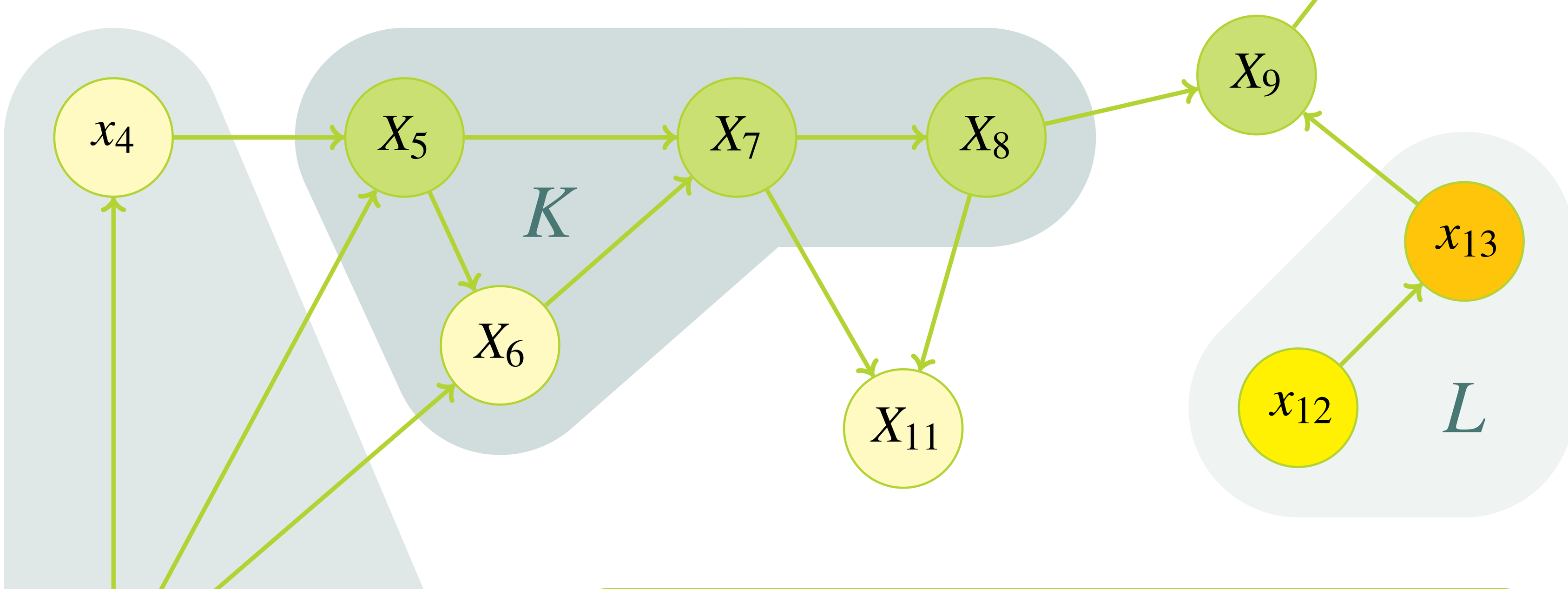
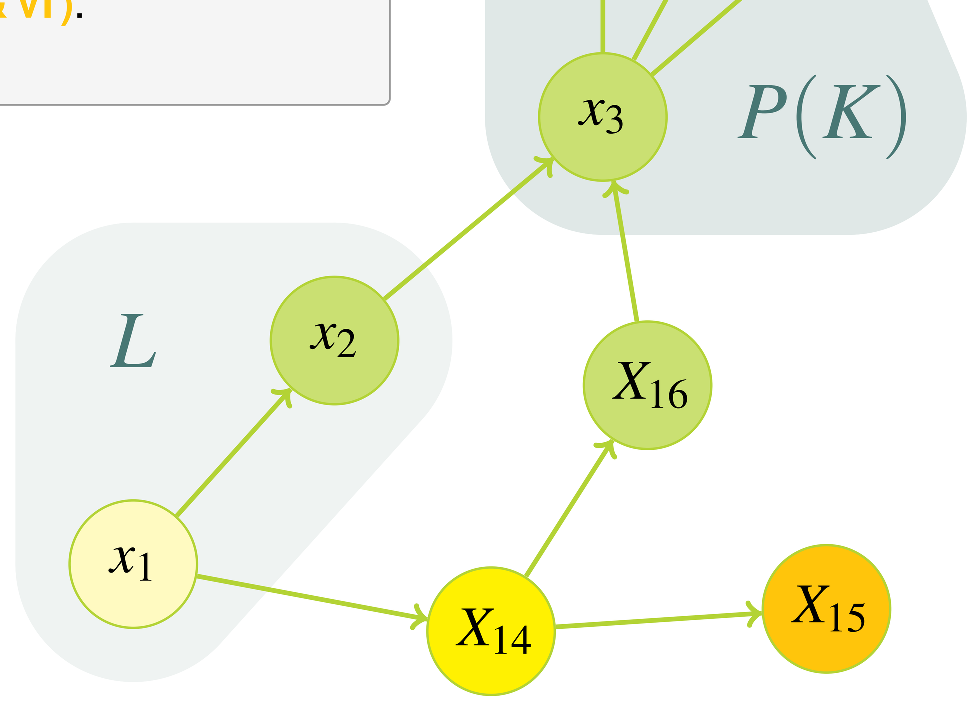
Credal networks under epistemic irrelevance using sets of desirable gambles

Jasper De Bock and Gert de Cooman
{Jasper.DeBock,Gert.deCooman}@UGent.be
SYSTeMS, Ghent University, Belgium





Abstract of the paper

We present a new approach to credal networks, which are graphical models that generalise Bayesian nets to deal with imprecise probabilities. Instead of applying the commonly used notion of strong independence, we replace it by the weaker notion of epistemic irrelevance (II). We show how assessments of epistemic irrelevance allow us to construct a global model out of given local uncertainty models (I), leading to an intuitive expression for the so-called irrelevant natural extension (III) of a network. In contrast with Cozman (2000) who introduced this notion in terms of credal sets, our main results are presented using the language of sets of desirable gambles (SDG). This has allowed us to derive a number of useful properties of the irrelevant natural extension. It has powerful marginalisation properties (IV) and satisfies all graphoid properties but symmetry, both in their direct and reverse forms (V & VI).

Sets of desirable gambles (SDG)

We will model a subject's beliefs about the value that a variable X, assumes in some set $\mathscr{X}$, by means of his behaviour: which gambles (real-valued maps) f on $\mathscr{X}$ does he strictly prefer to the status quo. This results in a **set of desirable gambles** $\mathscr{D} \subseteq \mathscr{G}(\mathscr{X})$, where $\mathscr{G}(\mathscr{X})$ is the set of all gambles on $\mathscr{X}$. $\mathscr{D}$ is called **coherent** if it satisfies the rationality requirements D1—D4 for all $f, f_1, f_2 \in \mathscr{G}(\mathscr{X})$ and all real $\lambda > 0$.

- D1 $f \leq 0 \Rightarrow f \notin \mathscr{D}$
- D2 $f > 0 \Rightarrow f \in \mathscr{D}$
- D3 $f \in \mathscr{D} \Rightarrow \lambda f \in \mathscr{D}$
- D4 $f_1, f_2 \in \mathscr{D} \Rightarrow f_1 + f_2 \in \mathscr{D}$

Although they are not as well known as other (imprecise) probability models, **sets of desirable gambles have definite advantages**. To give a few examples: they are operational, are easily **able to deal with conditioning on events with probability zero**, allow for intuitive **geometrically flavoured proofs** and are **more expressive than both credal sets and lower previsions** (see our papers or the second poster for credal networks under epistemic irrelevance that use these alternative models).

Local uncertainty models (I)

With every node s of a finite **directed acyclic graph (DAG)**, we associate a variable X_s taking values in some finite, non-empty set $\mathscr{X}_s$. The set of all nodes is denoted by G. For every subset $S \subseteq G$, the joint variable X_S takes values in $\mathscr{X}_S := \times_{s \in S} \mathscr{X}_s$. For every $s \in G$, we denote by $P(s)$ the set consisting of the parent nodes of s. Similar to what is done in classical Bayesian networks, **we attach local uncertainty models to the nodes of the network, conditional on the value of their parents**. For all $s \in G$ and every instantiation $x_{P(s)} \in \mathscr{X}_{P(s)}$ of $X_{P(s)}$, we require a coherent set $\mathscr{D}_{s \rfloor x_{P(s)}}$ of desirable gambles (SDG) on $\mathscr{X}_s$.

Epistemic irrelevance (II)

Consider a global set of desirable gambles $\mathscr{D}_G$ (SDG) on $\mathscr{X}_G$ (I) and disjoint subsets S and K of G. Then the marginal model for X_K, conditional on the information that X_S assumes a value $x_S \in \mathscr{X}_S$, is given by

$$\operatorname{marg}_K(\mathscr{D}_G \rfloor x_S) = \{f \in \mathscr{G}(\mathscr{X}_K) \colon \mathbb{I}_{\{x_S\}} f \in \mathscr{D}_G\}.$$

Consider now three subsets $C, I, O \subseteq G$, with $I \setminus C$ and $O \setminus C$ disjoint. We say that X_I is **epistemically irrelevant** to X_O conditional on X_C, denoted as $\mathrm{IR}(I, O|C)$, if and only if for all $x_{C \cup I} \in \mathscr{X}_{C \cup I}$ we have

$$\operatorname{marg}_{O \setminus C}(\mathscr{D}_G \rfloor x_{C \cup I}) = \operatorname{marg}_{O \setminus C}(\mathscr{D}_G \rfloor x_C).$$

Our paper also considers **epistemical subset-irrelevance**, which although interesting, is not discussed on this poster.

Irrelevant natural extension (III)

We define the **irrelevant natural extension** $\mathscr{D}_G^{\mathrm{irr}}$ of a credal network as the **most conservative** (smallest) **coherent** set of desirable gambles on $\mathscr{X}_G$ that (a) **marginalises to the given local uncertainty models** $\mathscr{D}_{s \rfloor x_{P(s)}}$ (I) and (b) for which, for any node s, its **non-parent non-descendant variables** $X_{N(s)}$ **are epistemically irrelevant** to X_s conditional on $X_{P(s)}$ (II). If we use $PN(s)$ as a shorthand notation for $P(s) \cup N(s)$, then requirements (a) and (b) are equivalent to requiring that for all $s \in G$ and $x_{PN(s)} \in \mathscr{X}_{PN(s)}$

$$\operatorname{marg}_s(\mathscr{D}_G^{\mathrm{irr}} \rfloor x_{PN(s)}) = \operatorname{marg}_s(\mathscr{D}_G^{\mathrm{irr}} \rfloor x_{P(s)}) = \mathscr{D}_{s \rfloor x_{P(s)}}.$$

We show that this irrelevant natural extension is **simple to construct**: $\mathscr{D}_G^{\mathrm{irr}} := \operatorname{posi}(\mathscr{A}_G^{\mathrm{irr}})$, where the 'posi'-operator generates the set of all finite positive linear combinations of elements in its argument set

$$\mathscr{A}_G^{\mathrm{irr}} := \{\mathbb{I}_{\{x_{PN(s)}\}} f \colon s \in G, x_{PN(s)} \in \mathscr{X}_{PN(s)}, f \in \mathscr{D}_{s \rfloor x_{P(s)}}\}.$$

Marginalisation properties (IV)

For any $K \subseteq G$, we **construct a sub-DAG of the original DAG** by eliminating the nodes $s \in G \setminus K$ and their associated edges. The parents of a node $s \in K$, with respect to this sub-DAG, are denoted by $P_K(s) := P(s) \cap K$. We derive local models $\mathscr{D}_{s \rfloor x_{P_K(s)}}$ for this sub-DAG from the original local models $\mathscr{D}_{s \rfloor x_{P(s)}}$ by fixing $x_{P(s) \setminus K}$. We do this consistently for all $s \in K$ at once by fixing $x_{P(K)}$, where $P(K) := (\bigcup_{s \in K} P(s)) \setminus K$. For any $x_{P(K)} \in \mathscr{X}_{P(K)}$, we use the resulting local models to **construct an irrelevant natural extension of the sub-DAG** and denote it by $\mathscr{D}_{K \rfloor x_{P(K)}}^{\mathrm{irr}}$.

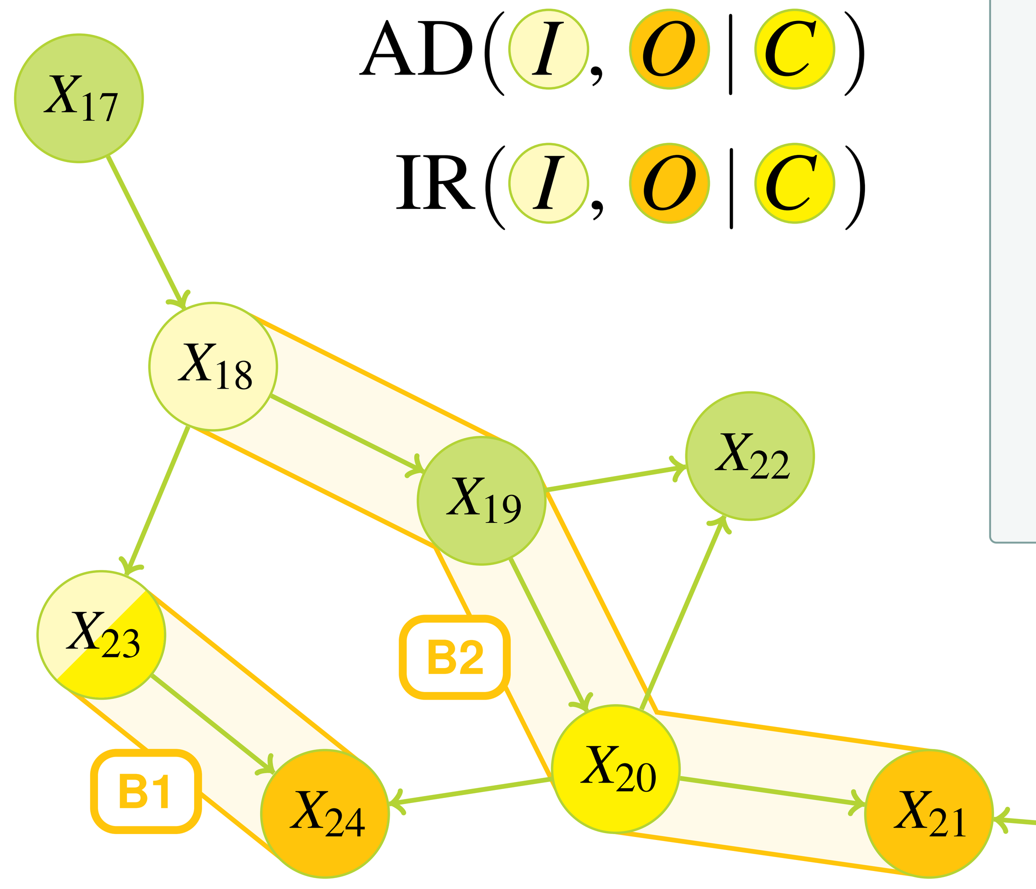
Suppose now that K is **closed** subset of G, meaning that $P(K) \cap D(K) = \emptyset$, where $D(K) := (\bigcup_{s \in K} D(s)) \setminus K$ and for all $s \in K$, $D(s)$ are the descendants of s. Then $\mathscr{D}_{K \rfloor x_{P(K)}}^{\mathrm{irr}}$ is related to $\mathscr{D}_G^{\mathrm{irr}}$ as follows: for all $L \subseteq N(K)$, with $N(K) := \bigcap_{s \in K} N(s)$, and for all $x_{P(K) \cup L} \in \mathscr{X}_{P(K) \cup L}$

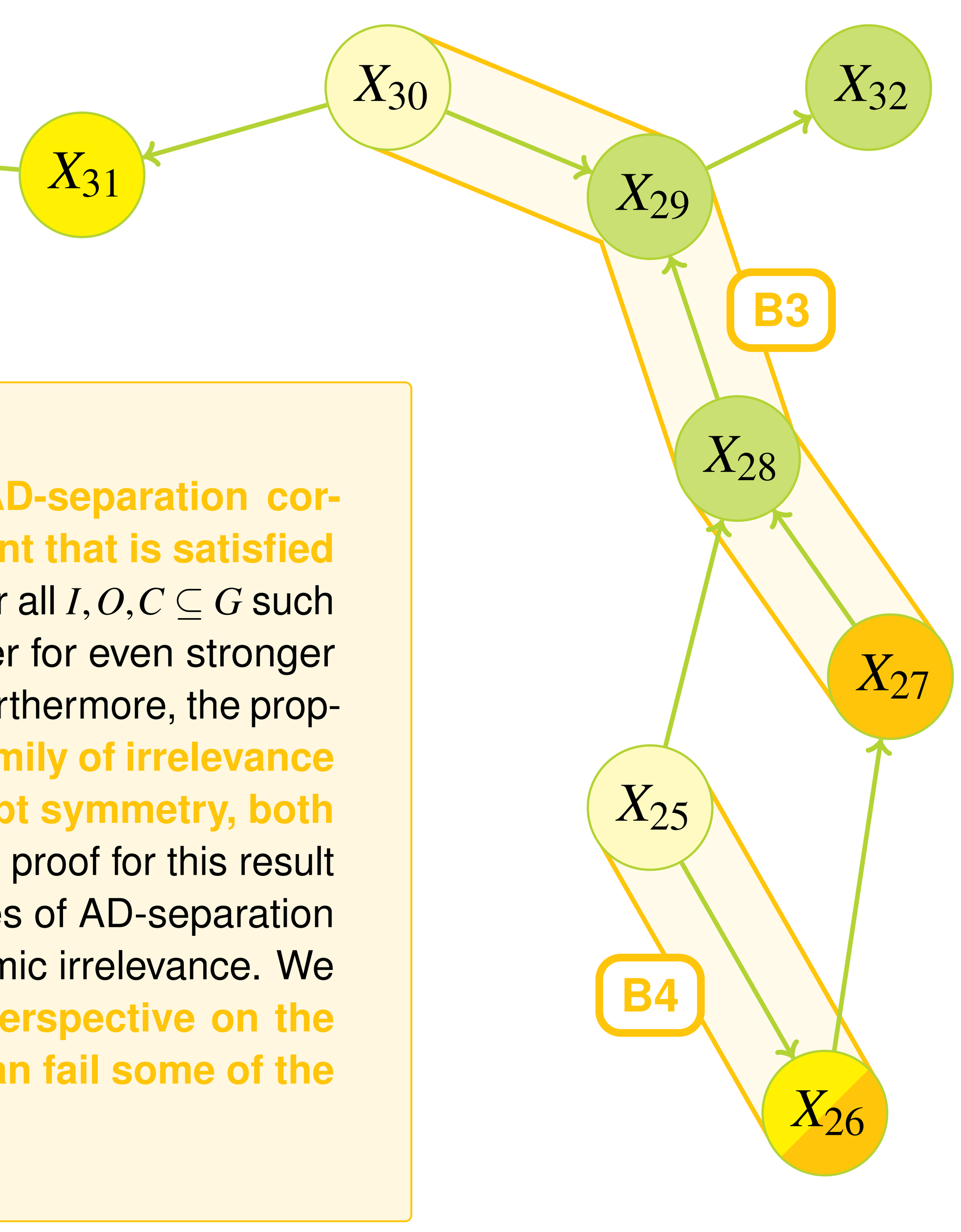
$$\operatorname{marg}_K(\mathscr{D}_G^{\mathrm{irr}} \rfloor x_{P(K) \cup L}) = \mathscr{D}_{K \rfloor x_{P(K)}}^{\mathrm{irr}}.$$

We refer to our paper for **marginalisation properties** that are even stronger than the one given above.

$\mathrm{AD}(I, O \mid C)$

$\mathrm{IR}(I, O \mid C)$





AD-separation (V)

Consider any path $s_1, \ldots, s_n$ in G, with $n \geq 1$. We say that this path is **blocked** by a set of nodes $C \subseteq G$ whenever at least one of the following four conditions holds:

- B1 $s_1 \in C$;
- B2 there is some $1 < i < n$ such that $s_i \to s_{i+1}$ and $s_i \in C$;
- B3 there is some $1 < i < n$ such that $s_{i-1} \to s_i \leftarrow s_{i+1}$, $s_i \notin C$ and $D(s_i) \cap C = \emptyset$;
- B4 $s_n \in C$.

Now consider (not necessarily disjoint) subsets I, O and C of G. We say that O is **AD-separated** from I by C, denoted as $\mathrm{AD}(I, O|C)$, **if every path** $i = s_1, \ldots, s_n = o$, $n \geq 1$, from a node $i \in I$ to a node $o \in O$, **is blocked** by C.

This asymmetrical version of D-separation is similar to, yet different from both Moral's (2005) version of AD-separation and the notion of L-separation, as introduced by Vantaggi (2002). Our reason for not using one of these existing concepts is that our version of **AD-separation** has stronger properties: it **satisfies all graphoid properties except symmetry: it satisfies redundancy, decomposition, weak union, contraction and intersection both in their direct and reverse forms.**

Graphoid properties (VI)

In a credal network under epistemic irrelevance, **every AD-separation corresponds to a conditional epistemic irrelevance statement that is satisfied by the corresponding irrelevant natural extension** $\mathscr{D}_G^{\mathrm{irr}}$: for all $I, O, C \subseteq G$ such that $\mathrm{AD}(I, O|C)$ we have $\mathrm{IR}(I, O|C)$ (II) (we refer to our paper for even stronger results regarding factorisation and subset-irrelevance (II)). Furthermore, the properties of AD-separation (V) imply that **the corresponding family of irrelevance statements is closed under all graphoid properties except symmetry, both in their direct and reverse form**. We want to stress that our proof for this result is based solely on marginalisation results (IV) and properties of AD-separation (V). At no point does it invoke graphoid properties of epistemic irrelevance. We believe that **this provides a new, rather more positive perspective on the well-known fact that epistemic irrelevance, in general, can fail some of the graphoid axioms** (Cozman 2005).