Title: Causal Inference Lecture - 230315

Speakers: Robert Spekkens

Collection: Causal Inference: Classical and Quantum

Date: March 15, 2023 - 10:00 AM

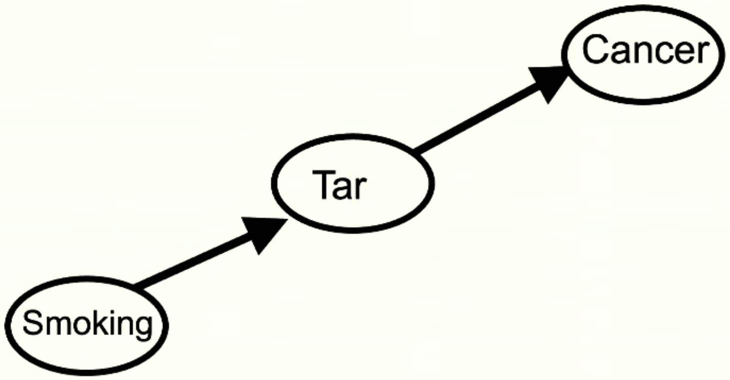
URL: https://pirsa.org/23030072

Abstract: zoom link: https://pitp.zoom.us/j/94143784665?pwd=VFJpajVIMEtvYmRabFYzYnNRSVAvZz09

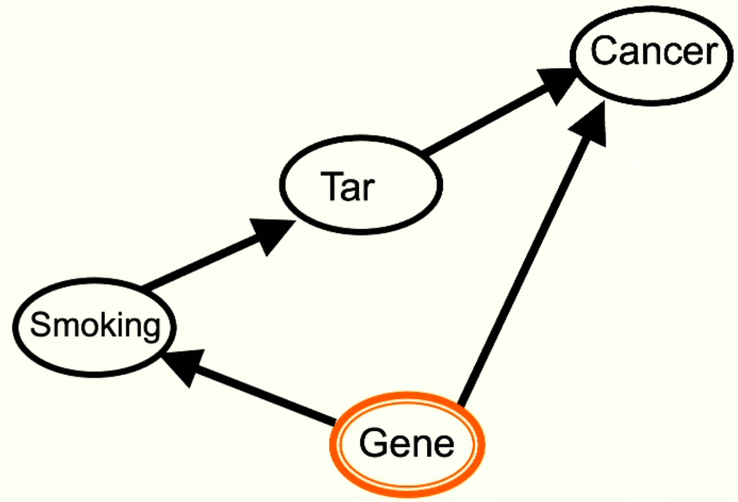
All the equality constraints for causal models

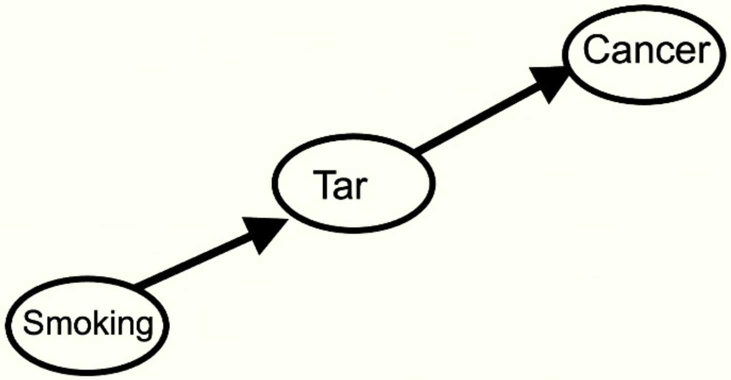
Causal structure ⇨ Constraint on observed probability distribution

Falsification of causal structure ⇦ Violation of constraint on observed probability distribution

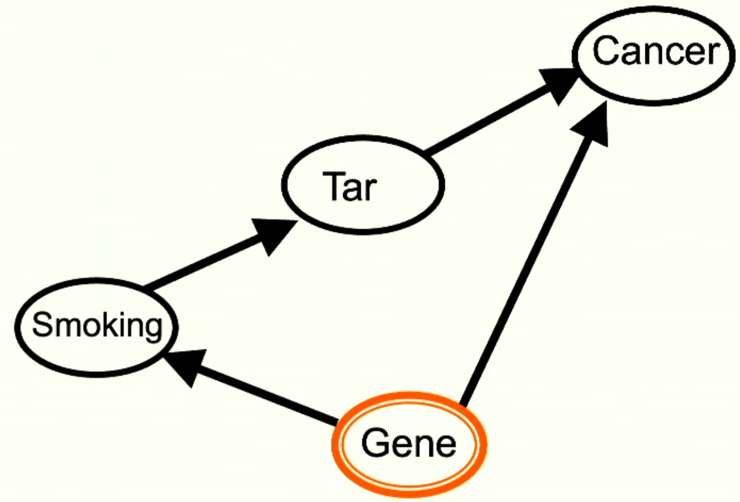


Vs.





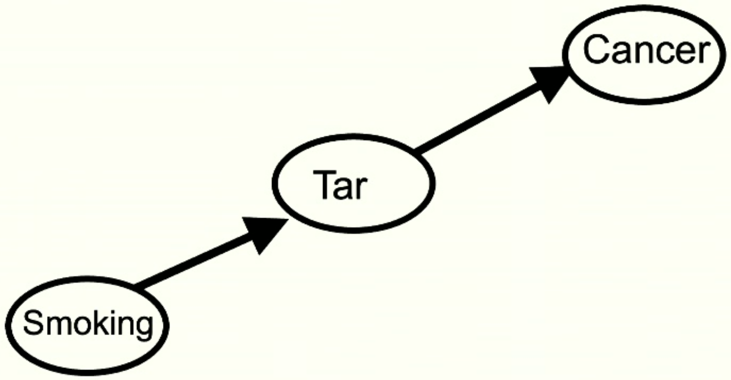
Vs.



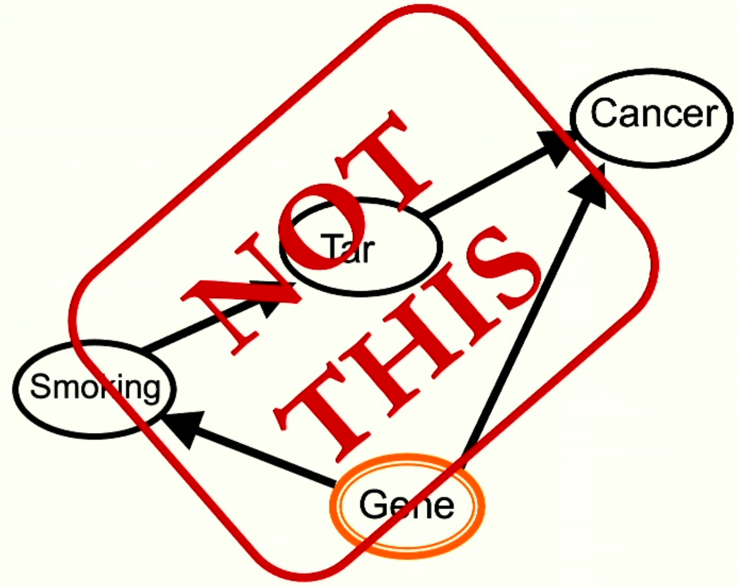
Observe P_{STC} such that

$$S \perp C | T$$

A distinction among interventional probing schemes



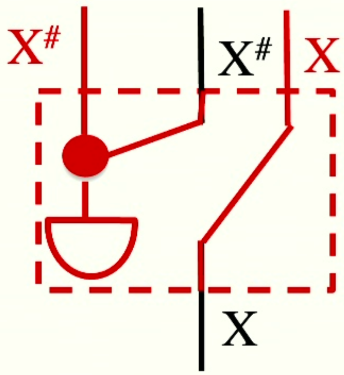
Vs.



Observe P_{STC} such that

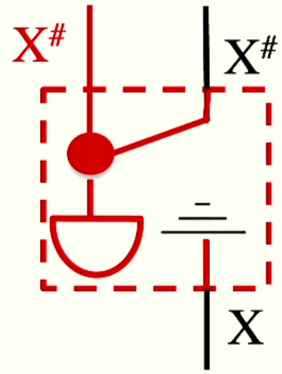
$$S \perp C | T$$

Split-node intervention



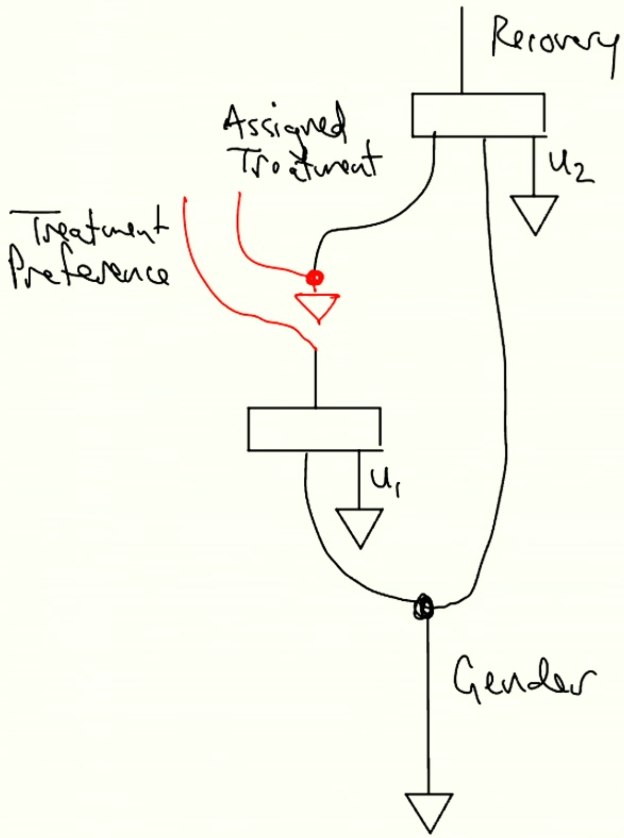
Observe and reprepare new random version with flag

Intervention

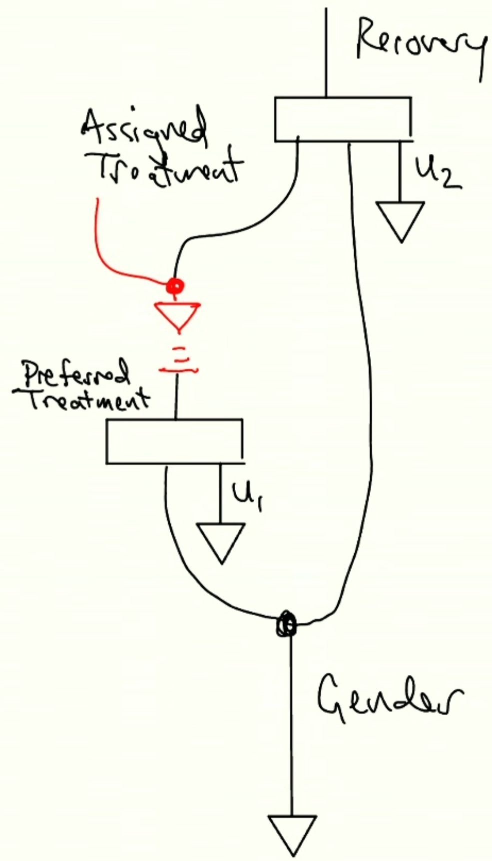


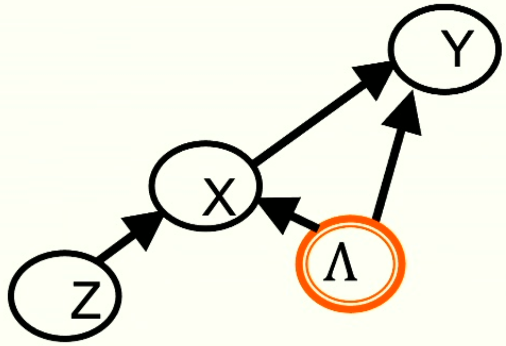
Marginalize and reprepare new version with flag

Split-node intervention



Plain intervention





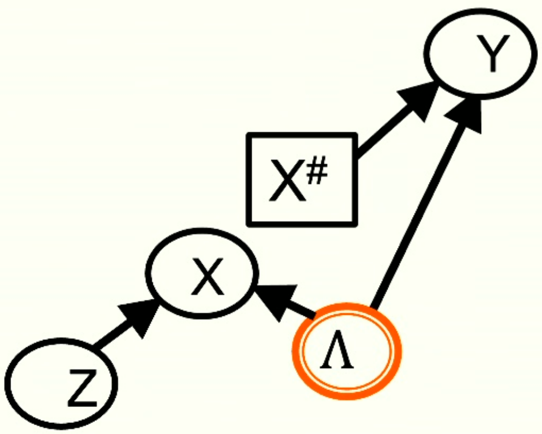
$P_{X|\Lambda Z}$

$P_{Y|\Lambda X}$

P_{Λ}

P_Z

Split-node intervention



$P_{Y|\Lambda X\#}$

$P_{X\#}$

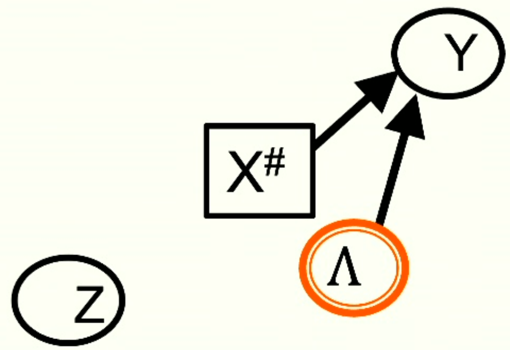
$P_{X|\Lambda Z}$

P_{Λ}

P_Z

$$P_{YX\#XZ} = \sum_{\Lambda} P_{Y|\Lambda X\#} P_{X\#} P_{X|\Lambda Z} P_{\Lambda} P_Z$$

Plain intervention



$P_{Y|\Lambda X\#}$

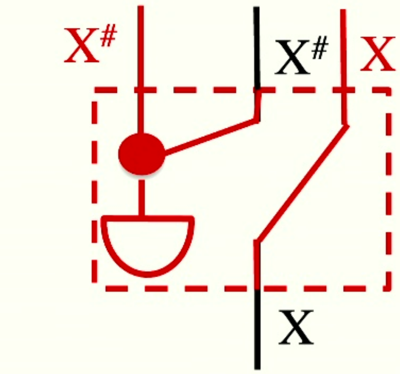
$P_{X\#}$

P_{Λ}

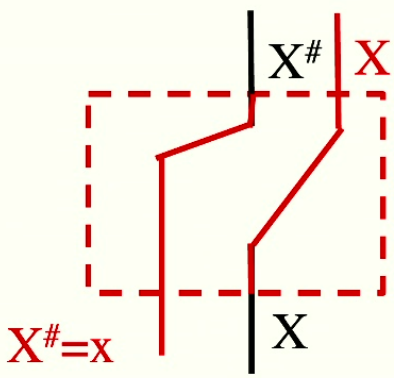
P_Z

$$P_{YX\#Z} = \sum_{\Lambda} P_{Y|\Lambda X\#} P_{X\#} P_{\Lambda} P_Z$$

Split-node intervention

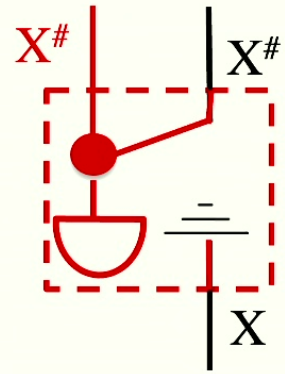


Observe and reprepare new random version with flag

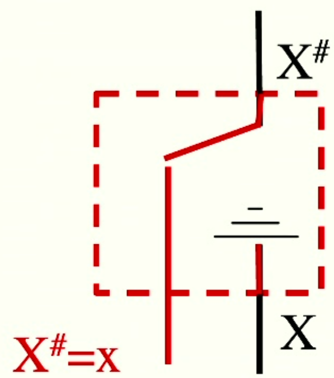


Observe and reprepare with fixed value x

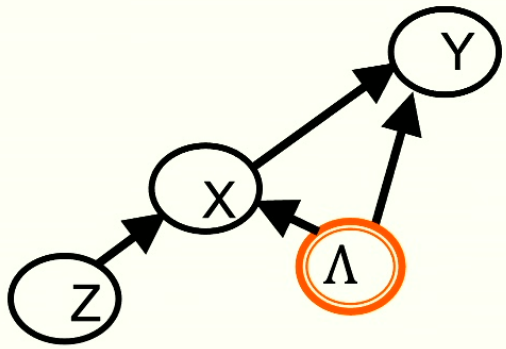
Intervention



Marginalize and reprepare new version with flag

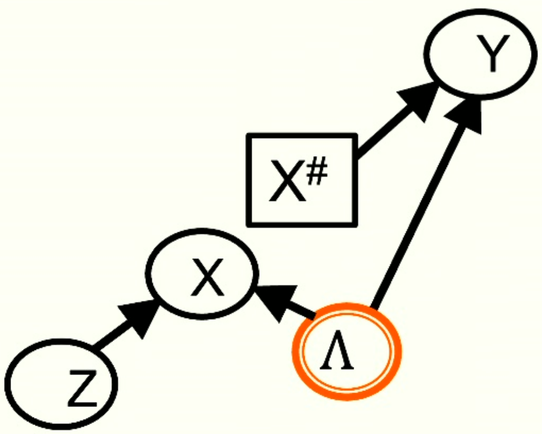


Marginalize and reprepare with fixed value x



$P_{X|\Lambda Z}$
$P_{Y|\Lambda X}$
P_{Λ}
P_{Z}

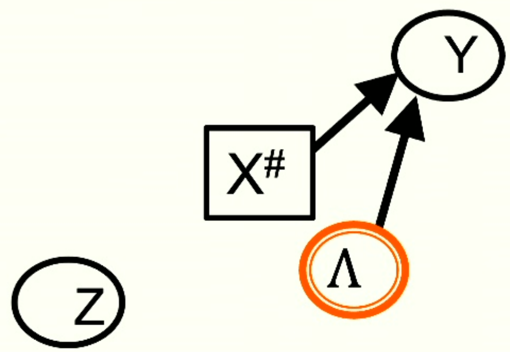
Split-node intervention



$P_{Y|\Lambda X^{\#}}$
$X^{\#} = x$
$P_{X|\Lambda Z}$
P_{Λ}
P_{Z}

$$P_{YXZ|X^{\#}=x} = \Sigma_{\Lambda} P_{Y|\Lambda X^{\#}=x} P_{X|\Lambda Z} P_{\Lambda} P_{Z}$$

Plain intervention



$P_{Y|\Lambda X^{\#}}$
$X^{\#} = x$
P_{Λ}
P_{Z}

$$P_{YZ|X^{\#}=x} = \Sigma_{\Lambda} P_{Y|\Lambda X^{\#}=x} P_{\Lambda} P_{Z}$$

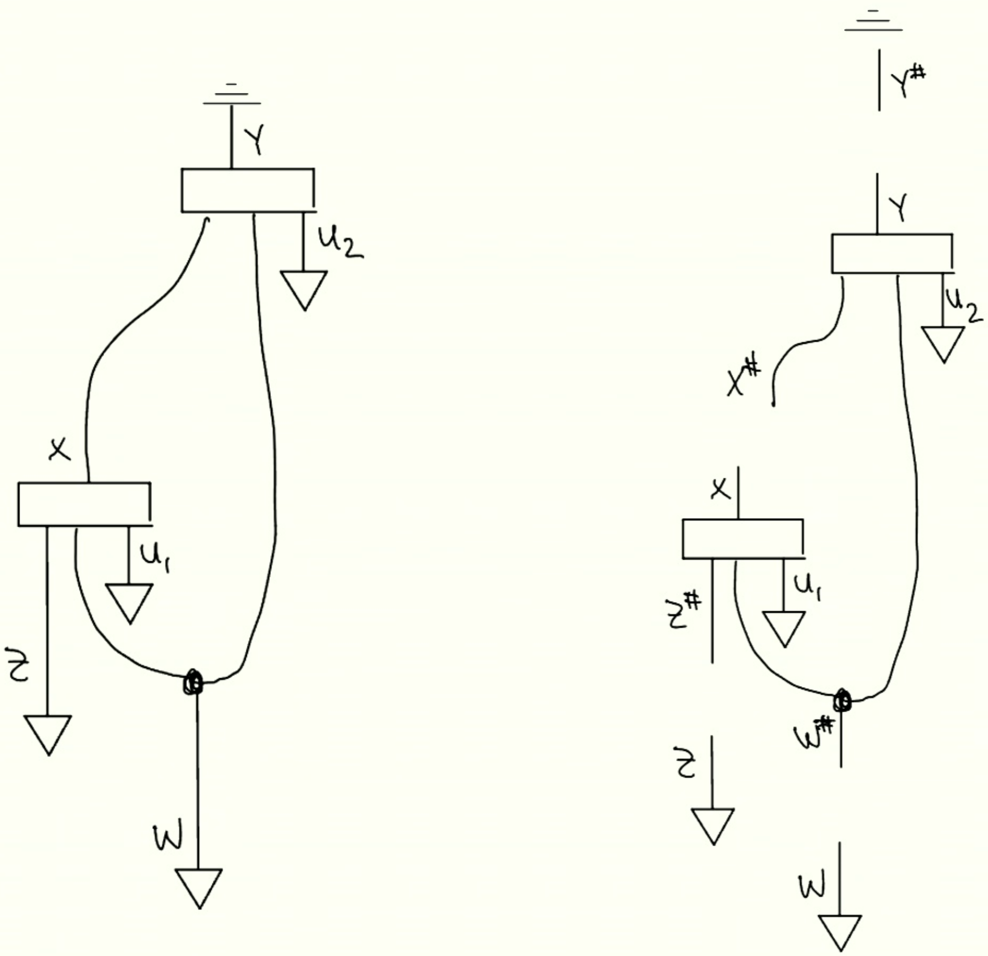
Latent-free causal models

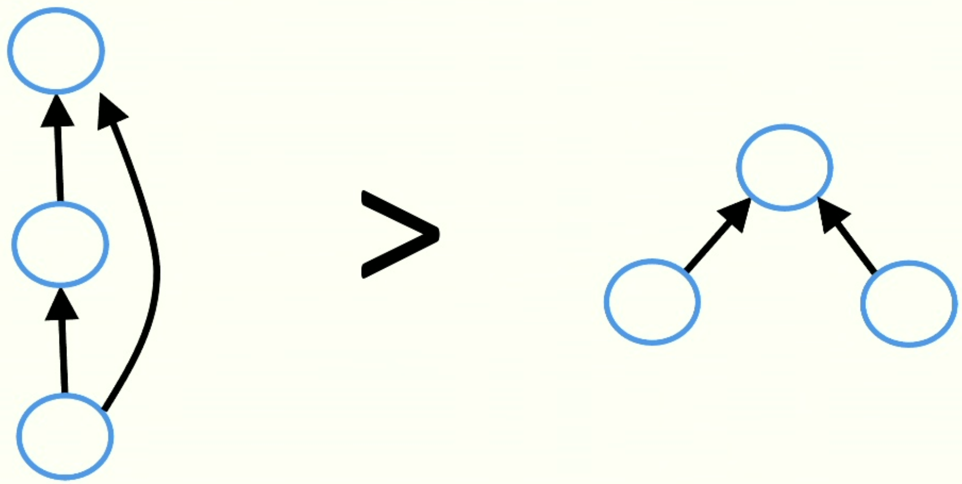
In this context, latent-free means no unobserved variable having more than one observed variable among its causal descendents

Interventional equivalence and dominance in latent-free causal models

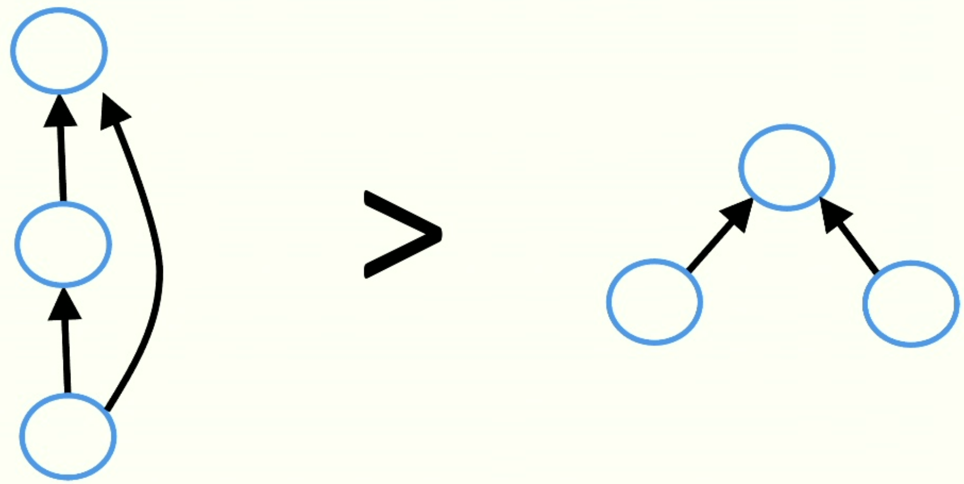
For a causal structure, we define the behaviours under all possible interventions:
For each valuation of the parameters in the model, find the tuple, ranging over subsets S of the full set V of nodes, of distributions over V arising from an intervention on the subset S.

Interventional equivalence of two causal structures: having the same behaviours under all possible interventions


Y
u2
X
u1
Z
W
Y#
Y
u2
X#
X
u1
Z#
W#
Z
W



Observational equivalence and dominance in latent-free causal models



Definition (path blocking) A path between node X and node Y is blocked by a set of vertices **Z** if at least one of the following conditions holds:

1. The path contains a chain whose mediary node is in **Z**
2. The path contains a fork whose tail node is in **Z**

2. The path contains a collider whose head node is **not** in **Z** and no descendant of which is in **Z**.

Definition (d-separation) Given a DAG G with nodes **V** and three disjoint sets of nodes, **X**, **Y, Z** ⊂ **V. X** and **Y** are d-separated by **Z** if and only if for every pair of vertices, X and Y, from the sets **X** and **Y**, every path between X and Y is blocked by **Z**.

The d-separation theorem:

Consider a causal structure G and three disjoint subsets of variables **X**, **Y** and **Z**.

Soundness

d-separation between **X** and **Y** given **Z** in G
implies $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$ in all distributions P compatible with G

Completeness

$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$ in all distributions P compatible with G
implies d-separation between **X** and **Y** given **Z** in G

The d-separation theorem:

Consider a causal structure G and three disjoint subsets of variables **X**, **Y** and **Z**.

Soundness

$$\mathbf{X} \perp_d \mathbf{Y}|\mathbf{Z} \text{ in } G \quad \Longrightarrow \quad \forall P \in \mathsf{Comp}_G : \mathbf{X} \perp \mathbf{Y}|\mathbf{Z} \text{ in } P$$

Completeness

$$\forall P \in \mathsf{Comp}_G : \mathbf{X} \perp \mathbf{Y}|\mathbf{Z} \text{ in } P \quad \Longrightarrow \quad \mathbf{X} \perp_d \mathbf{Y}|\mathbf{Z} \text{ in } G$$

If all the variables in a DAG are observed,
the conditional independence relations implied by the d-separation relations are **all** the constraints on the distribution

Proof:
Recall the Markov condition characterizing all compatible distributions for a latent-free DAG

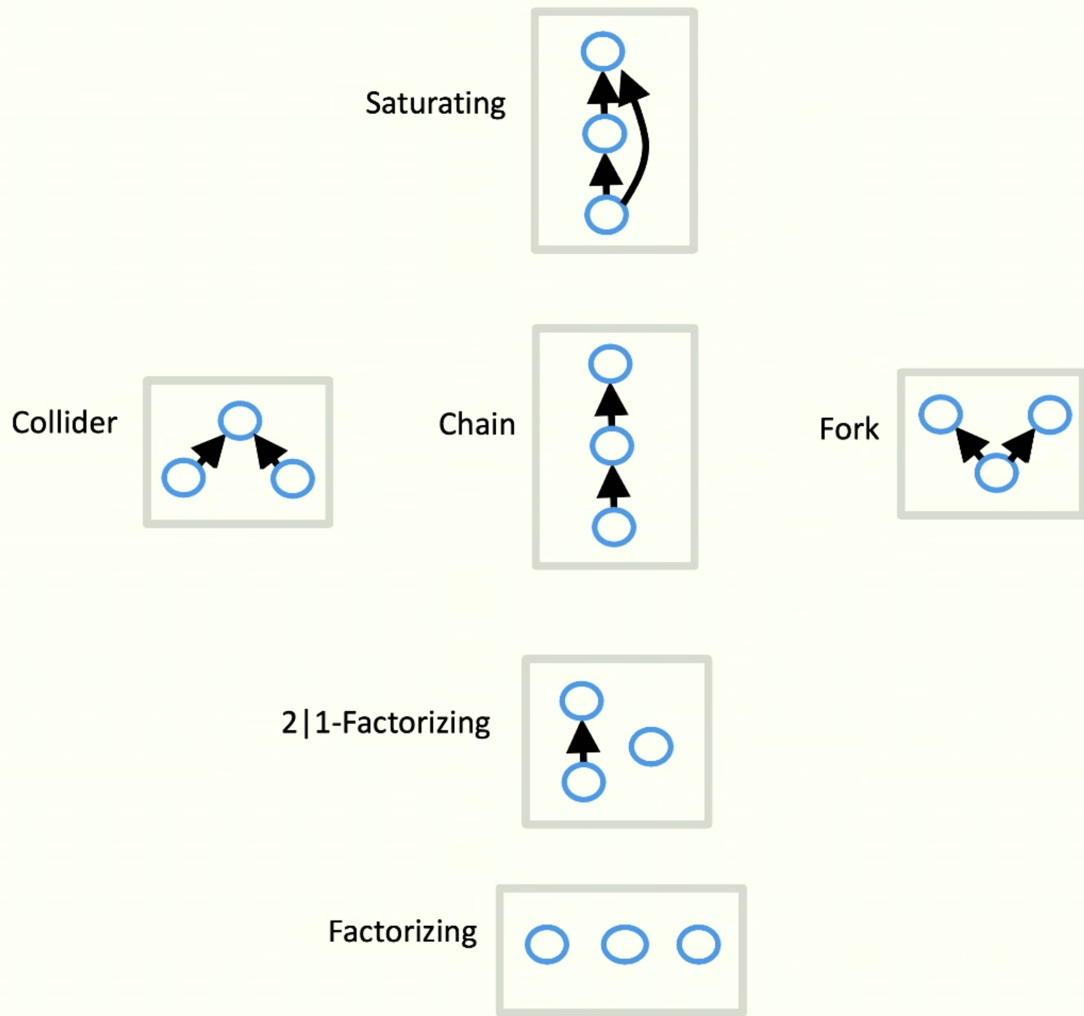
We saw that this set can also be characterized by the set of conditional independence relations described by the local Markov condition together with semi-graphoid axioms

If all the variables in a DAG are observed,
the conditional independence relations implied by the d-separation relations are **all** the constraints on the distribution

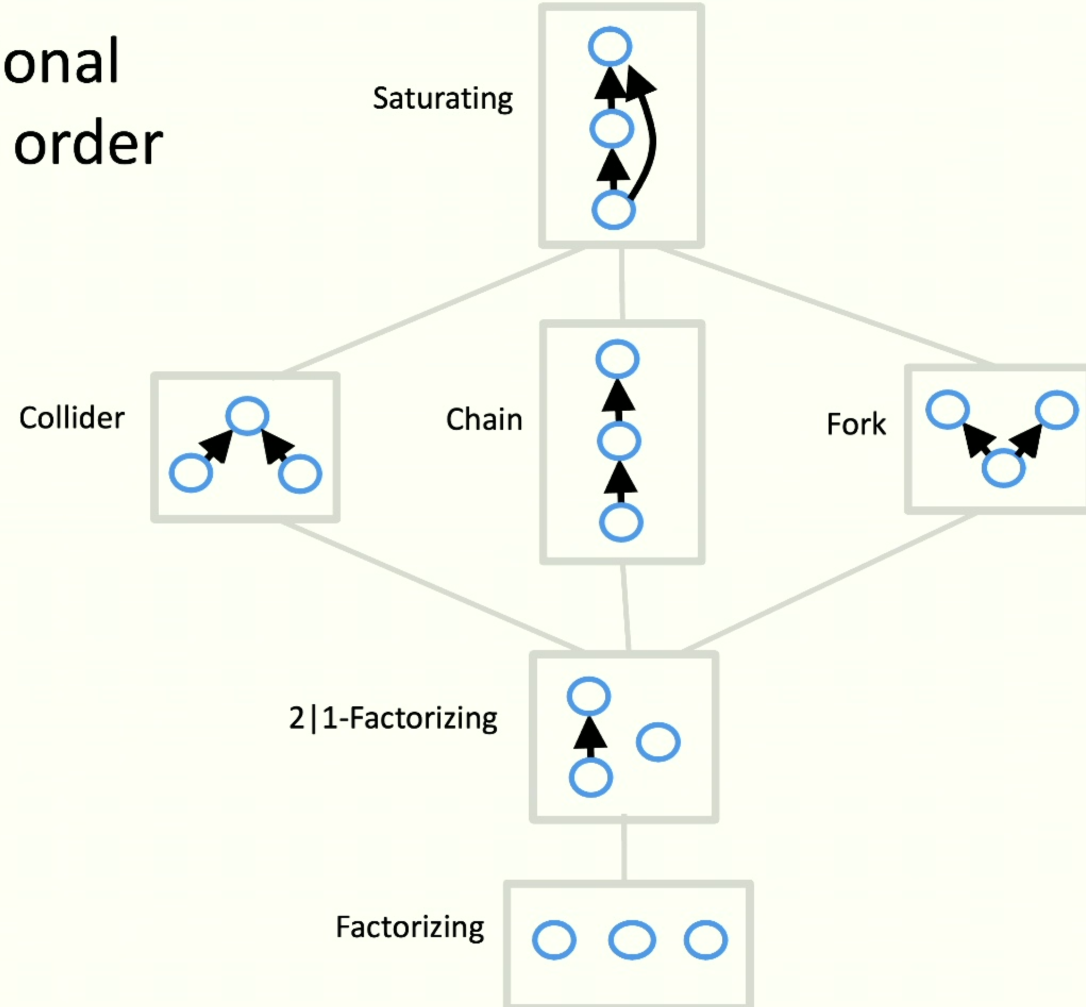
Proof:
Recall the Markov condition characterizing all compatible distributions for a latent-free DAG

We saw that this set can also be characterized by the set of conditional independence relations described by the local Markov condition together with semi-graphoid axioms

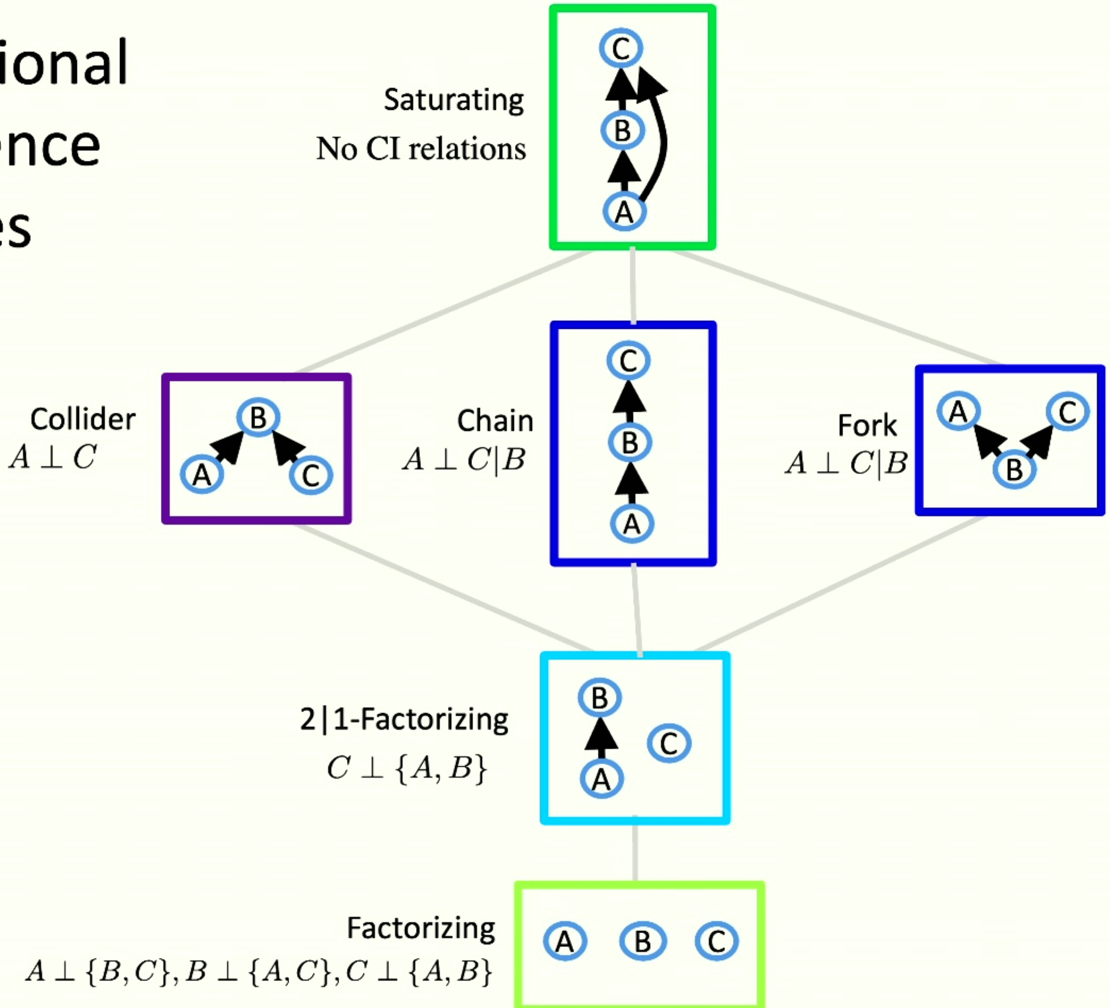
This set is equivalent to the one obtained from all d-separation relations


Saturating
Collider
Chain
Fork
2|1-Factorizing
Factorizing

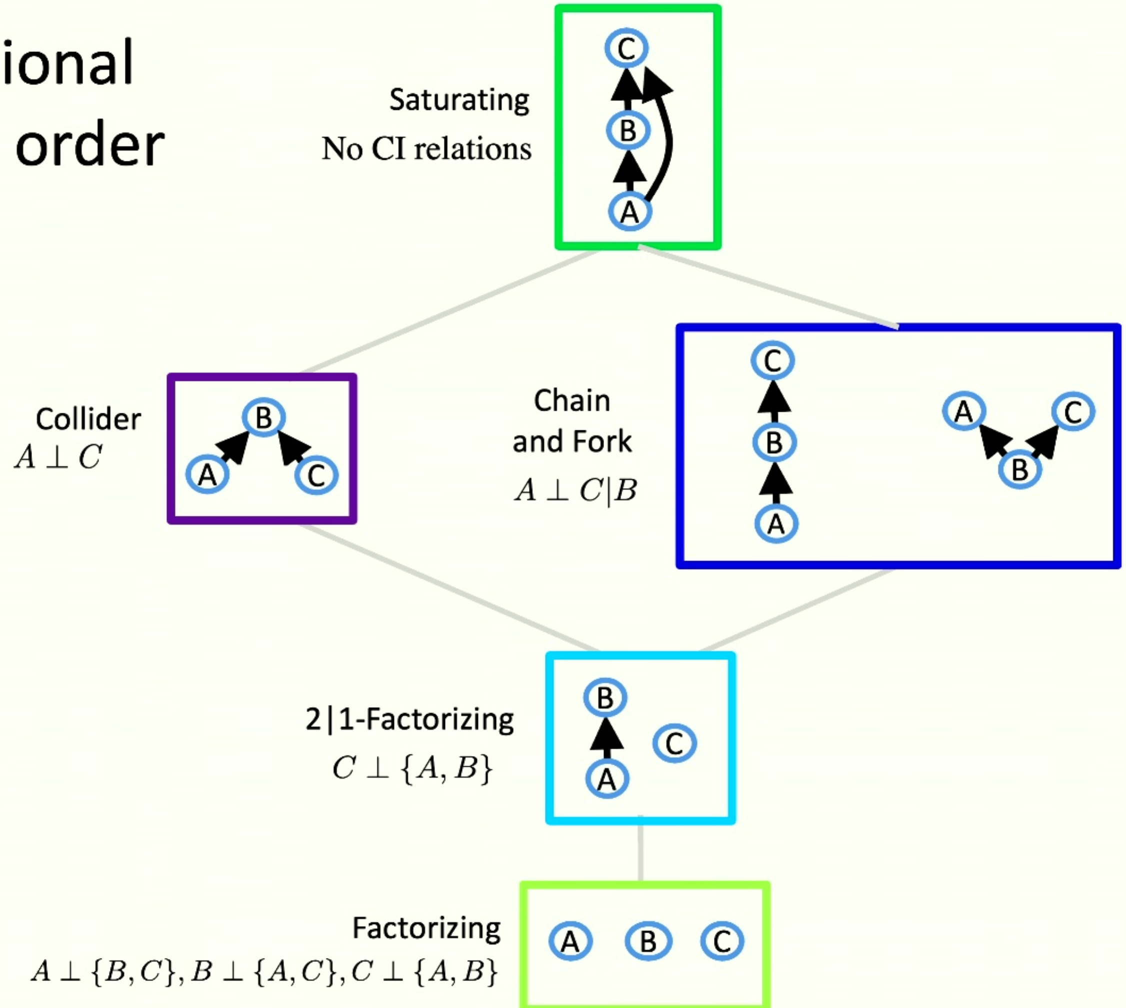
Interventional dominance order

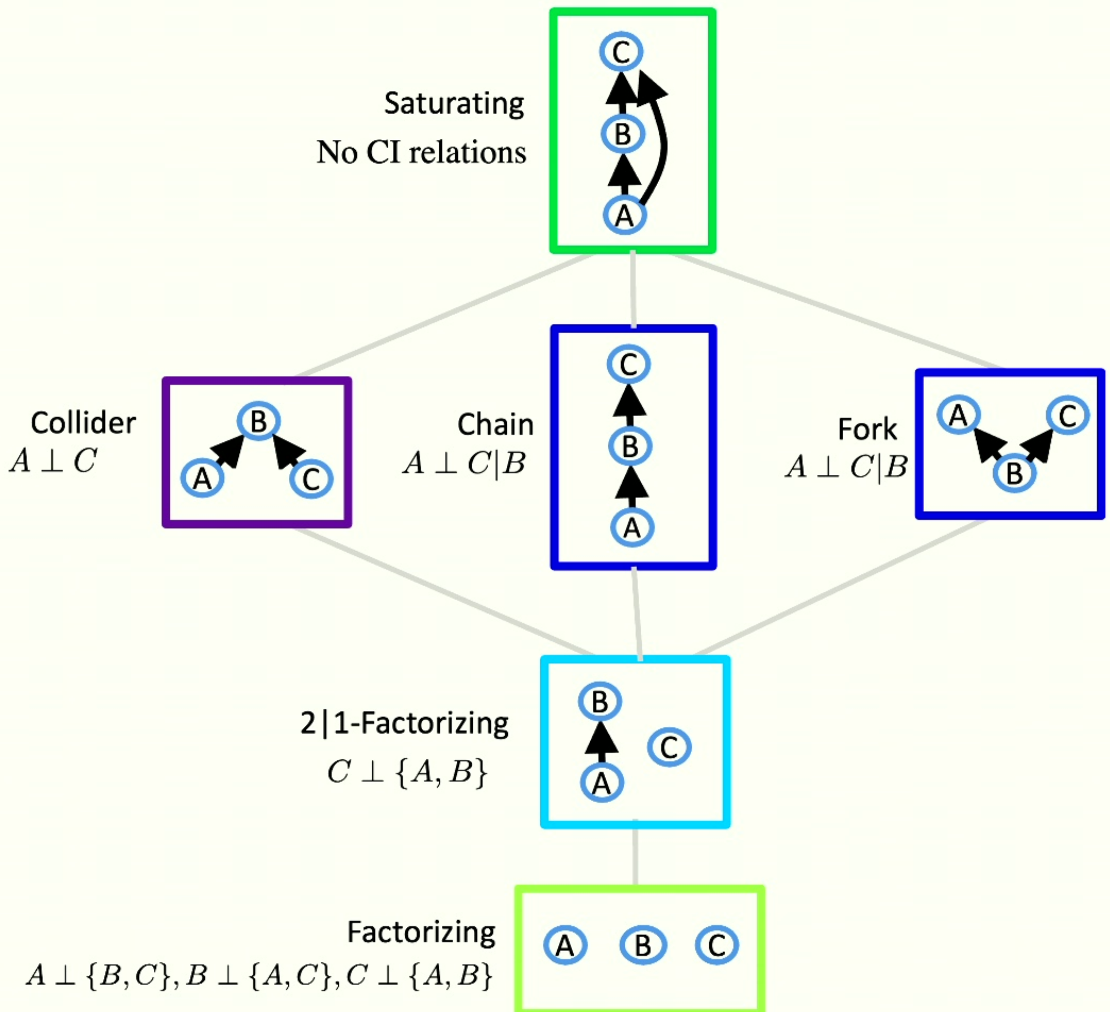


Observational Equivalence classes

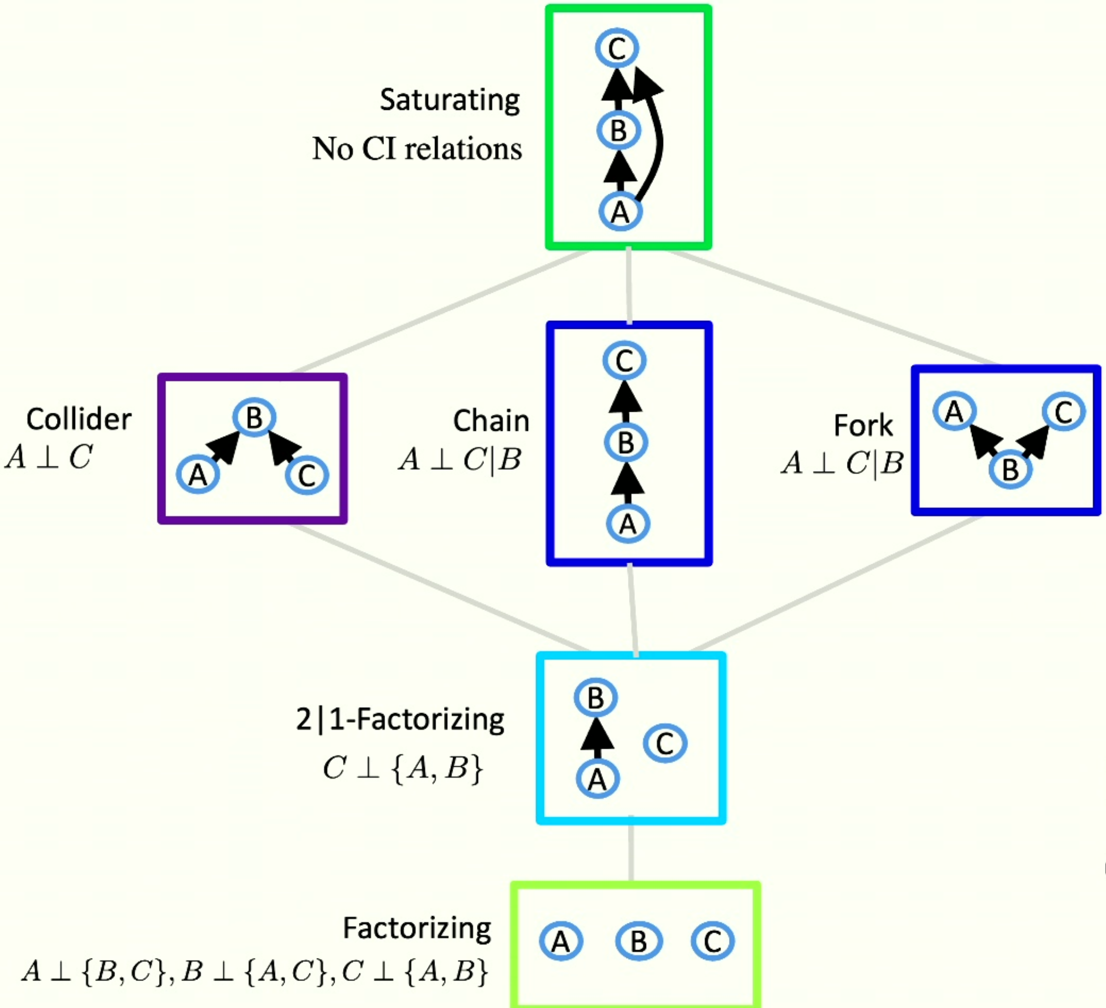


Observational Domiance order



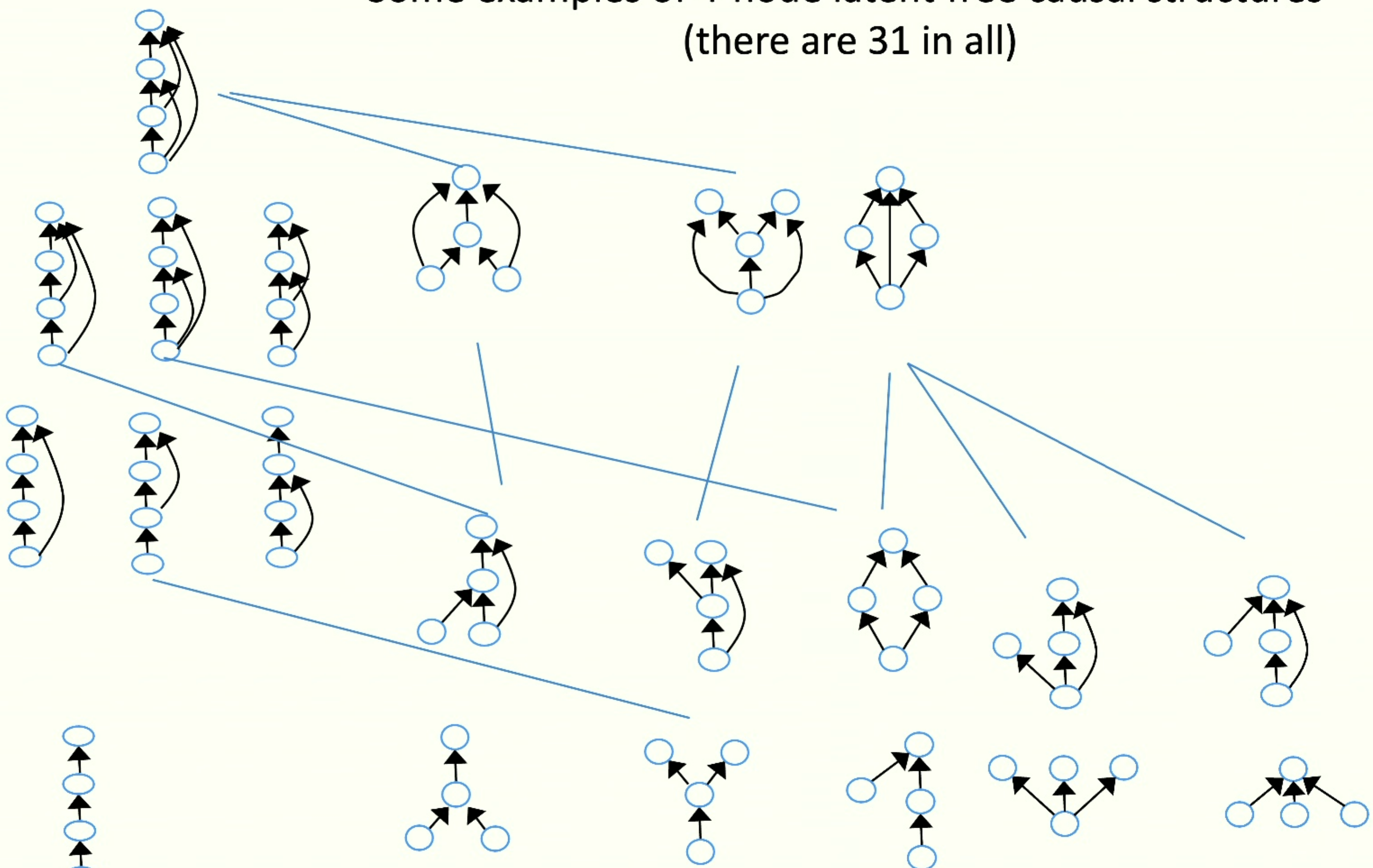


Saturating
No CI relations
C
B
A
Collider
$A \perp C$
B
A
C
Chain
$A \perp C|B$
C
B
A
Fork
$A \perp C|B$
A
C
B
2|1-Factorizing
$C \perp \{A, B\}$
B
C
A
Factorizing
$A \perp \{B, C\}, B \perp \{A, C\}, C \perp \{A, B\}$
A
B
C



Compare with conditional independences estimated from data

Some examples of 4-node latent-free causal structures (there are 31 in all)



IC* algorithm and PC algorithm

Set of conditional independence relations ⇨ Feasible and infeasible latent-free DAGs

Entropies

Shannon entropy

$$H(X) := -\sum_x P_X(x)\log P_X(x)$$

Conditional entropy

$$H(X|Y) := H(XY) - H(Y)$$

Mutual information

$$I(X:Y) = H(X) + H(Y) - H(XY)$$

Conditional mutual information

$$I(X:Y|Z) = H(XZ) + H(YZ) - H(XYZ) - H(Z)$$

Def'n: A and B are marginally independent

$$P_{AB} = P_A P_B$$
$$P_{B|A} = P_B$$
$$P_{A|B} = P_A$$
$$I(A:B) = 0$$

Denote this

$$(A \perp B)$$

Proof:

$$H(AB) := -\sum_{a,b} P_{AB}(ab) \log P_{AB}(ab)$$

If A and B are marginally independent

$$P_{AB} = P_A P_B$$

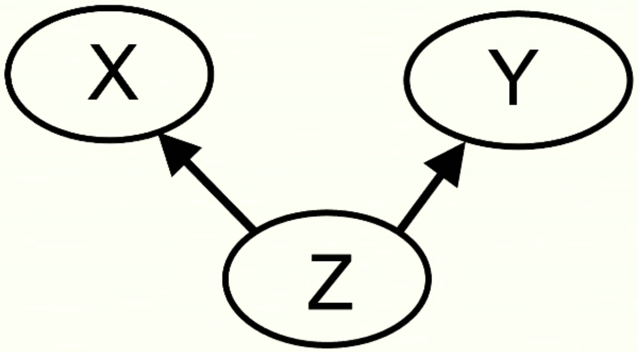
$$\begin{aligned} H(AB) &:= -\sum_{a,b} P_A(a) P_B(b) (\log P_A(a) + \log P_B(b)) \\ &= -\sum_a P_A(a) \log P_A(a) - \sum_b P_B(b) \log P_B(b) \\ &= H(A) + H(B) \end{aligned}$$

Def'n: A and B are conditionally independent given C

$$P_{AB|C} = P_{A|C}P_{B|C}$$
$$P_{B|AC} = P_{B|C}$$
$$P_{A|BC} = P_{A|C}$$

$$P_{ABC} = P_{A|C}P_{B|C}P_C$$

Denote this
$(A \perp B|C)$



$$X \perp Y | Z$$

$$P_{XY|Z} = P_{X|Z} P_{Y|Z}$$

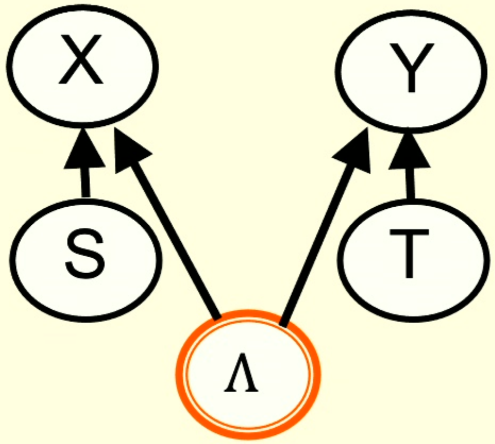
$$I(X : Y|Z) = 0$$

$$I(X : Y) \leq H(Z)$$

Latent-permitting causal models

The conventional type of latent-permitting causal model: unrestricted cardinality of latents

Causal structure



Parameters

$P_{X|S\Lambda}$

$P_{Y|T\Lambda}$

P_{Λ}

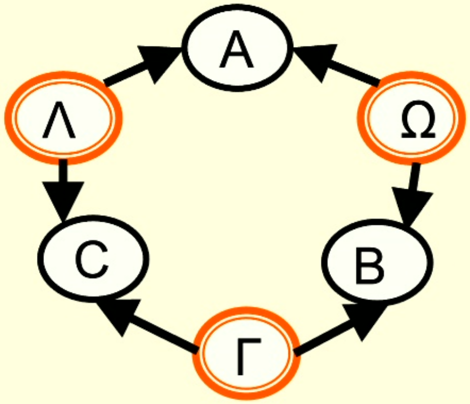
P_S

P_T

$$P_{XYST} = \sum_{\Lambda} P_{Y|T\Lambda} P_{X|S\Lambda} P_{\Lambda} P_S P_T$$

Λ of arbitrary cardinality

Causal structure



Parameters

$\mathrm{P}_{A|\Lambda\Omega}$

$\mathrm{P}_{B|\Omega\Gamma}$

$\mathrm{P}_{C|\Gamma\Lambda}$

P_{Λ}

P_{Ω}

P_{Γ}

$$P_{ABC} = \sum_{\Lambda\Omega\Gamma} \mathrm{P}_{A|\Lambda\Omega}\mathrm{P}_{B|\Omega\Gamma}\mathrm{P}_{C|\Gamma\Lambda}\mathrm{P}_{\Lambda}\mathrm{P}_{\Omega}\mathrm{P}_{\Gamma}$$

Λ,*Γ*,*Ω* of arbitrary cardinality

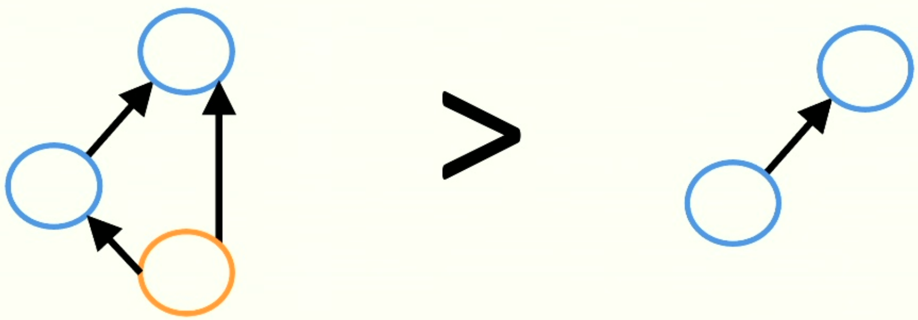
Note: Upcoming causal inference workshop at Perimeter has as a focus the case of latent variables with restricted cardinality

Interventional equivalence classes of latent-permitting causal models

For a causal structure, we define the behaviours under possible interventions on observed nodes:
For each valuation of the parameters in the model, find the tuple, ranging over subsets S of the full set V of *observed* nodes, of distributions over V arising from an intervention on the subset S.

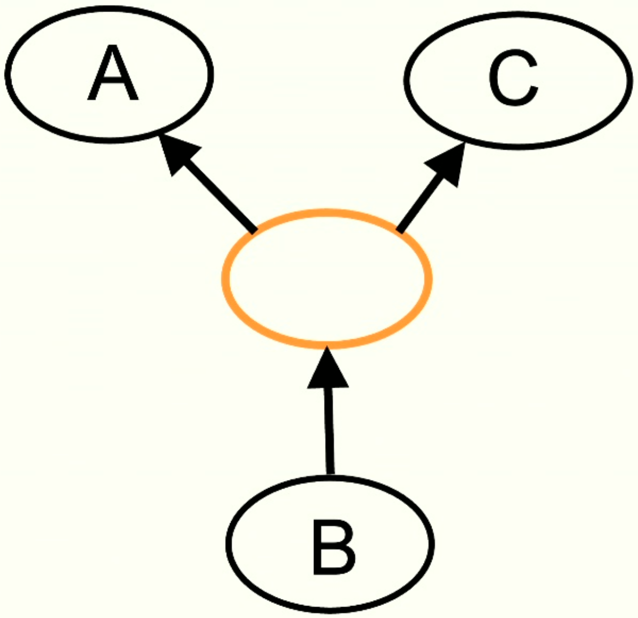
Interventional equivalence of two latent-permitting causal structures: having the same behaviours under possible interventions on observed nodes

Different interventional behaviours



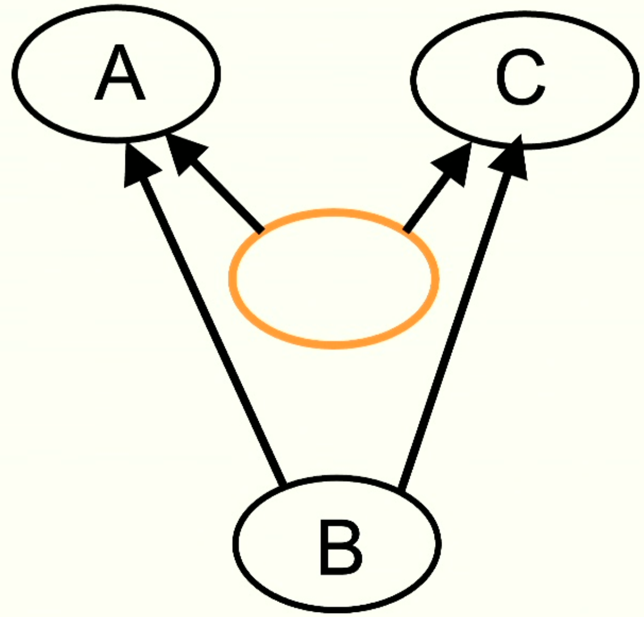
Rules for proving interventional equivalence

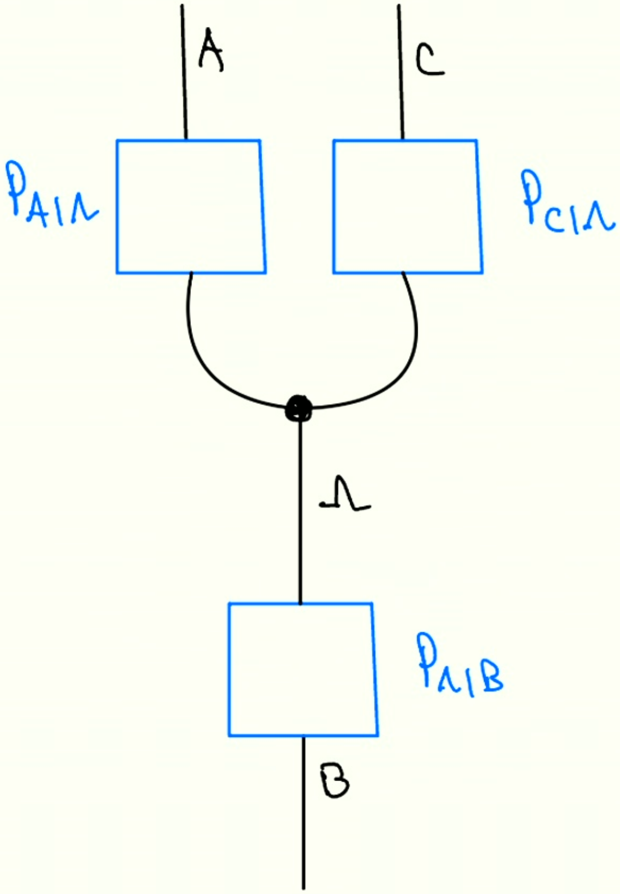
Exogenization rule



$\simeq$

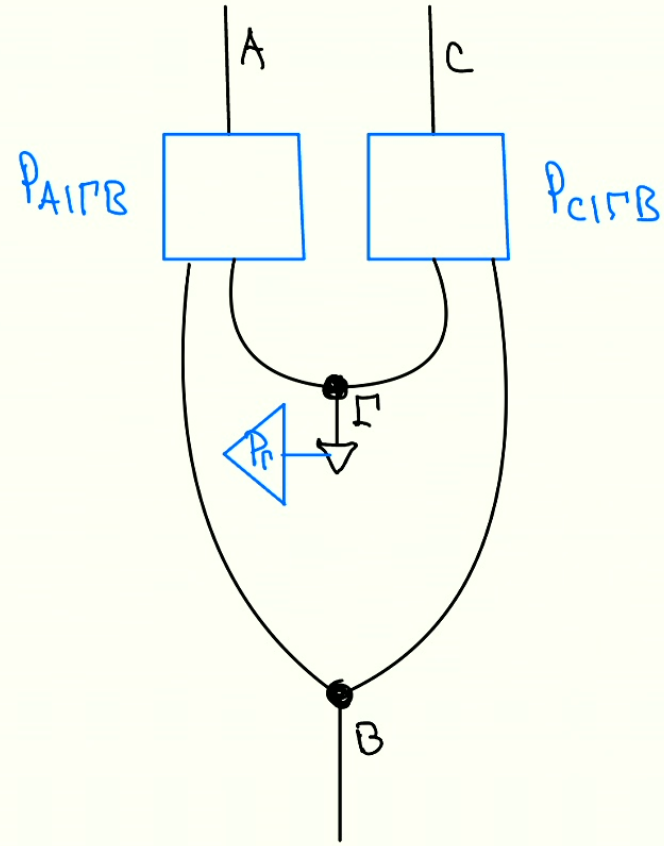
interventionally equivalent

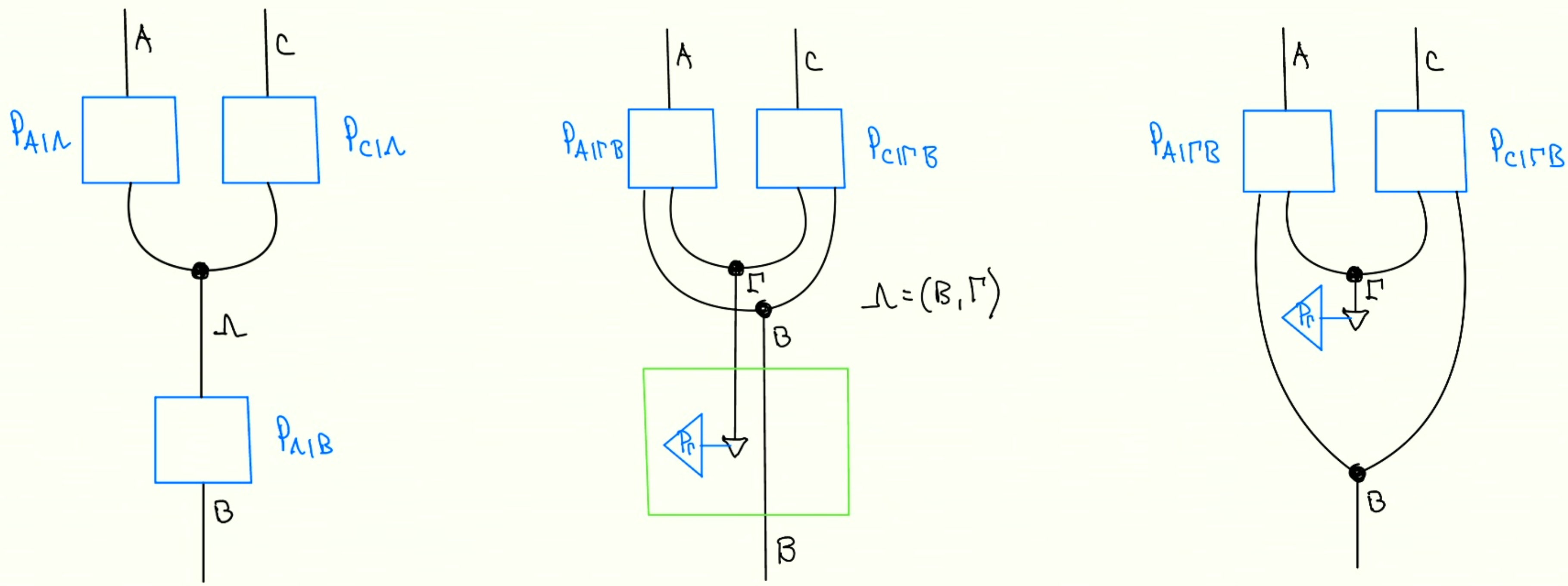


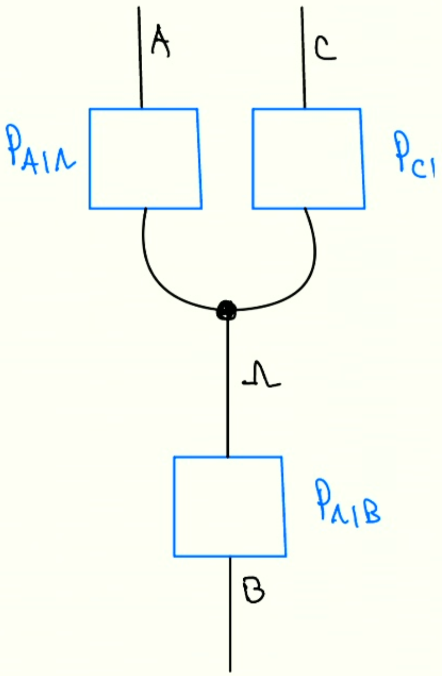


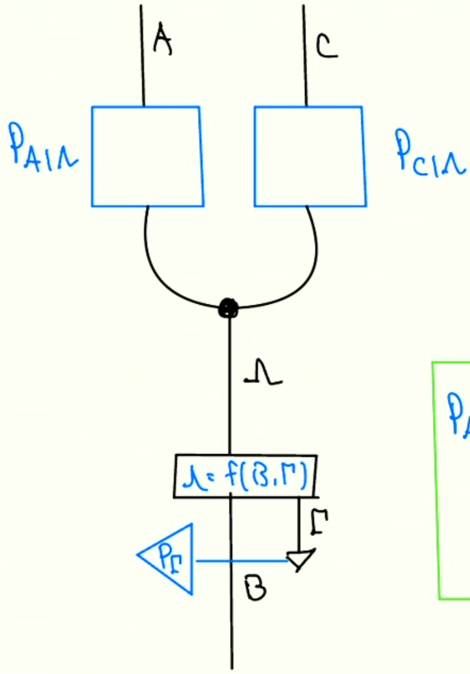
$\simeq$

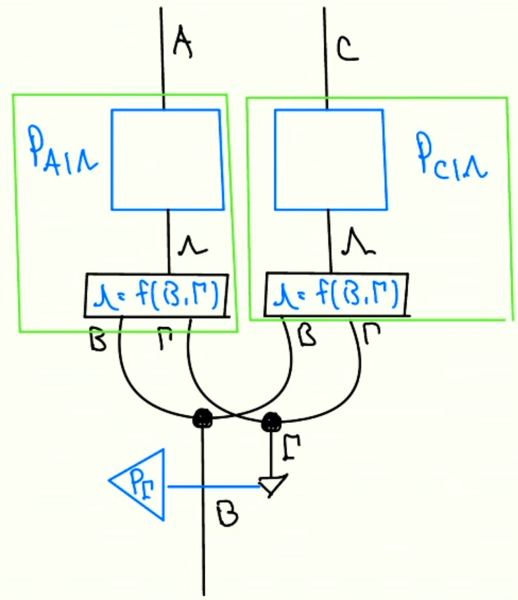
interventionally equivalent

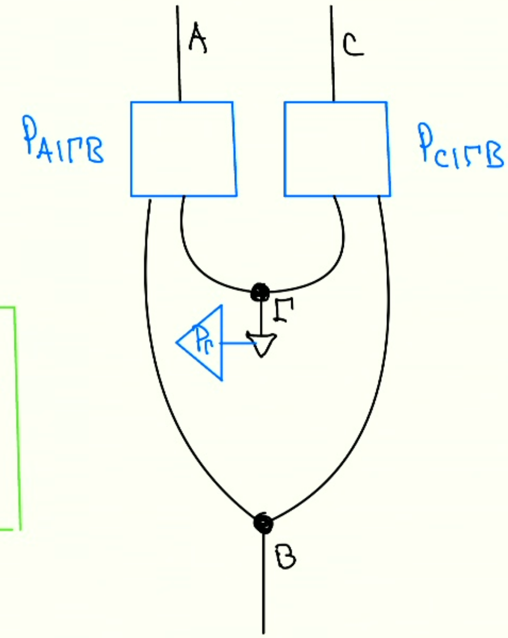



A
C
$P_{A|\Lambda}$
$P_{C|\Lambda}$
Λ
$P_{\Lambda|B}$
B
A
C
$P_{A|\Gamma B}$
$P_{C|\Gamma B}$
Γ
B
$\Lambda = (B, \Gamma)$
P_Γ
B
A
C
$P_{A|\Gamma B}$
$P_{C|\Gamma B}$
Γ
P_Γ
B

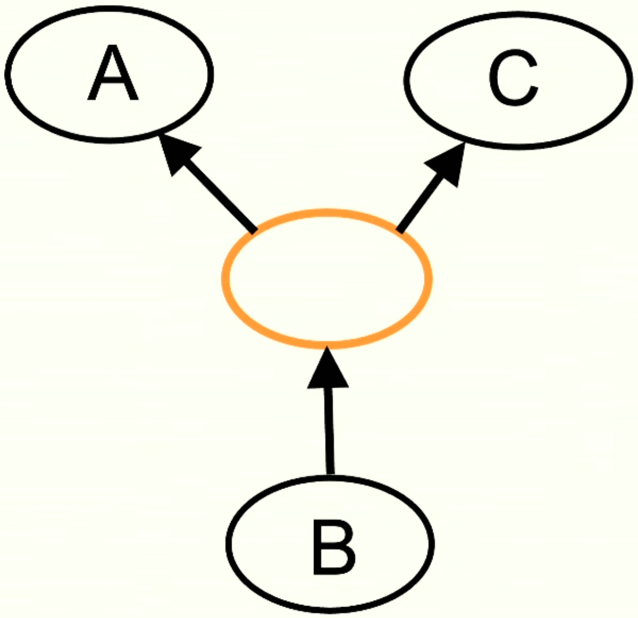

A
C
P_{A|Λ}
P_{C|}
Λ
P_{Λ|B}
B


A
C
P_{A|Λ}
P_{C|Λ}
Λ
Λ = f(B,Γ)
Γ
P_Γ
B


A
C
P_{A|Λ}
P_{C|Λ}
Λ
Λ
Λ = f(B,Γ)
Λ = f(B,Γ)
B
Γ
B
Γ
Γ
P_Γ
B

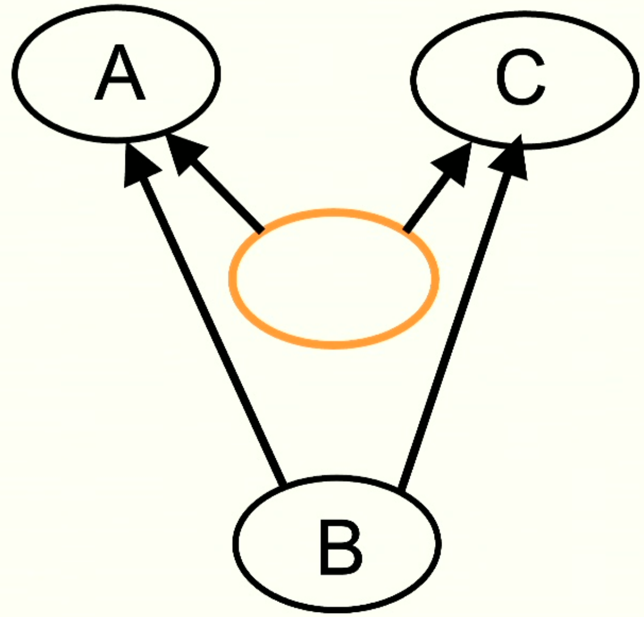

A
C
P_{A|ΓB}
P_{C|ΓB}
Γ
P_Γ
B

Exogenization rule

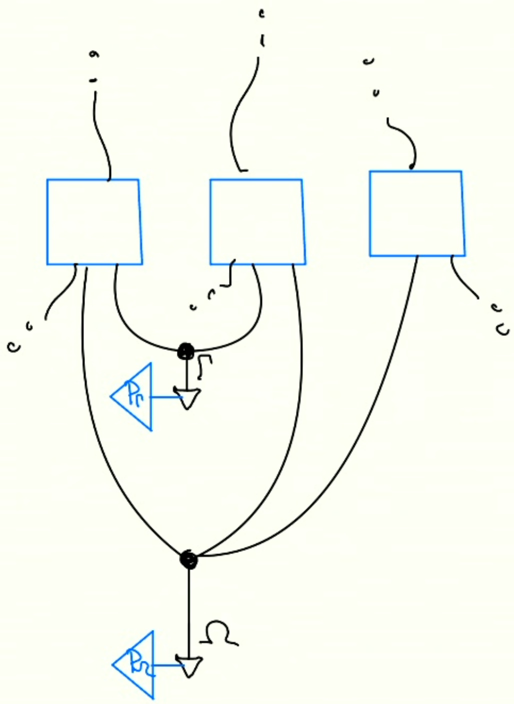


$\simeq$

interventionally equivalent

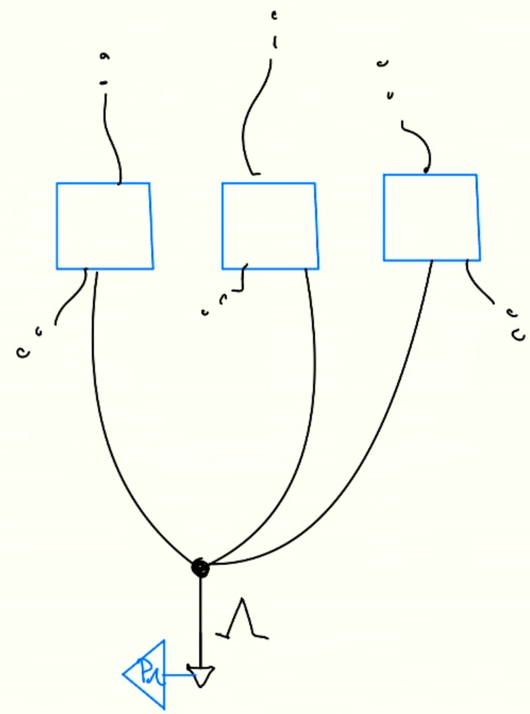


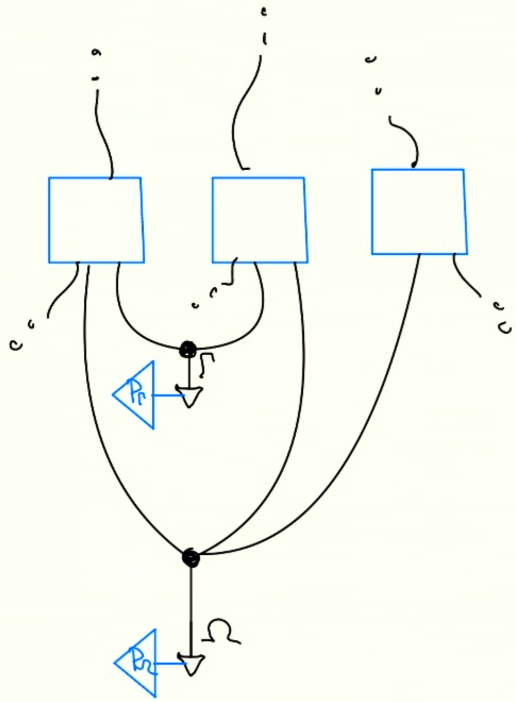
Eliminating redundant latents rule

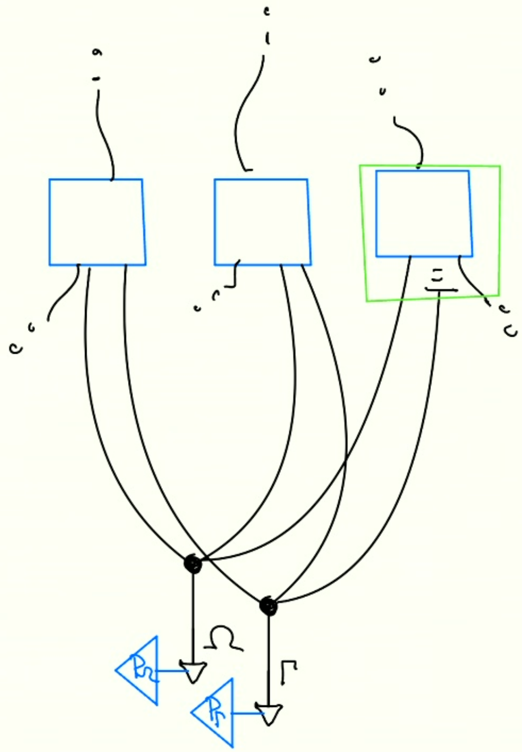


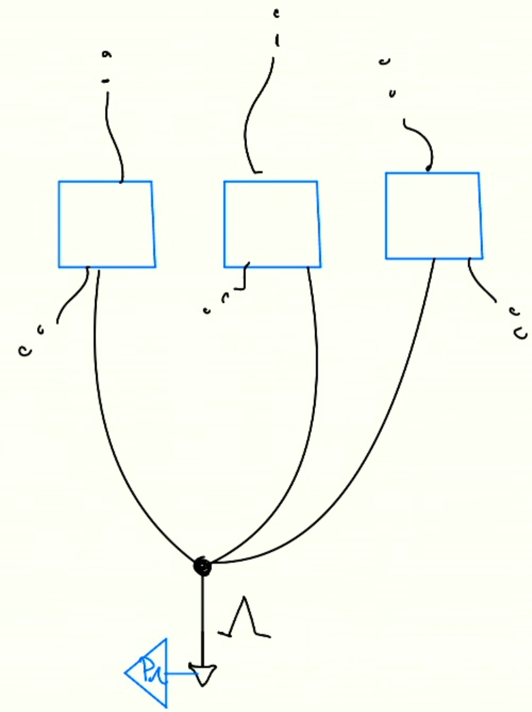
$\simeq$

interventionally equivalent





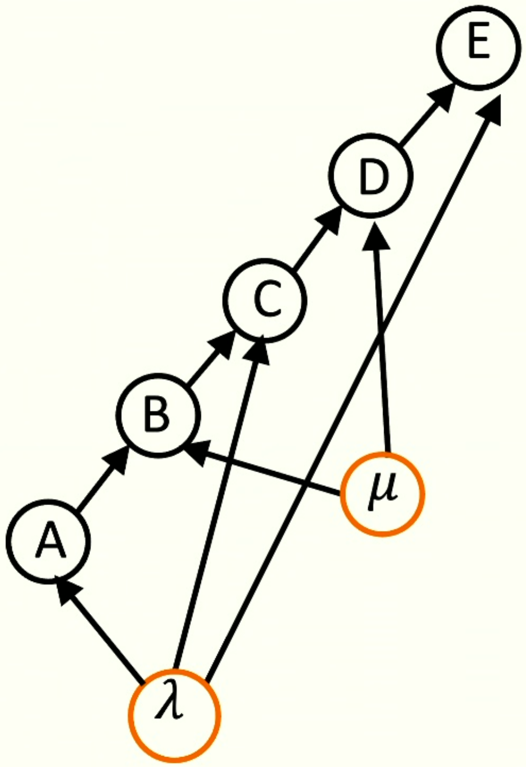


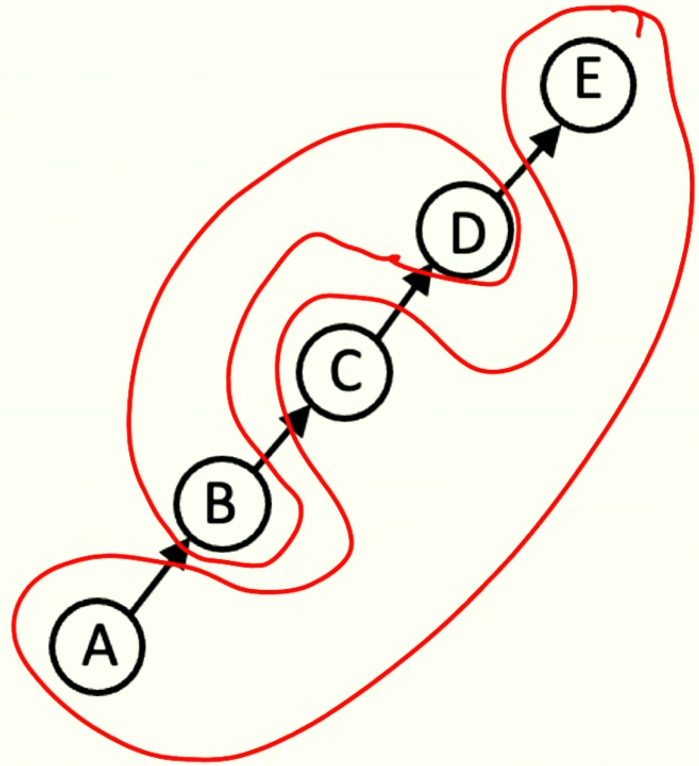


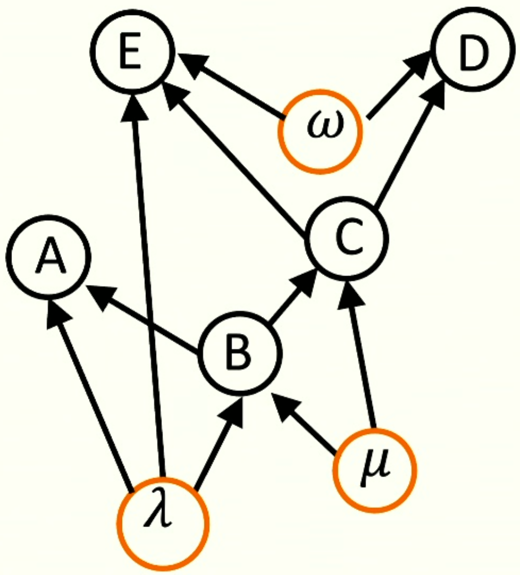
Definition: An abstract Simplicial Complex over a finite set V is a set B of subsets of V such that

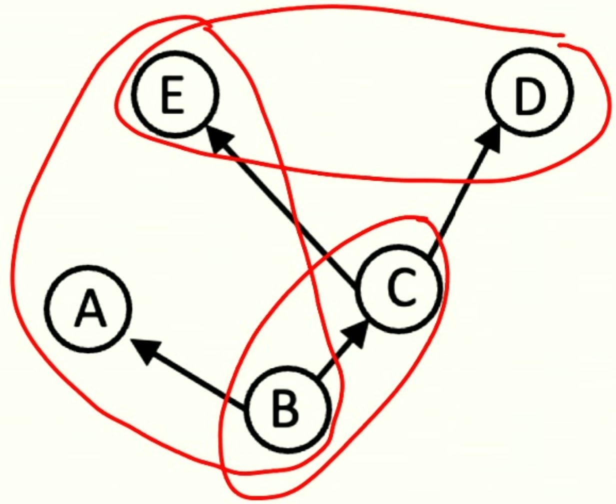
- $\{v\} \in B$ for all $v \in V$
- If $S \subseteq T \subseteq V$ and $T \in B$, then $S \in B$

Definition: an mDAG is a pair (D, *B*), where D is a latent-free DAG and *B* is an abstract simplicial complex


E
D
C
B
A
μ
λ


E
D
C
B
A

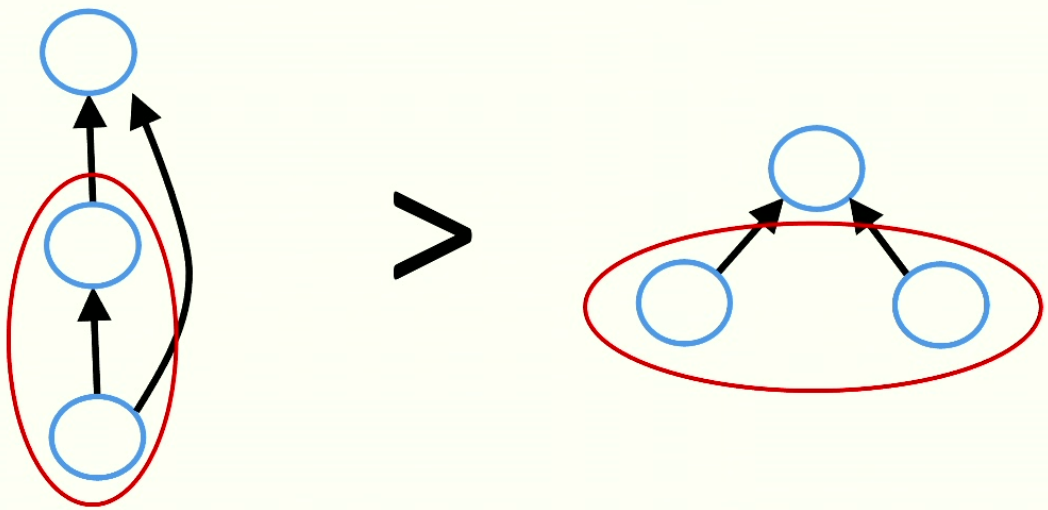

E
D
ω
A
C
B
μ
λ


E
D
A
C
B

Some classification results

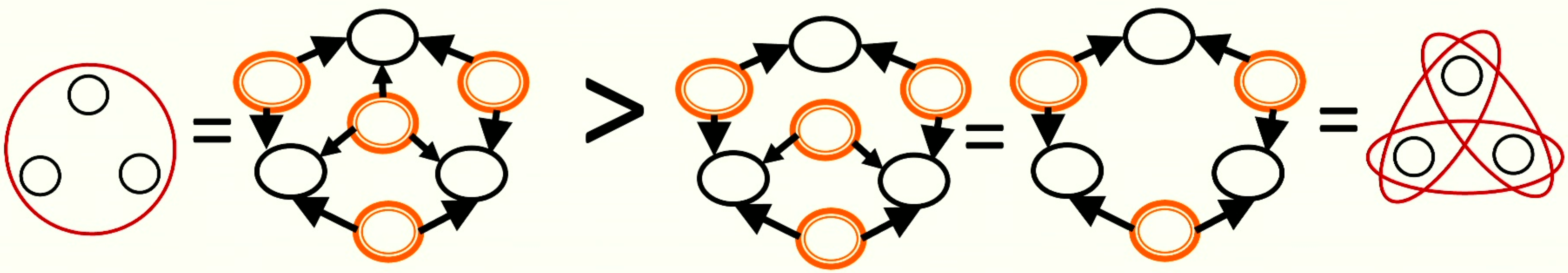
Interventional strict dominance:

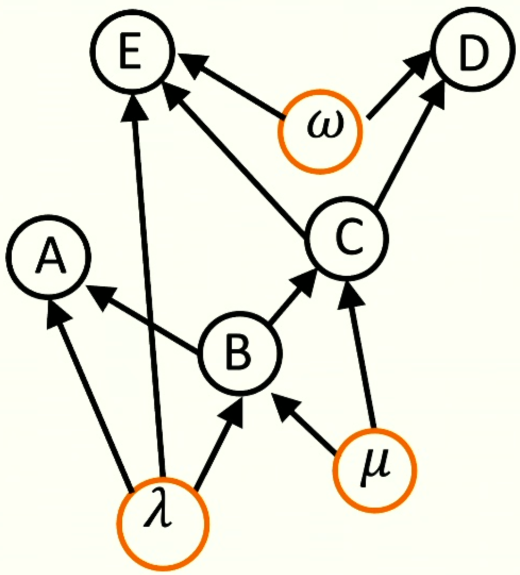
Dropping a directed edge in the directed graph

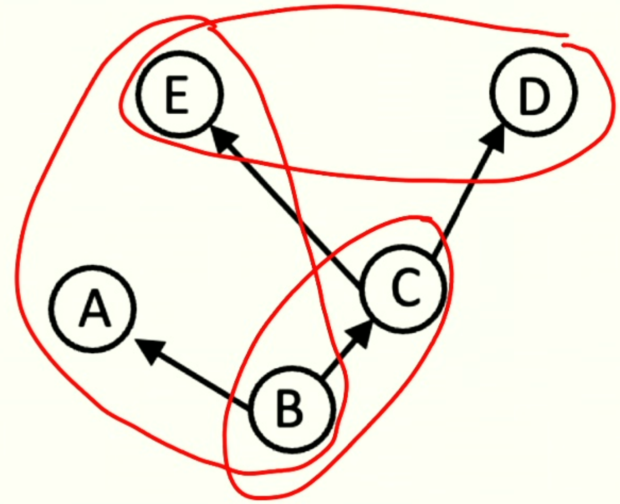


Interventional strict dominance:

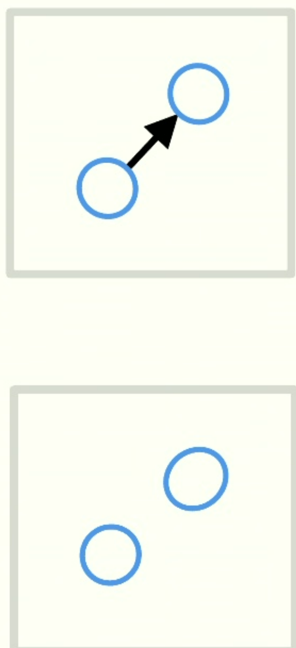
Reducing the abstract simplicial complex to a subcomplex



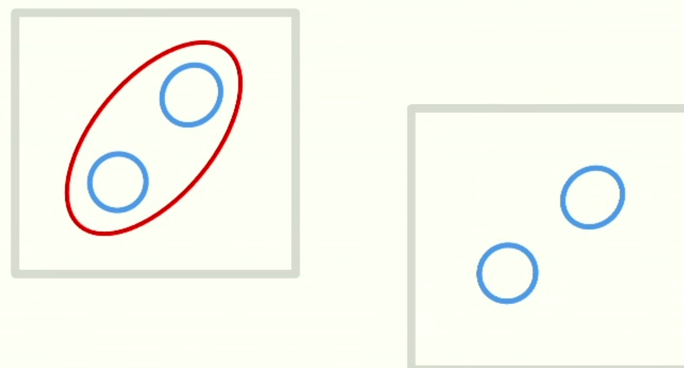

E
D
ω
A
C
B
μ
λ


E
D
A
C
B

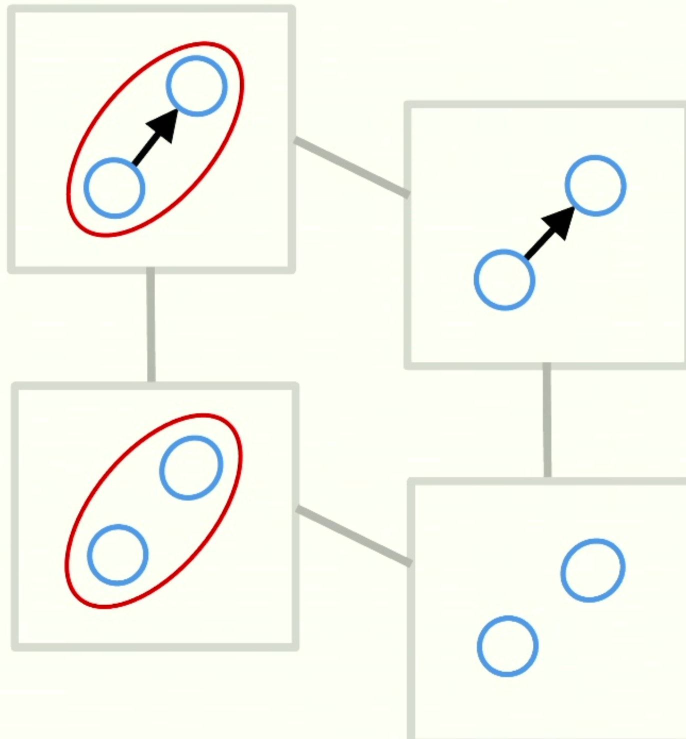
Directed structures



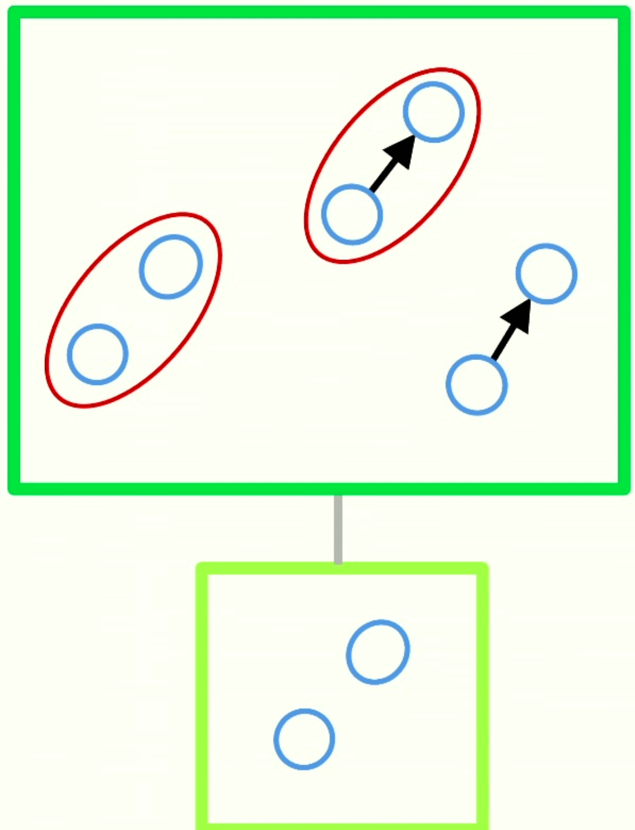
Simplicial complexes



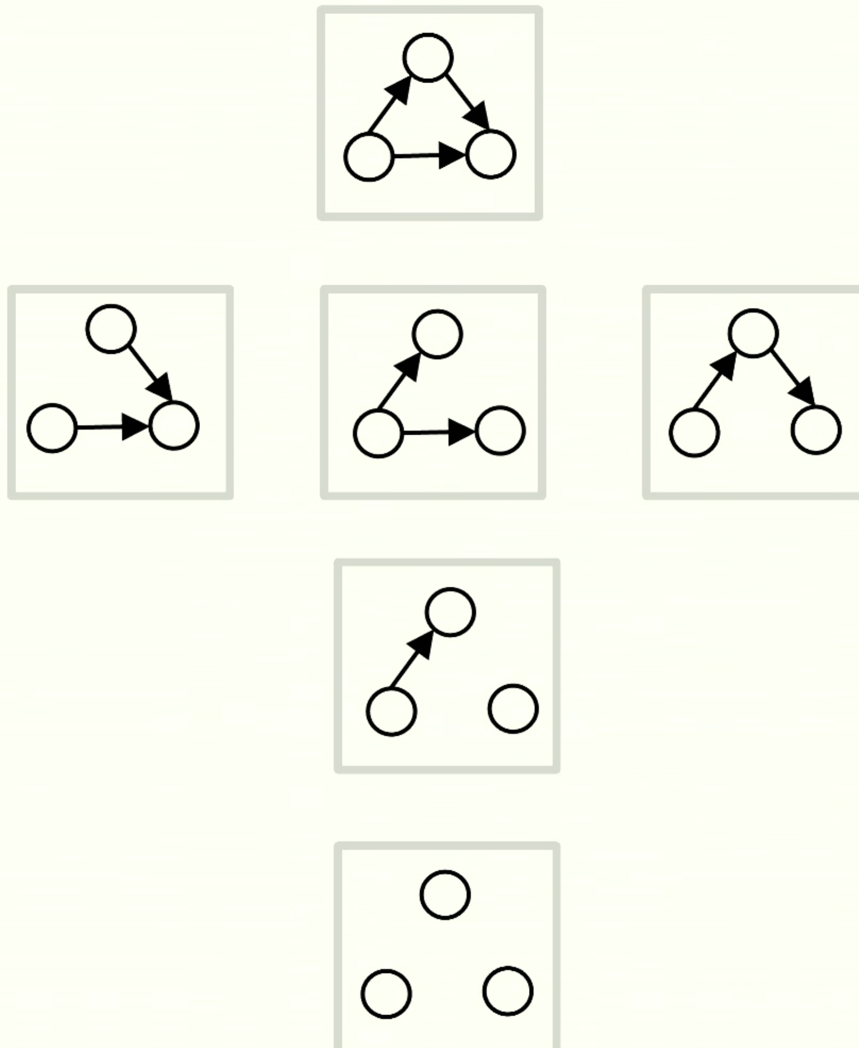
Interventional dominance order

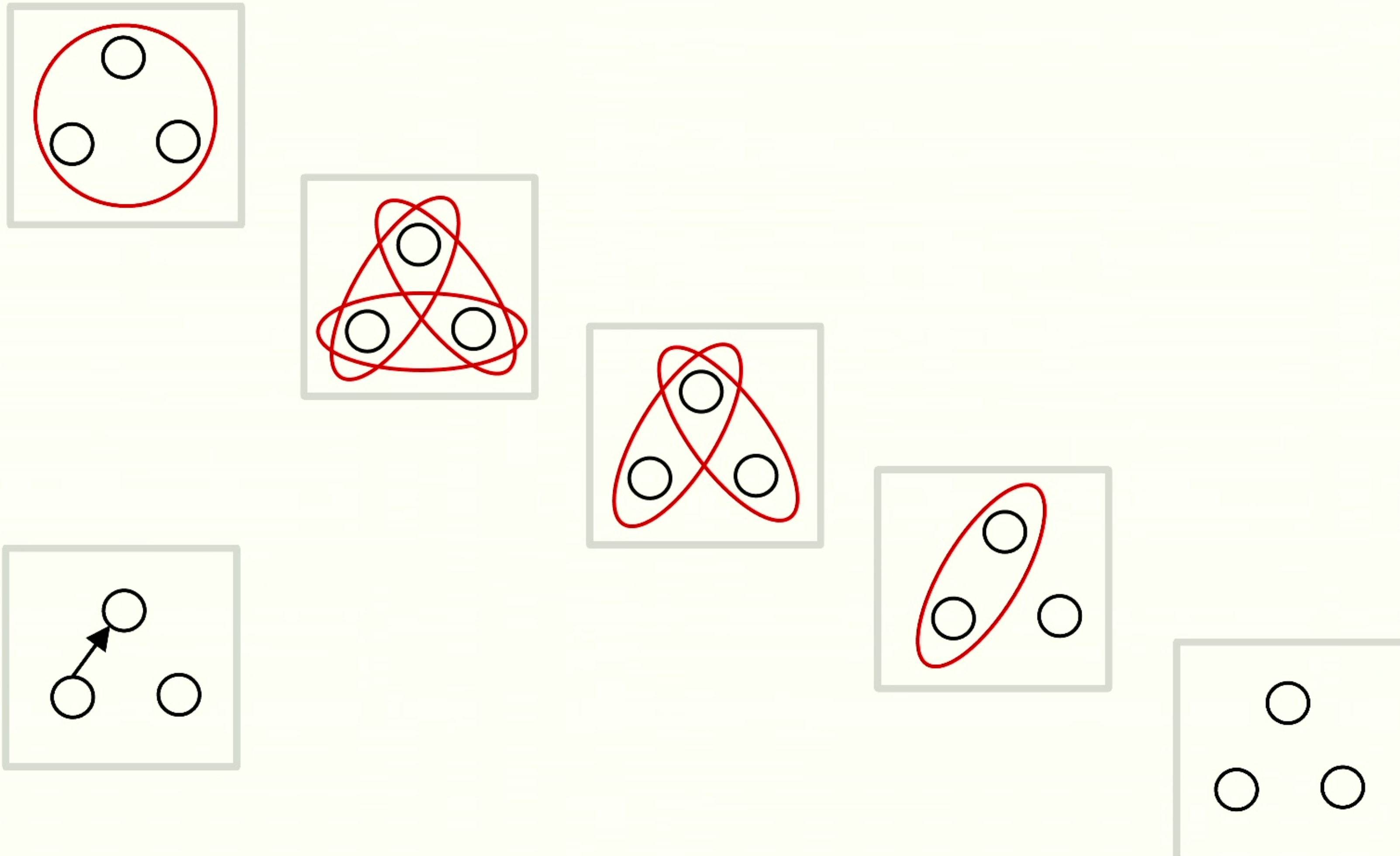


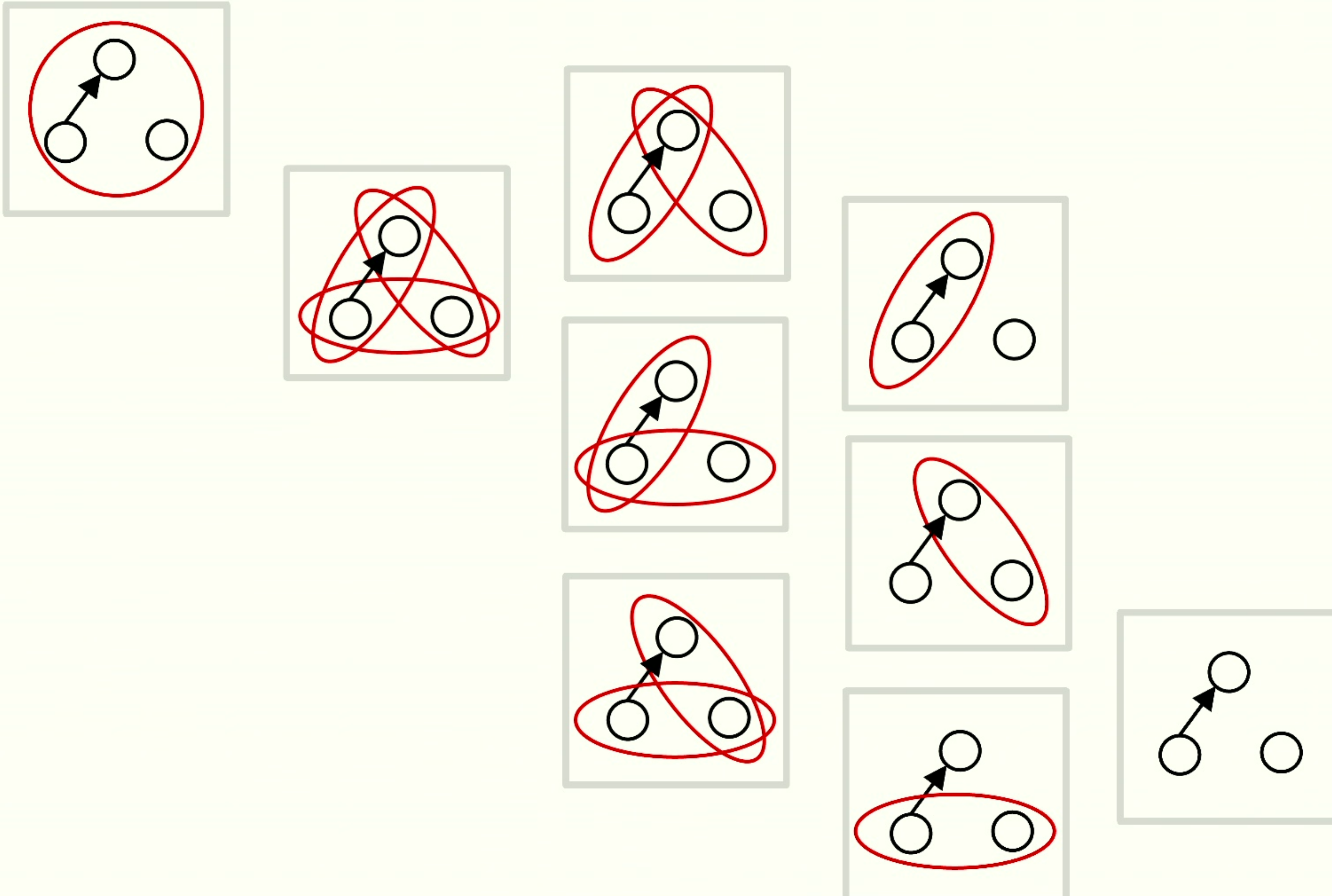
Observational dominance order

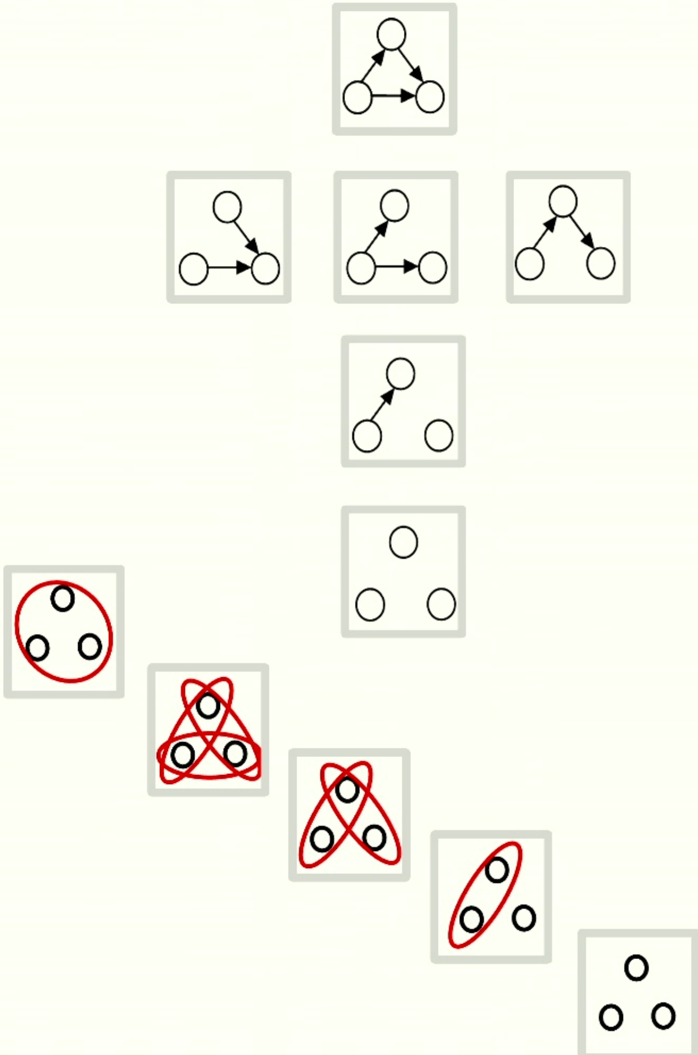


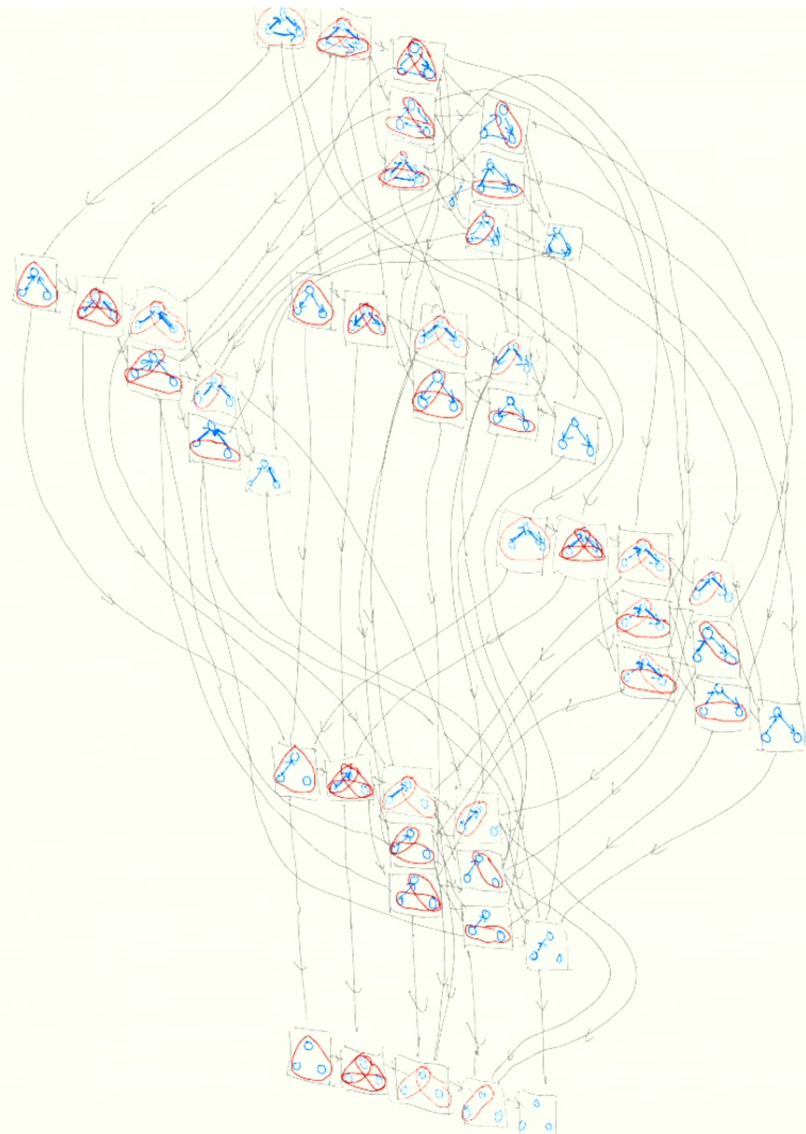
Three node mDAGs

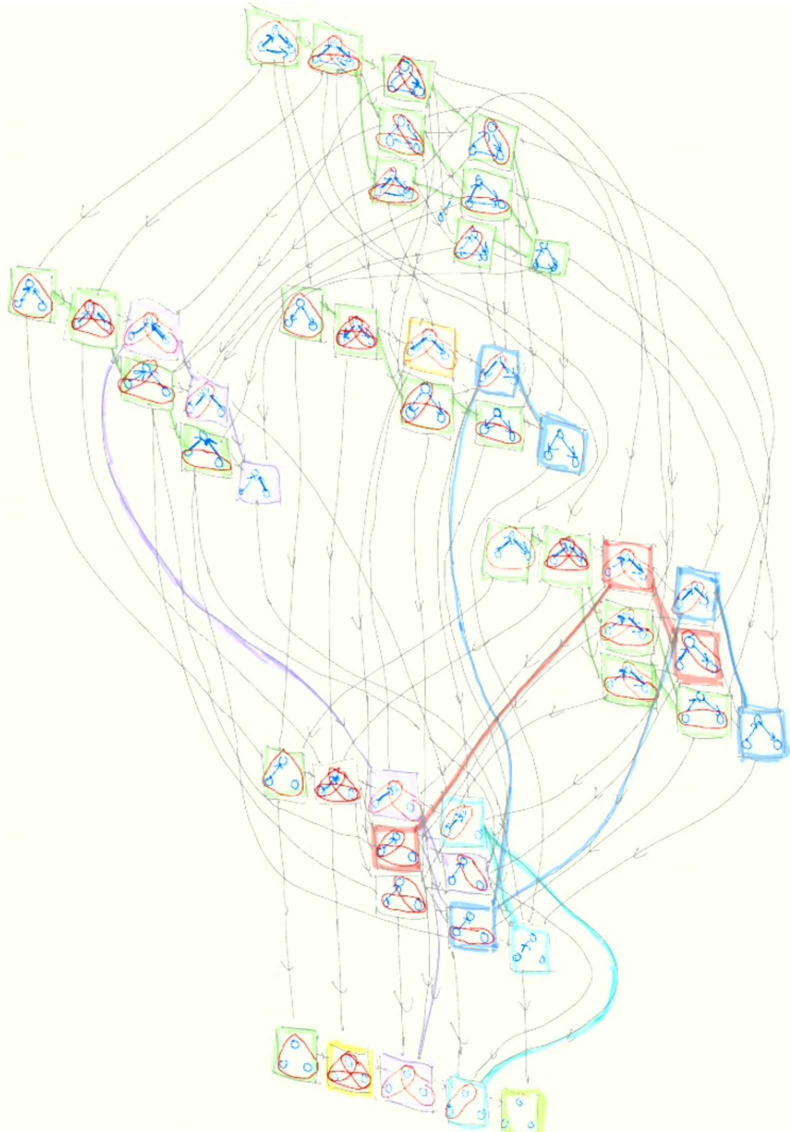


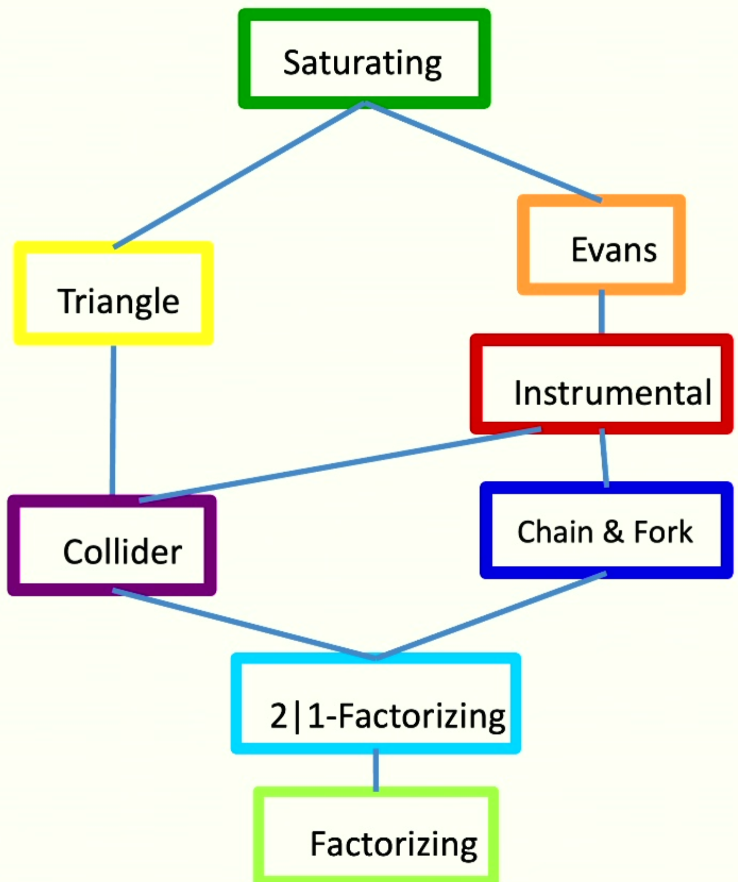




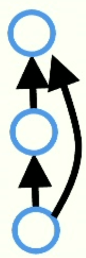


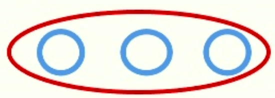





Saturating
Triangle
Evans
Instrumental
Collider
Chain & Fork
2|1-Factorizing
Factorizing

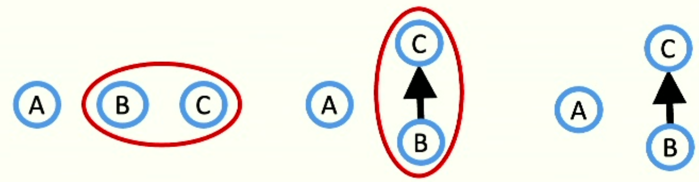
Saturating





+ 21 others

1|2-Factorizing

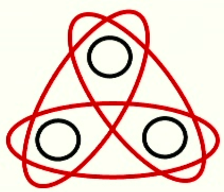


$$A \perp CB$$

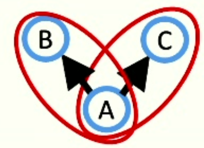
Factorizing

$$A \perp CB, B \perp AC, C \perp AB$$

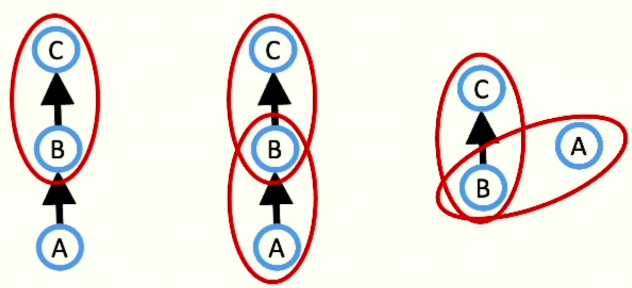
Triangle



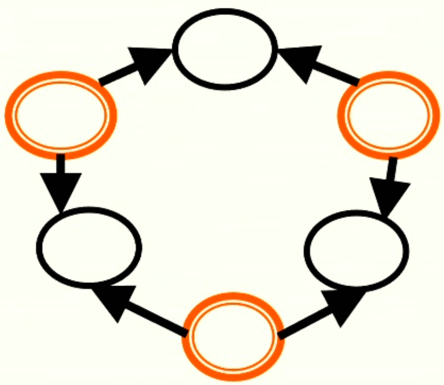
Evans



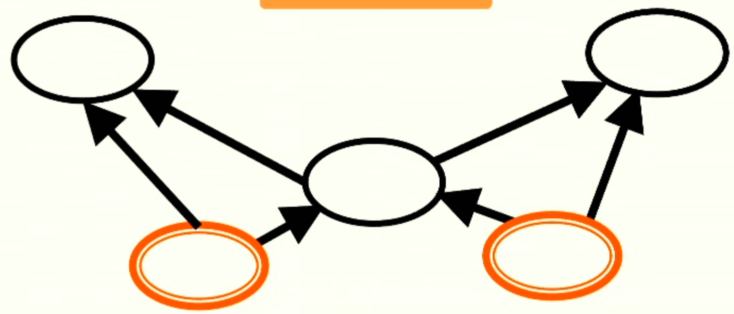
Instrumental



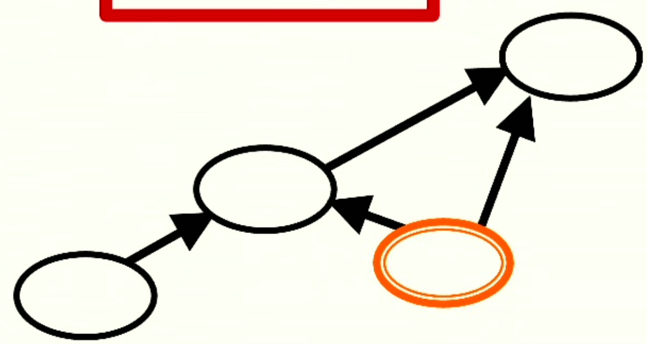
Triangle

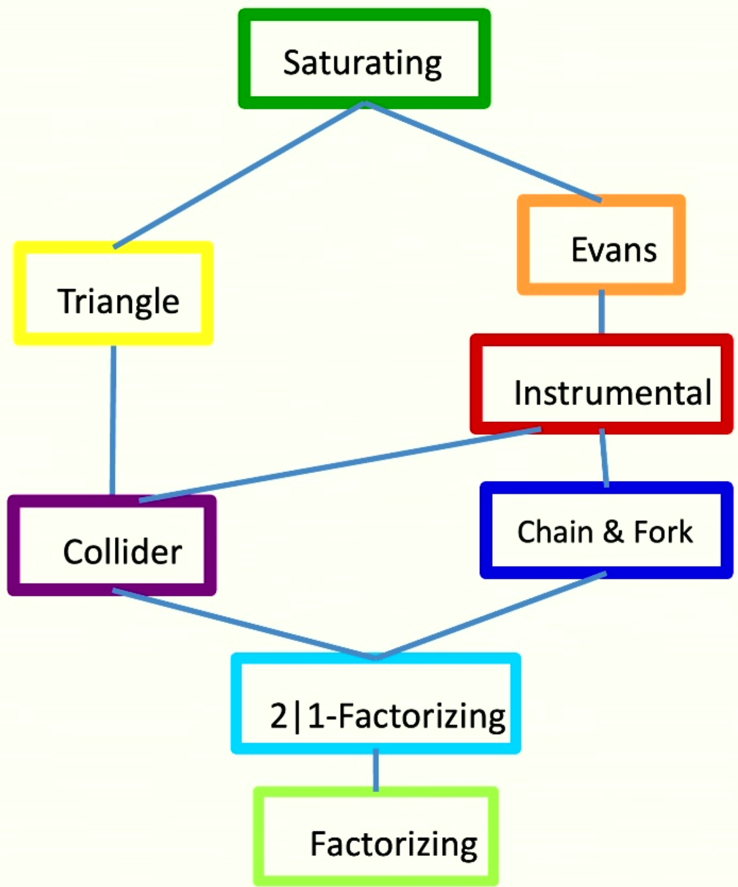


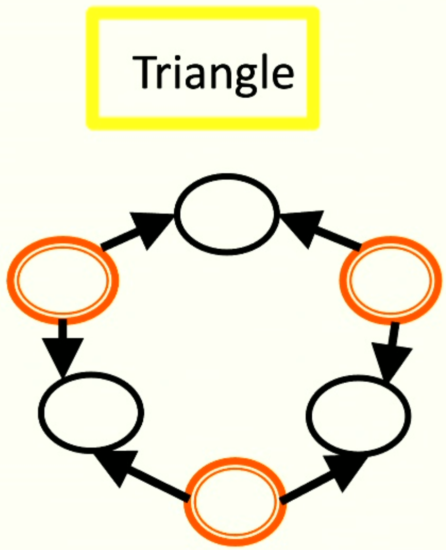
Evans

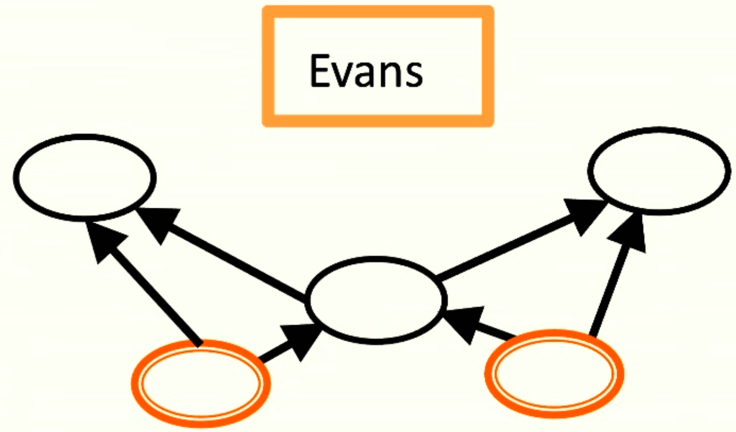


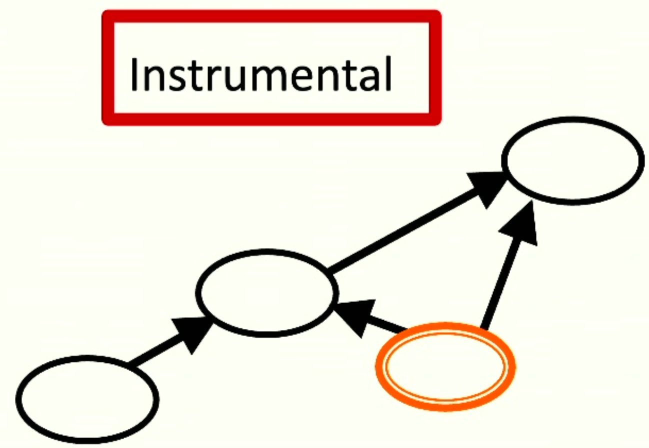
Instrumental




Saturating
Triangle
Evans
Instrumental
Collider
Chain & Fork
2|1-Factorizing
Factorizing


Triangle


Evans


Instrumental

Extending the d-separation theorem to latent-permitting causal models

Extension of d-separation theorem for latent-permitting causal models:

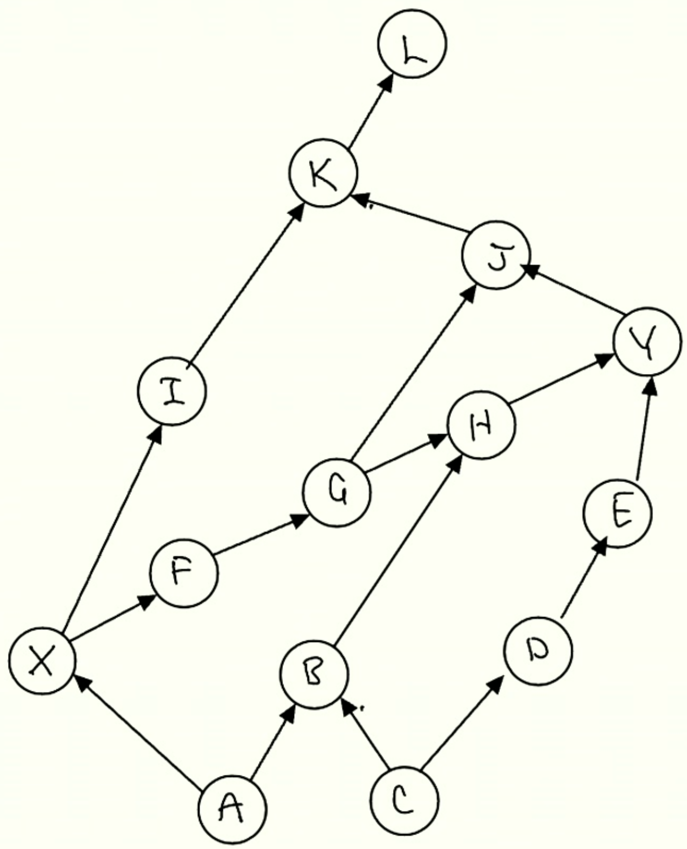
Consider a latent-permitting causal structure G and three disjoint subsets of observed variables **X**, **Y** and **Z**.

Soundness

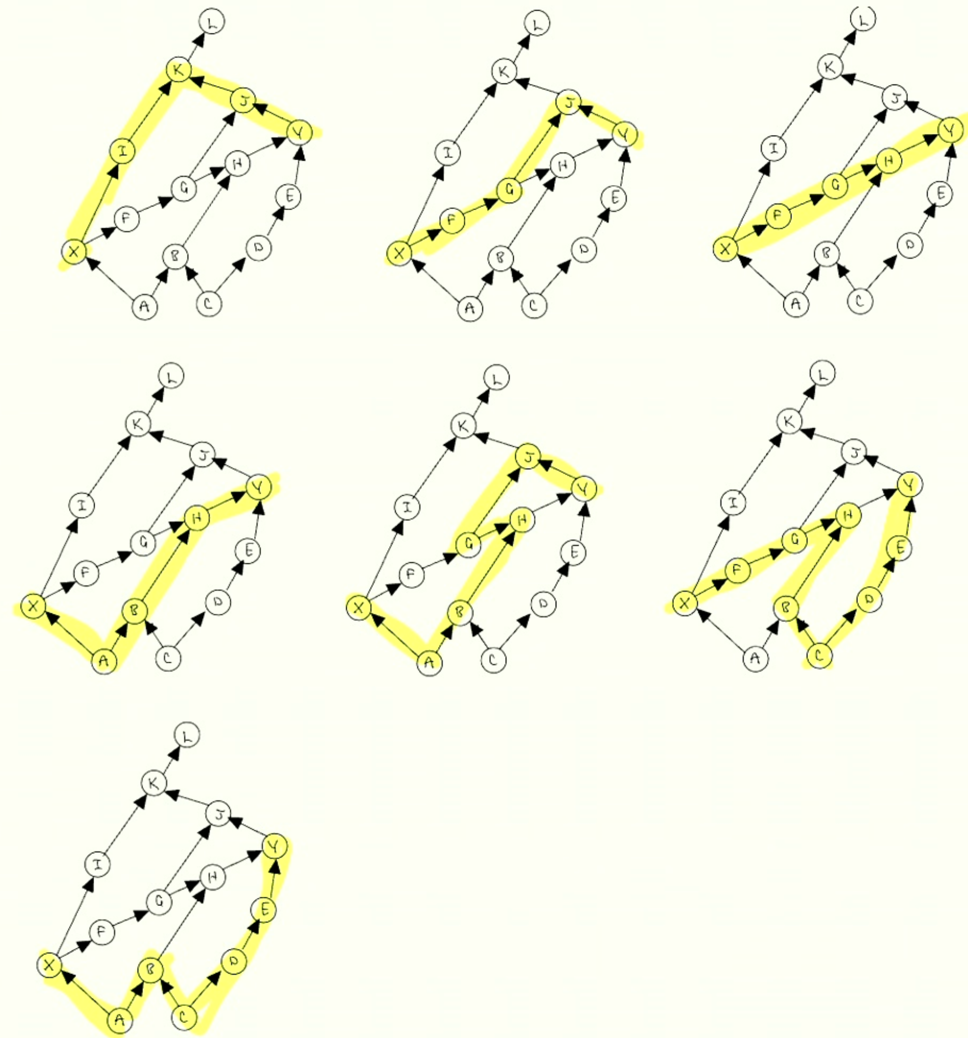
$$\mathbf{X} \perp_d \mathbf{Y}|\mathbf{Z} \text{ in } G \quad \Longrightarrow \quad \forall P \in \mathsf{Comp}_G : \mathbf{X} \perp \mathbf{Y}|\mathbf{Z} \text{ in } P$$

Completeness

$$\forall P \in \mathsf{Comp}_G : \mathbf{X} \perp \mathbf{Y}|\mathbf{Z} \text{ in } P \quad \Longrightarrow \quad \mathbf{X} \perp_d \mathbf{Y}|\mathbf{Z} \text{ in } G$$



$$X \perp Y | GA$$



IC* algorithm and PC algorithm

Set of conditional independence relations on observed variables

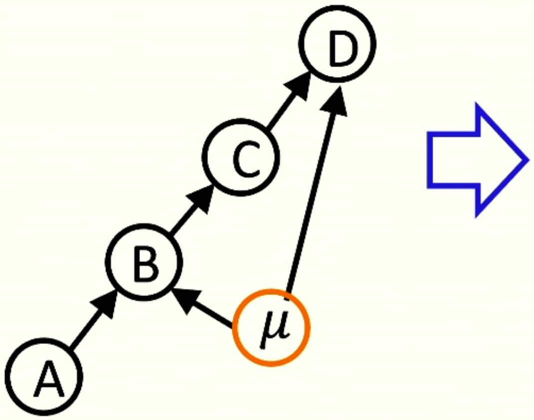
⇒

Find latent-free DAGs that have the right d-separation relations, but these DAGs might still fail to be compatible with the full distribution

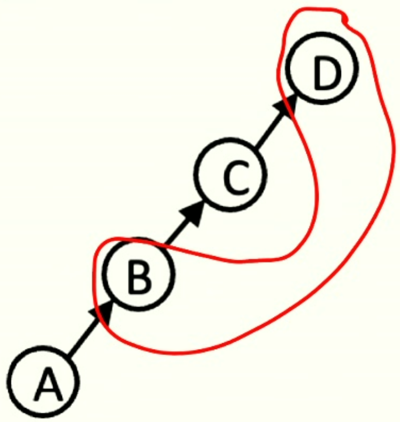
IC* algorithm and PC algorithm

Set of conditional independence relations on observed variables ⇒ Find latent-free DAGs that have the right d-separation relations, but these DAGs might still fail to be compatible with the full distribution

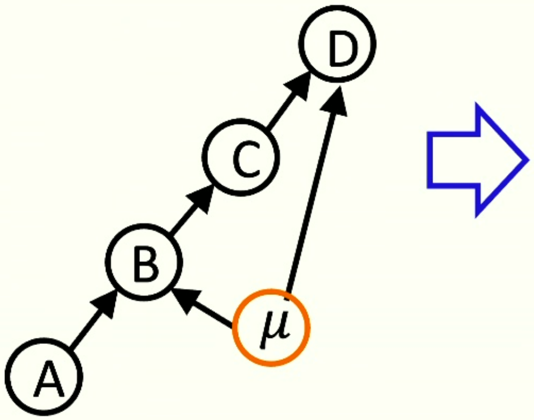
Example: CI relations of quantum-realizable Bell correlations yield classical Bell model



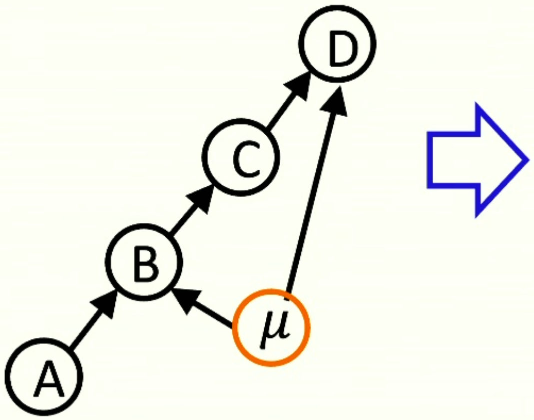
$$A \perp C \mid B$$



Verma graph



$$
\begin{aligned}
\mathrm{P}_{ABCD} &= \sum_{\mu} \mathrm{P}_{D|\mu C}\mathrm{P}_{C|B}\mathrm{P}_{B|A\mu}\mathrm{P}_{A}\mathrm{P}_{\mu} \\
&= \left(\sum_{\mu} \mathrm{P}_{D|\mu C}\mathrm{P}_{B|A\mu}\mathrm{P}_{\mu}\right)\mathrm{P}_{C|B}\mathrm{P}_{A} \\
&= \mathrm{Q}_{BD|AC}\mathrm{P}_{C|B}\mathrm{P}_{A}
\end{aligned}
$$



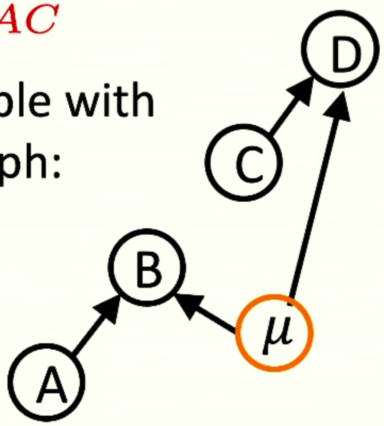
$$P_{ABCD} = \sum_\mu P_{D|\mu C} P_{C|B} P_{B|A\mu} P_A P_\mu$$

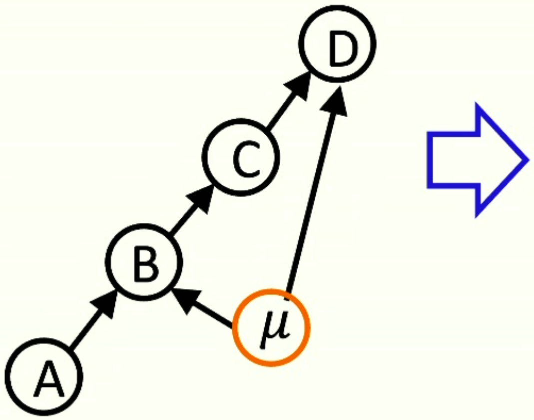
$$= \left(\sum_\mu P_{D|\mu C} P_{B|A\mu} P_\mu \right) P_{C|B} P_A$$

$$= Q_{BD|AC} P_{C|B} P_A$$

$Q_{BD|AC}$

Is compatible with the subgraph:

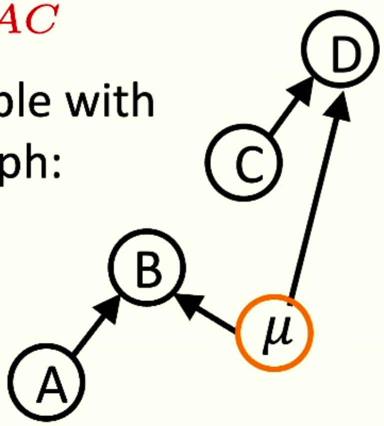




$$P_{ABCD} = \sum_{\mu} P_{D|\mu C} P_{C|B} P_{B|A\mu} P_A P_\mu$$
$$= \left(\sum_{\mu} P_{D|\mu C} P_{B|A\mu} P_\mu \right) P_{C|B} P_A$$
$$= Q_{BD|AC} P_{C|B} P_A$$

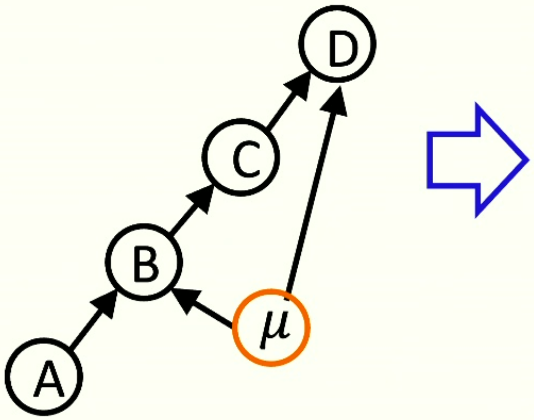
$Q_{BD|AC}$

Is compatible with the subgraph:



This subgraph has d-separation relations implying

$D \perp A|C$ or equivalently, $Q_{D|AC} = Q_{D|C}$



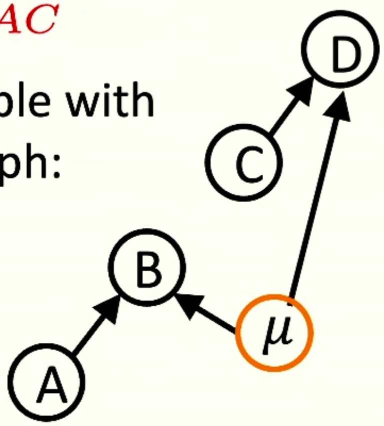
$$\mathrm{P}_{ABCD} = \sum_{\mu} \mathrm{P}_{D|\mu C}\mathrm{P}_{C|B}\mathrm{P}_{B|A\mu}\mathrm{P}_{A}\mathrm{P}_{\mu}$$

$$= \left(\sum_{\mu} \mathrm{P}_{D|\mu C}\mathrm{P}_{B|A\mu}\mathrm{P}_{\mu}\right)\mathrm{P}_{C|B}\mathrm{P}_{A}$$

$$= \mathrm{Q}_{BD|AC}\mathrm{P}_{C|B}\mathrm{P}_{A}$$

$\mathrm{Q}_{BD|AC}$

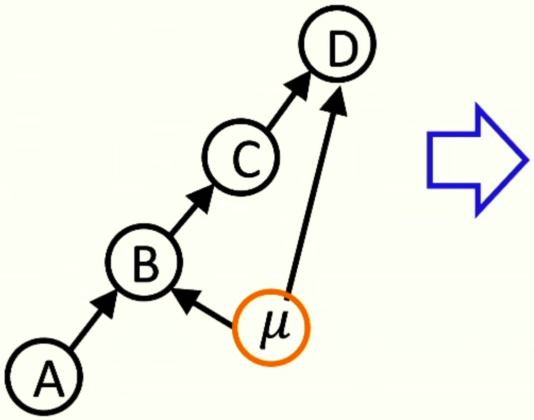
Is compatible with the subgraph:



This subgraph has d-separation relations implying

$D \perp A|C$ or equivalently, $\mathrm{Q}_{D|AC} = \mathrm{Q}_{D|C}$

This implies equality constraints on $\mathrm{Q}_{BD|AC}$ and hence equality constraints on $\mathrm{P}_{ABCD}/\mathrm{P}_{C|B}\mathrm{P}_{A}$



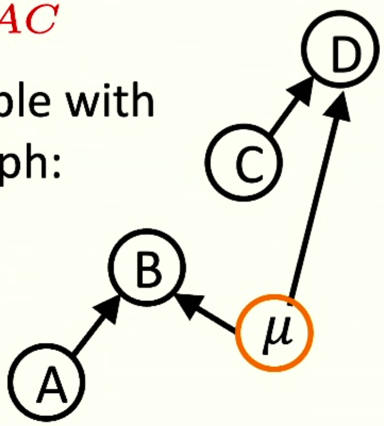
$$\mathrm{P}_{ABCD} = \sum_{\mu} \mathrm{P}_{D|\mu C}\mathrm{P}_{C|B}\mathrm{P}_{B|A\mu}\mathrm{P}_{A}\mathrm{P}_{\mu}$$

$$= \left(\sum_{\mu} \mathrm{P}_{D|\mu C}\mathrm{P}_{B|A\mu}\mathrm{P}_{\mu} \right) \mathrm{P}_{C|B}\mathrm{P}_{A}$$

$$= \mathrm{Q}_{BD|AC}\mathrm{P}_{C|B}\mathrm{P}_{A}$$

$\mathrm{Q}_{BD|AC}$

Is compatible with the subgraph:

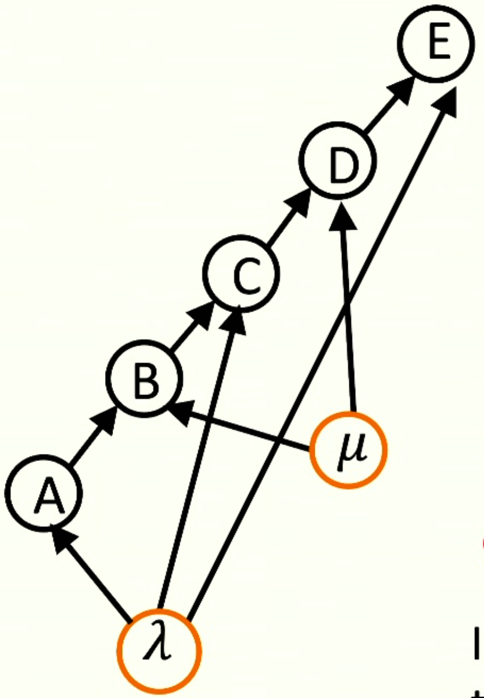


This subgraph has d-separation relations implying

$D \perp A|C$ or equivalently, $\mathrm{Q}_{D|AC} = \mathrm{Q}_{D|C}$

This implies equality constraints on $\mathrm{Q}_{BD|AC}$ and hence equality constraints on $\mathrm{P}_{ABCD}/\mathrm{P}_{C|B}\mathrm{P}_{A}$

Verma constraints

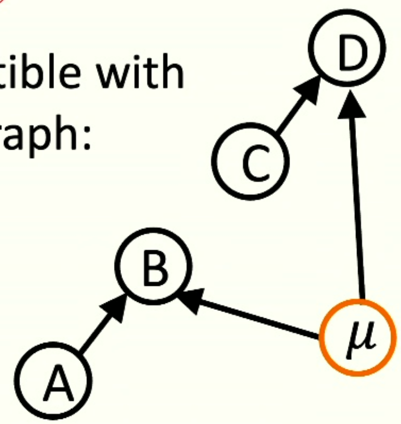


$$P_{ABCDE} = \left(\sum_{\mu} \mathrm{P}_{D|\mu C}\mathrm{P}_{B|A\mu}\mathrm{P}_{\mu}\right) \times \left(\sum_{\lambda} \mathrm{P}_{E|\lambda D}\mathrm{P}_{C|B\lambda}\mathrm{P}_{A|\lambda}\mathrm{P}_{\lambda}\right)$$

$$= \mathrm{Q}_{BD|AC}\mathrm{Q}_{ACE|BD}$$

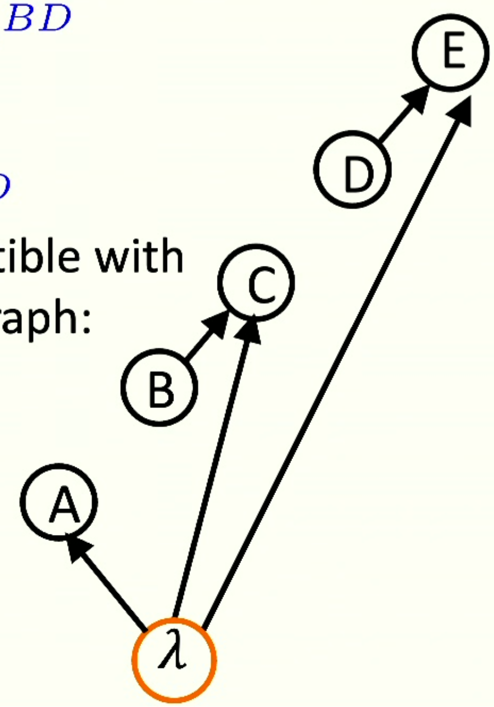
$\mathrm{Q}_{BD|AC}$

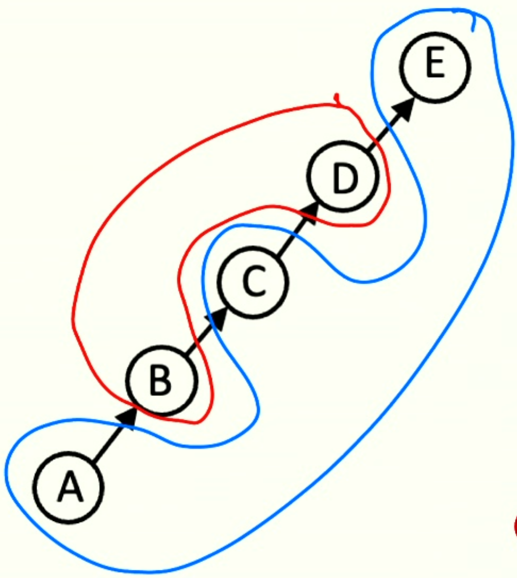
Is compatible with the subgraph:



$\mathrm{Q}_{ACE|BD}$

Is compatible with the subgraph:



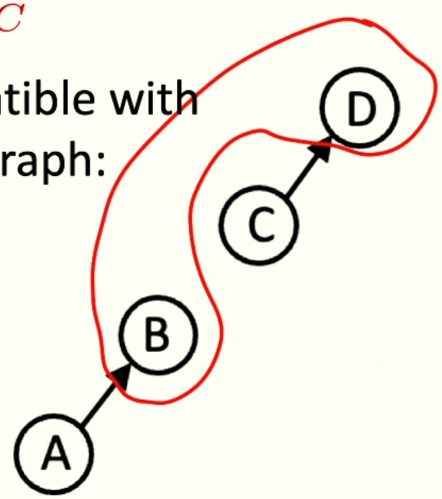


$$P_{ABCDE} = \left(\sum_{\mu} \mathrm{P}_{D|\mu C}\mathrm{P}_{B|A\mu}\mathrm{P}_{\mu}\right) \times \left(\sum_{\lambda} \mathrm{P}_{E|\lambda D}\mathrm{P}_{C|B\lambda}\mathrm{P}_{A|\lambda}\mathrm{P}_{\lambda}\right)$$

$$= \mathrm{Q}_{BD|AC}\mathrm{Q}_{ACE|BD}$$

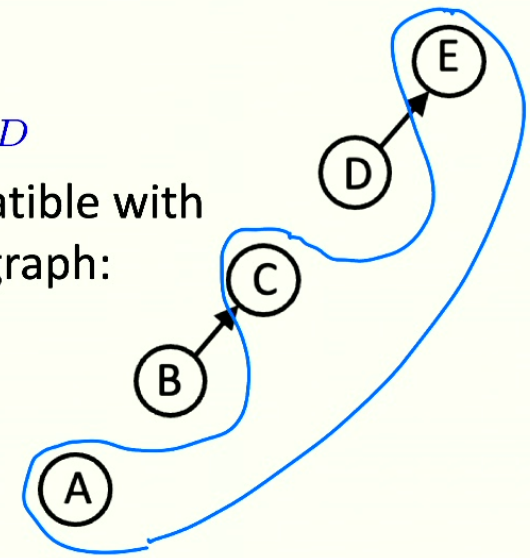
$\mathrm{Q}_{BD|AC}$

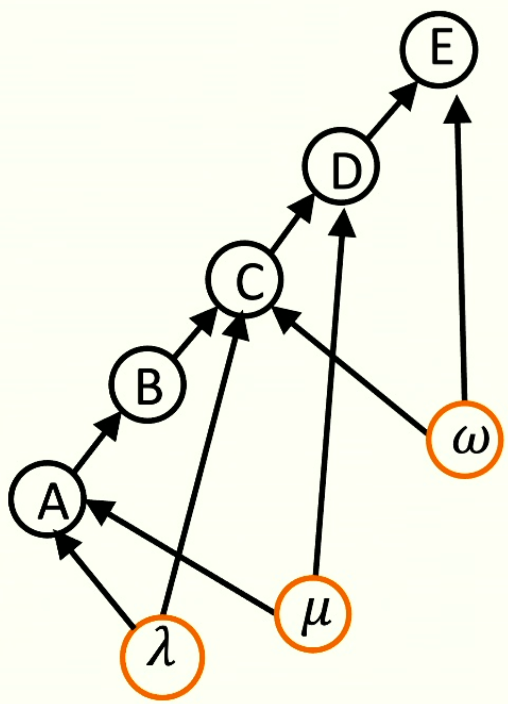
Is compatible with the subgraph:

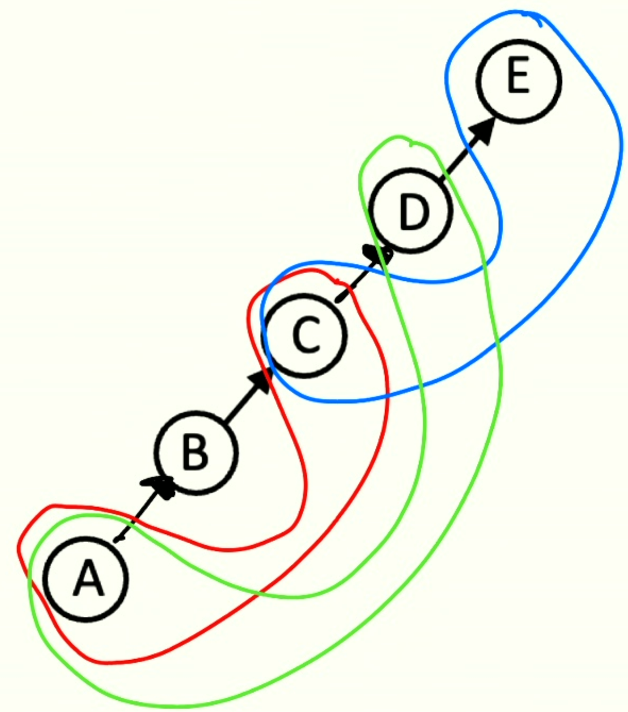


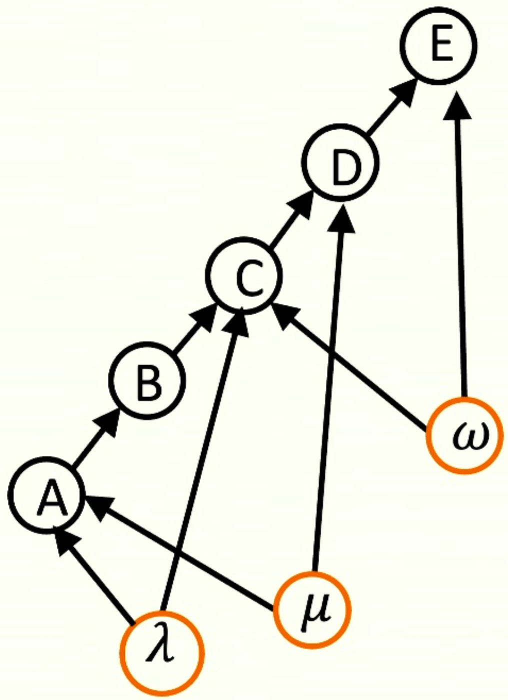
$\mathrm{Q}_{ACE|BD}$

Is compatible with the subgraph:



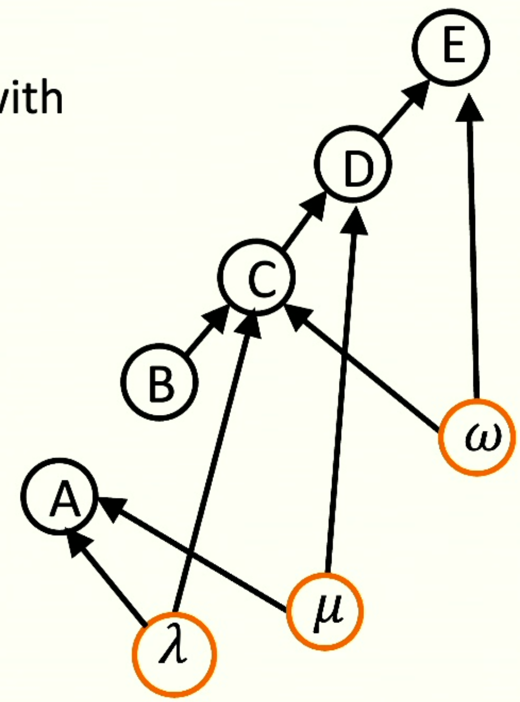

E
D
C
B
A
λ
μ
ω

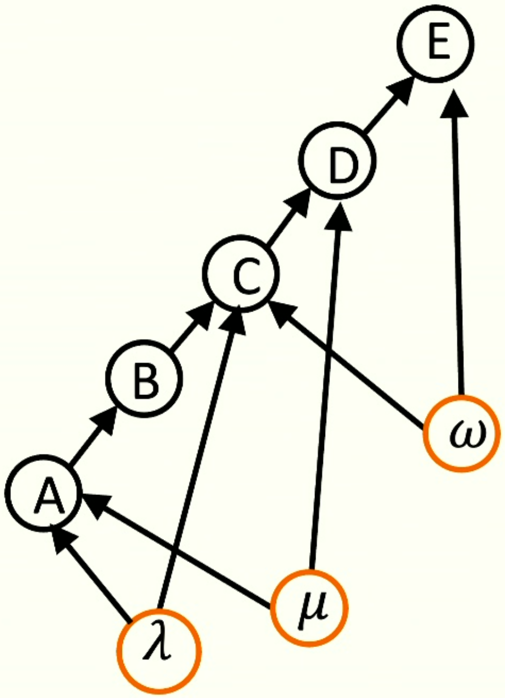

E
D
C
B
A



$$P_{ABCDE} = \left(\sum_{\omega,\mu,\lambda} \mathrm{P}_{E|\omega D}\mathrm{P}_{D|\mu C}\mathrm{P}_{C|\omega\lambda B}\mathrm{P}_{A|\lambda\mu}\mathrm{P}_{\omega}\mathrm{P}_{\mu}\mathrm{P}_{\lambda} \right) \mathrm{P}_{B|A}$$
$$= \mathrm{Q}_{ACDE|B}\mathrm{P}_{B|A}$$

$\mathrm{Q}_{ACDE|B}$
Is compatible with the subgraph:



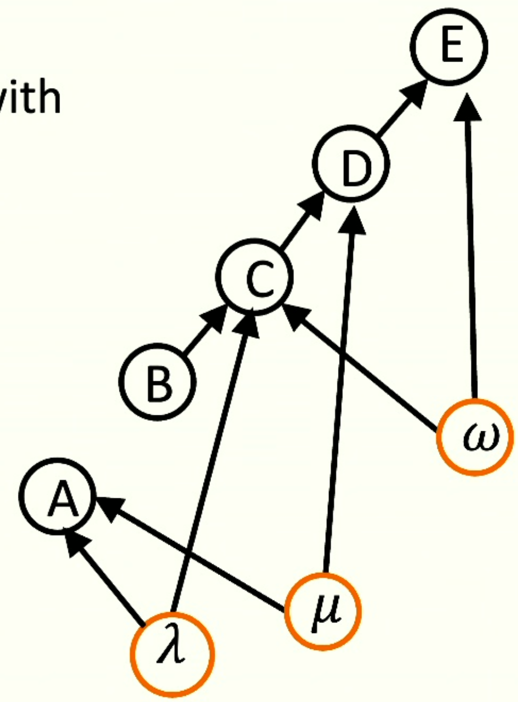


$$P_{ABCDE} = \left(\sum_{\omega,\mu,\lambda} \mathrm{P}_{E|\omega D} \mathrm{P}_{D|\mu C} \mathrm{P}_{C|\omega\lambda B} \mathrm{P}_{A|\lambda\mu} \mathrm{P}_{\omega} \mathrm{P}_{\mu} \mathrm{P}_{\lambda} \right) \mathrm{P}_{B|A}$$

$$= \mathrm{Q}_{ACDE|B} \mathrm{P}_{B|A}$$

$\mathrm{Q}_{ACDE|B}$

Is compatible with the subgraph:



d-separation implies

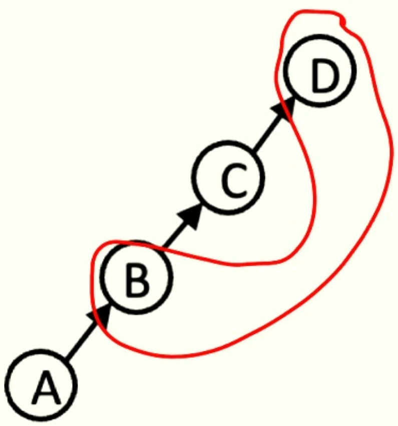
$$B \perp A$$

or equivalently,

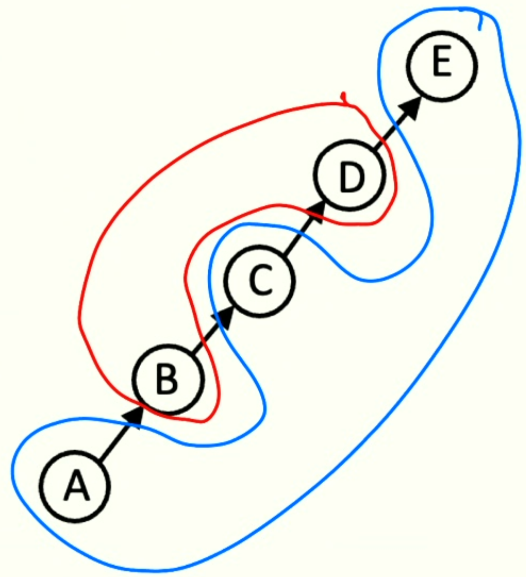
$$\mathrm{Q}_{A|B} = \mathrm{Q}_{A}$$

This implies equality constraint on $\mathrm{Q}_{ACDE|B}$ and hence equality constraints on $\mathrm{P}_{ABCDE} / \mathrm{P}_{B|A}$

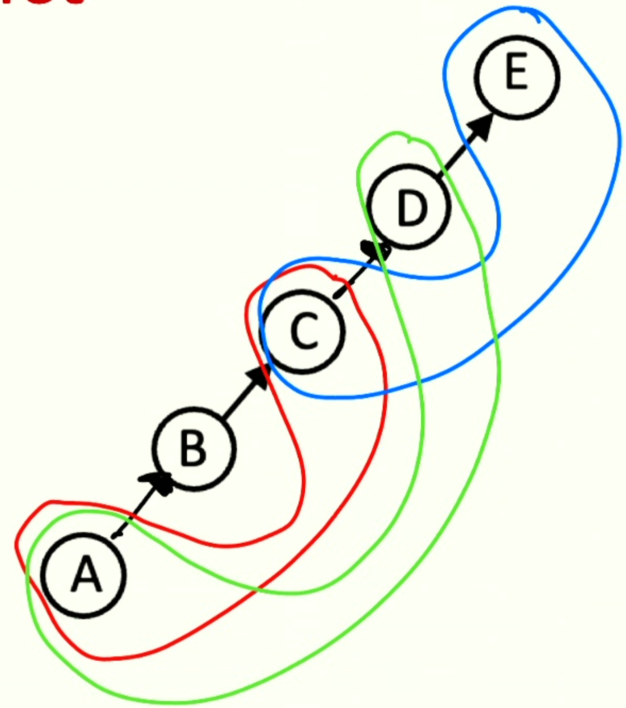
Definition of a district



3 districts:
BD, A, C



2 districts:
BD, ACE



2 districts:
ACDE, B

Algorithm for identifying all nested Markov constraints

Proceeds recursively as follows:

- identify the **districts** within a DAG, relative to which the full distribution factorizes into conditionals of districts given their parents. Demand that each factor satisfy any equality constrains implied by the subgraph associated to the corresponding district and its parents

- For each childless variable in the DAG, identify the subgraph obtained by eliminating it from the DAG and demanding that the distribution obtained by marginalizing over the childless variable satisfies any equality constraint implied by the subgraph

See: R. Evans, Annals of Statistics 46, 2623 (2018)

Next lecture: Inequality constraints for latent-permitting causal models