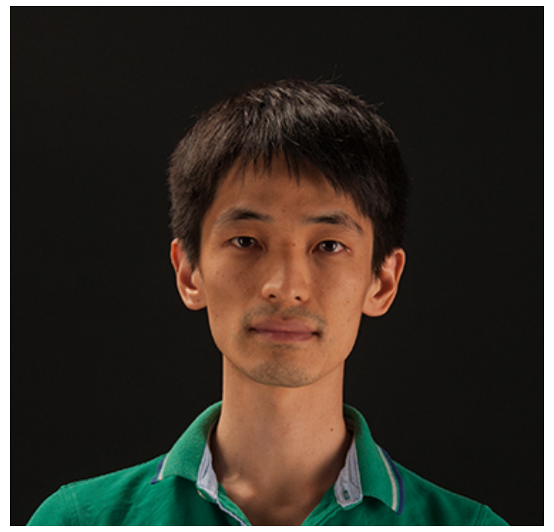
Testing Causal Models with Hidden Variables in Polynomial Delay via Conditional Independencies




Hyunchai Jeong*[1] , Adiba Ejaz*[2], Jin Tian[3] and Elias Bareinboim[2]

[1]Purdue University, [2]Columbia University, [3]Mohamed bin Zayed University of Artificial Intelligence

Motivation

"To find out what happens when you interfere with a system, you have to interfere with it (not just passively observe it)"

George Box, *Use and Abuse of Regression* (1996)

"Not exactly."

Wright, Neyman, Rubin, Pearl, etc.

Insight: can make causal inferences from observational studies using causal models.

Problem: how do we validate (or falsify) these causal models?

References

Paper

H. Jeong, A. Ejaz, J. Tian, E. Bareinboim. *Testing Causal Models with Hidden Variables in Polynomial Delay via Conditional Independencies.* AAAI 2025. https://causalai.net/r117.pdf

Code

https://github.com/CausalAILab/ListConditionalIndependencies

Outline

I. **Testable implications of causal models with hidden variables**

II. **The c-component local Markov property:** $O(4^n) \rightarrow O(n2^s)$ CI tests

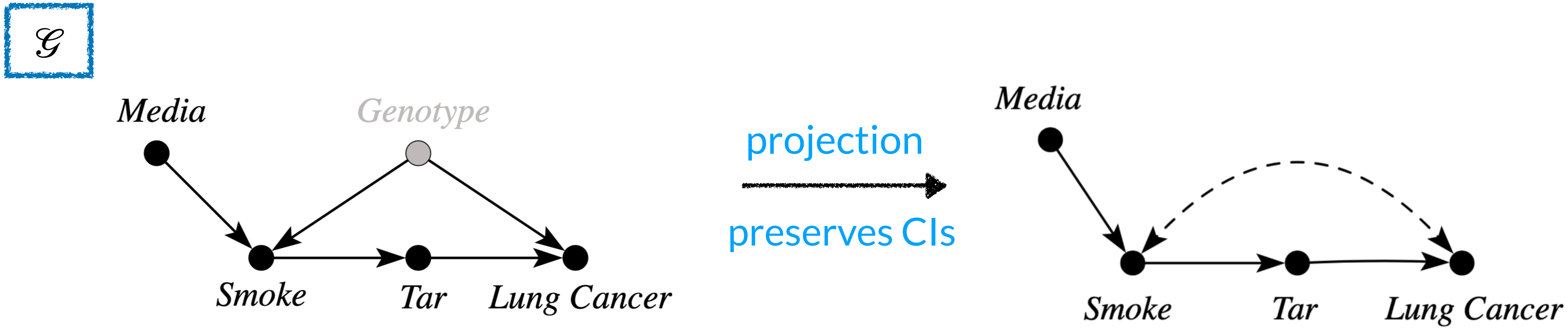
III. **The algorithm ListCI:** $\exp(n) \rightarrow \text{poly}(n)$ enumeration of CI tests

IV. **Empirical performance:** orders of magnitude speed-up

V. **Conclusions and future work**

Semi-Markovian Causal Graphs

- Causal graphs encode causal assumptions needed for inference (Pearl 1995, Bareinboim et al. 2020).



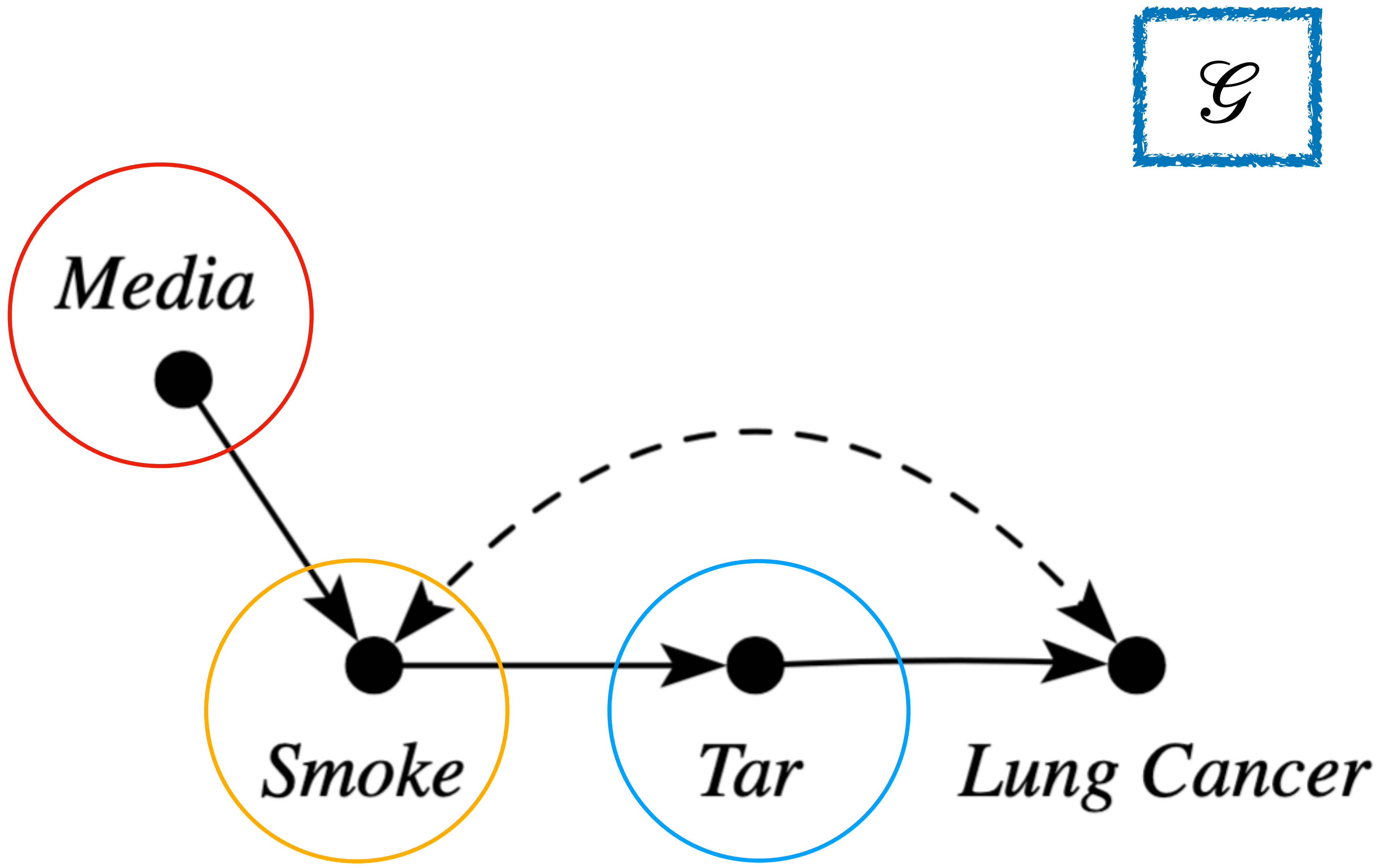
Bareinboim, E.; Correa, J. D.; Ibeling, D.; and Icard, T. 2022. On Pearl's Hierarchy and the Foundations of Causal Infer ence. In Probabilistic and Causal Inference: The Works of Judea Pearl, 507–556. New York, NY, USA: Association for Computing Machinery, 1st edition

Pearl, J. 1995. Causal diagrams for empirical research. Biometrika, 82(4): 669–688.

Causal Model Selection

How do we evaluate a causal graph using data?

Test if conditional independencies (CIs) implied by the model hold in the observational data!



$Media \perp Tar \mid Smoke.$

Problem Statement

Problem: Given a semi-Markovian graph $\mathcal{G}$ and an arbitrary observational distribution $P(\mathbf{v})$, do all CIs implied by $\mathcal{G}$ hold in $P(\mathbf{v})$?

Useful for: causal model selection, causal discovery algorithms, e.g. Sparsest Permutations, GES.

Which CIs do we need to test?

A basic approach

Definition 1 (GMP (Pearl 1988, Lauritzen et al. 1990))

Given a causal graph $\mathscr{G}$ over variables $\mathbf{V}$, a probability distribution $P(\mathbf{v})$ satisfies *the global Markov property* (LMP) for $\mathscr{G}$ if, for any disjoint $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathbf{V}$,

$$\mathbf{X} \perp_{\mathbf{d}} \mathbf{Y} \mid \mathbf{Z} \text{ in } \mathscr{G} \implies \mathbf{X} \perp_{\mu} \mathbf{Y} \mid \mathbf{Z} \text{ in } P(\mathbf{v})$$

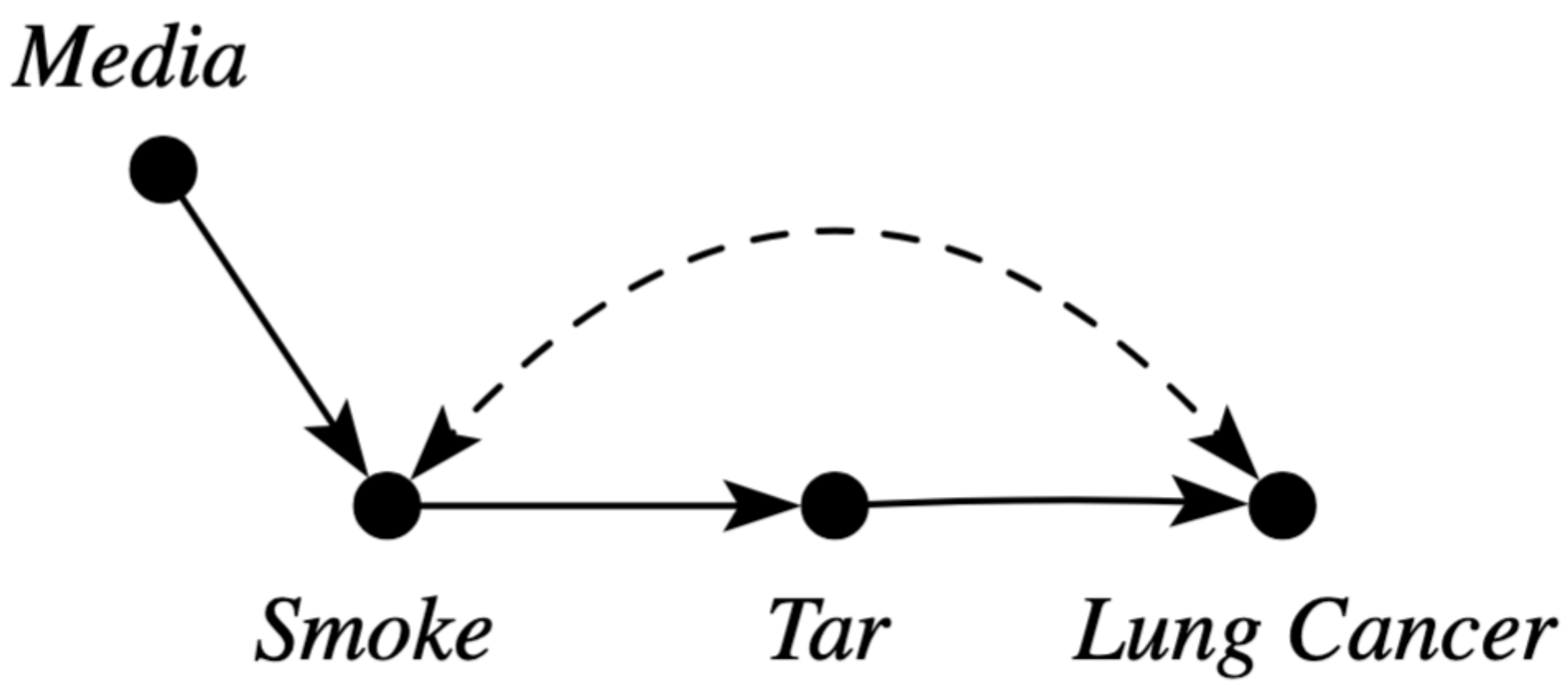
Problem: requires $\Theta(4^n)$ CI tests!

Lauritzen, S. L.; Dawid, A. P.; Larsen, B. N.; and Leimer, H. G. G. 1990. Independence Properties of Directed Markov Fields. Networks, 20(5): 491–505.
Pearl, J. 1988. Probabilistic Reasoning in Intelligent Systems. San Mateo, CA: Morgan Kaufmann

Global Markov Property

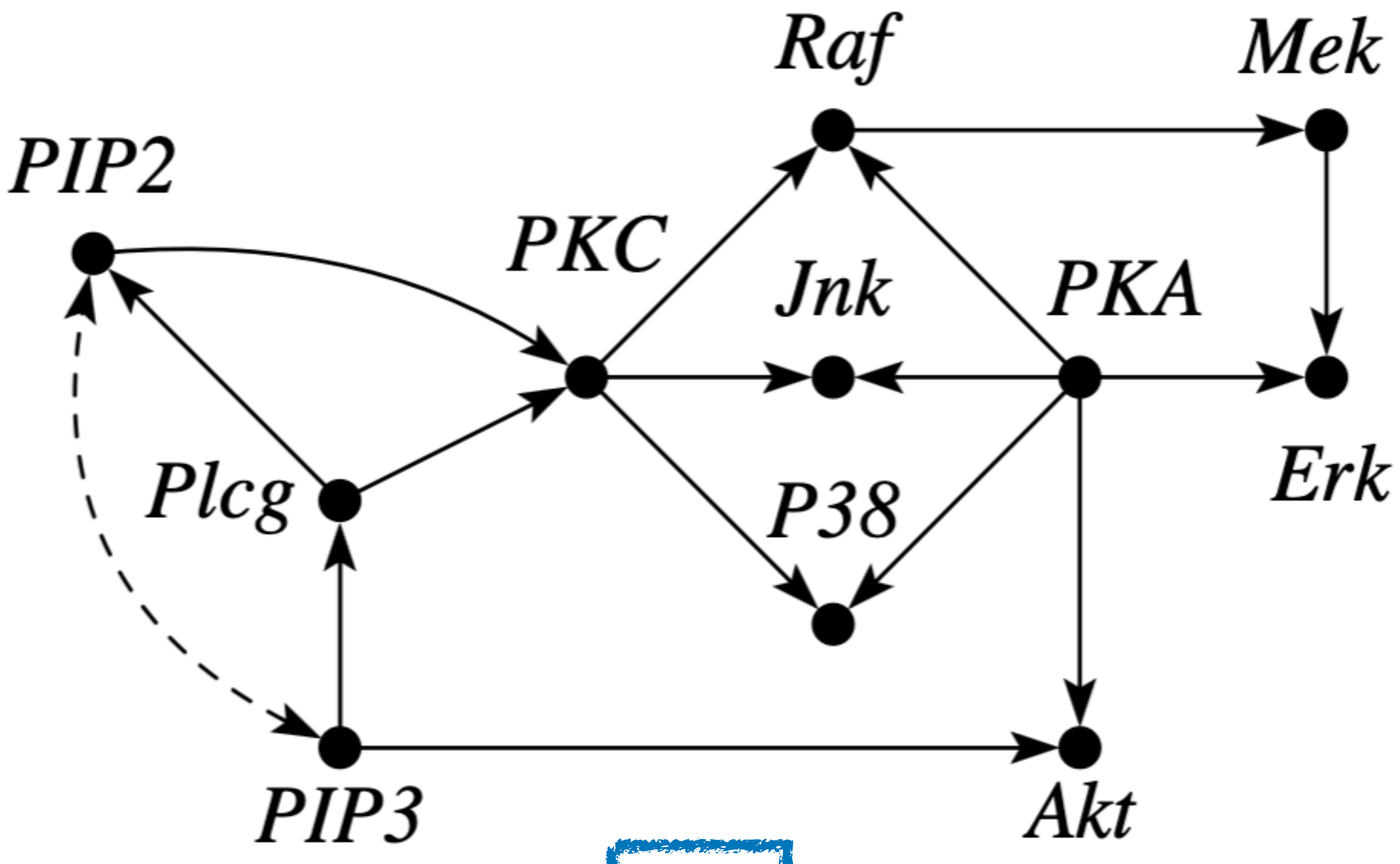
How many CIs do we need to test?

$n = 4, m = 4$. 1 CI encoded in $\mathscr{G}$



$\mathscr{G}$

$n = 11, m = 16$. 76,580 CIs encoded in $\mathscr{G}^S$



$\mathscr{G}^S$

Expert protein signalling network (Sachs et al. 2005)

Sachs, K.; Perez, O.; Pe'er, D.; Lauffenburger, D. A.; and Nolan, G. P. 2005. Causal protein-signaling networks derived from multiparameter single-cell data. Science, 308(5721): 523–529.

Testing Conditional Independencies

- Global Markov property (GMP): $\Theta(4^n)$ CIs.

Can we be more efficient?

- Simple observation: if we test $\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z}$, we don't need to test $\mathbf{Y} \perp \mathbf{X} \mid \mathbf{Z}$

Semi-Graphoid Axioms

- Conditional independence is a semi-graphoid *for any probability distribution* (Pearl and Paz 1985, Pearl 1988):
  - **Symmetry**

$$X \perp Y \mid Z \implies Y \perp X \mid Z$$

  - **Decomposition**

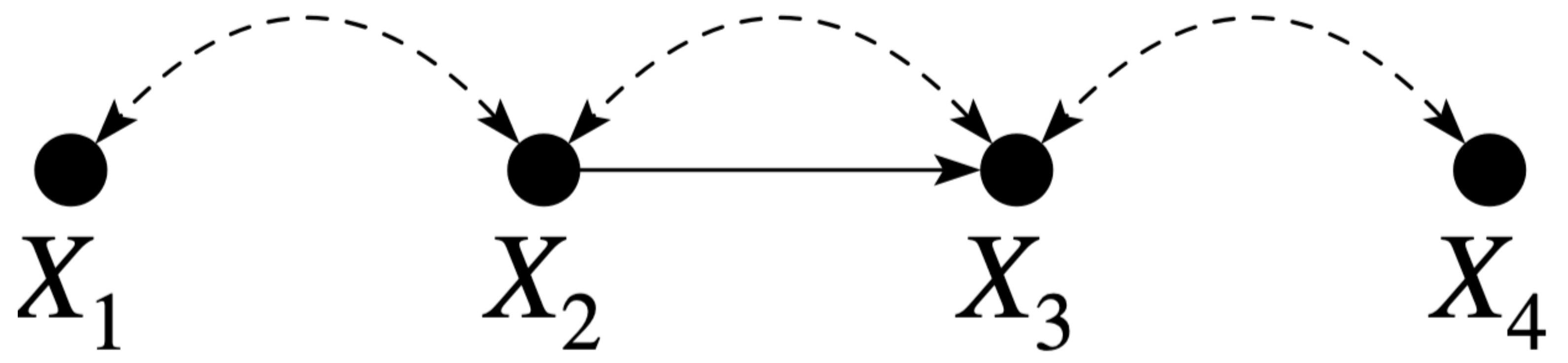
$$X \perp Y \cup W \mid Z \implies X \perp Y \mid Z \text{ and } X \perp W \mid Z$$

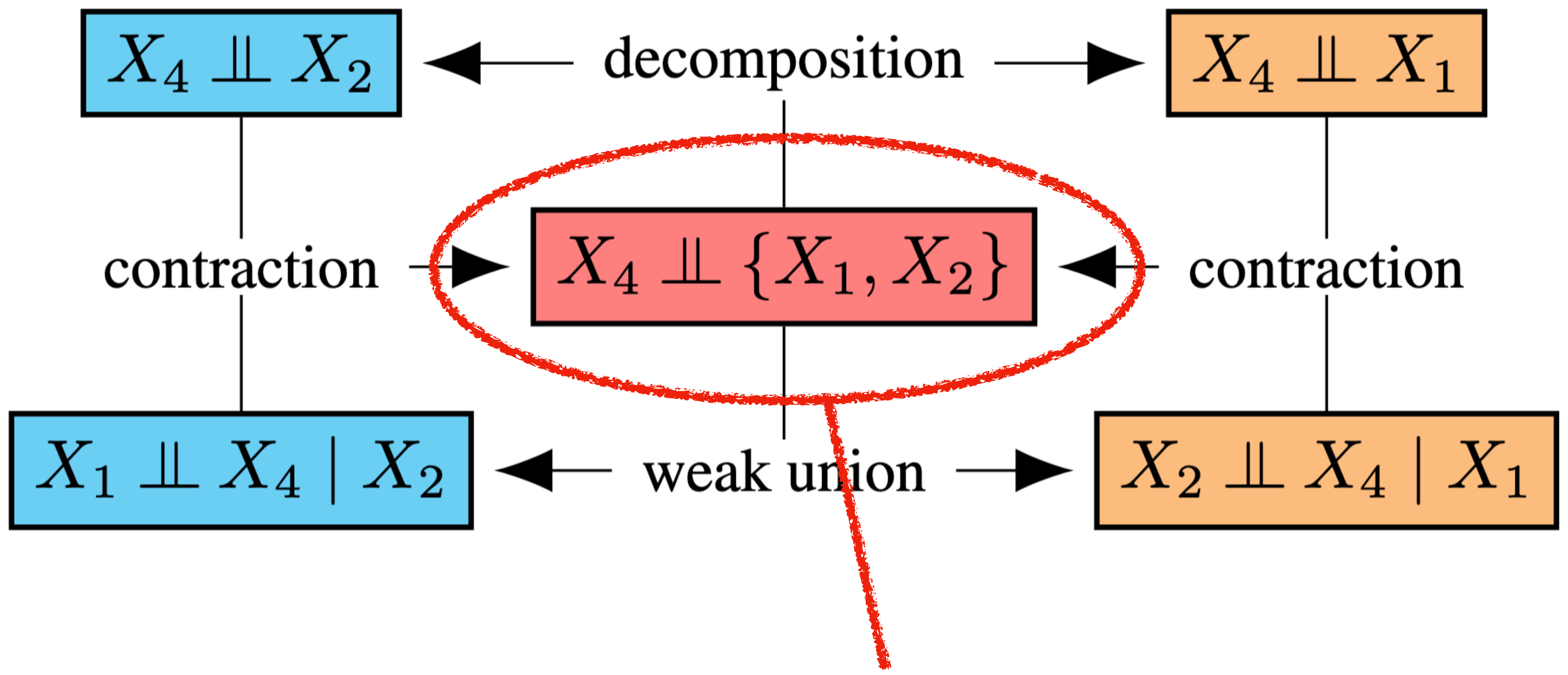
  - **Weak union**

$$X \perp Y \cup W \mid Z \implies X \perp Y \mid W \cup Z \text{ and } X \perp W \mid Y \cup Z$$

  - **Contraction**

$$X \perp Y \mid W \cup Z \text{ and } X \perp W \mid Z \implies X \perp Y \cup W \mid Z$$



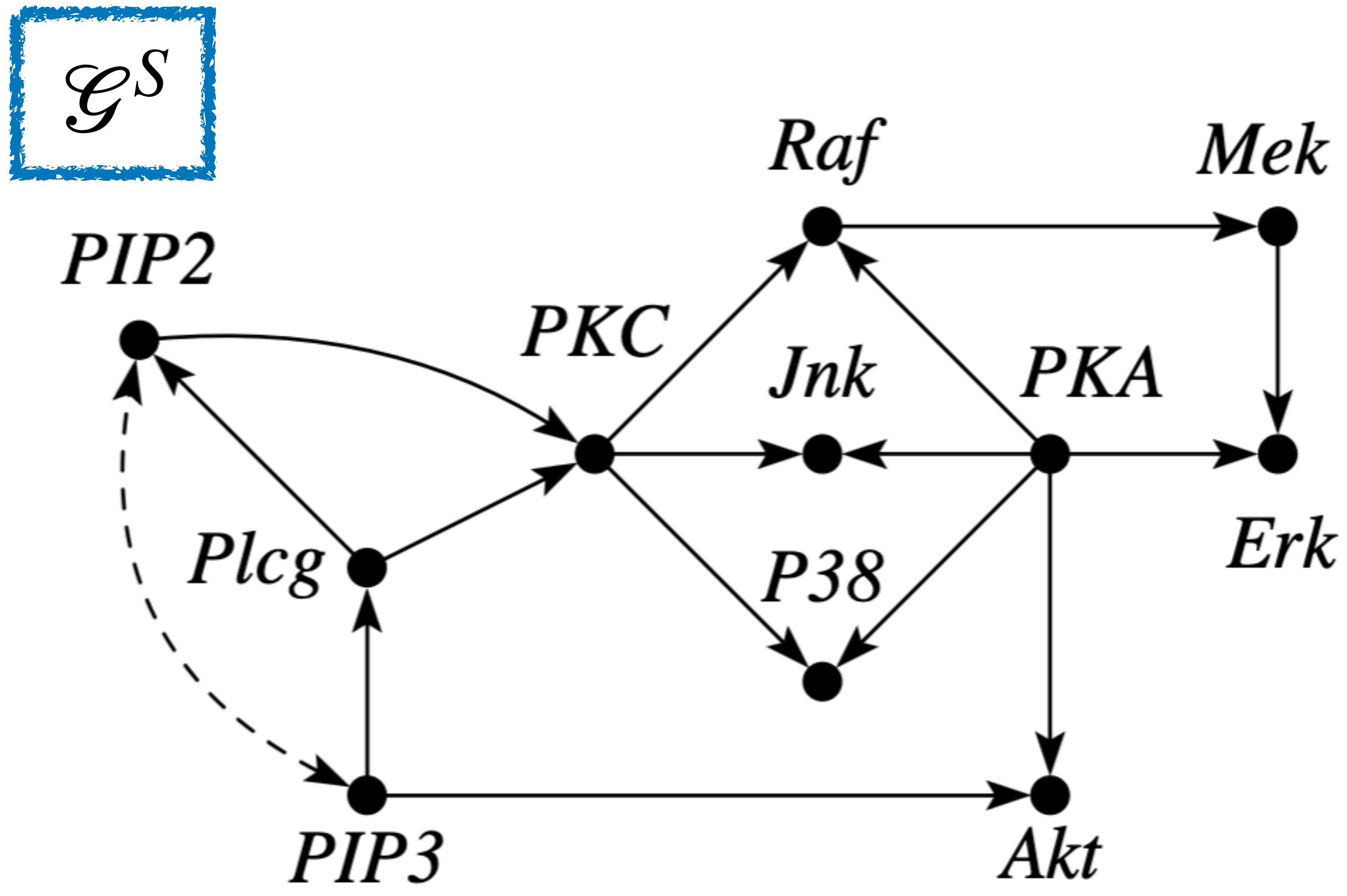


Test this one!

Pearl, J and Paz, A. 1986. Graphoids: graph-based logic for reasoning about relevance relations or when would x tell you more about y if you already know z? In Proceedings of the 7th European Conference on Artificial Intelligence - Volume 2 (ECAI'86). North-Holland, 357–363.
Pearl, J. 1988. Probabilistic Reasoning in Intelligent Systems. San Mateo, CA: Morgan Kaufmann

Local Markov Property

- Local Markov property (LMP): a 'basis' of CIs that implies all others
- Data consistent with GMP $\iff$ consistent with LMP
- But how do we find this basis systematically?



n = 11, m = 16. 76580 CIs encoded in $\mathcal{G}$!

10 CIs imply all other 76570 CIs.

Markov Factorization

Definition 1 (Markov factorization (Pearl 1988))

Given a causal graph $\mathcal{G}$ over variables $\mathbf{V}$, a probability distribution $P(\mathbf{v})$ is said to be *Markov* with respect to to $\mathcal{G}$ if

$$P(\mathbf{v}) = \prod_{V_i \in \mathbf{V}} P(v_i \mid pa_i)$$

Parsimonious encoding of the joint

$O(2^n)$ parameters $\rightarrow O(n2^k)$ parameters if $|pa_i| \leq k$

Pearl, J. 1988. Probabilistic Reasoning in Intelligent Systems. San Mateo, CA: Morgan Kaufmann

Local Markov Property

Definition 1 (LMP (Pearl 1988, Lauritzen et al. 1990))

Given a causal graph $\mathcal{G}$ over variables $\mathbf{V}$, a probability distribution $P(\mathbf{v})$ satisfies *the local Markov property* (LMP) for $\mathcal{G}$ if, for any $X \in \mathbf{V}$,

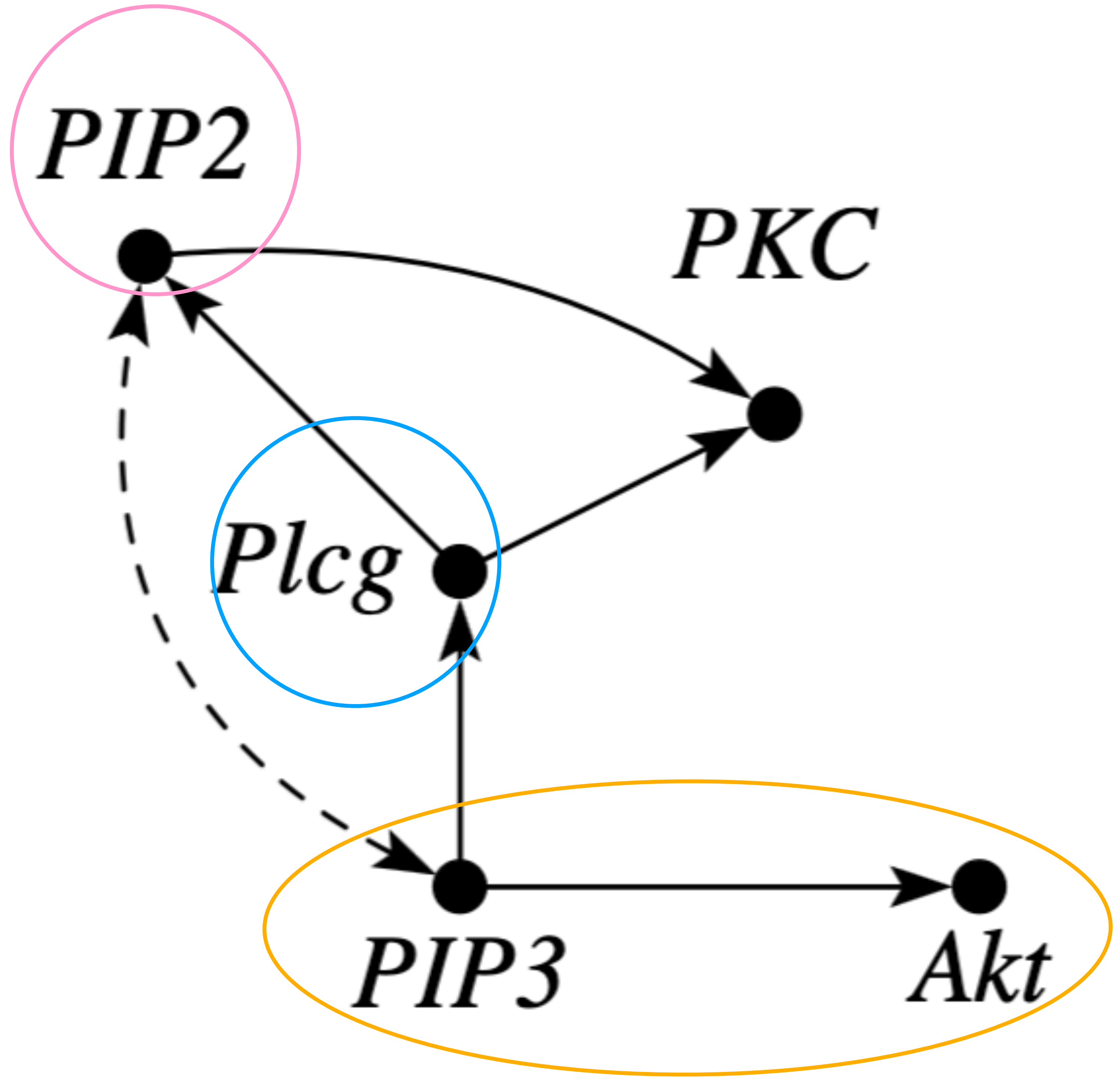
$$P(\mathbf{v}) = \prod_{V_i \in \mathbf{V}} P(v_i \mid pa_i)$$

$$X \perp_{\mu} nd_X \mid pa_X \text{ in } P(\mathbf{v}).$$

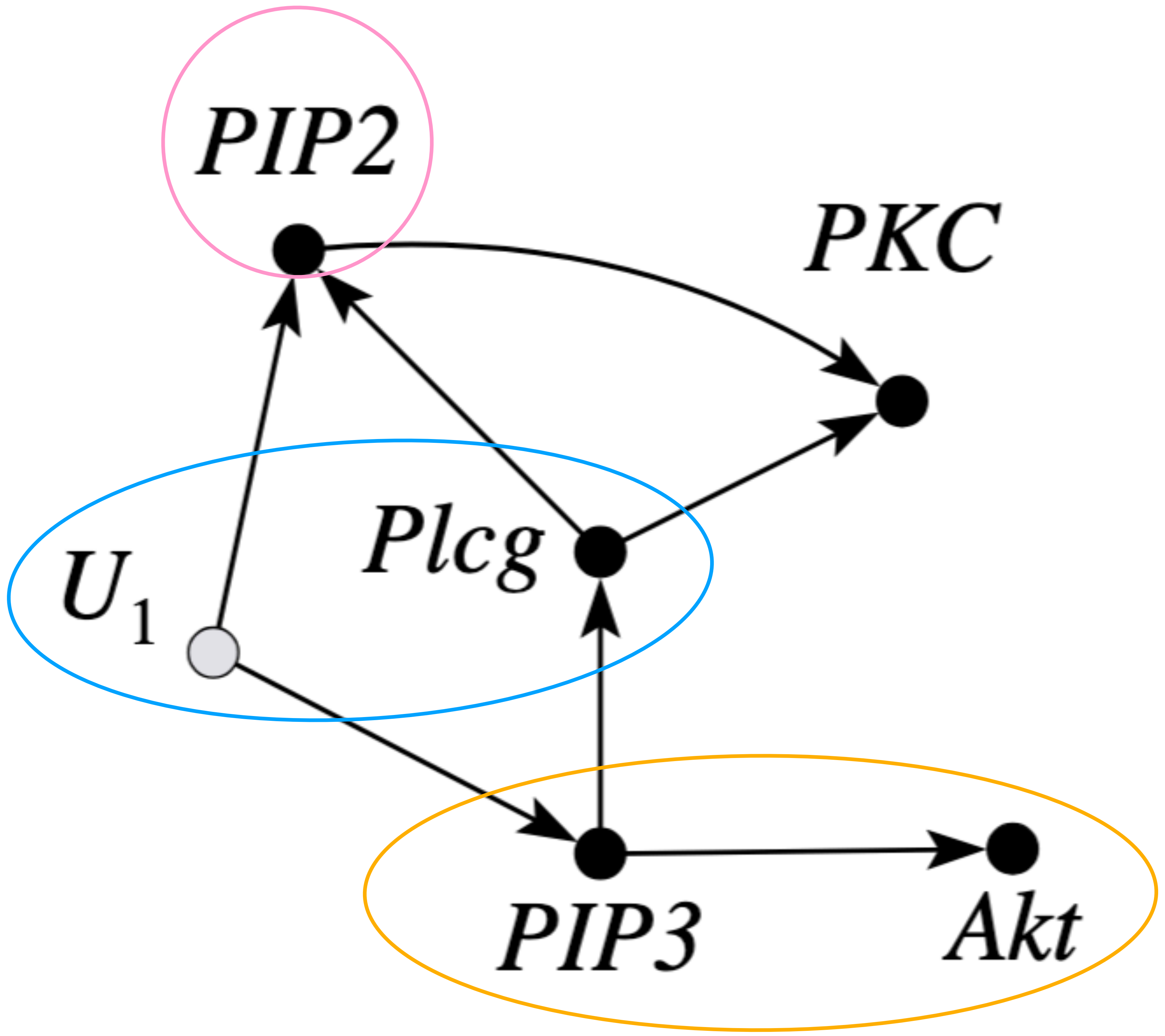
Solution requires $\Theta(n)$ CI tests?

Lauritzen, S. L.; Dawid, A. P.; Larsen, B. N.; and Leimer, H. G. G. 1990. Independence Properties of Directed Markov Fields. Networks, 20(5): 491–505.
Pearl, J. 1988. Probabilistic Reasoning in Intelligent Systems. San Mateo, CA: Morgan Kaufmann

Challenge: Hidden Variables



make unobserved variable explicit



$X \perp nd_X \mid pa_X$

$PIP2 \perp PIP3, Akt \mid Plcg$? **No**

$PIP2 \perp PIP3, Akt \mid Plcg, U_1$? Yes

Local Markov Property

Definition 1 (LMP (Pearl 1988, Lauritzen et al. 1990))

Given a causal graph $\mathcal{G}$ over variables $\mathbf{V}$, a probability distribution $P(\mathbf{v})$ satisfies *the local Markov property* (LMP) for $\mathcal{G}$ if, for any $X \in \mathbf{V}$,

$$X \perp nd_X \mid pa_X \text{ in } P(\mathbf{v}).$$

Assumes no hidden confounders!

Lauritzen, S. L.; Dawid, A. P.; Larsen, B. N.; and Leimer, H. G. G. 1990. Independence Properties of Directed Markov Fields. Networks, 20(5): 491–505.
Pearl, J. 1988. Probabilistic Reasoning in Intelligent Systems. San Mateo, CA: Morgan Kaufmann

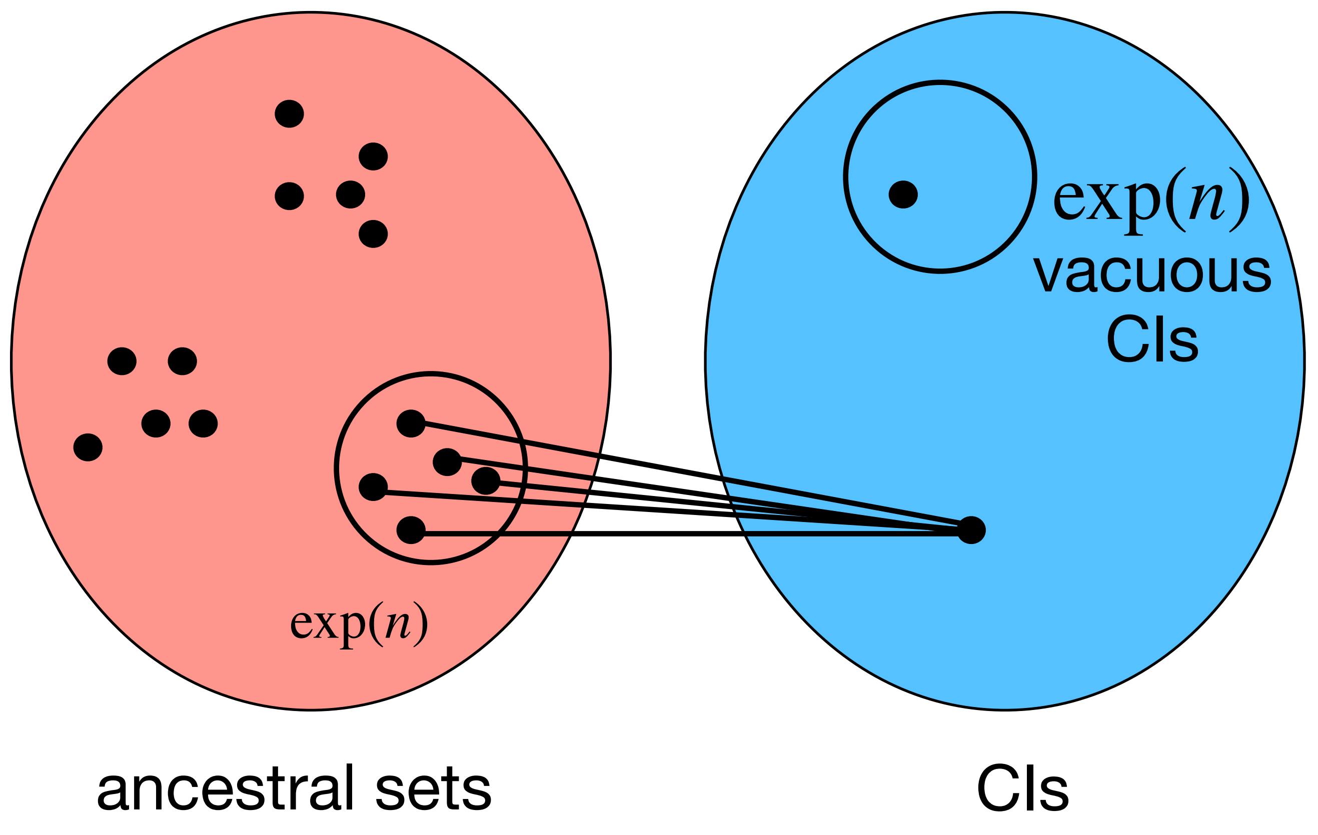
Questions

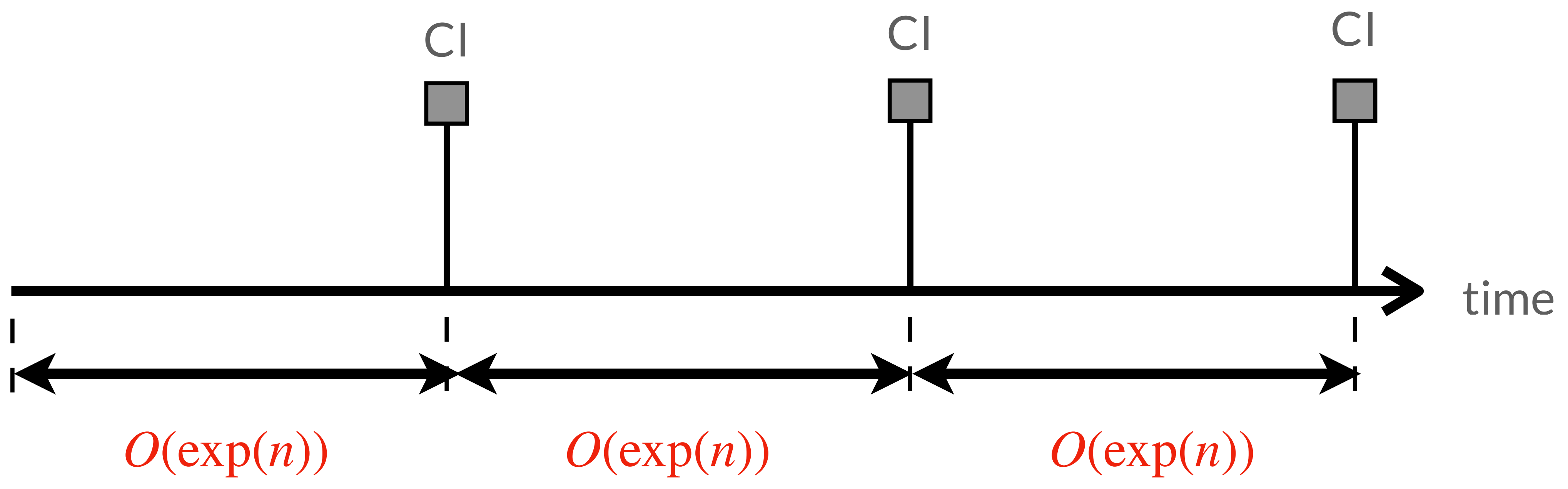
Q1. How to recover locality in semi-Markovian models?

Q2. How to enumerate the required CI tests?

	Markovian	Semi-Markovian
Global Markov Property	$O(4^n)$	$O(4^n)$
Local Markov Property	$O(n)$	?

Local Markov Properties exist (Richardson 2003), but naive enumeration is exponential delay!





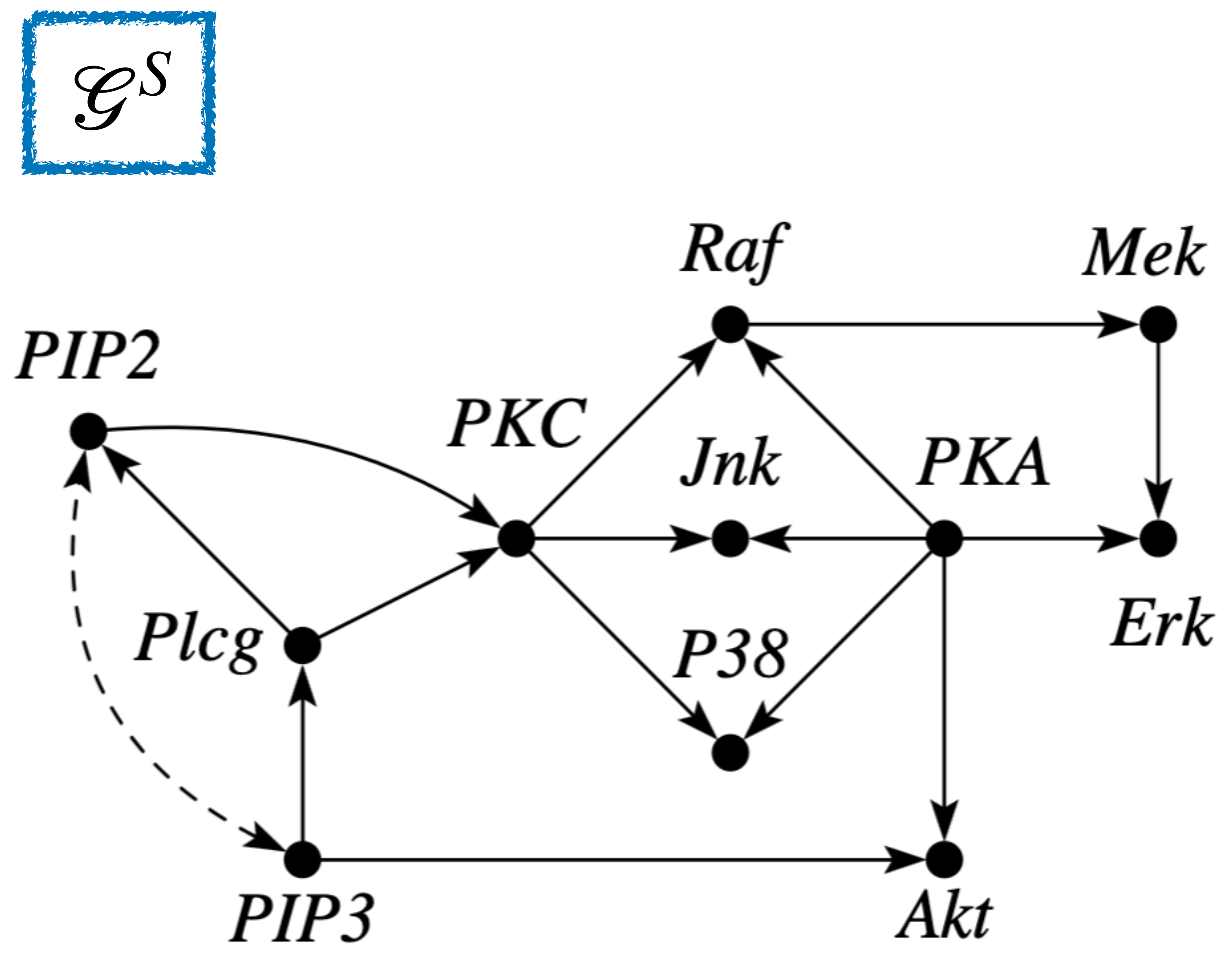
Richardson, T. 2003. Markov properties for acyclic directed mixed graphs. Scandinavian Journal of Statistics, 30(1): 145–157.

Result I - C-LMP

C-component local Markov property (C-LMP) for causal graphs with hidden variables.

- Equivalent to GMP for **arbitrary** probability distributions
- **Exponentially fewer CI tests** than GMP: $\mathcal{O}(n2^s)$ vs $\mathcal{O}(4^n)$ with $s \leq n$ (size of the largest c-component)

Takeaway: use C-LMP for model testing

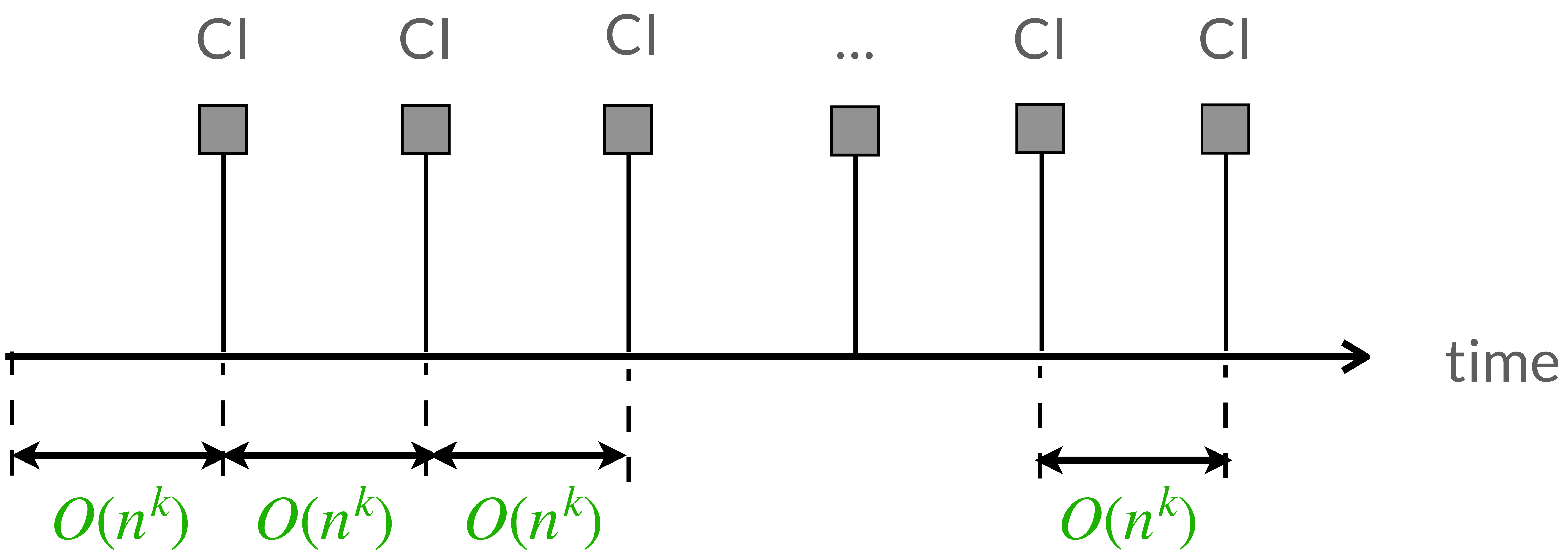


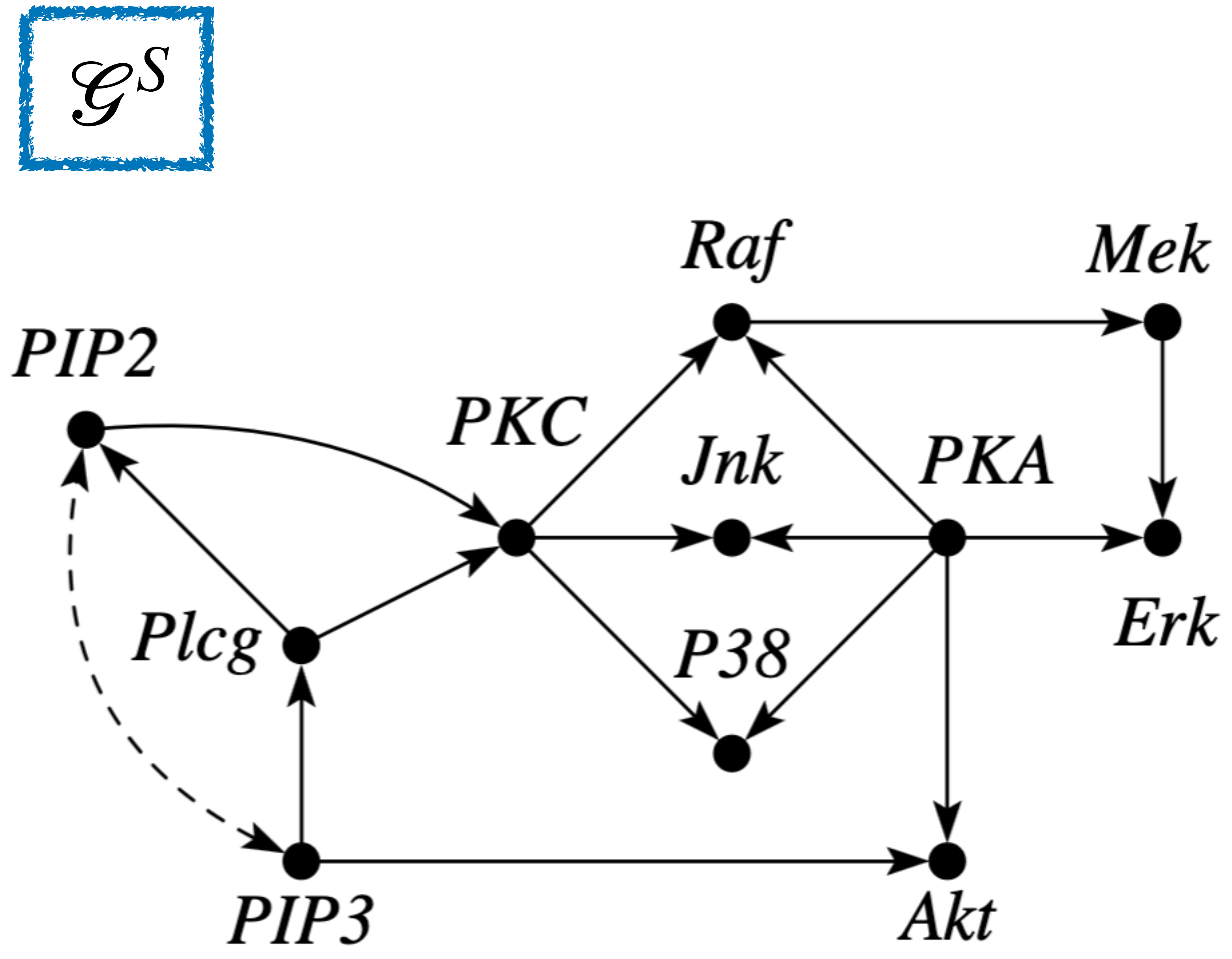
C-LMP: 10 CIs imply all other 76570 CIs

Result II - `ListCI`

`ListCI`, a polynomial delay (Johnson et al. 1988) algorithm for listing CIs invoked by C-LMP

- **First** efficient algorithm for local Markov properties in hidden variable settings





`ListCI` lists 10 CIs in poly-time intervals

Johnson, D. S.; Yannakakis, M.; and Papadimitriou, C. H. 1988. On generating all maximal independent sets. Information Processing Letters, 27(3): 119–123

Recovering Locality in Semi-Markovian Models

Why do semi-Markovian local Markov properties invoke exponentially many CIs?

Definition 1 (Semi-Markov factorization (Bareinboim et al 2020))

Given a causal graph $\mathcal{G}$ over variables $\mathbf{V}$, a probability distribution $P(\mathbf{v})$ is said to be *semi-Markov* with respect to to $\mathcal{G}$ if for every topological ordering $\prec$,

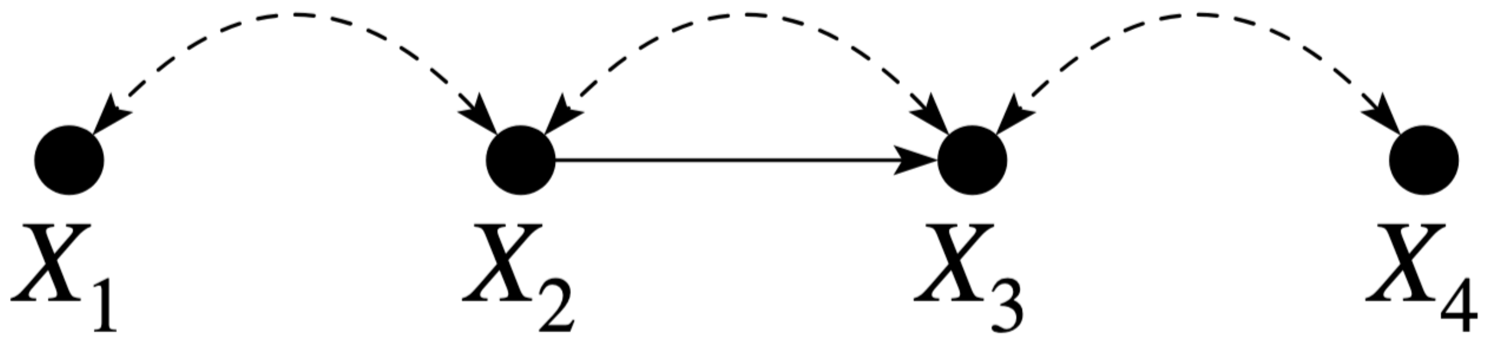
$$P(\mathbf{v}) = \prod_{V_i \in \mathbf{V}} P(v_i \mid pa^+_{\prec,i})$$

where $pa^+_{i,\prec} = Pa(\mathbf{C}(V_i)_{\mathcal{G}_{\mathbf{V} \leq X}}) \setminus \{V_i\}$.

For Markovian models, $pa^+_{i,\prec}$ is independent of $\prec$!

Recovering Locality in Semi-Markovian Models

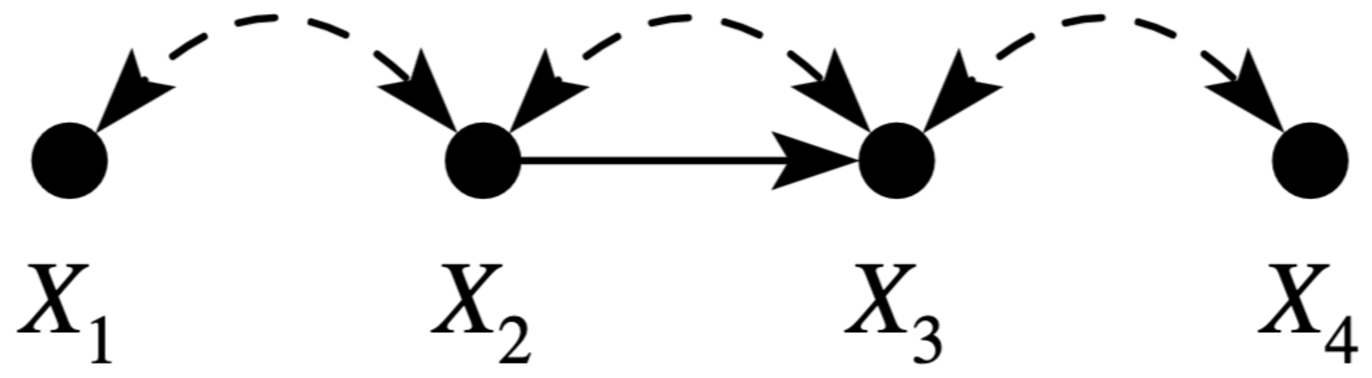
One ordering does not suffice.



$$P(\mathbf{v}) = \prod_{V_i \in \mathbf{V}} P(v_i \mid pa^{+}_{\prec,i})$$

$X_1 \prec X_2 \prec X_3 \prec X_4$

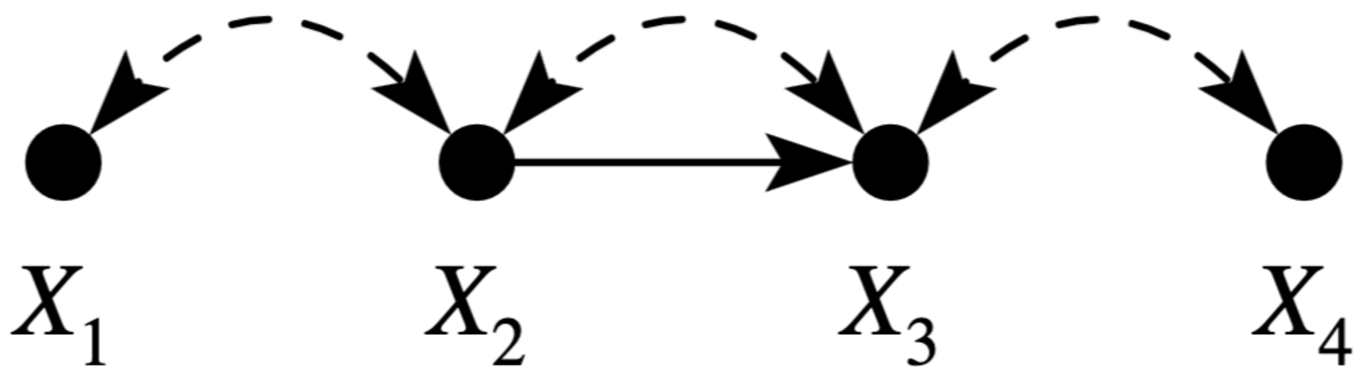
$$P(x_1, x_2, x_3, x_4) = P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_2, x_1)P(x_4 \mid x_3, x_2, x_2)$$



☹ chain rule, no CIs

$X_1 \prec X_4 \prec X_2 \prec X_3$

$$P(x_1, x_2, x_3, x_4) = P(x_1)P(x_4)P(x_2 \mid x_1)P(x_3 \mid x_2, x_1, x_4)$$

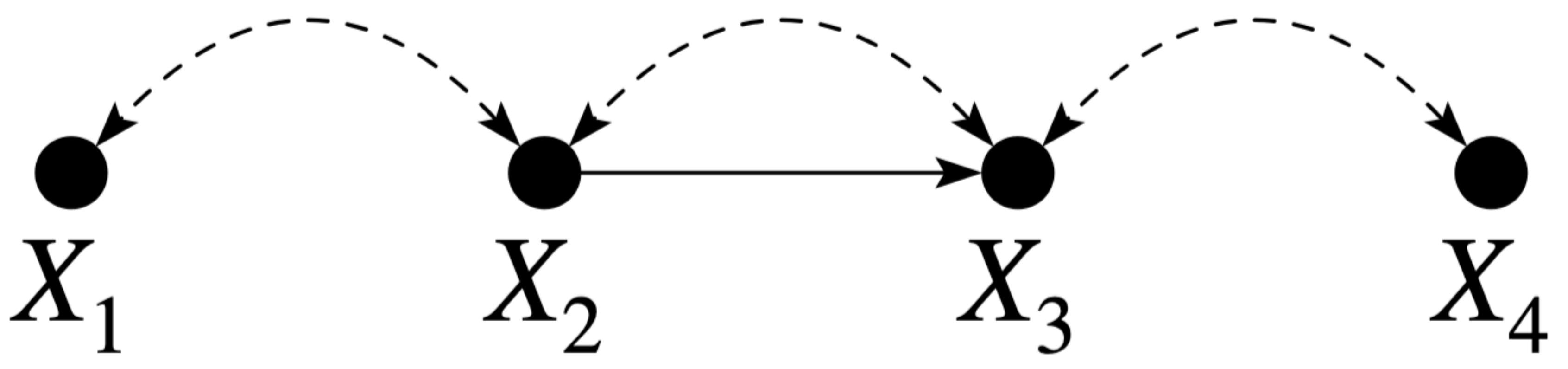


☺ $X_1 \perp X_4$

From Factorization to Markov Property

Semi-Markov factorization implies: for every topological ordering of $\mathscr{G}$ and every variable V_i: $\quad V_i \perp_\mu \mathbf{V}^{\leq X} \setminus pa^+_{i,\prec} \mid pa^+_{i,\prec}$.

Problem: this is not a basis!



$X_1 \prec X_2 \prec X_4 \prec X_3$: $\quad X_4 \perp X_1, X_2$ maximal

$\Downarrow$

$X_1 \prec X_4 \prec X_2 \prec X_3$: $\quad X_4 \perp X_1$ not maximal

C-component Local Markov Property

Definition 2 (C-LMP)

Given a causal graph $\mathcal{G}$ and a consistent ordering $\mathbf{V}^{\prec}$, a probability distribution $P(\mathbf{v})$ satisfies *the c-component local Markov property* (C-LMP) for $\mathcal{G}$ w.r.t $\mathbf{V}^{\prec}$ if, for any variable $X \in \mathbf{V}^{\prec}$ and ancestral c-component $\mathbf{C}$ relative to X,

$X \perp \mathbf{W} \mid \mathbf{Z}$ in $P(\mathbf{v})$ where

$$\mathbf{W} = \mathbf{V}^{\leq X} \setminus (De(Sp(\mathbf{C}) \setminus Pa(\mathbf{C})) \cup Pa(\mathbf{C}))$$ and

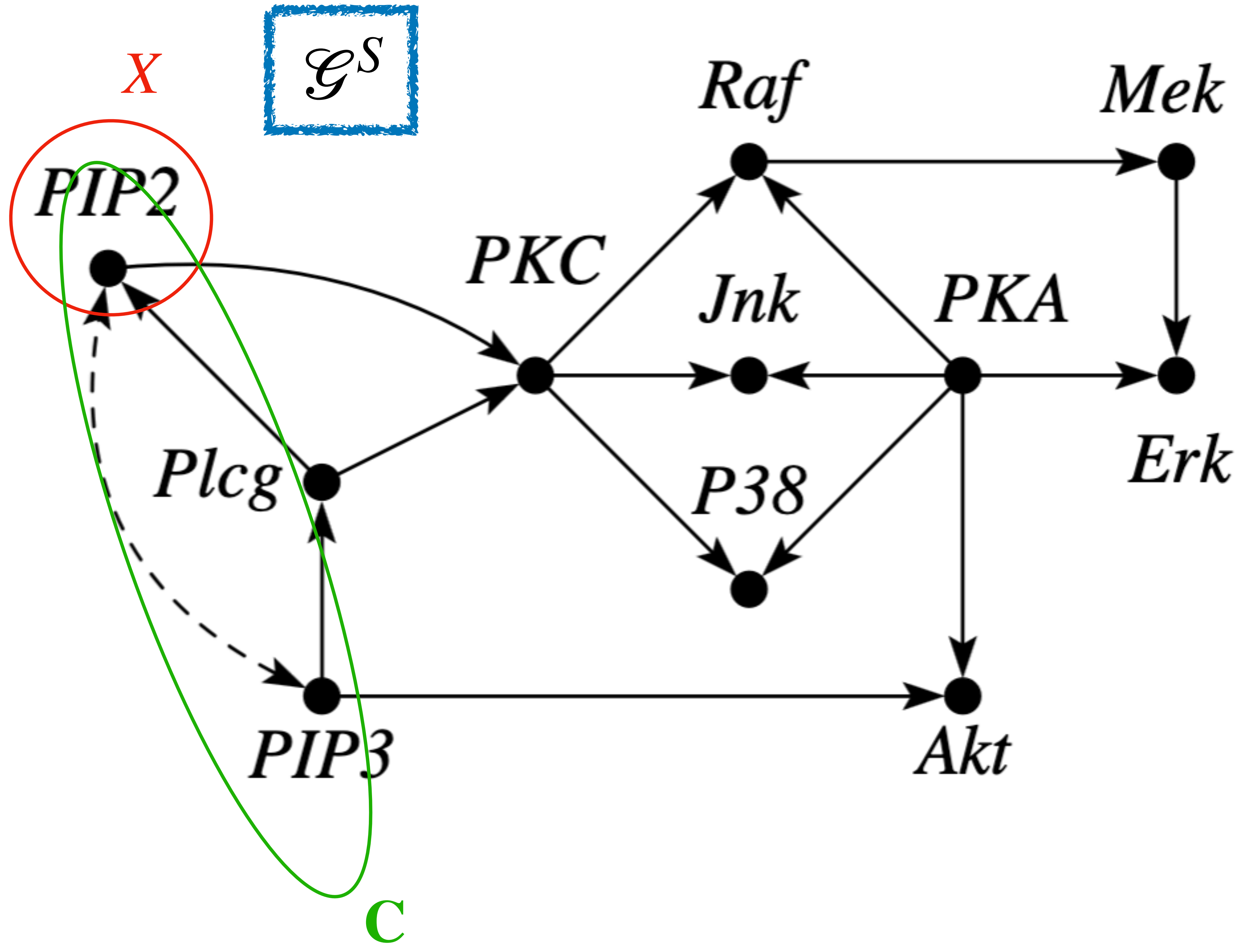
$$\mathbf{Z} = Pa(\mathbf{C}) \setminus \{X\}.$$

Compare with LMP: $X \perp nd_X \mid pa_X$

Example

- $X = PIP2$
- Ancestral c-component $\mathbf{C} = \{PIP2, PIP3\}$.
- CI: $X \perp \mathbf{W} \mid \mathbf{Z}$
- $\mathbf{W} = \mathbf{V}^{\leq X} \setminus (De(Sp(\mathbf{C}) \setminus Pa(\mathbf{C})) \cup Pa(\mathbf{C})) = \{Akt, PKA\}$.
- $\mathbf{Z} = Pa(\mathbf{C}) \setminus \{X\} = \{PIP3, Plcg\}$.

$$\therefore PIP2 \perp \{Akt, PKA\} \mid \{PIP3, Plcg\}$$



LMP: $PIP2 \perp \{PIP3, Akt, PKA\} \mid \{Plcg\}$? No

C-LMP: $PIP2 \perp \{Akt, PKA\} \mid \{PIP3, Plcg\}$? Yes

C-LMP $\iff$ GMP

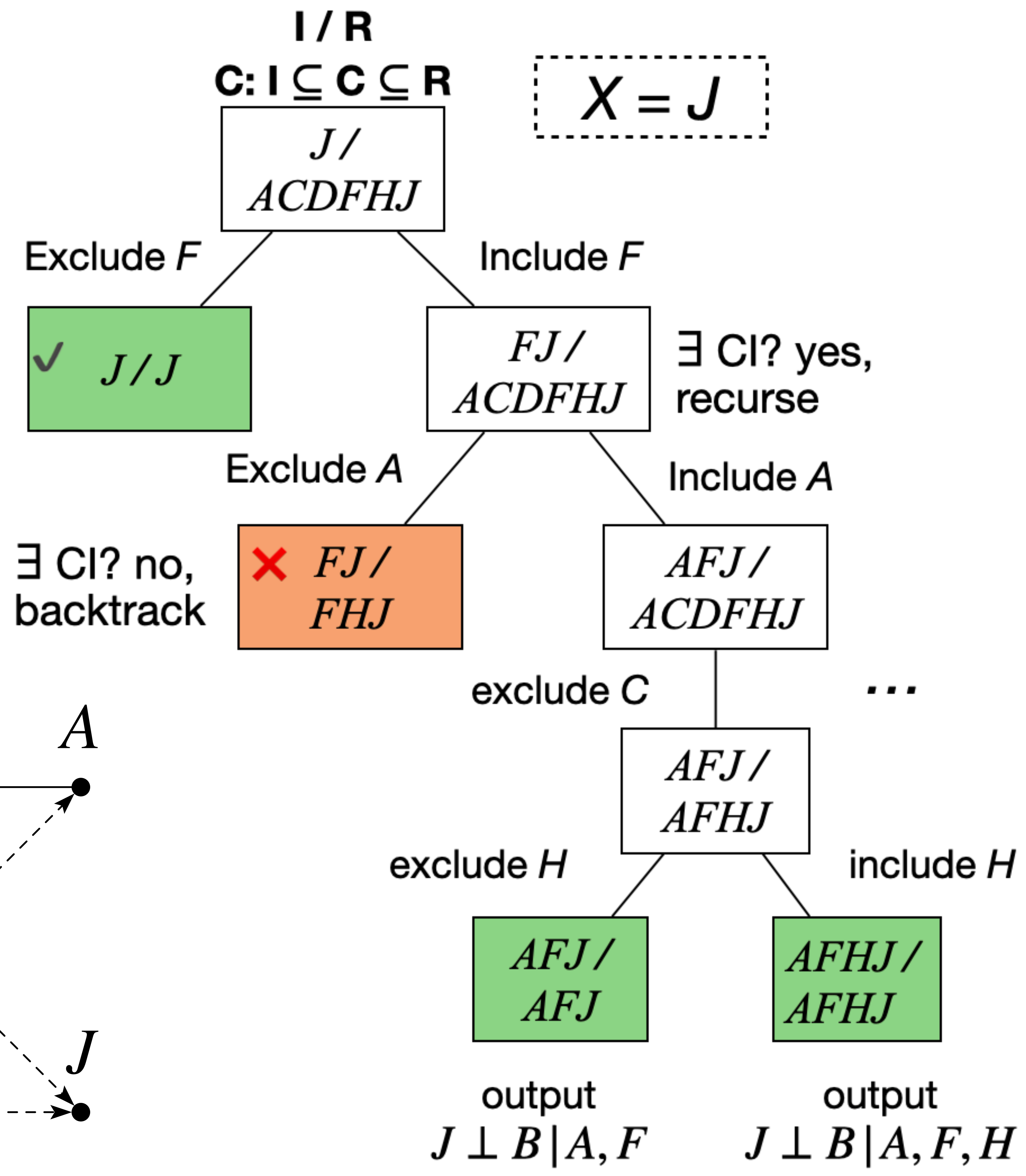
Theorem 1 (Equivalence of C-LMP and GMP)

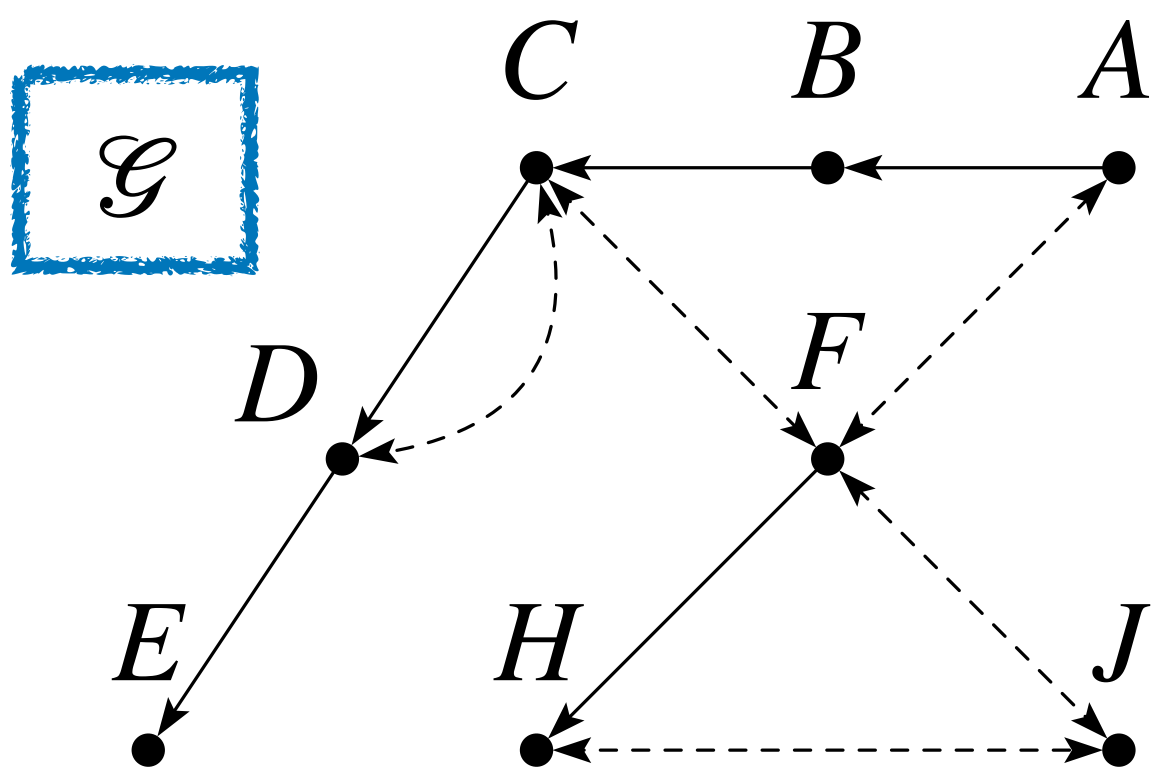
Given $\mathscr{G}$ and a consistent ordering $\mathbf{V}^{\prec}$, a probability distribution $P(\mathbf{v})$ satisfies *the c-component local Markov property* for $\mathscr{G}$ w.r.t $\mathbf{V}^{\prec}$ if and only if it satisfies the global Markov property for $\mathscr{G}$.

Test CIs invoked by C-LMP instead of GMP!

Listing CIs Invoked by C-LMP

`ListCI` enumerates non-vacuous CIs invoked by C-LMP in polynomial delay.





Correctness of `ListCI`

Theorem 2 (Correctness of `ListCI`)

Given $\mathcal{G}$ and a consistent ordering $\mathbf{V}^{\prec}$, `ListCI`$(\mathcal{G}, \mathbf{V}^{\prec})$ enumerates all and only all non-vacuous CIs invoked by the c-component local Markov property for $\mathcal{G}$ w.r.t $\mathbf{V}^{\prec}$.

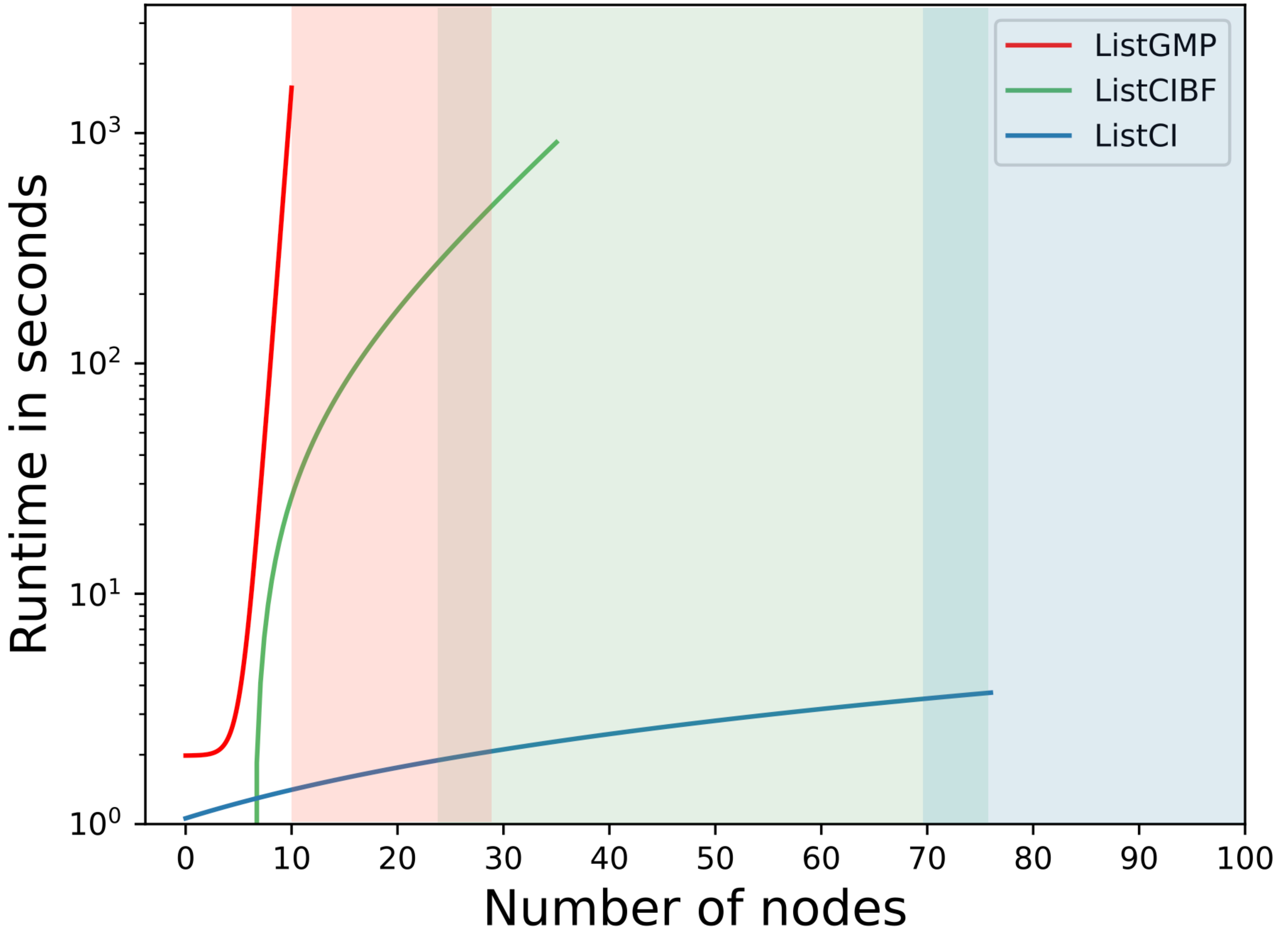
- Time: $O(n^2(n+m))$ delay.
  - To output k CIs, time $O(kn^2(n+m))$
- Space: $O(n(n+m))$.

Experiment 1: ListCI is fast

Comparison of `ListCI` with two baseline algorithms.

- Algorithms: `ListGMP`, `ListCIBF` (brute force C-LMP).
- Graphs: `bnlearn` graphs on up to 100 nodes (Scutari 2010)

`ListCI` is **orders of magnitude faster** than baselines!

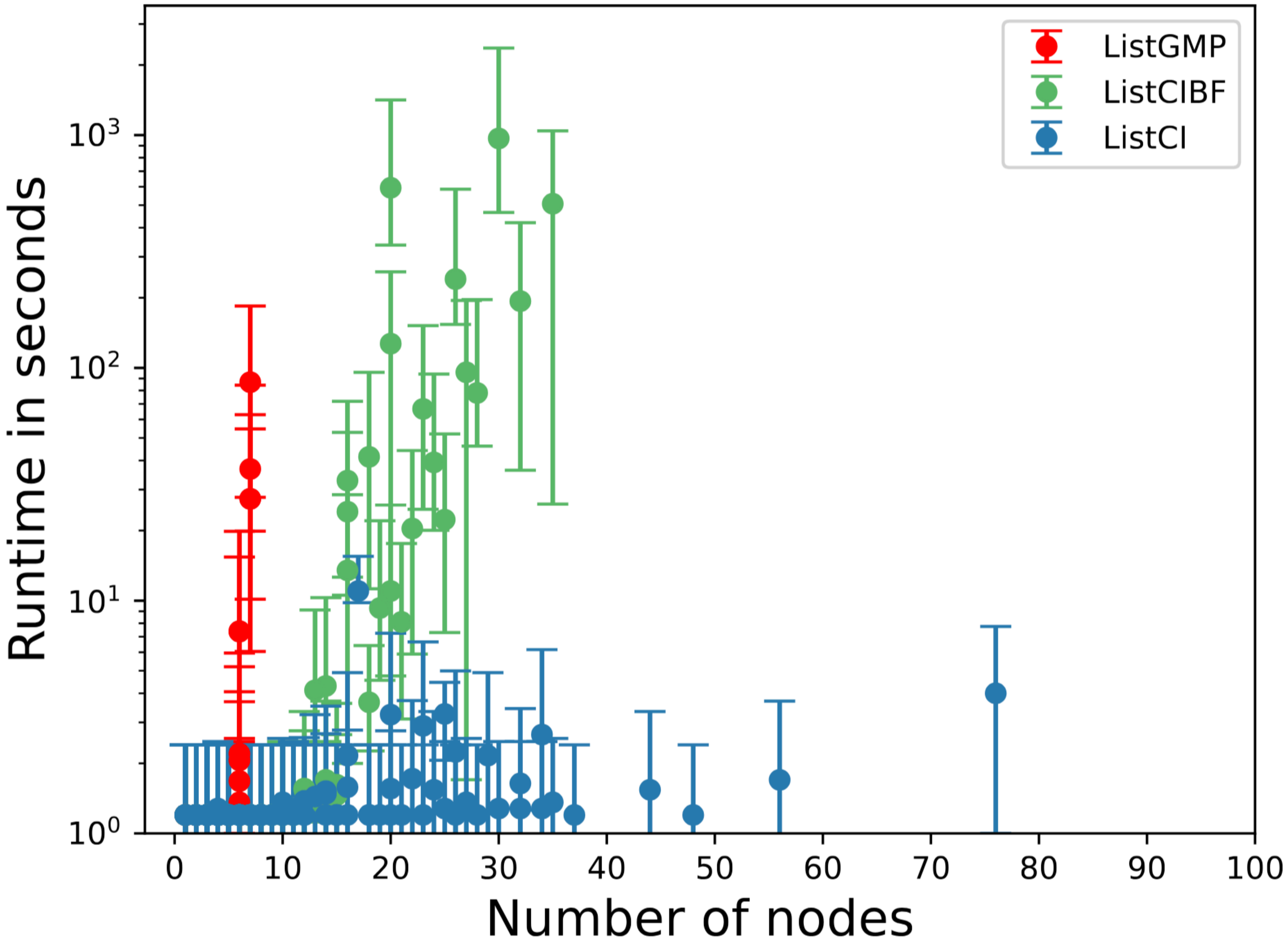


y-axis: log scale.

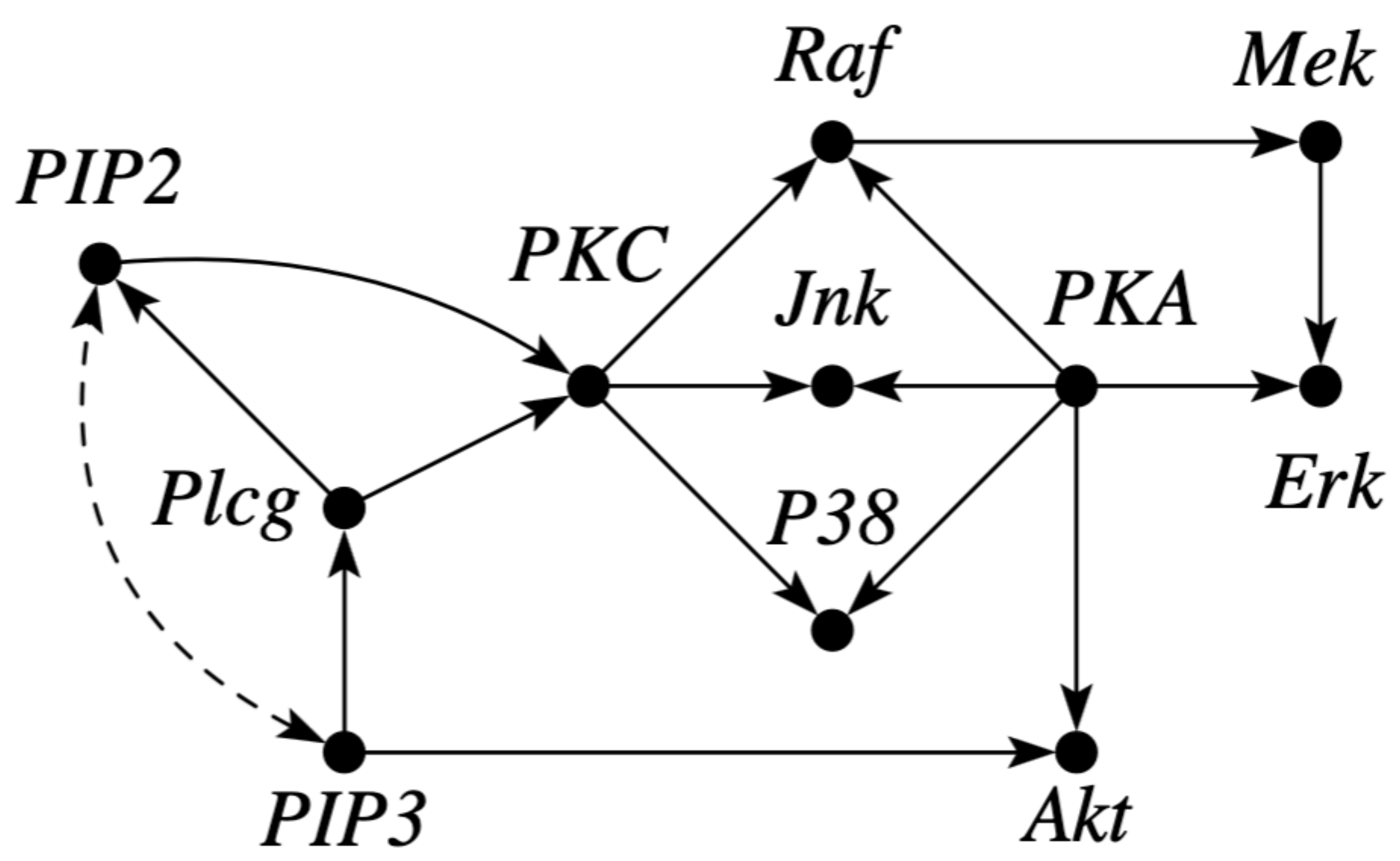
Colored region: algorithm timed out (i.e., spent > 1 hour).

Scutari, M. 2010. Learning Bayesian Networks with the bnlearn R Package. Journal of Statistical Software, 35(3): 1–22

Experiment 1: ListCI is fast



Experiment 2: Testing the Sachs graph



Sachs dataset (853 samples, continuous)

Result: **3 /10 CIs** resulted in $p < 0.05$, suggesting dependence

CIs invoked by C-LMP for the Sachs graph	p-value
PIP3 ⊥ {PKA}	0.175
Plcg ⊥ {PKA} | {PIP3}	0.081
Akt ⊥ {Plcg} | {PIP3, PKA}	0.370
PIP2 ⊥ {Akt, PKA} | {PIP3, Plcg}	0.648
PKC ⊥ {Akt, PIP3, PKA} | {PIP2, Plcg}	0.318
Raf ⊥ {Akt, PIP2, PIP3, Plcg} | {PKA, PKC}	0.036
P38 ⊥ {Akt, PIP2, PIP3, Plcg, Raf} | {PKA, PKC}	0.680
Jnk ⊥ {Akt, P38, PIP2, PIP3, Plcg, Raf} | {PKA, PKC}	0.002
Mek ⊥ {Akt, Jnk, P38, PIP2, PIP3, PKA, PKC, Plcg} | {Raf}	0.544
Erk ⊥ {Akt, Jnk, P38, PIP2, PIP3, PKC, Plcg, Raf} | {Mek, PKA}	0.000

Experiment 3: Phase Transitions

3-SAT: $\phi : (x_1 \vee \neg x_3 \vee x_4) \wedge (\neg x_1 \vee x_2 \vee \neg x_4)$
Does there exist (x_1, x_2, x_3, x_4) satisfying ϕ?

$\alpha = \frac{\#clauses}{\#vars}$ (Selman et al 1996)

- large $\alpha \rightarrow$ often unsatisfiable
- small $\alpha \rightarrow$ often satisfiable
- $\alpha =$ in 'critical region': hard to decide!

C-LMP: how many CIs are encoded in a $\mathcal{G}$ with n nodes and m edges?

$\alpha \approx$ # nodes / # edges

- large $\alpha \rightarrow$ small c-components → few CIs
- small $\alpha \rightarrow$ dense → few Cis
- α in 'critical region': $O(n2^s)$?

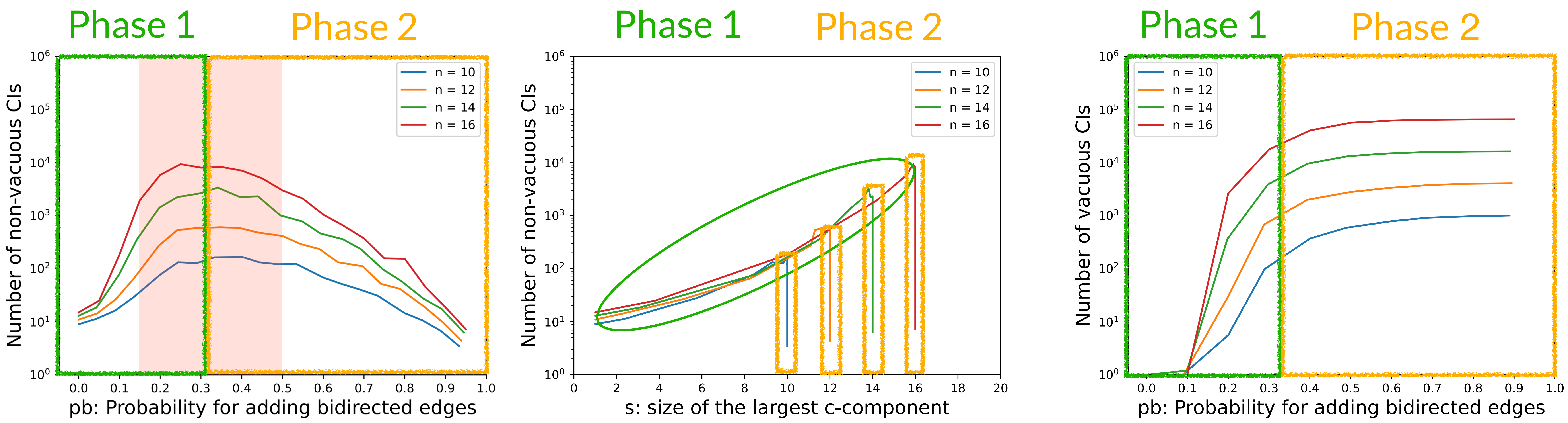
Selman, B., Mitchell, D.G., and Levesque H.J.. 1996. Generating hard satisfiability problems. Artificial Intelligence, 811-2): 17-29.

Experiment 3: Phase Transitions

nCI depends primarily on two factors:

A. Size of the largest $\mathbf{C}$ ($s \leq n$), and

B. Sparsity of $\mathbf{C}$ determined by number of bidirected edges



Phase 1

(i) more bidirected edges $\Rightarrow$ larger $\mathbf{C}$

(ii) larger $\mathbf{C} \Rightarrow$ exponentially more CIs: $\Theta(n2^s)$

Phase 2

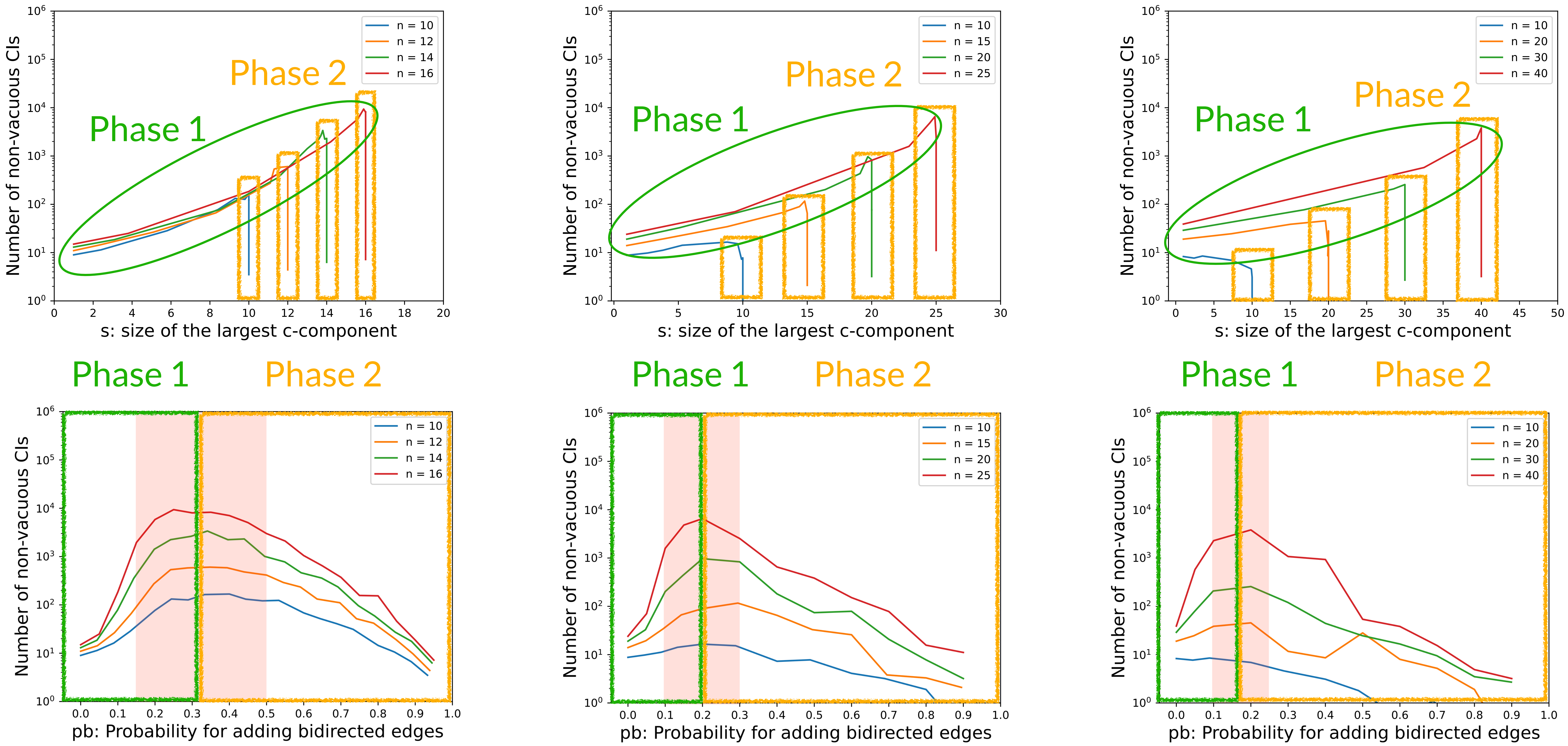
(i) s is constant

(ii) more bidirected edges $\Rightarrow$ denser connectivity $\Rightarrow$ fewer CIs

Experiment 3: Phase Transitions

expected number of directed edges

$md = 0$ | $md = n$ | $md = 2n$



Conclusions

Testing causal models before inference is good practice.

To support this, and other tasks, we introduce:

The c-component local Markov property, for testing hidden-variable graphs with with fewer CI tests against non-parametric probability distributions.

The algorithm ListCI, for testing models with C-LMP polynomial delay.

Future Work

- Repairing misspecified causal models

A. Ejaz, E. Bareinboim. *Less Greedy Equivalence Search.* 2025. In submission. https://causalai.net/r134.pdf

- Testing faithfulness
- Controlling the size of conditioning sets

Thank you! Questions?

References

- Bareinboim, E.; Correa, J. D.; Ibeling, D.; and Icard, T. 2022. On Pearl’s Hierarchy and the Foundations of Causal Inference. In Probabilistic and Causal Inference: The Works of Judea Pearl, 507–556. New York, NY, USA: Association for Computing Machinery, 1st edition
- Johnson, D. S.; Yannakakis, M.; and Papadimitriou, C. H. 1988. On generating all maximal independent sets. Information Processing Letters, 27(3): 119–123
- Lauritzen, S. L.; Dawid, A. P.; Larsen, B. N.; and Leimer, H. G. G. 1990. Independence Properties of Directed Markov Fields. Networks, 20(5): 491–505.
- Pearl, J and Paz, A. 1986. Graphoids: graph-based logic for reasoning about relevance relations or when would x tell you more about y if you already know z? In Proceedings of the 7th European Conference on Artificial Intelligence - Volume 2 (ECAI'86). North-Holland, 357–363.
- Pearl, J. 1995. Causal diagrams for empirical research. Biometrika, 82(4): 669–688.
- Pearl, J. 1988. Probabilistic Reasoning in Intelligent Systems. San Mateo, CA: Morgan Kaufmann
- Richardson, T. 2003. Markov properties for acyclic directed mixed graphs. Scandinavian Journal of Statistics, 30(1): 145–157.
- Sachs, K.; Perez, O.; Pe’er, D.; Lauffenburger, D. A.; and Nolan, G. P. 2005. Causal protein-signaling networks derived from multiparameter single-cell data. Science, 308(5721): 523–529.
- Scutari, M. 2010. Learning Bayesian Networks with the bnlearn R Package. Journal of Statistical Software, 35(3): 1–22
- Selman, B., Mitchell, D.G., and Levesque H.J.. 1996. Generating hard satisfiability problems. Artificial Intelligence, 81(1-2): 17-29.