Extending Idioms for Bayesian Network Construction with Qualitative Constraints

Annet Onnes[1], Mehdi Dastani[1], Roel Dobbe[2], and Silja Renooij[1]

[1] Department of Information and Computing Sciences
Utrecht University, The Netherlands
{a.t.onnes, m.m.dastani, s.renooij}@uu.nl
[2] Faculty of Technology, Policy and Management
Delft University of Technology, The Netherlands
r.i.j.dobbe@tudelft.nl

Abstract. Bayesian networks (BNs) are compact representations of probability distributions that allow for supporting reasoning and decision making under uncertainty. Their interpretable structure and probability parameters allow for integrating human knowledge in their construction and explanation. For BN construction, reusable building blocks, or idioms, exist that describe the dependencies and reasoning patterns among small sets of variables. In this paper we formalise the concept of an idiom, explicitly including qualitative constraints that capture the reasoning patterns among variables as stated in the informal descriptions that accompany the idioms in literature. Our proposed formalisation ensures that idioms can be applied more consistently and reliably, improving the BN's accountability.

Keywords: Bayesian Networks · Idioms · Reasoning Patterns · Qualitative Constraints · QPNs

1 Introduction

AI systems for decision support, such as diagnosis or treatment selection systems in a medical setting, take part in safety-critical processes in which humans also play a role. In these Hybrid Intelligence settings where AI systems and humans operate towards the same goal, we want both the process of constructing the model as well as the resulting model to be reliable and transparent. The aim is to maintain transparency and reliability by ensuring human knowledge is integrated into an AI system in an explicit representation where it can be traced and is not unintentionally deformed.

Various approaches to integrating human knowledge in AI systems exist. Recent examples include reinforcement learning using human feedback (RLHF) [8]. The disadvantage of methods such as RLHF is that the human contributions are reduced to data points, which remain implicit throughout the modelling process and in the resulting model. This paper focuses on transparent models in which the knowledge provided by humans can be captured more explicitly in the structure of the model. More specifically, we focus on Bayesian networks (BNs).

BNs are representations of probability distributions that combine an interpretable graph with conditional probability tables [6]. Bayesian networks can be learned from data, but more complex and detailed models are often constructed manually. To aid manual construction, generally applicable building blocks, or *idioms*, have been proposed [11]. Using reusable building blocks should help to reduce the number of subjective modelling decisions that have to be made and explained, and as such ensure that general patterns of reasoning that are natural to and common among human experts are easily modelled in a *consistent* way. To represent such general patterns, idioms were proposed to abstract away from domain-specific details such as probabilities. Hence, they differ from BN-fragments, which are actual, domain-specific sub-networks that include specified probabilities and can be automatically combined into a complete BN for a given application [9].

In this paper, we propose an extension of the concept of an idiom to find a middle ground in the trade-off between general and domain-specific building blocks. We first introduce a general definition for the original concept of idiom, based on example idioms encountered in the literature. We then further formalise the concept by explicitly adding *qualitative interaction patterns* to their definition. These qualitative constraints on the conditional probabilities capture the reasoning patterns among variables that are associated with the idiom according to the accompanying informal description. We will show that making such constraints explicit allows for guaranteeing the desired reasoning patterns upon probability specification, where an informal description does not. Any type of probability constraint can be included in our formalisation. In this paper we first exploit constraints that are available from the literature on *Qualitative Probabilistic Networks* (QPNs) [17, 18]. QPN constraints have been used to study the interaction patterns in various causal interaction models, designed specifically for facilitating probability elicitation for variables with many parents in a BN [10].

We will review BNs and QPNs in Section 2. Subsequently, in Section 3, we discuss and provide a definition for the original concept of idiom and motivate the necessity of explicitly extending this definition with qualitative interaction patterns. Section 4 applies this extended definition to existing idioms from literature. Finally, we conclude the paper in Section 5.

2 Preliminaries

In this section we briefly review Bayesian networks and Qualitative Probabilistic Networks, while introducing our notations.

2.1 Bayesian Networks

A Bayesian network (BN) is a probabilistic graphical model that compactly represents a joint probability distribution $\Pr(\mathbf{V})$ over a set of random variables, from which any prior or posterior probability of interest can be efficiently computed [6]. Formally, a BN $\mathfrak{B} = (G, \Pr)$ has two elements. Firstly, an acyclic directed graph (DAG) $G = (\mathbf{V}_G, \mathbf{A}_G)$ with nodes $\mathbf{V}_G = \mathbf{V}$ corresponding to the variables and directed edges, or arcs, $\mathbf{A}_G$ describing the (in)dependences among $\mathbf{V}$. Secondly, the BN has local distributions $\Pr(V \mid \mathbf{Pa}(V))$ specified in conditional probability tables (CPTs) for each variable $V \in \mathbf{V}$, conditional on its parents $\mathbf{Pa}(V)$ in the graph G. In this paper, we focus on binary variables V with possible values $\Omega(V) = \{true, false\}$. We assume an ordering on values such that *true* $>$ *false* and denote the assignments of $true$ and $false$ to a variable V as v and $\bar{v}$, respectively.

For the construction of the BN-graph, arcs have to be directed. A common *heuristic* is to use the causal direction for arcs that could be interpreted as causal relations [6]. However, it is important to realize that the arcs do not necessarily *represent* causal relations [13]. Formally, the BN-graph is a so-called I-map, which encodes the independence relation among its variables; it therefore captures probabilistic (in)dependence rather than causality. As a result, choosing the right arc directions is not always trivial.

To aid BN construction, we can exploit the observation that it sees reoccurring patterns in reasoning, about which we repeatedly have to answer the same questions, concerning e.g. arc direction. Neil et al. [11] argued that by identifying natural and reusable patterns, called 'idioms', these questions can be answered in a generalised way. Idioms were introduced as building blocks to make the process of constructing complex BNs more consistent by formalising the reoccurring patterns. As such, the idioms would also lead to more interpretable BNs.

2.2 Qualitative Probabilistic Networks

Qualitative Probabilistic Networks (QPNs) are abstractions of Bayesian networks, where the CPTs are replaced by signs that denote qualitative influences and interaction patterns, or synergies [17, 18]. More formally, a QPN is a tuple $(G, \mathbf{Q})$, where the graph $G = (\mathbf{V}_G, \mathbf{A}_G)$ is the same as in a BN and $\mathbf{Q}$ is a set of *qualitative influences* and *synergies*, defined as constraints on Pr. Each qualitative influence and synergy has an associated sign $\delta \in \{0, +, -, ?\}$. Qualitative influences are defined for each arc in a QPN graph G:

Definition 1. *Consider variables $A, B \in \mathbf{V}_G$ with $A \rightarrow B \in \mathbf{A}_G$. Then, A* positively *influences B, written $S^+(A, B)$, iff*

$$\Pr(b|a, \mathbf{x}) \geq \Pr(b|\bar{a}, \mathbf{x})$$

for any combination of values $\mathbf{x}$ for the set of variables $\mathbf{X} = \mathbf{Pa}(B) \setminus \{A\}$.

Analogously, variable A *negatively* influences variable B, written $S^-(A, B)$, iff the inequality is reversed. By replacing the inequality with an equality we get a *zero* qualitative influence, written $S^0(A, B)$. An influence can also be non-monotonic, when the sign of the influence depends on specific values assigned to variables in the set $\mathbf{Pa}(B) \setminus \{A\}$. If this is the case, or if the influence is otherwise unknown, we get $S^?(A, B)$.

A positive qualitative influence captures a monotonicity property indicating that higher values of one variable make higher values of another more likely. For example, having a medical condition makes an associated symptom more likely. Such properties are relatively easy for people to establish.

QPNs also enable assigning signs to interactions between more than two variables. These interactions are defined as *synergies*, of which two types have been defined: additive and product synergy [17, 18]. In this paper, we will only use the latter and make some simplifying assumptions to its definition that suffice for the current paper; a more general definition can be found in [2].

Definition 2. *Consider variables* $A, B, C \in \mathbf{V}_G$ *with* $A \to C$, $B \to C \in \mathbf{A}_G$ *and* $\mathbf{Pa}(C) \setminus \{A, B\} = \emptyset$*. Then,* A *and* B *exhibit a* negative product synergy *with regard to the value assignment* $c_0 \in \{c, \bar{c}\}$ *to variable* C*, written* $X^-(\{A, B\}, c_0)$*, iff*

$$\Pr(c_0|a, b) \cdot \Pr(c_0|\bar{a}, \bar{b}) \leq \Pr(c_0|a, \bar{b}) \cdot \Pr(c_0|\bar{a}, b)$$

Positive, zero and ambiguous synergies are defined analogously. Note that a separate product synergy is defined for each value assignment to child C.

The product synergy was introduced to capture a significant pattern in reasoning called *explaining away*, which is a form of intercausal reasoning between two causes of a common effect. More specifically, a negative product synergy $X^-(\{A, B\}, c_0)$ is a necessary and sufficient condition for characterizing an explaining away effect between two a priori independent variables A and B, upon observing c_0 [18]. We will further elaborate on the explaining away pattern in the next section.

We note that both qualitative influences and product synergies in a QPN are defined in terms of probabilities from the CPT of the variable with the incoming arcs in the associated BN.

3 Idioms and Associated Semantics: A Formalisation

In this section we first focus on idioms as described and applied in literature and propose an initial general definition. We then argue that some of the semantics associated with idioms can and should be formalised further using probability constraints. We extend the initial definition of idiom to include such constraints.

3.1 A First Attempt at Defining Idioms

Neil et al.[11] introduce idioms to help "identify the semantics and graph structure syntax of common modes of uncertain reasoning" with the purpose of being able to reuse these as building blocks across BN models in a consistent way. These building blocks concern the graph only and ignore the associated CPTs, since the CPTs will depend on the how idioms are combined in the final model and their probabilities will be problem-specific [11].

To further increase consistency and transparency in the use of idioms, we propose to explicitly define the general concept of idiom based upon the *examples* of idioms described in the literature [11, 4, 7]. These example idioms are all presented as a graph and the accompanying text typically contains an informal natural language description of the semantics of the intended reasoning pattern.

Definition 3. *An* idiom $I_o = (G_I, D_I)$ *consists of*

- *a BN-graph* $G_I = (\mathbf{V}_I, \mathbf{A}_I)$ *over random variables* $\mathbf{V}_I$ *with* semantically meaningful, *interpretable names;*
- *a textual description* D_I *of the reasoning patterns among the variables.*

Example 1. Neil et al. [11, p. 272] introduce five idioms for generic patterns in reasoning, among which is the Cause-Consequence idiom. It illustrates that an idiom is a template for a general pattern of reasoning rather than an actual BN-fragment. Using the above definition of an idiom, we can describe the Cause-Consequence idiom $I_o^{\mathsf{CC}} = (G_I, D_I)$ as follows:

- $G_I = (\mathbf{V}_I, \mathbf{A}_I)$ with $\mathbf{V}_I = \{Input, Output\}$ and $\mathbf{A}_I = \{Input \rightarrow Output\}$
- D_I: [models a causal process between causes - the *input* to the process - and consequences - the *output* of the process].

Note that the direction of the arc, from *Input* (cause) to *Output* (consequence), aligns with using causality as a guidance heuristic for directing arcs (see Section 2.1), yet carries no formal causal semantics [13]. □

3.2 Interpreting the Semantics Associated with Idioms

The BN-graph G_I that is part of Definition 3 of an idiom has a limited associated semantics: it is merely an I-map expressing the probabilistic independencies among its variables.

In the description D_I, however, much more can be addressed, including the semantics associated with the idiom's intended reasoning pattern. The description allows a human to interpret it and identify the reasoning pattern that the idiom represents in the structure. This means that the descriptions provide essential additional information beyond the graphical structure of the idiom. BN modellers using idioms will thus heavily rely on this description to determine how to use the idioms' graph to fully represent the intended pattern.

To illustrate the importance of the description D_I for the idiom's interpretation, we consider an idiom proposed for the medical domain, the Comorbidity Common Symptomology (CCS) idiom [7].

Example 2. The CCS idiom models the relationships between two medical conditions and a common manifestation. We define the idiom as $I_o^{\mathsf{CCS}} = (G_I, D_I)$, where

- $G_I = (\mathbf{V}_I, \mathbf{A}_I)$ with $\mathbf{V}_I = \{CC, C, M\}$ and $\mathbf{A}_I = \{C \rightarrow M, CC \rightarrow M\}$, where CC is short for *Comorbid Condition*, C for *Condition*, and M for *Manifestation*;
- D_I: [models uncertain relationships between two conditions that share the same manifestation, such that when observing the shared manifestation both conditions are updated and knowing, in addition, CC to be true reduces the likelihood of C being true, and vice-versa.] [7, p.11].

Kyrimi et al. [7] describe the modelled reasoning patterns in detail. Some of these patterns can be read directly from the graph, such as "the converging connection between the two conditions makes them dependent when their common consequences are observed". However, the specifics of the induced dependency between the conditions that is given in description D_I indicate a case of *explaining away*, where the probability of one condition is expected to decrease upon observation of the other. The direction of change in probabilities cannot be established from the graph alone. □

The above example illustrates the importance of the description D_I in conveying the intended reasoning pattern underlying an idiom. If the graph alone cannot capture an important reasoning pattern, then in the final BN this pattern must be modelled through the CPTs. Since idioms abstract away from any underlying probabilities [11, p. 268] there is no guarantee that the BN still exhibits these reasoning patterns after the quantification step in which the CPTs are specified. We illustrate this observation with an example.

Example 3. Kyrimi et al. apply the CCS idiom to model the relation between the comorbid condition *Lung cancer* (L), the condition *Coronary artery disease* (A) and their common manifestation *Chest pain* (P). We compare two BN fragments based on this idiom, where CPTs are specified for the variables involved; the two instances differ only in the conditional probabilities for the manifestation. The prior probabilities of the conditions are $\Pr(l) = 0.2$ and $\Pr(a) = 0.25$. For the first instance we specify the following probabilities for the CPT of P:

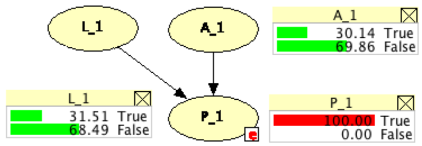
$$\Pr_1(p \mid l,\ a) = 0.8;\ \Pr_1(p \mid l,\ \bar{a}) = 0.5;\ \Pr_1(p \mid \bar{l},\ a) = 0.35;\ \Pr_1(p \mid \bar{l},\ \bar{a}) = 0.3.$$

and for the second instance we use these probabilities:

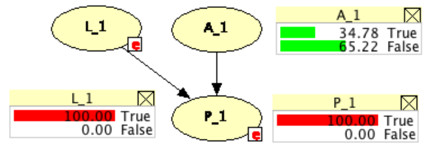
$$\Pr_2(p \mid l,\ a) = 0.9;\ \Pr_2(p \mid l,\ \bar{a}) = 0.75;\ \Pr_2(p \mid \bar{l},\ a) = 0.7;\ \Pr_2(p \mid \bar{l},\ \bar{a}) = 0.1.$$

In both cases we find that e.g. $\Pr(p \mid l) > \Pr(p)$ and $\Pr(p \mid a) > \Pr(p)$, so each condition increases the probability of having chest pain. Similarly, observing chest pain results in an increase of the probabilities of the conditions, compared to their priors (see Fig. 1(a),(c)). These patterns of reasoning are as expected. Fig. 1 further shows where the behaviour of the instances differs. If in addition to observing chest pain, we know the patient has lung cancer, the probability of coronary artery disease in the first instance increases even further (cf. Fig. 1(a) and (b)). That is, the *opposite* of explaining away occurs. For the second instance, contrarily, observing lung cancer decreases the probability of coronary artery disease (cf. Fig. 1(c) and (d)). Therefore, the second instance does exhibit an explaining away effect. □

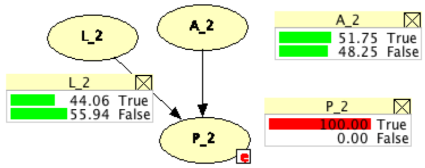
The difference between the two instances with identical graphical structures demonstrates that the DAG alone is not sufficient to guarantee the desired reasoning patterns upon quantification. We find that the description is an essential part of the idiom, as it is only there that we find these reasoning patterns. Constructing BNs that have to behave according to these patterns thus requires a



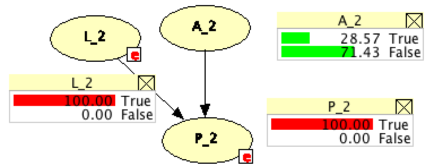
(a) Instance 1, p observed



(b) Instance 1, p and l observed



(c) Instance 2, p observed



(d) Instance 2, p and l observed

Fig. 1: BN inferences in Hugin[3]. The DAGs are based on the CCS idiom. The variable subscripts indicate two instances with different sets of CPTs, resulting in different reasoning patterns: (a) and (b) indicate "explaining in"; (c) and (d) indicate "explaining away".

strong grasp on ensuring that idioms that intend to model e.g. *explaining away* actually do so. The above example shows that this requires at least some information about the underlying probabilities. We therefore propose to further formalise the concept of an idiom by adding probability constraints that capture interaction patterns in a qualitative way.

3.3 Qualitative Interaction Patterns as Constraints

Idioms describe the structure and semantics of common reasoning patterns in a qualitative way. We propose to further formalise the interactions that are only available in the idiom's description D_I by introducing explicit *qualitative interaction patterns (QIPs)*. Many of these patterns, such as the explaining away pattern in the Comorbidity Common Symptomology idiom, implicitly imply that the probabilities associated with the idiom have to fulfil constraints. When using an idiom to construct a BN, we thus want to ensure that the associated CPTs adhere to these constraints. We would therefore like QIPs to have the following properties:

- they represent relative or absolute constraints on CPT probabilities;
- the constraints should suffice to guarantee that the intended reasoning pattern is indeed implemented upon quantification;

[3] Hugin Expert A/S – http://www.hugin.com/

– unlike CPTs in general, the constraints should be preserved upon combining idiom-based fragments into a larger network.

Explaining away and other qualitative influences, as introduced in Section 2.2, are examples of such qualitative patterns for which the probability constraints have been defined. For example, we recall that the 'explaining away' pattern, as identified in the description of the Comorbidity Common Symptomology idiom, is formalised by a negative product synergy [18]. More specifically, from Definition 2 we have that the following constraint should hold for the probabilities for the variables involved in the CCS idiom:

$$X^-(\{CC, C\}, m) \iff \Pr(m|cc, c) \cdot \Pr(m|\bar{c}c, \bar{c}) \leq \Pr(m|cc, \bar{c}) \cdot \Pr(m|\bar{c}c, c)$$

This constrains the CPT of the manifestation variable such that when $M = true$ (m), the probability of only one of the conditions (CC and C) is higher than when both would be *true*. Constraints like this can guide the challenging process of eliciting probabilities, as they limit which numerical values can be assigned while still exhibiting the desired behaviour [1, 14, 3].

3.4 Defining Idioms with Qualitative Interaction Patterns

In Section 3.1, we introduced a general definition for the concept of an idiom, based on the literature. Given our observations above, we explicitly extend this definition with formalised QIPs, derived from the idiom's description. Using QIPs enables us to place constraints on the CPTs to guarantee intended reasoning patterns, without the need for including actual CPTs as part of an idiom.

Definition 4. *An* idiom $I = (G_I, \mathbf{\Psi}_I, D_I)$ *consists of*

– *a BN-graph* $G_I = (\mathbf{V}_I, \mathbf{A}_I)$ *over random variables* $\mathbf{V}_I$ *with* semantically meaningful *names;*
– *a set of constraints* $\mathbf{\Psi}_I$ *that ensure the qualitative interaction patterns (QIPS);*
– *a textual description* D_I *of the reasoning patterns that refers to the variables and the QIPs.*

Note that we do not propose to *replace* the description D_I by the set of constraints $\mathbf{\Psi}_I$. Firstly, due to the informal nature of the description it is not precisely defined what information it can contain. Even in cases where it is possible to extract and formalise the reasoning patterns from the description in $\mathbf{\Psi}_I$, it might still be the case that D_I contains additional information relevant for the interpretation of the idiom. Secondly, the description can be used for explanation purposes. The description therefore still has benefits for transparency and interpretability in both the process of BN construction and its use.

4 Qualitative Interaction Patterns in Existing Idioms

The initial idioms for common reasoning patterns from Neil et al. have been extended upon to develop idioms helpful for BN construction in various domains,

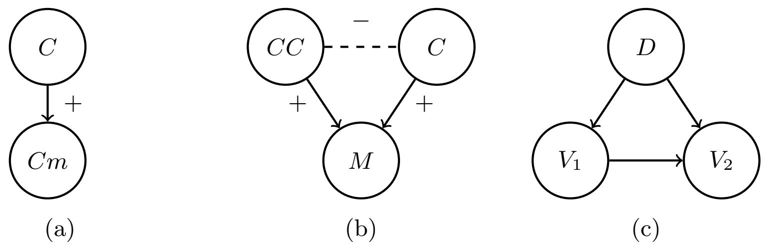


Fig. 2: Examples of idioms: (a) the Complication idiom, a variant of the Cause-Consequence idiom, (b) the Comorbidity Common Symptomology idiom, and (c) the Variation idiom for 2 variations.

such as the legal, risk safety and medical domains [11, 5, 7, 16]. In this section we will look at qualitative interaction patterns that can be identified in descriptions of such idioms.

4.1 Qualitative Influence QIPs in Medical Idioms

Recall from Section 2.2 that a qualitative influence in a QPN is the influence between two variables, which can be positive, negative, zero or ambiguous. We can identify such influences as qualitative interaction patterns in existing idioms. For example, in the Cause-Consequence idiom from Example 1 it appears that a cause occurring increases the probability of the consequence. This can be made explicit as a positive qualitative influence, adding this constraint to the idioms description: $\mathbf{\Psi} = \{S^+(Cause, Consequence)\}$.

Among the ten medical idioms introduced by Kyrimi et al. [7], we find the Complication idiom (Fig. 2a). This idiom is described as a variant of the Cause-Consequence idiom, where a condition (C) increases the probability of a complication (Cm). We can identify a positive qualitative influence constraint, $S^+(C, Cm)$. There are at least two more of these ten idioms where this constraint should hold, the Manifestation and Risk Factor idioms. A patient having a condition increases the probability of a manifestation (M) and a risk factor (RF) increases the probability of a condition. Both these changes in probability are QIPs, which can be formalised as $S^+(C, M)$ and $S^+(RF, C)$ respectively.

4.2 Qualitative Synergy QIP in Medical Idioms

In Section 3.2, we discussed the QIP contained in the description of the Comorbidity Common Symptomology idiom. It was described to exhibit *explaining away* behaviour. The involved variables represent two medical conditions, or causes more generally, and one manifestation, or an event more generally, connected as in the DAG of the CCS idiom as in Fig. 2b. For the CCS idiom, we can moreover indicate a positive influence for both arcs since the associated reasoning patterns are similar to that in the Cause-Consequence idiom. Explaining away, however, extends beyond influences on arcs because there is an induced intercausal interaction between the unconnected variables. The negative product synergy is indicated with a '$-$' above a dotted line that captures the intercausal interaction in Fig. 2b.

We can now extend the definition of the CCS idiom from Example 2 to explicitly include the QIPs. The Comorbidity Common Symptomology idiom is then defined by $I^{\mathsf{CCS}} = (G_I, \mathbf{\Psi}_I, D_I)$ where

- $G_I = (\mathbf{V}_I, \mathbf{A}_I)$ with $\mathbf{V}_I = \{CC, C, M\}$, where CC is a comorbid condition, C a condition and M a manifestation, and $\mathbf{A}_I = \{C \to M, CC \to M)\}$
- $\mathbf{\Psi}_I = \{S^+(CC, M), S^+(C, M), X^-(CC, C, M)\}$
- D_I: [models an uncertain relationship exhibiting an explaining away effect between two conditions sharing a manifestation, i.e. when observing the manifestation both conditions are updated and knowing, in addition, CC to be true reduces the likelihood of C being true, and vice-versa].

The constraints in $\mathbf{\Psi}_I$ cover what is described in D_I, such that by adhering to the constraints the intended explaining away behaviour of the idiom can be ensured.

4.3 Mutual Exclusivity Constraints in Legal idioms

So far, we have considered only QPN constraints, meaning that all extended idioms could be considered as QPN-fragments. Now we will consider another kind of constraint, appearing in an idiom from the wide range of idioms developed for the legal domain. Given the decisive nature of laws and rules, absolute constraints are often more fitting than relative constraints. The Variation idiom was introduced by Vlek et al. [16] for modelling variations in scenarios for criminal cases, e.g. two possible murder weapons leaving the same injury. It consists of an auxiliary disjunction variable (D) and $n \geq 2$ ordered variables V_i, one for each of the i variations. Fig. 2c shows the idiom's graph for $n = 2$. The idiom was later refined to capture mutual exclusivity among variations, by including additional constraints on the CPTs involved [15]. Based on these two definitions we define the mutual exclusivity QIP $M(V_1, \ldots, V_n \mid D)$ below. We first define, using Definition 4, the Variation idiom, for n variations by $I^{\mathsf{Vr}} = (G_I, \mathbf{\Psi}_I, D_I)$ where

- $G_I = (\mathbf{V}_I, \mathbf{A}_I)$ with $\mathbf{V}_I = \{D, V_1, \ldots, V_n\}$, where D is the Disjunction variable and $V_1, \ldots, V_n$ are the n variation variables, and $\mathbf{A}_I = \{D \to V_i \mid 1 \leq i \leq n\} \cup \{V_j \to V_i \mid 2 \leq i \leq n, 1 \leq j < i\}$.
- $\mathbf{\Psi}_I = \{M(V_1, \ldots, V_n \mid D)\}$
- D_I: [models mutually exclusive variations, where a disjunction variable represents the collection of variations, one of which is true].

The mutual exclusivity QIP $M(V_1, \ldots, V_n \mid D)$ in this idiom is defined by the following set of constraints [15]. Let $\mathbf{X_i} = \mathbf{Pa}(V_i) \setminus \{D\}$, $1 \leq i \leq n$. Then

$$\begin{array}{lll} \text{for all } V_i, 1 \leq i \leq n: & \Pr(v_i \mid \bar{d}, \mathbf{x}^*) = 0 & \text{for any assignment } \mathbf{x}^* \text{ for } \mathbf{X_i}; \\ \text{for all } V_i, 1 < i \leq n: & \Pr(v_i \mid d, \mathbf{x}^*) = 0 & \text{for any assignment } \mathbf{x}^* \text{ for } \mathbf{X_i} \text{ such that} \\ & & \exists V_j \in \mathbf{X_i} : V_j = \mathit{true}; \\ \text{and for } V_n: & \Pr(v_n \mid d, \overline{\mathbf{x}}) = 1 & \text{for assignment } \overline{\mathbf{x}} = \bar{v}_1, \ldots, \bar{v}_{n-1}. \end{array}$$

These constraints differ from QPN's relative qualitative constraints: instead of being expressed using inequalities between probabilities, these are absolute constraints on probabilities, but without completely defining the CPT. Since the constraints are flexible with respect to the number of variation variables included, we consider these mutual exclusivity constraints combined to be the QIP $M(V_1, \ldots, V_n \mid D)$. This QIP thus represents a pattern over all variations, enabled through an auxiliary variable D that is not a domain variable. We note that these mutual exclusivity constraints imply the following qualitative influences: $S^+(D, V_i)$ for each V_i and $S^-(V_j, V_i)$ on each arc $V_j \rightarrow V_i$ among the variation variables. $M(V_1, \ldots, V_n \mid D)$, however, represents a stronger set of constraints than the mentioned combination of qualitative influences.

5 Conclusion and Future Work

In this paper we formalised the concept of an idiom for BN construction, revealing a discrepancy between what an idiom intends to convey and what is represented by the graphical structure. More specifically, we showed that associated reasoning patterns cannot be captured by structure alone and proposed to extend idioms with probability constraints. We illustrated our proposal for several idioms available from literature. Probability constraints enhance idioms for the construction of BNs by further reducing the number of subjective modelling decisions about specifically assigning numerical values to the conditional probabilities. Hereby we increase transparency, efficiency and reliability of the modelling process and the models themselves.

The work in this paper can be continued in various directions. Firstly, for ease of exposition our examples had binary variables, while most idioms can also have non-binary variables. Using the generalised definitions of qualitative influences and synergies to study the constraints on non-binary variables is left for future work [17, 18]. The extended definition of idioms provides additional guidance, through the constraints, for determining the CPT probabilities. Future work can be done on how this additional guidance for using idioms can contribute to improving participatory methods for BN construction by domain experts directly [12].

Acknowledgements This research was supported by the Hybrid Intelligence Centre, a 10-year programme funded by the Dutch Ministry of Education, Culture and Science through the Netherlands Organisation for Scientific Research, https://hybrid-intelligence-centre.nl.

References

1. Druzdzel, M.J., Van der Gaag, L.C.: Building probabilistic networks: "where do the numbers come from?". IEEE Transactions on Knowledge and Data Engineering **12**(4), 481–486 (2000)
2. Druzdzel, M.J., Henrion, M.: Intercausal reasoning with uninstantiated ancestor nodes. In: Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence (1993)
3. Feelders, A.J., Van der Gaag, L.C.: Learning Bayesian network parameters under order constraints. International Journal of Approximate Reasoning pp. 37–53 (2006)
4. Fenton, N., Neil, M., Lagnado, D.A.: A general structure for legal arguments about evidence using Bayesian networks. Cognitive Science **37**(1), 61–102 (2013)
5. Hunte, J., Neil, M., Fenton, N.: Product safety idioms: a method for building causal Bayesian networks for product safety and risk assessment. arXiv **abs/2310.13595v2** (2022)
6. Jensen, F.V., Nielsen, T.D.: Bayesian Networks and Decision Graphs. Springer, 2nd edn. (2007)
7. Kyrimi, E., Neves, M.R., McLachlan, S., Neil, M., Marsh, W., Fenton, N.: Medical idioms for clinical Bayesian network development. Journal of Biomedical Informatics **108**, 103495 (2020)
8. Lambert, N., Gilbert, T.K., Zick, T.: The history and risks of reinforcement learning and human feedback. arXiv **abs/2310.13595** (2023)
9. Laskey, K.B., Mahoney, S.M.: Network fragments: Representing knowledge for constructing probabilistic models. In: Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence. pp. 334–341 (1997)
10. Lucas, P.J.F.: Bayesian network modelling through qualitative patterns. Artificial Intelligence **163**(2), 233–263 (2005)
11. Neil, M., Fenton, N., Nielson, L.: Building large-scale Bayesian networks. The Knowledge Engineering Review **15**(3), 257–284 (2000)
12. Nyberg, E.P., Nicholson, A.E., Korb, K.B., Wybrow, M., Zukerman, I., Mascaro, S., Thakur, S., Oshni Alvandi, A., Riley, J., Pearson, R., Morris, S., Herrmann, M., Azad, A., Bolger, F., Hahn, U., Lagnado, D.: BARD: A structured technique for group elicitation of Bayesian networks to support analytic reasoning. Risk Analysis **42**(6), 1155–1178 (2022)
13. Pearl, J.: Causality. Cambridge university press (2009)
14. Renooij, S., Van der Gaag, L.: From qualitative to quantitative probabilistic networks. In: Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence. pp. 422–429 (2002)
15. Vlek, C.: When stories and numbers meet in court: Constructing and Explaining Bayesian Networks for Criminal Cases with Scenarios. Ph.D. thesis, Rijksuniversiteit Groningen (2016)
16. Vlek, C.S., Prakken, H., Renooij, S., Verheij, B.: Unfolding crime scenarios with variations: A method for building a Bayesian network for legal narratives. In: Proceedings of the 26th International Conference on Legal Knowledge and Information Systems. pp. 145–154. IOS Press (2013)
17. Wellman, M.P.: Fundamental concepts of qualitative probabilistic networks. Artificial Intelligence **44**(3), 257–303 (1990)
18. Wellman, M., Henrion, M.: Explaining "explaining away". IEEE Transactions on Pattern Analysis and Machine Intelligence **15**(3), 287–292 (1993)