Proceedings of Machine Learning Research 215:259–269, 2023 ISIPTA 2023

On the Relationship between Graphical and Compositional Models for the Dempster-Shafer Theory of Belief Functions

Radim Jiroušek RADIM@UTIA.CAS.CZ
Václav Kratochvíl VELOREX@UTIA.CAS.CZ
Faculty of Management, Prague University of Economics and Business, Jindřichův Hradec, & Czech Academy of Sciences, Institute of Information Theory and Automation, Prague, Czechia

Prakash P. Shenoy PSHENOY@KU.EDU
University of Kansas School of Business, Lawrence, KS 66045 USA

Abstract

This paper studies the relationship between graphical and compositional models representing joint belief functions. In probability theory, the class of Bayesian networks (directed graphical models) is equivalent to compositional models. Such an equivalence does not hold for the Dempster-Shafer belief function theory. We show that each directed graphical belief function model can be represented as a compositional model, but the converse does not hold. As there are two composition operators for belief functions, there are two types of compositional models. In studying their relation to graphical models, they are closely connected. Namely, one is more specific than the other. A precise relationship between these two composition operators is described.

Keywords: joint belief functions, conditional independence, Markov models, composition operators, Dempster's combination rule, conditionals

1. Introduction

Bayesian networks are a popular way of representing joint probability distributions. Bayesian networks are equivalent to compositional models. Any joint probability distribution defined in a Bayesian network can also be expressed as a compositional model (with approximately the same number of parameters), and vice versa [3]. In this paper, we show that in the framework of Dempster-Shafer (D-S) belief function theory, such an equivalence does not hold. We show that a compositional model can represent any directed graphical belief function model (a belief function counterpart of a Bayesian network). The converse does not hold, i.e., not all compositional models can be represented as a directed graphical model.

There is extensive literature on graphical belief function models (see Almond [1]). Our first attempt [4] on this topic appeared to be a dead end; the proposal indiscriminately copied the ideas of undirected graphical models from probability theory. In this paper, we define both directed and undirected models. The former are belief function counterparts of Bayesian networks [2], and the latter corresponds to graphical models on undirected graphs (Markov networks [13]).

In the D-S belief function theory, compositional models were first introduced in [6], where the composition operator was defined. However, there is another way to define the composition operator, which better complies with the philosophy of the D-S theory [9]. The latter definition is based on Dempster's combination rule. Therefore, to distinguish it from the older version (called f-composition in this paper), we call it the d-composition operator.

Although two papers have been published on the relationship between these two composition operators (the first one presented at ISIPTA 2015 [5], the second one in [8]), the current study casts a new light also on this relation. This paper shows that one is more specific than the other and that the two composition operators often coincide for directed graphical models.

Section 2 introduces the notation and recalls the necessary basic notions of the D-S belief function theory. Section 3 describes conditional independence, which is crucial for graphical models. Section 4 describes the two composition operators and their main properties. The following two sections are devoted to the study of the relationship between compositional models and directed graphical models (Section 5) and undirected graphical models (Section 6).

2. Basics of D-S Belief Function Theory

Let $X, Y, Z, \ldots$ denote discrete (finite-valued) variables. Lower-case characters $r, s, t, \ldots$ denote the sets of variables. $\Omega_X, \Omega_Y, \ldots$ denote the state spaces of the corresponding variables. For a set of variables r, the corresponding state space is a Cartesian product $\Omega_r = \times_{X \in r} \Omega_X$.

A *basic probability assignment* (BPA) for r is a mapping $m : 2^{\Omega_r} \to [0, 1]$, such that $\sum_{\mathsf{a} \subseteq \Omega_r} m(\mathsf{a}) = 1$ and $m(\emptyset) = 0$. We will often call it a *joint* BPA to highlight that it is defined for a group of variables. We say that $\mathsf{a} \subseteq \Omega_r$ is a *focal element* of m if $m(\mathsf{a}) > 0$. A BPA with only one focal element is called *deterministic*. ι_r denotes the deterministic BPA for r, the focal element of which is the whole state space: $\iota_r(\Omega_r) = 1$. Since ι_r represents total ignorance, the corresponding BPA is said to be *vacuous*. BPA m is said to be Bayesian if all its focal elements are singletons ($m(\mathsf{a}) > 0 \Rightarrow |\mathsf{a}| = 1$).

A BPA m for r can also be specified by a corresponding *belief function* (BEL) or by *commonality function* (CF) [14]. Both these functions are also mappings $2^{\Omega_r} \to [0,1]$. They can be derived from BPA m as follows:

$$Bel_m(\mathsf{a}) = \sum_{\mathsf{b} \subseteq \Omega_r : \mathsf{b} \subseteq \mathsf{a}} m(\mathsf{b}), \quad Q_m(\mathsf{a}) = \sum_{\mathsf{b} \subseteq \Omega_r : \mathsf{b} \supseteq \mathsf{a}} m(\mathsf{b}).$$

All these representations are mutually equivalent; we can uniquely compute the others when one of these functions is given:

$$m(\mathsf{a}) = \sum_{\mathsf{b} \subseteq \mathsf{a}} (-1)^{|\mathsf{a} \setminus \mathsf{b}|} Bel_m(\mathsf{b}),$$
$$m(\mathsf{a}) = \sum_{\mathsf{b} \subseteq \Omega_r : \mathsf{b} \supseteq \mathsf{a}} (-1)^{|\mathsf{b} \setminus \mathsf{a}|} Q_m(\mathsf{b}). \tag{1}$$

Consider a BPA m for r, and suppose $s \subset r$. A marginal of m for s is denoted $m^{\downarrow s}$. A similar notation is used for *projections*. For $a \in \Omega_r$, $a^{\downarrow s}$ denote the element of Ω_s that is obtained from a by omitting the values of variables from $r \setminus s$. This notation is also used for the projections of subsets $\mathsf{b} \subseteq \Omega_r$: $\mathsf{b}^{\downarrow s} = \{a^{\downarrow s} : a \in \mathsf{b}\}$. The projection of sets enables us to define a *join* of two sets. Consider two arbitrary sets r and s of variables (they may be disjoint or overlapping, or one may be a subset of the other), and $\mathsf{a} \subseteq \Omega_r$, $\mathsf{b} \subseteq \Omega_s$. Their *join* is defined as:

$$\mathsf{a} \bowtie \mathsf{b} = \{c \in \Omega_{r \cup s} : c^{\downarrow r} \in \mathsf{a} \ \& \ c^{\downarrow s} \in \mathsf{b}\}.$$

Notice that if r and s are disjoint, then $\mathsf{a} \bowtie \mathsf{b} = \mathsf{a} \times \mathsf{b}$, if $r = s$, then $\mathsf{a} \bowtie \mathsf{b} = \mathsf{a} \cap \mathsf{b}$, and, in general, for $\mathsf{c} \subseteq \Omega_{r \cup s}$, c is a subset of $\mathsf{c}^{\downarrow r} \bowtie \mathsf{c}^{\downarrow s}$, which may be a proper one.

For BPA m for r and $s \subseteq r$, the marginal $m^{\downarrow s}$ is defined as follows:

$$m^{\downarrow s}(\mathsf{b}) = \sum_{\mathsf{a} \subseteq \Omega_r : \mathsf{a}^{\downarrow s} = \mathsf{b}} m(\mathsf{a}),$$

for all $\mathsf{b} \subseteq \Omega_s$.

An important operator of the D-S theory is *Dempster's combination rule*, which combines distinct belief functions. Consider two distinct BPAs m_1 and m_2, defined for r and s, respectively. Dempster's combination rule is defined for each $\mathsf{c} \subseteq \Omega_{r \cup s}$ as follows:

$$(m_1 \oplus m_2)(\mathsf{c}) = \frac{1}{K} \sum_{\mathsf{a} \subseteq \Omega_r,\, \mathsf{b} \subseteq \Omega_s : \mathsf{a} \bowtie \mathsf{b} = \mathsf{c}} m_1(\mathsf{a}) \cdot m_2(\mathsf{b}), \tag{2}$$

where the normalization constant

$$K = \sum_{\mathsf{a} \subseteq \Omega_r,\, \mathsf{b} \subseteq \Omega_s : \mathsf{a} \bowtie \mathsf{b} \neq \emptyset} m_1(\mathsf{a}) \cdot m_2(\mathsf{b}). \tag{3}$$

$(1-K)$ can be interpreted as the *amount of conflict* between m_1 and m_2. If $(1-K) = 1$, we say that BPAs m_1 and m_2 are in total conflict, and their Dempster's combination is undefined.

The assumption of distinct BPAs is essential. In general $m \oplus m \neq m$. Duble-counting of evidence by combining non-distinct BPAs may be misleading. In the following sections, we will study directed graphical belief function models consisting of priors and conditionals. Such BPAs are distinct if the respective conditional independencies hold in the models, so we can combine them using Dempster's combination rule.

It is known that Dempster's combination is commutative and associative [14]. Another important property of Dempster's combination rule relates to the marginalization of joint BPAs. This property is called *local computation* [18]. Suppose m_1 and m_2 be defined for r and s, respectively. If $s \subseteq t \subseteq r \cup s$, then

$$(m_1 \oplus m_2)^{\downarrow t} = m_1^{\downarrow t} \oplus m_2.$$

Thus, when we compute the marginal of $m_1 \oplus m_2$ for t by removing variables in $r \setminus t$ among which there is no variable from s, we can avoid combination on the state space of $r \cup s$ and do it instead on the smaller space of r.

Dempster's combination rule can also be described using the corresponding commonality functions. Consider two distinct BPAs m_1, m_2 defined for r, s, respectively, and the corresponding commonality functions Q_{m_1} and Q_{m_2}. Then, as showed in [14],

$$Q_{m_1 \oplus m_2}(\mathsf{c}) = \frac{1}{K}\, Q_{m_1}(\mathsf{c}^{\downarrow r})\, Q_{m_2}(\mathsf{c}^{\downarrow s}), \tag{4}$$

where K is the same normalization constant as that defined in Equation (3).

Equation (4) enables us to define the inverse of Dempster's combination rule called *removal* [16]. Since the combination is defined as the pointwise multiplication of CFs followed by normalization, the removal is defined as the pointwise division of CFs followed by normalization. Consider BPA m defined for variables $r \supset s$. Then, $Q_m \ominus Q_{m^{\downarrow s}}$, is defined as follows:

$$(Q_m \ominus Q_{m^{\downarrow s}})(\mathsf{a}) = L^{-1}\, Q_m(\mathsf{a}) / Q_{m^{\downarrow s}}(\mathsf{a}^{\downarrow s}), \tag{5}$$

for all nonempty $\mathsf{a} \subseteq \Omega_r$, where the normalization constant L equals

$$L = \sum_{\emptyset \neq \mathsf{a} \subseteq \Omega_V} (-1)^{|\mathsf{a}|+1} Q_m(\mathsf{a}) / Q_{m^{\downarrow s}}(\mathsf{a}^{\downarrow s}). \tag{6}$$

In [20], the removal operator is called the *decombination* operator. Notice that we define the removal only when we remove a marginal of Q_m from Q_m. Thus, if $Q_{m^{\downarrow s}}(\mathsf{a}^{\downarrow s}) = 0$, then also $Q_m(\mathsf{a}) = 0$. In this case, we define $0/0 = 0$. Though we define the removal operator for CFs, in what follows, we will also use it for BPAs. Thus, $m \ominus m^{\downarrow s}$ denotes the BPA corresponding to $Q_m \ominus Q_{m^{\downarrow s}}$. It means that $m \ominus m^{\downarrow s}$ can be computed from $Q_m \ominus Q_{m^{\downarrow s}}$ using Equation (1). In this context, it may happen that $m \ominus m^{\downarrow s}$ has negative masses – it is not a BPA (see Example 1 in the next section). For more details about the properties of the removal operator $\ominus$, see the following section and [16].

3. Conditional Independence

To cope with models with many variables, one must consider joint BPAs that can be represented with a limited number of parameters. Following the ideas employed in probability theory, it is achieved by constructing joint BPAs from smaller BPAs defined for a few variables. Such process introduces conditional independence relations among the variables. We will discuss two types of models representing joint BPAs using a collection of BPAs defined for a small number of variables. We will study graphical and compositional models. For both, the notion of conditional independence of variables is essential.

Definition 1 *Consider three disjoint sets of variables r, s, t, and a BPA m for $u \supseteq r \cup s \cup t$. Assume r and s are nonempty. We say r and s are conditionally independent given t, with respect to m, written as $r \perp\!\!\!\perp_m s | t$, if there exist BPAs m_1 for $r \cup t$ and m_2 for $s \cup t$ such that $m^{\downarrow r \cup s \cup t} = m_1 \oplus m_2$.*

In the above definition, if t is empty, r and s are said to be unconditionally independent, and the joint BPA $m^{\downarrow r \cup s}$ is equal to the combination of its marginals. Therefore, some authors call this independence *marginal independence*. The joint BPA can be reconstructed from its marginals even if $t \neq \emptyset$. For this, one has to use the d-composition operator derived from Dempster's combination rule in [9] defined as follows.

Definition 2 *Consider BPAs m_1 for r and m_2 for s. If $m_2 \ominus m_2^{\downarrow r \cap s}$ is a BPA, the* d-composition *$m_1 \triangleright_d m_2$ is a BPA*

$$m_1 \triangleright_d m_2 = m_1 \oplus (m_2 \ominus m_2^{\downarrow r \cap s}).$$

If $m_2 \ominus m_2^{\downarrow r \cap s}$ is not a BPA, then $m_1 \triangleright_d m_2$ is undefined.

Remark. In this paper, we exclude the possibility of composing BPAs for which $m_2 \ominus m_2^{\downarrow r \cap s}$ have negative values. We do not extend our consideration to the so-called *pseudo-BPAs*, even though, as shown in the following example, combining a BPA with a pseudo-BPA can sometimes yield a non-negative BPA. We keep this restriction in this paper because it is employed for proofs of some of the following assertions (Corollary 4, and Propositions 7, 8, 12).

Example 1 *Consider two simple BPAs m_1 and m_2. The former one defined for variables X, Y is deterministic with the focal element $m_1(\{(x\bar{y}\}) = 1$. The latter one, which is defined for variables Y, Z, has two focal elements $m_2(\{(\bar{y}\bar{z}), (\bar{y}z)\}) = 0.6$, and $m_2(\{(\bar{y}z), (y\bar{z}), (yz)\}) = 0.4$.*

Table 1: Computation of the conditional $m_2 \ominus m_2^{\downarrow Y}$ via the corresponding CFs

a	$Q_2(a)$	$Q_2^{\downarrow Y}(a^{\downarrow Y})$	$(Q_2/Q_2^{\downarrow Y})(a)$
$\{(\bar{y}\bar{z})\}$	0.60	1.00	0.60
$\{(\bar{y}z)\}$	1.00	1.00	1.00
$\{(y\bar{z})\}$	0.40	0.40	1.00
$\{(yz)\}$	0.40	0.40	1.00
$\{(\bar{y}\bar{z}), (\bar{y}z)\}$	0.60	1.00	0.60
$\{(\bar{y}z), (y\bar{z})\}$	0.40	0.40	1.00
$\{(\bar{y}z), (yz)\}$	0.40	0.40	1.00
$\{(y\bar{z}), (yz)\}$	0.40	0.40	1.00
$\{(\bar{y}z), (y\bar{z}), (yz)\}$	0.40	0.40	1.00

To compute the d-composition $m_1 \triangleright_d m_2$, we have to find the conditional $m_2 \ominus m_2^{\downarrow Y}$. To this end, we know no other way than to transform BPA m_2 and its marginal $m_2^{\downarrow Y}$ into the corresponding CFs Q_2 and $Q_2^{\downarrow Y}$. Their ratio $Q_2/Q_2^{\downarrow Y}$ (see Table 1) is a CF, which means that we apply Equation (5) with $L = 1$. Applying Eq, 1 to the last column of Table 1, we get the desired conditional $(m_2 \ominus m_2^{\downarrow Y})$

$$\begin{aligned}
(m_2 \ominus m_2^{\downarrow Y})(\{(\bar{y}z)\}) &= -0.6,\\
(m_2 \ominus m_2^{\downarrow Y})(\{(\bar{y}\bar{z}), (\bar{y}z)\}) &= 0.6,\\
(m_2 \ominus m_2^{\downarrow Y})(\{(\bar{y}z), (y\bar{z}), (yz)\}) &= 1.
\end{aligned} \tag{7}$$

Thus, we see that $(m_2 \ominus m_2^{\downarrow Y})$ is not a BPA; it achieves a negative value. Therefore, accepting Definition 2, the composition $m_1 \triangleright_d m_2$ is not defined, yet computing $m_1 \oplus (m_2 \ominus m_2^{\downarrow Y})$ yields a non-negative BPA with two focal elements:

$$\begin{aligned}
(m_1 \oplus (m_2 \ominus m_2^{\downarrow Y}))(\{(x\bar{y}z)\}) &= 0.4,\\
(m_1 \oplus (m_2 \ominus m_2^{\downarrow Y}))(\{(x\bar{y}\bar{z}), (x\bar{y}z)\}) &= 0.6.
\end{aligned}$$

From this example, we can also see that, even though $\left(m_2 \ominus m_2^{\downarrow Y}\right)^{\downarrow Y}$ is vacuous, $(m_2 \ominus m_2^{\downarrow Y})$ has focal elements, the projection of which to Y is not the full space which cannot happen when $(m_2 \ominus m_2^{\downarrow Y})$ is non-negative.

Even though d-composition is sometimes undefined, it is a valuable tool for constructing compositional models of joint

BPAs, for which a sufficiently rich system of conditional independence relations holds. The following assertion and its corollary theoretically support this fact.

Proposition 3 *Consider a BPA m for u, and three disjoint subsets r, s, t of u ($r \neq \emptyset \neq s$). Assume $m^{\downarrow s\cup t} \ominus m^{\downarrow t}$ is a BPA. Then, $r \perp\!\!\!\perp_m s|t$ if and only if $m^{\downarrow r\cup s\cup t} = m^{\downarrow r\cup t} \triangleright_d m^{\downarrow s\cup t}$.*

Proof Assume that for BPA m, r and s are conditionally independent given t, i.e., there exist valuations m_1 for $r \cup t$ and m_2 for $s \cup t$ such that $m^{\downarrow r\cup s\cup t} = m_1 \oplus m_2$. Using the local computation property, $m^{\downarrow r\cup t} = (m_1 \oplus m_2)^{\downarrow r\cup t} = m_1 \oplus m_2^{\downarrow t}$. Similarly, $m^{\downarrow s\cup t} = m_1^{\downarrow t} \oplus m_2$, and $m^{\downarrow t} = m_1^{\downarrow t} \oplus m_2^{\downarrow t}$. Therefore,

$$\begin{aligned} m^{\downarrow r\cup t} \triangleright_d m^{\downarrow s\cup t} &= m^{\downarrow r\cup t} \oplus m^{\downarrow s\cup t} \ominus m^{\downarrow t} \\ &= m_1 \oplus m_2^{\downarrow t} \oplus m_1^{\downarrow t} \oplus m_2 \ominus (m_1^{\downarrow t} \oplus m_2^{\downarrow t}) \\ &= m_1 \oplus m_2 = m^{\downarrow r\cup s\cup t}. \end{aligned}$$

The converse follows directly from Definition 2. ■

Corollary 4 *Consider two sets of variables r, s. Let m_1 and m_2 be defined for r and s, respectively. Then*

$$m_1 \oplus m_2 = (m_1 \oplus m_2)^{\downarrow r} \triangleright_d (m_1 \oplus m_2)^{\downarrow s}$$

if the composition is defined.

4. Composition Operators

As said already, the d-composition defined in Definition 2 is derived from Dempster's combination rule, and therefore, it is fully compatible with the D-S belief function theory. Nevertheless, [6] defines another composition operator as follows.

Definition 5 *Consider BPAs m_1 for r and m_2 for s. Their f-composition is a BPA $m_1 \triangleright_f m_2$ defined for each nonempty $\mathsf{c} \subseteq \Omega_{r\cup s}$ by one of the following expressions:*

(i) *if $m_2^{\downarrow r\cap s}(\mathsf{c}^{\downarrow r\cap s}) > 0$ and $\mathsf{c} = \mathsf{c}^{\downarrow r} \bowtie \mathsf{c}^{\downarrow s}$, then*

$$(m_1 \triangleright_f m_2)(\mathsf{c}) = \frac{m_1(\mathsf{c}^{\downarrow r}) \cdot m_2(\mathsf{c}^{\downarrow s})}{m_2^{\downarrow r\cap s}(\mathsf{c}^{\downarrow r\cap s})};$$

(ii) *if $m_2^{\downarrow r\cap s}(\mathsf{c}^{\downarrow r\cap s}) = 0$ and $\mathsf{c} = \mathsf{c}^{\downarrow r} \times \Omega_{s\setminus r}$, then $(m_1 \triangleright_f m_2)(\mathsf{c}) = m_1(\mathsf{c}^{\downarrow r})$;*

(iii) *in all other cases, $(m_1 \triangleright_f m_2)(\mathsf{c}) = 0$.*

Both composition operators introduced in Definitions 2 and 5 satisfy the properties expressed in the following statements (for proofs, see [6] and [9]).

Proposition 6 *For both composition operators (d-composition and f-composition) the following statements hold. Assume that BPAs m_r, m_s, and m_t are for r, s, and t, respectively, and that all the d-compositions are defined. Then,*

1. (Domain): *$m_r \triangleright m_s$ is a BPA for variables $r \cup s$.*
2. (Composition preserves first marginal): *$(m_r \triangleright m_s)^{\downarrow r} = m_r$.*
3. (Reduction:) *If $s \subseteq r$, then $m_r \triangleright m_s = m_r$.*
4. (Commutativity under consistency): *If m_r and m_s are consistent, i.e., $m_r^{\downarrow r\cap s} = m_s^{\downarrow r\cap s}$, then $m_r \triangleright m_s = m_s \triangleright m_r$.*
5. (Associativity under special condition): *If $r \supseteq (s \cap t)$, or, $s \supseteq (r \cap t)$ then, $(m_r \triangleright m_s) \triangleright m_t = m_r \triangleright (m_s \triangleright m_t)$.*
6. (Stepwise composition): *If $(r \cap s) \subseteq t \subseteq s$, then $(m_r \triangleright m_s^{\downarrow t}) \triangleright m_s = m_r \triangleright m_s$.*
7. (Exchangeability): *If $r \supseteq (s \cap t)$, then $(m_r \triangleright m_s) \triangleright m_t = (m_r \triangleright m_t) \triangleright m_s$.*
8. (Local computation): *If $(r \cap s) \subseteq t \subseteq (r \cup s)$, then $(m_r \triangleright m_s)^{\downarrow t} = m_r^{\downarrow r\cap t} \triangleright m_s^{\downarrow s\cap t}$.*

Before we study the role of the composition operators in representing joint BPAs in more detail, let us highlight the main differences between composition operators and Dempster's rule. The latter should be only applied to distinct belief functions representing independent pieces of evidence. On the other hand, the composition operator is typically used to compose two non-distinct marginals with a non-empty intersection, to assemble two pieces of evidence with some common knowledge. The composition operator is defined to avoid double counting of evidence from the two composed pieces of evidence. Thus, composition and Dempster's combination are designed for different purposes and possess different properties. While Dempster's rule is always commutative and associative, the composition operator meets these properties only in particular situations. On the other hand, Dempster's rule does not preserve the first marginal; it is not idempotent.

Consider BPAs m_1 for r and m_2 for s such that $m_2 \ominus m_2^{\downarrow r\cap s}$ is a BPA. In connection with Definition 2, we will identify situations when conditional BPA $m_2 \ominus m_2^{\downarrow r\cap s}$ is, in a way, "adapted" to BPA m_1. We say that $m_2 \ominus m_2^{\downarrow r\cap s}$ is *tight* with respect to m_1 if for all couples of focal elements a and b (a is a focal element of m_1, and b is a focal element of $m_2 \ominus m_2^{\downarrow r\cap s}$) the following condition holds:

$$\text{for } \forall\, b \in \mathsf{b},\ \exists\, a \in \mathsf{a},\ \text{such that } \{a\} \bowtie \{b\} \neq \emptyset. \quad (8)$$

In [7], we proved the following assertion.

Proposition 7 *Suppose BPAs m_1 for r, and m_2 for s are such that $m_2 \ominus m_2^{\downarrow r\cap s}$ is a BPA. Then, BPA $m_2 \ominus m_2^{\downarrow r\cap s}$ is* tight *with respect to m_1 if and only if*

$$m_1 \triangleright_f m_2 = m_1 \triangleright_d m_2.$$

Example 2 *In this example, we present a couple of BPAs that do not meet the assumptions of the previous assertion. Notice that $m_{Y,Z}$ from Table 2 is a conditional, because $m_{Y,Z}^{\downarrow Y}$ is vacuous, and thus $m_{Y,Z} \ominus (m_{Y,Z})^{\downarrow Y} = m_{Y,Z}$.*

Expression (ii) in Definition 5 applies to states for which the composed BPAs are, in a way, incompatible; the second argument does not bear the information on how to divide the mass assigned to a focal element of the first argument. Therefore, Expression (ii) assigns the respective value of a mass function to the least specific focal element. The acceptance of this idea makes the f-composition of any couple of BPAs possible. Notice that if the conditional of $m_{Y,Z}$ is tight with respect to $m_{X,Y}$, then Expression (ii) does not find its use.

Verify that the compositions $m_{X,Y} \triangleright_d m_{Y,Z}$ and $m_{X,Y} \triangleright_f m_{Y,Z}$ (see Table 2) differ only in the fact that the d-composition assigns mass 0.70 to $\{(\bar{x}\bar{y}z), (x\bar{y}z)\}$ and, in contrast, the f-composition assigns this mass to $\{(\bar{x}\bar{y}\bar{z}), (\bar{x}\bar{y}z), (x\bar{y}\bar{z}), (x\bar{y}z)\}$ by Expression (ii). Thus, the result of the f-composition is less specific than that of the d-composition. It is a general property, precisely formulated in Proposition 8 below. By the loss of specificity, we have to pay for the ability to combine any couple of BPAs. In other words, when we want to compose two BPAs whose d-composition is undefined, we can do it using f-composition, but we have to reconcile to a partial loss of information.

Table 2: Example illustrating Proposition 8

a	$m_{X,Y}(a)$
$\{(\bar{x}\bar{y}), (x\bar{y})\}$	0.70
$\{(\bar{x}\bar{y}), (\bar{x}y), (x\bar{y})\}$	0.30

a	$m_{Y,Z}(a)$
$\{(\bar{y}z), (y\bar{z})\}$	0.51
$\{(\bar{y}z), (yz)\}$	0.49

	$(m_{X,Y} \triangleright m_{Y,Z})(a)$	
a	$\triangleright_d$	$\triangleright_f$
$\{(\bar{x}\bar{y}z), (x\bar{y}z)\}$	0.70	
$\{(\bar{x}\bar{y}\bar{z}), (\bar{x}\bar{y}z), (x\bar{y}\bar{z}), (x\bar{y}z)\}$		0.70
$\{(\bar{x}\bar{y}z), (\bar{x}y\bar{z}), (x\bar{y}z)\}$	0.15	0.15
$\{(\bar{x}\bar{y}z), (\bar{x}yz), (x\bar{y}z)\}$	0.15	0.15

Proposition 8 *Suppose basic assignments m_1 for r and m_2 for s are such that $m_1 \triangleright_d m_2$ is defined. Then,*

$$Bel_{m_1 \triangleright_f m_2} \leq Bel_{m_1 \triangleright_d m_2}.$$

Proof The assumption that $m_1 \triangleright_d m_2$ is defined guarantees that $m_2 \ominus m_2^{\downarrow r\cap s}$ is a conditional, i.e., it is a BPA for which $\left(m_2 \ominus m_2^{\downarrow r\cap s}\right)^{\downarrow r\cap s}$ is vacuous. Therefore, for all its focal elements b, $\mathsf{b}^{\downarrow r\cap s} = \Omega_{r\cap s}$. This implies that for any couple of focal elements a and b of m_1 and $m_2 \ominus m_2^{\downarrow r\cap s}$, respectively, $\mathsf{a} \bowtie \mathsf{b} \neq \emptyset$. It implies that when computing (for any $\mathsf{c} \subseteq \Omega_{\{X\}\cup s}$)

$$\begin{aligned}(m_1 \triangleright_d m_2)(\mathsf{c}) &= (m_1 \oplus (m_2 \ominus m_2^{\downarrow r\cap s}))(\mathsf{c}) \\ &= \frac{1}{K} \sum_{\mathsf{a}\subseteq\Omega_r, \mathsf{b}\subseteq\Omega_s : \mathsf{a}\bowtie\mathsf{b}=\mathsf{c}} m_1(\mathsf{a}) \cdot (m_2 \ominus m_2^{\downarrow r\cap s})(\mathsf{b}), \quad (9)\end{aligned}$$

one gets

$$K = \sum_{\mathsf{a}\subseteq\Omega_r, \mathsf{b}\subseteq\Omega_s : \mathsf{a}\bowtie\mathsf{b}\neq\emptyset} m_1(\mathsf{a}) \cdot (m_2 \ominus m_2^{\downarrow r\cap s})(\mathsf{b}) = 1.$$

Now, consider a focal element a of m_1 and a focal element b of $m_2 \ominus m_2^{\downarrow r\cap s}$. The product $m_1(\mathsf{a}) \cdot (m_2 \ominus m_2^{\downarrow r\cap s})(\mathsf{b})$ plays its role in computations of both compositions. Considering d-composition first, this product contributes to the value assigned to focal element $\mathsf{a} \bowtie \mathsf{b}$. It means that either $(m_1 \triangleright_d m_2)(\mathsf{a} \bowtie \mathsf{b}) = m_1(\mathsf{a}) \cdot (m_2 \ominus m_2^{\downarrow r\cap s})(\mathsf{b})$, or, if there is yet another couple of focal elements a' and b' for which $\mathsf{a}' \bowtie \mathsf{b}' = \mathsf{a} \bowtie \mathsf{b}$, $(m_1 \triangleright_d m_2)(\mathsf{a} \bowtie \mathsf{b}) > m_1(\mathsf{a}) \cdot (m_2 \ominus m_2^{\downarrow r\cap s})(\mathsf{b})$. Considering the f-composition, notice that in Expression (i), the denominator equals 1, because the corresponding marginal is vacuous. Therefore, Expression (i) assigns value $(m_1 \triangleright_f m_2)(\mathsf{c}) = m_1(\mathsf{a}) \cdot (m_2 \ominus m_2^{\downarrow r\cap s})(\mathsf{b})$ for $\mathsf{c} \in \Omega_{r\cup s}$, for which $\mathsf{c}^{\downarrow r} = \mathsf{a}$ and $\mathsf{c}^{\downarrow s} = \mathsf{b}$. Since we know that $\mathsf{c} \subseteq \mathsf{c}^{\downarrow r} \bowtie \mathsf{c}^{\downarrow s}$, we see that the considered product contributes either to the same focal elements for both compositions, or to a larger focal element for f-composition than for d-composition. Trivially, the same property holds also when Expression (ii) applies in Definition 5. In this case, namely, $(m_1 \triangleright_f m_2)(\mathsf{a} \times \Omega_{s\setminus r}) = \sum_{\mathsf{d}} m_1(\mathsf{a}) \cdot (m_2 \ominus m_2^{\downarrow r\cap s})(\mathsf{d})$, where the summarization is realized for all focal elements d of $m_2 \ominus m_2^{\downarrow r\cap s}$.

Thus, we showed that focal elements of d-composition are smaller or equal to focal elements of f-composition. Precisely speaking we showed that for any $\mathsf{c} \subseteq \Omega_{r\cup s}$

$$\sum_{\mathsf{d}\subseteq\mathsf{c}} (m_1 \triangleright_d m_2)(\mathsf{d}) \geq \sum_{\mathsf{d}\subseteq\mathsf{c}} (m_1 \triangleright_f m_2)(\mathsf{d}),$$

■

5. Directed Graphical Models

Directed graphical models are a belief-function counterpart of Bayesian networks in probability theory.

Definition 9 *A belief function directed graphical model is a couple* $\left(G = (\{X_1, \ldots, X_n\}, \vec{E}), \{m_i\}_{i=1,\ldots,n}\right)$*, where G is an acyclic directed graph*[1]*, and* $\{m_i\}_{i=1,\ldots,n}$ *is a collection of conditional BPAs such that*

- *the nodes of G,* $\{X_1, \ldots, X_n\}$ *are discrete variables;*
- *for each node (variable)* X_i*, the corresponding conditional BPA* m_i *is defined for variable* X_i *and its parents; it is a conditional BPA for* X_i *given its parents* $Pa(X_i)$*, i.e.,* m_i *is defined* $\{X_i\} \cup Pa(X_i)$*, and* $m_i^{\downarrow Pa(X_i)}$ *is vacuous.*

Such a directed graphical model represents a joint BPA m for all variables: $m = \oplus_{i=1}^{n} m_i$.

Notice, that if $Pa(X_k) = \emptyset$, then the conditional for X_k is the prior belief function for X_k. If $Pa(X_k) \neq \emptyset$, then m_i is a conditional BPA for variable X_i given its parents. This requirement is fundamental. It guarantees that all BPAs m_i are mutually distinct and the expression $\oplus_{i=1}^{n} m_i$ is correctly used; there is no double counting of evidence.

There is extensive literature on conditional belief functions [15, 1, 21, 10]. In the following example, we employ the process of *conditional embedding* introduced in 1978 by Smets in [19], which is often useful for constructing directed graphical models.

Consider a variable X and its parents $Pa(X)$. Assume that if the parent variables are in a state $a \in \Omega_{Pa(X)}$, the behavior of variable X is described by BPA $m_{X|a}$ for X. We want to *embed* this BPA for X into a conditional BPA for $(\{X\} \cup Pa(X))$ denoted by $m_{a,Pa(X)}$, so that the following two conditions hold:

1. $m_{a,Pa(X)}$ tells us nothing about $Pa(X)$, i.e., $m_{a,Pa(X)}^{\downarrow Pa(X)}$ is vacuous.
2. Consider the deterministic BPA $m_{Pa(X)=a}$ (which is a BPA for variables $Pa(X)$ such that $m_{Pa(X)=a}(a) = 1$). If we combine $m_{a,Pa(X)}$ with $m_{Pa(X)=a}$ using Dempster's rule, and marginalize the result to X, we obtain $m_{X|a}$, i.e., $(m_{a,Pa(X)} \oplus m_{Pa(X)=a})^{\downarrow X} = m_{X|a}$.

The process called *Smets' conditional embedding* [19] (see also, [15], [21], and [1]) consists of taking each focal element $\mathsf{b} \subseteq \Omega_X$ of $m_{X|a}$ and converting it to the corresponding focal element $\mathsf{c} \subseteq \Omega_{\{X\} \cup Pa(X)}$ of $m_{a,Pa(X)}$, where

$$\mathsf{c} = (\{a\} \times \mathsf{b}) \cup ((\Omega_{Pa(X)} \setminus \{a\}) \times \Omega_X) \subseteq \Omega_{\{X\} \cup Pa(X)}, \tag{10}$$

and $m_{a,Pa(X)}(\mathsf{c}) = m_{X|a}(\mathsf{b})$. From this, one can immediately see that $m_{a,Pa(X)}$ has exactly the same number of focal elements as $m_{X|a}$, and that for each focal element c of $m_{a,Pa(X)}$, its projection $\mathsf{c}^{\downarrow Pa(X)} = \Omega_{Pa(X)}$.

This process is repeated for all $a \in \Omega_{Pa(X)}$ for which we know $m_{X|a}$, and eventually $m_{X|Pa(X)}$ is obtained as a Dempster's combination of all the constructed $m_{a,Pa(X)}$. In a more general form, the process can be realized not only for elements $a \in \Omega_{Pa(X)}$ but for a system of disjoint subsets of $\mathsf{a} \subseteq \Omega_{Pa(X)}$).

Example 3 (Changing tires) *Consider a graphical model with six variables* B, O, A, S, T, M*, the graph from Fig. 1, and six conditional BPAs* $m_B, m_{O|B}, m_{A|O}, m_S, m_{T|A,S}$*, and* $m_{M|B,T}$ *specified below. The model should answer how long it takes to change seasonal tires in a car repair shop. The total time is the sum of waiting time and time a car mechanic spends to perform the required work (time of service). The waiting time depends on whether the customer has made their online booking in advance or not. The booked customers are served with higher priority; their waiting time is zero. The working time depends on what is to be done. Partly it depends on whether the customer needs to change tires or the entire wheels. If the customer already has two sets of wheels on the discs, the mechanic need not remove and install tires; they disassemble the old and assemble new wheels (including balancing), which requires much less time. The mechanic needs another half an hour if the customer accepts a seasonal offer to carry out a winter test (a package covering checks of windscreen wipers and washers, tests of the brake system, battery, etc.) at a reduced price. The customer can accept this offer during the online booking. The unbooked customers are offered this service only if the capacity allows.*

The variables and state spaces are listed in Table 3.

Since we do not know the priors, both m_B *and* m_S *are vacuous.*

We assume that all booked customers and about one-half of the non-booked customers are offered the winter test. The conditional $m_{O|B}$ *is defined using Smets' conditional embedding to include this information in the model. First, consider* $m_{O|B=y}$ *as a deterministic Bayesian BPA with the focal elements* $m_{O|B=y}(\{O = y\}) = 1$*. Applying Equation (10), the embedded conditional* $m_{B=y,O}$ *is also deterministic BPA with one focal element* $m_{B=y,O}(\{(B = y, O = y), (B = n, O = n), (B = n, O = y)\}) = 1$.

The fact that half of the non-booked customers are also offered the winter test means that the conditional $m_{O|B=n}$ *is a*

[1] For a directed graph $G = (\mathcal{V}, \vec{E})$, and its node X we denote the set of parents of X, $Pa(X) = \{Y \in \mathcal{V} : (Y \to X) \in \vec{E}\}$. Such a graph is acyclic if its nodes can be ordered so that parents are always before their children. Such an ordering is called *topological* (for the directed graph G).

Table 3: The variables, their state spaces, and the meaning of the states.

Variable	Name	State space Ω	Meaning
B	**B**ooking	$\{y, n\}$	**y**es; **n**o
O	winter test - **O**ffer	$\{y, n\}$	**y**es; **n**o
A	winter test - **A**ccept	$\{y, n\}$	**y**es; **n**o
S	type of **S**ervice	$\{t, w\}$	changing **t**ires; changing entire **w**heels
T	**T**ime of service	$\{t_1, t_2, t_3\}$	30 min.; 60 min.; 90 min.
M	total ti**M**e	$\{p_1, p_2, p_3, p_4, p_5\}$	30 min.; 60 min.; 90 min.; two hours; three hours

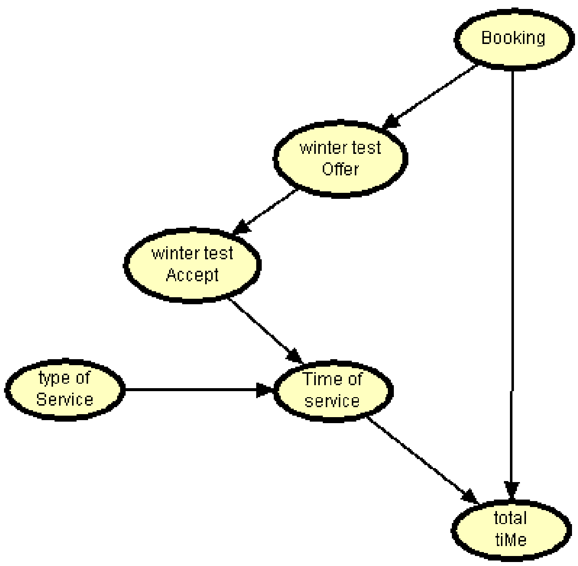


Figure 1: Graph for the Changing tires example

Bayesian BPA with two focal elements $m_{O|B=n}(\{O = n\}) = m_{O|B=n}(\{O = y\}) = 0.5$. *Applying Equation (10), the embedded conditional* $m_{B=n,O}$ *has also two focal elements* $m_{B=n,O}(\{(B = n, O = n), (B = y, O = n), (B = y, O = y)\}) = m_{B=n,O}(\{(B = n, O = y), (B = y, O = n), (B = y, O = y)\}) = 0.5$. *Thus, the conditional* $m_{O|B} = m_{B=y,O} \oplus m_{B=n,O}$ *has also two focal elements* $m_{O|B}(\{(B = n, O = n), (B = y, O = y)\}) = m_{O|B}(\{(B = n, O = y), (B = y, O = y)\}) = 0.5$.

Similarly, we get the conditional $m_{A|O}$ *describing that only the customers offered the winter test can accept it and that 80%* $m_{A|O}(\{(O = y, A = y), (O = n, A = n)\}) = 0.8$ *and* $m_{A|O}(\{(O = y, A = n), (O = n, A = n)\}) = 0.2$.

BPA $m_{T|A,S}$ *models the functional dependence of variable* T *on its parents. It encodes that the time of service is a sum of the time necessary to change tires and that necessary to realize winter test if required. Thus,* $m_{T|A,S}$ *encodes the following implications:*

- $(A = n \;\&\; S = w) \Longrightarrow T = t_1$;
- $(A = n \;\&\; S = t) \Longrightarrow T = t_2$;
- $(A = y \;\&\; S = w) \Longrightarrow T = t_2$;
- $(A = y \;\&\; S = t) \Longrightarrow T = t_3$.

So, $m_{T|A,S}$ *is a deterministic BPA with the focal element*

$$\mathsf{a} = \left\{ \begin{matrix} (A = n, S = w, T = t_1), (A = n, S = t, T = t_2), \\ (A = y, S = w, T = t_2), (A = y, S = t, T = t_3) \end{matrix} \right\},$$

which is a conditional BPA because $\mathsf{a}^{\downarrow\{A,S\}} = \Omega_{\{A,S\}}$

Not expecting that the waiting time for the unbooked customers depends on the considered variables, we define the last conditional BPA $m_{M|B,T}$ *to express the following implications:*

- $(B = y \;\&\; T = t_i) \Longrightarrow M = p_i$;
- $(B = n \;\&\; T = t_i) \Longrightarrow M = t_j$ *for some* $j \geq i$.

It is modeled by a deterministic BPA $m_{M|B,T}$, *the focal element* a *of which consists of the 15 states:*

$$\mathsf{a} = \left\{ \begin{matrix} (B = y, T = t_1, M = p_1), (B = y, T = t_2, M = p_2), \\ (B = y, T = t_3, M = p_3), (B = n, T = t_1, M = p_1), \\ (B = n, T = t_1, M = p_2), (B = n, T = t_1, M = p_3), \\ (B = n, T = t_1, M = p_4), (B = n, T = t_1, M = p_5), \\ (B = n, T = t_2, M = p_2), (B = n, T = t_2, M = p_3), \\ (B = n, T = t_2, M = p_4), (B = n, T = t_2, M = p_5), \\ (B = n, T = t_3, M = p_3), (B = n, T = t_3, M = p_4), \\ (B = n, T = t_3, M = p_5) \end{matrix} \right\}.$$

It is an easy task to check that $\mathsf{a}^{\downarrow\{B,T\}} = \Omega_{\{B,T\}}$, *which means that* $m_{M|B,T}$ *is a required conditional.*

The directed graphical model represents the joint BPA m *as follows:* [2]

$$\begin{aligned} m &= m_B \oplus m_S \oplus m_{O|B} \oplus m_{A|O} \oplus m_{T|A,S} \oplus m_{M|B,T} \\ &= m_B \triangleright_d m_S \triangleright_d m_{O|B} \triangleright_d m_{A|O} \triangleright_d m_{T|A,S} \triangleright_d m_{M|B,T} \,. \end{aligned}$$

[2] To keep the formulas uncluttered, we omit the parentheses when the operators of composition are performed successively from left to right.

In probability theory, the class of Bayesian networks is equivalent to the class of compositional models [3]. In the theory of belief functions, the relation of directed graphical models with compositional models is expressed in Proposition 10. It takes advantage of the fact that there exists a topological ordering of nodes of an acyclic directed graph.

Proposition 10 *Let* $\left(G = (\{X_1, \ldots, X_n\}, \vec{E}), \{m_i\}_{i=1,\ldots,n}\right)$ *be a directed graphical belief function model representing BPA* $m = \oplus_{i=1}^n m_i$*. If the ordering of variables* $(X_1, X_2, \ldots, X_n)$ *is topological with respect to G, then*

$$\begin{aligned} m &= m_1 \triangleright_d m_2 \triangleright_d \ldots \triangleright_d m_n \\ &= m^{\downarrow X_1} \triangleright_d m^{\downarrow \{X_2\} \cup Pa(X_2)} \triangleright_d \ldots \triangleright_d m^{\downarrow \{X_n\} \cup Pa(X_n)}. \end{aligned} \tag{11}$$

Proof The first equality in Equation (11) follows directly from the definition because for all $i = 2, \ldots, n$

$$\begin{aligned} (m_1 \triangleright_d \ldots \triangleright_d m_{i-1}) \triangleright_d m_i & \\ &= (m_1 \triangleright_d \ldots \triangleright_d m_{i-1}) \oplus m_i \ominus m_i^{\downarrow Pa(X_i)} \\ &= (m_1 \triangleright_d \ldots \triangleright_d m_{i-1}) \oplus m_i \end{aligned}$$

(notice that the last equality holds because the marginal $m_i^{\downarrow Pa(X_i)}$ is vacuous).

To prove that the second equality holds in Equation (11), realize that

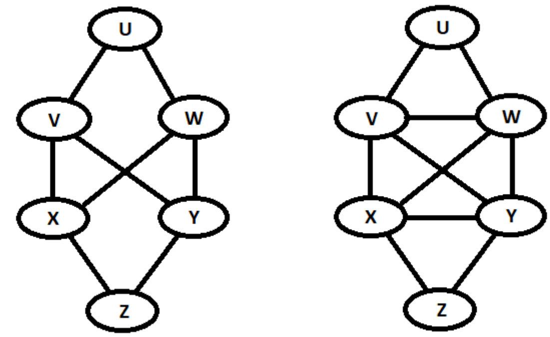
$$\begin{aligned} &(m^{\downarrow X_1} \triangleright_d \ldots \triangleright_d m^{\downarrow \{X_{i-1}\} \cup Pa(X_{i-1})}) \oplus m_i \\ &= (m^{\downarrow X_1} \triangleright_d \ldots \triangleright_d m^{\downarrow \{X_{i-1}\} \cup Pa(X_{i-1})}) \triangleright_d m^{\downarrow \{X_i\} \cup Pa(X_i)} \end{aligned}$$

follows from Corollary 4 (notice that the d-composition is defined because the conditional computed from $m^{\downarrow \{X_i\} \cup Pa(X_i)}$ equals m_i). ■

The assertion says that each directed graphical model can be represented equally efficiently by a compositional model. However, the converse does not hold for belief functions. It is because not all joint BPAs m defined for variables r can be expressed as Dempster's combination of its marginal $m^{\downarrow s}$ (for $s \subsetneq r$) with the respective conditional BPA ($m \ominus m^{\downarrow s}$). An example of such $m_{Y,Z}$ that cannot be expressed as Dempster's combination of its marginal $m_{Y,Z}^{\downarrow Y}$ with the respective conditional BPA $m_{Z|Y}$ is BPA m_2 in Example 1.

6. Undirected Graphical Models

Definition 11 *A belief function undirected graphical model is a couple* $\left(G = (\{X_1, \ldots, X_n\}, E), \{m_i\}_{i=1,\ldots,k}\right)$*, where G is an undirected graph with cliques*[3] $q_1, \ldots, q_k$*, and* $\{m_i\}_{i=1,\ldots,k}$ *is a collection of BPAs such that* m_i *is defined for variables* q_i*. Such an undirected graphical model represents a joint BPA m for* $\{X_1, \ldots, X_n\}$ *such that* $m = \oplus_{i=1}^k m_i$*.*



(*a*) Graph G (*b*) Triangulated graph $\hat{G}$

Figure 2: Undirected graphs for model from Example 4

Notice that implicit in the definition, the BPAs $\{m_i\}_{i=1,\ldots,k}$ are all distinct.

Example 4 *Consider a simple graph G with six nodes (binary variables) and eight edges from Figure 2(a). All cliques of this graph consist of two nodes; thus, the set of cliques equals the set of edges. A system of eight two-dimensional BPAs specifies an undirected graphical model with graph G, and the model represents a six-dimensional joint BPA*

$$\begin{aligned} m = m_1(U,V) \oplus m_2(U,W) \oplus m_3(V,X) \oplus m_4(V,Y) \\ \oplus m_5(W,X) \oplus m_6(W,Y) \oplus m_7(X,Z) \oplus m_8(Y,Z). \end{aligned}$$

From this expression, one can see that only a few parameters define such a graphical model. Their number is negligible in comparison with the potential number of focal elements of a general six-dimensional BPA ($2^{(2^6)} - 1$)*. Nevertheless, the only commonly known way to employ such an economically expressed joint BPA is described by Proposition 12.*

In probability theory, Markov networks (the term used by Pearl [13] for undirected graphical models) are much less popular than Bayesian networks. Nevertheless, their subclass, called decomposable models, are used in the Lauritzen-Spiegelhalter procedure for local computations. It employs the property that a probability distribution represented by a Bayesian network is a Markov model with

[3] A clique of a graph G is a maximal complete subset of nodes of G, i.e., a subset of mutually adjacent nodes of G that cannot be extended without violating the condition of mutual adjacency.

any graph, which is a supergraph of the moral graph[4] of the original acyclic directed graph. The following assertion shows that a similar result also holds for belief function models. The process of "moralization" and "triangulation" of an acyclic-directed graph can also be realized in the framework of belief functions to get a decomposable model equivalent to a directed graphical model. However, in contrast to probability theory, exploiting such a decomposable model for inference may fail because of the nonexistence of a necessary conditional BPA. This fact also manifests in the following assertion.

Proposition 12 *Suppose* $\big(G = (\{X_1,\ldots,X_n\},E),\{m_i\}_{i=1,\ldots,k}\big)$ *is an undirected graphical model representing BPA* $m = \oplus_{i=1}^{k} m_i$. *Consider an arbitrary decomposable supergraph* $\hat{G}$ *of* G, *i.e.,*

- *nodes of* $\hat{G}$ *are* $\{X_1,\ldots,X_n\}$;
- *all edges of* G *are also in* $\hat{G}$, *and*
- *cliques of* $\hat{G}$ *can be ordered to meet the running intersection property*[5] *(let it be the ordering* $\bar{q}_1,\ldots,\bar{q}_\ell$*).*

Then

$$m = m^{\downarrow \bar{q}_1} \triangleright_d m^{\downarrow \bar{q}_2} \triangleright_d \ldots \triangleright_d m^{\downarrow \bar{q}_\ell} \qquad (12)$$

if all the compositions are defined.

Proof To show that an undirected graphical model with graph G is also an undirected graphical model with any supergraph of G is trivial. Consider an undirected graphical model $\big(G = (\{X_1,\ldots,X_n\},E),\{m_i\}_{i=1,\ldots,k}\big)$. Let $\bar{G}$ with cliques $\bar{q}_1,\ldots,\bar{q}_\ell$ be a supergraph of G. Choose any mapping $h : \{1,...,k\} \to \{1,...,\ell\}$, such that $q_i \subseteq \bar{q}_{h(i)}$. Such a mapping exists because each clique q of G is a subset of (at least one) clique $\bar{q}$ of $\bar{G}$. For all $j = 1,\ldots,\ell$, define $\bar{m}_j = \big(\oplus_{i:h(i)=j} m_i\big) \oplus \iota_{q_j}$. Then, it is evident that $\big(\bar{G} = (\{X_1,\ldots,X_n\},\bar{E}),\{\bar{m}_i\}_{i=1,\ldots,\ell}\big)$ represents the same joint BPA as $\big(G = (\{X_1,\ldots,X_n\},E),\{m_i\}_{i=1,\ldots,k}\big)$ because $m = \oplus_{i=1}^{k} m_i = \oplus_{i=1}^{\ell} \bar{m}_i$.

Now assume $\bar{G}$ is decomposable and $\bar{q}_1,\ldots,\bar{q}_\ell$ are ordered to meet RIP. Consider an arbitrary $j \in \{2,\ldots,\ell\}$. Since the ordering $\bar{q}_1,\ldots,\bar{q}_\ell$ meets RIP, set $s = \bar{q}_j \cap (\bar{q}_1 \cup \ldots \cup \bar{q}_{j-1})$ is an articulation set in $\bar{G}$, and therefore the set of all cliques can be split into two disjoint parts $\{\bar{q}_i\}_{i\in I}$ and $\{\bar{q}_i\}_{i\in J}$, such that $(\cup_{i\in I}\bar{q}_i) \cap (\cup_{i\in J}\bar{q}_i) = s$.

Thus, $m = (\oplus_{i\in I}\bar{m}_i) \oplus (\oplus_{i\in J}\bar{m}_i)$, which corresponds with the definition of the conditional independence $\bar{q}_j \setminus s \perp\!\!\!\perp_m (\bar{q}_1 \cup \ldots \cup \bar{q}_{j-1}) \setminus s$. Applying Proposition 3 we get that

$$m^{\downarrow(\bar{q}_1 \cup \ldots \cup \bar{q}_j)} = m^{\downarrow(\bar{q}_1 \cup \ldots \cup \bar{q}_{j-1})} \triangleright_d m^{\downarrow \bar{q}_j} \qquad (13)$$

if this composition exists. Thus Equation (12) is proved because Equation (13) holds for all $j \in \{2,\ldots,\ell\}$, which finishes the proof. ∎

Example 4 (continued) *Graph* G *from Figure 2(a) can be triangulated by adding two edges. The resulting graph* $\hat{G}$ *(see Figure 2(b)) has three cliques:* $\{U,V,W\}$, $\{V,W,X,Y\}$, *and* $\{X,Y,Z\}$. *To get the corresponding decomposable compositional model specified in Proposition 12, one has to compute the respective two three-dimensional and one four-dimensional marginal BPAs. They can be computed using Shenoy-Shafer elimination algorithm ([18, 17]) getting:*

- $\bar{m}_1(U.V,W) = m_1(U,V) \oplus m_2(U,W) \oplus \Big(m_3(V,X) \oplus m_4(V,Y) \oplus m_5(W,X) \oplus m_6(W,Y) \oplus \big(m_7(X,Z) \oplus m_8(Y,Z)\big)^{\downarrow\{X,Y\}}\Big)^{\downarrow\{V,W\}}$.
- $\bar{m}_2(V,W,X,Y) = \big(m_1(U,V) \oplus m_2(U,W)\big)^{\downarrow\{V,W\}} \oplus m_3(V,X) \oplus m_4(V,Y) \oplus m_5(W,X) \oplus m_6(W,Y) \oplus \big(m_7(X,Z) \oplus m_8(Y,Z)\big)^{\downarrow\{X,Y\}}$.
- $\bar{m}_3(X,Y,Z) = \Big(\big(m_1(U,V) \oplus m_2(U,W)\big)^{\downarrow\{V,W\}} \oplus m_3(V,X) \oplus m_4(V,Y) \oplus m_5(W,X) \oplus m_6(W,Y)\Big)^{\downarrow\{X,Y\}} \oplus m_7(X,Z) \oplus m_8(Y,Z)$.

Due to Proposition 12, we get the original six-dimensional BPA m *composing these marginals in any ordering meeting RIP. So, if the following compositions are defined, then*

$$m = \bar{m}_1 \triangleright_d \bar{m}_2 \triangleright_d \bar{m}_3 = \bar{m}_3 \triangleright_d \bar{m}_2 \triangleright_d \bar{m}_1 = \bar{m}_3 \triangleright_d \bar{m}_2 \triangleright_d \bar{m}_1.$$

7. Summary, Conclusions, and Open Problems

The paper shows that the relationship between compositional and graphical models for belief functions is almost the same as in probability theory. As shown in Example 1, the difference results from that: not all joint BPAs can be decomposed as a combination of its marginal and the corresponding conditional. Avoiding this deficiency for compositional models is easier than for directed graphical models. It becomes evident also from studying the relationship between these two classes of models. As expressed in Proposition 10 and the ensuing comment, each directed graphical model can be represented as a compositional model, but the converse relation does not hold.

[4]An undirected graph $G = (V,E)$ is said to be a moral graph of an acyclic directed graph $\vec{G} = (V,\vec{E})$ if $E = \bigcup_{Z\in V} \{(X,Y) \in Pa(Z) \cup \{Z\}\}$.

[5]Ordered system of sets $\bar{q}_1,\bar{q}_2,\ldots,\bar{q}_\ell$ meets the *running intersection property* (RIP) if for all $i = 2,\ldots,\ell$ there exists j $(1 \le j < i)$ such that $\bar{q}_i \cap (\bar{q}_1 \cup \ldots \cup \bar{q}_{i-1}) \subseteq \bar{q}_j$.

Another interesting result concerns the two composition operators defined for belief functions. Recall that while f-composition ($\triangleright_f$) is always defined, the d-composition ($\triangleright_d$) remains undefined when the necessary conditional BPA does not exist. As expressed in Proposition 7, these two composition operators may but need not yield the same joint BPA. For example, if the Captain's problem described in the book by Almond [1] is described with a compositional model, both operators yield the same result. Therefore, for this example, one can use a (computationally) simpler f-composition instead of a d-composition. Notice that we designed the Changing tires problem in Example 3 so that the corresponding d-compositional and f-compositional models differ.

The relationship between the results of composition, when both the operators are defined, is expressed in Proposition 8. It says that the d-composition is more specific than the f-composition. This suggests using f-composition to approximate composition when the d-composition is undefined.

In practice, we never start with a large joint BPA for many variables. We usually construct one using a graphical or compositional model. It is typically intractable to explicitly compute the joint. Therefore, as in the probabilistic framework, the composition models are designed to enable inference with multidimensional BPAs. The application of the Shenoy-Shafer architecture for graphical models is straightforward, as no removals are involved.

The results presented in Sections 5 and 6 show that also Lauritzen-Spiegelhalter [12] and Hugin [11] architectures apply to the studied apparatus. A directed graphical model with graph G may easily be modified to get an undirected graphical model with a moral graph of G. Then, applying Proposition 12 to this undirected model, we obtain a decomposable model representing the same joint BPA as the original directed model. However, as already said several times, not all BPAs can be expressed as a combination of their marginal and the corresponding conditionals. This fact may cause problems with employing decomposable d-compositional models for inference.

Some of the problems may be avoided by extending the definition of a d-composition in Definition 2 to accept the situations when the conditional $m_2 \ominus m_2^{\downarrow r\cap s}$ is a pseudo-BPA (instead of $m_1 \triangleright_d m_2$ is undefined when $m_2 \ominus m_2^{\downarrow r\cap s}$ is not a BPA). It would be consistent with Lauritzen and Jensen [11], who show that the Hugin architecture (involving removals) also works for making inferences from belief-function directed graphical models. They show that the intermediate computations may result in pseudo-BPAs, but the marginals are always BPAs. However, a question remains — which assertions in this paper remain valid for the extended definition of d-composition?

The possibility of applying the graphical and compositional belief function models to practical problems would be enhanced by answering more open problems. We mention just two of them:

- We know that for some problems, the compositional models coincide whether we consider d-compositional or f-compositional models. Is it possible to characterize situations for which these compositional models coincide?
- As illustrated with the Changing Tires example, Smets' conditional embedding can be used to define conditionals required in directed graphical models. As mentioned prior to Example 3, Smets' approach can be used in more general situations than those illustrated in the example. Nevertheless, there are practical situations to which this approach is not applicable. It happens when one wants to embed several conditionals, and the respective conditions are not disjoint. Thus, one cannot use Smets' conditional embedding when, given three BPAs m_1, m_2, and m_3 for X, they want to find a conditional BPA $m_{X|Y}$ meeting the following three constraints

$$\left(m_{Y=0} \oplus m_{X|Y}\right)^{\downarrow X} = m_1,$$
$$\left(m_{Y=1} \oplus m_{X|Y}\right)^{\downarrow X} = m_2,$$
$$\left(m_{Y\in\{0,1\}} \oplus m_{X|Y}\right)^{\downarrow X} = m_3.$$

  Thus, there is a question about how to proceed in such situations.

Acknowledgments

This research was financially supported by the Czech Science Foundation Grant No. 21-07494S to the first two authors and by the Ronald G. Harper Professorship at the University of Kansas to the third author.

References

[1] R. G. Almond. *Graphical Belief Modeling*. Chapman & Hall, London, UK, 1995.

[2] F. V. Jensen and T. D. Nielsen. *Bayesian networks and decision graphs*, volume 2. Springer, 2007.

[3] R. Jiroušek. What is the difference between Bayesian networks and compositional models? In *Proceedings of the 7th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty*, pages 191–196, 2004. URL https://www.utia.cas.cz/cs/biblio?pub=0106358.

[4] R. Jiroušek. An attempt to define graphical models in Dempster-Shafer theory of evidence. In *Combining Soft Computing and Statistical Methods in Data Analysis*, pages 361–368. Springer, 2010.

[5] R. Jiroušek. On two composition operators in Dempster-Shafer theory. In *Proceedings of the 9th International Symposium on Imprecise Probability: Theories and Applications (ISIPTA 2015)*, pages 157–165, 2015.

[6] R. Jiroušek, J. Vejnarová, and M. Daniel. Compositional models for belief functions. In *Proceedings of ISIPTA*, volume 7, pages 243–252, 2007.

[7] R. Jiroušek, V. Kratochvíl, and P. P. Shenoy. Entropy for evaluation of Dempster-Shafer belief function models. *International Journal of Approximate Reasoning*, 151(12), 2022.

[8] R. Jiroušek, V. Kratochvíl, and P. P. Shenoy. Two composition operators for belief functions revisited. In *Proceedings of the 12th Workshop on Uncertainty Processing (WUPES-22)*, pages 123–134. MatfyzPress, 2022.

[9] R. Jiroušek and P. P. Shenoy. Compositional models in valuation-based systems. *International Journal of Approximate Reasoning*, 55(1):277–293, 2014.

[10] R. Jiroušek, V. Kratochvíl, and P. P. Shenoy. On conditional belief functions in the Dempster-Shafer theory. In S. Le Hégarat-Mascle, I. Bloch, and E. Aldea, editors, *Belief Functions: Theory and Applications, 7th International Conference, BELIEF 2022*, volume 13506 of *Lecture Notes in Artificial Intelligence*, pages 207–218, Switzerland, 2022. Springer Nature.

[11] S. L. Lauritzen and F. V. Jensen. Local computation with valuations from a commutative semigroup. *Annals of Mathematics and Artificial Intelligence*, 21(1): 51–69, 1997.

[12] S. L. Lauritzen and D. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, series B (Methodological)*, 50(2):157–224, 1988.

[13] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Elsevier, 2014.

[14] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[15] G. Shafer. Belief functions and parametric models. *Journal of the Royal Statistical Society, Series B (Methodological)*, 44(3):322–352, 1982.

[16] P. P. Shenoy. Conditional independence in valuation-based systems. *International Journal of Approximate Reasoning*, 10(3):203–234, 1994.

[17] P. P. Shenoy. Binary join trees for computing marginals in the shenoy-shafer architecture. *International Journal of approximate reasoning*, 17(2-3):239–263, 1997.

[18] P. P. Shenoy and G. Shafer. Axioms for probability and belief-function propagation. In R. D. Shachter, T. Levitt, J. F. Lemmer, and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 4*, volume 9 of *Machine Intelligence and Pattern Recognition Series*, pages 169–198. North-Holland, Amsterdam, Netherlands, 1990.

[19] P. Smets. *Un modele mathematico-statistique simulant le processus du diagnostic medical*. PhD thesis, Free University of Brussels, 1978.

[20] P. Smets. The canonical decomposition of a weighted belief. In *Proceedings of the 1995 IJCAI Conference*, volume 95, pages 1896–1901, 1995.

[21] H. Xu and P. Smets. Reasoning in evidential networks with conditional belief functions. *International Journal of Approximate Reasoning*, 14(2–3):155–185, 1996.