Title: Lecture - Causal Inference, PHYS 777

Speakers: Robert Spekkens

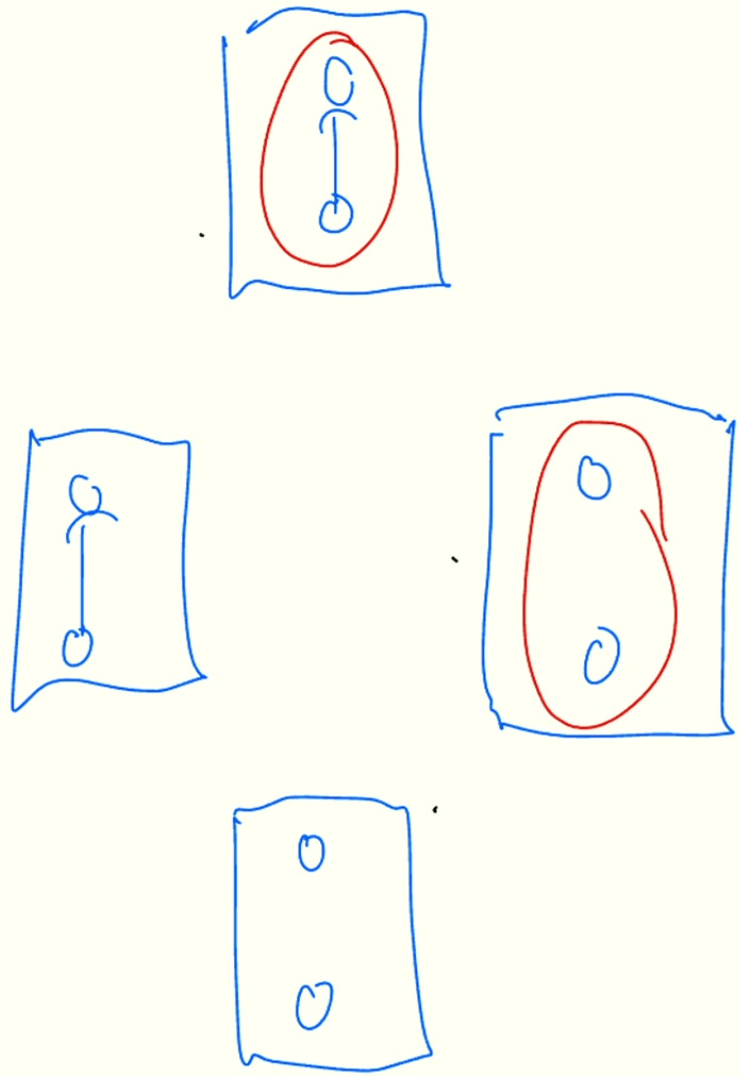
Collection/Series: Causal Inference (Elective), PHYS 777, March 31 - May 2, 2025

Subject: Quantum Foundations

Date: April 14, 2025 - 11:30 AM

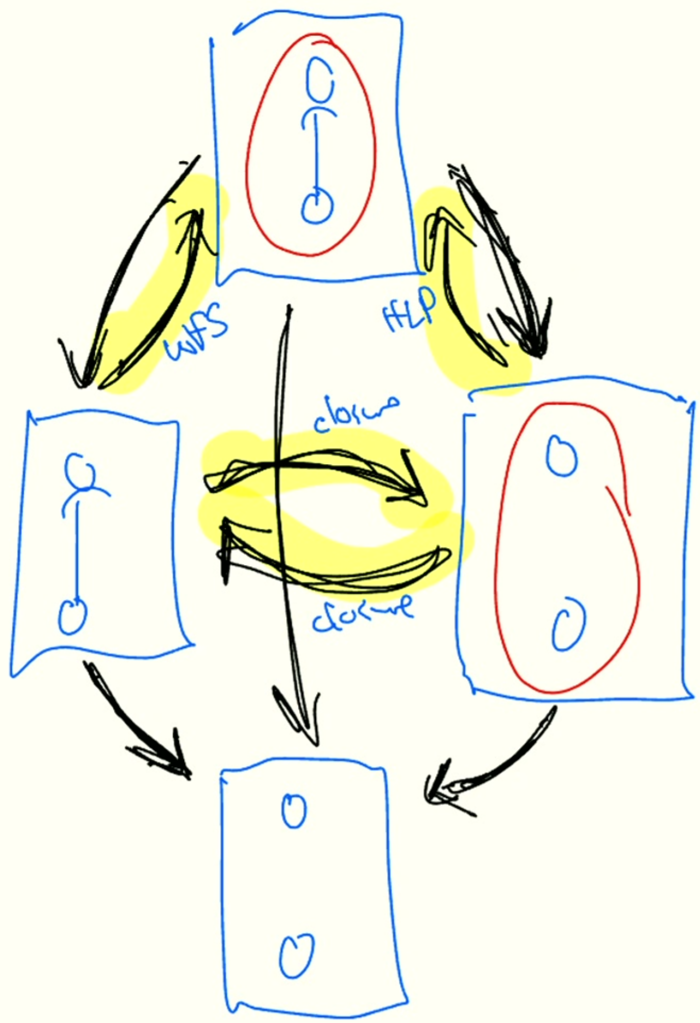
URL: https://pirsa.org/25040041

2-node mDAGs



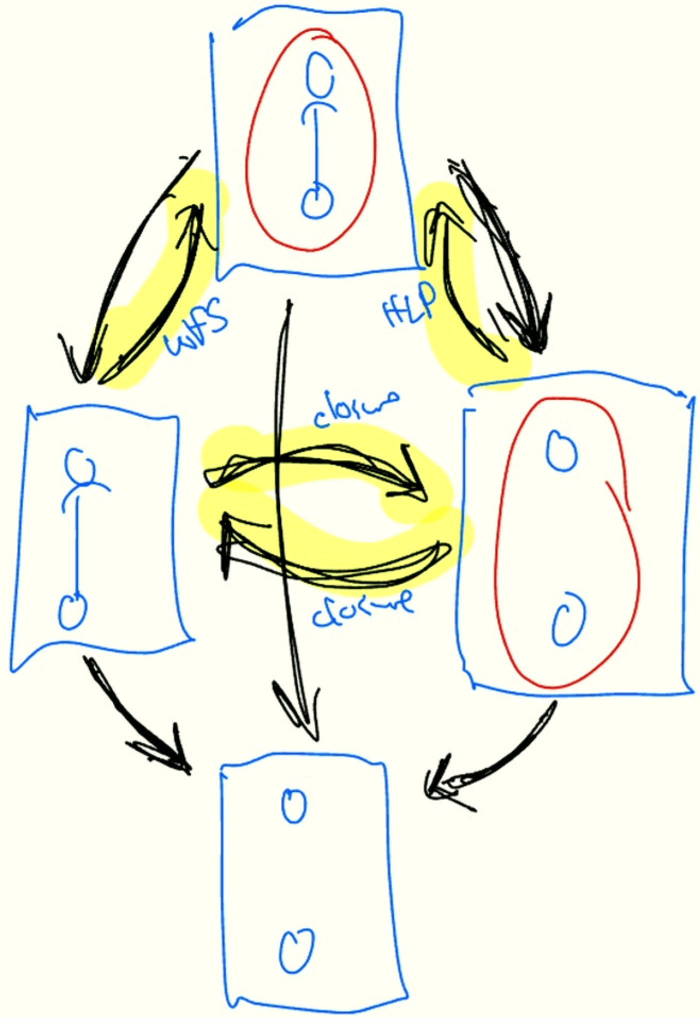
Dominance-proving rules

From HLP Edge-adding rule and Weak Facet-Merging rules, and the transitivity of equivalence, we infer:

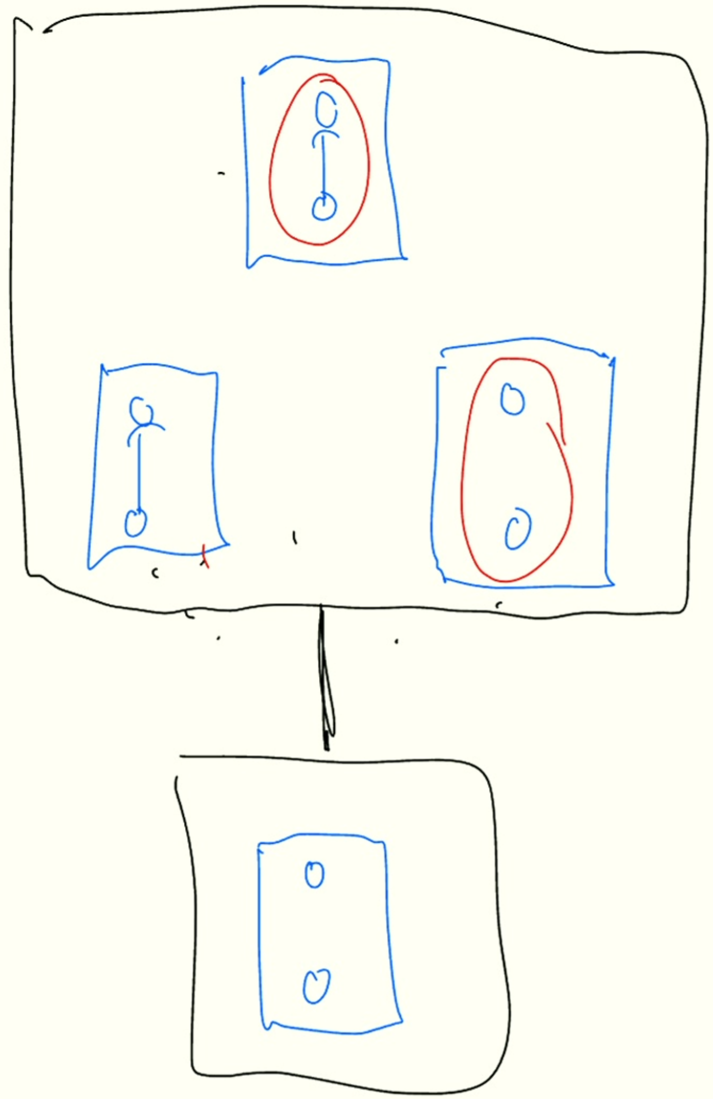


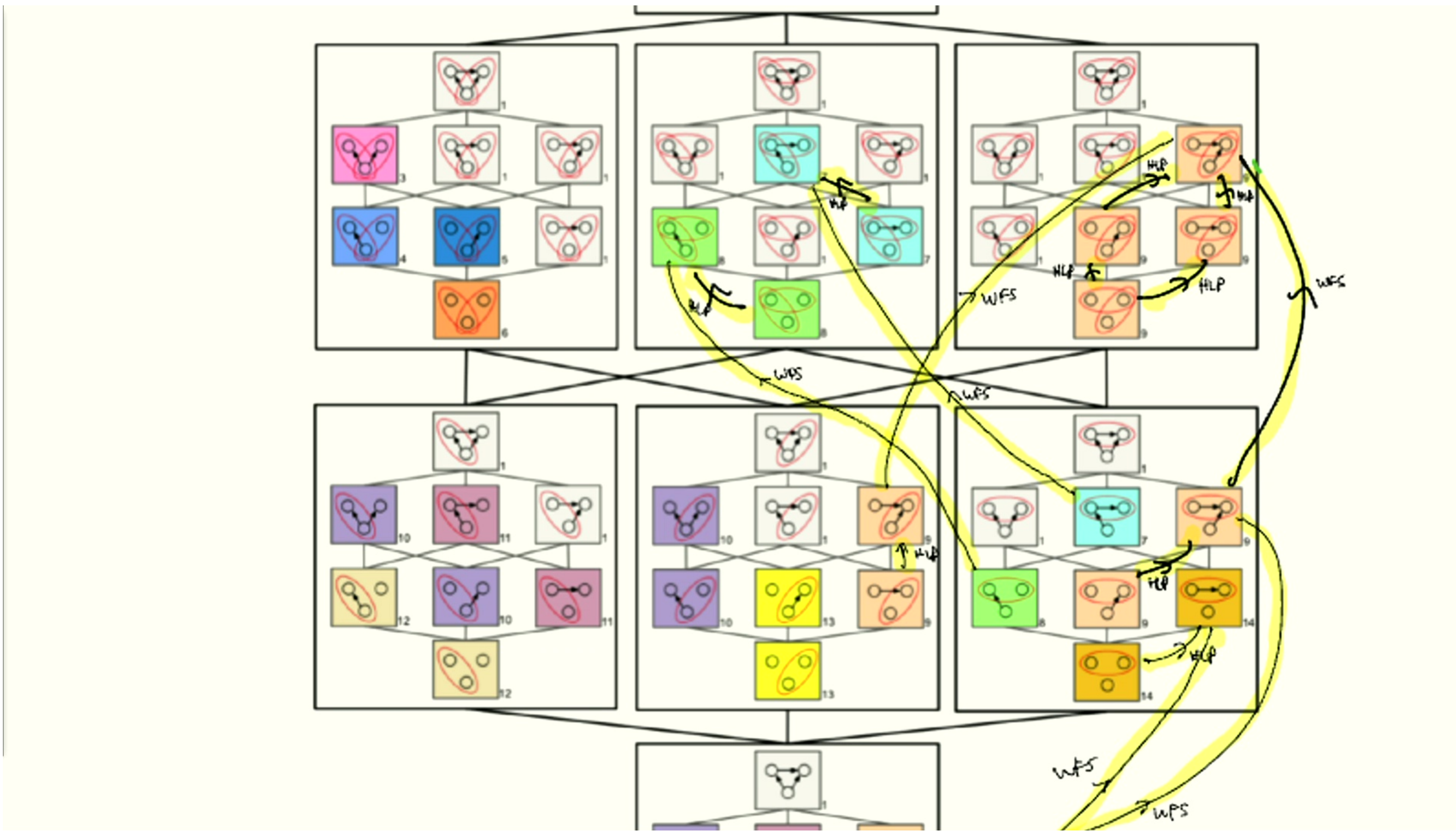
Dominance-proving rules

From HLP Edge-adding rule and Weak Facet-Merging rules, and the transitivity of equivalence, we infer:



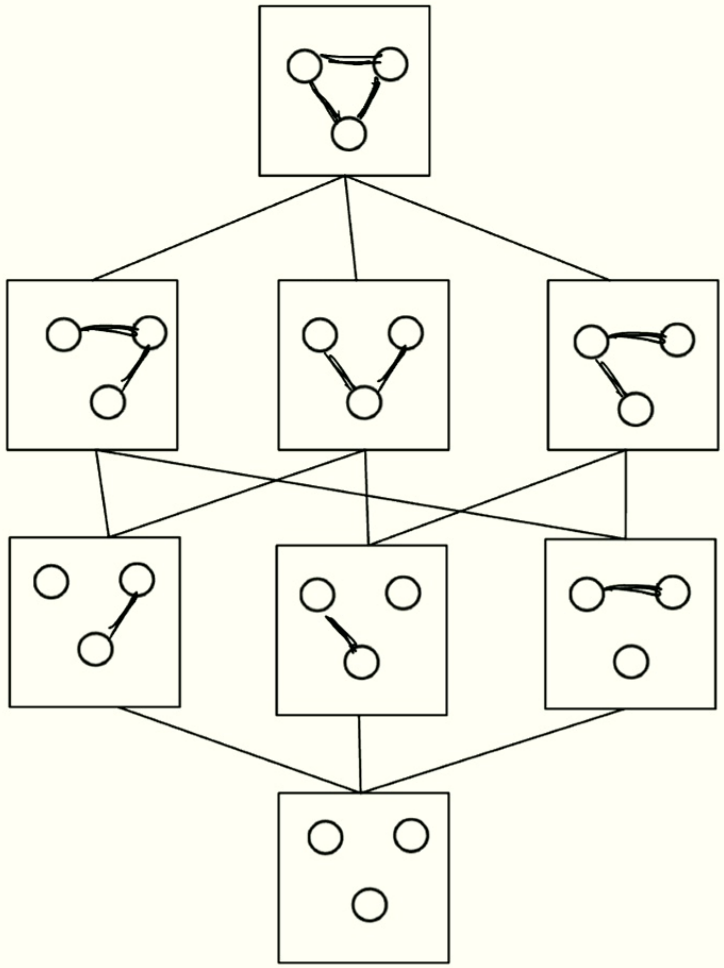
In all, we have:



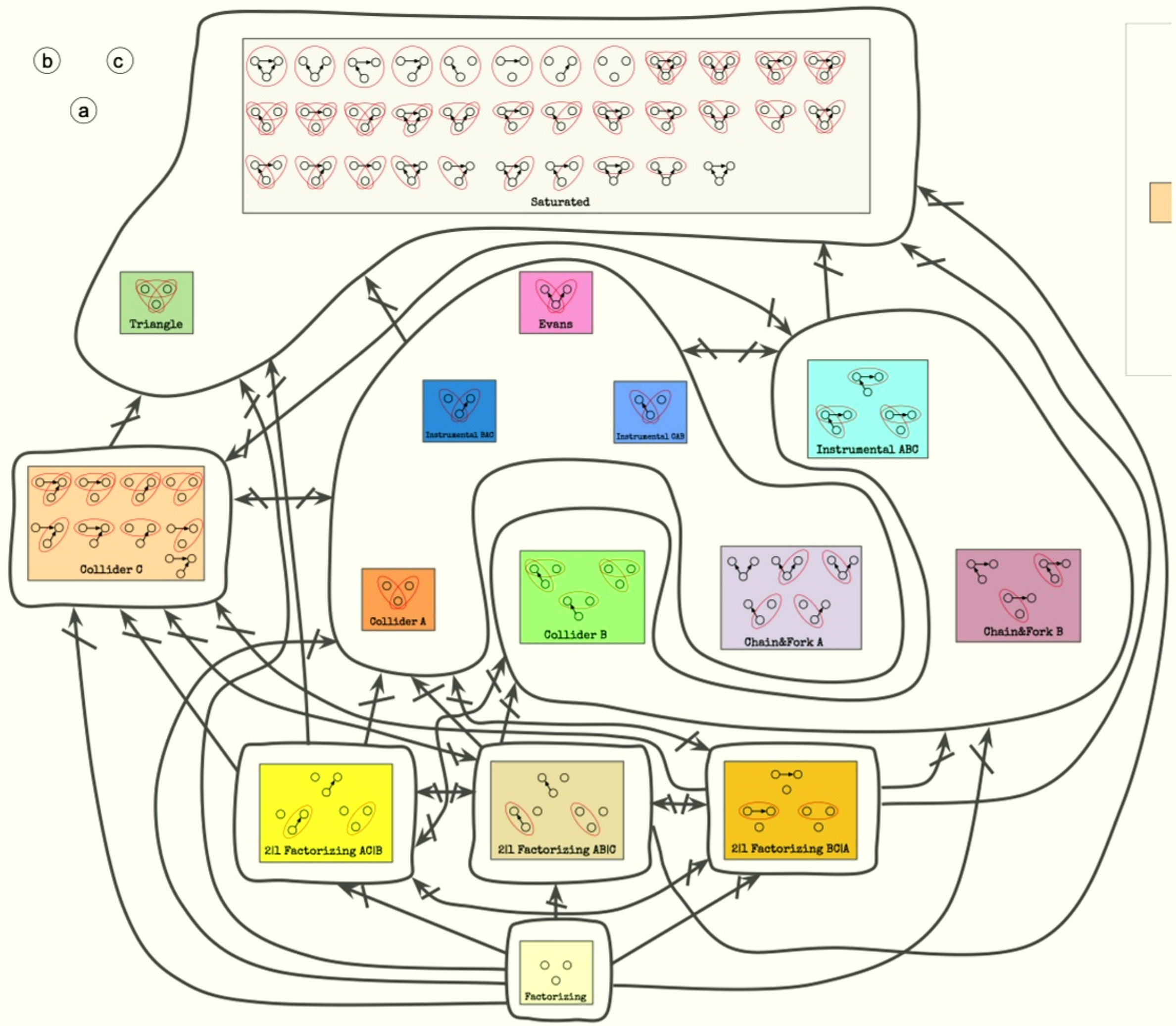


HLP
WFS

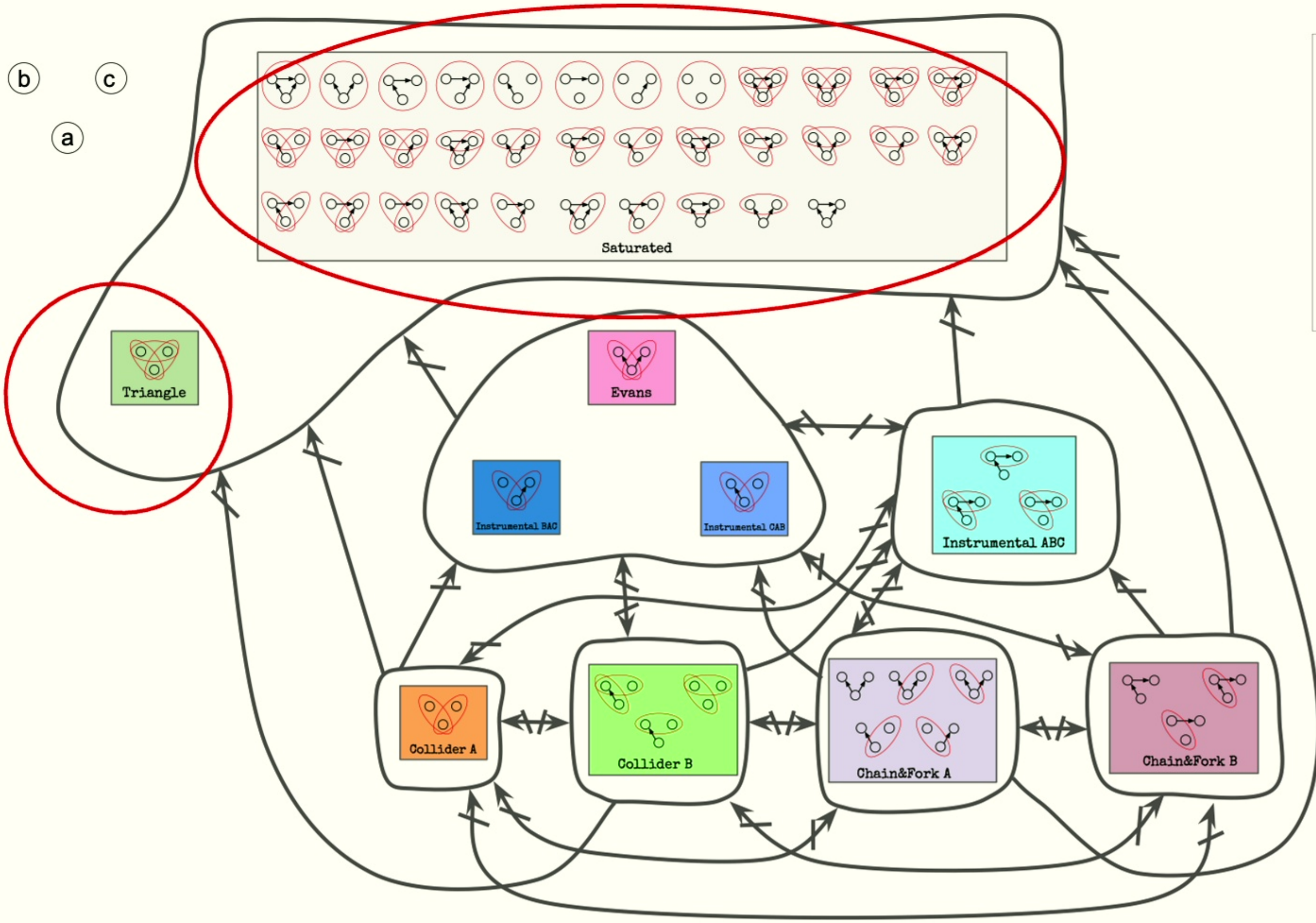
Comparison of skeletons rule

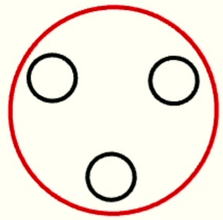


Comparison of skeletons rule



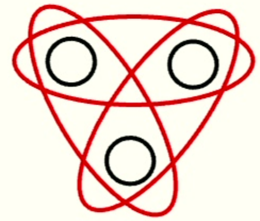
Directed-edge-free rule





Unrealizable supports:

none



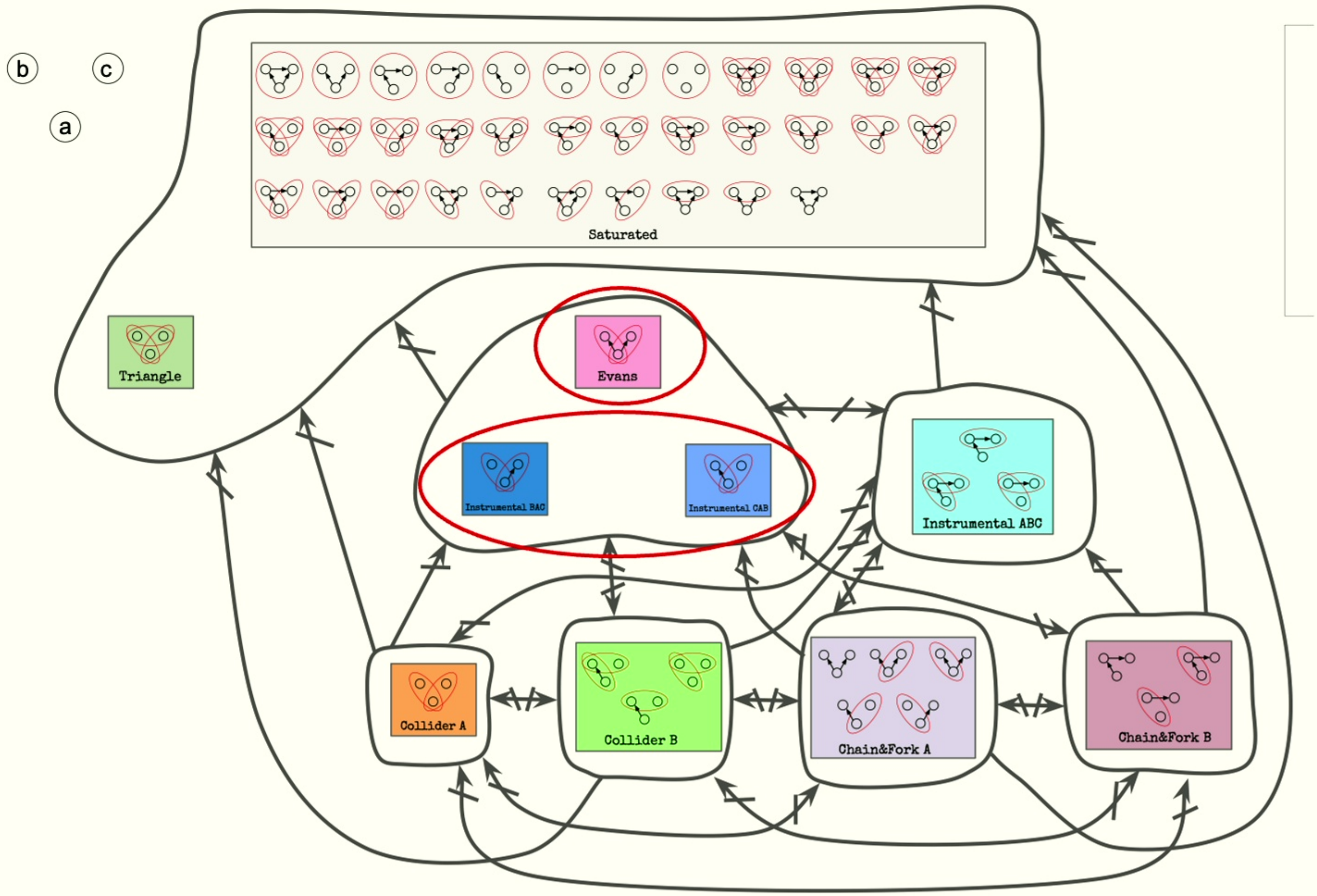
Unrealizable supports:

$$\{\{0,0,0\},\{1,1,1\}\}$$

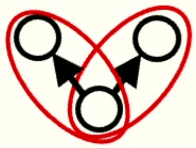
$$\{\{1,0,0\},\{0,1,0\},\{0,0,1\}\}$$

...

Comparison of unrealizable supports rule

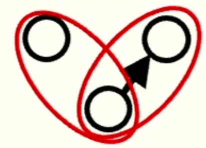


Evans



$\not\preceq_{\text{obs}}$

Instrumental BAC



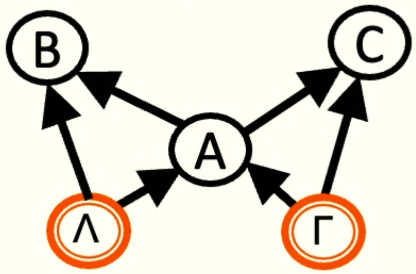
Realizable support

$$\{\{0,0,0\},\{0,1,1\},\{1,0,0\}\}$$

Unrealizable support

$$\{\{0,0,0\},\{0,1,1\},\{1,0,0\}\}$$

Evans



Realizable support on binary (A,B,C)

$$\{\{0,0,0\},\{0,1,1\},\{1,0,0\}\}$$

$$P_{ABC}^{\text{pinch2}} = w[0]_A\,(r[00]_{BC} + (1-r)[11]_{BC}) + (1-w)[1]_A[00]_{BC}$$

Proof: P_Λ is full support

P_Γ is full support

$A = \Lambda \oplus \Gamma$

$B = \Lambda(A \oplus 1)$

$C = \Gamma(A \oplus 1)$

$(\Lambda, \Gamma) \implies (A, B, C)$

$(0,0) \implies (0,0,0)$

$(0,1) \implies (1,0,0)$

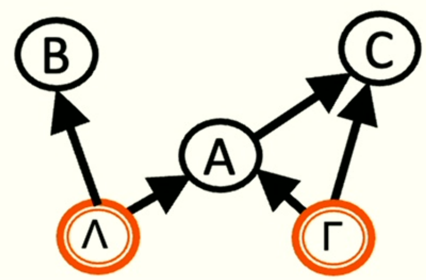
$(1,0) \implies (1,0,0)$

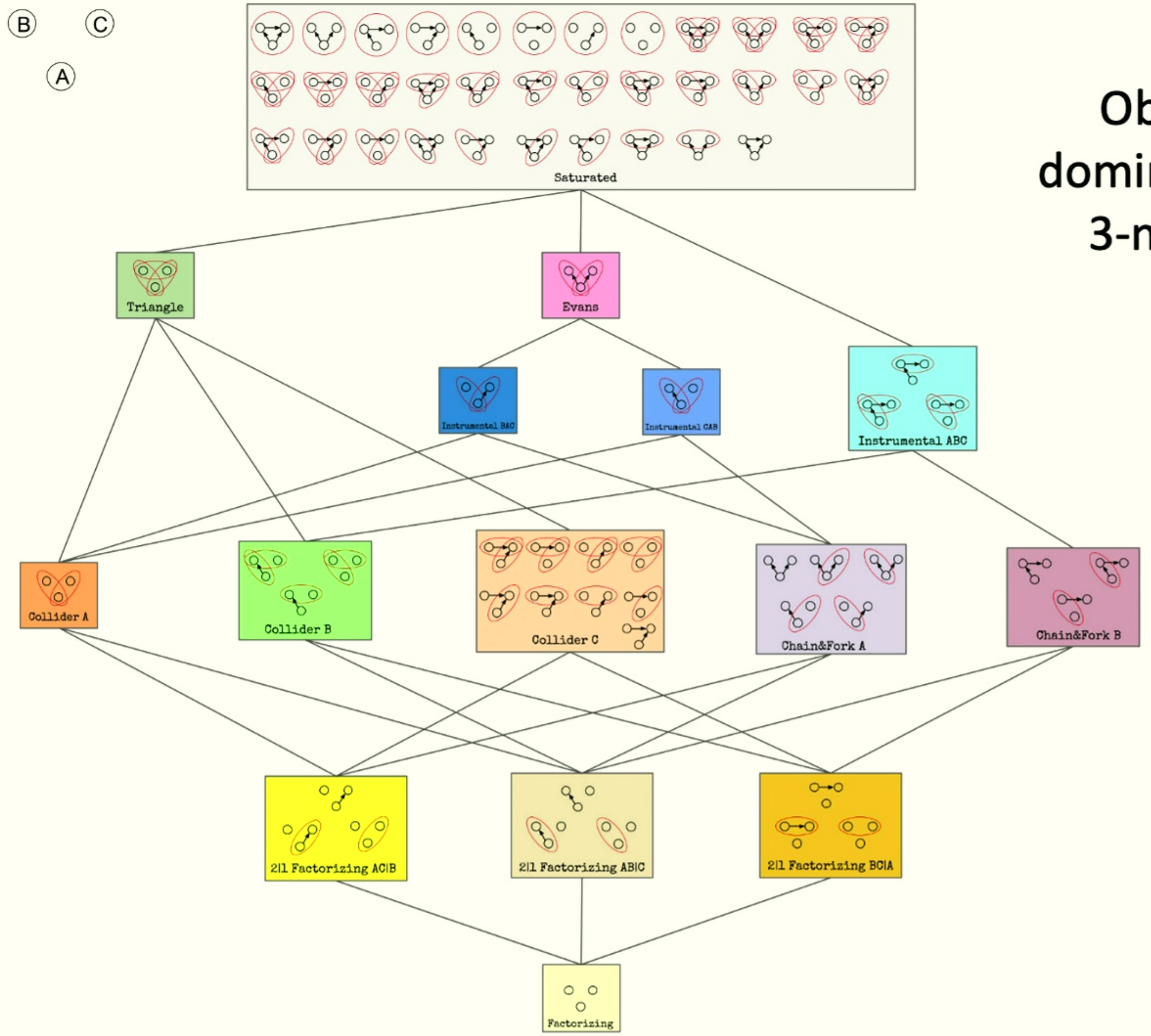
$(1,1) \implies (0,1,1)$

When B and C see that A=0, they track their respective latent variable
When B and C see that A=1, they both output 0

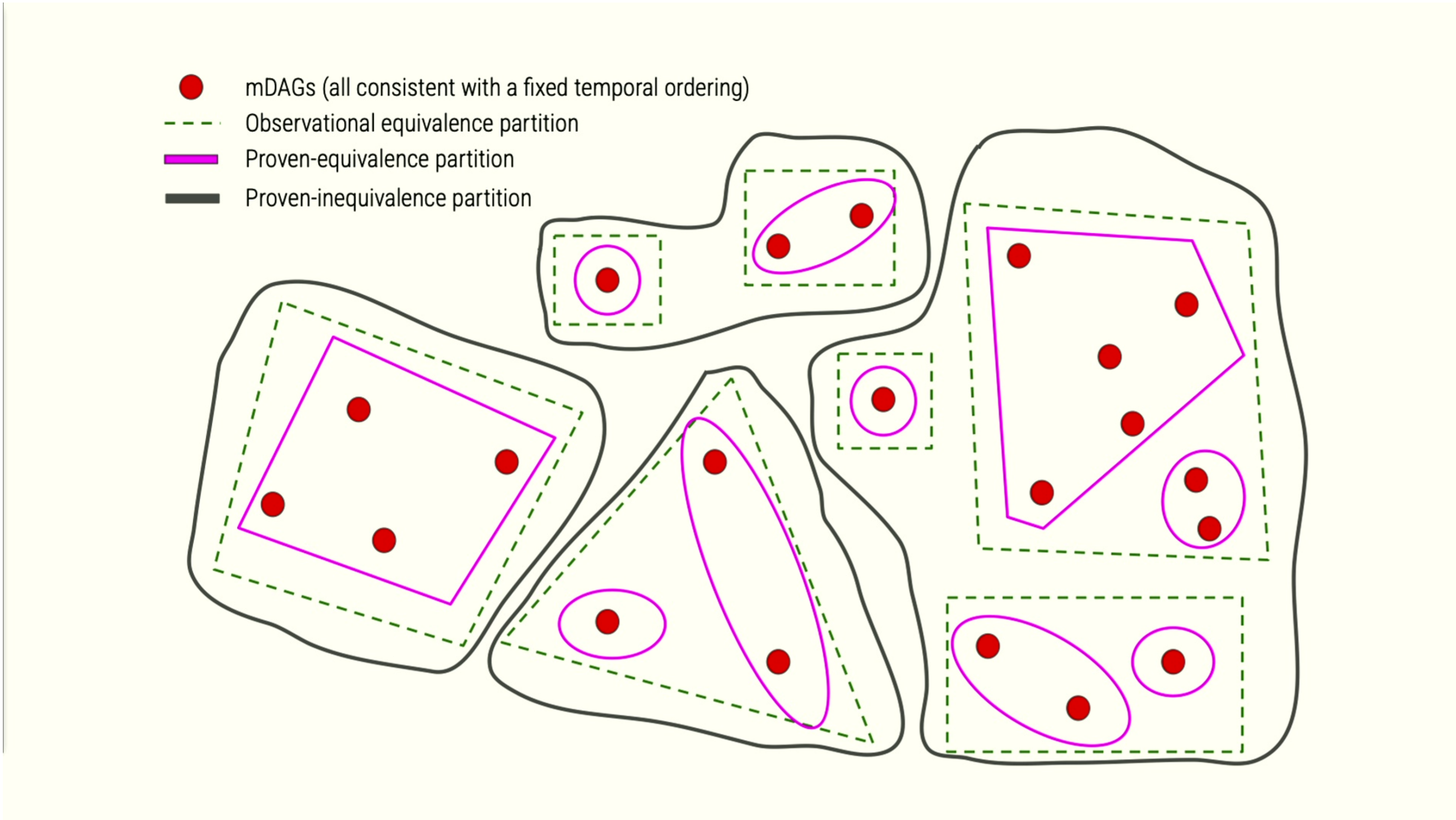
This is not possible if the arrow from A to B is absent

Instrumental BAC





Observational dominance order of 3-node mDAGs

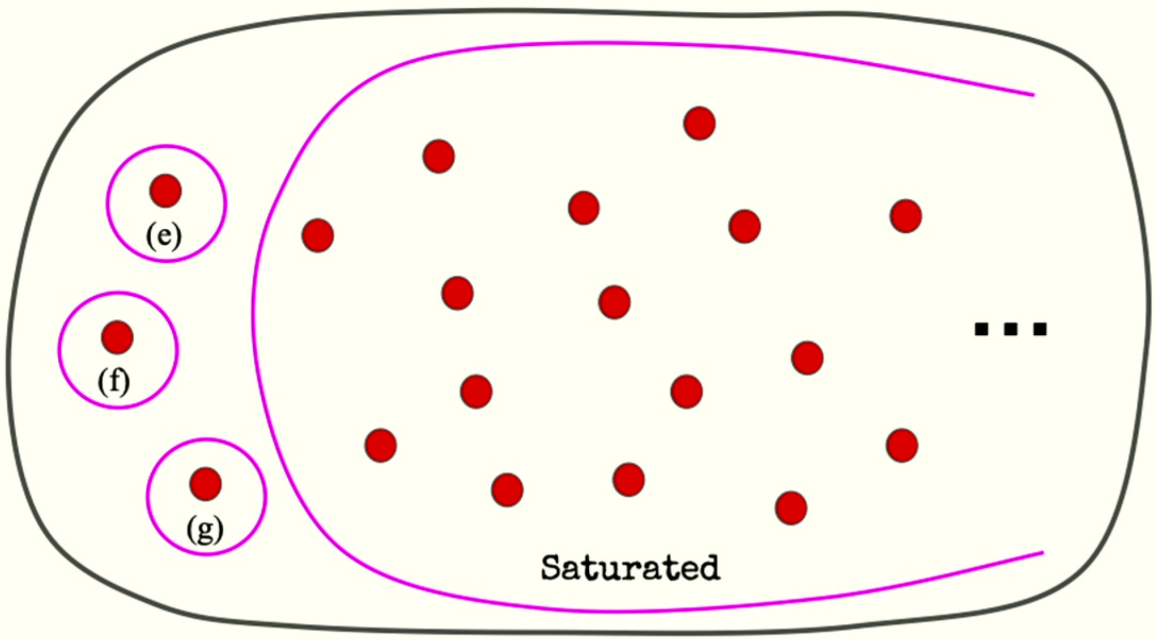

mDAGs (all consistent with a fixed temporal ordering)
Observational equivalence partition
Proven-equivalence partition
Proven-inequivalence partition

7296 mDAGs

Methods Applied	Cardinality of the induced proven-equivalence partition
None	7296
SD + HLP	4417
SD + HLP + Weak FM	1481
SD + HLP + Moderate FM	1466
SD + HLP + Strong FM	1444

Methods Applied	Cardinality of the induced proven-inequivalence partition
None	1
skel	64
d-sep+skel	259
e-sep	326
DC + e-sep	334
DEF + DC + e-sep	350
DEF + DC + e-sep + Supps up to 2 events	447
DEF + DC + e-sep + Supps up to 3 events	595
DEF + DC + e-sep + Supps up to 4 events	1054
DEF + DC + e-sep + Supps up to 5 events	1153
DEF + DC + e-sep + Supps up to 6 events	1243
DEF + DC + e-sep + Supps up to 7 events	1253
DEF + DC + e-sep + Supps up to 8 events	1253

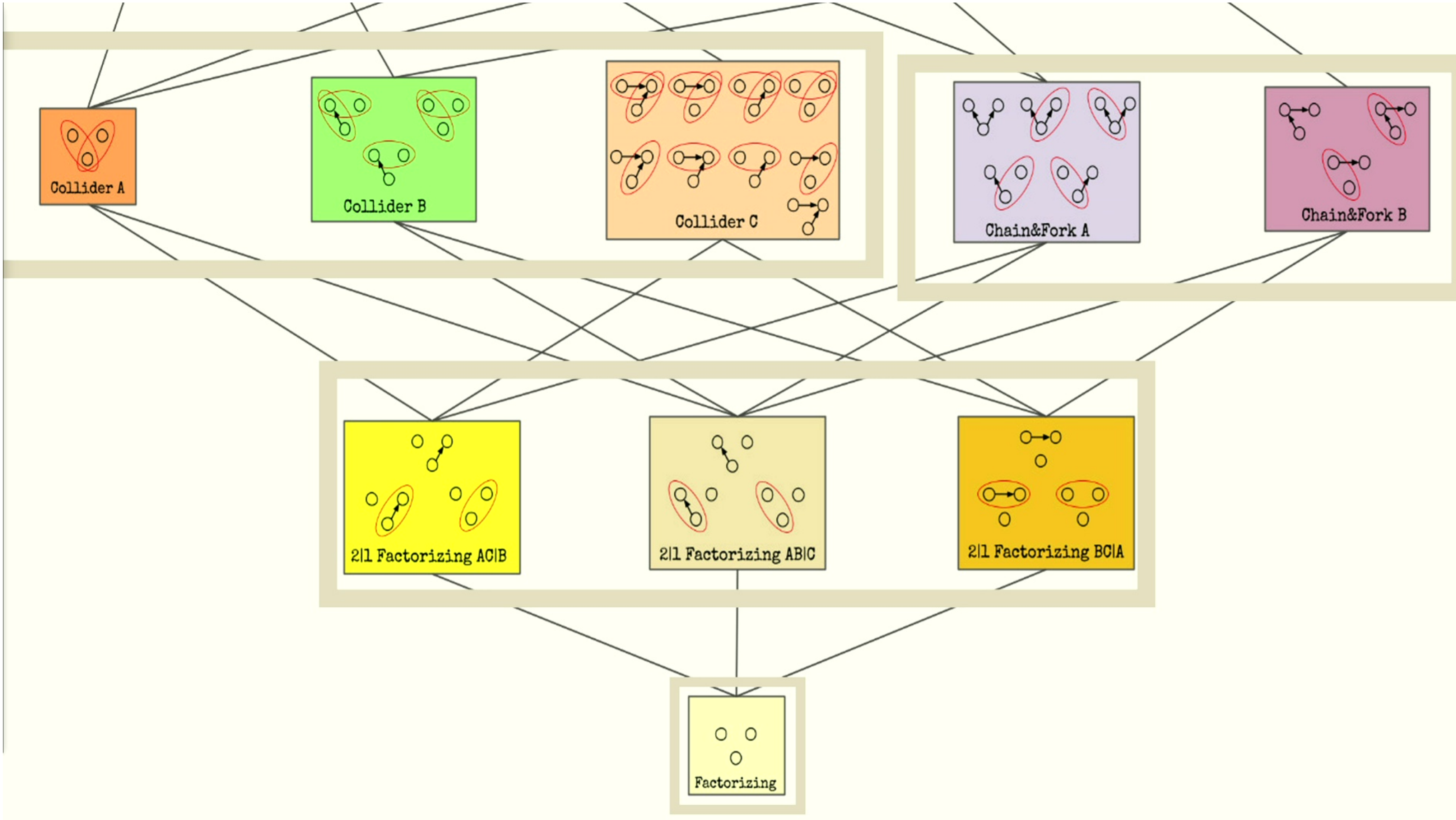
1253 ≤ # of equivalence classes ≤ 1444


(e)
(f)
(g)
Saturated
...

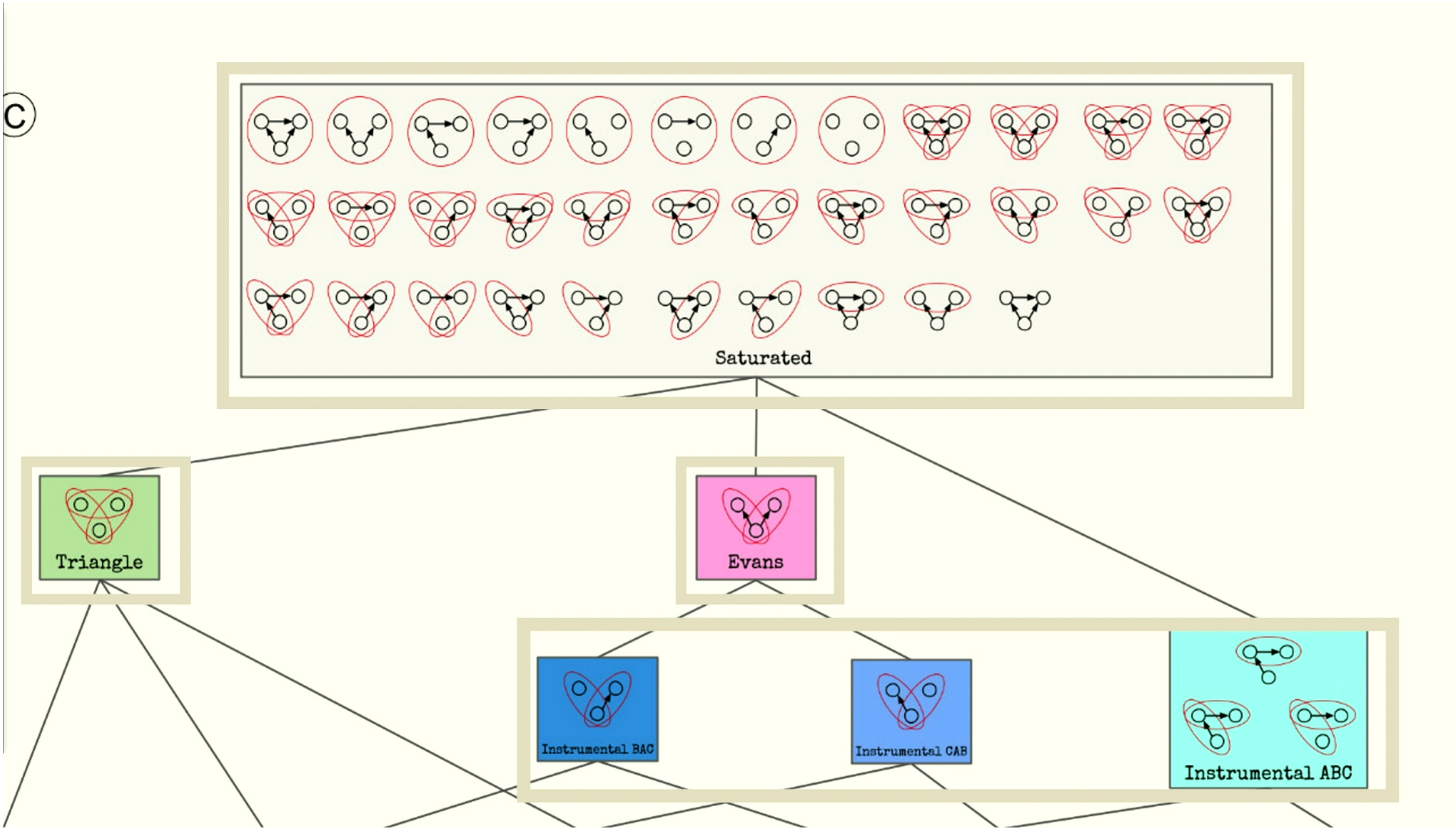
An observational equivalence class of mDAGs has no inequality constraints if it contains a confounder-free mDAG.

Proven in: Robin J. Evans. Latent-free equivalent mDAGs. arXiv:2209.06534, 2022.

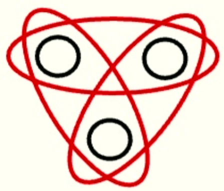
Any observational equivalence class of mDAGs that is related by symmetry to one that contains a confounder-free mDAG also has no inequality constraints



Collider A
Collider B
Collider C
Chain&Fork A
Chain&Fork B
2|1 Factorizing AC|B
2|1 Factorizing AB|C
2|1 Factorizing BC|A
Factorizing

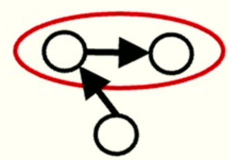

C
Saturated
Triangle
Evans
Instrumental BAC
Instrumental CAB
Instrumental ABC

Triangle

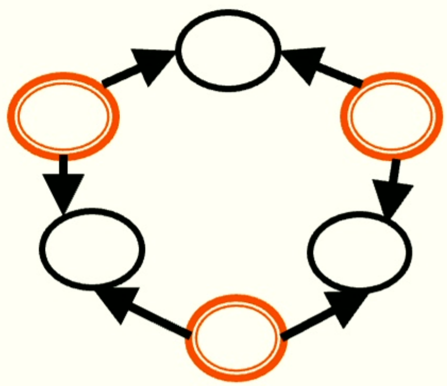


Evans

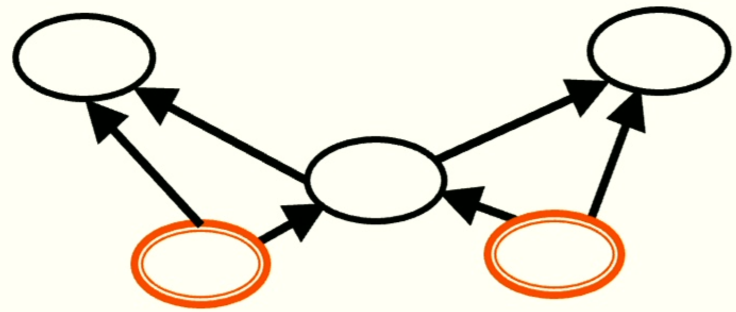
Instrumental



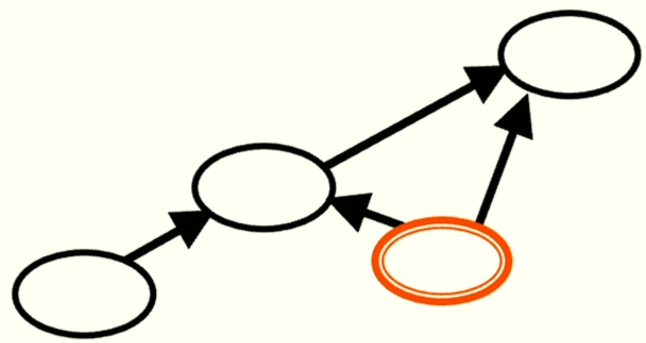
Triangle



Evans



Instrumental



Nonalgebraicness is ubiquitous

4-node mDAGs

Upper bound on the number of algebraic classes
185

Upper bound on the maximum fraction of classes that are algebraic
185/1253 ≈ 14.8%

Lower bound on the fraction of classes that are nonalgebraic
85.2%.

The weakness of conditional independence relations relative to all constraints

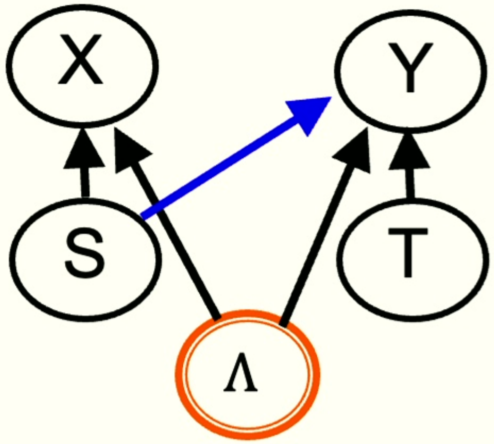
Could inequality constraints still be unimportant relative to conditional independence relations?
For instance, could most equivalence classes still be singled out by their set of conditional independence constraints alone?

The fraction of equivalence classes can be identified by d-separation relations alone:

For 2-node mDAGs: 2 of 2, i.e., 100%
For 3-node mDAGs: 9 of 15, i.e., 60%
For 4-node mDAGs: 115 of N>1253, i.e., less than 10%

Fine-tuning

Causal structure



Parameters

$P_{X|S\Lambda}$

$P_{Y|ST\Lambda}$

P_{Λ}

$$P_{XY|ST} = \sum_{\Lambda} P_{Y|ST\Lambda} P_{X|S\Lambda} P_{\Lambda}$$

Causal compatibility constraints:

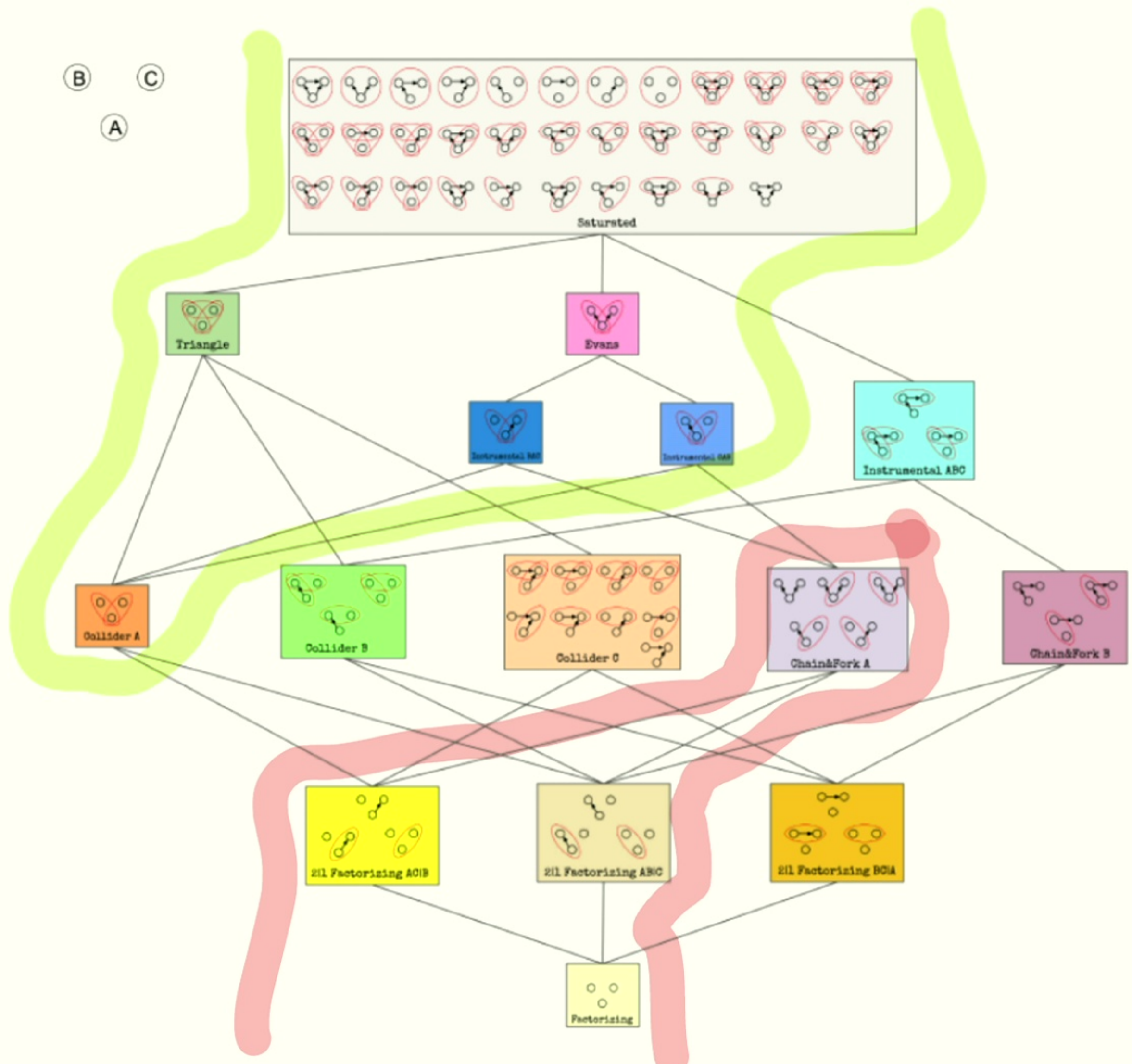
$P_{X|ST} = P_{X|S}$

But the data *also* satisfies $P_{Y|ST} = P_{Y|T}$

Reproducing this requires fine-tuning

Wood and RWS, New J. Phys. 17, 033002 (2015)

Applications of the observational dominance order


B
C
A
Saturated
Triangle
Evans
Instrumental ABC
Collider A
Collider B
Collider C
Chain&Fork A
Chain&Fork B
2|1 Factorizing AC|B
2|1 Factorizing AB|C
2|1 Factorizing BC|A
Factorizing

Applications to causal discovery

- More efficient search for the causal structures that neither underfit *nor overfit* the data

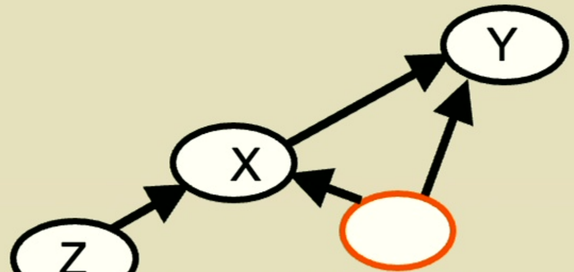
The evidence

Violates one or more inequality constraint

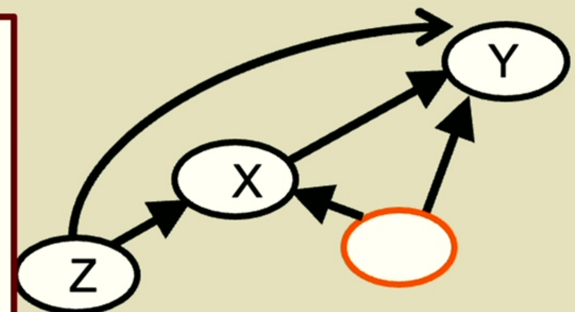
$$P_{XY|Z}(00|0) + P_{XY|Z}(01|1) - 1 \geq 0$$
$$P_{XY|Z}(01|0) + P_{XY|Z}(00|1) - 1 \geq 0$$
$$P_{XY|Z}(10|0) + P_{XY|Z}(11|1) - 1 \geq 0$$
$$P_{XY|Z}(10|0) + P_{XY|Z}(11|1) - 1 \geq 0$$

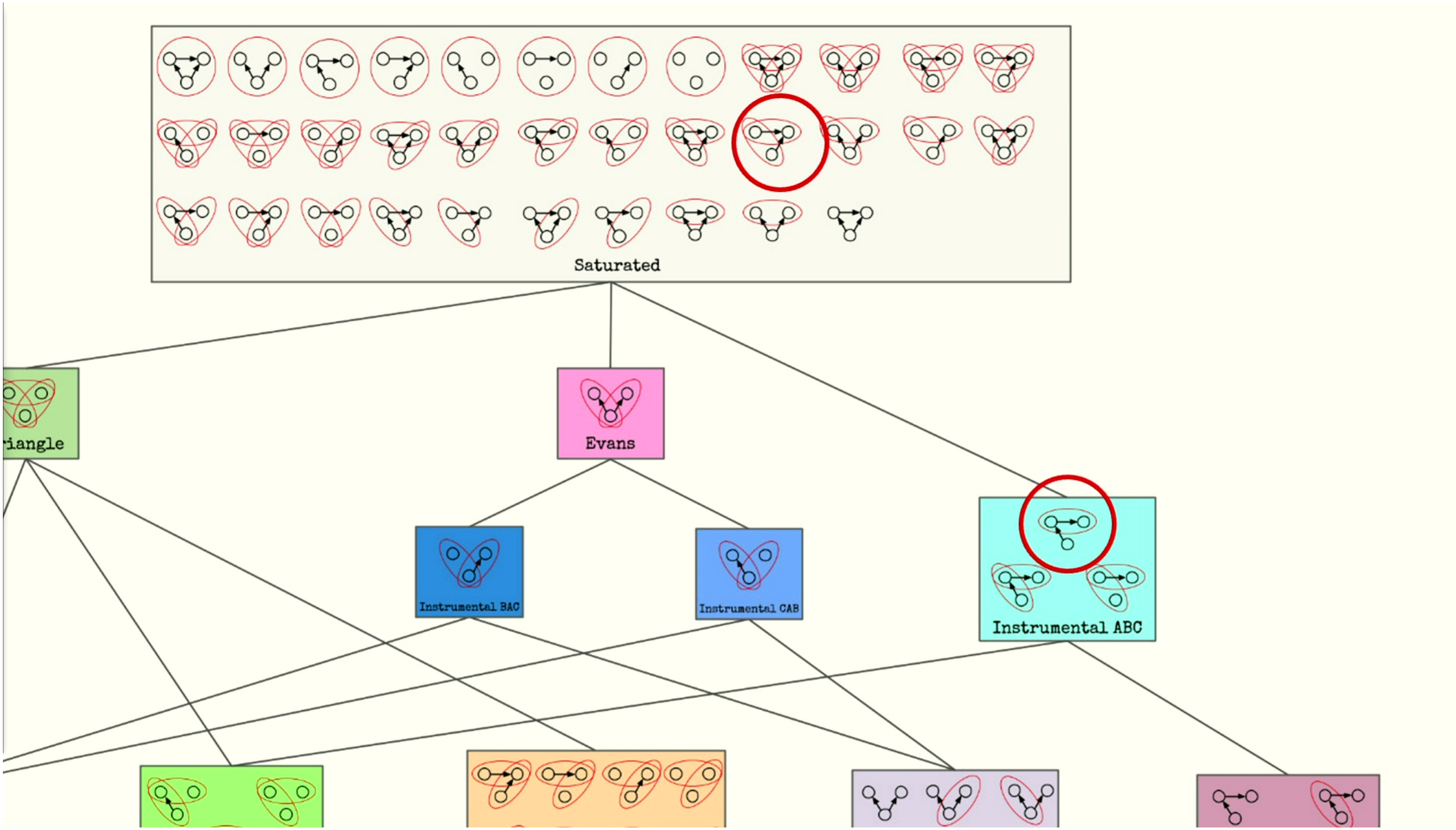
$$P_{F^{X=0}_{Z\to Y}}(\mathbb{I}) \geq \max\{0, P_{XY|Z}(00|0) + P_{XY|Z}(01|1) - 1\}$$
$$P_{F^{X=0}_{Z\to Y}}(\mathbb{F}) \geq \max\{0, P_{XY|Z}(01|0) + P_{XY|Z}(00|1) - 1\}$$
$$P_{F^{X=1}_{Z\to Y}}(\mathbb{I}) \geq \max\{0, P_{XY|Z}(10|0) + P_{XY|Z}(11|1) - 1\}$$
$$P_{F^{X=1}_{Z\to Y}}(\mathbb{F}) \geq \max\{0, P_{XY|Z}(11|0) + P_{XY|Z}(10|1) - 1\}$$

The hypotheses



$$P_{XY|Z}(00|0) + P_{XY|Z}(01|1) \leq 1$$
$$P_{XY|Z}(01|0) + P_{XY|Z}(00|1) \leq 1$$
$$P_{XY|Z}(10|0) + P_{XY|Z}(11|1) \leq 1$$
$$P_{XY|Z}(11|0) + P_{XY|Z}(10|1) \leq 1$$



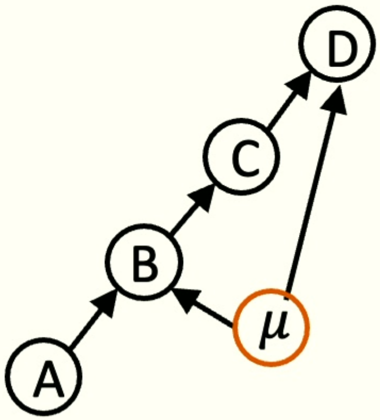

Saturated
iangle
Evans
Instrumental BAC
Instrumental CAB
Instrumental ABC

Note: Characterizing the set of compatible distributions for each mDAG is much harder than determining the dominance relations among mDAGs

But this is what is required for both the causal discovery problem and the problem of estimating causal effects

This problem is treated in subsequent lectures

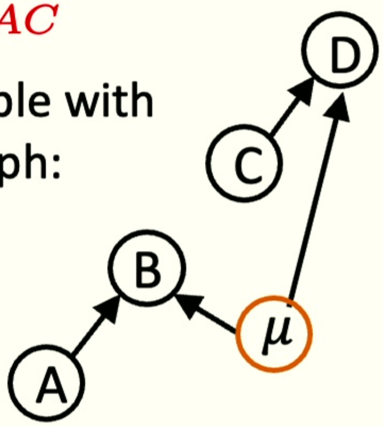
The complete set of equality constraints for latent-permitting causal models: Nested Markov constraints



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

$Q_{BD|AC}$

Is compatible with the subgraph:

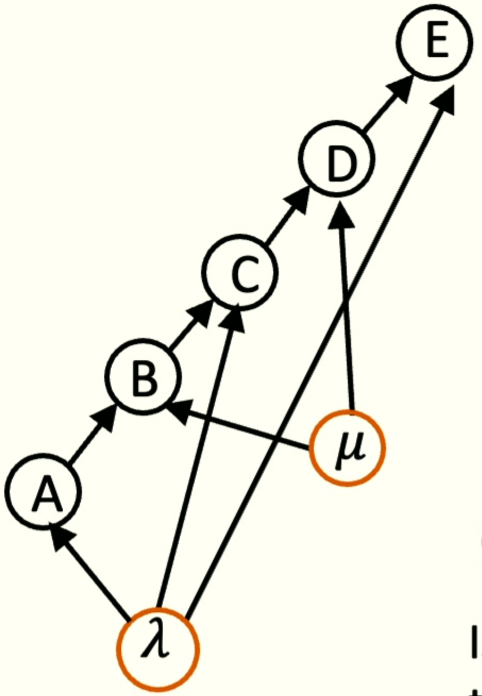


This subgraph has d-separation relations implying

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

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

This implies equality constraints on $Q_{BD|AC}$ and hence equality constraints on $P_{ABCD}/P_{C|B}P_A$

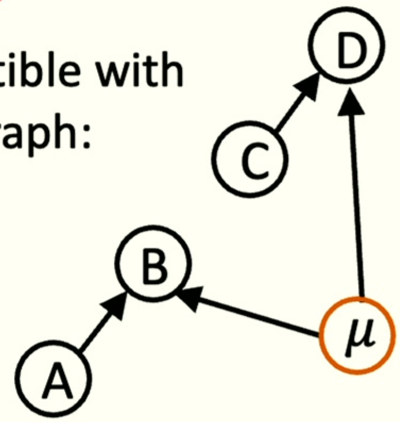


$$P_{ABCDE} = \left(\sum_{\mu} \mathrm{P}_{D|\mu C}\mathrm{P}_{B|A\mu}\mathrm{P}_{\mu}\right) \times \left(\sum_{\lambda} \mathrm{P}_{E|\lambda D}\mathrm{P}_{C|B\lambda}\mathrm{P}_{A|\lambda}\mathrm{P}_{\lambda}\right)$$

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

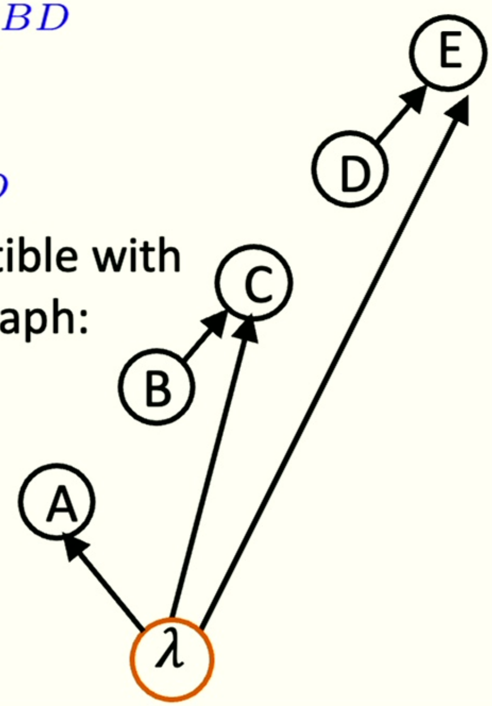
$\mathrm{Q}_{BD|AC}$

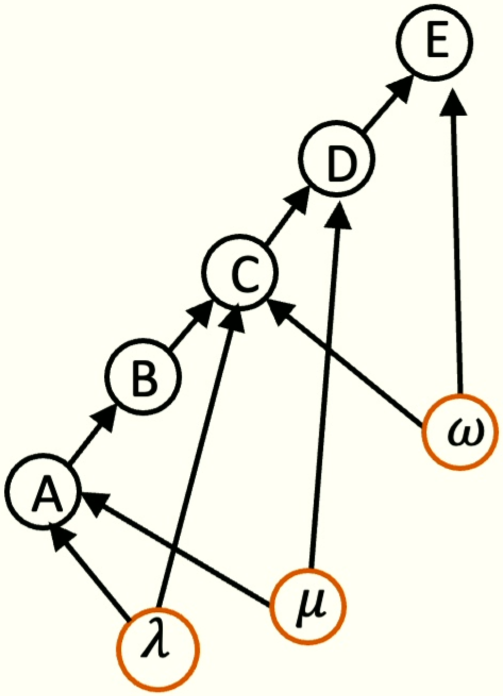
Is compatible with the subgraph:



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

Is compatible with the subgraph:

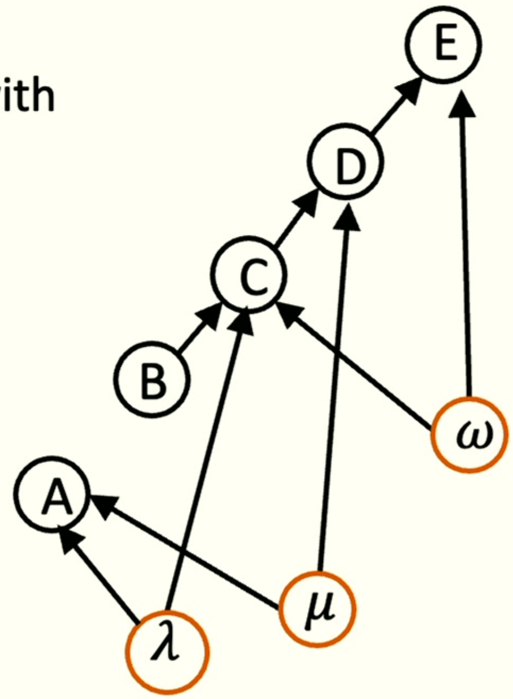




$$P_{ABCDE} = \left(\sum_{\omega,\mu,\lambda} \mathrm{P}_{E|\omega D} \mathrm{P}_{D|\mu C} \mathrm{P}_{C|\omega\lambda B} \mathrm{P}_{A|\lambda\mu} \mathrm{P}_{\omega} \mathrm{P}_{\mu} \mathrm{P}_{\lambda} \right) \mathrm{P}_{B|A}$$
$$= \mathrm{Q}_{ACDE|B} \mathrm{P}_{B|A}$$

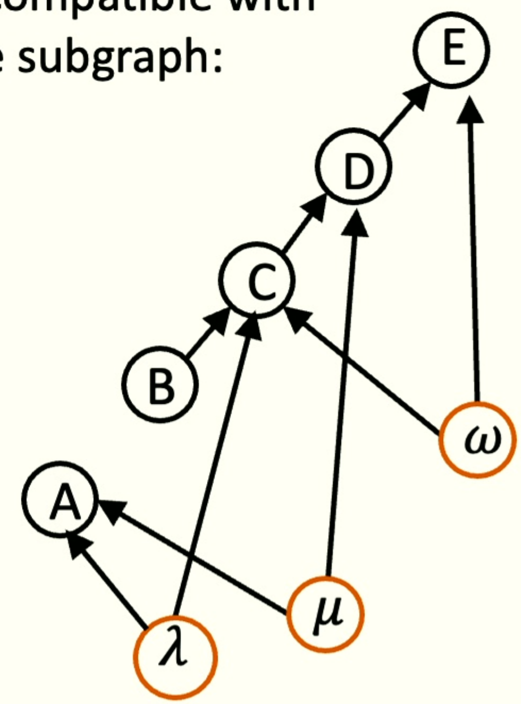
$\mathrm{Q}_{ACDE|B}$

Is compatible with the subgraph:



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

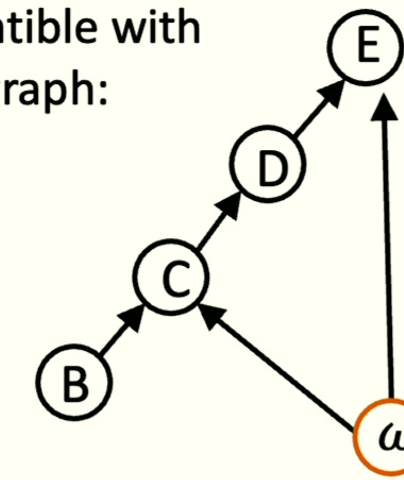
Is compatible with the subgraph:



A is childless, so we can marginalize without inducing any new correlations

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

Is compatible with the subgraph:



Verma graph

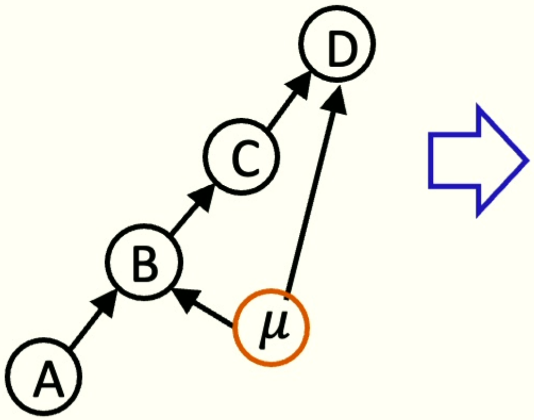
$$\mathrm{Q}_{CDE|B} = \mathrm{R}_{CE|BD}\mathrm{Q}_{D|C}\mathrm{Q}_{B}$$

Verma equality constraints on $\mathrm{R}_{CE|BD}$ imply equality constraints on $\mathrm{Q}_{CDE|B}/\mathrm{Q}_{D|D}\mathrm{Q}_{B}$

hence on $\mathrm{Q}_{ACDE|B}$ and hence on $\mathrm{P}_{ABCDE}/\mathrm{P}_{B|A}$

Such equality constraints are called nested Markov constraints

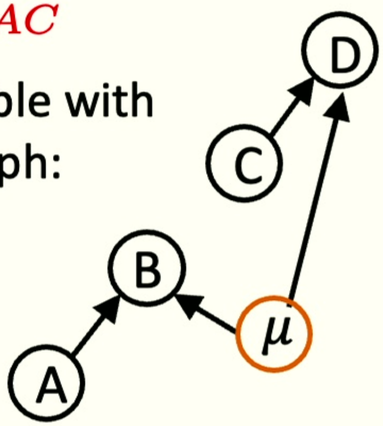
They include the equality constraints implied by d-separation and constitute *all* the equality constraints of a latent-permitting causal model



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

$Q_{BD|AC}$

Is compatible with the subgraph:



This subgraph has d-separation relations implying

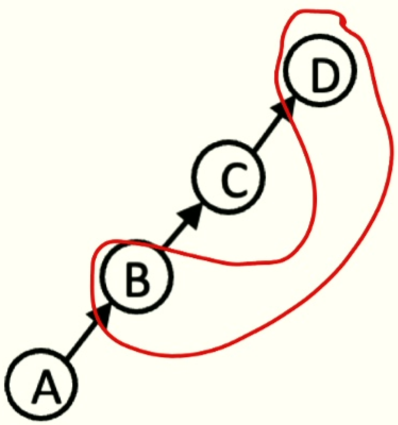
$D \perp A|C$ or equivalently, $Q_{D|AC} = Q_{D|C}$

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

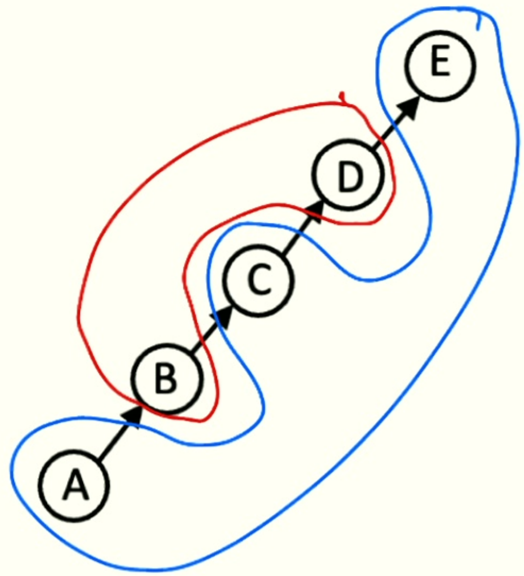
This implies equality constraints on $Q_{BD|AC}$ and hence equality constraints on $P_{ABCD}/P_{C|B}P_A$

Verma constraints

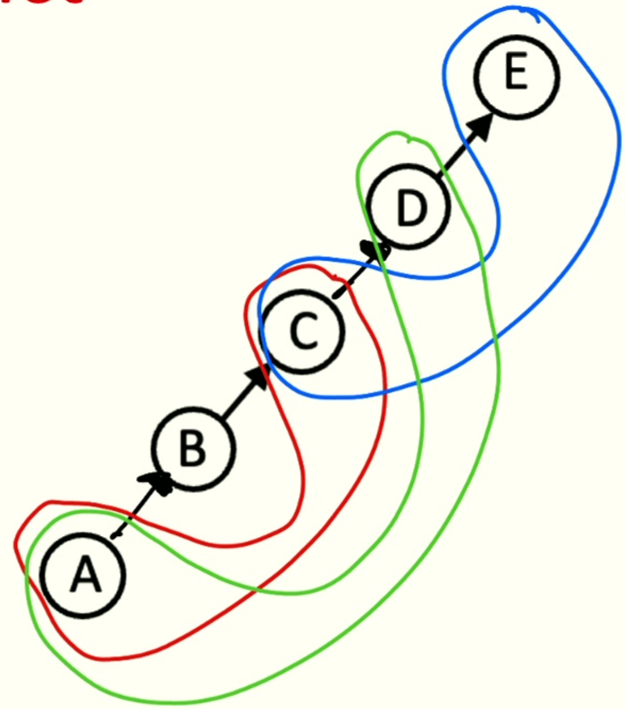
Definition of a district



3 districts:
BD, A, C

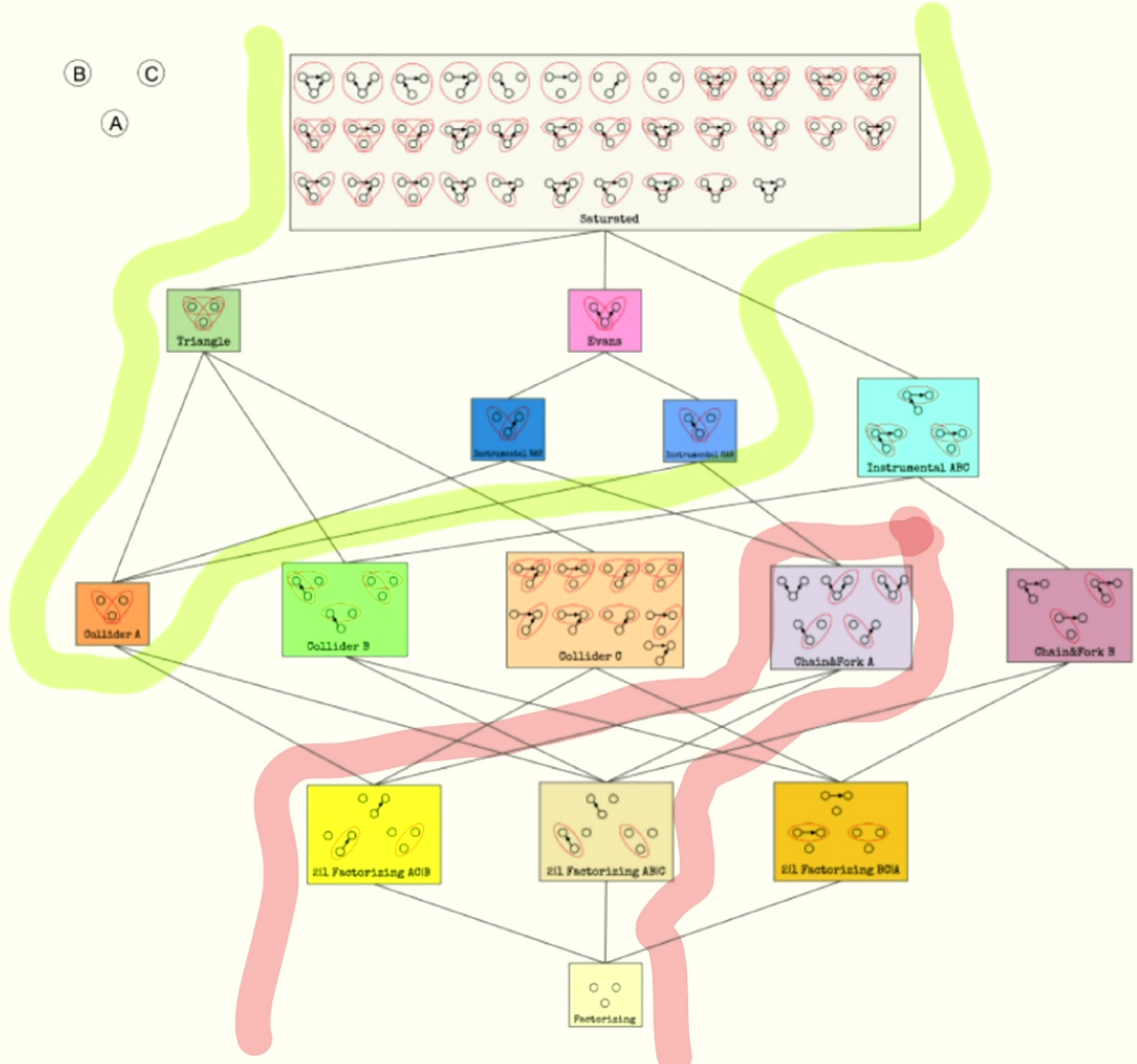


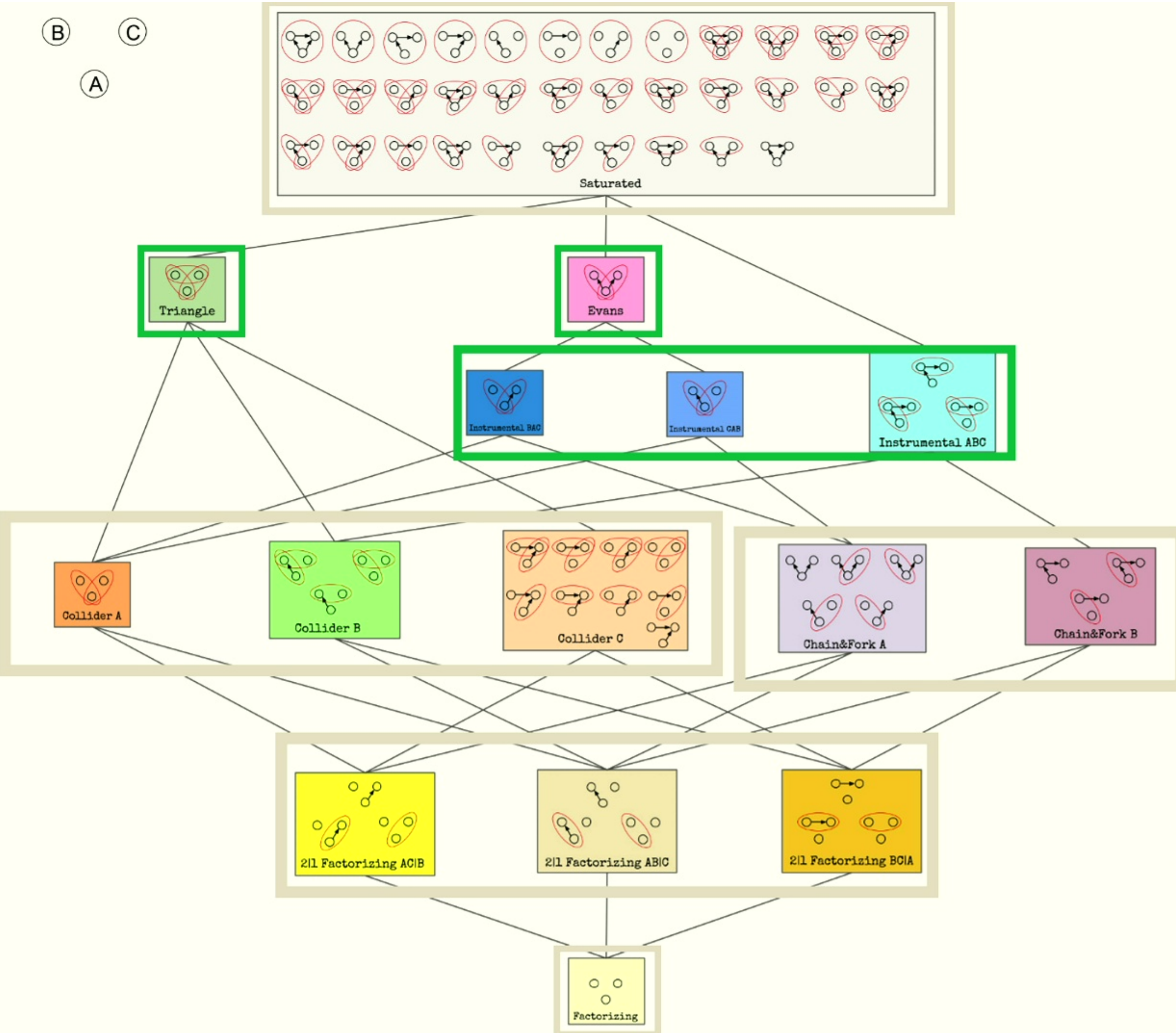
2 districts:
BD, ACE



2 districts:
ACDE, B

Note that district Factorization can also be used to simplify the characterization of the observational dominance order


B
C
A
Saturated
Triangle
Evans
Instrumental ABC
Collider A
Collider B
Collider C
Chain&Fork A
Chain&Fork B
2|1 Factorizing AC|B
2|1 Factorizing AB|C
2|1 Factorizing BC|A
Factorizing



Fraction of nonalgebraic classes = 5/15 (33%)