On the Foundations of Cycles
in Bayesian Networks[⋆]

Christel Baier[1], Clemens Dubslaff[1], Holger Hermanns[2,3], and Nikolai Käfer[1]

[1] TU Dresden, Dresden, Germany
[2] Saarland University, Saarbrücken, Germany
[3] Institute of Intelligent Software, Guangzhou, China

Abstract. Bayesian networks (BNs) are a probabilistic graphical model widely used for representing expert knowledge and reasoning under uncertainty. Traditionally, they are based on directed acyclic graphs that capture dependencies between random variables. However, directed cycles can naturally arise when cross-dependencies between random variables exist, e.g., for modeling feedback loops. Existing methods to deal with such cross-dependencies usually rely on reductions to BNs without cycles. These approaches are fragile to generalize, since their justifications are intermingled with additional knowledge about the application context. In this paper, we present a foundational study regarding semantics for cyclic BNs that are generic and conservatively extend the cycle-free setting. First, we propose constraint-based semantics that specify requirements for full joint distributions over a BN to be consistent with the local conditional probabilities and independencies. Second, two kinds of limit semantics that formalize infinite unfolding approaches are introduced and shown to be computable by a Markov chain construction.

1 Introduction

A *Bayesian network* (BN) is a probabilistic graphical model representing a set of random variables and their conditional dependencies. BNs are ubiquitous across many fields where reasoning under uncertainties is of interest [10]. Specifically, a BN is a directed acyclic graph with the random variables as nodes and edges manifesting conditional dependencies, quantified by *conditional probability tables* (CPTs). The probability of any random variable can then be deduced by the CPT entries along all its predecessors. Here, these probabilities are independent of all variables that are no (direct or transitive) predecessors in the graph. Acyclicity is hence crucial and commonly assumed to be rooted in some sort of causality [23]. A classical use of BNs is in expert systems [22] where BNs aggregate statistical data obtained by several independent studies. In the medical

---

Y	$X{=}\mathsf{T}$	$X{=}\mathsf{F}$
F	s_1	$1 - s_1$
T	s_2	$1 - s_2$

X	$Y{=}\mathsf{T}$	$Y{=}\mathsf{F}$
F	t_1	$1 - t_1$
T	t_2	$1 - t_2$
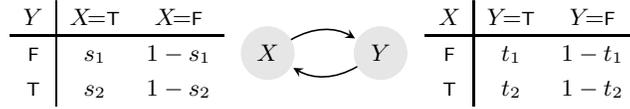
Fig. 1: A cyclic GBN with CPTs for X and Y

domain, e.g., they can capture the correlation of certain symptoms, diseases, and human factors [11,15,26].

Imagine for instance an expert system for supporting diagnosis of Covid-19, harvesting multiple clinical studies. One study might have investigated the percentage of patients who have been diagnosed with fever also having Covid-19, while another study in turn might have investigated among the Covid-19 patients whether they have fever, too. Clearly, both studies investigate the dependency between fever and Covid-19, but under different conditions. Fever may weaken the immune system and could increase the risk of a Covid-19 infection, while Covid-19 itself has fever as a symptom. In case there is uniform knowledge about "which symptom was first" in each of the constituent studies, then *dynamic Bayesian networks* (DBNs) [19] could be used as a model for the expert system, breaking the interdependence of fever and Covid-19 through a precedence relation. However, this implies either to rely only on studies where these temporal dependencies are clearly identified or to introduce an artificial notion of time that might lead to spurious results [18]. A naive encoding into the BN framework always yields a graph structure that contains cycles, as is the case in our small example shown in Fig. 1 where X and Y stand for the random variables of diagnosing Covid-19 and fever, respectively.

That cycles might be unavoidable has already been observed in seminal papers such as [22,15]. But acyclicity is crucial for computing the joint probability distribution of a BN, and thereby is a prerequisite for, e.g., routine inference tasks. Existing literature that considers cycles in BNs mainly recommends reducing questions on the probability values to properties in acyclic BNs. For instance, in [11] nodes are collapsed towards removing cycles, while [22] suggests to condition on each value combination on a cycle, generating a decomposition into tree-like BNs and then averaging over the results to replace cycles. Sometimes, application-specific methods that restructure the cyclic BN towards an acyclic BN by introducing additional nodes [26,8] or by unrolling cycles up to a bounded depth [17,2] have been reported to give satisfactory results. Other approaches either remove edges that have less influence or reverse edges on cycles (see, e.g., [10]). However, such approaches are highly application dependent and hinge on knowledge about the context of the statistical data used to construct the BN. Furthermore, as already pointed out by [30], they usually reduce the solution space of families of joint distributions to a single one, or introduce solutions not consistent with the CPTs of the original cyclic BN. While obviously many practitioners have stumbled on the problem how to treat cycles in BNs

and on the foundational question "What is the meaning of a cyclic BN?", there is very little work on the foundations of Bayesian reasoning with cycles.

In this paper, we approach this question by presenting general semantics for BNs with cycles, together with algorithms to compute families of joint distributions for such BNs. First, we investigate how the two main constituents of classical BNs, namely consistency with the CPTs and independencies induced by the graph structure, influence the joint distributions in the presence of cycles. This leads to *constraints semantics* for cyclic BNs that comprise all those joint distributions respecting the constraints, being either a single uniquely defined one, none, or infinitely many distributions. Second, we present semantics that formalize unfolding approaches and depend on the choice of a *cutset*, a set of random variables that break every cycle in a cyclic BN. Intuitively, such cutsets form the seams along which feedback loops can be unraveled. These semantics are defined in terms of the limit (or limit average) of a sequence of distributions at descending levels in the infinite unfolding of the BN. We show that the same semantics can be defined using a Markov chain construction and subsequent long-run frequency analysis, which enables both precise computation of the semantics and deep insights in the semantics' behavior. Among others, an immediate result is that the family of distributions induced with respect to the limit semantics is always non-empty. As we will argue, the limit semantics have obvious relations to a manifold of approaches that have appeared in the literature, yet they have not been spelled out and studied explicitly.

1.1 Notation

Let $\mathcal{V}$ be a set of Boolean random variables[4] over the domain $\mathbb{B} = \{\mathsf{F}, \mathsf{T}\}$. We usually denote elements of $\mathcal{V}$ by X, Y, or Z. An *assignment* over $\mathcal{V}$ is a function $b\colon \mathcal{V} \to \mathbb{B}$ which we may specify through set notation, e.g., $b = \{X{=}\mathsf{T}, Y{=}\mathsf{F}\}$ for $b(X) = \mathsf{T}$ and $b(Y) = \mathsf{F}$, or even more succinctly as $X\overline{Y}$. The set of all possible assignments over $\mathcal{V}$ is denoted by $Asg(\mathcal{V})$. We write $b_{\mathcal{U}}$ for the restriction of b to a subset $\mathcal{U} \subseteq \mathcal{V}$, e.g., $b_{\{X\}} = \{X{=}\mathsf{T}\}$, and may omit set braces, e.g., $b_{X,Y} = b_{\{X,Y\}}$.

A *distribution* over a set $\mathcal{S}$ is a function $\mu\colon \mathcal{S} \to [0,1]$ where $\sum_{s \in \mathcal{S}} \mu(s) = 1$. The set of all distributions over $\mathcal{S}$ is denoted by $Dist(\mathcal{S})$. For $|\mathcal{S}| = n$, μ will occasionally be represented as a vector of size n for some fixed order on $\mathcal{S}$. In the following, we are mainly concerned with distributions over assignments, that is distributions $\mu \in Dist(Asg(\mathcal{V}))$ for some set of random variables $\mathcal{V}$. Each such distribution μ induces a probability measure (also called μ) on $2^{Asg(\mathcal{V})}$. Thus, for a set of assignments $\phi \subseteq Asg(\mathcal{V})$, we have $\mu(\phi) = \sum_{b \in \phi} \mu(b)$. We are often interested in the probability of a *partial assignment* $d \in Asg(\mathcal{U})$ on a subset $\mathcal{U} \subsetneq \mathcal{V}$ of variables, which is given as the probability of the set of all full

---

[4] We use Boolean random variables for simplicity of representation, an extension of the proposed semantics over random variables with arbitrary finite state spaces is certainly possible.

assignments $b \in Asg(\mathcal{V})$ that agree with d on $\mathcal{U}$. As a shorthand, we define

$$\mu(d) \; := \; \mu\big(\{b \in Asg(\mathcal{V}) : b_{\mathcal{U}} = d\}\big) \; = \sum_{\substack{b \in Asg(\mathcal{V}) \\ \text{s.t. } b_{\mathcal{U}} = d}} \mu(b).$$

The special case $\mu(X{=}\top)$ is called the *marginal probability* of X. The restriction of $\mu \in Dist(Asg(\mathcal{V}))$ to $\mathcal{U}$, denoted $\mu|_{\mathcal{U}} \in Dist(Asg(\mathcal{U}))$, is given by $\mu|_{\mathcal{U}}(d) := \mu(d)$. For a set $\mathcal{W}$ disjoint from $\mathcal{V}$ and $\nu \in Dist(Asg(\mathcal{W}))$, the *product distribution* of μ and ν is given by $(\mu \otimes \nu)(c) := \mu(c_{\mathcal{V}}) \cdot \nu(c_{\mathcal{W}})$ for every $c \in Asg(\mathcal{V} \cup \mathcal{W})$. μ is called a *Dirac distribution* if $\mu(b) = 1$ for some assignment $b \in Asg(\mathcal{V})$ and thus $\mu(c) = 0$ for all other assignments $c \neq b$. A Dirac distribution derived from a given assignment b is denoted by $Dirac(b)$.

Graph Notations. For a graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ with nodes $\mathcal{V}$ and directed edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, we may represent an edge $(X, Y) \in \mathcal{E}$ as $X \to Y$ if $\mathcal{E}$ is clear from context. $Pre(X) := \{Y \in \mathcal{V} : Y \to X\}$ denotes the set of *parents* of a node $X \in \mathcal{V}$, and $Post^*(X) := \{Y \in \mathcal{V} : X \to \cdots \to Y\}$ is the set of nodes *reachable* from X. A node X is called *initial* if $Pre(X) = \varnothing$, and $Init(\mathcal{G})$ is the set of all nodes initial in $\mathcal{G}$. A graph $\mathcal{G}$ is *strongly connected* if each node in $\mathcal{V}$ is reachable from every other node. A set of nodes $\mathcal{D}$ is a *strongly connected component* (SCC) of $\mathcal{G}$ if all nodes in $\mathcal{D}$ can reach each other and $\mathcal{D}$ is not contained in another SCC, and a *bottom SCC* (BSCC) if no node in $\mathcal{V} \setminus \mathcal{D}$ can be reached from $\mathcal{D}$.

Markov Chains. A *discrete-time Markov chain* (DTMC) is a tuple $\mathcal{M} = \langle \mathcal{S}, \mathbf{P} \rangle$ where $\mathcal{S}$ is a finite set of states and $\mathbf{P} : \mathcal{S} \times \mathcal{S} \to [0, 1]$ a function such that $\mathbf{P}(s, \cdot) \in Dist(\mathcal{S})$ for all states $s \in \mathcal{S}$. The underlying graph $\mathcal{G}_{\mathcal{M}} = \langle \mathcal{S}, \mathcal{E} \rangle$ is defined by $\mathcal{E} = \{(s, t) \in \mathcal{S} \times \mathcal{S} : \mathbf{P}(s, t) > 0\}$. The transient distribution $\pi_n^{\iota} \in Dist(\mathcal{S})$ at step n is defined through the probability $\pi_n^{\iota}(s)$ to be in state s after n steps if starting with initial state distribution ι. It satisfies (in matrix-vector notation) $\pi_n^{\iota} = \iota \cdot \mathbf{P}^n$. We are also interested in the long-run frequency of state occupancies when n tends to infinity, defined as the Cesàro limit $\mathrm{lrf}^{\iota} : \mathcal{S} \to [0, 1]$:

$$\mathrm{lrf}^{\iota}(s) \; := \; \lim_{n \to \infty} \frac{1}{n+1} \sum_{i=0}^{n} \pi_n^{\iota}(s). \tag{LRF}$$

This limit always exists and corresponds to the long-run fraction of time spent in each state [12]. The limit probability $\lim_{n \to \infty} \pi_n^{\iota}$ is arguably more intuitive as a measure of the long-run behavior, but may not exist (due to periodicity). In case of existence, it agrees with the Cesàro limit lrf^{ι}. If $\mathcal{G}_{\mathcal{M}}$ forms an SCC, the limit is independent of the choice of ι and the superscript can be dropped. We denote this limit by $\mathrm{lrf}_{\mathcal{M}}$.

2   Generalized Bayesian Networks

We introduce *generalized Bayesian networks* (GBNs) as a BN model that does not impose acyclicity and comes with a distribution over initial nodes.

Definition 1 (Generalized BN). *A GBN $\mathcal{B}$ is a tuple $\langle \mathcal{G}, \mathcal{P}, \iota \rangle$ where*

- *$\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ is a directed graph with nodes $\mathcal{V}$ and an edge relation $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$,*
- *$\mathcal{P}$ is a function that maps all non-initial nodes $X \in \mathcal{V} \backslash Init(\mathcal{G})$ paired with each of their parent assignments $b \in Asg(Pre(X))$ to a distribution*

$$\mathcal{P}(X, b) \colon Asg(\{X\}) \to [0, 1],$$

- *ι is a distribution over the assignments for the initial nodes $Init(\mathcal{G})$, i.e., $\iota \in Dist\big(Asg(Init(\mathcal{G}))\big)$.*

The distributions $\mathcal{P}(X, b)$ have the same role as the entries in a *conditional probability table* (CPT) for X in classical BNs: they specify the probability for $X{=}\mathsf{T}$ or $X{=}\mathsf{F}$ depending on the assignments of the predecessors of X. To this end, for $X \in \mathcal{V} \backslash Init(\mathcal{G})$ and $b \in Asg(Pre(X))$, we also write $\Pr(X{=}\mathsf{T} \mid b)$ for $\mathcal{P}(X, b)(X{=}\mathsf{T})$. In the literature, initial nodes are often assigned a marginal probability via a CPT as well, assuming independence of all initial nodes. Differently, in our definition of GBNs, it is possible to specify an arbitrary distribution ι over all initial nodes. If needed, $\mathcal{P}$ can be easily extended to initial nodes by setting $\mathcal{P}(X, \varnothing) := \iota|_{\{X\}}$ for all $X \in Init(\mathcal{G})$. Hence, classical BNs arise as a special instance of GBNs where the graph $\mathcal{G}$ is acyclic and initial nodes are pairwise independent. In that case, the CPTs given by $\mathcal{P}$ are a compact representation of a single unique full joint distribution $dist_{\mathrm{BN}}(\mathcal{B})$ over all random variables $X \in \mathcal{V}$. For every assignment $b \in Asg(\mathcal{V})$, we can compute $dist_{\mathrm{BN}}(\mathcal{B})(b)$ by the so-called *chain rule*:

$$dist_{\mathrm{BN}}(\mathcal{B})(b) := \iota\big(b_{Init(\mathcal{G})}\big) \cdot \prod_{X \in \mathcal{V} \backslash Init(\mathcal{G})} \Pr\big(b_X \mid b_{Pre(X)}\big). \tag{CR}$$

In light of the semantics introduced later on, we define the *standard BN-semantics* of an acyclic GBN $\mathcal{B}$ as the set $[\![\mathcal{B}]\!]_{\mathrm{BN}} := \{dist_{\mathrm{BN}}(\mathcal{B})\}$, and $[\![\mathcal{B}]\!]_{\mathrm{BN}} := \varnothing$ if $\mathcal{B}$ contains cycles.

The distribution $dist_{\mathrm{BN}}(\mathcal{B})$ satisfies two crucial properties: First, it is consistent with the CPT entries given by $\mathcal{P}$ and the distribution ι, and second, it observes the independencies encoded in the graph $\mathcal{G}$. In fact, those two properties are sufficient to uniquely characterize $dist_{\mathrm{BN}}(\mathcal{B})$. We briefly review the notion of independence and formally define CPT consistency later on in Section 3.

Independence. Any full joint probability distribution $\mu \in Dist(Asg(\mathcal{V}))$ may induce a number of conditional independencies among the random variables in $\mathcal{V}$. For $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ disjoint subsets of $\mathcal{V}$, the random variables in $\mathcal{X}$ and $\mathcal{Y}$ are independent under μ given $\mathcal{Z}$ if the conditional probability of each assignment over the nodes in $\mathcal{X}$ given an assignment for $\mathcal{Z}$ is unaffected by further conditioning on any assignment of $\mathcal{Y}$. Formally, the set $Indep(\mu)$ contains the triple $(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$ iff for all $a \in Asg(\mathcal{X})$, $b \in Asg(\mathcal{Y})$, and $c \in Asg(\mathcal{Z})$, we have

$$\mu(a \mid b, c) = \mu(a \mid c) \quad \text{or} \quad \mu(b, c) = 0.$$

We also write $(\mathcal{X} \perp \mathcal{Y} \mid \mathcal{Z})$ for $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}) \in Indep(\mu)$ and may omit the set brackets of $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$.

d-separation. For classical BNs, the graph topology encodes independencies that are necessarily satisfied by any full joint distribution regardless of the CPT entries. Given two random variables X and Y as well as a set of observed variables $\mathcal{Z}$, then X and Y are conditionally independent given $\mathcal{Z}$ if the corresponding nodes in the graph are *d-separated* given $\mathcal{Z}$ [6]. To establish *d*-separation, all simple undirected paths[5] between X and Y need to be *blocked* given $\mathcal{Z}$. Let $\mathbb{W}$ denote such a simple path $W_0, W_1, \ldots, W_k$ with $W_0 = X$, $W_k = Y$, and either $W_i \to W_{i+1}$ or $W_i \leftarrow W_{i+1}$ for all $i < k$. Then $\mathbb{W}$ is blocked given $\mathcal{Z}$ if and only if there exists an index i, $0 < i < k$, such that one of the following two conditions holds: (1) W_i is in $\mathcal{Z}$ and is situated in a *chain* or a *fork* in $\mathbb{W}$, i.e.,

- $W_{i-1} \to W_i \to W_{i+1}$ (forward chain)
- $W_{i-1} \leftarrow W_i \leftarrow W_{i+1}$ (backward chain)   and $W_i \in \mathcal{Z}$,
- $W_{i-1} \leftarrow W_i \to W_{i+1}$ (fork)

(2) W_i is in a *collider* and neither W_i nor any descendant of W_i is in $\mathcal{Z}$, i.e.,

- $W_{i-1} \to W_i \leftarrow W_{i+1}$ (collider)   and $Post^*(W_i) \cap \mathcal{Z} = \varnothing$.

Two sets of nodes $\mathcal{X}$ and $\mathcal{Y}$ are *d*-separated given a third set $\mathcal{Z}$ if for each $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$, X and Y are *d*-separated given $\mathcal{Z}$. Notably, the *d*-separation criterion is applicable also in presence of cycles [28]. For a graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ of a GBN, we define the set $d\text{-}sep(\mathcal{G})$ as

$$d\text{-}sep(\mathcal{G}) := \big\{ (\mathcal{X}, \mathcal{Y}, \mathcal{Z}) \in (2^{\mathcal{V}})^3 : \mathcal{X} \text{ and } \mathcal{Y} \text{ are } d\text{-separated given } \mathcal{Z} \big\}.$$

For acyclic Bayesian networks it is well known that the independencies evident from the standard BN semantics' distribution include the independencies derived from the graph. That is, for acyclic GBNs $\mathcal{B}_{\varnothing} = \langle \mathcal{G}, \mathcal{P}, \iota \rangle$ where all initial nodes are pairwise independent under ι, we have

$$d\text{-}sep(\mathcal{G}) \ \subseteq \ Indep\big(dist_{\mathrm{BN}}(\mathcal{B}_{\varnothing})\big).$$

For an arbitrary initial distribution, the above relation does not necessarily hold. However, we can still find a set of independencies that are necessarily observed by the standard BN semantics and thus act as a similar lower bound. We do so by assuming the worst case, namely that each initial node depends on every other initial node under ι. Formally, given a graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$, we define a closure operation $Close(\cdot)$ as follows and compute the set $d\text{-}sep\big(Close(\mathcal{G})\big)$:

$$Close(\mathcal{G}) := \big\langle \mathcal{V}, \mathcal{E} \cup \{(A, B) \text{ for } A, B \in Init(\mathcal{G}), A \neq B\} \big\rangle.$$

Lemma 1. *Let $\mathcal{B}_{\varnothing} = \langle \mathcal{G}, \mathcal{P}, \iota \rangle$ be an acyclic GBN. Then*

$$d\text{-}sep\big(Close(\mathcal{G})\big) \ \subseteq \ Indep\big(dist_{\mathrm{BN}}(\mathcal{B}_{\varnothing})\big).$$

As intuitively expected, the presence of cycles in $\mathcal{G}$ generally reduces the number of graph independencies, though note that also in strongly connected graphs independencies may exist. For example, if $\mathcal{G}$ is a four-node cycle with nodes W, X, Y, and Z, then $d\text{-}sep(\mathcal{G}) = \big\{ (W \perp Y \mid X, Z), (X \perp Z \mid W, Y) \big\}$.

---

[5] A path is simple if no node occurs twice in the path. "Undirected" in this context means that edges in either direction can occur along the path.

3 Constraints Semantics

For classical acyclic BNs there is exactly one distribution that agrees with all CPTs and satisfies the independencies encoded in the graph. This distribution can easily be constructed by means of the chain rule (CR). For cyclic GBNs, applying the chain rule towards a full joint distribution is not possible in general, as the result is usually not a valid probability distribution. Still, we can look for distributions consistent with a GBN's CPTs and the independencies derived from its graph. Depending on the GBN, we will see that there may be none, exactly one, or even infinitely many distributions fulfilling these constraints.

3.1 CPT-consistency

We first provide a formal definition of CPT consistency in terms of linear constraints on full joint distributions.

Definition 2 (Strong and weak CPT-consistency). *Let $\mathcal{B}$ be a GBN with nodes $\mathcal{V}$ and $X \in \mathcal{V}$. Then μ is called* strongly CPT-consistent *for X in $\mathcal{B}$ (or simply* CPT-consistent*) if for all $c \in Asg(Pre(X))$*

$$\mu(X{=}\top, c) \;=\; \mu(c) \cdot \Pr(X{=}\top \mid c). \tag{CPT}$$

We say that μ is weakly CPT-consistent *for X in $\mathcal{B}$ if*

$$\mu(X{=}\top) \;=\; \sum_{c \in Asg(Pre(X))} \mu(c) \cdot \Pr(X{=}\top \mid c). \tag{wCPT}$$

Intuitively, the constraint (CPT) is satisfied for μ if the conditional probability $\mu(X{=}\top \mid c)$ equals the entry in the CPT for X under assignment c, i.e., $\mu(X{=}\top \mid c) = \Pr(X{=}\top \mid c)$. In the weak case (wCPT), only the resulting marginal probability of X needs to agree with the CPTs.

Definition 3 (CPT and wCPT semantics). *For a GBN $\mathcal{B} = \langle \mathcal{G}, \mathcal{P}, \iota \rangle$, the CPT-semantics $[\![\mathcal{B}]\!]_{\mathrm{CPT}}$ is the set of all distributions $\mu \in Dist(Asg(\mathcal{V}))$ where $\mu|_{Init(\mathcal{G})} = \iota$ and μ is CPT-consistent for every node $X \in \mathcal{V} \backslash Init(\mathcal{G})$. The weak CPT-semantics $[\![\mathcal{B}]\!]_{\mathrm{wCPT}}$ is defined analogously.*

Clearly, we have $[\![\mathcal{B}]\!]_{\mathrm{CPT}} \subseteq [\![\mathcal{B}]\!]_{\mathrm{wCPT}}$ for all $\mathcal{B}$. The next example shows that depending on the CPT values, the set $[\![\mathcal{B}]\!]_{\mathrm{CPT}}$ may be empty, a singleton, or of infinite cardinality.

Example 1. To find CPT-consistent distributions for the GBN from Fig. 1, we construct a system of linear equations whose solutions form distributions $\mu \in Dist\big(Asg(\{X, Y\})\big)$, represented as vectors in the space $[0, 1]^4$:

$$\begin{pmatrix} s_1 & 0 & s_1{-}1 & 0 \\ 0 & s_2 & 0 & s_2{-}1 \\ t_1 & t_1{-}1 & 0 & 0 \\ 0 & 0 & t_2 & t_2{-}1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} \mu_{\overline{XY}} \\ \mu_{\overline{X}Y} \\ \mu_{X\overline{Y}} \\ \mu_{XY} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

where, e.g., $\mu_{X\overline{Y}}$ abbreviates $\mu(X{=}\mathsf{T}, Y{=}\mathsf{F})$. The first line of the matrix states the (CPT) constraint for node X and the parent assignment $c = \{Y{=}\mathsf{F}\}$:

$$
\begin{aligned}
0 &= s_1 \cdot \mu_{\overline{XY}} + 0 \cdot \mu_{\overline{X}Y} + (s_1{-}1) \cdot \mu_{X\overline{Y}} + 0 \cdot \mu_{XY} \\
\mu_{X\overline{Y}} &= (\mu_{X\overline{Y}} + \mu_{\overline{XY}}) \cdot s_1 \\
\mu_{X\overline{Y}} &= \mu_{\overline{Y}} \cdot \Pr(X{=}\mathsf{T} \mid Y{=}\mathsf{F}) \\
\mu(X{=}\mathsf{T}, c) &= \mu(c) \cdot \Pr(X{=}\mathsf{T} \mid c).
\end{aligned}
$$

Analogously, the following three rows encode the CPT constraints for X, Y, and their remaining parent assignments. The last row ensures that solutions are indeed probability distributions satisfying $\sum_c \mu(c) = 1$.

The number of solutions for the system now depends on the CPT entries s_1, s_2, t_1, and t_2. For $s_1 = t_2 = 0$ and $s_2 = t_1 = 1$, no solution exists as the first four equations require $\mu(b) = 0$ for all $b \in Asg(\{X, Y\})$, while the last equation ensures $\mu_{\overline{XY}} + \mu_{\overline{X}Y} + \mu_{X\overline{Y}} + \mu_{XY} = 1$. For $s_1 = t_1 = 0$ and $s_2 = t_2 = 1$, all distributions with $\mu_{XY} = 1 - \mu_{\overline{XY}}$ and $\mu_{X\overline{Y}} = \mu_{\overline{X}Y} = 0$ are solutions. Finally, e.g., for $s_1 = t_1 = 3/4$ and $s_2 = t_2 = 1/2$, there is exactly one solution with $\mu_{\overline{XY}} = 1/10$ and $\mu(b) = 3/10$ for the other three assignments.

3.2   Independence-consistency

We extend CPT semantics with a set of independencies that need to be observed by all induced distributions.

Definition 4 (CPT-$\mathcal{I}$ semantics). *For a GBN $\mathcal{B} = \langle \mathcal{G}, \mathcal{P}, \iota \rangle$ and a set of independencies $\mathcal{I}$, the* CPT-$\mathcal{I}$ *semantics $[\![\mathcal{B}]\!]_{\mathrm{CPT}\text{-}\mathcal{I}}$ is defined as the set of all CPT-consistent distributions μ for which $\mathcal{I} \subseteq Indep(\mu)$ holds.*

Technically, the distributions in $[\![\mathcal{B}]\!]_{\mathrm{CPT}\text{-}\mathcal{I}}$ have to fulfill the following polynomial constraints in addition to the CPT-consistency constraints:

$$
\mu(b) \cdot \mu(b_{\mathcal{W}}) = \mu(b_{\{X\} \cup \mathcal{W}}) \cdot \mu(b_{\mathcal{U} \cup \mathcal{W}}) \tag{CPT-$\mathcal{I}$}
$$

for each independence $(X \perp\!\!\!\perp \mathcal{U} \mid \mathcal{W}) \in \mathcal{I}$ with variable $X \in \mathcal{V}$ and sets of variables $\mathcal{U}, \mathcal{W} \subseteq \mathcal{V}$, and for each assignment $b \in Asg(\{X\} \cup \mathcal{U} \cup \mathcal{W})$. Note that in case $\mu(b_{\mathcal{W}}) > 0$, (CPT-$\mathcal{I}$) is equivalent to the constraint $\mu(b_X \mid b_{\mathcal{U} \cup \mathcal{W}}) = \mu(b_X \mid b_{\mathcal{W}})$.

We can now formally state the alternative characterization of the standard BN semantics as the unique CPT-consistent distribution that satisfies the d-separation independencies of the graph. For each classical BN $\mathcal{B}$ with acyclic graph $\mathcal{G}$ and $\mathcal{I} = d\text{-}sep(\mathcal{G})$, we have $[\![\mathcal{B}]\!]_{\mathrm{BN}} = \{dist_{\mathrm{BN}}(\mathcal{B})\} = [\![\mathcal{B}]\!]_{\mathrm{CPT}\text{-}\mathcal{I}}$. Thus, the CPT-$\mathcal{I}$ semantics provides a conservative extension of the standard BN semantics to GBNs with cycles. However, in practice, its use is limited since there might be no distribution that satisfies all constraints. In fact, the case where $[\![\mathcal{B}]\!]_{\mathrm{CPT}\text{-}\mathcal{I}} = \varnothing$ is to be expected for most cyclic GBNs, given that the resulting constraint systems tend to be heavily over-determined.

The next section introduces semantics that follow a more constructive approach. We will see later on in Section 5.1 that the families of distributions induced by these semantics are always non-empty and usually singletons.
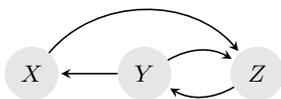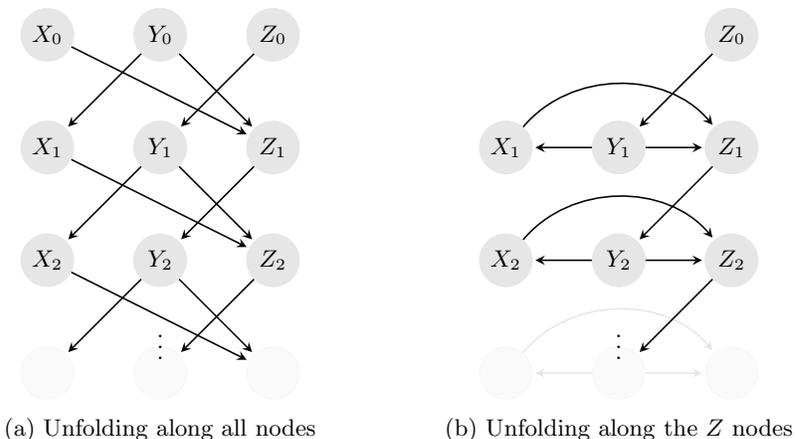
Fig. 2: The graph of a strongly connected GBN

(a) Unfolding along all nodes   (b) Unfolding along the Z nodes

Fig. 3: Two infinite unfoldings of the graph in Fig. 2

4   Limit and Limit Average Semantics

We first develop the basic ideas underling the semantics by following an example, before giving a formal treatment in Section 4.2.

4.1   Intuition

Consider the GBN $\mathcal{B}$ whose graph $\mathcal{G}$ is depicted in Fig. 2. One way to get rid of the cycles is to construct an infinite unfolding of $\mathcal{B}$ as shown in Fig. 3a. In this new graph $\mathcal{G}_\infty$, each level contains a full copy of the original nodes and corresponds to some $n \in \mathbb{N}$. For any edge $X \to Y$ in the original graph, we add edges $X_n \to Y_{n+1}$ to $\mathcal{G}_\infty$, such that each edge descends one level deeper. Clearly any graph constructed in this way is acyclic, but this fact alone does not aid in finding a matching distribution since we dearly bought it by giving up finiteness. However, we can consider what happens when we plug in some initial distribution μ_0 over the nodes X_0, Y_0, and Z_0. Looking only at the first two levels, we then get a fully specified acyclic BN by using the CPTs given by $\mathcal{P}$ for X_1, Y_1, and Z_1. For this sub-BN, the standard BN semantics yields a full joint distribution over the six nodes from X_0 to Z_1, which also induces a distribution μ_1 over the three nodes at level 1. This procedure can then be repeated to construct a distribution μ_2 over the nodes X_2, Y_2, and Z_2, and, more generally, to get a distribution μ_{n+1} given a distribution μ_n. Recall that

each of those distributions can be viewed as vector of size 2^3. Considering the sequence $\mu_0, \mu_1, \mu_2, \ldots$, the question naturally arises whether a limit exists, i.e., a distribution/vector μ such that

$$\mu \;=\; \lim_{n \to \infty} \mu_n.$$

Example 2. Consider the GBN from Fig. 1 with CPT entries $s_1 = t_2 = 1$ and $s_2 = t_1 = 0$, which intuitively describe the contradictory dependencies "X iff not Y" and "Y iff X". For any initial distribution $\mu_0 = \langle e\ f\ g\ h \rangle$, the construction informally described above yields the following sequence of distributions μ_n:

$$\mu_0 = \begin{pmatrix} e \\ f \\ g \\ h \end{pmatrix}, \ \mu_1 = \begin{pmatrix} f \\ h \\ e \\ g \end{pmatrix}, \ \mu_2 = \begin{pmatrix} h \\ g \\ f \\ e \end{pmatrix}, \ \mu_3 = \begin{pmatrix} g \\ e \\ h \\ f \end{pmatrix}, \ \mu_4 = \begin{pmatrix} e \\ f \\ g \\ h \end{pmatrix}, \ \ldots$$

As $\mu_4 = \mu_0$, the sequence starts to cycle infinitely between the first four distributions. The series converges for $e = f = g = h = 1/4$ (in which case the sequence is constant), but does not converge for any other initial distribution.

The example shows that the existence of the limit depends on the given initial distribution. In case no limit exists because some distributions keep repeating without ever converging, it is possible to determine the *limit average* (or *Cesàro limit*) of the sequence:

$$\tilde{\mu} \;=\; \lim_{n \to \infty} \frac{1}{n+1} \sum_{i=0}^{n} \mu_i.$$

The limit average has three nice properties: First, if the regular limit μ exists, then the limit average $\tilde{\mu}$ exists as well and is identical to μ. Second, in our use case, $\tilde{\mu}$ in fact always exists for any initial distribution μ_0. And third, as we will see in Section 5, the limit average corresponds to the long-run frequency of certain Markov chains, which allows us both to explicitly compute and to derive important properties of the limit distributions.

Example 3. Continuing Ex. 2, the limit average of the sequence $\mu_0, \mu_1, \mu_2, \ldots$ is the uniform distribution $\tilde{\mu} = \langle 1/4\ 1/4\ 1/4\ 1/4 \rangle$, regardless of the choice of μ_0.

Before we formally define the infinite unfolding of GBNs and the resulting limit semantics, there is one more observation to be made. To ensure that the unfolded graph $\mathcal{G}_\infty$ is acyclic, we redirected every edge of the GBN $\mathcal{B}$ to point one level deeper, resulting in the graph displayed in Fig. 3a. As can be seen in Fig. 3b, we also get an acyclic unfolded graph by only redirecting the edges originating in the Z nodes to the next level and keeping all other edges on the same level. The relevant property is to pick a set of nodes such that for each cycle in the original GBN $\mathcal{B}$, at least one node in the cycle is contained in the set. We call such sets the *cutsets* of $\mathcal{B}$.

Definition 5 (Cutset). *Let $\mathcal{B}$ be an GBN with graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$. A subset $\mathcal{C} \subseteq \mathcal{V}$ is a* cutset *for $\mathcal{B}$ if every cycle in $\mathcal{G}$ contains at least one node from $\mathcal{C}$.*

Example 4. The GBN in Fig. 2 has the following cutsets: $\{Y\}$, $\{Z\}$, $\{X, Y\}$, $\{X, Z\}$, $\{Y, Z\}$, and $\{X, Y, Z\}$. Note that $\{X\}$ does not form a cutset as no node from the cycle $Y \to Z \to Y$ is contained.

So far we implicitly used the set $\mathcal{V}$ of all nodes for the unfolding, which always trivially forms a cutset. The following definitions will be parameterized with a cutset, as the choice of cutsets influences the resulting distributions as well as the time complexity.

4.2 Formal Definition

Let $\mathcal{V}_n := \{X_n : X \in \mathcal{V}\}$ denote the set of nodes on the nth level of the unfolding in $\mathcal{G}_\infty$. For $\mathcal{C} \subseteq \mathcal{V}$ a cutset of the GBN, the subset of cutset nodes on that level is given by $\mathcal{C}_n := \{X_n \in \mathcal{V}_n : X \in \mathcal{C}\}$. Then a distribution $\gamma_n \in Dist(Asg(\mathcal{C}_n))$ for the cutset nodes in $\mathcal{C}_n$ suffices to get a full distribution $\mu_{n+1} \in Dist(Asg(\mathcal{V}_{n+1}))$ over all nodes on the next level, $n + 1$: We look at the graph fragment $\mathcal{G}_{n+1}$ of $\mathcal{G}_\infty$ given by the nodes $\mathcal{C}_n \cup \mathcal{V}_{n+1}$ and their respective edges. In this fragment, the cutset nodes are initial, so the cutset distribution γ_n can be combined with the initial distribution ι to act as new initial distribution. For the nodes in $\mathcal{V}_{n+1}$, the corresponding CPTs as given by $\mathcal{P}$ can be used, i.e., $\mathcal{P}_n(X_n, \cdot) = \mathcal{P}(X, \cdot)$ for $X_n \in \mathcal{V}_n$. Putting everything together, we obtain an acyclic GBN $\mathcal{B}_{n+1} = \langle \mathcal{G}_{n+1}, \mathcal{P}_{n+1}, \iota \otimes \gamma_n \rangle$. However, GBNs constructed in this way for each level $n > 0$ are all isomorphic and only differ in the given cutset distribution γ. For simplicity and in light of later use, we thus define a single representative GBN $Dissect(\mathcal{B}, \mathcal{C}, \gamma)$ that represents a dissection of $\mathcal{B}$ along a given cutset $\mathcal{C}$, with $\iota \otimes \gamma$ as initial distribution.

Definition 6 (Dissected GBN). *Let $\mathcal{B} = \langle \mathcal{G}, \mathcal{P}, \iota \rangle$ be a GBN with graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ and $\mathcal{C} \subseteq \mathcal{V}$ a cutset for $\mathcal{B}$ with distribution $\gamma \in Dist(Asg(\mathcal{C}))$. Then, the $\mathcal{C}$-dissected GBN $Dissect(\mathcal{B}, \mathcal{C}, \gamma)$ is the acyclic GBN $\langle \mathcal{G}_\mathcal{C}, \mathcal{P}_\mathcal{C}, \iota \otimes \gamma \rangle$ with graph $\mathcal{G}_\mathcal{C} = \langle \mathcal{V} \cup \mathcal{C}', \mathcal{E}_\mathcal{C} \rangle$ where*

- $\mathcal{C}' := \{X' : X \in \mathcal{C}\}$ *extends $\mathcal{V}$ by fresh copies of all cutset nodes;*
- *incoming edges to nodes in $\mathcal{C}$ are redirected to their copies, i.e.,*

$$\mathcal{E}_\mathcal{C} := \big\{(X, Y') : (X, Y) \in \mathcal{E}, Y \in \mathcal{C}\big\} \cup \big\{(X, Y) : (X, Y) \in \mathcal{E}, Y \notin \mathcal{C}\big\};$$

- *the function $\mathcal{P}_\mathcal{C}$ uses the CPT entries given by $\mathcal{P}$ for the cutset nodes as entries for their copies and the original entries for all other nodes, i.e., we have $\mathcal{P}_\mathcal{C}(Y', a) = \mathcal{P}(Y, a)$ for each node $Y' \in \mathcal{C}'$ and parent assignment $a \in Asg(Pre(Y'))$, and $\mathcal{P}_\mathcal{C}(X, b) = \mathcal{P}(X, b)$ for $X \in \mathcal{V} \backslash \mathcal{C}$ and $b \in Asg(Pre(X))$.*

Fig. 4 shows two examples of dissections on the GBN of Fig. 2. As any dissected GBN is acyclic by construction, the standard BN semantics yields a full joint distribution over all nodes in $\mathcal{V} \cup \mathcal{C}'$. We restrict this distribution to the nodes
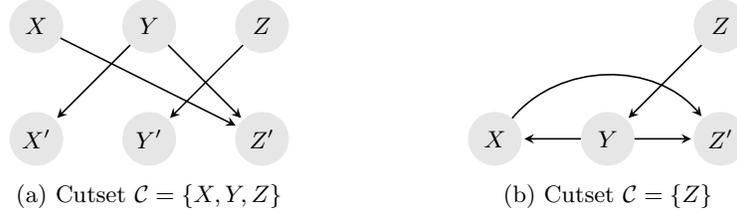
(a) Cutset $\mathcal{C} = \{X, Y, Z\}$ (b) Cutset $\mathcal{C} = \{Z\}$

Fig. 4: Dissections of the GBN in Fig. 2 for two cutsets

in $(\mathcal{V} \setminus \mathcal{C}) \cup \mathcal{C}'$, as those are the ones on the "next level" of the unfolding, while re-identifying the cutset node copies with the original nodes to get a distribution over $\mathcal{V}$. Formally, we define the distribution $Next(\mathcal{B}, \mathcal{C}, \gamma)$ for each assignment $b \in Asg(\mathcal{V})$ as

$$Next(\mathcal{B}, \mathcal{C}, \gamma)(b) \;:=\; dist_{\mathrm{BN}}\big(Dissect(\mathcal{B}, \mathcal{C}, \gamma)\big)(b')$$

where the assignment $b' \in Asg\big((\mathcal{V} \setminus \mathcal{C}) \cup \mathcal{C}'\big)$ is given by $b'(X) = b(X)$ for all $X \in \mathcal{V} \setminus \mathcal{C}$ and $b'(Y') = b(Y)$ for all $Y \in \mathcal{C}$. In the unfolded GBN, this allows us to get from a cutset distribution γ_n to the next level distribution $\mu_{n+1} = Next(\mathcal{B}, \mathcal{C}, \gamma_n)$. The next cutset distribution γ_{n+1} is then given by restricting the full distribution to the nodes in $\mathcal{C}$, i.e., $\gamma_{n+1} = Next(\mathcal{B}, \mathcal{C}, \gamma_n)|_{\mathcal{C}}$.[6] Vice versa, a cutset distribution γ suffices to recover the full joint distribution over all nodes $\mathcal{V}$. Again using the standard BN semantics of the dissected GBN, we define the distribution $Extend(\mathcal{B}, \mathcal{C}, \gamma) \in Dist(Asg(\mathcal{V}))$ as

$$Extend(\mathcal{B}, \mathcal{C}, \gamma) \;:=\; dist_{\mathrm{BN}}\big(Dissect(\mathcal{B}, \mathcal{C}, \gamma)\big)\big|_{\mathcal{V}}.$$

With these definitions at hand, we can formally define the limit and limit average semantics described in the previous section.

Definition 7 (Limit and limit average semantics). *Let $\mathcal{B}$ be a GBN over nodes $\mathcal{V}$ with cutset $\mathcal{C}$. The* limit semantics *of $\mathcal{B}$ w.r.t. $\mathcal{C}$ is the partial function*

$$Lim(\mathcal{B}, \mathcal{C}, \cdot) \;:\; Dist\big(Asg(\mathcal{C})\big) \rightharpoonup Dist\big(Asg(\mathcal{V})\big)$$

from initial cutset distributions γ_0 to full distributions $\mu = Extend(\mathcal{B}, \mathcal{C}, \gamma)$ where

$$\gamma \;=\; \lim_{n \to \infty} \gamma_n \qquad and \qquad \gamma_{n+1} = Next(\mathcal{B}, \mathcal{C}, \gamma_n)|_{\mathcal{C}}.$$

The set $[\![\mathcal{B}]\!]_{\mathrm{LIM}\text{-}\mathcal{C}}$ is given by the image of $Lim(\mathcal{B}, \mathcal{C}, \cdot)$, i.e.,

$$[\![\mathcal{B}]\!]_{\mathrm{LIM}\text{-}\mathcal{C}} := \{Lim(\mathcal{B}, \mathcal{C}, \gamma_0) : \gamma_0 \in Dist(Asg(\mathcal{C})) \text{ s.t. } Lim(\mathcal{B}, \mathcal{C}, \gamma_0) \text{ is defined}\}.$$

The limit average semantics *of $\mathcal{B}$ w.r.t. $\mathcal{C}$ is the partial function*

$$LimAvg(\mathcal{B}, \mathcal{C}, \cdot) \;:\; Dist\big(Asg(\mathcal{C})\big) \rightharpoonup Dist\big(Asg(\mathcal{V})\big)$$

---

[6] Recall that we may view distributions as vectors which allows us to equate distributions over different but isomorphic domains.
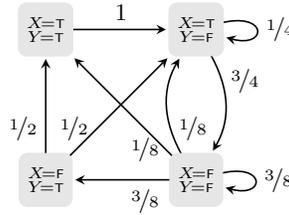
Fig. 5: A cutset Markov chain for a cutset $\mathcal{C} = \{X, Y\}$

from γ_0 to distributions $\mu = Extend(\mathcal{B}, \mathcal{C}, \gamma)$ where

$$\gamma \;=\; \lim_{n \to \infty} \frac{1}{n+1} \sum_{i=0}^{n} \gamma_n \qquad and \qquad \gamma_{n+1} = Next(\mathcal{B}, \mathcal{C}, \gamma_n)|_{\mathcal{C}}.$$

The set $[\![\mathcal{B}]\!]_{\textsc{LimAvg-}\mathcal{C}}$ is likewise given by the image of $LimAvg(\mathcal{B}, \mathcal{C}, \cdot)$.

We know that the limit average coincides with the regular limit if the latter exists, so for every initial cutset distribution γ_0, we have $Lim(\mathcal{B}, \mathcal{C}, \gamma_0) = LimAvg(\mathcal{B}, \mathcal{C}, \gamma_0)$ if $Lim(\mathcal{B}, \mathcal{C}, \gamma_0)$ is defined. Thus, $[\![\mathcal{B}]\!]_{\textsc{Lim-}\mathcal{C}} \subseteq [\![\mathcal{B}]\!]_{\textsc{LimAvg-}\mathcal{C}}$.

5 Markov Chain Semantics

While we gave some motivation for the limit and limit average semantics, their definitions do not reveal an explicit way to compute their member distributions. In this section we introduce the *(cutset) Markov chain semantics* which offers explicit construction of distributions and is shown to coincide with the limit average semantics. It further paves the way for proving several properties of both limit semantics in Section 5.1.

At the core of the cutset Markov chain semantics lies the eponymous *cutset Markov chain* which captures how probability mass flows from one cutset assignment to the others. To this end, the Dirac distributions corresponding to each assignment are used as initial distributions in the dissected GBN. With the *Next* function we then get a new distribution over all cutset assignments, and the probabilities assigned by this distribution are used as transition probabilities for the Markov chain.

Definition 8 (Cutset Markov chain). *Let $\mathcal{B}$ be a GBN with cutset $\mathcal{C}$. The cutset Markov chain $CMC(\mathcal{B}, \mathcal{C}) = \langle Asg(\mathcal{C}), \mathbf{P} \rangle$ w.r.t. $\mathcal{B}$ and $\mathcal{C}$ is a DTMC where the transition matrix $\mathbf{P}$ is given for all cutset assignments $b, c \in Asg(\mathcal{C})$ by*

$$\mathbf{P}(b, c) \;:=\; Next\big(\mathcal{B}, \mathcal{C}, Dirac(b)\big)(c).$$

Example 5. Fig. 5 shows the cutset Markov chain for the GBN from Fig. 1 with CPT entries $s_1 = {}^1\!/_4$, $s_2 = 1$, $t_1 = {}^1\!/_2$, $t_2 = 0$, and cutset $\mathcal{C} = \{X, Y\}$. Exemplarily, the edge at the bottom from assignment $b = \{X{=}\mathsf{F}, Y{=}\mathsf{F}\}$ to assignment

$c = \{X{=}\mathsf{F}, Y{=}\mathsf{T}\}$ with label $3/8$ is derived as follows:

$$
\begin{aligned}
\mathbf{P}(b,c) &= Next\big(\mathcal{B},\mathcal{C},Dirac(b)\big)(c) = dist_{\mathrm{BN}}\big(Dissect(\mathcal{B},\mathcal{C},Dirac(b))\big)(c') \\
&= \sum_{\substack{a \in Asg(\mathcal{V_C}) \\ \text{s.t. } c' \subseteq a}} dist_{\mathrm{BN}}\big(Dissect(\mathcal{B},\mathcal{C},Dirac(b))\big)(a) \\
&= \sum_{\substack{a \in Asg(\mathcal{V_C}) \\ \text{s.t. } c' \subseteq a}} Dirac(b)(a_{X,Y}) \cdot \Pr(X'{=}\mathsf{F} \mid a_Y) \cdot \Pr(Y'{=}\mathsf{T} \mid a_X) \\
&= \Pr(X'{=}\mathsf{F} \mid Y{=}\mathsf{F}) \cdot \Pr(Y'{=}\mathsf{T} \mid X{=}\mathsf{F}) = (1-s_1)\cdot t_1 = 3/8.
\end{aligned}
$$

Note that in the second-to-last step, in the sum over all full assignments a which agree with the partial assignment c', only the assignment which also agrees with b remains as for all other assignments we have $Dirac(b)(a_{X,Y}) = 0$.

Given a cutset Markov chain with transition matrix $\mathbf{P}$ and an initial cutset distribution γ_0, we can compute the uniquely defined long-run frequency distribution lrf^{γ_0} (see Section 1.1). Then the Markov chain semantics is given by the extension of this distribution over the whole GBN.

Definition 9 (Markov chain semantics). *Let $\mathcal{B}$ be a GBN over nodes $\mathcal{V}$ with a cutset $\mathcal{C} \subseteq \mathcal{V}$ and cutset Markov chain $CMC(\mathcal{B},\mathcal{C}) = \langle Asg(\mathcal{C}), \mathbf{P} \rangle$. Then the* Markov chain semantics *of $\mathcal{B}$ w.r.t. $\mathcal{C}$ is the function*

$$
MCS(\mathcal{B},\mathcal{C},\cdot) : \ Dist\big(Asg(\mathcal{C})\big) \to Dist\big(Asg(\mathcal{V})\big)
$$

from cutset distributions γ_0 to full distributions $\mu = Extend(\mathcal{B},\mathcal{C},\mathrm{lrf}^{\gamma_0})$ where

$$
\mathrm{lrf}^{\gamma_0} = \lim_{n \to \infty} \frac{1}{n+1} \sum_{i=0}^{n} \gamma_i \qquad and \qquad \gamma_{i+1} = \gamma_i \cdot \mathbf{P}.
$$

The set $[\![\mathcal{B}]\!]_{\mathrm{MC}\text{-}\mathcal{C}}$ is defined as the image of $MCS(\mathcal{B},\mathcal{C},\cdot)$.

In the following lemma, we give four equivalent characterizations of the long-run frequency distributions of the cutset Markov chain.

Lemma 2. *Let $\mathcal{B}$ be a GBN with cutset $\mathcal{C}$, cutset distribution $\gamma \in Dist(Asg(\mathcal{C}))$, and $\mathcal{M} = \langle Asg(\mathcal{C}), \mathbf{P} \rangle$ the cutset Markov chain $CMC(\mathcal{B},\mathcal{C})$. Then the following statements are equivalent:*

(a) $\gamma = \gamma \cdot \mathbf{P}$.
(b) There exists $\gamma_0 \in Dist(Asg(\mathcal{C}))$ such that for $\gamma_{i+1} = \gamma_i \cdot \mathbf{P}$, we have

$$
\gamma = \lim_{n \to \infty} \frac{1}{n+1} \sum_{i=0}^{n} \gamma_i.
$$

(c) γ belongs to the convex hull of the long-run frequency distributions $\mathrm{lrf}_{\mathcal{D}}$ of the bottom SCCs $\mathcal{D}$ of $\mathcal{M}$.
(d) $\gamma = Next(\mathcal{B},\mathcal{C},\gamma)|_{\mathcal{C}}$.

Following Lemma 2, we can equivalently define the cutset Markov chain semantics as the set of extensions of all stationary distributions for $\mathbf{P}$:

$$[\![\mathcal{B}]\!]_{\text{MC-}\mathcal{C}} := \big\{ Extend(\mathcal{B}, \mathcal{C}, \gamma) : \gamma \in Dist\big(Asg(\mathcal{C})\big) \text{ s.t. } \gamma = \gamma \cdot \mathbf{P} \big\}.$$

Example 6. Continuing Ex. 5, there is a unique stationary distribution γ with $\gamma = \gamma \cdot \mathbf{P}$ for the cutset Markov chain in Fig. 5: $\gamma = \langle ^{48}/_{121}\ ^{18}/_{121}\ ^{40}/_{121}\ ^{15}/_{121} \rangle$. As in this case the cutset $\mathcal{C} = \{X, Y\}$ equals the set of all nodes $\mathcal{V}$, we have $Extend(\mathcal{B}, \mathcal{C}, \gamma) = \gamma$ and thus $[\![\mathcal{B}]\!]_{\text{MC-}\{X,Y\}} = \{\gamma\}$.

As shown by Lemma 2, the behavior of the *Next* function is captured by multiplication with the transition matrix $\mathbf{P}$. Both the distributions in the limit average semantics and the long-run frequency distributions of the cutset Markov chain are defined in terms of a Cesàro limit, the former over the sequence of distributions obtained by repeated application of *Next*, the latter by repeated multiplication with $\mathbf{P}$. Thus, both semantics are equivalent.

Theorem 1. *Let $\mathcal{B}$ be a GBN. Then for any cutset $\mathcal{C}$ of $\mathcal{B}$ and initial distribution $\gamma_0 \in Dist(Asg(\mathcal{C}))$, we have*

$$MCS(\mathcal{B}, \mathcal{C}, \gamma_0) = LimAvg(\mathcal{B}, \mathcal{C}, \gamma_0).$$

We know that $Lim(\mathcal{B}, \mathcal{C}, \gamma_0)$ is not defined for all initial distributions γ_0. However, the set of all limits that do exist contains exactly the distributions admitted by the Markov chain and limit average semantics.

Lemma 3. *Let $\mathcal{B}$ be a GBN. Then for any cutset $\mathcal{C}$ of $\mathcal{B}$, we have*

$$[\![\mathcal{B}]\!]_{\text{MC-}\mathcal{C}} = [\![\mathcal{B}]\!]_{\text{LimAvg-}\mathcal{C}} = [\![\mathcal{B}]\!]_{\text{Lim-}\mathcal{C}}.$$

5.1 Properties

By the equivalences established in Theorem 1 and Lemma 3, we gain profound insights about the limit and limit average distributions by Markov chain analysis. As every finite-state Markov chain has at least one stationary distribution, it immediately follows that $[\![\mathcal{B}]\!]_{\text{MC-}\mathcal{C}}$—and thus $[\![\mathcal{B}]\!]_{\text{LimAvg-}\mathcal{C}}$ and $[\![\mathcal{B}]\!]_{\text{Lim-}\mathcal{C}}$—is always non-empty. Further, if the cutset Markov chain is *irreducible*, i.e., the graph is strongly connected, the stationary distribution is unique and $[\![\mathcal{B}]\!]_{\text{MC-}\mathcal{C}}$ is a singleton. The existence of the limit semantics for a given initial distribution γ_0 hinges on the *periodicity* of the cutset Markov chain.

Example 7. We return to Example 2 and construct the cutset Markov chain $CMC(\mathcal{B}, \mathcal{C}) = \langle Asg(\mathcal{C}), \mathbf{P} \rangle$ for the (implicitly used) cutset $\mathcal{C} = \{X, Y\}$:
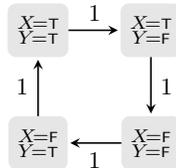
The chain is strongly connected and has a period of length 4, which explains the observed behavior that for any initial distribution γ_0, we got the sequence

$$\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_0, \gamma_1, \cdots$$

This sequence obviously converges only for initial distributions that are stationary, i.e., if we have $\gamma_0 = \gamma_0 \cdot \mathbf{P}$.

The following lemma summarizes the implications that can be drawn from close inspection of the cutset Markov chain.

Lemma 4 (Cardinality). *Let $\mathcal{B}$ be a GBN with cutset $\mathcal{C}$ and cutset Markov chain $CMC(\mathcal{B}, \mathcal{C}) = \langle Asg(\mathcal{C}), \mathbf{P} \rangle$. Further, let $k > 0$ denote the number of bottom SCCs $\mathcal{D}_1, \ldots, \mathcal{D}_k$ of $CMC(\mathcal{B}, \mathcal{C})$. Then*

1. *the cardinality of the cutset Markov chain semantics is given by*

$$\left| [\![\mathcal{B}]\!]_{\text{MC-}\mathcal{C}} \right| = \begin{cases} 1 & if \ \ k = 1, \\ \infty & if \ \ k > 1; \end{cases}$$

2. *$Lim(\mathcal{B}, \mathcal{C}, \gamma_0)$ is defined for all $\gamma_0 \in Dist(Asg(\mathcal{C}))$ if all $\mathcal{D}_i$ are aperiodic;*
3. *$Lim(\mathcal{B}, \mathcal{C}, \gamma)$ is only defined for stationary distributions γ with $\gamma = \gamma \cdot \mathbf{P}$ if $\mathcal{D}_i$ is periodic for any $1 \leqslant i \leqslant k$.*

A handy sufficient (albeit not necessary) criterion for both aperiodicity and the existence of a single bottom SCC in the cutset Markov chain is the absence of zero and one entries in the CPTs and the initial distribution of a GBN.

Definition 10 (Smooth GBNs). *A GBN $\mathcal{B} = \langle \mathcal{G}, \mathcal{P}, \iota \rangle$ is called* smooth *iff all CPT entries as given by $\mathcal{P}$ and all values in ι are in the open interval $]0, 1[$.*

Lemma 5. *Let $\mathcal{B}$ be a smooth GBN and $\mathcal{C}$ a cutset of $\mathcal{B}$. Then the graph of the cutset Markov chain $CMC(\mathcal{B}, \mathcal{C})$ is a complete digraph.*

Corollary 1. *The limit semantics of a smooth GBN $\mathcal{B}$ is a singleton for every cutset $\mathcal{C}$ of $\mathcal{B}$ and $Lim(\mathcal{B}, \mathcal{C}, \gamma_0)$ is defined for all $\gamma_0 \in Dist(Asg(\mathcal{C}))$.*

As noted in [14], one rarely needs to assign a probability of zero (or, conversely, of one) in real-world applications; and doing so in cases where some event is extremely unlikely but not impossible is a common modeling error. This observation gives reason to expect that most GBNs encountered in practice are smooth, and their semantics is thus, in a sense, well-behaved.

5.2   Relation to Constraints Semantics

We take a closer look at how the cutset semantics relates to the CPT-consistency semantics defined in Section 3. CPTs of nodes outside cutsets remain unaffected in the dissected BNs from which the Markov chain semantics is computed. Since there are cyclic GBNs for which no CPT-consistent distribution exists (cf. Example 1) while Markov chain semantics always yields at least one solution due to Lemma 4, it cannot be expected that cutset nodes are necessarily CPT-consistent. However, they are always weakly CPT-consistent.

Lemma 6. *Let $\mathcal{B}$ be a GBN over nodes $\mathcal{V}$, $\mathcal{C} \subseteq \mathcal{V}$ a cutset for $\mathcal{B}$, and $\mu \in \llbracket\mathcal{B}\rrbracket_{\mathrm{MC}\text{-}\mathcal{C}}$. Then μ is strongly CPT-consistent for all nodes in $\mathcal{V}\backslash\mathcal{C}$ and weakly CPT-consistent for the nodes in $\mathcal{C}$.*

The lemma shows a way to find fully CPT consistent distributions: Consider there is a distribution $\mu \in \llbracket\mathcal{B}\rrbracket_{\mathrm{MC}\text{-}\mathcal{C}} \cap \llbracket\mathcal{B}\rrbracket_{\mathrm{MC}\text{-}\mathcal{D}}$ for two disjoint cutsets $\mathcal{C}$ and $\mathcal{D}$. Then by Lemma 6 the nodes in $\mathcal{V} \setminus \mathcal{C}$ and $\mathcal{V} \setminus \mathcal{D}$ are CPT consistent, so in fact μ is CPT consistent. In general, we get the following result.

Lemma 7. *Let $\mathcal{B}$ be a GBN over nodes $\mathcal{V}$ and $\mathcal{C}_1, \ldots, \mathcal{C}_k$ cutsets of $\mathcal{B}$ s.t. for each node $X \in \mathcal{V}$ there is an $i \in \{1, \ldots, k\}$ with $X \notin \mathcal{C}_i$. Then*

$$\bigcap_{0 \leqslant i \leqslant k} \llbracket\mathcal{B}\rrbracket_{\mathrm{MC}\text{-}\mathcal{C}_i} \ \subseteq \ \llbracket\mathcal{B}\rrbracket_{\mathrm{CPT}}.$$

We take a look at which independencies are necessarily observed by the distributions in $\llbracket\mathcal{B}\rrbracket_{\mathrm{MC}\text{-}\mathcal{C}}$. Let $\gamma \in Dist(Asg(\mathcal{C}))$ be the cutset distribution and let $\mathcal{G}[\mathcal{C}]$ denote the graph of $Dissect(\mathcal{B}, \mathcal{C}, \gamma)$ restricted to the nodes in $\mathcal{V}$ such that the cutset nodes in $\mathcal{C}$ are initial. Then by Lemma 1, the d-separation independencies of the closure of $\mathcal{G}[\mathcal{C}]$ hold in all distributions $\mu \in \llbracket\mathcal{B}\rrbracket_{\mathrm{MC}\text{-}\mathcal{C}}$, i.e., $d\text{-}sep\big(Close(\mathcal{G}[\mathcal{C}])\big) \subseteq Indep(\mu)$. The next lemma states that any CPT-consistent distribution that satisfies these independence constraints for some cutset $\mathcal{C}$ also belongs to $\llbracket\mathcal{B}\rrbracket_{\mathrm{MC}\text{-}\mathcal{C}}$.

Lemma 8. *Let $\mathcal{B}$ be a GBN with cutset $\mathcal{C}$ and $\mathcal{I}_{\mathcal{C}} = Close(\mathcal{G}[\mathcal{C}])$. Then we have*

$$\llbracket\mathcal{B}\rrbracket_{\mathrm{CPT}\text{-}\mathcal{I}_{\mathcal{C}}} \ \subseteq \ \llbracket\mathcal{B}\rrbracket_{\mathrm{MC}\text{-}\mathcal{C}}.$$

Combining Lemma 7 and Lemma 8 yields the following equivalence.

Corollary 2. *For a GBN $\mathcal{B}$ with cutsets $\mathcal{C}_1, \ldots, \mathcal{C}_k$ as in Lemma 7 and the independence set $\mathcal{I} = \bigcup_{0 \leqslant i \leqslant k} Close(\mathcal{G}[\mathcal{C}_i])$, we have*

$$\bigcap_{0 \leqslant i \leqslant k} \llbracket\mathcal{B}\rrbracket_{\mathrm{MC}\text{-}\mathcal{C}_i} \ = \ \llbracket\mathcal{B}\rrbracket_{\mathrm{CPT}\text{-}\mathcal{I}}.$$

5.3   Overview

Fig. 6 gives an overview of the relations between all proposed semantics. Boxes represent the set of distributions induced by the respective semantics and arrows stand for set inclusion. For the non-trivial inclusions the arrows are annotated with the respective lemma or theorem. As an example, CPT→wCPT states that $\llbracket\mathcal{B}\rrbracket_{\mathrm{CPT}} \subseteq \llbracket\mathcal{B}\rrbracket_{\mathrm{wCPT}}$ holds for all GBNs $\mathcal{B}$. The three semantics in the top row parameterized with a cutset $\mathcal{C}$ and a distribution γ stand for the singleton set containing the respective function applied to γ, i.e., $\llbracket\mathcal{B}\rrbracket_{\mathrm{LIM}\text{-}\mathcal{C}\text{-}\gamma} = \{Lim(\mathcal{B}, \mathcal{C}, \gamma)\}$. $\bigcap_{\mathcal{C}} \mathrm{MC}\text{-}\mathcal{C}$ stands for the intersection of the Markov chain semantics for various cutsets as in Lemma 7, and the incoming arrow from CPT-$\mathcal{I}_{\mathcal{C}}$ holds for the set of independencies $\mathcal{I}_{\mathcal{C}}$ as in Lemma 8.
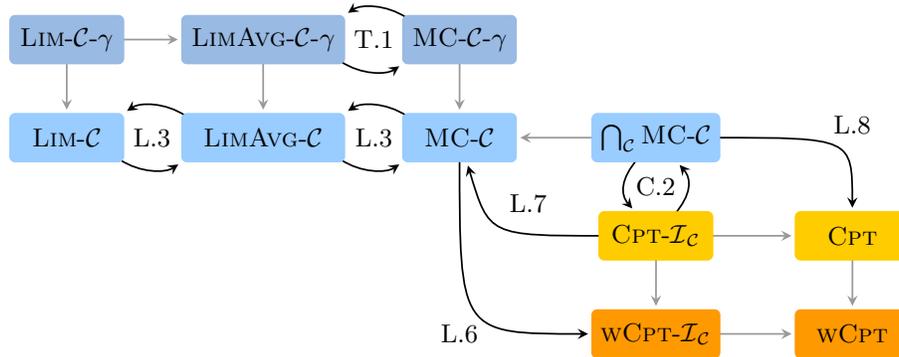
Fig. 6: Relations between different variations of limit, limit average, and Markov chain semantics (blue) as well as strong and weak CPT-consistency semantics (yellow resp. orange)

6   Related Work

That cycles in a BN might be unavoidable when learning its structure is well known for more than 30 years [15,22]. During the learning process of BNs, cycles might even be favorable as demonstrated in the context of *gene regulatory networks* where cyclic structures induce monotonic scores [32]. That work only discusses learning algorithms, but does not deal with evaluating the joint distribution of the resulting cyclic BNs. In most applications, however, cycles have been seen as a phenomenon to be avoided to ease the computation of the joint distribution in BNs. By an example BN comprising a single isolated cycle, [30] showed that reversing or removing edges to avoid cycles may reduce the solution space from infinitely many joint distributions that are (weakly) consistent with the CPTs to a single one. In this setting, our results on weak CPT-semantics also provide that wCPT cannot express conditions on the relation of variables like implications or mutual exclusion. This is rooted in the fact that the solution space of weak CPT-semantics always contains at least one full joint distribution with pairwise independent variables. An example where reversing edges led to satisfactory results has been considered in [3], investigating the impact of reinforced defects by steel corrosion in concrete structures.

Unfolding cycles up to a bounded depth has been applied in the setting of a robotic sensor system by [2]. In their use case, only cycles of length two may appear, and only the nodes appearing on the cycles are implicitly used as cutset for the unfolding. In [13], the set of all nodes is used for unfolding (corresponding to a cutset $\mathcal{C} = \mathcal{V}$ in our setting) and subsequent limit construction, but restricted to cases where the limit exists.

There have been numerous variants of BNs that explicitly or implicitly address cyclic dependencies. *Dynamic Bayesian networks* (DBNs) [19] extend BNs by an explicit notion of discrete time steps that could break cycles through

timed ordering of random variables. Cycles in BNs could be translated to the DBN formalism by introducing a notion of time, e.g., following [13]. Our cutset approach is orthogonal, choosing a time-abstract view on cycles and treating them as stabilizing feedback loops. Learning DBNs requires "relatively large time-series data" [32] and thus, may be computationally demanding. In [18] *activator random variables* break cycles in DBNs to circumvent spurious results in DBN reasoning when infinitesimal small time steps would be required.

Causal BNs [23] are BNs that impose a meaning on the direction of an edge in terms of causal dependency. Several approaches have been proposed to extend causal BNs for modeling feedback loops. In [25], an equilibrium semantics is sketched that is similar to our Markov chain semantics, albeit based on variable oderings rather than cutsets. Determining independence relations, Markov properties, and joint distributions are central problems addressed for cyclic causal BNs [2,5,20,24,29]. Markov properties and joint distributions for extended versions of causal BNs have been considered recently, e.g., in *directed graphs with hyperedges* (HEDGes) [5] and *cyclic structural causal models* (SCMs) [2]. Besides others, they show that in presence of cycles, there might be multiple solutions for a joint distribution or even no solution at all [7]. While we consider all random variables to be observable, the latter approaches focus on models with latent variables. Further, while our focus in this paper is not on causality, our approach is surely also applicable to causal BNs with cycles.

Recursive relational Bayesian networks (RRBNs) [9] allow representing probabilistic relational models where the random variables are given by relations over varying domains. The resulting first-order dependencies can become quite complex and may contain cycles, though semantics are given only for the acyclic cases by the construction of corresponding standard BNs.

Bayesian attack graphs (BAGs) [16] are popular to model and reason about security vulnerabilities in computer networks. Learned graphs and thus their BN semantics frequently contain cycles, e.g., when using the tool MuLVAL [21]. In [27], "handling cycles correctly" is identified as "a key challenge" in security risk analysis. Resolution methods for cyclic patterns in BAGs [1,4,17,31] are mainly based on context-specific security considerations, e.g., to break cycles by removing edges. The semantic foundations for cyclic BNs laid in this paper do not require graph manipulations and decouple the probability theoretic basis from context-specific properties.

7   Conclusion

This paper has developed a foundational perspective on the semantics of cycles in Bayesian networks. Constraint-based semantics provide a conservative extension of the standard BN semantics to the cyclic setting. While conceptually important, their practical use is limited by the fact that for many GBNs, the induced constraint system is unsatisfiable. On the other hand, the two introduced limit semantics echo in an abstract and formal way what practitioners have been devising across a manifold of domain-specific situations. In this abstract perspective,

cutsets are the ingredients that enable a controlled decoupling of dependencies. The appropriate choice of cutsets is where, in our view, domain-specific knowledge is confined to enter the picture. Utilizing the constructively defined Markov chain semantics, we established key results relating and demarcating the different semantic notions and showed that for the ubiquitous class of smooth GBNs a unique full joint distribution always exists.

References

1. Aguessy, F., Bettan, O., Blanc, G., Conan, V., Debar, H.: Bayesian attack model for dynamic risk assessment. CoRR (2016)
2. Castellano-Quero, M., Fernández-Madrigal, J.A., García-Cerezo, A.: Improving bayesian inference efficiency for sensory anomaly detection and recovery in mobile robots. Expert Systems with Applications **163** (2021)
3. Castillo, E., Gutiérrez, J.M., Hadi, A.S.: Modeling probabilistic networks of discrete and continuous variables. Journal of Multivariate Analysis **64**(1), 48–65 (1998)
4. Doynikova, E., Kotenko, I.: Enhancement of probabilistic attack graphs for accurate cyber security monitoring. In: 2017 IEEE SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI. pp. 1–6 (2017)
5. Forré, P., Mooij, J.M.: Markov properties for graphical models with cycles and latent variables (2017)
6. Geiger, D., Verma, T., Pearl, J.: d-separation: From theorems to algorithms. In: Machine Intelligence and Pattern Recognition, vol. 10, pp. 139–148. Elsevier (1990)
7. Halpern, J.Y.: Axiomatizing causal reasoning. J. Artif. Int. Res. **12**, 317–337 (2000)
8. Han, S.H.: A top-down iteration algorithm for monte carlo method for probability estimation of a fault tree with circular logic. NET **50**(6), 854–859 (2018)
9. Jaeger, M.: Complex probabilistic modeling with recursive relational bayesian networks. Ann. Math. Artif. Intell. **32**(1-4), 179–220 (2001)
10. Jensen, F., Nielsen, T.: Bayesian Network and Decision Graphs. Springer (2007)
11. Jensen, F.V., Andersen, S.K., Kjærulff, U., Andreassen, S.: Munin — on the case for probabilities in medical expert systems — a practical exercise. In: AIME 87. pp. 149–160. Springer Berlin Heidelberg (1987)
12. Kemeny, J., Snell, J.: Finite Markov chains. University series in undergraduate mathematics, VanNostrand, New York, repr edn. (1969)
13. Kłopotek, M.A.: Cyclic Bayesian network: Markov process approach. Studia Informatica: systems and information technology **1**(7), 47–55 (2006)
14. Koller, D., Friedman, N.: Probabilistic Graphical Models: Principles and Techniques. Adaptive Computation and Machine Learning, MIT Press (2009)
15. Lauritzen, S.L., Spiegelhalter, D.J.: 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)
16. Liu, Y., Man, H.: Network vulnerability assessment using Bayesian networks. In: Data Mining, Intrusion Detection, Information Assurance, and Data Networks Security. vol. 5812, pp. 61–71. SPIE (2005)
17. Matthews, I., Mace, J., Soudjani, S., van Moorsel, A.: Cyclic bayesian attack graphs: A systematic computational approach (2020)
18. Motzek, A.: Indirect Causes, Dependencies and Causality in Bayesian Networks. Ph.D. thesis, University of Lübeck (2017)

19. Murphy, K.P.: Dynamic Bayesian Networks: Representation, Inference and Learning. Ph.D. thesis, University of California, Berkeley (2002)
20. Neal, R.M.: On deducing conditional independence from d-separation in causal graphs with feedback. Journal of Artificial Intelligence Research **12**, 87–91 (2000)
21. Ou, X., Govindavajhala, S., Appel, A.W.: Mulval: A logic-based network security analyzer. In: Proceedings of the 14th Conference on USENIX Security Symposium - Volume 14. p. 8. SSYM'05, USENIX Association, USA (2005)
22. Pearl, J.: Fusion, Propagation, and Structuring in Belief Networks, pp. 366–413. Morgan Kaufmann Publishers Inc. (1990)
23. Pearl, J.: Causality: Models, Reasoning and Inference. CUP (2009)
24. Pearl, J., Dechter, R.: Identifying independencies in causal graphs with feedback. In: Proceedings of the Twelfth International Conference on Uncertainty in Artificial Intelligence. pp. 420–426. UAI'96, Morgan Kaufmann Publishers Inc. (1996)
25. Poole, D., Crowley, M.: Cyclic causal models with discrete variables: Markov chain equilibrium semantics and sample ordering. In: IJCAI International Joint Conference on Artificial Intelligence (2013)
26. Robert F. Nease, J., Owens, D.K.: Use of influence diagrams to structure medical decisions. Medical Decision Making **17**(3), 263–275 (1997)
27. Singhal, A., Ou, X.: Security Risk Analysis of Enterprise Networks Using Probabilistic Attack Graphs, pp. 53–73. Springer International Publishing, Cham (2017)
28. Spirtes, P.: Conditional independence in directed cyclic graphical models for feedback. Tech. Rep. CMU-PHIL-53, Carnegie Mellon University (1994)
29. Spirtes, P.: Directed cyclic graphical representations of feedback models. In: Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence. pp. 491–498. UAI'95, Morgan Kaufmann Publishers Inc. (1995)
30. Tulupyev, A.L., Nikolenko, S.I.: Directed cycles in bayesian belief networks: Probabilistic semantics and consistency checking complexity. In: MICAI 2005: Advances in Artificial Intelligence. pp. 214–223. Springer Berlin Heidelberg (2005)
31. Wang, L., Islam, T., Long, T., Singhal, A., Jajodia, S.: An attack graph-based probabilistic security metric. In: Data and Applications Security XXII. pp. 283–296. Springer Berlin Heidelberg (2008)
32. Wiecek, W., Bois, F.Y., Gayraud, G.: Structure learning of bayesian networks involving cyclic structures (2020)

A   Appendix

The appendix contains the proofs omitted from the body of the submission "On the Foundations of Cycles in Bayesian Networks" due to space constraints.

Lemma 1. *Let* $\mathcal{B}_\emptyset = \langle \mathcal{G}, \mathcal{P}, \iota \rangle$ *be an acyclic GBN. Then*

$$d\text{-}sep\big(Close(\mathcal{G})\big) \ \subseteq \ Indep\big(dist_{\mathrm{BN}}(\mathcal{B}_\emptyset)\big).$$

Proof. The idea is to show that the dependencies of every possible BN structure for the initial distribution ι are covered by the closure operation. Let the graph $\mathcal{G}_\iota^\star = \langle Init(\mathcal{G}), \mathcal{E}^\star \rangle$ be a DAG that is an I-map for ι, i.e., $d\text{-}sep(\mathcal{G}_\iota^\star) \subseteq Indep(\iota)$. Then ι factorizes according to $\mathcal{G}_\iota^\star$, that is for every assignment $b \in Asg(Init(\mathcal{G}))$, we have

$$\iota(b) \ = \ \prod_{X \in Init(\mathcal{G})} \iota\big(b_X \mid b_{Pre^{\mathcal{G}_\iota^\star}(X)}\big).$$

Now consider the BN $\mathcal{B}_\emptyset^\star$ with graph $\mathcal{G}^\star = \langle \mathcal{V}, \mathcal{E} \cup \mathcal{E}^\star \rangle$ where we add the edges of $\mathcal{G}_\iota^\star$ to $\mathcal{G}$. The CPTs for the nodes in $\mathcal{V} \setminus Init(\mathcal{G})$ are given by $\mathcal{P}$ whereas the new CPTs (according to the structure in $\mathcal{G}_\iota^\star$) for the nodes in $Init(\mathcal{G})$ are derived from ι. Then for every assignment $c \in Asg(\mathcal{V})$:

$$
\begin{aligned}
dist_{\mathrm{BN}}(\mathcal{B}_\emptyset^\star)(c) \ &= \ \prod_{X \in \mathcal{V}} \Pr\big(c_X \mid c_{Pre(X)}\big) \\
&= \ \prod_{X \in Init(\mathcal{G})} \iota\big(c_X \mid c_{Pre^{\mathcal{G}_\iota^\star}(X)}\big) \ \cdot \ \prod_{X \in \mathcal{V} \setminus Init(\mathcal{G})} \Pr\big(c_X \mid c_{Pre(X)}\big) \\
&= \ \iota\big(c_{Init(\mathcal{G})}\big) \ \cdot \ \prod_{X \in \mathcal{V} \setminus Init(\mathcal{G})} \Pr\big(c_X \mid c_{Pre(X)}\big) \\
&= \ dist_{\mathrm{BN}}(\mathcal{B}_\emptyset)(c).
\end{aligned}
$$

As $\mathcal{B}_\emptyset^\star$ is a regular BN without an initial distribution, we have $d\text{-}sep(\mathcal{G}^\star) \subseteq Indep(dist_{\mathrm{BN}}(\mathcal{B}_\emptyset^\star))$.

We proceed to show $d\text{-}sep(Close(\mathcal{G})) \subseteq d\text{-}sep(\mathcal{G}^\star)$. Let $(X \perp Y \mid \mathcal{Z}) \in d\text{-}sep(Close(\mathcal{G}))$. Then each path from X to Y in $Close(\mathcal{G})$ is blocked by the nodes in $\mathcal{Z}$. As $Close(\mathcal{G})$ contains all possible edges between the nodes $Init(\mathcal{G})$ but $\mathcal{G}^\star$ only a subset thereof, it is clear that each path in $\mathcal{G}^\star$ also exists in $Close(\mathcal{G})$. Thus, there cannot be an unblocked path from X to Y given $\mathcal{Z}$ in $\mathcal{G}^\star$ either, so $(X \perp Y \mid \mathcal{Z}) \in d\text{-}sep(\mathcal{G}^\star)$. Altogether, we have

$$d\text{-}sep\big(Close(\mathcal{G})\big) \subseteq d\text{-}sep(\mathcal{G}^\star) \subseteq Indep\big(dist_{\mathrm{BN}}(\mathcal{B}_\emptyset^\star)\big) = Indep\big(dist_{\mathrm{BN}}(\mathcal{B}_\emptyset)\big).$$

$\square$

Lemma 2. *Let* $\mathcal{B}$ *be a GBN with cutset* $\mathcal{C}$*, cutset distribution* $\gamma \in Dist(Asg(\mathcal{C}))$*, and* $\mathcal{M} = \langle Asg(\mathcal{C}), \mathbf{P} \rangle$ *the cutset Markov chain* $CMC(\mathcal{B}, \mathcal{C})$*. Then the following statements are equivalent:*

(a) $\gamma = \gamma \cdot \mathbf{P}$.

(b) There exists $\gamma_0 \in Dist(Asg(\mathcal{C}))$ such that for $\gamma_{i+1} = \gamma_i \cdot \mathbf{P}$, we have

$$\gamma \;=\; \lim_{n \to \infty} \frac{1}{n+1} \sum_{i=0}^{n} \gamma_i.$$

(c) γ belongs to the convex hull of the long-run frequency distributions $\mathrm{lrf}_{\mathcal{D}}$ of the bottom SCCs $\mathcal{D}$ of $\mathcal{M}$.

(d) $\gamma = Next(\mathcal{B}, \mathcal{C}, \gamma)|_{\mathcal{C}}$.

Proof. (a) $\Longrightarrow$ (b): If we have $\gamma = \gamma \cdot \mathbf{P}$, then statement (b) is obtained by considering $\gamma_0 = \gamma$, as then $\gamma_i = \gamma$ for all i.

(b) $\Longrightarrow$ (c): The proof of the implication relies on the following standard facts about finite-state Markov chains. Given a BSCC $\mathcal{D}$ and an arbitrary distribution $\nu_0 \in Dist(Asg(\mathcal{D}))$, the distribution $\mathrm{lrf}_{\mathcal{D}}$ agrees with the Cesàro limit of the sequence $(\nu_i)_{i \geqslant 0}$ where $\nu_{i+1} = \nu_i \cdot \mathbf{P}_{\mathcal{D}}$ and $\mathbf{P}_{\mathcal{D}}$ denotes the restriction of $\mathbf{P}$ to assignments on $\mathcal{D}$. That is,

$$\mathrm{lrf}_{\mathcal{D}} = \lim_{n \to \infty} \frac{1}{n+1} \sum_{i=0}^{n} \nu_i.$$

Vice versa, for $\gamma_0 \in Dist(Asg(\mathcal{C}))$ and $\gamma_{i+1} = \gamma_i \cdot \mathbf{P}$, then the Cesàro limit γ of the sequence $(\gamma_i)_{i \geqslant 0}$ has the form

$$\gamma = \sum_{\mathcal{D}} \lambda(\mathcal{D}) \cdot \mathrm{lrf}_{\mathcal{D}}$$

where $\mathcal{D}$ ranges over all BSCCs of $\mathcal{M}$, $\lambda(\mathcal{D})$ is the probability for reaching $\mathcal{D}$ in $\mathcal{M}$ with the initial distribution γ_0, and all vectors $\mathrm{lrf}_{\mathcal{D}}$ are padded with zero entries to range over the whole state space. In particular, γ is a convex combination of the distributions $\mathrm{lrf}_{\mathcal{D}}$ as $0 \leqslant \lambda(\mathcal{D}) \leqslant 1$ and $\sum_{\mathcal{D}} \lambda(\mathcal{D}) = 1$ (because every finite-state Markov chain almost surely reaches a BSCC).

(c) $\Longrightarrow$ (a): Suppose $\gamma = \sum_{\mathcal{D}} \lambda(\mathcal{D}) \cdot \mathrm{lrf}_{\mathcal{D}}$ where $0 \leqslant \lambda(\mathcal{D}) \leqslant 1$, $\sum_{\mathcal{D}} \lambda(\mathcal{D}) = 1$, and each $\mathrm{lrf}_{\mathcal{D}}$ is padded appropriately as before. Then:

$$\gamma \cdot \mathbf{P} \;=\; \sum_{\mathcal{D}} \lambda(\mathcal{D}) \cdot \mathrm{lrf}_{\mathcal{D}} \cdot \mathbf{P} \;=\; \sum_{\mathcal{D}} \lambda(\mathcal{D}) \cdot \mathrm{lrf}_{\mathcal{D}} \;=\; \gamma$$

where we use the fact that $\mathrm{lrf}_{\mathcal{D}} = \mathrm{lrf}_{\mathcal{D}} \cdot \mathbf{P}$.

(a) $\Longleftrightarrow$ (d): Because γ can be represented as convex combination of Dirac distributions as $\gamma = \sum_{c \in Asg(\mathcal{C})} \gamma(c) \cdot Dirac(c)$, we know:

$$Next(\mathcal{B}, \mathcal{C}, \gamma) \;=\; \sum_{c \in Asg(\mathcal{C})} \gamma(c) \cdot Next\big(\mathcal{B}, \mathcal{C}, Dirac(c)\big).$$

As $\mathbf{P}(c, b) = Next\big(\mathcal{B}, \mathcal{C}, Dirac(c)\big)(b)$ for any assignment $b \in Asg(\mathcal{C})$, and assuming $\gamma = \gamma \cdot \mathbf{P}$, we get

$$Next(\mathcal{B}, \mathcal{C}, \gamma)(b) \;=\; \sum_{c \in Asg(\mathcal{C})} \gamma(c) \cdot \mathbf{P}(c, b) \;=\; (\gamma \cdot \mathbf{P})(b) \;=\; \gamma(b).$$

Conversely, assuming $Next(\mathcal{B}, \mathcal{C}, \gamma)|_{\mathcal{C}} = \gamma$, we yield $\gamma = \gamma \cdot \mathbf{P}$. $\qquad\square$

Lemma 3. *Let $\mathcal{B}$ be a GBN. Then for any cutset $\mathcal{C}$ of $\mathcal{B}$, we have*

$$[\![\mathcal{B}]\!]_{\text{MC-}\mathcal{C}} \;=\; [\![\mathcal{B}]\!]_{\text{LimAvg-}\mathcal{C}} \;=\; [\![\mathcal{B}]\!]_{\text{Lim-}\mathcal{C}}.$$

Proof. We have $[\![\mathcal{B}]\!]_{\text{MC-}\mathcal{C}} = [\![\mathcal{B}]\!]_{\text{LimAvg-}\mathcal{C}}$ by Theorem 1 and know $[\![\mathcal{B}]\!]_{\text{Lim-}\mathcal{C}} \subseteq [\![\mathcal{B}]\!]_{\text{LimAvg-}\mathcal{C}}$, so it remains to show $[\![\mathcal{B}]\!]_{\text{MC-}\mathcal{C}} \subseteq [\![\mathcal{B}]\!]_{\text{Lim-}\mathcal{C}}$. Let $\mu \in [\![\mathcal{B}]\!]_{\text{MC-}\mathcal{C}}$. Then there exists a cutset distribution γ s.t. $\mu = Extend(\mathcal{B}, \mathcal{C}, \gamma)$. We need to show there exists an initial distribution $\gamma_0 \in Dist(Asg(\mathcal{C}))$ such that $\gamma = \lim_{n\to\infty} \gamma_i$ where $\gamma_{i+1} = Next(\mathcal{B}, \mathcal{C}, \gamma_i)|_{\mathcal{C}}$. Let us choose $\gamma_0 = \gamma$. Then we know $\gamma_0 = Next(\mathcal{B}, \mathcal{C}, \gamma_0)|_{\mathcal{C}}$ by Lemma 2, so $\gamma_i = \gamma_0$ for all $i \in \mathbb{N}$. Thus, $\gamma = \lim_{i\to\infty} \gamma_i$ and therefore $\mu \in [\![\mathcal{B}]\!]_{\text{Lim-}\mathcal{C}}$. □

Lemma 4 (Cardinality). *Let $\mathcal{B}$ be a GBN with cutset $\mathcal{C}$ and cutset Markov chain $CMC(\mathcal{B}, \mathcal{C}) = \langle Asg(\mathcal{C}), \mathbf{P} \rangle$. Further, let $k > 0$ denote the number of bottom SCCs $\mathcal{D}_1, \ldots, \mathcal{D}_k$ of $CMC(\mathcal{B}, \mathcal{C})$. Then*

1. *the cardinality of the cutset Markov chain semantics is given by*

$$\left| [\![\mathcal{B}]\!]_{\text{MC-}\mathcal{C}} \right| \;=\; \begin{cases} 1 & \text{if } k = 1, \\ \infty & \text{if } k > 1; \end{cases}$$

2. *$Lim(\mathcal{B}, \mathcal{C}, \gamma_0)$ is defined for all $\gamma_0 \in Dist(Asg(\mathcal{C}))$ if all $\mathcal{D}_i$ are aperiodic;*
3. *$Lim(\mathcal{B}, \mathcal{C}, \gamma)$ is only defined for stationary distributions γ with $\gamma = \gamma \cdot \mathbf{P}$ if $\mathcal{D}_i$ is periodic for any $1 \leqslant i \leqslant k$.*

Proof. (1.) By Lemma 2, every cutset distribution with $\gamma = \gamma \cdot \mathbf{P}$ is a convex combination of the steady-state distributions for the BSCCs. Thus, for $k = 1$ a unique distribution γ exists, whereas for $k > 1$, there are infinitely many real-valued distributions in the convex hull.

(2.) A Markov chain is aperiodic if all its BSCCs are aperiodic. Aperiodicity suffices for the limit $\lim_{n\to\infty} \gamma_n$ with $\gamma_{n+1} = \gamma_n \cdot \mathbf{P}$ to exist for every γ_0. Then $\lim_{n\to\infty} \gamma'_n$ with $\gamma'_{n+1} = Next(\mathcal{B}, \mathcal{C}, \gamma'_n)|_{\mathcal{C}}$ exists as well by Lemma 2.

(3.) Assume some BSCC $\mathcal{D}$ is periodic with a period of p. Then, for any $\gamma_0 \in Dist(Asg(\mathcal{C}))$, $\gamma_{n+1} = \gamma_n \cdot \mathbf{P}$, and $\nu_n = \gamma_n|_{\mathcal{D}}$, we have $\nu_{p\cdot n} = \nu_n$. Now consider γ_0 and $\gamma_1 = \gamma_0 \cdot \mathbf{P}$. If $\gamma_0 = \gamma_1$, then $\gamma_0 = \gamma_n$ for all $n \in \mathbb{N}$ and $\gamma_0 = \lim_{n\to\infty} \gamma_n$ holds. Otherwise, if $\gamma_0 \neq \gamma_1$, the following non-convergent sequence exists:

$$\nu_0, \nu_1, \ldots, \nu_p, \nu_{p+1}, \ldots, \nu_{2p}, \nu_{2p+1}, \ldots$$

Then $\lim_{n\to\infty} \gamma_n$ cannot converge either, so $Lim(\mathcal{B}, \mathcal{C}, \gamma_0)$ is undefined. □

Lemma 5. *Let $\mathcal{B}$ be a smooth GBN and $\mathcal{C}$ a cutset of $\mathcal{B}$. Then the graph of the cutset Markov chain $CMC(\mathcal{B}, \mathcal{C})$ is a complete digraph.*

Proof. The graph of $CMC(\mathcal{B}, \mathcal{C}) = \langle Asg(\mathcal{C}), \mathbf{P} \rangle$ is a complete digraph iff each entry in $\mathbf{P}$ is positive. Thus, for each two assignments $b, c \in Asg(\mathcal{C})$, we need to

show $\mathbf{P}(b, c) > 0$. Let $\mathcal{B}_b = Dissect(\mathcal{B}, \mathcal{C}, Dirac(b))$. Then from Definition 8, we have

$$\mathbf{P}(b, c) = Next(Dirac(b), \mathcal{B}, \mathcal{C})(c)$$
$$= dist_{\mathrm{BN}}(\mathcal{B}_b)(c').$$

The probability $dist_{\mathrm{BN}}(\mathcal{B}_b)(c')$ is given by the sum over all full assignments $v \in Asg(\mathcal{V})$ that agree with c' on the assignment of the cutset node copies $\mathcal{C}'$. Further, the sum can be partitioned into those v that agree with assignment b on $\mathcal{C}$ and those that do not:

$$dist_{\mathrm{BN}}(\mathcal{B}_b)(c') \;=\; \sum_{\substack{v \in Asg(\mathcal{V}) \\ s.t. \; c' \subset v, \; b \subset v}} dist_{\mathrm{BN}}(\mathcal{B}_b)(v) \;+\; \sum_{\substack{v \in Asg(\mathcal{V}) \\ s.t. \; c \subset v, \; b \not\subset v}} dist_{\mathrm{BN}}(\mathcal{B}_b)(v).$$

By the definition of the standard BN-semantics, we have

$$dist_{\mathrm{BN}}(\mathcal{B}_b)(v) \;=\; \iota\big(v_{Init(\mathcal{G})}\big) \cdot Dirac(b)(v_{\mathcal{C}}) \cdot \prod_{X \in \mathcal{V} \setminus \mathcal{C}} \Pr\big(v_X \mid v_{Pre(X)}\big).$$

Now consider the second sum in the previous equation where $b \not\subset v$. For those assignments, $Dirac(b)(v_{\mathcal{C}}) = 0$ and thus the whole sum equals zero. For the first sum, we have $v_{\mathcal{C}} = b$, so $Dirac(b)(v_{\mathcal{C}}) = 1$ and we only need to consider the product with $X \in \mathcal{V} \setminus \mathcal{C}$ and the initial distribution over $Init(\mathcal{G})$. By the construction of $\mathcal{B}_b$, the CPTs of all $X \in \mathcal{V} \setminus \mathcal{C}$ are the original CPTs from $\mathcal{B}$, thus their entries all fall within the open interval $]0, 1[$ by the smoothness assumption of $\mathcal{B}$. The same holds for the value $\iota\big(v_{Init(\mathcal{G})}\big)$. Thus, the whole product resides in $]0, 1[$ as well. Finally, note that the sum is non-empty as $\mathcal{C}'$ and $\mathcal{C}$ are disjoint, so there exists at least one $v \in Asg(\mathcal{V})$ with $c \subset v$ and $b \subset v$. As a non-empty sum over values in $]0, 1[$ is necessarily positive, we have $dist_{\mathrm{BN}}(\mathcal{B}_b)(c') > 0$ and the claim follows. $\qquad\square$

Corollary 1. *The limit semantics of a smooth GBN $\mathcal{B}$ is a singleton for every cutset $\mathcal{C}$ of $\mathcal{B}$ and $Lim(\mathcal{B}, \mathcal{C}, \gamma_0)$ is defined for all $\gamma_0 \in Dist(Asg(\mathcal{C}))$.*

Proof. Follows from Lemma 4 and Lemma 5 because every complete graph forms a single bottom SCC and is necessarily aperiodic. $\qquad\square$

Lemma 6. *Let $\mathcal{B}$ be a GBN over nodes $\mathcal{V}$, $\mathcal{C} \subseteq \mathcal{V}$ a cutset for $\mathcal{B}$, and $\mu \in [\![\mathcal{B}]\!]_{\mathrm{MC}\text{-}\mathcal{C}}$. Then μ is strongly CPT-consistent for all nodes in $\mathcal{V} \setminus \mathcal{C}$ and weakly CPT-consistent for the nodes in $\mathcal{C}$.*

Proof. By definition, $\mu = Extend(\mathcal{B}, \mathcal{C}, \gamma)$ for some $\gamma \in Dist(Asg(\mathcal{C}))$ with $\gamma = \gamma \cdot \mathbf{P}$. As $Extend(\mathcal{B}, \mathcal{C}, \gamma)$ is the standard BN semantics for the acyclic BN $Dissect(\mathcal{B}, \mathcal{C}, \gamma)$ without the copies of the cutset nodes, CPT-consistency for the nodes in $\mathcal{V} \setminus \mathcal{C}$ follows directly from the CPT-consistency of the standard semantics for acyclic BNs.

It remains to prove weak CPT-consistency for the cutset nodes. Let $\delta = dist_{\text{BN}}(Dissect(\mathcal{B}, \mathcal{C}, \gamma)) \in Dist(Asg(\mathcal{V} \cup \mathcal{C}'))$. Thus, $\mu = \delta|_{\mathcal{V}}$ and $\gamma = \delta|_{\mathcal{C}}$. Then for each assignment $b \in Asg(\mathcal{C})$, we have

$$\mu(b) \;=\; \gamma(b) \;=\; (\gamma \cdot \mathbf{P})(b) \;=\; \delta(b')$$

where $b' \in Asg(\mathcal{C}')$ is given by $b'(Y') = b(Y)$ for all $Y \in \mathcal{C}$. In particular, for each $Y \in \mathcal{C}$:

$$\mu(Y{=}\mathsf{T}) \;=\; \delta(Y'{=}\mathsf{T})$$

Let $D = Asg(Pre(Y))$ where $Pre(\cdot)$ refers to the original scGBN. For $c \in Asg(\mathcal{C})$, we write D_c for the set of all assignments $d \in D$ that comply with c in the sense that if $Z \in \mathcal{C} \cap Pre(Y)$ then $c(Z) = d(Z)$. In this case, c and d can be combined to an assignment for $\mathcal{C} \cup Pre(Y)$. Similarly, if $d \in D$, then the notation $Asg_d(\mathcal{C})$ is used for the set of assignments $c \in Asg(\mathcal{C})$ that comply with d. Then:

$$\delta(Y'{=}\mathsf{T}) = \sum_{c \in Asg(\mathcal{C})} \delta(Y'{=}\mathsf{T} \mid c) \cdot \mu(c)$$

$$= \sum_{c \in Asg(\mathcal{C})} \sum_{d \in D_c} \underbrace{\delta(Y'{=}\mathsf{T} \mid c, d)}_{\Pr(Y{=}\mathsf{T}\mid d)} \cdot \underbrace{\delta(d \mid c)}_{\mu(d\mid c)} \cdot \underbrace{\delta(c)}_{\mu(c)}$$

$$= \sum_{d \in D} \Pr(Y{=}\mathsf{T} \mid d) \cdot \sum_{c \in Asg_d(\mathcal{C})} \mu(d \mid c) \cdot \mu(c)$$

$$= \sum_{d \in D} \Pr(Y{=}\mathsf{T} \mid d) \cdot \mu(d).$$

Putting everything together, we obtain:

$$\mu(Y{=}\mathsf{T}) \;=\; \delta(Y'{=}\mathsf{T}) \;=\; \sum_{d \in D} \Pr(Y{=}\mathsf{T} \mid d) \cdot \mu(d).$$

Thus, μ is weakly CPT-consistent for $Y \in \mathcal{C}$.   □

Lemma 7. *Let $\mathcal{B}$ be a GBN over nodes $\mathcal{V}$ and $\mathcal{C}_1, \ldots, \mathcal{C}_k$ cutsets of $\mathcal{B}$ s.t. for each node $X \in \mathcal{V}$ there is an $i \in \{1, \ldots, k\}$ with $X \notin \mathcal{C}_i$. Then*

$$\bigcap_{0 \leqslant i \leqslant k} [\![\mathcal{B}]\!]_{\text{MC-}\mathcal{C}_i} \;\subseteq\; [\![\mathcal{B}]\!]_{\text{CPT}}.$$

Proof. We need to show CPT-consistency for every node under $\mu \in \bigcap_i [\![\mathcal{B}]\!]_{\text{MC-}\mathcal{C}_i}$. Let $X \in \mathcal{V}$. Then we choose a cutset $\mathcal{C}_i$ s.t. $X \notin \mathcal{C}_i$ and CPT consistency follows from Lemma 6.   □

Lemma 8. *Let $\mathcal{B}$ be a GBN with cutset $\mathcal{C}$ and $\mathcal{I}_{\mathcal{C}} = d\text{-}sep\big(Close(Close(\mathcal{G})[\mathcal{C}])\big)$. Then we have*

$$[\![\mathcal{B}]\!]_{\text{CPT-}\mathcal{I}_{\mathcal{C}}} \;\subseteq\; [\![\mathcal{B}]\!]_{\text{MC-}\mathcal{C}}.$$

Proof. Let $\mu \in [\![\mathcal{B}]\!]_{\text{CPT-}\mathcal{I}_\mathcal{C}}$ and $\gamma = \mu|_\mathcal{C}$. The task is to show that γ satisfies the fixed point equation $\gamma = \gamma \cdot \mathbf{P}$.

The standard BN semantics $\delta = dist_{\text{BN}}(Dissect(\mathcal{B}, \mathcal{C}, \gamma))$ of the dissected BN is the unique distribution over $Asg(\mathcal{V} \cup \mathcal{C}')$ that

- is CPT-consistent w.r.t. the conditional probability tables in $Dissect(\mathcal{B}, \mathcal{C}, \gamma)$,
- agrees with γ when restricted to the assignments for $\mathcal{C}$, and
- satisfies the conditional independencies in $\mathcal{I}_\mathcal{C}$.

Consider the distribution $\tilde{\mu} \in Dist\big(Asg(\mathcal{V} \cup \mathcal{C}')\big)$ defined as follows for $b \in Asg(\mathcal{V})$ and $c' \in Asg(\mathcal{C}')$:

$$\tilde{\mu}(b, c') \; := \; \mu(b) \cdot \prod_{Y \in \mathcal{C}} \Pr\big(Y{=}c'(Y') \mid b_{Pre(Y)}\big).$$

Then, $\tilde{\mu}$ satisfies the above three constraints. Hence, $\tilde{\mu} = \delta$.

For $c \in Asg(\mathcal{C})$, let $c' \in Asg(\mathcal{C}')$ denote the corresponding assignment with $c'(Y') = c(Y)$ for $Y \in \mathcal{C}$.

$$\begin{aligned}
(\gamma \cdot \mathbf{P})(c) \; &= \; \delta(c') \; = \; \tilde{\mu}(c') \\
&= \; \sum_{d \in Asg(Pre(\mathcal{C}))} \mu(d) \cdot \prod_{Y \in \mathcal{C}} \underbrace{\Pr(Y{=}c'(Y') \mid d)}_{\Pr(Y{=}c(Y)\mid d)} \; = \; \mu(c) \; = \; \gamma(c).
\end{aligned}$$

Hence, $\gamma = \gamma \cdot \mathbf{P}$ and $\mu \in [\![\mathcal{B}]\!]_{\text{MC-}\mathcal{C}}$. $\qquad\qquad\square$