bar{t}$
a	.0926	0.1955
$\bar{a}$	.0	0

$\hat{f}_{C_2}$

		t	$\bar{t}$
e	l	.0103	0.1955
	$\bar{l}$	0.0926	0
$\bar{e}$	l	0	0
	$\bar{l}$	0	.0770

$\hat{f}_{C_3}$

	x	$\bar{x}$
e	.1955	0
$\bar{e}$	.0770	0

$\hat{f}_{C_4}$

		l	$\bar{l}$
b	e	.1955	.09261
	$\bar{e}$	0	.0770
$\bar{b}$	e	.0931	.0441
	$\bar{e}$	0	.0770

$\hat{f}_{C_5}$

		s	$\bar{s}$
b	l	.1955	0
	$\bar{l}$	.0926	0
$\bar{b}$	l	.0931	0
	$\bar{l}$	.0770	0

$\hat{f}_{C_6}$

		e	$\bar{e}$
b	d	.1955	.0770
	$\bar{d}$	.0838	.0513
$\bar{b}$	d	.0931	.0090
	$\bar{d}$	.0931	.0770

Separator-potentials:

$\hat{f}_{S_1}$

t	$\bar{t}$
.0926	.1955

$\hat{f}_{S_2}$

e	$\bar{e}$
.1955	.0770

$\hat{f}_{S_3}$

	l	$\bar{l}$
b	.1955	.0926
$\bar{b}$	.0931	.0770

$\hat{f}_{S_4}$

	l	$\bar{l}$
b	.1955	.0926
$\bar{b}$	.0931	.0770

$\hat{f}_{S_5}$

	e	$\bar{e}$
b	.1955	.0770
$\bar{b}$	.0931	.0770

collection $\{\xi_{C_1}, \ldots, \xi_{C_6}\}$, from which the maximizing configuration, x^1 is given. Figure 2 shows that $x^1 = (a, \bar{t}, e, l, x, b, s, d)$. □

4. The M most probable configurations

Suppose that a function f, factorizing on $\mathscr{T}$, is specified by the charge $(\{a_C : C \in \mathscr{C}\}, \{b_S : S \in \mathscr{S}\})$. In this section we describe how max-flow propagation (MFP) finds the M highest values of f, and identifies the maximizing values.

Let x^L, for any positive integer L, denote the configuration, where the Lth highest value of f is achieved. Note that when f is a joint probability function of the variables $(X_u : u \in U)$, the configurations $x^1, \ldots, x^M$ are the M most probable configurations of the joint distribution. x^L may not be unique, in which case we arbitrarily associate x^L with one of the configurations that maximizes f over the subset

$$\mathscr{H} \backslash \{x^1, x^2, \ldots, x^{L-1}\}.$$

Accordingly $f(x^1) \geq f(x^2)$, $f(x^2) \geq f(x^3)$ and so forth.

MFP uses Theorem 2 to find the maximum of f. Subsequently x^L for $L > 1$ is found by going through the following three phases (by induction, we may suppose that the configurations $x^1, \ldots, x^{L-1}$ have been identified).

Partition-phase. Here we partition $\mathscr{H} \backslash \{x^1, x^2, \ldots, x^{L-1}\}$ in subsets. Many possible partitionings exists; however a clever partitioning has two (conflicting) properties. Firstly the partitioning should consist of subsets that we can 'conquer'; that is the maximum of f over each subset in the partitioning should be easily obtained. Secondly the number of subsets in the partitioning should be as low as possible, since eventually we have to compare the maximum of f over each subset in the partitioning (and for efficiency reasons we want to compare as few numbers as possible).

Candidate-phase. This phase finds the maximum of f over the subsets generated in the partition-phase. The maximum of these maxima is the Lth highest value of f.

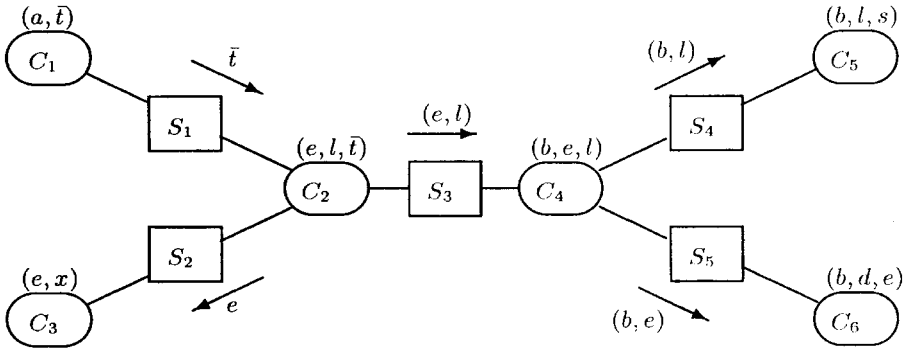

Fig. 2. *Max-distribution in $\mathscr{T}$ for ASIA gives $(a, \bar{t}, e, l, x, b, s, d)$*

Identification-phase. Here we identify the configuration where the Lth highest value is achieved.

4.1. *Introduction to the algorithm*

First we discuss a simplified version of the algorithm. Order the elements in the index-set U in some way, say $1, \ldots, n$. With this ordering we can write f as

$$f(x_1, \ldots, x_n) = \frac{\prod_{C \in \mathscr{C}} a_C(x_C)}{\prod_{S \in \mathscr{S}} b_S(x_S)}.$$

The maximizing value of f is denoted $x^1 = (x_1^1, \ldots, x_n^1)$, and found after the propagation of max-flows in $\mathscr{T}$. Now the second highest value of f must differ from x^1 in the state of at least one variable. Therefore, the n subsets

$$\begin{aligned}
&\{x \in \mathscr{H} : x_1 \neq x_1^1\} \\
&\ldots \\
&\{x \in \mathscr{H} : x_1 = x_1^1, \ldots, x_{i-1} = x_{i-1}^1, x_i \neq x_i^1\} \qquad (6) \\
&\ldots \\
&\{x \in \mathscr{H} : x_1 = x_1^1, \ldots, x_{n-1} = x_{n-1}^1, x_n \neq x_n^1\}
\end{aligned}$$

is a partitioning of $\mathscr{H} \backslash \{x^1\}$. Associated with the subsets in (6) we define functions

$$h^1(x) = \begin{cases} f(x) & \text{if } x_1 \neq x_1^1 \\ 0 & \text{otherwise} \end{cases}$$

$$\ldots$$

$$h^i(x) = \begin{cases} f(x) & \text{if } x_1 = x_1^1, \ldots, x_{i-1} = x_{i-1}^1, \\ & \quad \text{and } x_i \neq x_i^1 \\ 0 & \text{otherwise} \end{cases}$$

$$\ldots$$

$$h^n(x) = \begin{cases} f(x) & \text{if } x_1 = x_1^1, \ldots, x_{n-1} = x_{n-1}^1, \\ & \quad \text{and } x_n \neq x_n^1 \\ 0 & \text{otherwise.} \end{cases}$$

It is clear that the maximum of h^i over $\mathscr{H}$ is equal to the maximum of f over the ith subset in (6). Furthermore, since the subsets in (6) partition $\mathscr{H} \backslash \{x^1\}$ we have that

$$\max_{\mathscr{H} \backslash \{x^1\}} f(x) = \max_i \max_{\mathscr{H}} h^i(x).$$

So, if we can find the maximum of each h^i, then the second highest value of f is given by comparing these maxima. The maximum of h^i can be found as follows: after incorporating the (hypothetical) evidence

$$X_1 = x_1^1, \ldots, X_{i-1} = x_{i-1}^1 \quad \text{and} \quad X_i \neq x_i^1,$$

we propagate max-flows through $\mathscr{T}$ until equilibrium has reached, and obtain the max-marginal charge of h^i. The maximum of h^i can now be found through Theorem 2, and the maximizing configuration of h^i is found by max-distribution in $\mathscr{T}$. Accordingly, we can determine the second highest value of f and the associated configuration $x^2 = (x_1^2, \ldots, x_n^2)$. In order to identify the third highest value of f, we apply a similar method to the one above. First we partition $\mathscr{H} \backslash \{x^1, x^2\}$ in subsets by refining the partitioning in (6): if x^2 belongs to the ith subset in (6) then we partition the ith subset, excluding x^2, in the following $(n + 1 - i)$ subsets

$$\begin{aligned}
&\{x \in \mathscr{H} : x_1 = x_1^2, \ldots, x_{i-1} = x_{i-1}^2, x_i \notin \{x_i^1, x_i^2\}\} \\
&\{x \in \mathscr{H} : x_1 = x_1^2, \ldots, x_i = x_i^2, x_{i+1} \neq x_{i+1}^2\} \qquad (7) \\
&\ldots \\
&\{x \in \mathscr{H} : x_1 = x_1^2, \ldots, x_{n-1} = x_{n-1}^2, x_n \neq x_n^2\}.
\end{aligned}$$

The subsets in (7) and the subsets in (6), excluding the ith subset, is a partitioning of $\mathscr{H} \backslash \{x^1, x^2\}$. It only remains to find the maximum of f over each subset in (7). This is done by inserting the (hypothetical) evidence associated with each subset and propagate max-flows through $\mathscr{T}$ until equilibrium has reached. From the resulting charge, we can find the maximum of f and the maximizing configuration.

Using this idea, we can find the M highest values of f. However the method does require a large number of max-flows to be passed in the junction tree, and we now show how this can be avoided by using another partitioning scheme than the one above.

4.2. *The ordering of the cliques*

In the remaining part of the paper, we assume the junction tree $\mathscr{T}$ consists of p cliques, say $C_1, \ldots, C_p$ and $p-1$ separators, say $S_1, \ldots, S_{p-1}$. Furthermore, the ordering of the cliques obeys the following *running intersection property*: let S_{j-1} be the separator between the clique C_j and the lower numbered cliques $C_1, \ldots, C_{j-1}$. Then there exists (at least) one clique in $C_1, \ldots, C_{j-1}$ that contains S_{j-1}. The junction tree property ensures that such an ordering exists. For instance, the ordering of the cliques in the junction tree for ASIA have the running intersection property.

We let $\mathscr{T}_j$ denote the subtree induced by the cliques $C_1, \ldots, C_j$. The junction tree property guarantees that the subtree $\mathscr{T}_j$ is connected. The subset $(C_1 \cup \cdots \cup C_j) \subseteq U$ is likewise denoted by $\mathscr{T}_j$. We define S_0 and $\mathscr{T}_0$ to be empty sets. It is convenient to introduce the sets $R_j = C_j \backslash S_{j-1}$ for $j = 1, \ldots, p$ (note that $R_1 = C_1$). The sets S_{j-1} separate *the residual* R_j from $\mathscr{T}_{j-1} \backslash S_{j-1}$.

The ordering of the cliques (and separators) with the running intersection property provides a way to collect active flows towards C_1:

$$C_p \to \mathscr{T}_{p-1}, C_{p-1} \to \mathscr{T}_{p-2}, \ldots, C_3 \to \mathscr{T}_2, C_2 \to C_1,$$

that is, first we pass a flow from clique C_p towards its neighbouring clique in $\mathscr{T}_{p-1}$ (note that C_p is a leaf of $\mathscr{T}$), then we pass a flow from C_{p-1} towards its neighbouring clique in $\mathscr{T}_{p-2}$, and so forth.

We shall use that the same ordering applies, when we wish to collect active max-flows towards clique C_j:

$$\begin{aligned} &C_p \to \mathscr{T}_{p-1}, C_{p-1} \to \mathscr{T}_{p-2}, \ldots, C_{j+1} \to \mathscr{T}_j \\ &C_1 \to C_2, \mathscr{T}_2 \to C_3, \ldots, \mathscr{T}_{j-1} \to C_j. \end{aligned} \tag{8}$$

For instance, $\mathscr{T} \to C_3$ means that a flow is passed from C_3's neighbour in $\mathscr{T}_2$ towards C_3.

Similarly we can apply the ordering, if we wish to distribute active flows, starting from clique C_j:

$$\begin{aligned} &C_j \to \mathscr{T}_{j-1}, C_{j-1} \to \mathscr{T}_{j-2}, \ldots, C_2 \to C_1 \\ &\mathscr{T}_j \to C_{j+1}, \mathscr{T}_{j+1} \to C_{j+2}, \ldots, \mathscr{T}_{p-1} \to C_p. \end{aligned} \tag{9}$$

4.3. *The phases of the algorithm*

Recall that in Section 4.1 we partitioned $\mathscr{H} \backslash \{x^1\}$ in certain subsets. By inserting the (hypothetical) evidence associated with each subset and propagating max-flows until equilibrium has reached, we are able to identify the maximum of f over each subset. Now we provide another partitioning scheme, and show that the maximum of f over each subset we generate in this partitioning scheme, can be found directly from the max-marginal charge of f.

The algorithm uses the fact that we do not need the max-marginal for a function f, that factorizes on $\mathscr{T}$, in order to find the maximum of f. When we start from any representation of f, and collect active max-flows towards an arbitrary root-clique C^*, say, we obtain the potential $\hat{f}_{C^*}$ on C^*. From this potential, the maximum of f can be determined (Theorem 2). To identify the maximizing configuration, we start from the new charge on $\mathscr{T}$, obtained after the collection of max-flows towards C^*, and distribute 'simple' max-flows through $\mathscr{T}$: starting from clique C^*, we identify a configuration ξ_{C^*}, say, which maximizes $\hat{f}_{C^*}$ (the potential on C^*) over $\mathscr{H}_{C^*}$. Now proceed towards a neighbouring clique C', say. Let S^* be the separator between the two cliques, and let ξ_{S^*} be the configuration on $\mathscr{H}_{S^*}$, obtained after extracting the relevant coordinates from ξ_{C^*}. Now, let $\xi_{C'}$ be a maximizing configuration of the potential on C' and such that the coordinates of $\xi_{C'}$ agree with ξ_{S^*} on S^*. Proceeding in this way, with the distribution of simple max-flows towards the periphery of $\mathscr{T}$, we obtain a collection $\{\xi_C : C \in \mathscr{C}\}$, and the junction tree property ensures that there exists a configuration $v \in \mathscr{H}$ such that $v_C = \xi_C$. Then v maximizes f over $\mathscr{H}$.

Now we are ready to describe the different phases of the algorithm.

Partition-phase. We need a short notation for the subsets that the algorithm generates: if $x^* \in \mathscr{H}$ and $\mathscr{H}^*_{R_j} \subseteq \mathscr{H}_{R_j}$ then the couple $(x^*, \mathscr{H}^*_{R_j})$ denotes the subset

$$\{z \in \mathscr{H} : z_{\mathscr{T}_{j-1}} = x^*_{\mathscr{T}_{j-1}}, z_{R_j} \in \mathscr{H}^*_{R_j}\}.$$

For instance, if $\mathscr{H}^*_{R_j} = \mathscr{H}_{R_j} \backslash \{x^1_{R_j}\}$, then $z \in (x^1, \mathscr{H}^*_{R_j})$ if and only if the coordinates of z for variables in $\mathscr{T}_{j-1}$ agree with x^1 and differ from x^1 for variables in R_j.

It turns out, as shown in the Candidate-phase, to be easy to compute the maximum of f over any subset of the form $(x^*, \mathscr{H}^*_{R_j})$. With this in mind we now construct a partitioning of $\mathscr{H} \backslash \{x^1\}$ such that each subset in the partitioning has the form $(x^*, \mathscr{H}^*_{R_j})$. If $\mathscr{H}^*_{R_i} = \mathscr{H}_{R_i} \backslash \{x^1_{R_i}\}$ then the subsets

$$(x^1, \mathscr{H}^*_{R_1}), \ldots, (x^1, \mathscr{H}^*_{R_p})$$

is such a partitioning. This partitioning is denoted $\mathscr{P}^1$.

Suppose we are able to determine x^2. Since $\mathscr{P}^1$ is a partitioning of $\mathscr{H} \backslash \{x^1\}$, x^2 must belong to one (and only one) of the subsets in $\mathscr{P}^1$, say $x^2 \in (x^1, \mathscr{H}^*_{R_j})$. Now we construct a partitioning of $\mathscr{H} \backslash \{x^1, x^2\}$ by refining the partitioning already performed. First we partition $(x^1, \mathscr{H}^*_{R_j}) \backslash \{x^2\}$ in subsets. If

$$\mathscr{H}'_{R_i} = \begin{cases} \mathscr{H}^*_{R_i} \backslash \{x^2_{R_i}\} & \text{for } i = j \\ \mathscr{H}_{R_i} \backslash \{x^2_{R_i}\} & \text{for } i > j \end{cases}$$

then the subsets

$$(x^2, \mathscr{H}'_{R_j}), \ldots, (x^2, \mathscr{H}'_{R_p}) \tag{10}$$

partition $(x^1, \mathscr{H}^*_{R_j}) \backslash \{x^2\}$. Thus the subsets in (10) united with $\mathscr{P}^1 \backslash \{(x^1, \mathscr{H}^*_{R_j})\}$ is a partitioning of $\mathscr{H} \backslash \{x^1, x^2\}$.

It is straightforward to extend this to the general case: by induction we assume that we are given a partitioning of $\mathscr{H} \backslash \{x^1, \ldots, x^L\}$, say $\mathscr{P}^L$ and such that every subset in $\mathscr{P}^L$ has the form $(x^*, \mathscr{H}^*_{R_j})$. Thus x^{L+1} maximizes f over some subset, say $(x^*, \mathscr{H}^*_{R_j})$ in $\mathscr{P}^L$. Let $\mathscr{H}'_{R_i} = \mathscr{H}^*_{R_i} \backslash \{x^{L+1}_{R_i}\}$ for

$i = j$ and $\mathscr{H}'_{R_i} = \mathscr{H}_{R_i} \backslash \{x^{L+1}_{R_i}\}$ for $i > j$, and note that the subsets

$$(x^{L+1}, \mathscr{H}'_{R_j}), \ldots, (x^{L+1}, \mathscr{H}'_{R_p}) \tag{11}$$

partition $(x^*, \mathscr{H}^*_{R_j}) \backslash \{x^{L+1}\}$. Therefore the subsets in (11) united with $\mathscr{P}^L \backslash \{(x^*, \mathscr{H}^*_{R_j})\}$ partition $\mathscr{H} \backslash \{x^1, \ldots, x^{L+1}\}$.

Candidate-phase. Every subset we generate in partition-phase, has the form $(x^*, \mathscr{H}^*_{R_j})$, where $x^* \in \mathscr{H}$ and $\mathscr{H}^*_{R_j} \subset \mathscr{H}_{R_j}$. In order to compute the maximum of f over the subset $(x^*, \mathscr{H}^*_{R_j})$ we may insert the evidence

$$X_{C_1} = x^*_{c_1}, \ldots, X_{C_{j-1}} = x^*_{C_{j-1}}, \tag{12}$$

and

$$X_{C_j} \in \{z_{C_j} : z_{S_{j-1}} = x^*_{S_{j-1}}, z_{R_j} \in \mathscr{H}^*_{R_j}\}$$

and subsequently collect active max-flows towards an arbitrary root-clique C^*, say. From the new potential on C^* we can compute the maximum of f over $(x^*, \mathscr{H}^*_{R_j})$. But the method does require a large number of max-flows to be passed in the junction tree. However Theorem 3 provides a method to find the maximum of f over every subset $(x^*, \mathscr{H}^*_{R_j})$ directly from the max-marginal charge of f.

Theorem 3. The maximum of f over $(x^*, \mathscr{H}^*_{R_j})$ is given by

$$\prod_{i=1}^{j-1} \frac{\hat{f}_{C_i}(x^*_{C_i})}{\hat{f}_{S_i}(x^*_{S_i})} \max_{z_{C_j}} \tilde{f}_{C_j}(z_{C_j}),$$

where $\tilde{f}_{C_j}$ is defined as

$$\tilde{f}_{C_j}(z_{C_j}) = \begin{cases} \hat{f}_{C_j}(z_{C_j}) & \text{if } z_{S_{j-1}} = x^*_{S_{j-1}},\ z_{R_j} \in \mathscr{H}^*_{R_j} \\ 0 & \text{otherwise.} \end{cases}$$

Recall that the subsets

$$(x^1, \mathscr{H}^*_{R_1}), \ldots, (x^1, \mathscr{H}^*_{R_p}),$$

where $\mathscr{H}^*_{R_j} = \mathscr{H}_{R_j} \backslash \{x^1_{R_j}\}$, is a partitioning of $\mathscr{H} \backslash \{x^1\}$. In this special case Theorem 3 reduces to

Corollary 1. The maximum of f over the subset $(x^1, \mathscr{H}_{R_j} \backslash \{x^1_{R_j}\})$ is given by

$$\max_{z \in \mathscr{H}_{C_j}} \{\hat{f}_{C_j}(z) : z_{S_{j-1}} = x^1_{S_{j-1}}, z_{R_j} \neq x^1_{R_j}\}.$$

Proof. Follows from Theorem 3, since $\hat{f}_{S_i}(x^1_{S_i}) = \hat{f}_{C_i}(x^1_{C_i}) = f(x^1)$ for $i = 1, \ldots, p$. □

Proof (Theorem 3). Without loss of generality, assume that the charge on $\mathscr{T}$ is the max-marginal charge of f, $(\{\hat{f}_C : C \in \mathscr{C}\}, \{\hat{f}_S : S \in \mathscr{S}\})$. In order to find the maximum of f over the subset $(x^*, \mathscr{H}^*_{R_j})$, we incorporate the evidence (12) in $\mathscr{T}$ as follows: define a new potential $\tilde{f}_{C_i}$ on clique $C_i (i < j)$ as

$$\tilde{f}_{C_i}(z_{C_i}) = \begin{cases} \hat{f}_{C_i}(z_{C_i}) & \text{if } z_{C_i} = x^*_{C_i} \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

and define a new potential $\tilde{f}_{C_j}$ on clique C_j as

$$\tilde{f}_{C_j}(z_{C_j}) = \begin{cases} \hat{f}_{C_j}(z_{C_j}) & \text{if } z_{S_{j-1}} = x^*_{S_{j-1}},\ z_{R_j} \in \mathscr{H}^*_{R_j} \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

All other potentials are unaltered, that is, the potential on clique $C_i (i > j)$ is $\hat{f}_{C_i}$, and the potential on separator S is $\hat{f}_S$ for all $S \in \mathscr{S}$.

Now we collect max-flows towards clique C_j, starting from the charge above. We may use the ordering in (8) for this purpose: note that the passage of max-flows from $C_p, \ldots, C_{j+1}$ towards C_j, leaves the potentials unaltered, because of max-consistency. Furthermore, by an induction argument we note that after the passage of max-flows from $C_1, \ldots, C_{j-1}$ towards C_j, we obtain the update factor $\lambda_{S_{j-1}}$ on S_{j-1} given by

$$\lambda_{S_{j-1}}(z_{S_{j-1}}) = \begin{cases} \prod_{i=1}^{j-1} \frac{\hat{f}_{C_i}(x^*_{C_i})}{\hat{f}_{S_i}(x^*_{S_i})} & \text{if } z_{S_{j-1}} = x^*_{S_{j-1}} \\ 0 & \text{otherwise} \end{cases}$$

such that the new potential on C_j becomes

$$\tilde{f}_{C_j}(z_{C_j}) \lambda_{S_{j-1}}(z_{S_{j-1}}) = \prod_{i=1}^{j-1} \frac{\hat{f}_{C_i}(x^*_{C_i})}{\hat{f}_{S_i}(x^*_{S_i})} \tilde{f}_{C_j}(z_{C_j}). \tag{15}$$

Now the result follows, since the maximum of f over $(x^*, \mathscr{H}^*_{R_j})$ is given by the maximum of (15), the potential on clique C_j after collecting max-flows towards C_j. □

Identification-phase. Theorem 3 states how to find the maximum of f over any subset of the form $(x^*, \mathscr{H}^*_{R_j})$. Now we show how to identify the maximizing configuration in $(x^*, \mathscr{H}^*_{R_j})$.

Without loss of generality, we assume that the charge on $\mathscr{T}$ is the max-marginal charge of f. In the proof of Theorem 3, we showed that after incorporating the evidence (12) in $\mathscr{T}$, we obtain the following charge on $\mathscr{T}$: the potential on C_j is $\tilde{f}_{C_j}$ given in (14), the potential on clique $C_i (i < j)$ is $\tilde{f}_{C_i}$ given in (13), and the potentials on the other cliques are unaltered. Finally, the potential on the separator S is $\hat{f}_S$ for all $S \in \mathscr{S}$.

Subsequently we collect max-flows towards root-clique C_j and obtain the following potentials: the potential on $C_i (i > j)$ is unaltered, and the potential on C_j is given in (15). The potential on $C_i (i < j)$ is of no interest to us for the following reason: in order to identify the maximizing configuration of f over $(x^*, \mathscr{H}^*_{R_j})$, we distribute max-flows (from the charge, obtained after collecting max-flows towards C_j) starting from the root-clique C_j; however, the flows from C_j towards $C_1, \ldots, C_{j-1}$ are vacuous since for all configurations ξ in $(x^*, \mathscr{H}^*_{R_j})$ we have that

$$\xi_{C_1} = x^*_{C_1}, \ldots, \xi_{C_{j-1}} = x^*_{C_{j-1}}.$$

So the routine starts by identifying a configuration, ξ_{C_j}, say, which maximizes (15), the potential on C_j (that depends on $\hat{f}_{C_j}$ only), and proceed towards clique C_{j+1}: let

C^* be the clique in $\mathscr{T}_j$ that is a neighbour to C_{j+1} (separator S_j). Define ξ_{S_j} by extracting the relevant coordinates from ξ_{C^*}. Now, let $\xi_{C_{j+1}}$ be the maximizing configuration of $\hat{f}_{C_{j+1}}$ (the potential on C_{j+1}) that agree with ξ_{S_j} for coordinates on S_j. Proceeding in this way, we distribute simple max-flows towards the remaining cliques $C_{j+2}, \ldots, C_p$, and obtain a collection $\{\xi_C : C \in \mathscr{C}\}$. The configuration v, say, given as $v_C = \xi_C$ for $C \in \mathscr{C}$ maximizes f over $(x^*, \mathscr{H}^*_{R_j})$.

The above routine is termed *max-distribution* $(x^*, \mathscr{H}^*_{R_j})$ *in* $\mathscr{T}$. So, max-distribution $(x^*, \mathscr{H}^*_{R_j})$ in $\mathscr{T}$ can be regarded as a routine, that starts with the max-marginal charge of f on $\mathscr{T}$, and pass 'simple' max-flows from clique C_j towards $C_{j+1}, \ldots, C_p$.

4.4. *The algorithm*

We will now show how the algorithm MFP finds $x^1, \ldots, x^M$. A pseudo-code for our algorithm is given below.

Step 1.

- Find the max-marginal charge (Theorem 1).
- Find max f (Theorem 2).
- Identify x^1 by max-distribution in $\mathscr{T}$.
- Let $\mathscr{P}^0 \equiv \mathscr{H}$ and $L = 1$.

Step 2. Partition-phase:

- If $L = 1$ then let $\mathscr{H}'_{R_i} = \mathscr{H}_{R_i} \backslash \{x^1_{R_i}\}$, and note that the subsets $(x^1, \mathscr{H}'_{R_i})$ for $i \geq 1$ partition $\mathscr{H} \backslash \{x^1\}$. Let this collection of subsets be denoted $\mathscr{P}^1$, and go to **Step 3.**
- If $L > 1$ then x^L must belong to some subset $(x^*, \mathscr{H}^*_{R_j})$, say, where $(x^*, \mathscr{H}^*_{R_j}) \in \mathscr{P}^{L-1}$.
- Define

$$\mathscr{H}'_{R_i} = \begin{cases} \mathscr{H}^*_{R_i} \backslash \{x^L_{R_i}\} & \text{for } i = j \\ \mathscr{H}_{R_i} \backslash \{x^L_{R_i}\} & \text{for } i > j \end{cases}$$

  and note that the subsets $(x^L, \mathscr{H}'_{R_i})$ for $i \geq j$ partition $(x^*, \mathscr{H}^*_{R_j}) \backslash \{x^L\}$.
- Let the collection of subsets

$$\left(\mathscr{P}^{L-1} \backslash \{(x^*, \mathscr{H}^*_{R_j})\}\right) \cup \left(\{(x^L, \mathscr{H}'_{R_i}) : i \geq j\}\right)$$

  be denoted $\mathscr{P}^L$.

Step 3. Candidate-phase:

- Find the maximum of f over the subsets $(x^L, \mathscr{H}'_{R_i})$ for $i \geq j$ (Theorem 3).
- Determine $f(x^{L+1})$. (Note that the maximum of f over each subsets in $\mathscr{P}^L$ is known from which we can determine $f(x^{L+1})$.)

Step 4. Identification-phase:

- x^{L+1} belongs to some subset $(x^*, \mathscr{H}^*_{R_j})$, say, where $(x^*, \mathscr{H}^*_{R_j}) \in \mathscr{P}^L$.
- Identify x^{L+1} by max-distribution $(x^*, \mathscr{H}^*_{R_j})$ in $\mathscr{T}$.

Step 5. If $L + 1 < M$ then $L := L + 1$ and go back to **Step 2.**

Example 3. Now we will identify the four most probable configurations in ASIA with the algorithm in Section 4.4.

Recall from Example 2 that $x^1 = (a, \bar{t}, e, l, x, b, s, d)$ is the most probable configuration consistent with the evidence $\alpha = a$, $\xi = x$ and $\sigma = s$.

Partition-phase: if $\mathscr{H}^*_{R_i} = \mathscr{H}_{R_i} \backslash \{x^1_{R_i}\}$ then the subsets $(x^1, \mathscr{H}^*_{R_i})$ for $i = 1, \ldots, 6$ are given by

$$\begin{aligned} (x^1, \mathscr{H}^*_{R_1}) &= \{z : z_{C_1} \neq x^1_{C_1}\} \\ (x^1, \mathscr{H}^*_{R_2}) &= \{z : z_{C_1} = x^1_{C_1}, z_{R_2} \neq x^1_{R_2}\} \\ &\ldots \\ (x^1, \mathscr{H}^*_{R_6}) &= \{z : z_{\mathscr{T}_5} = x^1_{\mathscr{T}_5}, z_{R_6} \neq x^1_{R_6}\} \end{aligned}$$

and the subsets partition $\mathscr{H} \backslash \{x^1\}$.

Note that a configuration belongs to $(x^1, \mathscr{H}^*_{R_i})$ if and only if its coordinates for variables in $\mathscr{T}_{i-1}$ agree with the coordinates of x^1, and its coordinates for variables in R_i differ from x^1. The subsets are given in Table 3.

It is clear that $(x^1, \mathscr{H}^*_{R_6})$ consists of one configuration, $(x^1, \mathscr{H}^*_{R_5})$ consists of two configurations, $(x^1, \mathscr{H}^*_{R_4})$ consists of four configurations, $(x^1, \mathscr{H}^*_{R_3})$ consists of eight configurations, $(x^1, \mathscr{H}^*_{R_2})$ consists of 48 configurations, and $(x^1, \mathscr{H}^*_{R_1})$ consists of 192 configurations. So the total number of configurations in the subsets is 255, which is one less than the total number of configurations in $\mathscr{H}$, as expected, since the subsets partition $\mathscr{H} \backslash \{x^1\}$.

Candidate-phase: Corollary 1 shows that the maximum of f over $(x^1, \mathscr{H}^*_{R_j})$ is given by

$$\max_{z_{C_j} \in \mathscr{H}_{C_j}} \{\hat{f}_{C_j}(z_{C_j}) : z_{S_{j-1}} = x^1_{S_{j-1}}, z_{R_j} \neq x^1_{R_j}\}.$$

For instance we have the maximum of f over $(x^1, \mathscr{H}^*_{R_2})$ is given by

$$\begin{aligned} \max\{\hat{f}_{C_2}(z_{C_2}) : z_{S_1} = \bar{t}, z_{R_2} \neq (e, l)\} &= \hat{f}_{C_2}(\bar{t}, \bar{e}, \bar{l}) \\ &= 0.770 \cdot 10^{-4}. \end{aligned}$$

The maximum of f over every subset is shown in Fig. 3, from which we see that the maximum of f over $\mathscr{H} \backslash \{x^1\}$ is the maximum of f over $(x^1, \mathscr{H}^*_{R_4})$, and given by $0.931 \cdot 10^{-4}$.

Table 3. *Partition-phase in ASIA:* $(x^1, \mathscr{H}^*_{R_1}), \ldots, (x^1, \mathscr{H}^*_{R_6})$ *partition* $\mathscr{H} \backslash \{(a, \bar{t}, e, l, x, b, s, d)\}$

$(x^1, \mathscr{H}^*_{R_1}) = \{z \in \mathscr{H} : z_{C_1} \neq (a, \bar{t})\}$
$(x^1, \mathscr{H}^*_{R_2}) = \{z \in \mathscr{H} : z_{C_1} = (a, \bar{t}), z_{R_2} \neq (e, l)\}$
$(x^1, \mathscr{H}^*_{R_3}) = \{z \in \mathscr{H} : z_{\mathscr{T}_2} = (a, \bar{t}, l, e), z_{R_3} \neq x\}$
$(x^1, \mathscr{H}^*_{R_4}) = \{z \in \mathscr{H} : z_{\mathscr{T}_3} = (a, \bar{t}, l, e, x), z_{R_4} \neq b\}$
$(x^1, \mathscr{H}^*_{R_5}) = \{z \in \mathscr{H} : z_{\mathscr{T}_4} = (a, \bar{t}, l, e, x, b), z_{R_5} \neq s\}$
$(x^1, \mathscr{H}^*_{R_6}) = \{z \in \mathscr{H} : z_{\mathscr{T}_5} = (a, \bar{t}, l, e, x, b, s), z_{R_6} \neq d\}$

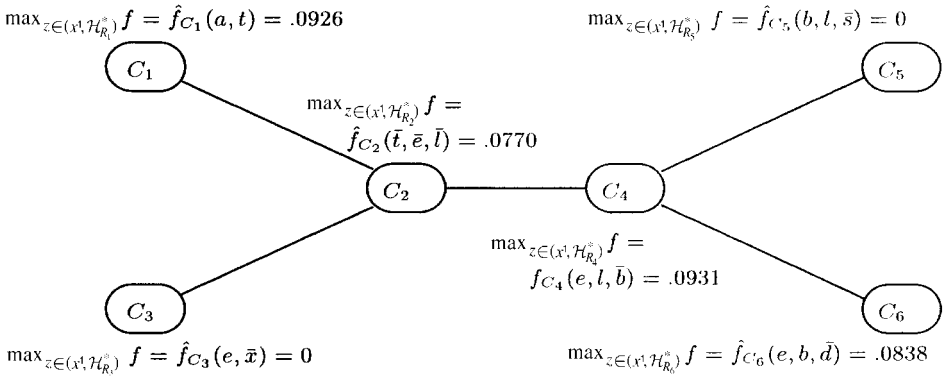

Fig. 3. *Candidate-phase for ASIA: Maximum of f over $(x^1, \mathcal{H}^*_{R_1}),\ldots,(x^1, \mathcal{H}^*_{R_6})$ (multiplied by 10^3)*

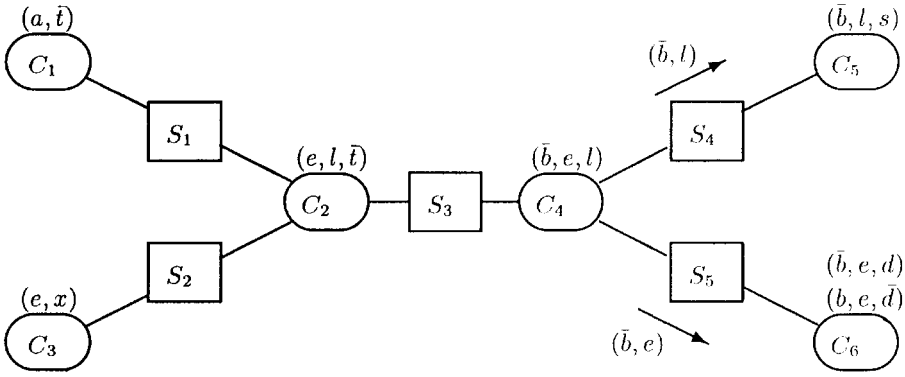

Fig. 4. *Identification-phase in ASIA: Max-distribution $(x^1, \mathcal{H}^*_{R_4})$ in $\mathcal{T}$ gives $(a, \bar{t}, e, l, x, \bar{b}, s, \bar{d})$ and $(a, \bar{t}, e, l, x, \bar{b}, s, d)$*

Identification-phase: From the candidate-phase we know that x^2 maximizes f over $(x^1, \mathcal{H}^*_{R_4})$. x^2 is identified by max-distribution $(x^1, \mathcal{H}^*_{R_4})$ in $\mathcal{T}$ (Fig. 4). Note that the maximizing configuration is not unique: both $(a,\bar{t},e,l,x,\bar{b},s,d)$ and $(a,\bar{t},e,l,x,\bar{b},s,\bar{d})$ maximizes f. In this case, we arbitrarily associate x^2 with any of the two configurations, say the former, and let x^3 be the latter.

At this stage we have found the three most probable configurations. We will now find the fourth most probable configuration by going through the same phases as above.

Partition-phase: We now partition $\mathcal{H}\backslash\{x^1,x^2,x^3\}$ by refining the partition already performed. Since x^2 and x^3 are configurations in $(x^1, \mathcal{H}^*_{R_4})$ we partition $(x^1, \mathcal{H}^*_{R_4})\backslash\{x^2,x^3\}$. This is done in two steps. First we partition $(x^1, \mathcal{H}^*_{R_4})\backslash\{x^2\}$ (see (10)). Let

Table 4. *Partition-phase in ASIA: $(x^2, \bar{cal}H_{R_4})$, $(x^2, \bar{\mathcal{H}}_{R_5})$, $(x^2, \bar{\mathcal{H}}_{R_6})$ partition $(x^1, \bar{\mathcal{H}}^*_{R_4})\backslash\{(a, \bar{t}, e, l, x, \bar{b}, s, d)\}$*

$(x^2, \bar{\mathcal{H}}_{R_4}) = \{z \in \mathcal{H} : z_{\mathcal{T}_3} = (a,\bar{t},e,l,x), z_{R_4} \notin \{\bar{b}, b\}\}$
$(x^2, \bar{\mathcal{H}}_{R_5}) = \{z \in \mathcal{H} : z_{\mathcal{T}_4} = (a,\bar{t},e,l,x,\bar{b}), z_{R_5} \neq s\}$
$(x^2, \bar{\mathcal{H}}_{R_6}) = \{z \in \mathcal{H} : z_{\mathcal{T}_5} = (a,\bar{t},e,l,x,\bar{b},s), z_{R_6} \neq \bar{d}\}$

$$\bar{\mathcal{H}}_{R_j} = \begin{cases} \mathcal{H}^*_{R_j}\backslash\{x^2_{R_j}\} & \text{for } j = 4 \\ \mathcal{H}_{R_j}\backslash\{x^2_{R_j}\} & \text{for } j = 5,6 \end{cases}$$

and note that $(x^2, \bar{\mathcal{H}}_{R_4}), (x^2, \bar{\mathcal{H}}_{R_5})$, and $(x^2, \bar{\mathcal{H}}_{R_6})$ partition $(x^2, \mathcal{H}^*_{R_4})\backslash\{x^2\}$ (see Table 4). However $(x^2, \bar{\mathcal{H}}_{R_4}) = \emptyset$ (since $\bar{\mathcal{H}}_{R_4} = \emptyset$), and is therefore ignored in the following analysis.

Secondly we note that $x^3 \in (x^2, \bar{\mathcal{H}}_{R_6})$ and therefore we partition $(x^2, \bar{\mathcal{H}}_{R_6})\backslash\{x^3\}$. This is easy, since $(x^2, \bar{\mathcal{H}}_{R_6}) = \{x^3\}$, such that $(x^2, \bar{\mathcal{H}}_{R_6})\backslash\{x^3\} = \emptyset$. Accordingly the partitioning of $(x^1, \mathcal{H}^*_{R_4})\backslash\{x^2,x^3\}$ consists of only one subset, namely $(x^2, \bar{\mathcal{H}}_{R_5})$. Thus the subsets in Table 3, excluding $(x^1, \mathcal{H}^*_{R_4})$, and $(x^2, \bar{\mathcal{H}}_{R_5})$ is a partitioning of $\mathcal{H}\backslash\{x^1,x^2,x^3\}$.

Candidate-phase: the maximum of f over $\mathcal{H}\backslash\{x^1,x^2,x^3\}$ is the maximum of f over the subsets in Table 3 (excluding $(x^1, \mathcal{H}^*_{R_4})$) and $(x^2, \bar{\mathcal{H}}_{R_5})$. Figure 3 shows the maximum of f over the subsets in Table 3, and the maximum of f over $(x^2, \bar{\mathcal{H}}_{R_5})$ is given from Theorem 3 by

$$\prod_{i=1}^{4} \frac{\hat{f}_{C_i}(x^2_{C_i})}{\hat{f}_{S_i}(x^2_{S_i})} \max\{\hat{f}_{C_5}(z_{C_5}) : z_{S_4} = x^2_{S_4}, z_{R_5} \in \bar{\mathcal{H}}_{R_5}\},$$

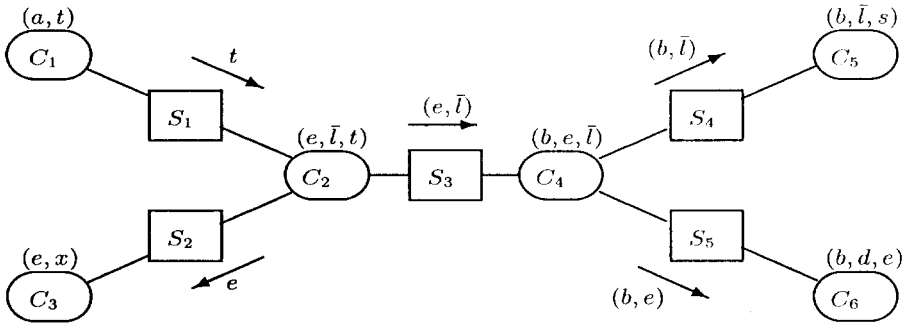

Fig. 5. *Identification-phase in ASIA: Max-distribution $(x^1, \mathscr{H}^*_{R_1})$ in $\mathscr{T}$ gives (a, t, e, $\bar{l}$, x, b, s, d)*

which is zero, since $x^2_{S_4} = (\bar{b}, l)$ and $\bar{\mathscr{H}}_{R_5} = \mathscr{H}_{R_5} \backslash \{x^2_{R_5}\} = \bar{s}$, and $\hat{f}_{C_5}(\bar{b}, l, \bar{s}) = 0$ (see Table 2).

Therefore the maximum of f over $\mathscr{H} \backslash \{x^1, x^2, x^3\}$ is the maximum of f over the subsets in Table 3, excluding $(x^1, \mathscr{H}^*_{R_4})$, which is $0.926 \cdot 10^{-4}$, and the maximizing configuration belongs to $(x^1, \mathscr{H}^*_{R_1})$.

Identification-phase: the candidate phase shows that the fourth most probable configuration belongs to $(x^1, \mathscr{H}^*_{R_1})$, and the maximizing configuration of f over $(x^1, \mathscr{H}^*_{R_1})$ is $(a, t, e, \bar{l}, x, b, s, d)$, obtained by max-distribution $(x^1, \mathscr{H}^*_{R_1})$ in $\mathscr{T}$ (see Fig. 5). □

4.5. *Fast max-flow propagation*

The algorithm MFP as described in the last subsection uses the max-marginal charge of f for finding $x^1, \ldots, x^M$. However we need not compute the max-marginal charge for finding these configurations.

Let $K^* = (\{f^*_C : C \in \mathscr{C}\}, \{f^*_S : S \in \mathscr{S}\})$ denote the charge, obtained after collecting active max-flows from the leaves towards any root-clique, say C_1.

It may be seen that all the considerations and results in Section 4.3 continue to apply, as long as we everywhere replace $\hat{f}_C$ (and $\hat{f}_S$) with f^*_C (and f^*_S).

For instance, Theorem 3, with this modification, becomes

Theorem 4. The maximum of f over $(x^*, \mathscr{H}^*_{R_j})$ is given by

$$\prod_{i=1}^{j-1} \frac{f^*_{C_i}(x^*_{C_i})}{f^*_{S_i}(x^*_{S_i})} \max_{z_{C_j}} \tilde{f}_{C_j}(z_{C_j}),$$

where $\tilde{f}_{C_j}$ is defined as

$$\tilde{f}_{C_j}(z_{C_j}) = \begin{cases} f^*_{C_j}(z_{C_j}) & \text{if } z_{S_{j-1}} = x^*_{S_{j-1}},\ z_{R_j} \in \mathscr{H}^*_{R_j} \\ 0 & \text{otherwise.} \end{cases}$$

Proof. The proof is similar to the proof of Theorem 3. □

Furthermore, the routine max-distribution $(x^*, \mathscr{H}^*_{R_j})$ in $\mathscr{T}$, described in Section 4.3 also works when we start from the charge K^* instead of the max-marginal charge of f on $\mathscr{T}$.

5. Complexity analysis

Suppose that a function f is factorizing on $\mathscr{T}$. Finding the maximizing value x^1 of f has for almost a decade been known to be feasible. We would claim that MFP, described in Section 4.4, provides a computationally feasible way of computing $x^1, \ldots, x^M$, even for large M.

We shall discuss the computational complexity of the fast version of MFP described in Section 4.5.

We introduce the quantities:

$$|C| = |\mathscr{H}_C|$$
$$= \text{no. of configurations in } \mathscr{H}_C.$$
$$|\hat{C}| = \max_C |\mathscr{H}_C|$$
$$= \text{ largest configuration space of a clique.}$$
$$|\bar{C}| = \frac{\sum_C |\mathscr{H}_C|}{p}$$
$$= \text{average no. of configurations of a clique.}$$

For the number of configurations in the separators S and the residuals R we use a similar notation.

5.1. *The complexity of max-flow propagation*

The complexity of MFP can be divided into three parts:

- Find the charge K^*.
- Candidate-phase: the maximum of f over the subsets $(x^*, \mathscr{H}^*_R)$ generated in the partition-phase is found, and the candidates are compared for each step $L = 1, \ldots, M$.
- Identification-phase: $x^1, \ldots, x^M$ are identified by performing M max-distributions $(x^*, \mathscr{H}^*_R)$ in $\mathscr{T}$.

Passing a max-flow from C to C' involves $|C| + |C'|$ computations. Thus the total number of computations for finding K^* is less than

$$2p|\bar{C}|.$$

For each step L in the algorithm, we generate at most p subsets of the form $(x^*, \mathscr{H}^*_R)$. Finding the maximum of f over $(x^*, \mathscr{H}^*_R)$, takes place via Theorem 4. So computing all the candidates demands at most

$$Mp|\bar{R}|$$

computations.

The total number of candidates in step L does not exceed pL elements. If we store the elements in a 2-3 search tree then the computations needed for inserting p new elements, and update the search tree uses $O(p \log pL)$ comparisons. Doing this for $L = 1, \ldots, M$, the number of comparisons needed is in the order of

$$\sum_{L=1}^{M} p \log(pL) = O(Mp \log pM).$$

Finally, performing 'simple' max-distributions $(x^*, \mathscr{H}_R^*)$ in $\mathscr{T}$, is a worst case when $R = R_1$. Then the number of comparisons needed is no larger than $p|\bar{R}|$. Thus the total number of comparisons we need to perform M max-distributions cannot exceed

$$Mp|\bar{R}|.$$

Adding up these terms, we obtain that the whole process of finding $x^1, \ldots, x^M$ using MFP is in the order of

$$2p|\bar{C}| + 2Mp|\bar{R}| + pM \log pM. \tag{16}$$

For small M the terms $2Mp|\bar{R}|$ and $2p|\bar{C}|$ are the dominant ones. In this case it may be seen from (16) that we can find $x^1, \ldots, x^M$ when $M = |\bar{S}|$, the average number of configurations in $\mathscr{H}_S$ for $S \in \mathscr{S}$, using $2p|\bar{C}| + 2|\bar{S}|p|\bar{R}| \approx 4p|\bar{C}|$ computations. This is similar to the number of computations needed for finding the max-marginal charge!

Now we shall compare the complexity of MFP with another algorithm, named Fast Belief Revision (Seroussi and Golmard, 1994). Fast Belief Revision (FBR) has complexity of the order

$$pM|\hat{S}|\big(|\hat{R}| + M|\hat{N}|\big), \tag{17}$$

where $|\hat{N}|$ is the maximum number of children to any clique in the junction tree.

Table 5 shows a comparison between MFP and FBR as a function of the quantities $M, p, |\hat{N}|, |\hat{R}|$ and $|\hat{S}|$. The comparison is such that four of the quantities are fixed at a time and one of the quantities varies.

The complexity of MFP is computed from (16) after replacing $|\bar{R}|$ with $|\hat{R}|$, and replacing $|\bar{C}|$ with $|\hat{S}|\,|\hat{R}|$ (note that $|\bar{C}| \leq |\hat{S}|\,|\hat{R}|$).

The complexity of FBR is computed from (17). The parameter $|\hat{N}|$, which only plays a role for the complexity of FBR, is chosen as small as possible, namely 1. Consequently the numbers in Table 5 are chosen to benefit FBR rather than MFP. Nevertheless Table 5 clearly shows that MFP is superior to FBR with the chosen values of the quantities. Other values give a similar result. When we compare (16) with (17) we see that for large M, MFP is of the order

$$\frac{|\hat{S}|\,|\hat{N}|M}{\log M}$$

faster than FBR.

Indeed MFP is faster than any other algorithm known to the author.

6. Solving related problems

Suppose that a function f, factorizing on $\mathscr{T}$, is specified by a representation K^0. We shall now solve some problems that are related to finding the maxima of f.

Recall that f_A for $A \subset U$ is defined as

$$f_A = \sum_{U \setminus A} f.$$

6.1. *Maximization of f_A*

Suppose we want to find the M highest values of f_A, for some $A \subset U$. If A is a subset of a clique C, then this is trivial, since we are given f_C after passing active sum-flows in the junction tree and equilibrium has been reached. f_A is then found by further marginalization and the M highest values of f_A can be found by examining f_A further.

If the subset A corresponds to the variables in a subtree $\mathscr{T}_A$, say, then the task is also easy: when we pass active sum-flows from the leaves towards $\mathscr{T}_A$ we obtain a new charge on $\mathscr{T}_A$ which is a representation of f_A (see Dawid, 1992). Subsequently we can find the M highest values of f_A using MFP on $\mathscr{T}_A$.

Finally, if A is associated with several disconnected subtrees, Xu (1994) provides a method for transforming the initial junction tree $\mathscr{T}$ into another junction tree, say $\mathscr{T}_A$ and such that f_A factorizes on $\mathscr{T}_A$. The new charge on $\mathscr{T}_A$ is

Table 5. *The complexity of MFP and FBR*

M	$\lvert\hat{N}\rvert$	p	$\lvert\hat{R}\rvert$	$\lvert\hat{S}\rvert$	MFP	FBR
M	1	10	10	10	$2000 + 200M + 10M \log 10M$	$1000M + 100M^2$
10	$\lvert\hat{N}\rvert$	10	10	10	4461	$10000 + 10000\lvert\hat{N}\rvert$
10	1	p	10	10	$400p + 10p \log 10p$	$2000p$
10	1	10	$\lvert\hat{R}\rvert$	10	$461 + 400\lvert\hat{R}\rvert$	$10000 + 1000\lvert\hat{R}\rvert$
10	1	10	10	$\lvert\hat{S}\rvert$	$2461 + 200\lvert\hat{S}\rvert$	$2000\lvert\hat{S}\rvert$

a representation of f_A. The M highest values of f_A can now be found using MFP on $\mathscr{T}_A$. Example 4 shows how Xu's method works on ASIA.

Example 4. Suppose we are interested in finding the marginal joint probability function of α, β and τ in ASIA. We shall write this probability function as $f_{\alpha\beta\tau}$, and we use this notation throughout this example. So the joint probability function can be written as

$$f_{\alpha\beta\tau\xi\lambda\sigma\delta\epsilon} = \frac{a_{\alpha\tau}a_{\epsilon\lambda\tau}a_{\epsilon\xi}a_{\beta\epsilon\lambda}a_{\beta\lambda\sigma}a_{\beta\delta\epsilon}}{b_\tau b_\epsilon b_{\epsilon\lambda} b_{\beta\epsilon} b_{\beta\lambda}}, \tag{18}$$

where the a and b are known functions. $f_{\alpha\beta\tau}$ is found by marginalizing out $\epsilon, \lambda, \sigma, \delta$ and ξ from (18). First we marginalize σ, δ and ξ out by computing

$$b^*_{\beta\lambda} = \sum_\sigma a_{\beta\lambda\sigma} \qquad b^*_{\beta\epsilon} = \sum_\delta a_{\beta\delta\epsilon} \qquad b^*_\epsilon = \sum_\xi a_{\epsilon\xi}. \tag{19}$$

Let

$$a^*_{\epsilon\lambda\tau} = a_{\epsilon\lambda\tau}\frac{b^*_\epsilon}{b_\epsilon} \qquad a^*_{\beta\epsilon\lambda} = a_{\beta\epsilon\lambda}\frac{b^*_{\beta\lambda}}{b_{\beta\lambda}}\frac{b^*_{\beta\epsilon}}{b_{\beta\epsilon}} \tag{20}$$

such that

$$\begin{aligned} f_{\alpha\beta\tau\epsilon\lambda} &= \sum_\sigma \sum_\delta \sum_\xi f_{\alpha\beta\tau\xi\lambda\sigma\delta\epsilon} \\ &= \frac{a_{\alpha\tau}a^*_{\beta\epsilon\lambda}a^*_{\epsilon\lambda\tau}}{b_\tau b_{\epsilon\lambda}}. \end{aligned} \tag{21}$$

Then we marginalize ϵ and λ out from (21) by computing

$$a^{**}_{\tau\beta} = \sum_\epsilon \sum_\lambda \frac{a^*_{\epsilon\lambda\tau}a^*_{\beta\epsilon\lambda}}{b_{\epsilon\lambda}} \tag{22}$$

and we get

$$\begin{aligned} f_{\alpha\beta\tau} &= \sum_\epsilon \sum_\lambda f_{\alpha\beta\tau\epsilon\lambda} \\ &= \frac{a^{**}_{\tau\beta}a_{\alpha\tau}}{b_\tau}. \end{aligned} \tag{23}$$

The above computations can be described using the junction tree in Fig. 1: the computations (19) and (20) correspond to passing sum-flows towards the smallest (connected) subtree that contains α, β and τ (that is the subtree containing the cliques C_1, C_2 and C_4). Now the new charge on the subtree, consisting of the cliques C_1, C_2, and C_4, is a representation of $f_{\alpha\beta\tau\epsilon\lambda}$ (see (21)).

The computations (22) correspond to 'deleting' variables, which are different from α, β and τ in the subtree.

Finally the factorization (23) of $f_{\alpha\beta\tau}$ can be used for finding the M most probable configurations of $f_{\alpha\beta\tau}$ using MFP. □

6.2. *No unique maximum*

If there is no unique maximum, it may be of interest to find the number of configurations where the maximum is achieved. Robert Cowell suggested to me a method for doing this. Suppose we are given the max-marginal charge of f, $(\{\hat{f}_C : C \in \mathscr{C}\}, \{\hat{f}_S : S \in \mathscr{S}\})$. Define f'_C for all $C \in \mathscr{C}$ as

$$f'_C(x_C) = \begin{cases} 1 & \text{if } \hat{f}_C(x_C) = f(x^1) \\ 0 & \text{otherwise} \end{cases}$$

and $f'_S(x_S) \equiv 1$, for all $S \in \mathscr{S}$ and all $x_S \in \mathscr{H}_S$.

Starting from the charge $K' = (\{f'_C : C \in \mathscr{C}\}, \{f'_S : S \in \mathscr{S}\})$ and collecting sum-flows towards any root-clique, say C_1, gives a new clique-potential on C_1, say $f^*_{C_1}$. It can be seen that $\sum_{C_1} f^*_{C_1}$ gives the number of configurations where the maximum is achieved.

6.3. *Minimization*

As long as f is strictly positive, MFP, with obvious modifications, can be applied to find the M smallest values of f, and where these values are achieved.

If f can be zero, however, we can find the M smallest non-zero values of f, simply by ignoring configurations of potentials whose value is zero.

Acknowledgements

I would like to express my sincere gratitude to Finn V. Jensen for critical readings of earlier versions of this article leading to several improvements. I am also indebted to Phil Dawid and Steffen L. Lauritzen for their encouragement. Finally, a special thanks to Robert Cowell and the anonymous referees for constructive criticism and good advice.

References

Andersen, S. K., Olesen, K. G., Jensen, F. V. and Jensen, F. (1990) HUGIN – a shell for building Bayesian belief universes for expert systems. In G. Shafer and J. Pearl (eds) *Readings in Uncertainty*, pp. 332–337. San Francisco: Morgan Kaufmann.

Cowell, R. G. (1992) BAIES – a probabilistic expert system shell with qualitative and quantitative learning. In J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (eds) *Bayesian Statistics* 4th edn, pp. 595–600. Oxford: Clarendon Press.

Dawid, A. P. (1992) Applications of a general propagation algorithm for probabilistic expert systems. *Statistics and Computing*, **2**, 25–36.

Jensen, F. V. (1988) Junction trees and decomposable hypergraphs. Judex Research Report, Judex Datasystemer A/S, Aalborg, Denmark F. V. (1990) Calculation in HUGIN of probabilities.

Jensen, F. V., Olesen, K. G. and Andersen, S. K. (1990) An algebra of Bayesian belief universes for knowledge-based systems. *Networks*, **20**, 637–59.

Jensen, F. V. (1996) *Introduction to Bayesian Networks*. London: UCL Press.

Lauritzen, S. L., Speed, T. P. and Vijayan, K. (1984) Decomposable graphs and hypergraphs. *Journal of the Australian Mathematical Society*, *Series A*, **36**, 12–29.

Lauritzen, S. L. and Spiegelhalter, D. J. (1988) Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *Journal of the Royal Statistical Society Series B*, **50**, 157–224.

Li, Z. and D'Ambrosio, B. (1993) An efficient approach for finding MAP assignments to belief networks. In Proceedings of the 9th Conference on Uncertainty in AI, pp. 342–9. Washington, Morgan Kaufmann.

Nilsson, D. (1994) An algorithm for finding the *M* most probable configurations of discrete variables that are specified in probabilistic expert systems. MSc thesis, University of Copenhagen.

Pearl, J. (1986) Fusion, propagation and structuring in belief networks. *Artificial Intelligence*, **29**, 241–88.

Pearl, J. (1988) *Probabilistic Reasoning in Intelligence Systems*. San Mateo, California: Morgan Kaufmann.

Seroussi, B. and Golmard, J. L. (1994) An algorithm for finding the *K* most probable configurations in Bayesian networks. *International Journal of Approximate Reasoning*, **1**, 205–33.

Sy, B. K. (1993) A recurrence local computation approach towards ordering composite beliefs in Bayesian belief networks. *International Journal of Approximate Reasoning*, **8**, 17–50.

Xu, H. (1994) Computing marginals from the marginal representation in Markov Trees. Proceedings from the International Conference on Information Processing and Management of Uncertainty in Knowledge-based systems (IPMU), Paris, France, pp. 275–80. Paris, Cite Internationale Universitaire.