Representation of context-specific causal models with observational and interventional data

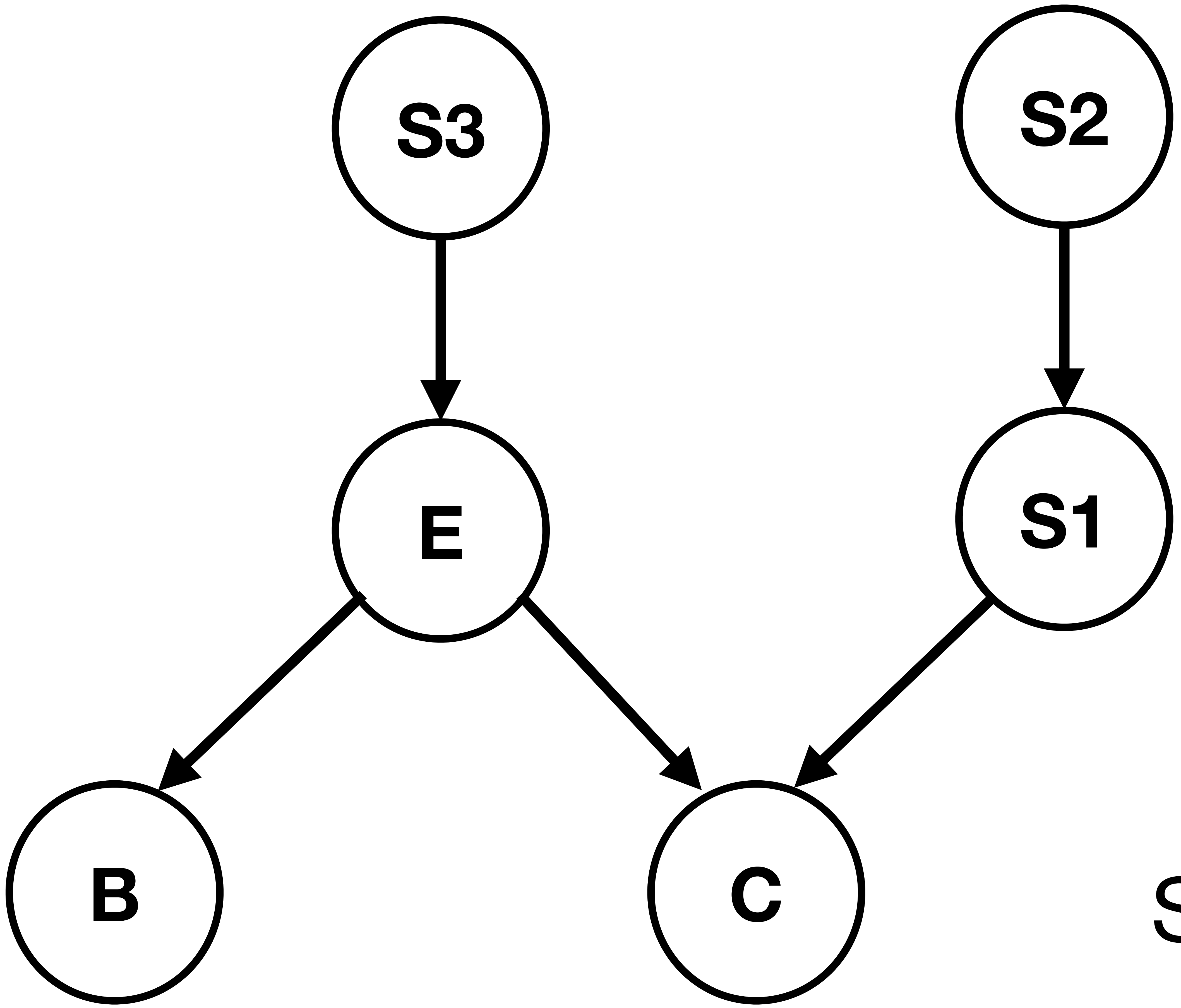
Eliana Duarte

CENTRO DE MATEMÁTICA UNIVERSIDADE DO PORTO

FCT Fundação para a Ciência e a Tecnologia

LASI Intelligent Systems Associate Laboratory

Graphical models

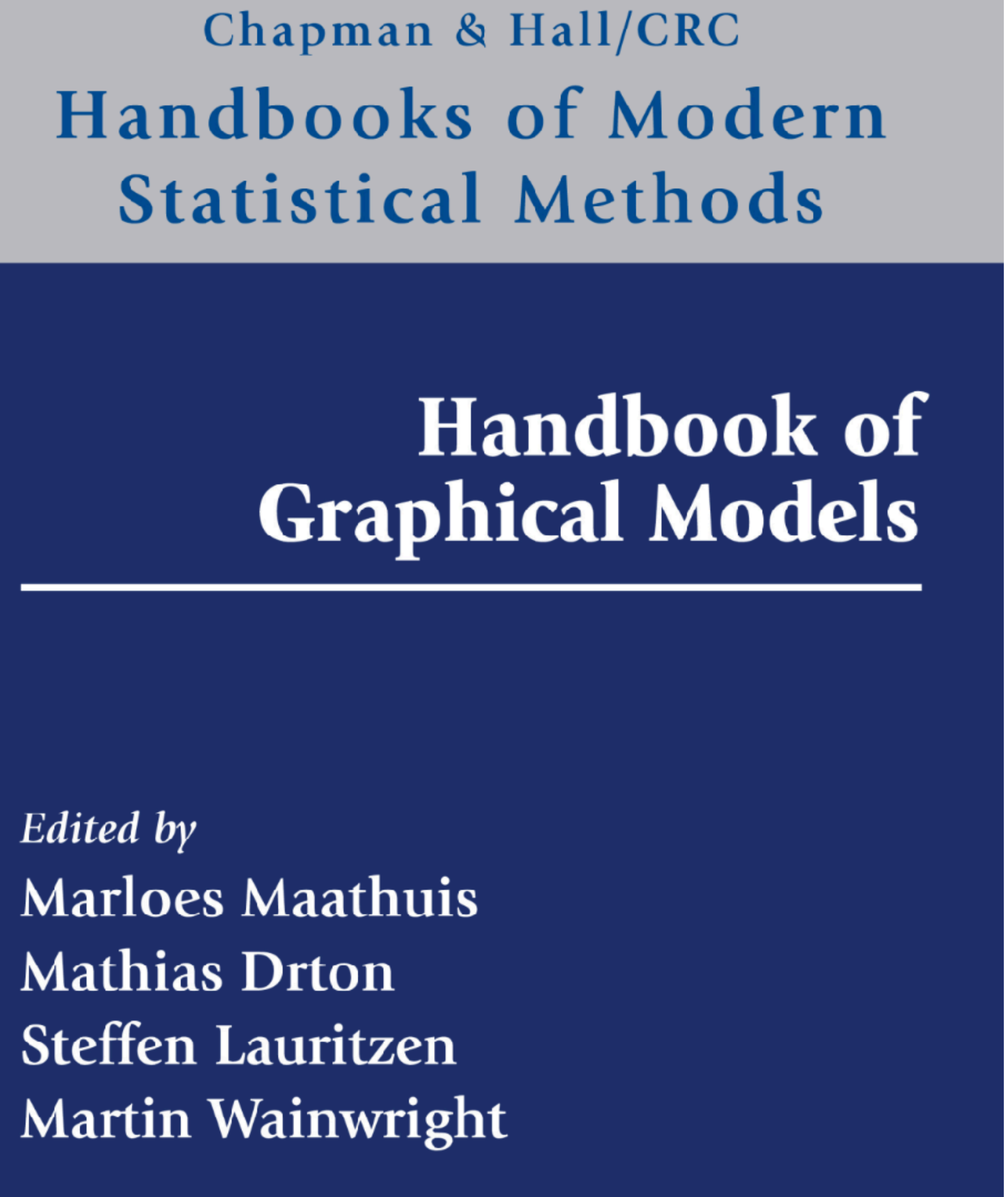


B = biomarker
C = cancer
E = environment
S1 = SNP 1
S2 = SNP 2
S3 = SNP 3

Single Nucleotide Polymorphism (SNP)

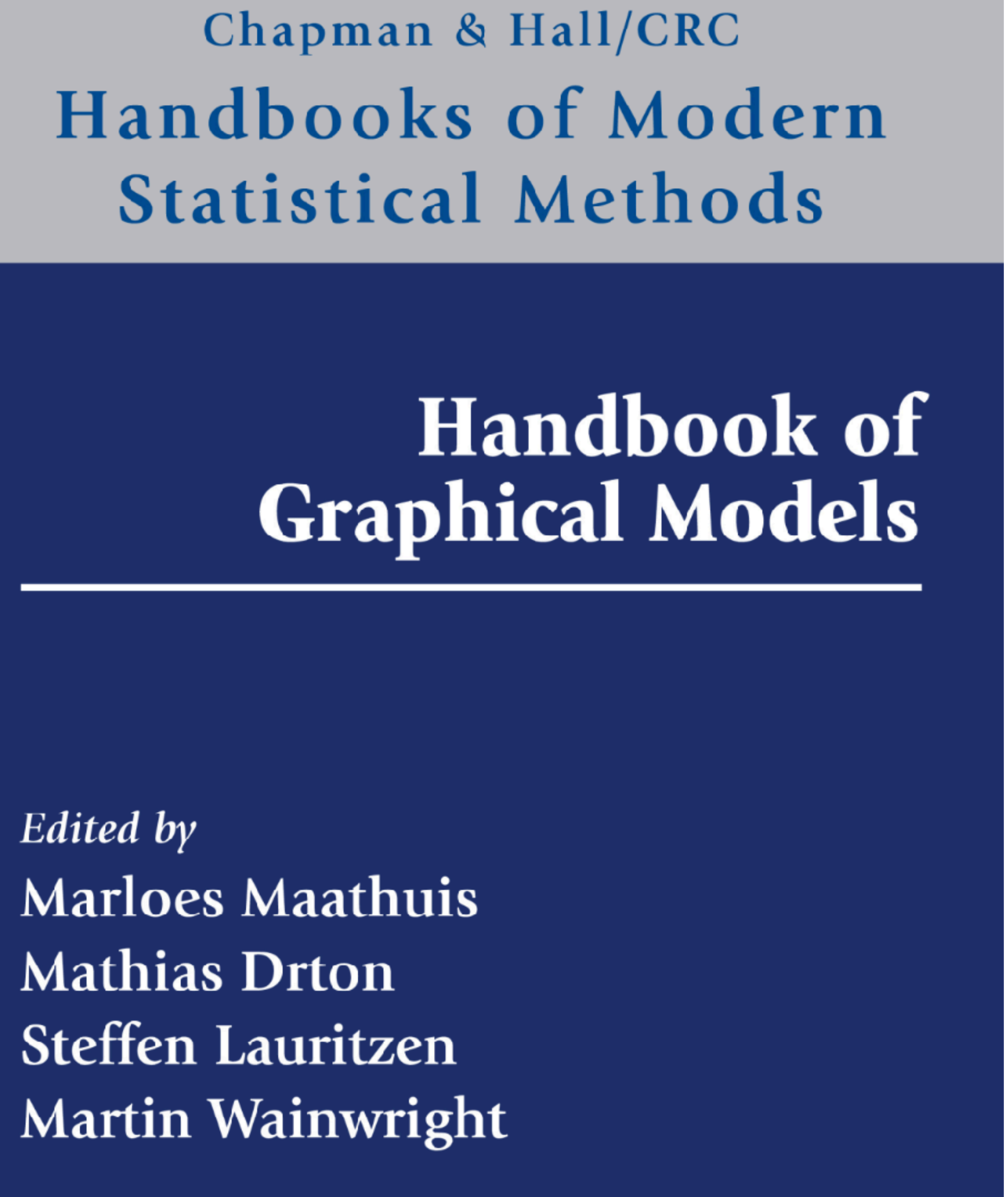
DAG = Directed **A**cyclic **G**raph

Su, C., Andrew, A., Karagas, M.R. and Borsuk, M.E., 2013. Using Bayesian networks to discover relations between genes, environment, and disease. *BioData mining*, *6*(1), pp.1-21.



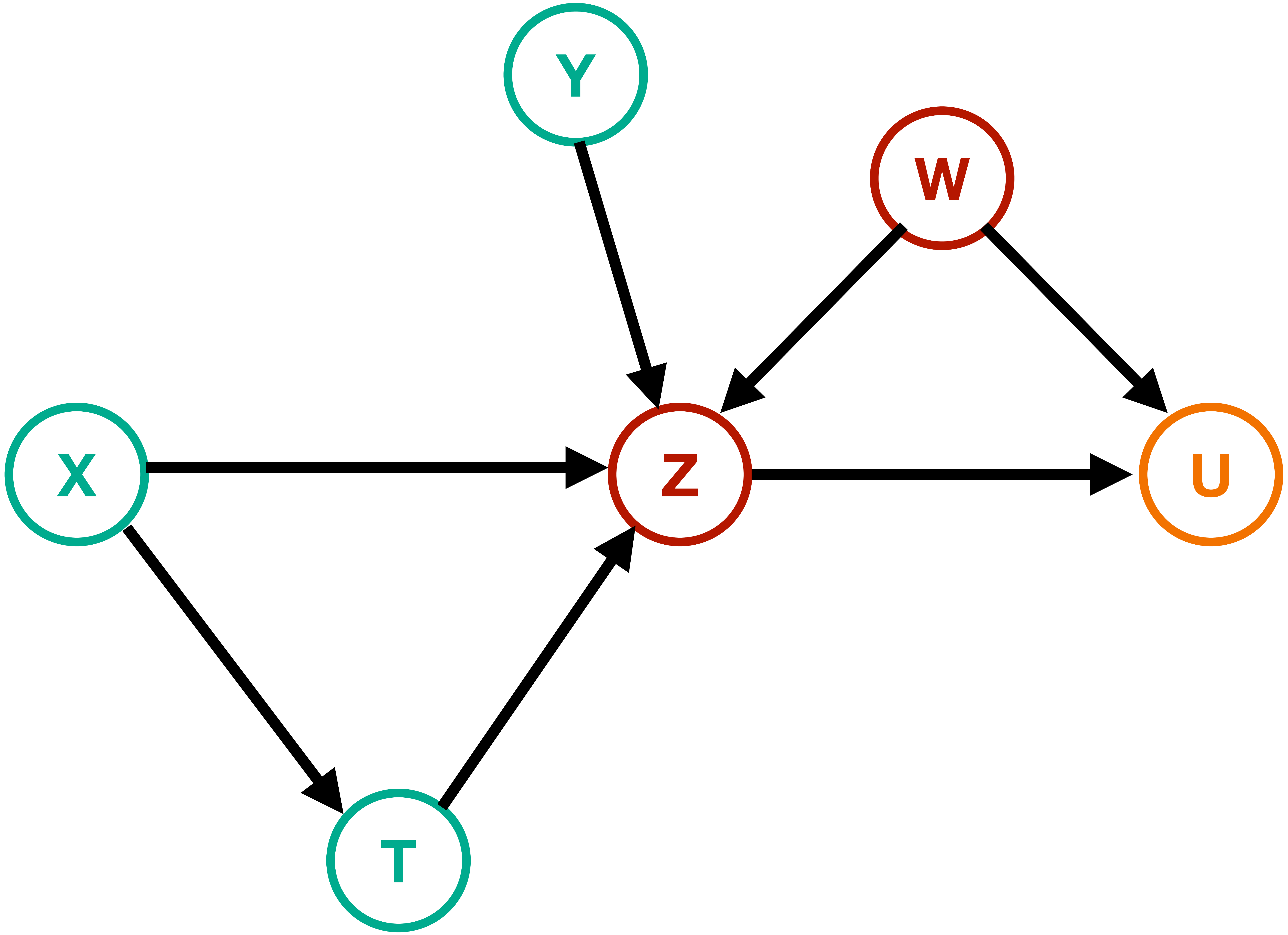
A broad perspective on Graphical Models

- Representation
- Structure learning
- Identifiability
- Estimation
- Model selection



A broad perspective on Graphical Models

- Representation
- Structure learning
- Identifiability
- Estimation
- Model selection



$$U \perp\!\!\!\perp (X, Y, T) \mid (W, Z)$$

Recap of DAG models (Bayesian networks)

$G = ([p], E)$ a DAG

$X_{[p]} = (X_1, \ldots, X_p)$ a vector of discrete random variables

$\mathscr{R} = \prod_{i=1}^{p} [d_i]$, $n = |\mathscr{R}|$, $[d_i]$ = outcome space of X_i

$f = (f_{\mathbf{x}} : \mathbf{x} \in \mathscr{R}), \quad f_{\mathbf{x}} \in \mathbb{R}_{>0} \qquad \Delta^{\circ}_{n-1} = \{f : \sum_{\mathbf{x} \in \mathscr{R}} f_{\mathbf{x}} = 1\},$

Implicit

$\mathscr{M}(G) := \{f \in \Delta^{\circ}_{n-1} : f$ satisfies the local Markov property for $G\}$

Parametric

$:= \{f \in \Delta^{\circ}_{n-1} : f$ satisfies the recursive factorization property w.r.t $G\}$

Example

$$X_1 \to X_2 \to X_3$$

$X_{[3]} = (X_1, X_2, X_3)$ a vector of binary random variables

$\mathscr{R} = \{000, 001, \ldots, 111\}, \; n = |\mathscr{R}| = 8$

$\Delta^{\circ}_{n-1} = \{(p_{000}, \ldots, p_{111}) : \sum_{\mathbf{x} \in \mathscr{R}} p_{\mathbf{x}} = 1\},$

local Markov property for G : $\quad X_3 \perp\!\!\!\perp X_1 \,|\, X_2$

$\mathscr{M}(G) = \{p \in \Delta^0_7 : p_{000}p_{101} - p_{100}p_{001} = p_{010}p_{111} - p_{110}p_{011} = 0\}$

$X_3 \perp\!\!\!\perp X_1 \,|\, X_2 = 0 \qquad\qquad X_3 \perp\!\!\!\perp X_1 \,|\, X_2 = 1$

Implicit

Example

$$X_1 \longrightarrow X_2 \longrightarrow X_3$$

$X_{[3]} = (X_1, X_2, X_3)$ a vector of binary random variables

$$\mathscr{R} = \{000, 001, \ldots, 111\}, \; n = |\mathscr{R}| = 8$$

$$\Delta^{\circ}_{n-1} = \{(p_{000}, \ldots, p_{111}) : \sum_{\mathbf{x} \in \mathscr{R}} p_{\mathbf{x}} = 1\},$$

Conditional Independence Statement

local Markov property for G : $\quad X_3 \perp\!\!\!\perp X_1 \mid X_2$

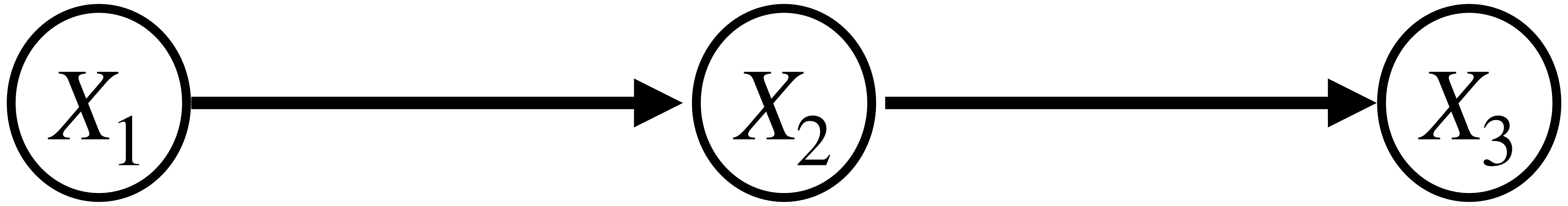
$$\mathscr{M}(G) = \{p \in \Delta^0_7 : p_{000}p_{101} - p_{100}p_{001} = p_{010}p_{111} - p_{110}p_{011} = 0\}$$

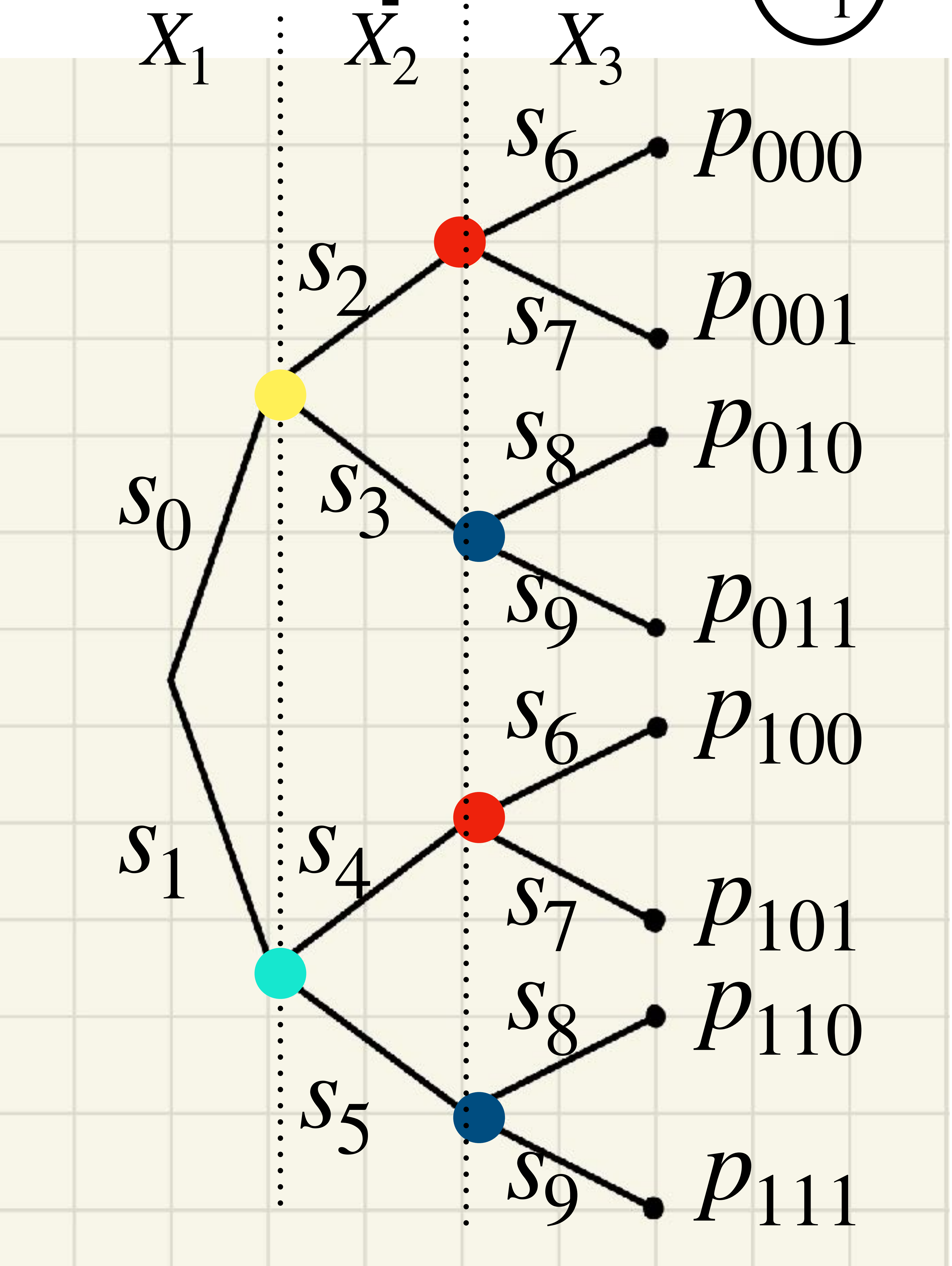
Context-Specific Conditional Independence Statement

$X_3 \perp\!\!\!\perp X_1 \mid X_2 = 0$ $\qquad$ $X_3 \perp\!\!\!\perp X_1 \mid X_2 = 1$

Implicit

Example





$\mathscr{M}(G) := \{(p_{000}, \ldots, p_{111}) \in \Delta_7^\circ : p$ satisfies the recursive factorization property$\}$

$$f(X_{[p]}) = \prod_{i=1}^{p} f(X_i \mid X_{\mathrm{pa}(i)})$$

$$f(X_{[3]}) = f(X_1)f(X_2 \mid X_1)f(X_3 \mid X_2)$$

Parametric

$$p_{000} = s_0 s_2 s_6, \ldots, p_{111} = s_1 s_5 s_9$$

$$s_0 + s_1 = \cdots = s_8 + s_9 = 1$$

Implicit

$$\mathscr{M}(G) = \{p \in \Delta_7^0 : p_{000}p_{101} - p_{100}p_{001} = 0$$
$$p_{010}p_{111} - p_{110}p_{011} = 0\}$$

$X_{[p]} = (X_1, \ldots, X_p)$ vector of discrete random variables

$[d_i] =$ outcome space of X_i

$\mathscr{R} =$ outcome space of $X_{[p]}$

Conditional independence statement: $X_A \perp\!\!\!\perp X_B \,|\, X_C$

$X_{[p]} = (X_1, \ldots, X_p)$ vector of discrete random variables

$[d_i] =$ outcome space of X_i

$\mathscr{R} =$ outcome space of $X_{[p]}$

Conditional independence statement: $X_A \perp\!\!\!\perp X_B \,|\, X_C$

Let A, B, C, S be disjoint subsets of $[p]$

A **and** B **are contextually independent given** S **in the context** $X_C = \mathbf{x}_C$ if

$$f(\mathbf{x}_A \,|\, \mathbf{x}_B, \mathbf{x}_C, \mathbf{x}_S) = f(\mathbf{x}_A \,|\, \mathbf{x}_C, \mathbf{x}_S)$$

for all $(\mathbf{x}_A, \mathbf{x}_B, \mathbf{x}_S) \in \mathscr{R}_A \times \mathscr{R}_B \times \mathscr{R}_S$

$$X_A \perp\!\!\!\perp X_B \,|\, X_S, X_C = \mathbf{x}_C$$

$X_{[p]} = (X_1, \ldots, X_p)$ vector of discrete random variables

$[d_i] =$ outcome space of X_i

$\mathscr{R} =$ outcome space of $X_{[p]}$

Conditional independence statement: $X_A \perp\!\!\!\perp X_B \,|\, X_C$ **CI statement**

Let A, B, C, S be disjoint subsets of $[p]$

A **and** B **are contextually independent given** S **in the context** $X_C = \mathbf{x}_C$ if

$$f(\mathbf{x}_A \,|\, \mathbf{x}_B, \mathbf{x}_C, \mathbf{x}_S) = f(\mathbf{x}_A \,|\, \mathbf{x}_C, \mathbf{x}_S)$$

for all $(\mathbf{x}_A, \mathbf{x}_B, \mathbf{x}_S) \in \mathscr{R}_A \times \mathscr{R}_B \times \mathscr{R}_S$

$X_A \perp\!\!\!\perp X_B \,|\, X_S, X_C = \mathbf{x}_C$ **CSI statement**

$G = ([p], E)$ and A, B, S, C subsets of $[p]$.

X_A is conditionally independent of X_B given X_S in the context $X_C = \mathbf{x}_C$ if

$$f(\mathbf{x}_A \,|\, \mathbf{x}_B, \mathbf{x}_S, \mathbf{x}_C) = f(\mathbf{x}_A \,|\, \mathbf{x}_S, \mathbf{x}_C), \;\; \forall(\mathbf{x}_A, \mathbf{x}_B, \mathbf{x}_S) \in \mathscr{R}_A \times \mathscr{R}_B \times \mathscr{R}_S$$

$$X_A \perp\!\!\!\perp X_B \,|\, X_S, X_C = \mathbf{x}_C$$

<u>Question:</u> How to encode context-specific conditional independence statements in DAG models?

Similarity Networks (Heckerman 1990),
Bayesian Multinets (Geiger, Heckerman 1996),
CPTs with regularity structure (Boutelier et. al. 1996),
Staged Trees (Smith, Anderson 2008),
LDAGS (Pensar et. al. 2015)

Axioms for conditional independence

(i) *(symmetry)* $X_A \perp\!\!\!\perp X_B \,|\, X_C \implies X_B \perp\!\!\!\perp X_A \,|\, X_C$;

(ii) *(decomposition)* $X_A \perp\!\!\!\perp X_{B\cup D} \,|\, X_C \implies X_A \perp\!\!\!\perp X_B \,|\, X_C$;

(iii) *(weak union)* $X_A \perp\!\!\!\perp X_{B\cup D} \,|\, X_C \implies X_A \perp\!\!\!\perp X_B \,|\, X_{C\cup D}$;

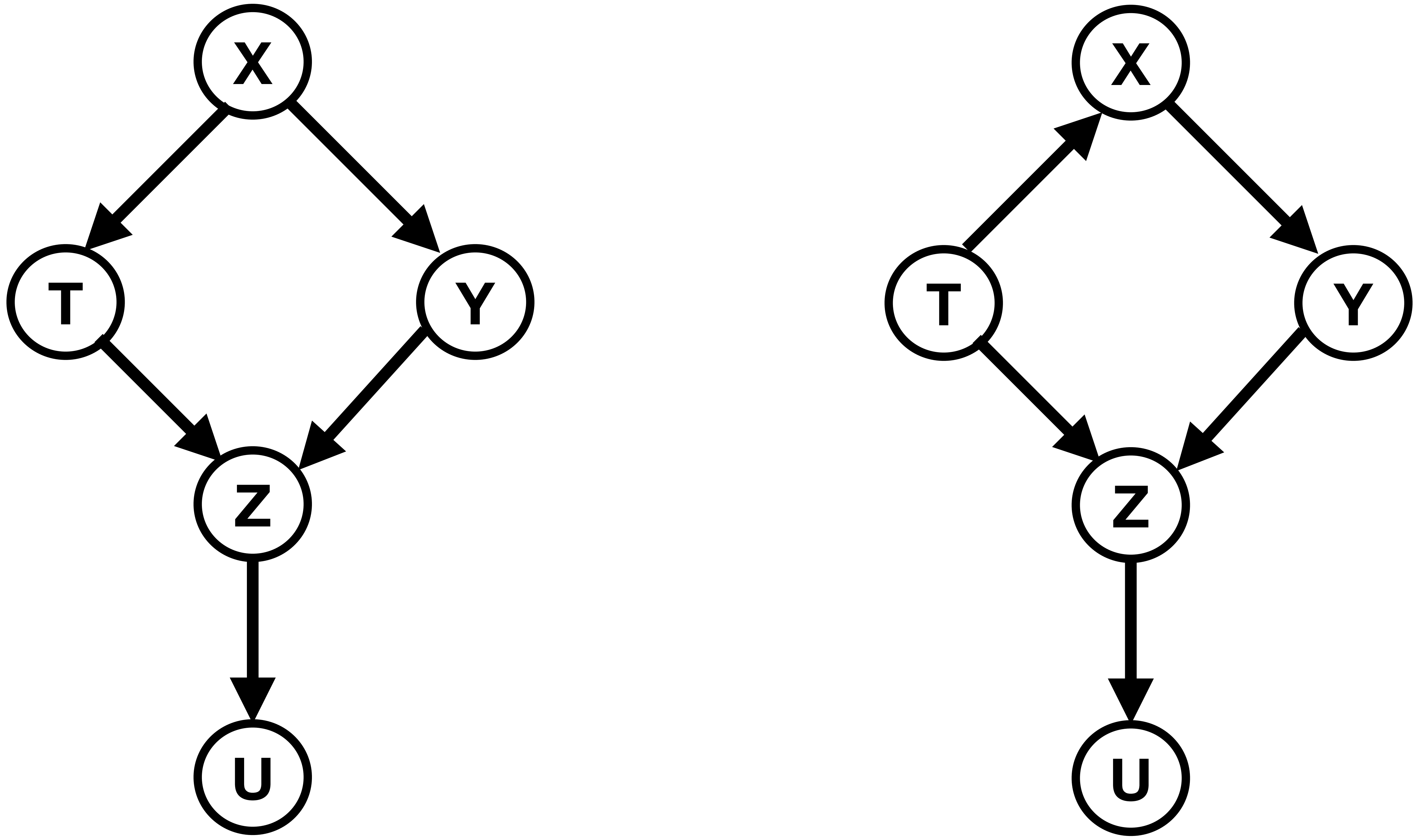
(iv) *(contraction)* $X_A \perp\!\!\!\perp X_B \,|\, X_{C\cup D}$ *and* $X_A \perp\!\!\!\perp X_D \,|\, X_C \implies X_A \perp\!\!\!\perp X_{B\cup D} \,|\, X_C$.

$$\begin{aligned} \mathscr{I}(G) &= \{\text{all CI statements implied by the local Markov property}\} \\ &= \{\text{all d-separation statements in } G\} \end{aligned}$$

<u>Theorem</u> (Verma and Pearl): $\mathscr{I}(G_1) = \mathscr{I}(G_2)$ if and only if G_1 and G_2 have the same skeleton and v-structures.

$$\mathcal{I}(G) = \{\text{all CI statements implied by the local Markov property}\}$$
$$= \{\text{all d-separation statements in } G\}$$

<u>Theorem</u> (Verma and Pearl): $\mathcal{I}(G_1) = \mathcal{I}(G_2)$ if and only if G_1 and G_2 have the same skeleton and v-structures.



$$G = ([p], E) \text{ and } A, B, S, C \text{ subsets of } [p].$$

X_A is conditionally independent of X_B given X_S in the context $X_C = \mathbf{x}_C$ if

$$f(\mathbf{x}_A \mid \mathbf{x}_B, \mathbf{x}_S, \mathbf{x}_C) = f(\mathbf{x}_A \mid \mathbf{x}_S, \mathbf{x}_C), \ \ \forall (\mathbf{x}_A, \mathbf{x}_B, \mathbf{x}_S) \in \mathcal{R}_A \times \mathcal{R}_B \times \mathcal{R}_S$$

$$X_A \perp\!\!\!\perp X_B \mid X_S, X_C = \mathbf{x}_C$$

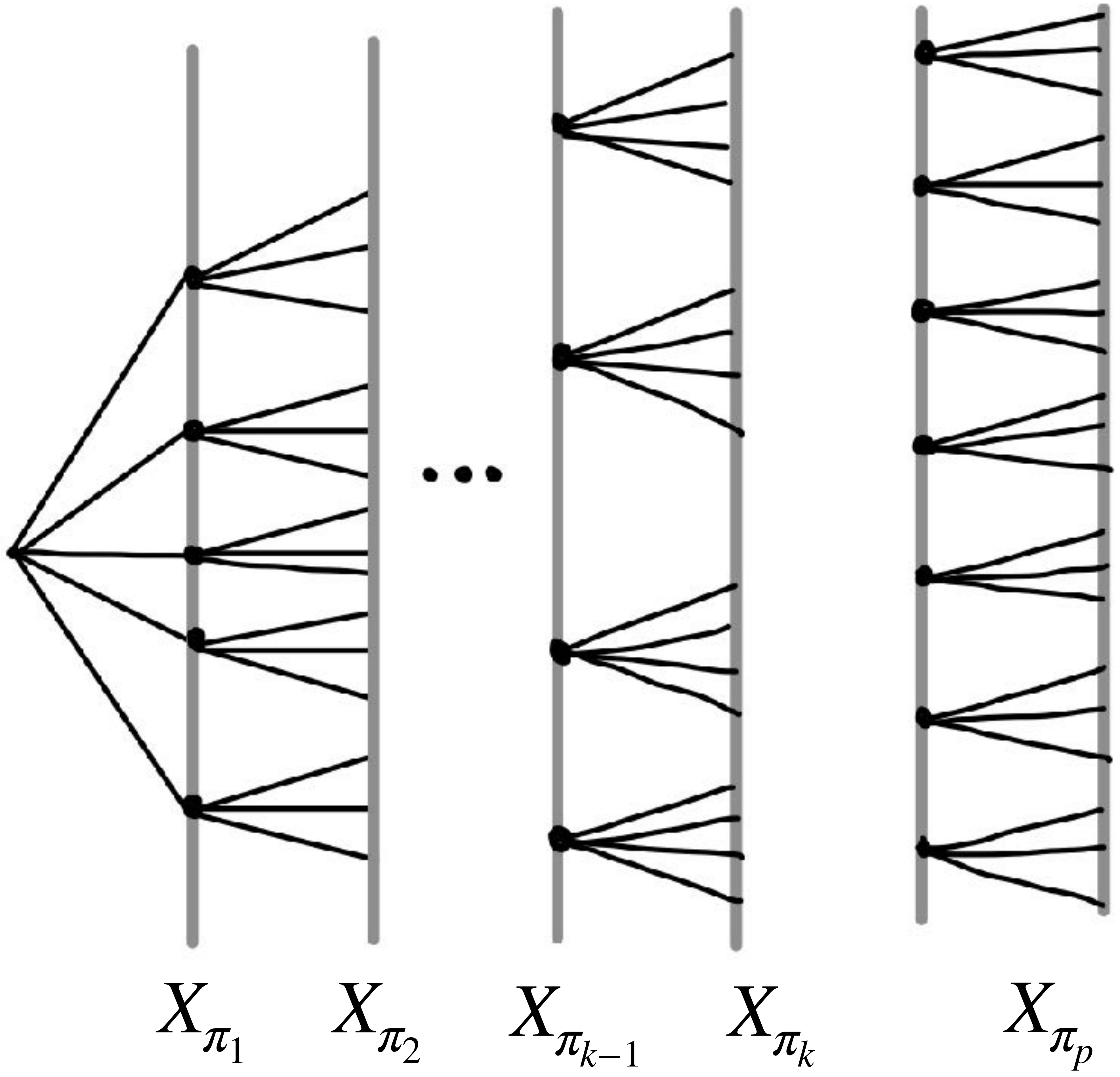
Question: How to encode context-specific conditional independence statements in DAG models?

Similarity Networks (Heckerman 1990),
Bayesian Multinets (Geiger, Heckerman 1996),
CPTs with regularity structure (Boutelier et. al. 1996),
Staged Trees (Smith, Anderson 2008),
LDAGS (Pensar et. al. 2015)

$\pi = \pi_1\pi_2\cdots\pi_p$ an ordering of $[p]$

$\mathscr{T} = (V, E) =$ outcome space of $X_{[p]}$ represented as a sequence of events

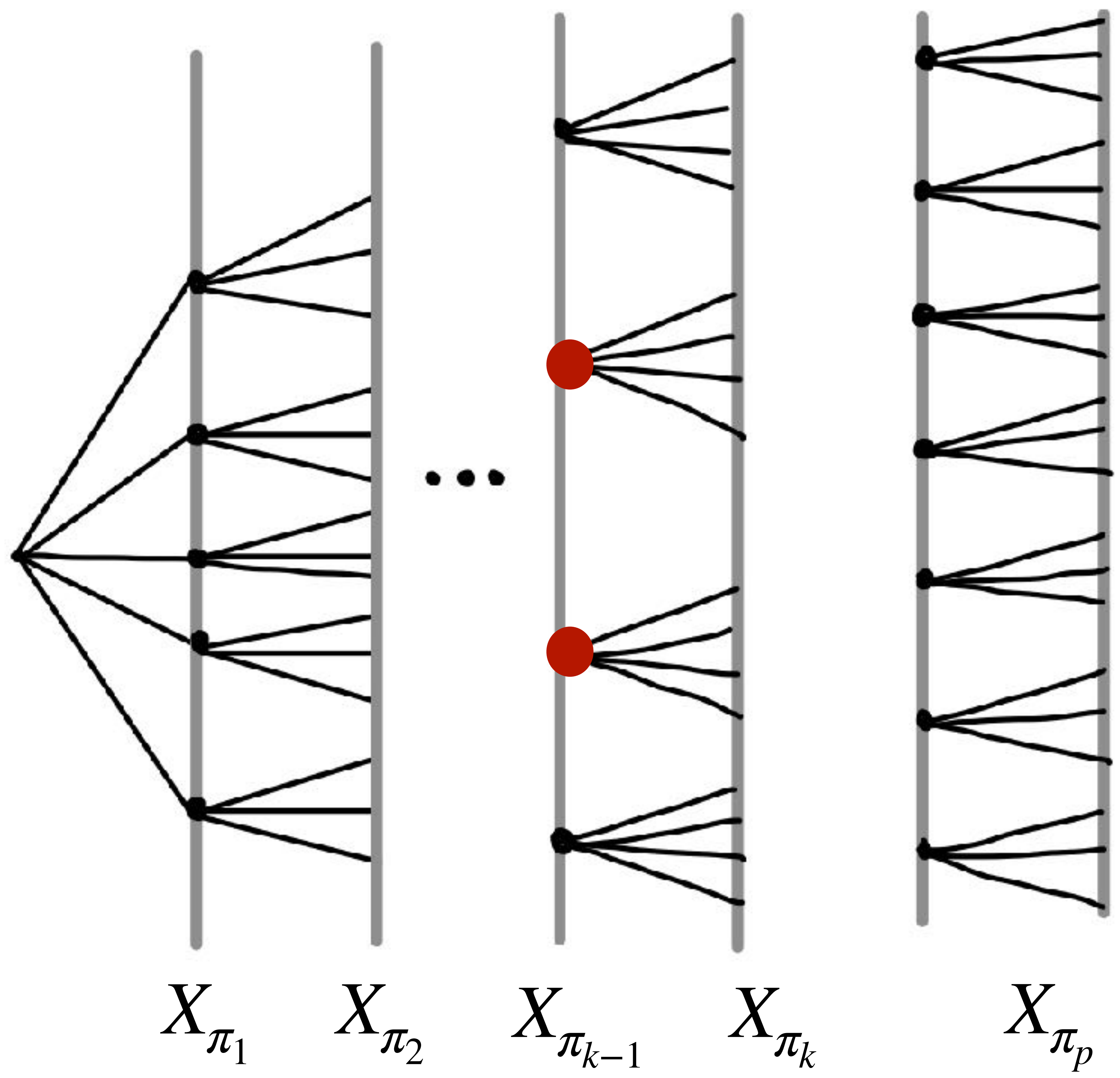
$\mathscr{R} = \{$leaves of $\mathscr{T}\}$



$\pi = \pi_1\pi_2\cdots\pi_p$ an ordering of $[p]$, w.l.o.g $\pi = 12\cdots p$

$\mathscr{T} = (V, E) =$ outcome space of $X_{[p]}$ represented as a sequence of events

$\mathscr{R} = \{$leaves of $\mathscr{T}\}$



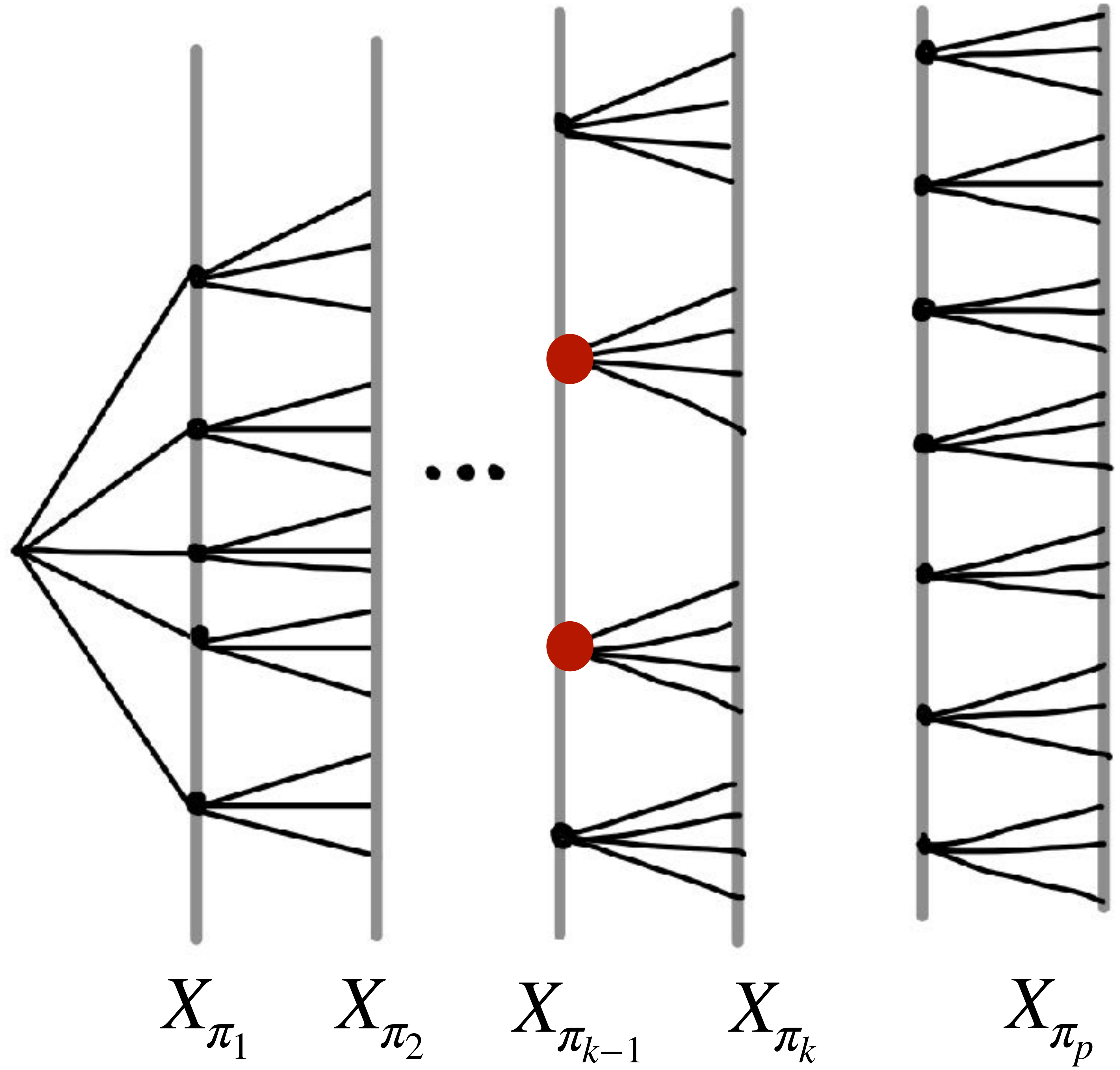
In a **staged tree** you color vertices on the same level to represent conditional distributions that are equal

$x_1\cdots x_{k-1}$ and $y_1\cdots y_{k-1}$ are in the same stage $\Leftrightarrow$

$$f(X_k \mid x_1\cdots x_{k-1}) = f(X_k \mid y_1\cdots y_{k-1})$$

The **staged tree model**:

$$\mathscr{M}(\mathscr{T}) = \{f \in \Delta^{\circ}_{n-1} : f \text{ factor according to } \mathscr{T}\}$$

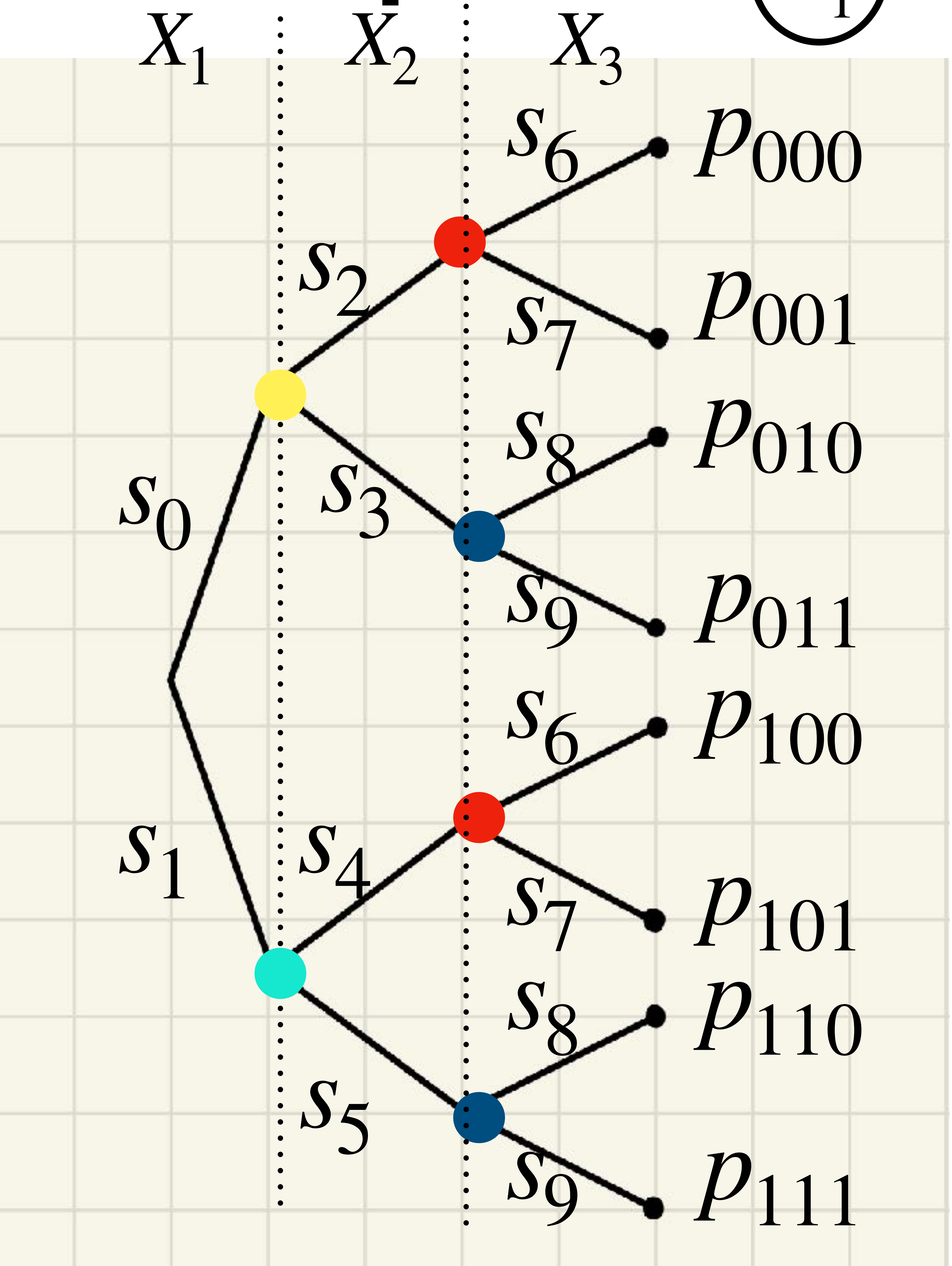


- $\mathscr{M}(\mathscr{T})$ is parametrized by polynomials
- $\mathscr{M}(\mathscr{T})$ is an algebraic variety
- All discrete DAG models are staged trees

Implicit description of staged tree models

Duarte, Görgen, (2020)

Example



$$X_1 \longrightarrow X_2 \longrightarrow X_3$$

$\mathcal{M}(G) := \{(p_{000}, \ldots, p_{111}) \in \Delta_7^\circ : p$ satisfies the recursive factorization property$\}$

$$f(X_{[p]}) = \prod_{i=1}^{p} f(X_i \,|\, X_{\mathrm{pa}(i)})$$

$$f(X_{[3]}) = f(X_1)f(X_2 \,|\, X_1)f(X_3 \,|\, X_2)$$

$$p_{000} = s_0 s_2 s_6, \ldots, p_{111} = s_1 s_5 s_9$$

$$s_0 + s_1 = \cdots = s_8 + s_9 = 1$$

$$\mathcal{M}(G) = \{p \in \Delta_7^0 : p_{000}p_{101} - p_{100}p_{001} = 0$$
$$p_{010}p_{111} - p_{110}p_{011} = 0\}$$

Stages

- $\{0\}$
- $\{1\}$
- $\{00,10\}$
- $\{01,11\}$

A **CStree** is a staged tree $\mathcal{T}$ such that for every stage S there exists a context $\mathbf{x}_C \in \mathcal{R}_C$, $C \subset \{1,2,\ldots,k-1\}$

$$S = \bigcup_{\mathbf{y} \in \mathcal{R}_{\{1,2,\ldots,k-1\} \setminus C}} \{\mathbf{x}_C \mathbf{y}\}$$

The stage S implies the equality

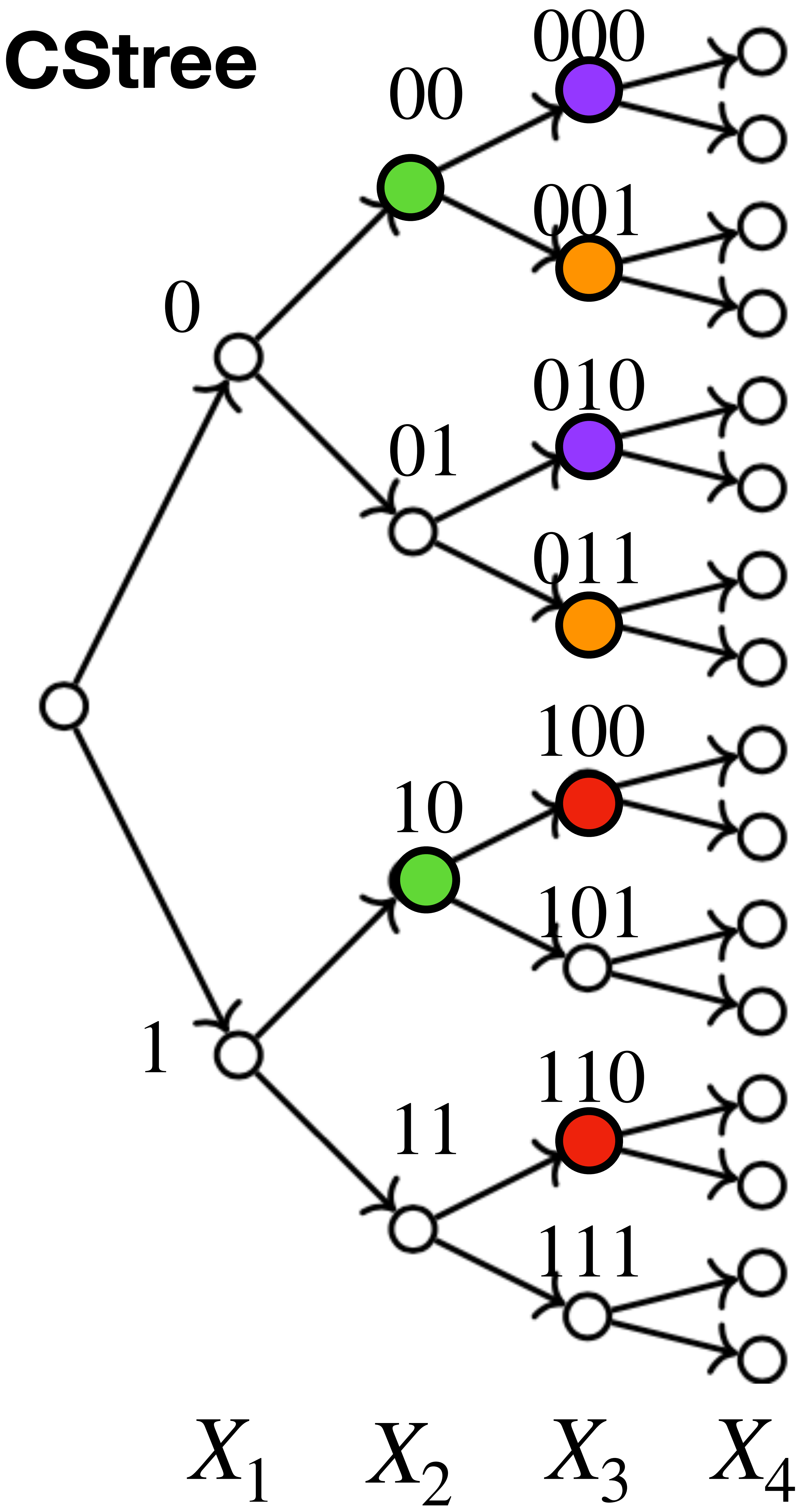
$$f(X_K | \mathbf{x}_C \mathbf{y}) = f(X_k | \mathbf{x}_C \mathbf{y}')$$

for all $\mathbf{y}, \mathbf{y}' \in \mathcal{R}_{\{1,2,\ldots,k-1\} \setminus C}$, which yields

$$X_k \perp\!\!\!\perp X_{\{1,2,\ldots,k-1\} \setminus C} | X_C = \mathbf{x}_C$$

D, Solus, (2022)

CStree



$X_3 \perp\!\!\!\perp X_1 \,|\, X_2 = 0$

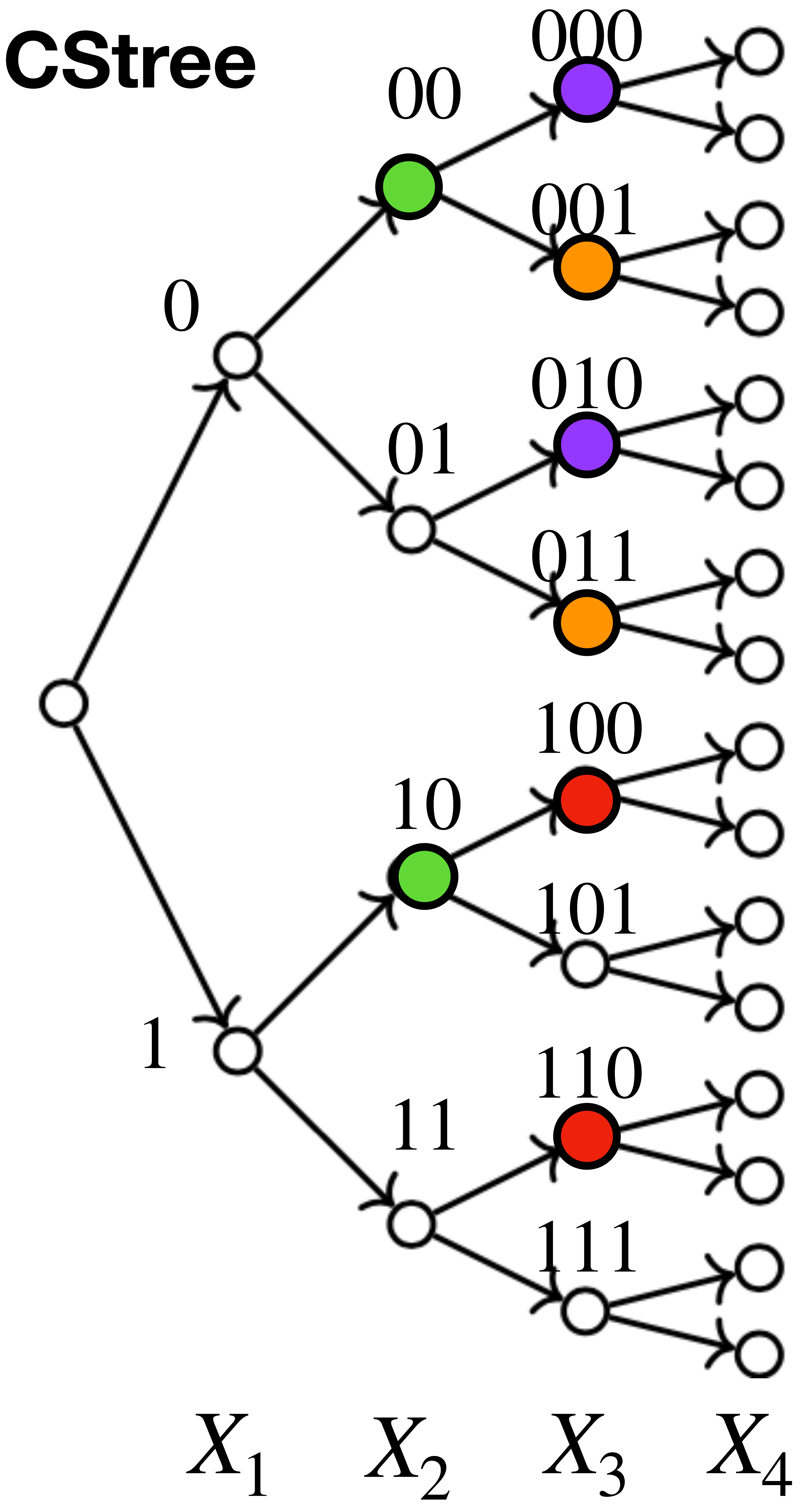
$X_4 \perp\!\!\!\perp X_2 \,|\, X_1 = 0, X_3 = 0$

$X_4 \perp\!\!\!\perp X_2 \,|\, X_1 = 0, X_3 = 1$

$X_4 \perp\!\!\!\perp X_2 \,|\, X_1 = 1, X_3 = 0$

D, Solus, (2022)

CStree



$X_3 \perp\!\!\!\perp X_1 \mid X_2 = 0$

$X_4 \perp\!\!\!\perp X_2 \mid X_1 = 0, X_3 = 0$

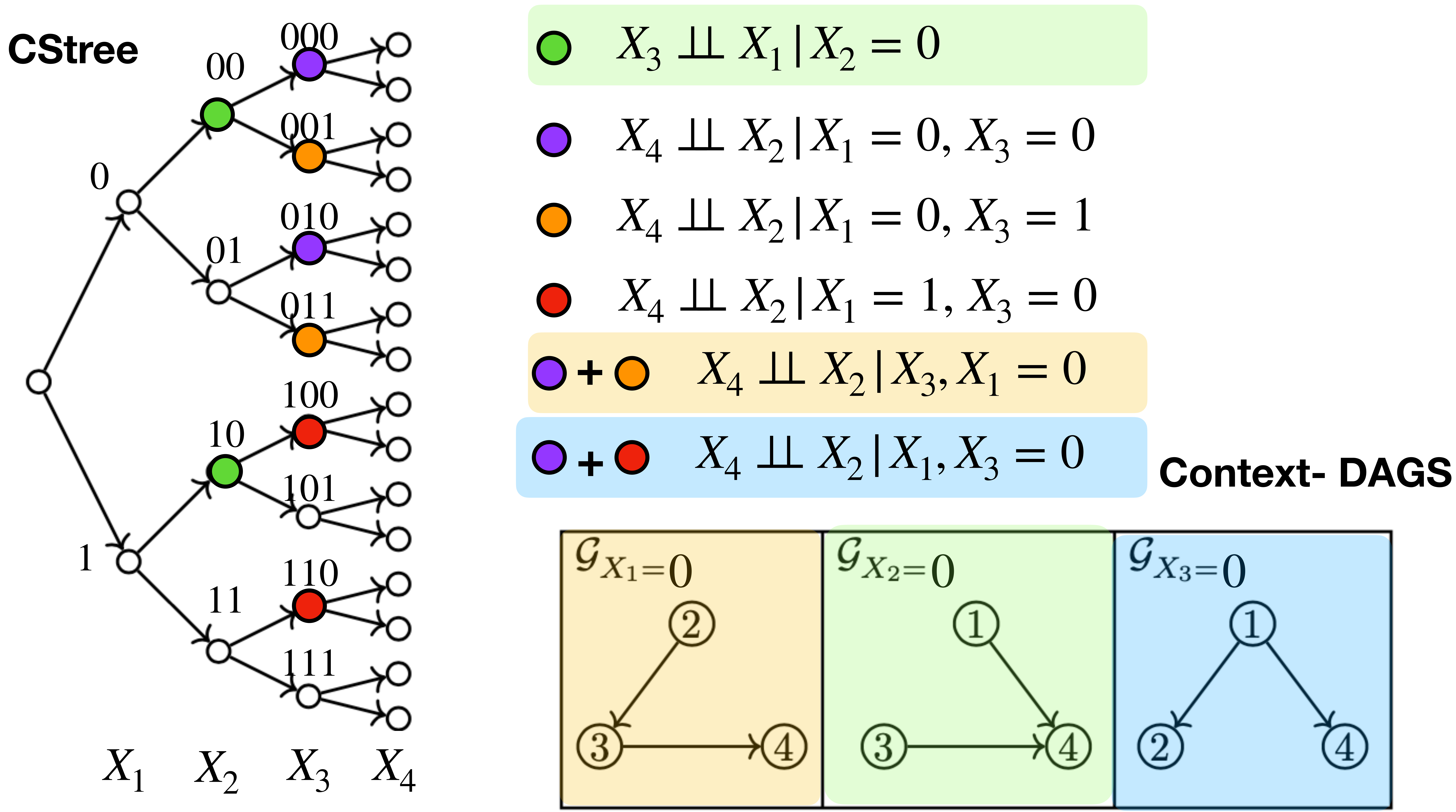
$X_4 \perp\!\!\!\perp X_2 \mid X_1 = 0, X_3 = 1$

$X_4 \perp\!\!\!\perp X_2 \mid X_1 = 1, X_3 = 0$

+ $X_4 \perp\!\!\!\perp X_2 \mid X_3, X_1 = 0$

+ $X_4 \perp\!\!\!\perp X_2 \mid X_1, X_3 = 0$

D, Solus, (2022)



CStree
0
00
01
1
10
11
000
001
010
011
100
101
110
111
X_1 X_2 X_3 X_4
$X_3 \perp\!\!\!\perp X_1 \mid X_2 = 0$
$X_4 \perp\!\!\!\perp X_2 \mid X_1 = 0, X_3 = 0$
$X_4 \perp\!\!\!\perp X_2 \mid X_1 = 0, X_3 = 1$
$X_4 \perp\!\!\!\perp X_2 \mid X_1 = 1, X_3 = 0$
+ $X_4 \perp\!\!\!\perp X_2 \mid X_3, X_1 = 0$
+ $X_4 \perp\!\!\!\perp X_2 \mid X_1, X_3 = 0$
Context- DAGS
$\mathcal{G}_{X_1=0}$
$\mathcal{G}_{X_2=0}$
$\mathcal{G}_{X_3=0}$

(1) *symmetry.* If $\langle A, B \mid S, X_C = \mathbf{x}_C \rangle \in \mathcal{J}$ then $\langle B, A \mid S, X_C = \mathbf{x}_C \rangle \in \mathcal{J}$.

(2) *decomposition.* If $\langle A, B \cup D \mid S, X_C = \mathbf{x}_C \rangle \in \mathcal{J}$ then $\langle A, B \mid S, X_C = \mathbf{x}_C \rangle \in \mathcal{J}$.

(3) *weak union.* If $\langle A, B \cup D \mid S, X_C = \mathbf{x}_C \rangle \in \mathcal{J}$ then $\langle A, B \mid S \cup D, X_C = \mathbf{x}_C \rangle \in \mathcal{J}$.

(4) *contraction.* If $\langle A, B \mid S \cup D, X_C = \mathbf{x}_C \rangle \in \mathcal{J}$ and $\langle A, D \mid S, X_C = \mathbf{x}_C \rangle \in \mathcal{J}$ then $\langle A, B \cup D \mid S, X_C = \mathbf{x}_C \rangle \in \mathcal{J}$.

(5) *intersection.* If $\langle A, B \mid S \cup D, X_C = \mathbf{x}_C \rangle \in \mathcal{J}$ and $\langle A, S \mid B \cup D, X_C = \mathbf{x}_C \rangle \in \mathcal{J}$ then $\langle A, B \cup S \mid D, X_C = \mathbf{x}_C \rangle \in \mathcal{J}$.

(6) *specialization.* If $\langle A, B \mid S, X_C = \mathbf{x}_C \rangle \in \mathcal{J}$, $T \subseteq S$ and $\mathbf{x}_T \in \mathcal{R}_T$, then $\langle A, B \mid S \setminus T, X_{T \cup C} = \mathbf{x}_{T \cup C} \rangle \in \mathcal{J}$.

(7) *absorption.* If $\langle A, B \mid S, X_C = \mathbf{x}_C \rangle \in \mathcal{J}$, $T \subseteq C$ for which $\langle A, B \mid S, X_{C \setminus T} = \mathbf{x}_{C \setminus T}, X_T = \mathbf{x}_T \rangle \in \mathcal{J}$ for all $\mathbf{x}_T \in \mathcal{R}_T$, then $\langle A, B \mid S \cup T, X_{C \setminus T} = \mathbf{x}_{C \setminus T} \rangle \in \mathcal{J}$.

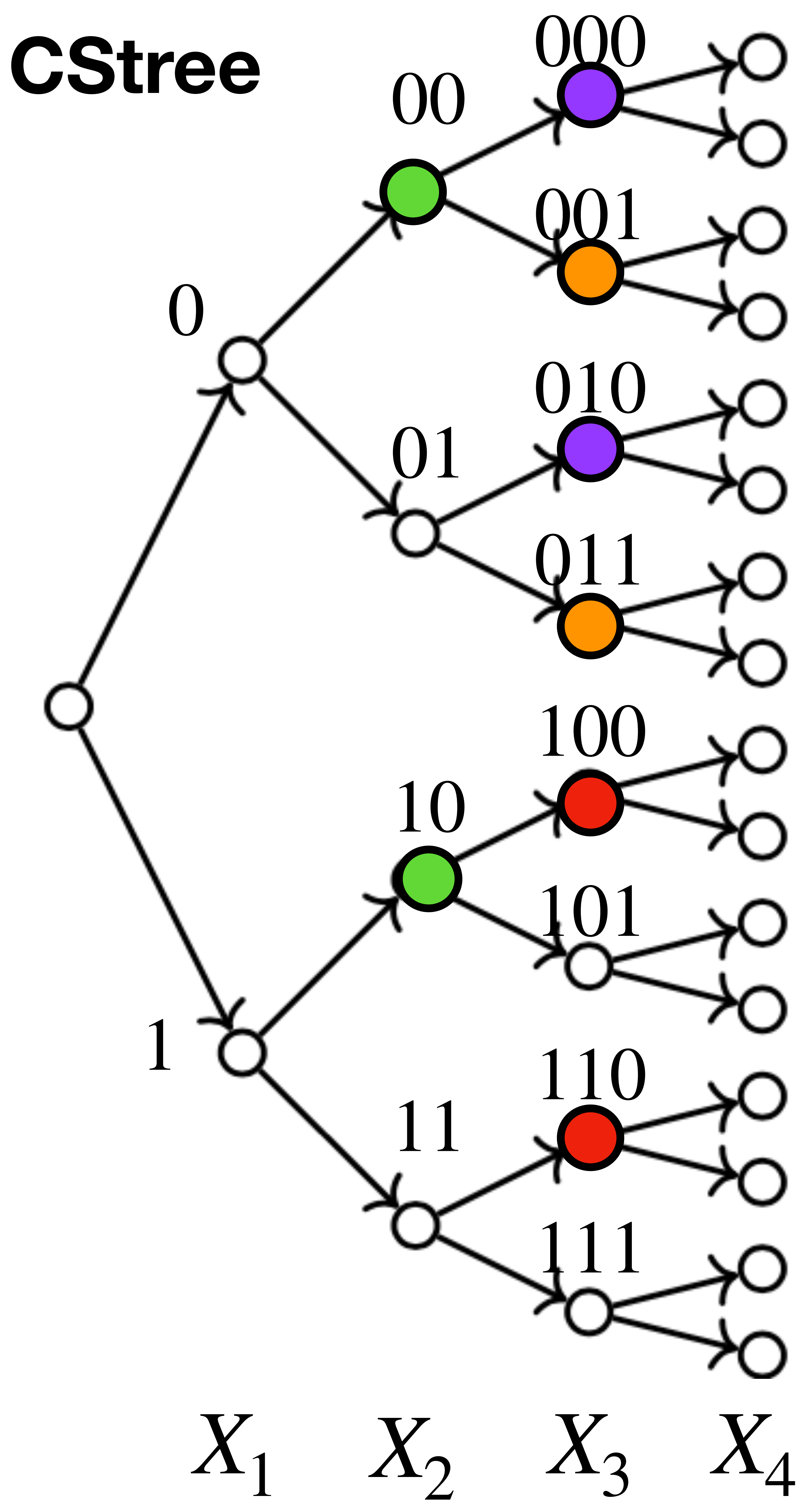
$$\mathcal{J}(\mathcal{T}) = \{ \text{ all CSI statements implied by } \mathcal{T} \}$$

Absorption $\Rightarrow$ $\mathscr{C}_{\mathcal{T}} = \{ \text{ minimal contexts } \} = \{X_c = \mathbf{x}_C\}$

$$\mathcal{J}(\mathcal{T}) = \bigcup_{X_C = \mathbf{x}_C} \mathcal{J}_{X_C = \mathbf{x}_C} \Rightarrow$$

Context DAGS

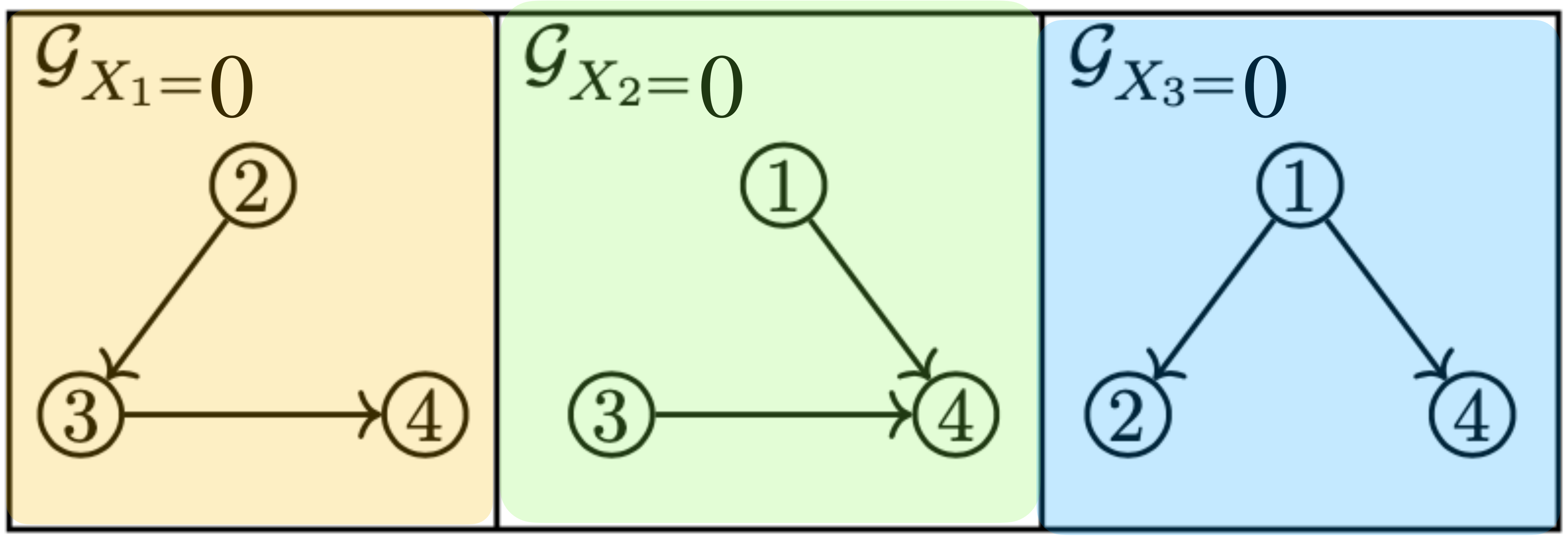
$$\{G_{X_C = \mathbf{x}_C} : X_C = \mathbf{x}_c \in \mathscr{C}_{\mathcal{T}}\}$$

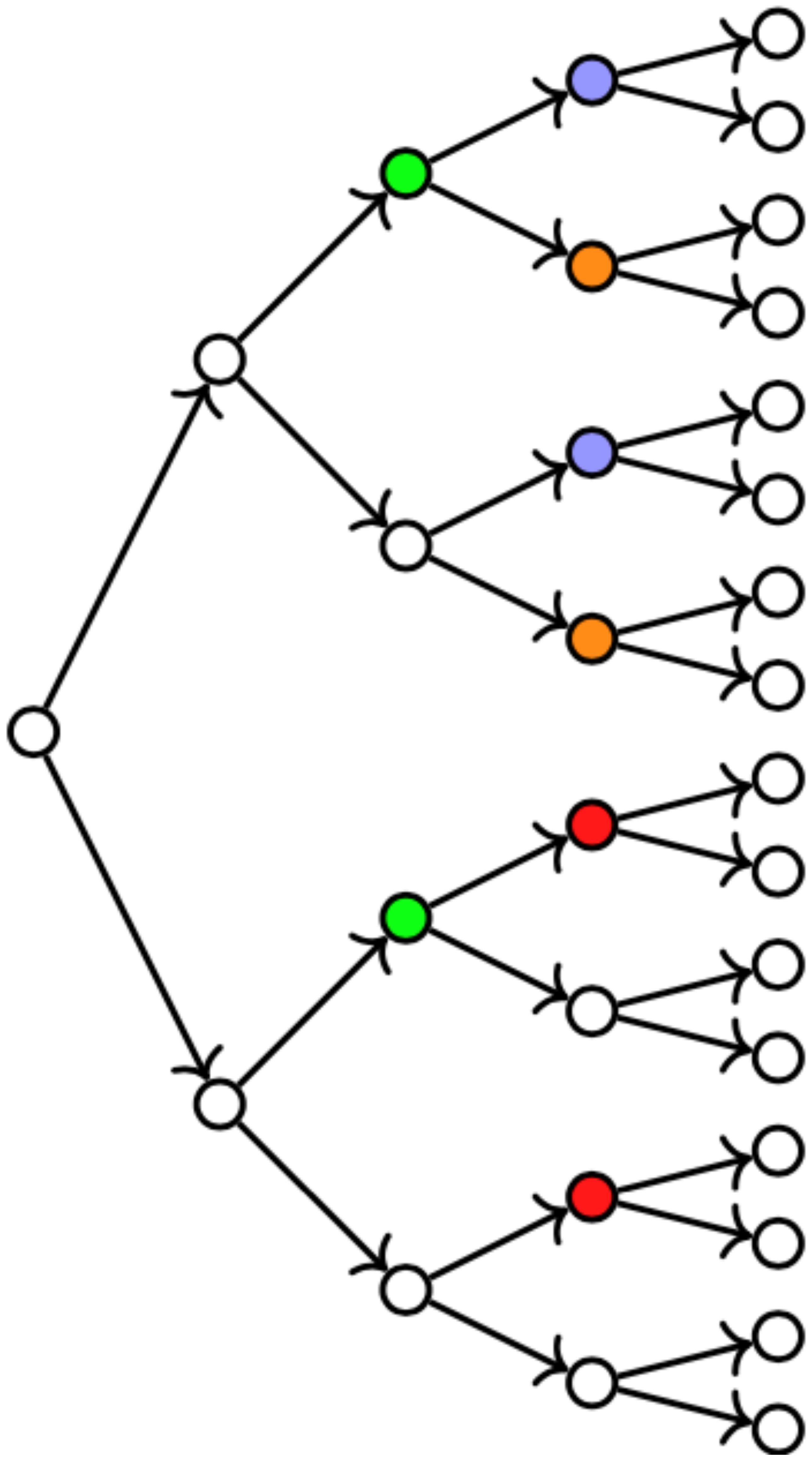


CStrees

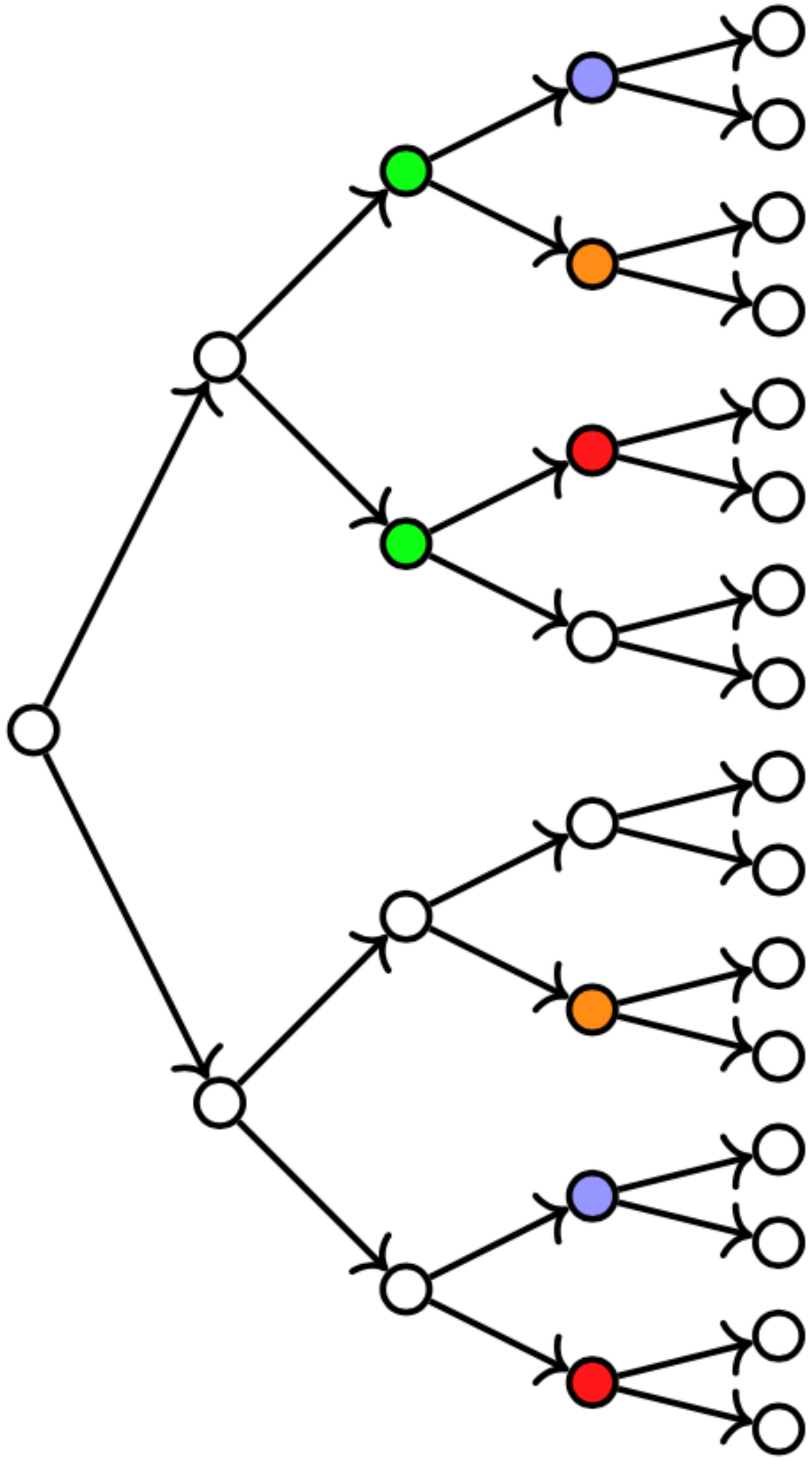
$\Longleftrightarrow$

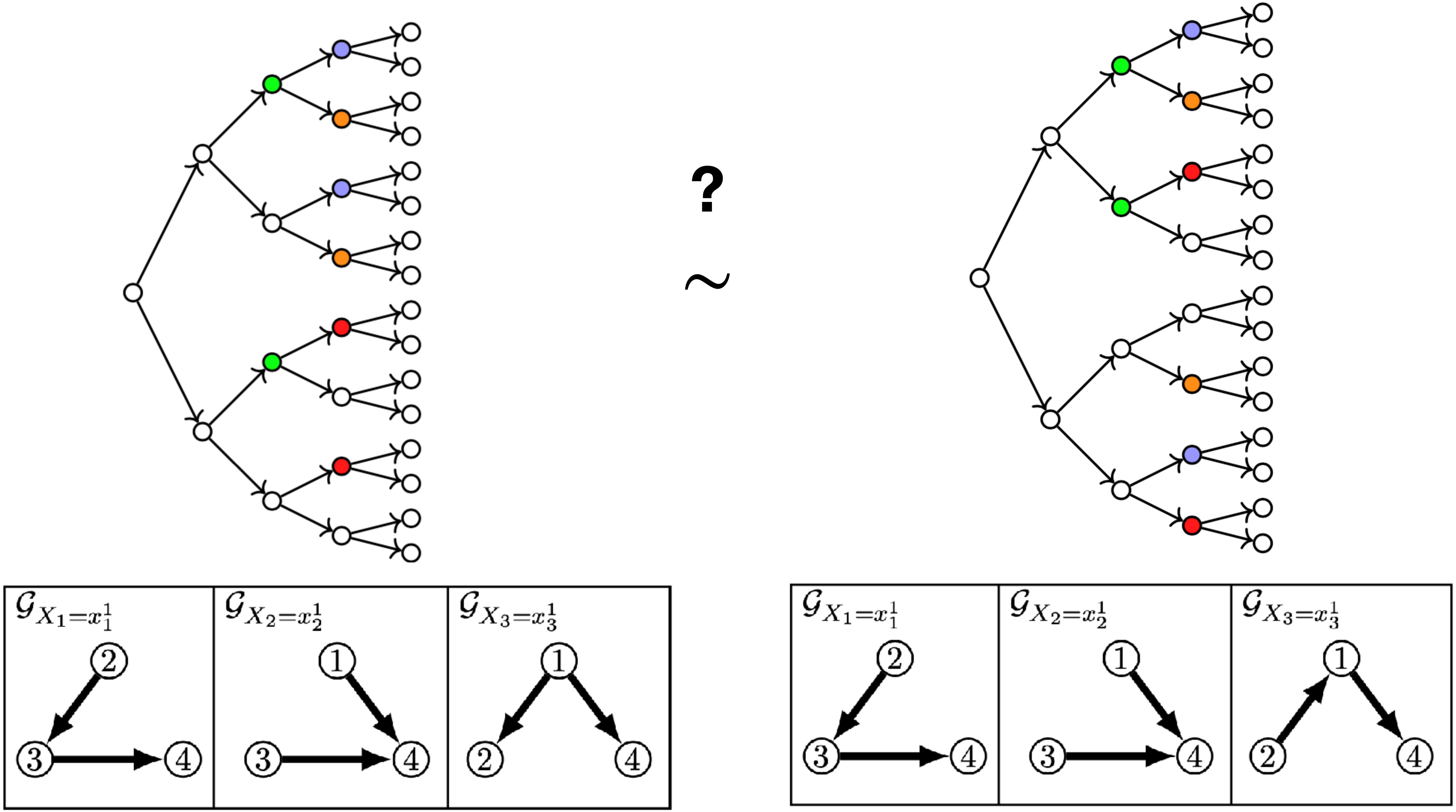
Context DAGs



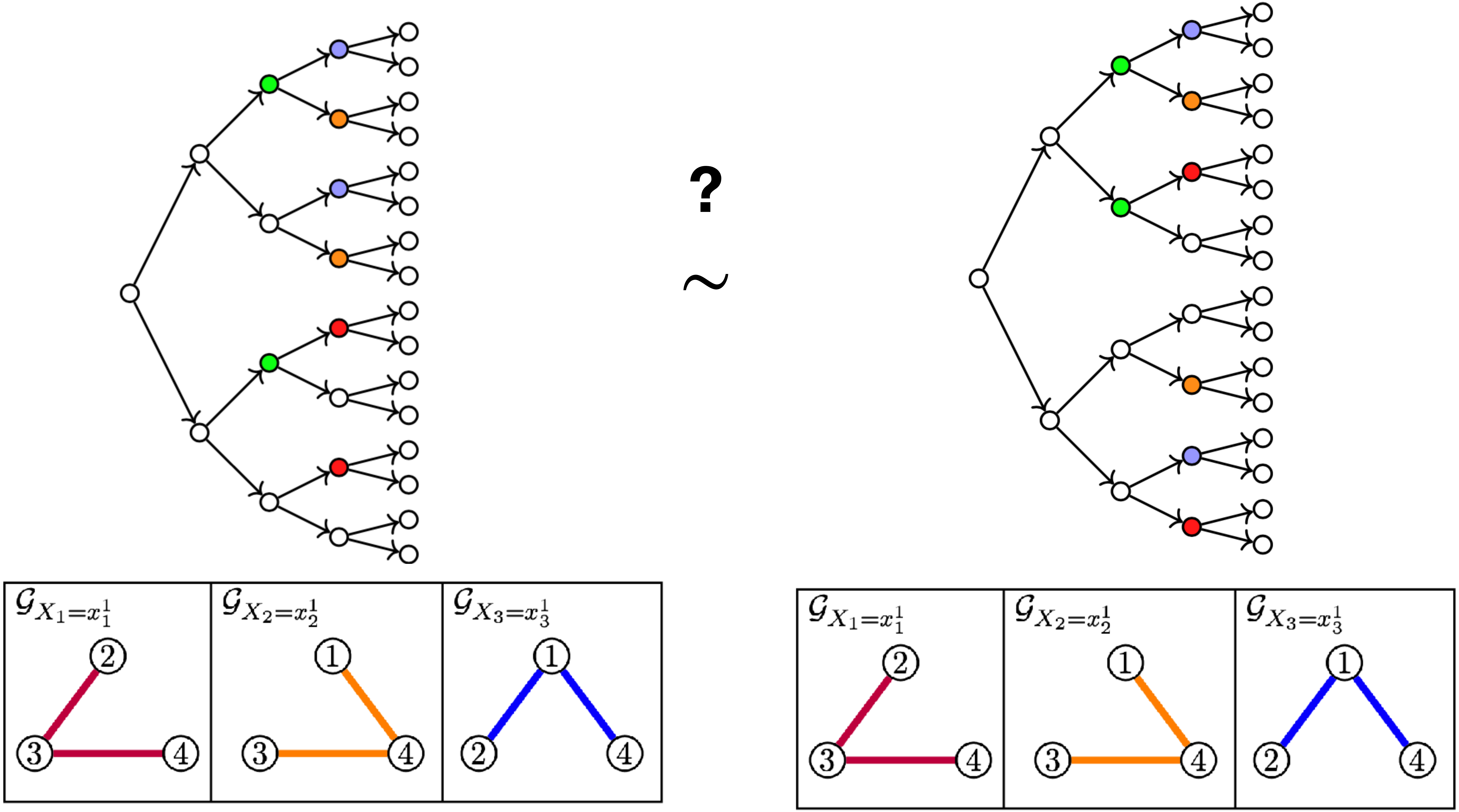


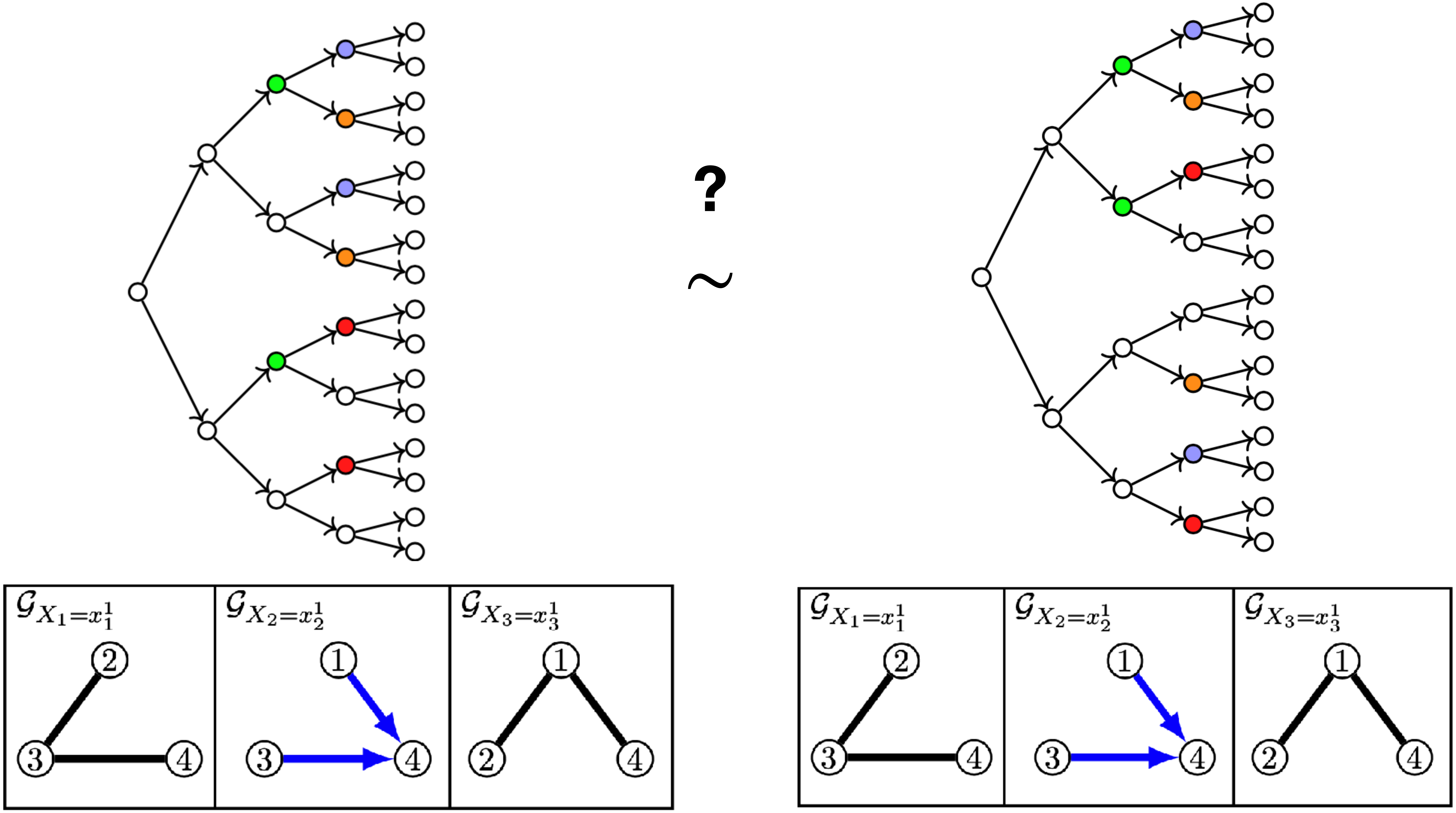
$\overset{?}{\sim}$





?
~
$\mathcal{G}_{X_1=x_1^1}$
$\mathcal{G}_{X_2=x_2^1}$
$\mathcal{G}_{X_3=x_3^1}$
$\mathcal{G}_{X_1=x_1^1}$
$\mathcal{G}_{X_2=x_2^1}$
$\mathcal{G}_{X_3=x_3^1}$


?
∼
$\mathcal{G}_{X_1=x_1^1}$
$\mathcal{G}_{X_2=x_2^1}$
$\mathcal{G}_{X_3=x_3^1}$
1
2
3
4
$\mathcal{G}_{X_1=x_1^1}$
$\mathcal{G}_{X_2=x_2^1}$
$\mathcal{G}_{X_3=x_3^1}$



?
$\sim$
$\mathcal{G}_{X_1=x_1^1}$
$\mathcal{G}_{X_2=x_2^1}$
$\mathcal{G}_{X_3=x_3^1}$
$\mathcal{G}_{X_1=x_1^1}$
$\mathcal{G}_{X_2=x_2^1}$
$\mathcal{G}_{X_3=x_3^1}$

<u>Theorem </u>(D,Solus 2021): Let $\mathscr{T}, \mathscr{T}'$ be two CStrees. The two CStrees encode the same CSI statements, $\mathscr{T} = \mathscr{T}' \quad \Leftrightarrow$ their minimal contexts are equal $\mathscr{C}_{\mathscr{T}} = \mathscr{C}_{\mathscr{T}'}$ and for each minimal context $X_C = \mathbf{x}_C \in \mathscr{C}_{\mathscr{T}}$ the context DAGs $G_{X_C=\mathbf{x}_C}, G'_{X_C=\mathbf{x}_C}$ have the same skeleton and v-structures.

Question: How to encode context-specific conditional independence statements in DAG models?

Similarity Networks (Heckerman 1990),
Bayesian Multinets (Geiger, Heckerman 1996),
CPTs with regularity structure (Boutelier et. al. 1996),
Staged Trees (Smith, Anderson 2008),
LDAGS (Pensar et. al. 2015)

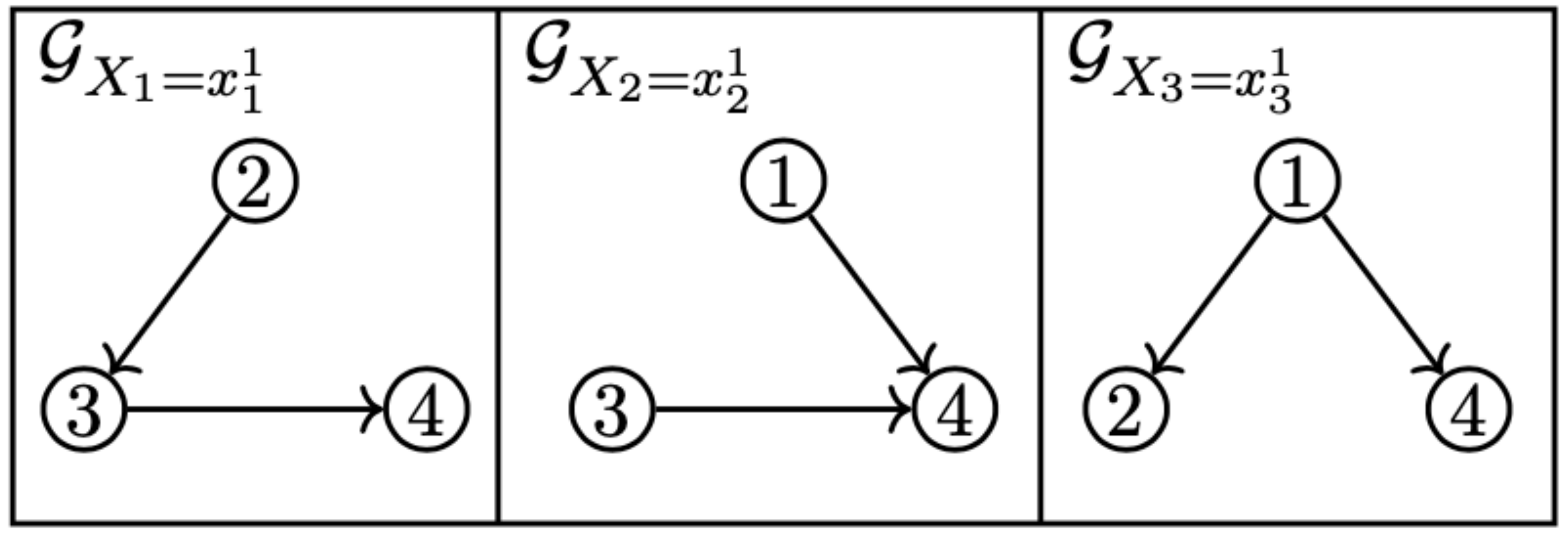
- CStrees are a subclass of LDAGs and of Staged Trees
- LDAGs and Staged Trees are too general, determining model equivalence is difficult.
- CStrees are the first to model interventions in the context-specific setting.

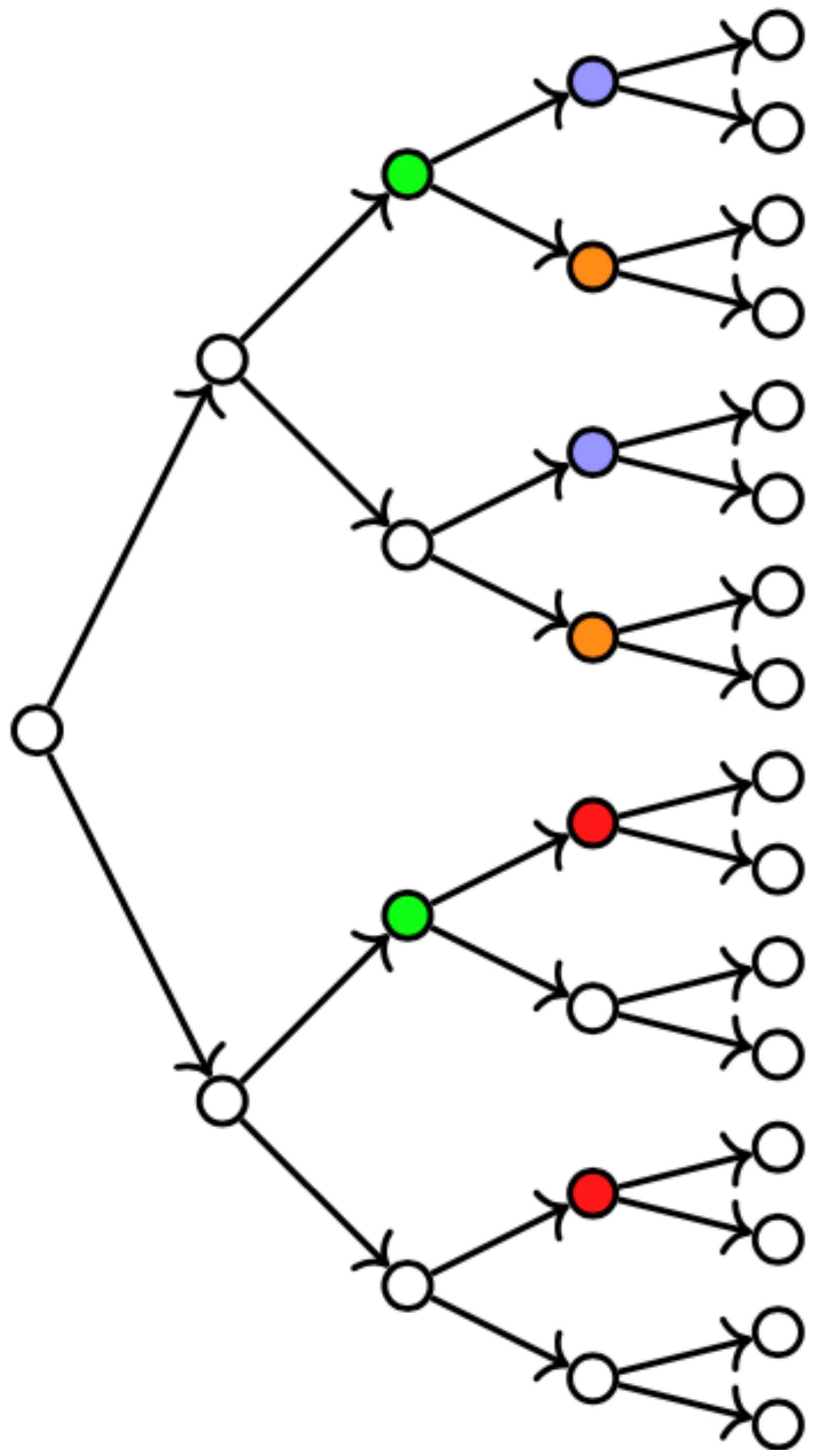
Question: How to encode context-specific conditional independence statements in DAG models?

Similarity Networks (Heckerman 1990),
Bayesian Multinets (Geiger, Heckerman 1996),
CPTs with regularity structure (Boutelier et. al. 1996),
Staged Trees (Smith, Anderson 2008),
LDAGS (Pensar et. al. 2015)

- A similar result extends the results on DAGs with soft interventions to soft interventions in CStrees
- Learning CStrees https://cstrees.readthedocs.io/en/latest/index.html
- R package for staged trees Varando, Leonelli https://cran.r-project.org/web/packages/stagedtrees/index.html

Eliana Duarte
Universidade do Porto
Portugal





Liam Solus
KTH
Sweden

Representation of Context-Specific Causal Models with Observational and Interventional Data

https://arxiv.org/abs/2101.09271