International Journal of Approximate Reasoning 161 (2023) 108984

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Computing the decomposable entropy of belief-function graphical models

Radim Jiroušek[a,b], Václav Kratochvíl[a,b], Prakash P. Shenoy[c,*]

[a] Faculty of Management, University of Economics, Jindrichuv Hradec, Czechia
[b] Institute of Information Theory and Automation, Czech Academy of Sciences, Prague, Czechia
[c] School of Business, University of Kansas, Lawrence, KS 66045, USA

A R T I C L E I N F O

Article history:
Received 16 February 2023
Received in revised form 5 June 2023
Accepted 7 July 2023
Available online 13 July 2023

Keywords:
Dempster-Shafer theory of belief functions
Decomposable entropy
Belief-function directed graphical models
Belief-function undirected graphical models

A B S T R A C T

In 2018, Jiroušek and Shenoy proposed a definition of entropy for Dempster-Shafer (D-S) belief functions called *decomposable entropy* (d-entropy). This paper provides an algorithm for computing the d-entropy of directed graphical D-S belief function models. We illustrate the algorithm using Almond's *Captain's Problem* example. For belief function undirected graphical models, assuming that the set of belief functions in the model is non-informative, the belief functions are distinct. We illustrate this using Haenni-Lehmann's *Communication Network* problem. As the joint belief function for this model is quasi-consonant, it follows from a property of d-entropy that the d-entropy of this model is zero, and no algorithm is required. For a class of undirected graphical models, we provide an algorithm for computing the d-entropy of such models. Finally, the d-entropy coincides with Shannon's entropy for the probability mass function of a single random variable and for a large multi-dimensional probability distribution expressed as a directed acyclic graph model called a Bayesian network. We illustrate this using Lauritzen-Spiegelhalter's *Chest Clinic* example represented as a belief-function directed graphical model.

1. Introduction

Jiroušek and Shenoy [14] proposes a definition of entropy for Dempster-Shafer (D-S) belief functions called *decomposable entropy* (d-entropy). Jiroušek and Shenoy [16] describes some basic properties of d-entropy. One of the basic properties of this entropy is as follows. Suppose we construct a joint basic probability assignment (BPA) $m_{X,Y}$ for $\{X, Y\}$ from a BPA m_X for X, a conditional $m_{Y|X}$ for Y given X, and $m_{X,Y} = m_X \oplus m_{Y|X}$, where $\oplus$ is Dempster's combination rule. Then, the joint d-entropy of $m_{X,Y}$, denoted by $H(m_{X,Y})$, is equal to $H(m_X) + H(m_{Y|X})$, where $H(m_{Y|X})$ denotes the conditional d-entropy of $m_{Y|X}$. This decomposable property is analogous to the decomposable property of Shannon's entropy for joint probability mass functions that is the basis of its definition [31]. There are numerous other definitions of entropy for the D-S theory (see [15] for a review). Still, none satisfy the decomposable property, and therefore, the computation of these entropies for large graphical models may be intractable.

The decomposable property lets us compute the d-entropy of graphical belief function models. Graphical belief function models can be either directed or undirected. This article provides an algorithm for computing the d-entropy of belief-

* Corresponding author.
E-mail address: pshenoy@ku.edu (P.P. Shenoy).

https://doi.org/10.1016/j.ijar.2023.108984

function directed graphical models, and we illustrate it using an example called the *Captain's decision problem* [1]. This problem has eight variables, and the joint state space of the eight variables has 3,584 states.

Two belief functions are said to be *mutually non-informative* if the marginals of these two belief functions to the intersection of their domains are vacuous. A set of belief functions is said to be *non-informative* if every pair of belief functions from the set is mutually non-informative. This is illustrated using the *Communication Network* example [10]. This problem has thirty-one binary variables, a joint state space of 2^{31} states, and forty-eight mutually non-informative belief functions. Assuming the conditional independence conditions implied by the factorization of the joint BPA are valid, and the set of BPAs is non-informative, the BPAs in the model are distinct [34]. As the joint BPA of this model is quasi-consonant, its d-entropy is 0. For general undirected graph models, we provide an algorithm for converting a class of such models to a directed graph model. We can use the algorithm for directed graph models to compute its d-entropy. A small example with six variables illustrates the algorithm.

Finally, the d-entropy generalizes Shannon's entropy for the probability of a single random variable and for large distributions expressed as directed acyclic graph models called Bayesian networks. We illustrate this using the *Chest Clinic* Bayesian network example [22]. First, we convert all probability potentials in the example to belief functions. In particular, we use Smets' conditional embedding to convert the conditional probability tables (CPTs) to conditional belief functions. These conditional belief functions are not Bayesian. Next, we compute the d-entropy of the directed graphical belief function model and show that it is the same as Shannon's entropy of this probability model. This example has eight binary variables with a joint state space of $2^8 = 256$ states.

An outline of the remainder of the article is as follows. Section 2 sketches the basic definitions in the D-S theory and reviews conditional belief functions, conditional independence, and distinct belief functions. Section 3 reviews d-entropy's basic definitions and properties. This section also contains a new property of d-entropy for two non-informative belief functions. Section 4 describes directed graphical belief function models and an algorithm for computing the d-entropy of large directed graphical belief function models using local computation. Section 5 defines undirected graphical belief function models. For a class of decomposable undirected graphical models, we describe an algorithm for converting such models to a directed graphical model and then computing its entropy using the algorithm described in Section 4. Section 6 describes three graphical belief function models discussed earlier. Section 7 describes some implementation details and tools used to implement the algorithm. Finally, Section 8 provides a summary and states some unresolved issues for future research.

2. Dempster-Shafer's belief function theory

In this section, we sketch the basics of Dempster-Shafer's theory of belief functions [7,26].

2.1. Representations

There are several representations in the D-S theory of belief functions. Here we focus on basic probability assignments and commonality functions.

Notation Let $\mathcal{V}$ denote the set of all variables. For each $X \in \mathcal{V}$, let Ω_X denote its finite state space. Subsets of $\mathcal{V}$ will be denoted by r, s, t, etc. For $r \subseteq \mathcal{V}$, let $\Omega_r = \times_{X \in r} \Omega_X$ denote the state space of r. Let 2^{Ω_r} denote the set of all subsets of Ω_r. Thus, 2^{Ω_r} is the space for defining belief functions.

Basic probability assignment A basic probability assignment (BPA) m for r is a function $m : 2^{\Omega_r} \to [0, 1]$ such that:

$$m(\emptyset) = 0, \text{and} \tag{1}$$

$$\sum_{\emptyset \neq \mathsf{a} \in 2^{\Omega_r}} m(\mathsf{a}) = 1. \tag{2}$$

$\emptyset$ denotes the empty set. For $\mathsf{a} \in 2^{\Omega_r}$, $m(\mathsf{a})$ represents the probability that is assigned exactly to subset a. Thus, no probability is assigned to the empty subset (Eq. (1)), and the total probability assigned to all non-empty subsets is 1 (Eq. (2)).

We say r is the *domain* of m The non-empty subsets $\mathsf{a} \in 2^{\Omega_r}$ such that $m(\mathsf{a}) > 0$ are called *focal elements* of m. A BPA m with only one focal element (with mass 1) is called *deterministic*. A deterministic BPA with focal element Ω_r is called *vacuous*. The vacuous BPA for r is sometimes denoted by ι_r. We say m is *Bayesian* if its focal elements are singleton subsets. We say m is *consonant* if the focal elements of m are nested, i.e., if they can be ordered such that $\mathsf{a}_1 \subset \mathsf{a}_2 \subset \ldots \subset \mathsf{a}_m$, where $\{\mathsf{a}_1, \ldots, \mathsf{a}_m\}$ denotes the set of all focal elements of m. Deterministic BPAs are consonant. We say m is *quasi-consonant* if the intersection of all focal elements of m is non-empty. A consonant BPA is also quasi-consonant, but not vice-versa.

Commonality function The information in a BPA m for X can also be represented by a corresponding commonality function (CF) Q_m for r that is defined as follows:

$$Q_m(\mathsf{a}) = \sum_{\mathsf{b}\in 2^{\Omega_r}:\mathsf{b}\supseteq\mathsf{a}} m(\mathsf{b}), \tag{3}$$

for all $\mathsf{a} \in 2^{\Omega_r}$. $Q_m(\mathsf{a})$ represents the probability mass that could move to every state in a.

From Eqs. (1)–(3), it follows that $0 \le Q_m \le 1$. From Eqs. (1)–(3), it follows that $Q_m(\emptyset) = 1$. If m is a vacuous BPA for r, then $Q_m(\mathsf{a}) = 1$ for all $\mathsf{a} \in 2^{\Omega_r}$. CFs are non-increasing in the sense that if $\mathsf{a} \subseteq \mathsf{b}$, then $Q_m(\mathsf{a}) \ge Q_m(\mathsf{b})$. The CF Q_m has the same information as the corresponding BPA m. Given a CF Q_m, we can recover the corresponding BPA m as follows [26]:

$$m(\mathsf{a}) = \sum_{\mathsf{b}\in 2^{\Omega_r}:\mathsf{b}\supseteq\mathsf{a}} (-1)^{|\mathsf{b}\setminus\mathsf{a}|} Q_m(\mathsf{b}). \tag{4}$$

Thus, it follows that $Q : 2^{\Omega_r} \to [0, 1]$ is a *well-defined CF* for r iff

$$Q(\emptyset) = 1, \tag{5}$$

$$\sum_{\mathsf{b}\in 2^{\Omega_r}:\mathsf{b}\supseteq\mathsf{a}} (-1)^{|\mathsf{b}\setminus\mathsf{a}|} Q(\mathsf{b}) \ge 0, \quad \text{for all } \mathsf{a} \in 2^{\Omega_r}, \text{ and} \tag{6}$$

$$\sum_{\emptyset\ne\mathsf{a}\in 2^{\Omega_r}} (-1)^{|\mathsf{a}|+1} Q(\mathsf{a}) = 1. \tag{7}$$

The left-hand side (LHS) of Eq. (6) is $m_Q(\mathsf{a})$, the BPA corresponding to CF Q, and the LHS of Eq. (7) can be shown to be $\sum_{\emptyset\ne\mathsf{a}\in 2^{\Omega_X}} m_Q(\mathsf{a})$. Eq. (7) can be regarded as a normalization condition for a CF. Thus, if we have a function $Q : 2^{\Omega_r} \to [0, 1]$ that satisfies Eqs. (5) and (6), but not (7), then we can divide each of the values of the function for non-empty subsets in 2^{Ω_r} by $K = \sum_{\emptyset\ne\mathsf{a}\in 2^{\Omega_r}} (-1)^{|\mathsf{a}|+1} Q_m(\mathsf{a})$, and the resulting function will then qualify as a CF.

In some cases, we could have a CF that doesn't satisfy Eq. (6) but does satisfy Eqs. (5) and (7). We will call such CFs *quasi-CFs*. If we convert a quasi-CF to a BPA using Eq. (4), then such a BPA will have negative masses that add to 1. We will call such BPAs *quasi-BPAs*. Quasi-CFs have been studied in [18,21].

2.2. Marginalization and combination

In the D-S theory, we reduce the domain of a joint belief function using the marginalization operation. We combine distinct (or independent) belief functions using Dempster's combination rule [7].

Marginalization Marginalization in D-S theory is the summation of values of BPAs over the states of the variables being marginalized to determine their contribution to the marginal.

Projection of states means dropping extra coordinates; for example, if (x, y) is a state of $\{X, Y\}$, then the projection of (x, y) to X, denoted by $(x, y)^{\downarrow X}$, is simply x, which is a state of X.

The projection of subsets of states is achieved by projecting every state in the subset. Suppose $\mathsf{b} \in 2^{\Omega_{X,Y}}$. Then $\mathsf{b}^{\downarrow X} = \{x \in \Omega_X : (x, y) \in \mathsf{b}\}$. Notice that $\mathsf{b}^{\downarrow X} \in 2^{\Omega_X}$.

Suppose m is a BPA for s and $r \subseteq s$. Then, the marginal of m for r, denoted by $m^{\downarrow r}$, is a BPA for r such that for each $\mathsf{a} \in 2^{\Omega_r}$,

$$m^{\downarrow r}(\mathsf{a}) = \sum_{\mathsf{b}\in 2^{\Omega_s}:\mathsf{b}^{\downarrow r}=\mathsf{a}} m(\mathsf{b}). \tag{8}$$

It follows from Eq. (8), that if $m(\mathsf{b}) > 0$, then $m^{\downarrow r}(\mathsf{b}^{\downarrow r}) > 0$, for all $\mathsf{b} \in 2^{\Omega_s}$.

Marginalization can also be defined in terms of CFs. Suppose Q is a CF for s and $r \subseteq s$. Then, for all $\mathsf{a} \in 2^{\Omega_r}$,

$$Q^{\downarrow r}(\mathsf{a}) = \sum_{\mathsf{b}\in 2^{\Omega_s}:\mathsf{b}^{\downarrow r}=\mathsf{a}} (-1)^{(|\mathsf{b}|-|\mathsf{a}|)} Q(\mathsf{b}). \tag{9}$$

As in the case of a BPA, it can be shown that if $Q(\mathsf{b}) > 0$, then $Q^{\downarrow r}(\mathsf{b}^{\downarrow r}) > 0$.

Dempster's combination rule We will define Dempster's combination rule in terms of CFs. Suppose r_1 and r_2 are arbitrary sets of variables, and Q_1 and Q_2 are distinct CFs for r_1 and r_2, respectively. Then $Q_1 \oplus Q_2$ is a CF for $r = r_1 \cup r_2$ given by:

$$(Q_1 \oplus Q_2)(\mathsf{a}) = \begin{cases} 1 & \text{if } \mathsf{a} = \emptyset, \\ K^{-1} Q_1(\mathsf{a}^{\downarrow r_1})\, Q_2(\mathsf{a}^{\downarrow r_2}) & \text{otherwise,} \end{cases} \tag{10}$$

for all $\mathsf{a} \in 2^{\Omega_r}$, where K is a normalization constant given by:

$$K = \sum_{\emptyset \neq \mathsf{a} \in 2^{\Omega_r}} (-1)^{|\mathsf{a}|+1} Q_1(\mathsf{a}^{\downarrow r_1})\, Q_2(\mathsf{a}^{\downarrow r_2}). \tag{11}$$

The definition of Dempster's rule assumes that the normalization constant K in Eq. (11) is non-zero. $(1-K)$ can be interpreted as a measure of *conflict* in the two CFs. Thus, if $1-K=1$, i.e., $K=0$, it represents total conflict, and the two CFs cannot be combined. If $K=1$, i.e., $1-K=0$, we say Q_1 and Q_2 are *non-conflicting*.

In general, $Q \oplus Q \neq Q$. Thus, it should be emphasized that $Q_1 \oplus Q_2$ in Eq. (10) only makes sense if Q_1 and Q_2 are distinct (or independent). Essentially, Q_1 and Q_2 are distinct if and only if $Q_1 \oplus Q_2$ doesn't involve double counting of non-idempotent knowledge. See Section 2.5 for a discussion on what constitutes distinct belief functions.

Vacuous extension Suppose m is a BPA for r and $s \supseteq r$. The vacuous extension of m to s, denoted by $m^{\uparrow s}$, is a BPA for s such that $m^{\uparrow s} = m \oplus \iota_{s \setminus r}$, where $\iota_{s\setminus r}$ is the vacuous BPA for $s \setminus r$. The vacuous extension doesn't add knowledge; it only changes the domain. $m^{\uparrow s}$ has the same number of focal elements as m with the same probabilities. The focal elements of $m^{\uparrow s}$ are vacuous extensions of the focal elements of m, i.e., $\mathsf{a} \times \Omega_{s\setminus r}$, where a is a focal element of m. Notice that $(m^{\uparrow s})^{\downarrow r} = m$.

Local computation Suppose m_1 is a BPA for r_1, m_2 is a BPA for r_2, and let r denote $r_1 \cup r_2$. Suppose $X \in r_1$ and $X \notin r_2$. Then,

$$(m_1 \oplus m_2)^{\downarrow r \setminus \{X\}} = m_1^{\downarrow r_1 \setminus \{X\}} \oplus m_2 \tag{12}$$

This property is called *local computation* [35] and is the basis of computing marginals of joint belief functions expressed as graphical belief function models. Giang and Shenoy [9] describes an implementation of a local computation algorithm, called Belief Function Machine (BFM), in Matlab for computing marginals of graphical belief function models.

Mutually non-informative belief functions Suppose m_1 and m_2 are two distinct BPAs for r_1 and r_2, respectively. We say m_1 and m_2 are *mutually non-informative* if $m_1^{\downarrow r_1 \cap r_2}$ and $m_2^{\downarrow r_1 \cap r_2}$ are vacuous BPAs for $r_1 \cap r_2$. Notice that if m_1 and m_2 are mutually non-informative, then $(m_1 \oplus m_2)^{\downarrow r_1} = m_1$ and $(m_1 \oplus m_2)^{\downarrow r_2} = m_2$. This follows from the definition of non-informative belief functions and the local computation property in Eq. (12).

Intuitively, m_1 does not tell us anything about m_2 and vice-versa. If r_1 and r_2 are disjoint, then they are trivially mutually non-informative. The definition of mutually non-informative belief functions can be generalized to sets of BPAs. A set of BPAs is said to be *non-informative* if every pair of BPAs from the set is mutually non-informative. Of course, checking only those pairs with a non-empty intersection of their domains is sufficient.

2.3. Conditional independence

Shenoy [32] describes conditional independence relation in the framework of valuation-based systems using factorization semantics. Here, we describe it for the D-S theory of belief functions.

Definition 1 *(Conditional independence).* Suppose $\mathcal{V}$ denotes the set of all variables, and suppose r, s, and t are disjoint subsets of $\mathcal{V}$. Suppose m is a joint BPA for $\mathcal{V}$. We say r and s are conditionally independent given t with respect to BPA m, written as $r \perp\!\!\!\perp_m s \,|\, t$, if and only if $m^{\downarrow r\cup s\cup t} = m_{r\cup t} \oplus m_{s\cup t}$, where $m_{r\cup t}$ is a BPA for $r \cup t$, $m_{s\cup t}$ is a BPA for $s \cup t$, and $m_{r\cup t}$ and $m_{s\cup t}$ are distinct.

This definition generalizes the CI relation in probability theory [5]. There are other definitions of conditional independence in the D-S theory (e.g., [19,38,2,3]). Definition 1 is closest to the definition in [19]. The definitions in [38,2,3] are based on the notion of non-interactivity, which are not useful in describing CI in belief-function graphical models.

The definition of CI in Definition 1 satisfies the graphoid properties of probabilistic conditional independence [25].

2.4. Conditional belief functions

This subsection defines a conditional belief function similar to a conditional probability table in probability theory. Conditional belief functions were initially studied by Smets [36], who introduced the notion of conditional embedding.[1] They have been further explored in [27,1,42]. The content of this subsection is taken from [13].

Definition 2 *(Conditionals).* Suppose r and s are disjoint subsets of variables and suppose $r' \subseteq r$. Suppose $m_{s|r'}$ is a BPA for $r' \cup s$. We say $m_{s|r'}$ is a conditional BPA for s given r' if and only if

1. $(m_{s|r'})^{\downarrow r'}$ is a vacuous BPA for r', and
2. If BPA m_r for r and $m_{s|r'}$ are distinct, then $m_r \oplus m_{s|r'}$ is a BPA for $r \cup s$.

[1] The 'conditional embedding' terminology is due to [27].

We call s the *head* of the conditional, and r the *tail*.

Some comments regarding this definition.

1. In graphical models, the joint is constructed from the conditionals. We don't start with a joint. The definition of a conditional belief function in Definition 2 reflects this fact. Other definitions of conditional belief functions start from a joint and then factor the joint into a marginal and a conditional (see, e.g., [1]). These other definitions do not help in constructing graphical models. Our definition, however, is consistent with these other definitions for the joint that a graphical belief function model implicitly defines [13].
2. In the second condition of Definition 2, m_r and $m_{s|r'}$ are distinct if and only if $s \perp\!\!\!\perp_{(m_r \oplus m_{s|r'})} (r \setminus r') \mid r'$. This is explained further in Section 2.5.
3. In a directed graphical belief function model, we have a conditional associated with each variable X. The head of the conditional is X, and the tail consists of the parents of X. For variables with no parents, we have priors associated with such variables. For convenience, priors can be regarded as conditionals with empty tails. For such BPAs, the first condition in the definition is trivially true as the marginal for the empty set is a vacuous BPA $\iota_\emptyset$.

Consider a BPA m for X and $x \in \Omega_X$. Suppose that there is a BPA for Y that expresses our belief about Y if we know that variable $X = x$, and denote it by m_{Y_x}. m_{Y_x} does not include the context $(X = x)$ in which m_{Y_x} is valid. So, we embed m_{Y_x} for Y into a conditional BPA for Y given X (whose domain is $\{X, Y\}$), denoted by $m_{Y|x}$, so that the following two conditions hold:

1. $m_{Y|x}$ tells us nothing about X, i.e., $(m_{Y|x})^{\downarrow X}(\Omega_X) = 1$.
2. If we combine $m_{Y|x}$ with the deterministic BPA $m_{X=x}$ for X such that $m_{X=x}(\{x\}) = 1$ using Dempster's rule, and marginalize the result to Y we obtain m_{Y_x}, i.e., $(m_{Y|x} \oplus m_{X=x})^{\downarrow Y} = m_{Y_x}$.

One way to obtain such an embedding is suggested by Smets [36] (see also, [27,42,1]), called *conditional embedding*. It consists of taking each focal element $\mathsf{b} \in 2^{\Omega_Y}$ of m_{Y_x} and converting it to the corresponding focal element

$$(\{x\} \times \mathsf{b}) \cup ((\Omega_X \setminus \{x\}) \times \Omega_Y) \in 2^{\Omega_{X,Y}} \tag{13}$$

of $m_{Y|x}$ with the same mass. It is easy to confirm that this embedding method satisfies both conditions mentioned above. We will illustrate conditional embedding using a small 2-variable Bayesian network.

Example 1 *(Representing a CPT by a conditional BPA).* Consider a two-variable Bayesian network consisting of variables A and T with state spaces $\Omega_A = \{a, \bar{a}\}$ and $\Omega_T = \{t, \bar{t}\}$. The directed acyclic graph is $A \to T$. The prior for A is PMF P_A such that $P_A(a) = 0.01$, $P_A(\bar{a}) = 0.99$, and the conditional probability table (CPT) for T, denoted by $P_{T|A}$, is as follows: $P_{T|A}(a, t) = 0.05$, $P_{T|A}(a, \bar{t}) = 0.95$, $P_{T|A}(\bar{a}, t) = 0.01$, and $P_{T|A}(\bar{a}, \bar{t}) = 0.99$.

$P(A)$ can be represented by the Bayesian BPA m_A as follows: $m_A(\{a\}) = 0.01$, $m_A(\{\bar{a}\}) = 0.99$.

Consider the conditional probability distribution for T when $A = a$. We can represent this conditional probability distribution by a Bayesian BPA m_{T_a} for T as follows: $m_{T_a}(\{t\}) = 0.05$, and $m_{T_a}(\{\bar{t}\}) = 0.95$. m_{T_a} is not a conditional. That this probability distribution is valid only when $A = a$ is not included in the BPA m_{T_a}. So, we embed this BPA in a conditional BPA $m_{T|a}$ for $\{A, T\}$ using Eq. (13) as follows:

$$m_{T|a}(\{(a, t), (\bar{a}, t), (\bar{a}, \bar{t})\}) = 0.05,$$
$$m_{T|a}(\{(a, \bar{t}), (\bar{a}, t), (\bar{a}, \bar{t})\}) = 0.95.$$

$m_{T|a}$ is a conditional BPA for T given A as $(m_{T|a})^{\downarrow A}$ is the vacuous BPA for A. Notice that $m_{T|a}$ is not a Bayesian BPA for $\{A, T\}$.

Similarly, the conditional probability distribution for T when $A = \bar{a}$ is modeled by the conditional BPA $m_{T|\bar{a}}$ for T given A as follows:

$$m_{T|\bar{a}}(\{(a, t), (a, \bar{t}), (\bar{a}, t)\}) = 0.01,$$
$$m_{T|\bar{a}}(\{(a, t), (a, \bar{t}), (\bar{a}, \bar{t})\}) = 0.99.$$

BPAs $m_{T|a}$ and $m_{T|\bar{a}}$ are distinct (as the contexts are disjoint). If we combine $m_{T|a}$ and $m_{T|\bar{a}}$ using Dempster's combination rule, we obtain $m_{T|A}$ as follows:

$$m_{T|A}(\{(a, t), (\bar{a}, t)\}) = 0.05 \cdot 0.01 = 0.0005,$$
$$m_{T|A}(\{(a, t), (\bar{a}, \bar{t})\}) = 0.05 \cdot 0.99 = 0.0495,$$
$$m_{T|A}(\{(a, \bar{t}), (\bar{a}, t)\}) = 0.95 \cdot 0.01 = 0.0095,$$

$$m_{T|A}(\{(a,\bar{t}),(\bar{a},\bar{t})\}) = 0.95 \cdot 0.99 = 0.9405.$$

$m_{T|A}$ is a conditional BPA for T given A as $(m_{T|A})^{\downarrow A}$ is a vacuous BPA for A. It is the belief-function representation of CPT $P_{T|A}$. Notice that $m_{T|A}$ is not a Bayesian BPA for $\{A,T\}$.

Finally, if we combine distinct BPAs m_A for A and $m_{T|A}$ for $\{A,T\}$ using Dempster's combination rule, we obtain a Bayesian joint BPA for $\{A,T\}$ representing the Bayesian joint distribution $P_A \otimes P_{T|A}$. $\square$

Conditionals can also be described using CFs. Suppose we start with a CF Q_X for X and want a conditional CF $Q_{Y|X}$ for $\{X,Y\}$. The conditional CF $Q_{Y|X}$ may include only those (non-vacuous) conditional CF $Q_{Y|x}$ for $\{X,Y\}$ such that[2] $Q_X(\{x\}) > 0$. If only one such conditional exists, then $Q_{Y|X} = Q_{Y|x}$. If we have more than one, and these are distinct, then $Q_{Y|X}$ is obtained by Dempster's combination of all such conditionals:

$$Q_{Y|X} = \bigoplus_{x\in\Omega_X : Q_X(\{x\})>0} Q_{Y|x}. \tag{14}$$

Next, we combine CFs Q_X for X and $Q_{Y|X}$ for $\{X,Y\}$ using Dempster's rule to obtain the joint CF $Q_{X,Y}$ for $\{X,Y\}$, i.e., $Q_{X,Y} = Q_X \oplus Q_{Y|X}$. First, from constructing $Q_{X,Y}$, the normalization constant K in Dempster's combination of Q_X and $Q_{Y|X}$ equals one. It follows from the definition of Dempster's rule in Eq. (10) that for all $\mathsf{b} \in 2^{\Omega_{X,Y}}$,

$$Q_{X,Y}(\mathsf{b}) = Q_X(\mathsf{b}^{\downarrow X}) \cdot Q_{Y|X}(\mathsf{b}). \tag{15}$$

If $\mathsf{b} \in 2^{\Omega_{X,Y}}$ is such that $Q_X(\mathsf{b}^{\downarrow X}) > 0$, then it follows from Eq. (15) that for all $\mathsf{b} \in 2^{\Omega_{X,Y}} \setminus \emptyset$,

$$Q_{Y|X}(\mathsf{b}) = \begin{cases} \frac{Q_{X,Y}(\mathsf{b})}{Q_X(\mathsf{b}^{\downarrow X})} & \text{if } Q_X(\mathsf{b}^{\downarrow X}) > 0, \\ \text{undefined} & \text{if } Q_X(\mathsf{b}^{\downarrow X}) = 0. \end{cases} \tag{16}$$

Notice that Eq. (16) is only valid for those joint CFs $Q_{X,Y}$ for $\{X,Y\}$ that are constructed using Eq. (15). If we start with an arbitrary CF Q for $\{X,Y\}$, compute the marginal CF $Q^{\downarrow X}$ for X (using Eq. (9)), and then construct a function $Q_{Y|X}$ using Eq. (16) for those $\mathsf{a} \in 2^{\Omega_{X,Y}}$ such that $Q^{\downarrow X}(\mathsf{a}^{\downarrow X}) > 0$, then $Q_{Y|X}$ may fail to be a CF because the condition in Eq. (6) is violated. Intuitively, the division operation in Eq. (16) can be regarded as a *removal* operation, an inverse of the combination operator [32]. Eq. (16) can be written as:

$$Q_{Y|X} = Q_{X,Y} \ominus Q_X, \tag{17}$$

where $\ominus$ denotes the removal operator. Thus, the right-hand side (RHS) of Eqs. (16) and (17) can be interpreted as removing the knowledge of the marginal Q_X from the joint knowledge $Q_{X,Y}$. Unlike probability theory, we can only remove knowledge from a joint that is explicitly included in it (in the sense that $Q_{X,Y} = Q_X \oplus Q$, where Q is some BPA for $\{X,Y\}$ [11].

Suppose Q_1 is a CF for r_1 and Q_2 is a CF for r_2 such that Q_1 and Q_2 are mutually non-informative.[3] Then, Q_2 can be considered a conditional for $r_2 \setminus (r_1 \cap r_2)$ given $r_1 \cap r_2$ (because $Q_2^{r_1 \cap r_2}$ is vacuous). Similarly, we can consider Q_1 as a conditional for $r_1 \setminus (r_1 \cap r_2)$ given $r_1 \cap r_2$.

2.5. Distinct belief functions

This material in this subsection is taken from Shenoy [34]. Distinct belief functions are also called independent belief functions in the D-S literature.[4] Dempster's combination rule is only applicable to combining distinct BPAs. So, what are distinct BPAs? Dempster [6] provides a definition. Consider the multi-valued semantics of BPAs as shown in Fig. 1.

Suppose we have a probability mass function (PMF) $P(X_1)$ for X_1, a multivalued function $\Gamma_1 : \Omega_{X_1} \to 2^{s_1} \setminus \emptyset$ that defines the BPA m_1 for s_1. Similarly, suppose we have a probability mass function (PMF) $P(X_2)$ for X_2, a multivalued function $\Gamma_2 : \Omega_{X_2} \to 2^{s_2} \setminus \emptyset$ that defines the BPA m_2 for s_2. BPAs m_1 and m_2 are distinct if and only if X_1 and X_2 are independent random variables, i.e., $P(X_1, X_2) = P(X_1) \otimes P(X_2)$, where $\otimes$ is the probabilistic combination operator, point-wise multiplication followed by normalization.

Some comments about Dempster's definition.

1. In practice, not every belief function in a belief function model is associated with a multi-valued mapping. Thus, Dempster's definition cannot be used directly in practice.

---

2 We can define conditionals $Q_{X,Y}$ for $\{X,Y\}$ such that $Q_X(\{x\}) = 0$, but as we will see shortly, such conditionals will have no impact on the joint CF $Q_{X,Y}$.

3 It is argued in [34] that Q_1 and Q_2 are distinct because they are mutually non-informative.

4 The terminology of 'distinct' belief functions is due to Smets [37]. As independence is usually associated with random variables, we prefer the terminology of distinct belief functions.

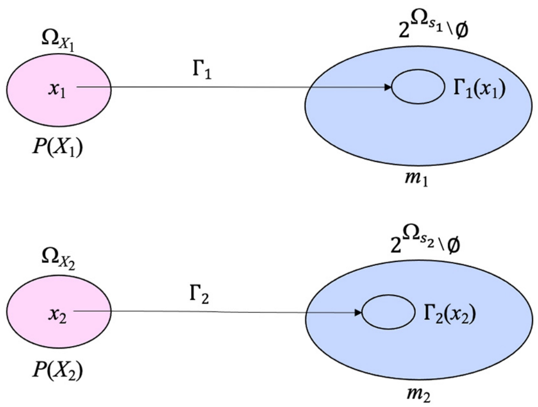


Fig. 1. Dempster's multi-valued semantics for BPAs.

2. We say BPA m is idempotent if $m \oplus m = m$. Idempotent knowledge is knowledge encoded in a BPA m that is idempotent. For example, if m is deterministic, then m is idempotent. Thus, double-counting idempotent knowledge is not a problem; double-counting non-idempotent knowledge is.
3. If we assume independence of random variables X_1 and X_2 when they are not, and we combine m_1 and m_2, then we are double-counting common knowledge encoded in m_1 and m_2. If the common knowledge encoded in these two BPAs is non-idempotent, then we have a problem. Thus, the spirit of Dempster's definition is that two belief functions are distinct if, when combining them using Dempster's combination rule, we are not double-counting non-idempotent knowledge.
4. BPA m_X for X and conditional BPA $m_{Y|X}$ for Y given X are always distinct (regardless of the numeric values of these BPAs). Notice that $(m_X \oplus m_{Y|X})^{\downarrow X} = m_X$, and $(m_X \oplus m_{Y|X}) \ominus m_X = m_{Y|X}$.
5. BPAs m_X for X and m_Y for Y are distinct if and only if $X \perp\!\!\!\perp_{m_X \oplus m_Y} Y$.
6. BPA $m_{X,Y}$ for $\{X, Y\}$ and conditional BPA $m_{Z|Y}$ for Z given Y are distinct if and only if $X \perp\!\!\!\perp_{m_{X,Y} \oplus m_{Z|Y}} Z \mid Y$.
7. The discussion of distinct belief functions is valid more broadly to many uncertainty calculi, including probability theory [28].
8. Some references to the literature on distinct belief functions are as follows: [28],[5] and [37].

3. Decomposable entropy of D-S belief functions

This section reviews the definitions of d-entropy and conditional d-entropy of belief functions in the D-S theory [14] and describes its properties [16]. We also describe a new property of d-entropy.

3.1. Decomposable entropy

Definition 3 *(d-entropy of a CF Q).* Suppose Q is a CF for r. Then, the d-entropy of Q, denoted by $H(Q)$, is defined as follows:

$$H(Q) = \sum_{\mathsf{a} \in 2^{\Omega_r} : Q(\mathsf{a}) > 0} (-1)^{|\mathsf{a}|} Q(\mathsf{a}) \log(Q(\mathsf{a})). \tag{18}$$

The definition of entropy of Q in Definition (3) is well-defined as the summation in the RHS of Eq. (18) is only for $Q(\mathsf{a}) > 0$.

In Subsection 2.4, we showed that the conditional commonality function, if it exists, can be expressed as $Q_{Y|X}(\mathsf{a}) = Q_{X,Y}(\mathsf{a})/Q_X(\mathsf{a}^{\downarrow X})$ (see Eq. (16)). This subsection will define the conditional entropy of a conditional CF. It would be incorrect to use Eq. (18) to compute the entropy of $Q_{Y|X}$ as our belief of X is not included in conditional CF $Q_{Y|X}$. The definition of conditional d-entropy of $Q_{Y|X}$ is analogous to the definition of Shannon's conditional entropy of a conditional probability distribution [31]. Here, we define the conditional d-entropy of a conditional CF $Q_{s|r}$ where r and s are disjoint subsets of variables.

Definition 4 *(Conditional d-entropy of $Q_{s|r}$).* Suppose r and s are disjoint subsets. Suppose Q_r is a CF for r, and suppose $Q_{s|r}$ is a conditional CF for s given r. Then, the conditional d-entropy of $Q_{s|r}$, denoted by $H(Q_{s|r})$, is defined as follows:

[5] This was published unchanged as [29].

$$H(Q_{s|r}) = \sum_{\mathsf{a} \in 2^{\Omega_{r\cup s}}: Q_{s|r}(\mathsf{a})>0} (-1)^{|\mathsf{a}|}\, Q_r(\mathsf{a}^{\downarrow r})\, Q_{s|r}(\mathsf{a}) \log(Q_{s|r}(\mathsf{a})). \tag{19}$$

Notice that as $Q_r(\mathsf{a}^{\downarrow r})\, Q_{s|r}(\mathsf{a}) = Q_{r\cup s}(\mathsf{a})$ for all $\mathsf{a} \in 2^{\Omega_{r\cup s}}$, we can rewrite Eq. (19) as follows:

$$H(Q_{s|r}) = \sum_{\mathsf{a} \in 2^{\Omega_{r\cup s}}: Q_{s|r}(\mathsf{a})>0} (-1)^{|\mathsf{a}|}\, Q_{r\cup s}(\mathsf{a}) \log(Q_{s|r}(\mathsf{a})) \tag{20}$$

Some comments on Definitions 3 and 4 are as follows:

1. There is a slight disconnect between the notation used for conditional d-entropy and the definition of conditional d-entropy. We use the notation $H(Q_{s|r})$ for the d-entropy of the conditional CF $Q_{s|r}$. But the definition of $H(Q_{s|r})$ also includes CF Q_r for r. So, when we write $H(Q_{s|r})$, we mean *in the context of some joint CF $Q_{r\cup s}$ for $r \cup s$ whose marginal for r (the tail of the conditional) is Q_r*. We can have any marginal CF Q_r for r, including the vacuous CF, and this does impact the value of $H(Q_{s|r})$. At the risk of causing slight confusion, we have decided to keep the notation simple (instead of including Q_r or $Q_{r\cup s}$ in the notation).
2. If the tail of a conditional is $\emptyset$, then the definition of $H(Q_{r|\emptyset})$ reduces to the (unconditional) d-entropy of Q_r, $H(Q_r)$ defined in Eq. (18). We are assuming the $\emptyset$ has state space $\Omega_\emptyset = \{\blacklozenge\}$, and a BPA $m_\emptyset$ for $\emptyset$ is such that $m(\{\blacklozenge\}) = 1$. The BPA for the $\emptyset$ represents the constant 1. To be consistent with our notation, we assume that $\Omega_\emptyset \times \Omega_r = \Omega_r$.
3. The intuition behind the definition of d-entropy is as follows. Shannon [31] showed that the definition of entropy of PMFs is motivated by the compound distribution axiom—if $P(X,Y) = P(X) \otimes P(Y|X)$, then the entropy of joint PMF $P(X,Y)$ is the same as the entropy of $P(X)$ plus the conditional entropy of $P(Y|X)$ ($\otimes$ is the Bayesian combination rule consisting of point-wise multiplication of probability potentials). A commonality function is a generalization of a PMF where we also have values of non-singleton subsets. For commonality functions, Dempster's rule is point-wise multiplication of commonality functions (followed by normalization), which is a generalization of the Bayesian combination rule. As we will see in the next subsection, our definition of d-entropy in terms of commonality functions shares the compound distribution property of Shannon's entropy.

3.2. Properties of decomposable entropy

A list of relevant properties of the d-entropy is as follows. For formal proofs, see [16].

Property 1 *(Compound distributions). Suppose m_r is a BPA for r, and suppose $m_{s|r}$ is a conditional BPA for s given r. Let $m_{r\cup s} = m_r \oplus m_{s|r}$. Then,*

$$H(m_{r\cup s}) = H(m_r) + H(m_{s|r}). \tag{21}$$

This is the most important property that characterizes this entropy. It is why the entropy is called *decomposable*. The assumption that $m_{s|r}$ is a conditional is essential. It guarantees that (1) $m_{s|r}$ does not have any information about r, i.e., $(m_{s|r})^{\downarrow r\cap s}$ is vacuous, and (2) that m_r and $m_{s|r}$ are distinct BPAs.

Property 2 *(Quasi-consonant BPAs). Suppose m is a quasi-consonant BPA. Then $H(m) = 0$. As vacuous, deterministic, and consonant BPAs are quasi-consonant, their decomposable entropies are 0.*

Property 3 *(Vacuous extension). Suppose m is a BPA for r, and suppose $m^{\uparrow(r\cup s)}$ denotes the vacuous extension of m to $r \cup s$. Then,*

$$H(m^{\uparrow(r\cup s)}) = H(m).$$

Suppose P_X is a probability mass function (PMF) for X such that $P_X(x) > 0$ for all $x \in \Omega_X$, and $P_{Y|X}$ is a conditional probability table (CPT) for Y given X, i.e., $P_{Y|X}(x,y) = P_{Y|x}(y)$, where $P_{Y|x}$ is the conditional PMF for Y given $X = x$ for all $(x,y) \in \Omega_{X,Y}$. Let $P_{X,Y} = P_X \otimes P_{Y|X}$ ($\otimes$ denotes probabilistic combination, which is pointwise multiplication followed by normalization). Let m_X denote the Bayesian BPA corresponding to P_X, and let m_{Y_x} denote the Bayesian BPA for Y corresponding to the conditional PMF $P_{Y|x}$ for Y given $X = x$. Let $m_{Y|x}$ denote the conditional BPA for $\{X,Y\}$ obtained by Smets' conditional embedding of m_{Y_x}. Let $m_{Y|X}$ denote $\bigoplus_{x\in\Omega_X} m_{Y|x}$. Let $m_{X,Y}$ denote $m_X \oplus m_{Y|X}$. Notice that although m_{Y_x} is a Bayesian BPA, $m_{Y|x}$ and $m_{Y|X}$ are not Bayesian BPAs (see Example 1 in Section 2.4).

Property 4 *(Strong probability consistency). Consider the situation described in the preceding paragraph. Let $H_s(P_{X,Y})$ and $H_s(P_X)$ denote Shannon's entropy of PMFs $P_{X,Y}$ and P_X, respectively, and let $H_s(P_{Y|X})$ denote Shannon's conditional entropy of the CPT $P_{Y|X}$. Then, $m_{X,Y}$ is a Bayesian BPA for $\{X,Y\}$ corresponding to PMF $P_{X,Y}$ such that*

Table 1
m_1 and m_2 and their entropies.

$\mathsf{a} \in 2^{\Omega_{\{X,Y\}}}$	$m_1(\mathsf{a})$	$(-1)^{\lvert\mathsf{a}\rvert+1} Q_1(\mathsf{a})$	$h(Q_1(\mathsf{a}))$
$\{(x, y)\}$		0.7	0.3602
$\{(x, \bar{y})\}$		0.7	0.3602
$\{(\bar{x}, y)\}$		0.3	0.5211
$\{(\bar{x}, \bar{y})\}$		0.3	0.5211
$\{(x, y), (x, \bar{y})\}$	0.7	−0.7	−0.3602
$\{(\bar{x}, y), (\bar{x}, \bar{y})\}$	0.3	−0.3	−0.5211
$\sum$	1	1	0.8813

$\mathsf{a} \in 2^{\Omega_{\{Y,Z\}}}$	$m_2(\mathsf{a})$	$(-1)^{\lvert\mathsf{a}\rvert+1} Q_2(\mathsf{a})$	$h(Q_2(\mathsf{a}))$
$\{(y, z)\}$		0.6	0.4422
$\{(y, \bar{z}), \}$		0.4	0.5288
$\{(\bar{y}, z)\}$		0.6	0.4422
$\{(\bar{y}, \bar{z})\}$		0.4	0.5288
$\{(y, z), (\bar{y}, z)\}$	0.6	−0.6	−0.4422
$\{(y, \bar{z}), (\bar{y}, \bar{z})\}$	0.4	−0.4	−0.5288
$\sum$	1	1	0.9710

$$H(m_{X,Y}) = H_s(P_{X,Y}), \tag{22}$$

$$H(m_X) = H_s(P_X), \tag{23}$$

$$H(m_{Y|X}) = H_s(P_{Y|X}). \tag{24}$$

The following theorem generalizes Property 1. It is a new property not discussed in [16].

Theorem 1 *(Mutually non-informative). Suppose Q_1 and Q_2 are mutually non-informative CFs. Then,*

$$H(Q_1 \oplus Q_2) = H(Q_1) + H(Q_2) \tag{25}$$

Proof of Theorem 1. As $Q_2^{\downarrow r_1 \cap r_2}$ is a vacuous CF for $r_1 \cap r_2$, it follows that $(Q_1 \oplus Q_2)^{\downarrow r_1} = Q_1$. Similarly, as $Q_1^{\downarrow r_1 \cap r_2}$ is a vacuous CF for $r_1 \cap r_2$, it follows that $(Q_1 \oplus Q_2)^{\downarrow r_2} = Q_2$. Thus, Q_1 and Q_2 are non-conflicting. Let r denote $r_1 \cup r_2$.

$$\begin{aligned}
&H(Q_1 \oplus Q_2)\\
&= \sum_{\emptyset \neq \mathsf{a} \in 2^{\Omega_r}} (-1)^{|\mathsf{a}|} (Q_1 \oplus Q_2)(\mathsf{a}) \log((Q_1 \oplus Q_2)(\mathsf{a}))\\
&= \sum_{\emptyset \neq \mathsf{a} \in 2^{\Omega_r}} (-1)^{|\mathsf{a}|} Q_1(\mathsf{a}^{\downarrow r_1})\, Q_2(\mathsf{a}^{\downarrow r_2}) \log(Q_1(\mathsf{a}^{\downarrow r_1})\, Q_2(\mathsf{a}^{\downarrow r_2}))\\
&= \sum_{\emptyset \neq \mathsf{a} \in 2^{\Omega_r}} (-1)^{|\mathsf{a}|} Q_1(\mathsf{a}^{\downarrow r_1})\, Q_2(\mathsf{a}^{\downarrow r_2}) (\log(Q_1(\mathsf{a}^{\downarrow r_1}) + \log(Q_2(\mathsf{a}^{\downarrow r_2}))\\
&= \sum_{\emptyset \neq \mathsf{a} \in 2^{\Omega_r}} (-1)^{|\mathsf{a}|} Q_1(\mathsf{a}^{\downarrow r_1})\, Q_2(\mathsf{a}^{\downarrow r_2}) \log(Q_1(\mathsf{a}^{\downarrow r_1})) +\\
&\qquad \sum_{\emptyset \neq \mathsf{a} \in 2^{\Omega_r}} (-1)^{|\mathsf{a}|} Q_1(\mathsf{a}^{\downarrow r_1})\, Q_2(\mathsf{a}^{\downarrow r_2}) \log(Q_2(\mathsf{a}^{\downarrow r_2}))
\end{aligned} \tag{26}$$

The first term in the RHS of Eq. (26) can be simplified as follows:

$$\begin{aligned}
&= \sum_{\emptyset \neq \mathsf{b} \in 2^{\Omega_{r_1}}} (-1)^{|\mathsf{b}|}\, Q_1(\mathsf{b}) \log(Q_1(\mathsf{b})) \sum_{\emptyset \neq \mathsf{c} \in 2^{\Omega_{r_2}} : \mathsf{c}^{\downarrow r_1 \cap r_2} = \mathsf{b}^{\downarrow r_1 \cap r_2}} (-1)^{|\mathsf{a}|-|\mathsf{b}|} Q_2(\mathsf{c})\\
&= \sum_{\emptyset \neq \mathsf{b} \in 2^{\Omega_{r_1}}} (-1)^{|\mathsf{b}|}\, Q_1(\mathsf{b}) \log(Q_1(\mathsf{b}))\, Q_2^{\downarrow r_1 \cap r_2}(\mathsf{b}^{\downarrow r_1 \cap r_2})\\
&= H(Q_1) \quad \text{(because } Q_2^{\downarrow r_1 \cap r_2} \text{ is vacuous)}
\end{aligned}$$

Similarly, it can be shown that the second term in the RHS of Eq. (26) simplifies to $H(Q_2)$.

Example 2. To illustrate Theorem 1, consider three binary variables X, Y, Z with state spaces $\Omega_X = \{x, \bar{x}\}$, $\Omega_Y = \{y, \bar{y}\}$, and $\Omega_Z = \{z, \bar{z}\}$, respectively. Suppose m_1 and m_2 are BPAs for (X, Y) and (Y, Z), respectively, as shown in Table 1. Notice that $m_1^{\downarrow Y}$ and $m_2^{\downarrow Y}$ are vacuous BPAs for Y. In Tables 1 and 2, let $h(Q(\mathsf{a}))$ denote $(-1)^{|\mathsf{a}|} Q(\mathsf{a}) \log Q(\mathsf{a})$. Only non-empty subsets with non-zero commonality values are shown. Notice that

$$H(m_1 \oplus m_2) = 1.8522 = 0.8813 + 0.9710 = H(m_1) + H(m_2).$$

Table 2
$m = m_1 \oplus m_2$ and its entropy.

$\mathsf{a} \in 2^{\Omega_{\{X,Y,Z\}}}$	$m(\mathsf{a})$	$Q(\mathsf{a})$	$h(Q(\mathsf{a}))$
$\{(x,y,z)\}$		0.42	0.5256
$\{(x,y,\bar{z})\}$		0.28	0.5142
$\{(x,\bar{y},z)\}$		0.42	0.5256
$\{(x,\bar{y},\bar{z})\}$		0.28	0.5142
$\{(\bar{x},y,z)\}$		0.18	0.4453
$\{(\bar{x},y,\bar{z})\}$		0.12	0.3671
$\{(\bar{x},\bar{y},z)\}$		0.18	0.4453
$\{(\bar{x},\bar{y},\bar{z})\}$		0.12	0.3671
$\{(x,y,z),(x,\bar{y},z)\}$	0.42	0.42	−0.5256
$\{(x,y,\bar{z}),(x,\bar{y},\bar{z})\}$	0.28	0.24	−0.5142
$\{(\bar{x},y,z),(\bar{x},\bar{y},z)\}$	0.18	0.18	−0.4453
$\{(\bar{x},y,\bar{z}),(\bar{x},\bar{y},\bar{z})\}$	0.12	0.12	−0.3671
$\sum$	1	1	1.8522

4. Belief-function directed graphical models

This section describes a belief-function directed graphical model and an algorithm for computing its d-entropy.

4.1. Definition

Graphical models allow us to construct large models by specifying the joint using small factors. Pearl [24] and Lauritzen and Spiegelhalter [22] use directed graphical models to represent probabilistic graphical models. The factorization of the joint probability distribution implies conditional independence (CI) assumptions of the model. As the definition of CI in D-S theory is similar to the definition of CI in probability theory, the CI assumptions of a belief-function directed graphical model are similar to the CI assumptions of a probabilistic graphical model. Belief-function graphical models are described in, e.g., [30,35,1,10].

First, we introduce some notation. A directed graph G is a pair $(\mathcal{V},\mathcal{E})$, where $\mathcal{V} = \{X_1,\ldots,X_n\}$ denotes the set of *nodes* and $\mathcal{E}$ denotes the set of *directed edges* (X_i,X_j) between two distinct variables in $\mathcal{V}$. For any node X_i, let $Pa_G(X_i) = \{X_j \in \mathcal{V} : (X_j,X_i) \in \mathcal{E}\}$. A directed graph is said to be acyclic if and only if there exists a sequence of the nodes of the graph, say $(X_1,\ldots,X_n)$ such that if there is a directed edge $(X_i,X_j) \in \mathcal{E}$ then X_i must precede X_j in the sequence. Such a sequence is called a *topological* sequence.

Definition 5 *(BF directed graphical model).* Suppose we have a directed acyclic graph $G = (\mathcal{V},\mathcal{E})$ with n nodes. A belief-function directed graphical model (BFDGM) is a pair $(G,\{m_1,\ldots,m_n\})$ such that BPA m_i associated with node X_i is a conditional BPA for X_i given $Pa_G(X_i)$ for $i = 1,\ldots,n$. A fundamental assumption of a BFDGM is that $m_1,\ldots,m_n$ are all distinct, and the joint BPA m for $\mathcal{V}$ associated with the model is given by

$$m = \bigoplus_{i=1}^{n} m_i. \tag{27}$$

Some comments about this definition:

1. The definition of a belief-function directed graphical model closely follows the definition of a Bayesian network. The main differences are that we can have missing priors and conditionals (or partial information about these), and belief function analogs of conditional probability tables describe the conditionals. Bayesian inference cannot handle missing (or partial information about) priors/conditionals [24, Ch. 9, p. 415], whereas, in the belief-function case, we can omit the missing information or model partial information as a BPA [33].
2. Although we have defined a BFDGM where each node is a single variable, the definition can be generalized to cases where each node is a subset of variables and the nodes (subsets) are disjoint. What is important is that we have a conditional belief function at each node. We will encounter such models when we convert decomposable undirected graphical models to directed graphical models as discussed in Section 5.
3. The assumption that $m_1,\ldots,m_n$ are all distinct allows the Dempster's combination in Eq. (27).
4. Eq. (27) implies conditional independence relations in the model. It follows from Definition 2.3, starting with a topological sequence $(X_1,\ldots,X_n)$, $X_i \perp\!\!\!\perp_m \{X_1,\ldots,X_{i-1}\} \setminus Pa_G(X_i) \mid Pa_G(X_i)$ for $i = 2,\ldots,n$.
5. The assumption that $m_1,\ldots,m_n$ are all distinct means that the conditional independence relations implied by the model are all valid [34].

If n is large, it may be intractable to explicitly compute the joint BPA m (m is a BPA for $\mathcal{V}$). However, depending on the graphical structure G, it may be tractable to compute the d-entropy of m using the properties of d-entropy sketched in Section 3, especially the compound distributions property.

Notice that we must disregard observations/likelihoods if we have these for a variable that is different from priors or conditionals in a directed graphical belief function model. For example, suppose we have a directed acyclic graph $X \to Y$ with a BPA m_1 for X, a BPA m_2 for $\{X, Y\}$ that constitutes a conditional for Y given X, and a BPA m_3 for Y that represents some observation or likelihood for Y. It follows from the compound distributions property that $H(m_1 \oplus m_2) = H(m_1) + H(m_2)$. But, in general, $H(m_1 \oplus m_2 \oplus m_3) \neq H(m_1) + H(m_2) + H(m_3)$. For this reason, we need to disregard observations/likelihoods in computing the d-entropy of a directed graphical belief function model.

4.2. An algorithm to compute d-entropy of BFDGMs

We start with a topological sequence $(X_1, \ldots, X_n)$. As G is acyclic, such a sequence always exists, but it may not be unique.

Do $i = 1, \ldots, n$:

- If $Pa_G(X_i) = \emptyset$, then $H(m_i)$ is computed using Definition 3.
- If $Pa_G(X_i) \neq \emptyset$, then first we find the marginal $(\bigoplus_{j=1}^{i} m_j)^{\downarrow \{X_i\} \cup Pa_G(X_i)}$ using local computation [35]. Thus, the conditional m_i and the corresponding marginal $\bar{m}_i$ are defined for the same variables $r = \{X_i\} \cup Pa_G(X_i)$. Next, we find the conditional d-entropy of m_i, $H(m_i)$, using Eq. (20) as follows. Let $\bar{m}_i$ denote the computed marginal for $\{X_i\} \cup Pa_G(X_i)$ and let $r = \{X_i\} \cup Pa_G(X_i)$ denote the domain of this marginal. Using Eq. (20),

$$H(m_i) = \sum_{\mathsf{a} \in 2^{\Omega_r} : Q_{m_i}(\mathsf{a}) > 0} (-1)^{|\mathsf{a}|} Q_{\bar{m}_i}(\mathsf{a}) \log(Q_{m_i}(\mathsf{a})) \tag{28}$$

End Do;

The d-entropy of the joint belief function $H(\bigoplus_{i=1}^{n} m_i) = \sum_{i=1}^{n} H(m_i)$. This follows from the compound distribution property of d-entropy.

Although the algorithm is described for directed models with variables as nodes, it can be easily generalized to directed graphical models where the nodes are disjoint subsets of variables.

5. Belief-function undirected graphical models

This section describes a belief-function undirected graphical model and an algorithm for computing its d-entropy that only works for a certain class of such models.

5.1. Definition

In the probabilistic graphical model literature, Darroch et al. [4], Whittaker [40], and Lauritzen [20] describe undirected graphical models. As the definition of CI for belief functions in Definition 2.3 is similar to probabilistic CI based on factorization semantics, a belief-function undirected graph model is similar to a probabilistic graphical model where we replace probability potentials by BPAs on the same domains. As in the directed case, belief-function graphical models can accommodate missing or partial information. Belief function undirected graphical models are described, e.g., in [1,10].

First, we introduce some notation. Consider an undirected graph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{X_1, \ldots, X_n\}$ denotes the set of *nodes* and $\mathcal{E}$ denotes the set of (undirected) *edges* $\{X_i, X_j\}$ between two distinct variables in $\mathcal{V}$. Consider node X_i in G. The Markov boundary of X_i, denoted by $Ma_G(X_i)$, is as follows $Ma_G(X_i) = \{X_j \in \mathcal{V} : \{X_i, X_j\} \in \mathcal{E}\}$. The Markov boundary of X_i consists of other nodes directly connected to X_i. A *clique* in G is a maximal completely connected subset of nodes of G. Fig. 2 shows the cliques of two undirected graphs. Suppose we have k cliques in the graph.

Definition 6 *(BF undirected graphical model).* A *belief-function undirected graphical model* (BFUGM) is $\big(G = (\mathcal{V}, \mathcal{E}), \{m_1, \ldots, m_k\}\big)$, where G is an undirected graph with cliques $r_1, \ldots, r_k$, and for each $i = 1, \ldots, k$, m_i is a BPA for r_i. A fundamental assumption of the BFUGM is that the BPAs are all distinct. Thus, a belief-function undirected graphical model corresponds to the joint BPA m for $\mathcal{V}$ defined as follows:

$$m = \bigoplus_{i=1}^{k} m_i, \tag{29}$$

assuming that m as defined in Eq. (29) is a well-defined BPA, i.e., the normalization constant K in Dempster's combination (Eq. (11)) is non-zero.

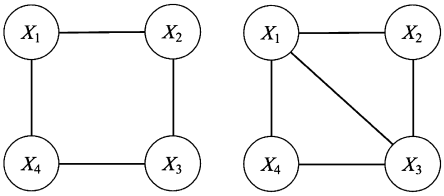


Fig. 2. The undirected graph on the left has four cliques $\{X_1, X_2\}$, $\{X_2, X_3\}$, $\{X_3, X_4\}$, and $\{X_1, X_4\}$, and the one on the right has two cliques $\{X_1, X_2, X_3\}$, and $\{X_1, X_3, X_4\}$.

Some comments regarding this definition.

1. The definition of a BFUGM closely follows the definition of a probabilistic undirected graphical model [20]. The main differences are that we can have missing/partial information about clique potentials modeled using BPAs.
2. The assumption that the BPAs are all distinct allows Dempster's combination in Eq. (29).
3. It follows from the definition of CI in Definition 1 and Eq. (29) that the following set of CI assumptions holds:

$$X_i \perp\!\!\!\perp_m (\mathcal{V} \setminus (Ma_G(X_i) \cup \{X_i\})) \mid Ma_G(X_i). \tag{30}$$

   In words, given the Markov boundary of X_i, X_i is conditionally independent of all other variables in the model [20].
4. The assumption that the BPAs in the model are all distinct includes the assumption that the conditional independence conditions encoded by the undirected graph G implied by Definition 1 are all valid. But, unlike directed graphical models with a conditional BPA associated with each node, more may be needed for BFUGMs. We have to ensure that there is no double-counting of non-idempotent knowledge when we use Dempster's combination in Eq. (29), which may depend on the details of the BPAs in the model. It is argued in [34] that if the set of BPAs is non-informative, and the CI assumptions of the model are all valid, then the BPAs are all distinct. In specific BFUGMs, we may have information that leads us to believe that the BPAs in a model are distinct.

5.2. Computing the d-entropy of decomposable BFUGMs

In section 4.2, we described an algorithm for computing the d-entropy of a large BFDGM. In this section, we describe a similar algorithm for decomposable BFUGMs. We start by defining decomposable undirected graphs.

Suppose G is an undirected graph with k cliques. We say G is *decomposable* if and only if there is an ordering of the cliques $(r_1, \ldots, r_k)$ of G that satisfies the *running intersection property* (RIP) [22] as follows:

$$\forall i = 2, \ldots, k \;\; \exists j \; (1 \le j \le i) \;\text{ such that }\; r_i \cap (r_1 \cup \ldots \cup r_{i-1}) \subseteq r_j. \tag{31}$$

If, for each i, there are several j's that satisfy the RIP property, we pick the smallest one and call r_j the *parent* clique of r_i. Notice that r_j always precedes r_i in the ordering $(r_1, \ldots, r_k)$. Also, we refer to the ordering satisfying the RIP as an *RIP* ordering. A BFUGM is said to be decomposable if its graph G is decomposable. A decomposable graph may have more than one RIP ordering. RIP ordering also applies to graphs with several connected components.

If the undirected graph is not decomposable, we make it decomposable by triangulating it using the *maximum cardinality search* algorithm [39]. Triangulating the graph involves adding edges. Thus, if $\big(G = (\mathcal{V}, \mathcal{E}), \{m_1, \ldots, m_k\}\big)$ denotes the original non-decomposable BFUGM, then let $\big(\bar{G} = (\mathcal{V}, \bar{\mathcal{E}}), \{\bar{m}_1, \ldots, \bar{m}_\ell\}\big)$ denote the decomposable BFUGM such that

1. $\bar{G}$ is decomposable;
2. $\mathcal{E} \subseteq \bar{\mathcal{E}}$;
3. $\bigoplus_{i=1}^{\ell} \bar{m}_i = \bigoplus_{i=1}^{k} m_i$, where $\ell \le k$.

Notice that item three says that the joint BPA associated with the decomposable model is the same as that associated with the non-decomposable model.

Example 3 *(Non-decomposable/decomposable BFUGMs).* Consider a BFUGM consisting of the undirected graph on the left in Fig. 2 and BPAs $m_{12}, m_{23}, m_{34}, m_{14}$ for cliques $\{X_1, X_2\}, \{X_2, X_3\}, \{X_3, X_4\}$, $\{X_1, X_4\}$, respectively. This BFUGM is not decomposable as there is no RIP ordering of the four cliques. However, it can be made decomposable by adding an edge $\{X_1, X_3\}$. This decomposable graph is shown on the right in Fig. 2. A RIP ordering of the two cliques in this model is (r_1, r_2), where $r_1 = \{X_1, X_2, X_3\}$, and $r_2 = \{X_1, X_3, X_4\}$. The BFUGM model associated with this graph includes BPAs $m_{12} \oplus m_{23}$ and $m_{34} \oplus m_{14}$ for cliques r_1 and r_2, respectively. Notice that the joint BPA is the same for the two BFUGM models. □

Let $(r_1, \ldots, r_\ell)$ denote an RIP ordering of the cliques of $\bar{G}$. In the next step, we convert the sequence of BPAs $(\bar{m}_1, \ldots, \bar{m}_\ell)$ into $(\bar{\bar{m}}_1, \ldots, \bar{\bar{m}}_\ell)$, so that for all $j = 1, \ldots, \ell$

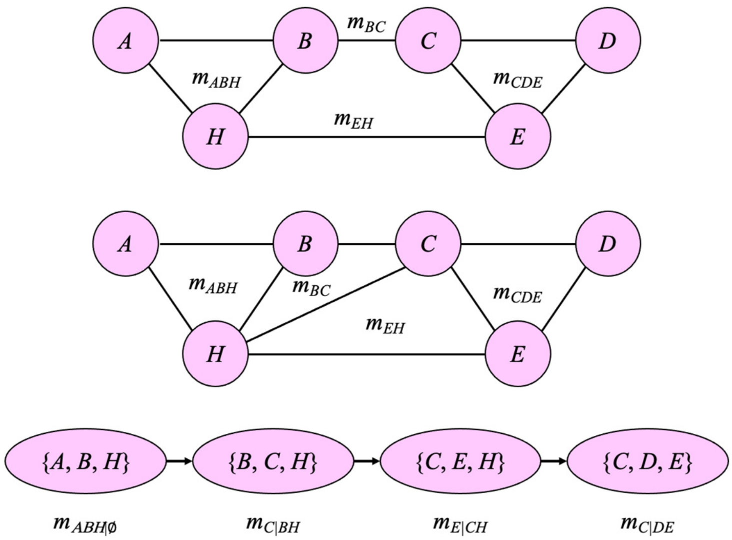


Fig. 3. Top: The original non-decomposable BFUGM. Middle: A decomposable BFUGM. Bottom: The BFDGM derived from the decomposable BFUGM.

$$\left(\bigoplus_{i=1}^{\ell} \bar{m}_i\right)^{\downarrow r_1 \cup \ldots \cup r_j} = \bigoplus_{i=1}^{j} \bar{\bar{m}}_i. \tag{32}$$

The conversion is realized in a cycle for $i = \ell, \ell-1, \ldots, 1$, do:

1. We compute BPA $\bar{m}_i \ominus (\bar{m}_i)^{\downarrow r_i \cap r_j}$, where r_j is the parent clique of r_i. Notice that if this BPA is well-defined (has non-negative values), then it is a conditional BPA for $r_i \setminus (r_i \cap r_j)$ given $r_i \cap r_j$ because $(\bar{m}_i \ominus (\bar{m}_i)^{\downarrow r_i \cap r_j})^{\downarrow r_i \cap r_j}$ is a vacuous BPA for $r_i \cap r_j$.
2. Let $\bar{\bar{m}}_i = \bar{m}_i \ominus (\bar{m}_i)^{\downarrow r_i \cap r_j}$.
3. The BPA $\bar{m}_j$ associated with parent clique r_j is replaced by $\bar{m}_j \oplus (\bar{m}_i)^{\downarrow r_i \cap r_j}$.

Notice that at the end of step 3, for each i in the do-loop, assuming the conditional BPA $\bar{m}_{i|j}$ is well-defined, the joint BPA associated with the undirected graph is unchanged. At any step in the do-loop, if $r_i \cap r_j = \emptyset$, then the BPAs associated with r_i and its parent clique r_j remain unchanged. At the end of the do-loop, we have a conditional associated with each clique of the decomposable graph. Thus, we have a BFDGM whose nodes are cliques of the original decomposable model and for which all the BPAs are conditionals. Moreover, notice that due to Eq. (32), $\bar{\bar{m}}_1$ is marginal of the joint model. We can use the Algorithm 4.2 to compute the d-entropy of the directed graphical model.

The class of BF undirected graphical models for which the algorithm described above works is where all the conditional BPAs computed in the first step of the do-loop are well-defined (not quasi-BPAs). Do we have a characterization of this class? Assuming all the conditionals computed in the algorithm are well-defined BPAs, at the end of the do-loop, we have a BFDGM where all the BPAs are well-defined conditionals. If we moralize this directed graphical model [22], we have a BFUGM. So this class of BFUGMs is non-empty. In [11], we have a characterization of when the removal of marginal results in a well-defined BPA. If the marginal being removed is explicitly included in the BPA from which the marginal is removed,[6] then removal will result in a well-defined BPA. Based on this characterization, we conjecture that the class of BF undirected graphical models for which the algorithm described here works are those whose joint BPAs are the same as the joints of the BFUGMs obtained from moralizing a BFDGM. The following example illustrates the algorithm described above.

Example 4 *(Converting a BFUGM to a BFDGM).* Consider six characteristics of a complex ecosystem: A, B, C, D, E, and H. Assume that biology describes the relationship among the characteristics A, B, and H, and it is encoded by the BPA m_{ABH} for $\{A, B, H\}$. Assume that chemistry explains the relationship among the remaining characteristics encoded by the BPA m_{CDE} for $\{C, D, E\}$. The thermodynamic laws describe the relationship between B and C encoded in the BPA m_{BC} for $\{B, C\}$, and economics explains the relationship between E and H, encoded as BPA m_{EH} for $\{E, H\}$. The diversity of the sources of knowledge justifies that the four BPAs are distinct. Thus, we have a BFUGM described by the undirected graph G in the top part of Fig. 3 with four cliques and the four associated BPAs. The joint BPA of the model is:

$$m = m_{ABH} \oplus m_{CDE} \oplus m_{BC} \oplus m_{EH} \tag{33}$$

6 Consider a joint BPA m for $\{X, Y\}$. We say that $m^{\downarrow X}$ is explicitly included in m if and only if $m = m^{\downarrow X} \oplus m'$, where m' is some BPA for $\{X, Y\}$.

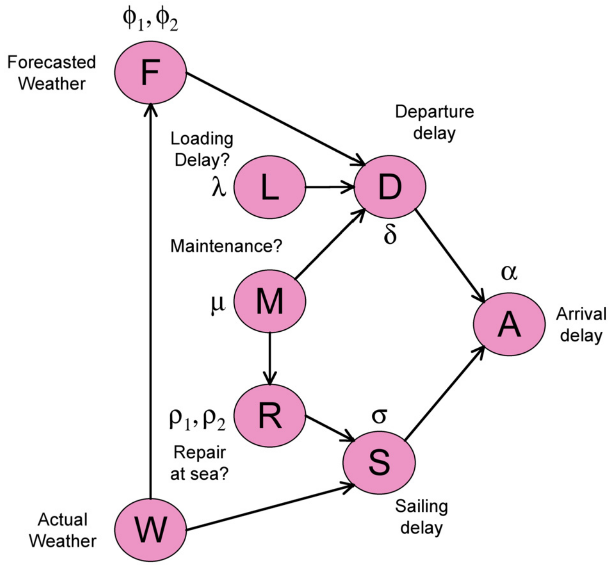


Fig. 4. The directed acyclic graph for the Captain's Problem. The Greek alphabets adjacent to a variable denote the prior or conditional evidence associated with the variable.

Notice that G is not decomposable.

Triangulating G means adding an edge $\{B, E\}$ or $\{C, H\}$. Suppose we choose the latter option. So, we now have four cliques: $r_1 = \{A, B, H\}$, $r_2 = \{B, C, H\}$, $r_3 = \{C, E, H\}$, and $r_4 = \{C, D, E\}$ as shown in the middle part of Fig. 3, with associated BPAs $m_1 = m_{ABH}$, $m_2 = m_{BC}$, $m_3 = m_{EH}$, and $m_4 = m_{CDE}$. This is a decomposable model as (r_1, r_2, r_3, r_4) is a RIP ordering, whose joint is the same as the original non-decomposable model.

Next, we compute the conditionals for each clique as follows. Using the cycle $i = 4, 3, 2, 1$ we get:

for $i = 4$**:** $m_4' = m_4 \ominus m_4^{\downarrow\{C,E\}} = m_{C|DE}$, and we replace m_3 by $m_3' = m_3 \oplus m_4^{\downarrow\{C,E\}}$.
for $i = 3$**:** $m_3'' = m_3' \ominus m_3'^{\downarrow\{C,H\}} = m_{E|CH}$, and we replace m_2 by $m_2' = m_2 \oplus m_3'^{\downarrow\{C,H\}}$.
for $i = 2$ $m_2'' = m_2' \ominus m_2'^{\downarrow\{B,H\}} = m_{C|BH}$, and we must replace m_1 by $m_1' = m_1 \oplus m_2'^{\downarrow\{B,H\}}$;
for $i = 1$ $m_1' = m_{ABH|\emptyset}$ is the prior associated with clique r_1.

Assuming the removal operation results in a well-defined BPA at each step, the joint BPA remains unchanged as we remove and combine the same BPA. Also, the final BPA associated with each clique is a conditional. In the end, we have conditionals for each clique in the BFDGM.

6. Three examples

In this section, we compute the d-entropy of three graphical belief function models.

6.1. Captain's problem

The Captain's Problem is from [1]. A ship's captain is concerned about how many days his ship may be delayed before arrival at a destination. The arrival delay is the sum of the departure delay and sailing delay. Departure delay may be a result of maintenance (at most one day), loading delay (at most one day), or a forecast of bad weather (at most one day). Sailing delays may result from bad weather (at most one day) and whether repairs are needed at sea (at most one day). If maintenance is done before sailing, chances of repairs at sea are less likely. The weather forecast says a slight chance of bad weather (0.2) and a good chance of good weather (0.6). The forecast is 80% reliable. The captain knows the loading delay and whether maintenance is done before departure. Fig. 4 shows the directed acyclic graph associated with this problem. What is the d-entropy of this belief function model?

Table 3 shows the variables and their states. The BPAs are as follows.

1. Weather forecast is 80% accurate. Notice that ϕ_1 is a conditional for F given W.

$$\phi_1(\{(g_w, g_f), (b_w, b_f)\}) = 0.8,$$
$$\phi_1(\Omega_W \times \Omega_F) = 0.2.$$

Table 3
The variables, their state spaces, and associated conditionals in the Captain's Problem.

Variable	Name	State Space, Ω	Assoc. Conditional
W	Actual weather	$\{g_w, b_w\}$	vacuous for W
F	Forecasted weather	$\{g_f, b_f\}$	ϕ_1 for $F\|W$
L	Loading delay?	$\{t_l, f_l\}$	λ for L
M	Maintenance done?	$\{t_m, f_m\}$	μ for M
R	Repair at sea needed?	$\{t_r, f_r\}$	$\rho_1' \oplus \rho_2'$ for $R\|M$
D	Dep. delay (in days)	$\{0, \ldots, 3\}$	δ for $D\|\{F, L, M\}$
S	Sailing delay (in days)	$\{0, \ldots, 3\}$	σ for $S\|\{W, R\}$
A	Arrival delay (in days)	$\{0, \ldots, 6\}$	α for $A\|\{D, S\}$

2. Forecast predicts bad weather with a chance of 0.2 and good weather with a chance of 0.6. Notice that ϕ_2 is an observation/likelihood for F, and it is not a conditional; therefore, it is not included in the computation of the d-entropy of the graphical model.

$$\begin{aligned}
\phi_2(\{b_f\}) &= 0.2,\\
\phi_2(\{g_f\}) &= 0.6,\\
\phi_2(\Omega_F) &= 0.2.
\end{aligned}$$

3. Loading is delayed with a chance of 0.3 and on schedule with a chance of 0.5. λ is a prior for L, which can be considered as a conditional for L given $\emptyset$.

$$\begin{aligned}
\lambda(\{t_l)\}) &= 0.3,\\
\lambda(\{f_l)\}) &= 0.5,\\
\lambda(\Omega_L) &= 0.2.
\end{aligned}$$

4. Maintenance is not done. μ is a prior for M, which can be considered as a conditional for M given $\emptyset$.

$$\mu(\{f_m\}) = 1.$$

5. If maintenance is done before sailing, the chances of repair at sea are between 10 and 30%. ρ_1 is a BPA for R (in the context $M = t_m$):

$$\begin{aligned}
\rho_1(\{t_r\}) &= 0.1,\\
\rho_1(\{f_r\}) &= 0.7,\\
\rho_1(\Omega_R) &= 0.2.
\end{aligned}$$

ρ_1 must be conditionally embedded to obtain a conditional for R given M. Let ρ_1' denote such a conditional.

$$\begin{aligned}
\rho_1'(\{(t_m, t_r), (f_m, t_r), (f_m, f_r)\}) &= 0.1,\\
\rho_1'((t_m, f_r), (f_m, t_r), (f_m, f_r)) &= 0.7,\\
\rho_1'(\Omega_{M,R}) &= 0.2.
\end{aligned}$$

6. If maintenance is not done before sailing, the chances of repair at sea are between 20 and 80%. ρ_2 is a BPA for R (in the context $M = f_m$):

$$\begin{aligned}
\rho_2(\{t_r\}) &= 0.2,\\
\rho_2(\{f_r\}) &= 0.2,\\
\rho_2(\Omega_R) &= 0.6.
\end{aligned}$$

ρ_2 must be conditionally embedded to obtain a conditional for R given M. Let ρ_2' denote such a conditional:

$$\begin{aligned}
\rho_2'(\{(t_m, t_r), (t_m, f_r), (f_m, t_r)\}) &= 0.2,\\
\rho_2'((t_m, t_r), (t_m, f_r), (f_m, f_r)) &= 0.2,\\
\rho_2'(\Omega_{M,R}) &= 0.6.
\end{aligned}$$

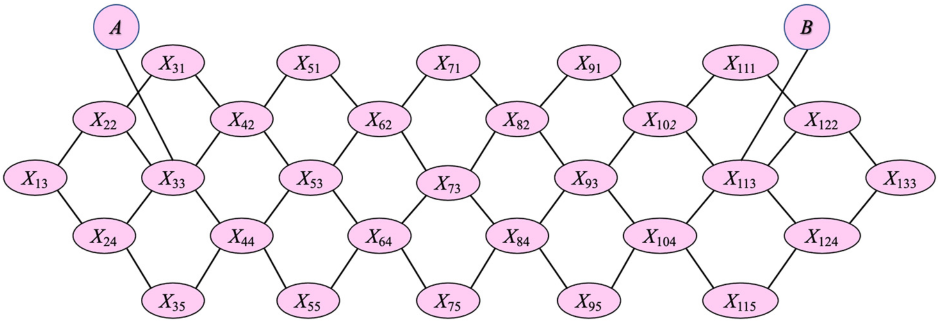


Fig. 5. An undirected graph for the Communication Network example.

7. Bad weather and repair at sea each add a day to sailing delay. This proposition is true 90% of the time. σ is a conditional for S given (W, R).

$$\sigma(\{(g_w, f_r, 0), (b_w, f_r, 1), (g_w, t_r, 1), (b_w, t_r, 2)\}) = 0.9,$$
$$\sigma(\Omega_W \times \Omega_R \times \Omega_S) = 0.1.$$

8. Departure delay may be a result of maintenance (at most one day), loading delay (at most one day), or a forecast of bad weather (at most one day). δ is a deterministic conditional for D given $\{F, L, M\}$.

$$\delta(\{(g_f, f_l, f_m, 0), (b_f, f_l, f_m, 1), (g_f, t_l, f_m, 1), (g_f, f_l, t_m, 1), (b_f, t_l, f_m, 2), (b_f, f_l, t_m, 2), (g_f, t_l, t_m, 2), (b_f, t_l, t_m, 3)\}) = 1.$$

9. The arrival delay is the sum of departure and sailing delays. α is a deterministic conditional for A given $\{D, S\}$.

$$\alpha(\{(0,0,0), (0,1,1), (0,2,2), (0,3,3), (1,0,1), (1,1,2), (1,2,3), (1,3,4), (2,0,2), (2,1,3), (2,2,4), (2,3,5), (3,0,3), (3,1,4), (3,2,5), (3,3,6)\}) = 1.$$

As ϕ_2 is observation/likelihood for F and not a conditional defining the joint BPA m, we ignore this belief function. A topological sequence for this directed graph is (W, F, L, M, D, R, S, A). First, notice that ϕ_1 and σ are consonant, and μ, δ, and α are deterministic. So the decomposable entropies of these five BPAs are zeroes. The decomposable entropies of the remaining BPAs are as follows (using the Algorithm 4.2):

$$H(\lambda) \approx 0.3958, \tag{34}$$
$$H(\rho_1' \oplus \rho_2') \approx 0.0729. \tag{35}$$

The d-entropy in Eq. (34) is computed using the prior BPA λ for L using Eq. (18). The conditional d-entropy in Eq. (35) is computed using $\rho_1' \oplus \rho_2'$ for {M, R} and the marginal of $\{M, R\}$ of the joint (shown below), and Eq. (19). Let m denote the joint BPA for all eight variables. Then

$$m^{\downarrow\{M,R\}}(\mathsf{a}) = \begin{cases} 0.6 & \text{if } \mathsf{a} = \{(f_m, t_r), (f_m, f_r)\}, \\ 0.2 & \text{if } \mathsf{a} = \{(f_m, t_r)\}, \\ 0.2 & \text{if } \mathsf{a} = \{(f_m, f_r)\}. \end{cases} \tag{36}$$

The marginal $m^{\downarrow\{M,R\}}$ was computed using Belief Function Machine Matlab software. Thus, the d-entropy of the Captain's Problem (ignoring the observation/likelihood ϕ_2) is $\approx 0.3958 + 0.0729 = 0.4687$.

6.2. Communication network

This example is from [10]. Fig. 5 shows an undirected graph associated with this example. We have a grid of $29 = 5 + 6 + 7 + 6 + 5$ communication nodes arranged in 13 columns and 5 rows. There are 46 links, as shown in Fig. 5, each has 90% reliability. Nodes A and B are connected to the grid with links having 80% reliability. What is the d-entropy of this graphical model?

We will not use the Algorithm from Sec. 5.2 as we can use a property of d-entropy to compute its d-entropy.

Consider the variables in the grid with 5 rows and 13 columns. Let X_{13} denote the first variable in row 3, and let X_{22} denote the first variable in row 2, etc. – see Fig. 5. Let $\Omega_{X_{13}} = \{t_{13}, f_{13}\}$, and let $\Omega_{X_{22}} = \{t_{22}, f_{22}\}$. The BPA m_{13-22} associated with the edge $\{X_{13}, X_{22}\}$ is as follows:

$$m_{13-22}(\{(t_{13}, t_{22}), (f_{13}, f_{22})\}) = 0.9,$$
$$m_{13-22}(\Omega_{\{X_{13}, X_{22}\}}) = 0.1.$$

The BPAs associated with the remaining 67 links are similar. The edges between A and X_{33} and between B and X_{113} are also similar, except that the reliability is 0.8 instead of 0.9.

Each edge in this network is a communication link between two nodes. The BPA associated with a link is a measure of the reliability of the link. Notice that the set of 48 BPAs in this model is non-informative, i.e., each pair of BPAs in the set are mutually non-informative. Thus, the BPAs in this model are distinct.

We will argue that the joint BPA of this model is quasi-consonant. Therefore, the d-entropy is 0.

Why is the joint BPA quasi-consonant? We will change the notation slightly to keep the exposition simple. Let X_1 denote X_{13}, let X_2 denote X_{22}, let X_3 denote X_{33} and let X_4 denote X_{24} (see Fig. 5). Let m_{12} denote the BPA m_{13-22}, let m_{23} denote m_{22-33}, let m_{34} denote m_{33-24}, and m_{14} denote m_{13-24}. Thus, m_{12} is a BPA for $\{X_1, X_2\}$, ..., m_{14} is a BPA for $\{X_1, X_4\}$. These four BPAs $\{m_{12}, m_{23}, m_{34}, m_{14}\}$ and four nodes $\{X_1, \ldots, X_4\}$ form a BF undecomposable undirected network model. Suppose $\Omega_{X_i} = \{t_i, f_i\}$ for $i = 1, \ldots, 4$. Then, BPA m_{12} has two focal sets: $\{(t_1, t_2), (f_1, f_2)\}$ and $\Omega_{\{X_1, X_2\}}$, $\ldots$, BPA m_{14} has focal sets $\{(t_1, t_4), (f_1, f_4)\}$, and $\Omega_{\{X_1, X_4\}}$. Thus, all four BPAs are consonant.

First, consider $m_{12} \oplus m_{23}$. The domain of this BPA is $\{X_1, X_2, X_3\}$, one more than the domain of m_{12}. This BPA has four focal elements as follows:

1. $e_1 = \{(t_1, t_2, t_3), (f_1, f_2, f_3)\}$,
2. $e_2 = \{(t_1, t_2, t_3), (f_1, f_2, f_3), (t_1, t_2, f_3), (f_1, f_2, t_3)\}$,
3. $e_3 = \{(t_1, t_2, t_3), (f_1, f_2, f_3), (f_1, t_2, t_3), (t_1, f_2, f_3)\}$, and
4. $e_4 = \Omega_{\{X_1, X_2, X_3\}}$

$e_1 = (\{(t_1, t_2), (f_1, f_2)\} \times \Omega_{X_3}) \cap (\Omega_1 \times \{(t_2, t_3), (f_2, f_3)\})$. $e_2 = (\{(t_1, t_2), (f_1, f_2)\} \times \Omega_{X_3}) \cap (\Omega_{\{X_1, X_2, X_3\}})$. $e_3 = (\Omega_{\{X_1, X_2, X_3\}}) \cap (\Omega_1 \times \{(t_2, t_3), (f_2, f_3)\})$. $e_4 = \Omega_{\{X_1, X_2, X_3\}} \cap \Omega_{\{X_1, X_2, X_3\}}$. Notice that $m_{12} \oplus m_{23}$ is not consonant, but it is quasi-consonant as the intersection of all four focal elements is the first focal element e_1. Also, the normalization constant K in Dempster's combination of these two BPAs is 1.

Next, consider $(m_{12} \oplus m_{23}) \oplus m_{34}$. As in the previous case, the domain of this BPA is one more variable than the domain of $(m_{12} \oplus m_{23})$. This BPA has $4 \cdot 2 = 8$ focal elements as follows:

1. $e_1 = \{(t_1, t_2, t_3, t_4), (f_1, f_2, f_3, f_4)\}$,
2. $\{e_2, e_3\} = \{(t_1, t_2, t_3), (f_1, f_2, f_3), (t_1, t_2, f_3), (f_1, f_2, t_3)\} \times \{t_4, f_4\}$
3. $\{e_4, e_5\} = \{(t_1, t_2, t_3), (f_1, f_2, f_3), (t_1, t_2, f_3), (f_1, f_2, t_3)\} \times \{t_4, f_4)\}$,
4. $\{e_6, e_7\} = \{(t_1, t_2, t_3), (f_1, f_2, f_3), (f_1, t_2, t_3), (t_1, f_2, f_3)\} \times \{t_4, f_4)$, and
5. $e_8 = \Omega_{\{X_1, X_2, X_3, X_4\}}$

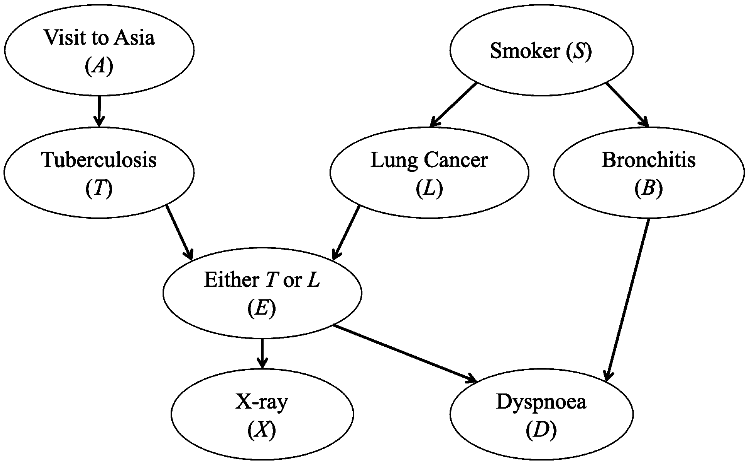
Notice that $(m_{12} \oplus m_{23}) \oplus m_{34}$ is quasi-consonant as the intersection of all eight focal sets is the first focal element e_1. Also, the normalization constant K in Dempster's combination of these two BPAs is 1.

Finally, consider $(m_{12} \oplus m_{23} \oplus m_{34}) \oplus m_{14}$. The domain of this BPA is the same as the domain of $(m_{12} \oplus m_{23} \oplus m_{34})$. This BPA has 15 focal elements. We will describe the focal elements instead of listing them. The first eight focal elements of $(m_{12} \oplus m_{23} \oplus m_{34}) \oplus m_{14}$ are the eight focal elements of $m_{12} \oplus m_{23} \oplus m_{34}$ intersected with $\{(t_1, t_4), (f_1, f_4)\} \times \Omega_{\{X_2, X_3\}}$. The next eight focal elements of $(m_{12} \oplus m_{23} \oplus m_{34}) \oplus m_{14}$ are the intersection of the eight focal elements of $m_{12} \oplus m_{23} \oplus m_{34}$ intersected with $\Omega_{\{X_1, X_2, X_3, X4\}}$. These eight focal elements are the same as the focal elements of $m_{12} \oplus m_{23} \oplus m_{34}$. The focal element e_1 appears twice, and its mass will be the sum of the two masses. Again, we have a quasi-consonant BPA with focal element e_1 included in all 15 focal elements of the joint BPA $m_{12} \oplus m_{23} \oplus m_{34} \oplus m_{14}$.

If we proceed sequentially by combining one adjacent BPA (whose domain intersects at one or two existing variables) at a time, there are two cases. We either extend the domain of the joint by one variable or not. In either case, the quasi-consonant property is retained. Given the nature of the BPAs (two focal elements, with one being $\Omega_{(X_i, X_j)}$, and one being $\{(t_i, t_j), (f_i, f_j)\}$ for link $\{X_i, X_j\}$), there is never any conflict. By induction, we argue that the focal element $\{(t_1, \ldots, t_{70}), (f_1, \ldots, f_{70})\}$ will be a subset of every focal element of the joint BPA of the entire model. Thus, the joint is quasi-consonant.

6.3. Chest clinic

The Chest Clinic example is from [22], a Bayesian network represented as a directed graphical belief function model, see Fig. 6. There are eight binary variables; some of the probabilities in the joint distribution are zeroes (as the CPT for E has many zeroes). The conditional probability tables (CPTs) in this example, also shown in Fig. 6, are represented as BPAs using Smets' conditional embedding, and these BPAs are not Bayesian (as demonstrated in Example 1 in Sec. 2.4). The priors for A and S are represented by Bayesian BPAs. It is proved in [33, Theorem 2] that the joint BPA of the belief function model corresponding to a Bayesian network is a Bayesian BPA corresponding to the joint PMF of the Bayesian network.



$P_A(a) = 0.01$, $P_A(\bar{a}) = 0.99$	$P_{E\mid L,T}(l, t, e) = 1$, $P_{E\mid L,T}(l, t, \bar{e}) = 0$ $P_{E\mid L,T}(l, \bar{t}, e) = 1$, $P_{E\mid L,T}(l, \bar{t}, \bar{e}) = 0$
$P_S(s) = 0.50$, $P_A(\bar{s}) = 0.50$	$P_{E\mid L,T}(\bar{l}, t, e) = 1$, $P_{E\mid L,T}(\bar{l}, t, \bar{e}) = 0$ $P_{E\mid L,T}(\bar{l}, \bar{t}, e) = 0$, $P_{E\mid L,T}(\bar{l}, \bar{t}, \bar{e}) = 1$
$P_{T\mid A}(a, t) = 0.05$, $P_{T\mid A}(a, \bar{t}) = 0.95$ $P_{T\mid A}(\bar{a}, t) = 0.01$, $P_{T\mid A}(\bar{a}, \bar{t}) = 0.99$	$P_{X\mid E}(e, x) = 0.98$, $P_{X\mid E}(e\ \bar{x}) = 0.02$ $P_{X\mid E}(\bar{e}, x) = 0.05$, $P_{X\mid E}(\bar{e}, \bar{x}) = 0.95$
$P_{L\mid S}(s, l) = 0.10$, $P_{L\mid S}(s, \bar{l}) = 0.90$ $P_{L\mid S}(\bar{s}, l) = 0.01$, $P_{L\mid S}(\bar{s}, \bar{l}) = 0.99$	$P_{D\mid E,B}(e, b, d) = 0.90$, $P_{D\mid E,B}(e, b, \bar{d}) = 0.10$ $P_{D\mid E,B}(e, \bar{b}, d) = 0.70$, $P_{D\mid E,B}(e, \bar{b}, \bar{d}) = 0.30$
$P_{B\mid S}(s, b) = 0.80$, $P_{B\mid S}(s, \bar{b}) = 0.20$ $P_{B\mid S}(\bar{s}, b) = 0.30$, $P_{B\mid S}(\bar{s}, \bar{b}) = 0.70$	$P_{D\mid E,B}(\bar{e}, b, d) = 0.80$, $P_{D\mid E,B}(\bar{e}, b, \bar{d}) = 0.20$ $P_{D\mid E,B}(\bar{e}, \bar{b}, d) = 0.10$, $P_{D\mid E,B}(\bar{e}, \bar{b}, \bar{d}) = 0.90$

Fig. 6. The directed acyclic graph and the CPTs for the Chest Clinic example.

The strong probability consistency (Property 4) of d-entropy states that the d-entropy of a 2-variable BN is the same as Shannon's entropy of the BN. This can be generalized to a BN of any size. The decomposable entropies of the conditionals are as follows (computed using the Algorithm 4.2 in Section 4):

$$H(m_A) \approx 0.0808, \tag{37}$$

$$H(m_{T|A}) \approx 0.0828, \tag{38}$$

$$H(m_S) = 1, \tag{39}$$

$$H(m_{L|S}) \approx 0.2749, \tag{40}$$

$$H(m_{B|S}) \approx 0.9261, \tag{41}$$

$$H(m_{E|L,T}) = 0, \tag{42}$$

$$H(m_{X|E}) \approx 0.2770, \tag{43}$$

$$H(m_{B|DE}) \approx 0.6471. \tag{44}$$

The d-entropy in Eqs. (37) and (39) were computed using Eq. (18) of the Bayesian priors for A and S, respectively. The d-entropy of the belief-function representation of the CPT for E in Eq. (42) is 0 as it is deterministic: Suppose ϵ is a BPA for $\{L, T, E\}$ as follows:

$$\epsilon(\{(l, t, e), (l, \bar{t}, e), (\bar{l}, t, e), (\bar{l}, \bar{t}, \bar{e})\}) = 1. \tag{45}$$

Notice that ϵ is a conditional BPA for E given $\{T, L\}$. If we conditionally embed the four Bayesian conditional distributions of E and combine these with Dempster's rule, we obtain BPA ϵ. Of course, we can write down ϵ as in Eq. (45) without going through this process. The d-entropies of the other conditionals were computed using Eq. (19). For example, to compute the d-entropy of the CPT of T, first, we compute the conditional BPA for T given A as shown in Example 1 in Sec. 2.4. Next, we compute the marginal of the joint for $\{A, T\}$, which is a Bayesian BPA for $\{A, T\}$. We skip the details of the computation of the remaining conditional d-entropies.

Thus, the d-entropy of the directed graphical belief function model is ≈ 3.2887, the same as Shannon's entropy of the corresponding Bayesian network.

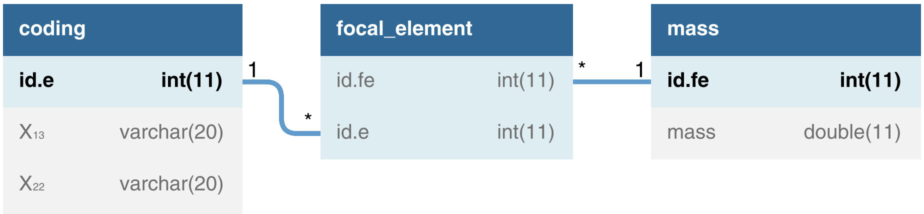


Fig. 7. Structure of database tables designed to store BPAs defined over two variables X_{13}, X_{22}.

Table 4
Representation of the BPA m_{13-22} for $\{X_{13}, X_{22}\}$ from Section 6.2.

id.e	X_{13}	X_{22}
1	t_{13}	t_{22}
2	t_{13}	f_{22}
3	f_{13}	t_{22}
4	f_{13}	f_{22}

(a) *coding*

id.fe	id.e
1	1
1	4
2	1
2	2
2	3
2	4

(b) focal_element

id.fe	mass
1	0.9
2	0.1

(c) mass

7. Notes on implementation

The power of any belief function software is determined by how the BPAs are represented in the computer's memory. Because of the superexponential growth of the number of subsets of the state space, we store only the focal elements of BPAs, i.e., the subsets of the frame of discernment with non-zero BPA values. A list of its elements defines a subset, and each element of the frame of discernment is essentially a record of random variable states. Thus, a focal element is a set of records assigned with one number – the corresponding BPA value. Computer scientists have developed a potent tool for records processing: a relational database. That is why we represent each BPA by a relational database satisfying a set of recommendations referred to as the third normal form (3NF) [17].

We did all experiments in *R*. We have created an *R* package,[7] which is based on relational databases as implemented in the *R* package *data.table* [8]. Each belief function is an object with three different relations (tables). Fig. 7 shows the relationship between the tables.

Consider a BPA for r. The first table, called *coding* in our implementation, contains all elements from the state space of r and assigns a unique identifier *id.e* to each of them. As we work with subset r, the columns correspond to individual variables, and rows to the elements of their joint state space $\times_{X \in r} \Omega_X$. For an example, see Table 4a, which contains all combinations of states of two binary variables X_{13} and X_{22} described in the Communication Network example in Section 6.2.

The second table, called *focal_element*, stores each focal element as a set of states using identifiers from the *coding* table. This table always has two columns. The first corresponds to the identifier of the given focal element *id.fe*, and the second refers to the identifier *id.e* from *coding* table. Thus, each focal element is defined by one or several rows in this table. As shown in Table 4b, the BPA m_{13-22} has two focal elements. The first one, $\{(t_{13}, t_{22}), (f_{13}, f_{22})\}$ is of cardinality two, while the second one contains the whole frame of discernment $\Omega_{\{X_{13}, X_{22}\}}$.

The third table, called *mass*, assigns a BPA value to each focal element. This table also has two columns. The first contains the focal element identifiers *id.fe* - as in the *focal_element* table. The second column defines the BPA value assigned to the corresponding focal element. Each row corresponds to one focal element. Thus, $m_{13-22}(\{(t_{13}, t_{22}), (f_{13}, f_{22})\}) = 0.9$, and $m_{13-22}(\Omega_{\{X_{13}, X_{22}\}}) = 0.1$.

Dempster's combination (in the case of BPAs) corresponds to the classic INNER JOIN operation of the focal element tables and the corresponding UPDATE of the mass table. The advantage of this approach is that, for example, the marginalization corresponds to the SELECT operation in the *coding table* and the corresponding aggregation in the focal element and mass tables. Therefore, the computational limitations of the package are not the number of random variables and their state spaces but the number of focal elements. SELECT, JOIN, UPDATE, and other SQL commands are standard terms from structured query language to write and query data in a database [41]. The corresponding software is very efficient since, for many years, SQL has been widely used as the language for database queries.

From an implementation point of view, this paper focuses on computing the d-entropy of the $\{m_i\}_{i=1}^n$ model. The word "decomposable" is essential. To calculate other entropies defined in the literature, it is usually necessary to combine elements of the model into a single BPA m - which the model represents. So, first, gradually apply Dempster's rule $\oplus$ and

[7] When completed, it will be published and available as an open source software.

compute $\bigoplus_{i=1}^{n} m_i$ into a single BPA. However, this is often not possible in practice. That is why the ability to calculate entropy using local computations directly from the model is crucial.

The implementation of the decomposable model entropy follows the algorithm presented in Section 4.2. However, upon closer examination, we find that the model's marginalization procedure fundamentally influences the calculation speed. Marginalization of the model employs the method of local calculations (12). Using this method, one can remove one variable at a time.

Let $\{m_i\}_{i\in I_X}$ be the set of BPAs with variable X in their domains and r as the respective domains union. Having variable X to eliminate (marginalize), we combine all BPAs from $\{m_i\}_{i\in I_X}$ and marginalize X from the combination $m_X = (\bigoplus_{i\in I_X} m_i)^{\downarrow r\setminus X}$. Then, we replace all $\{m_i\}_{i\in I_X}$ by m_X.

Since we usually have more variables to eliminate, we must decide on the elimination order, significantly affecting computing time. Unfortunately, the optimal ordering is not known. Our implementation corresponds to the one in the tool *Belief Function Machine* environment in Matlab [9] in that it uses the *one-step-look-ahead* heuristic for finding the sequence in which variables are eliminated. In this heuristic, at each step, we pick a variable that leads to a combination on the smallest domain r, and ties are broken arbitrarily.

We use the *R* package *igraph* [23] to work with graphs.

Calculating the d-entropy of a single BPA is challenging, as it requires the conversion of the BPA into the corresponding CF. Notice that usually $Q(\mathsf{a}) > 0$ for many $\mathsf{a} \in 2^{\Omega_s}$. It reveals weak points of our approach based on relational databases. In general, we cannot store the CF for more than five binary variables, which limits graphical models in terms of the maximum size of the parent set for each node. Fortunately, these sets are mostly limited. Similarly, to calculate the d-entropy, we do not need to store the entire CF; it is enough to save the intermediate results while generating it.

8. Summary & conclusions

The primary goal is to describe the computation of d-entropy of directed and undirected graphical belief function models. The d-entropy has a property that if we construct a joint BPA for two variables $\{X, Y\}$ by Dempster's combination of a BPA for m_X for X and a conditional BPA $m_{Y|X}$ for Y given X, then the d-entropy of $m_{X,Y} = m_X \oplus m_{Y|X}$ is equal to the d-entropy of m_X plus the conditional d-entropy of $m_{Y|X}$. This property is analogous to the decomposition property that is a basis for Shannon's entropy for probability distributions. We describe an algorithm for computing the d-entropy of any BFDGM. We apply this algorithm to Almond's Captain's problem [1] consisting of eight variables with a joint state space of 3,584 states.

Two BPAs m_1 for r_1 and m_2 for r_2 are said to be mutually non-informative if $m_1^{\downarrow r_1\cap r_2}$ and $m_2^{\downarrow r_1\cap r_2}$ are vacuous BPAs for $r_1 \cap r_2$. We show that the d-entropy of the Dempster combination of two such BPAs satisfies the property: $H(m_1 \oplus m_2) = H(m_1) + H(m_2)$. Suppose we have a BF undirected graphical belief function model whose belief functions are mutually non-informative, i.e., every pair of belief functions in the model is mutually non-informative. Then, BPAs in this BF undirected graphical model are all distinct. We describe an algorithm for computing a decomposable BFUGM to a BFDGM and then use the algorithm for computing the d-entropy of directed graphical models to compute its d-entropy. Haenni-Lehmann's Communication Network model [10] consists of 31 binary variables and 48 BPAs. The set of all BPAs is non-informative. Thus, the BPAs are distinct. The joint BPA of this BFUGM is quasi-consonant. Therefore, the d-entropy of this model is 0. No algorithm is needed for this problem.

The d-entropy also has a property that if we start with a Bayesian network (BN) probability model and encode all the priors and conditionals as belief functions, then the d-entropy of the corresponding belief function model is equal to Shannon's entropy of the Bayesian network model. We illustrate this using Lauritzen-Spiegelhalter's Chest Clinic example [22] consisting of eight binary variables.

The d-entropy is defined using commonality functions. If a graphical model has a clique whose state space is large, then computing the d-entropy of the clique may be intractable. For example, in the Captain's Problem, the conditional for arrival delay has three variables with a joint space of $4 \times 4 \times 7 = 112$ states. Fortunately, this conditional is deterministic, and the d-entropy of deterministic BPAs is 0. However, if this conditional wasn't deterministic or consonant or quasi-consonant, and the joint commonality function for these three variables had non-zero values for each of the 2^{112} subsets, the computation of the exact d-entropy of the conditional would be intractable. In such cases, we may have to resort to some approximate methods. This will be the focus of future work.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

This study was supported by the Czech Science Foundation Grant No. 21-07494S to the first two authors and by the Ronald G. Harper Professorship in the School of Business at the University of Kansas to the third author. A short version of this paper appeared in [12]. This paper has benefited from constructive comments from two anonymous reviewers of the *International Journal of Approximate Reasoning*.

References

[1] R.G. Almond, Graphical Belief Modeling, Chapman & Hall, London, UK, 1995.
[2] B. Ben-Yaghlane, P. Smets, K. Mellouli, Belief function independence: I. The marginal case, Int. J. Approx. Reason. 29 (2002) 47–70.
[3] B. Ben-Yaghlane, P. Smets, K. Mellouli, Belief function independence: II. The conditional case, Int. J. Approx. Reason. 31 (2002) 31–75.
[4] J.N. Darroch, S.L. Lauritzen, T.P. Speed, Markov fields and log-linear interaction models for contingency tables, Ann. Math. Stat. 8 (1980) 522–539.
[5] P. Dawid, Conditional independence in statistical theory, J. R. Stat. Soc. Ser. B 41 (1979) 1–15.
[6] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. Math. Stat. 38 (1967) 325–339.
[7] A.P. Dempster, A generalization of Bayesian inference, J. R. Stat. Soc. Ser. B 30 (1968) 205–247.
[8] M. Dowle, A. Srinivasan, data.table: extension of 'data.frame', https://CRAN.R-project.org/package=data.table, 2021, r package version 1.14.2.
[9] P. Giang, S. Shenoy, The belief function machine: An environment for reasoning with belief functions in Matlab, Lawrence, KS 66045, https://pshenoy.ku.edu/Papers/BFM072503.zip, 2003.
[10] R. Haenni, N. Lehmann, Resource bounded and anytime approximation of belief function computation, Int. J. Approx. Reason. 31 (2002) 103–154.
[11] R. Jiroušek, V. Kratochvíl, P.P. Shenoy, Entropy for evaluation of Dempster-Shafer belief function models, Int. J. Approx. Reason. 151 (2022) 164–181.
[12] R. Jiroušek, V. Kratochvíl, P.P. Shenoy, Computing the decomposable entropy of graphical belief function models, in: M. Studený, N. Ay, G. Coletti, G.D. Kleiter, P.P. Shenoy (Eds.), Proceedings of the 12th Workshop on Uncertain Processing (WUPES'22), MatfyzPress, Prague, Czechia, 2022, pp. 111–122, https://pshenoy.ku.edu/Papers/WUPES22a.pdf.
[13] R. Jiroušek, V. Kratochvíl, P.P. Shenoy, On conditional belief functions in directed graphical models in the Dempster-Shafer theory, Int. J. Approx. Reason. 160 (2023) 108976, https://pshenoy.ku.edu/Papers/IJAR23b.pdf.
[14] R. Jiroušek, P.P. Shenoy, A decomposable entropy for belief functions in the Dempster-Shafer theory, in: S. Destercke, T. Denoeux, F. Cuzzolin, A. Martin (Eds.), Belief Functions: Theory and Applications, in: Lecture Notes in Artificial Intelligence, vol. 11069, Springer Nature, Switzerland, 2018, pp. 146–154.
[15] R. Jiroušek, P.P. Shenoy, A new definition of entropy of belief functions in the Dempster-Shafer theory, Int. J. Approx. Reason. 92 (2018) 49–65.
[16] R. Jiroušek, P.P. Shenoy, On properties of a new decomposable entropy of Dempster-Shafer belief functions, Int. J. Approx. Reason. 119 (2020) 260–279.
[17] W. Kent, A simple guide to five normal forms in relational database theory, Commun. ACM 26 (1983) 120–125.
[18] C.T.A. Kong, Multivariate belief functions and graphical models, Ph.D. thesis, Department of Statistics, Harvard University, Cambridge, Massachusetts, 1986.
[19] C.T.A. Kong, A belief function generalization of Gibbs ensembles, Technical Report 239, University of Chicago, Department of Statistics, Chicago, IL, 1988.
[20] S.L. Lauritzen, Graphical Models, Oxford University Press, Oxford, UK, 1996.
[21] S.L. Lauritzen, F.V. Jensen, Local computation with valuations from a commutative semigroup, Ann. Math. Artif. Intell. 21 (1997) 51–69.
[22] S.L. Lauritzen, D. Spiegelhalter, Local computations with probabilities on graphical structures and their application to expert systems, J. R. Stat. Soc. Ser. B 50 (1988) 157–224.
[23] T. Nepusz, igraph: network analysis and visualization, https://CRAN.R-project.org/package=igraph, 2021, r package version 1.2.11.
[24] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Mateo, CA, 1988.
[25] J. Pearl, A. Paz, Graphoids: graph-based logic for reasoning about relevance relations, in: B.D. Boulay, D. Hogg, L. Steele (Eds.), Advances in Artificial Intelligence II, North-Holland, Amsterdam, 1987, pp. 357–363.
[26] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.
[27] G. Shafer, Belief functions and parametric models, J. R. Stat. Soc. Ser. B 44 (1982) 322–352.
[28] G. Shafer, The problem of dependent evidence, Working Paper 164, University of Kansas School of Business, Lawrence, KS 66045, 1984.
[29] G. Shafer, The problem of dependent evidence, Int. J. Approx. Reason. 79 (2016) 41–44.
[30] G. Shafer, P.P. Shenoy, K. Mellouli, Propagating belief functions in qualitative Markov trees, Int. J. Approx. Reason. 1 (1987) 349–400.
[31] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (1948) 379–423.
[32] P.P. Shenoy, Conditional independence in valuation-based systems, Int. J. Approx. Reason. 10 (1994) 203–234.
[33] P.P. Shenoy, Making inferences in incomplete Bayesian networks: A Dempster-Shafer belief-function approach, Int. J. Approx. Reason. 160 (2023) 108967, https://pshenoy.ku.edu/Papers/IJAR23a.pdf.
[34] P.P. Shenoy, On distinct belief functions in the Dempster-Shafer theory, in: E. Miranda, I. Montes, E. Quaeghebeur, B. Vantaggi (Eds.), Proceedings of the Thirteenth International Symposium on Imprecise Probability: Theories and Applications, PMLR 215, pp. 426–437, 2023, https://pshenoy.ku.edu/Papers/ISIPTA23a.pdf.
[35] P.P. Shenoy, G. Shafer, Axioms for probability and belief-function propagation, in: R.D. Shachter, T. Levitt, J.F. Lemmer, L.N. Kanal (Eds.), Uncertainty in Artificial Intelligence 4, North-Holland, Amsterdam, 1990, pp. 169–198.
[36] P. Smets, Un modele mathematico-statistique simulant le processus du diagnostic medical, Ph.D. thesis, Free University of Brussels, 1978.
[37] P. Smets, The concept of distinct evidence, in: B. Bouchon-Meunier (Ed.), Proceedings of the 4th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'92), 1992, pp. 789–794.
[38] P. Smets, Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem, Int. J. Approx. Reason. 9 (1993) 1–35.
[39] R. Tarjan, M. Yannakakis, Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs, SIAM J. Comput. 13 (1984) 566–579.
[40] J. Whittaker, Graphical Models in Applied Multivariate Statistics, Wiley, 1990.
[41] Wikipedia contributors, SQL — Wikipedia, the free encyclopedia [Online]. https://en.wikipedia.org/w/index.php?title=SQL&oldid=1110294038, 2022. (Accessed 15 September 2022).
[42] H. Xu, P. Smets, Reasoning in evidential networks with conditional belief functions, Int. J. Approx. Reason. 14 (1996) 155–185.