The Query Complexity of Verifying a Causal Graph from Single-Node Interventions

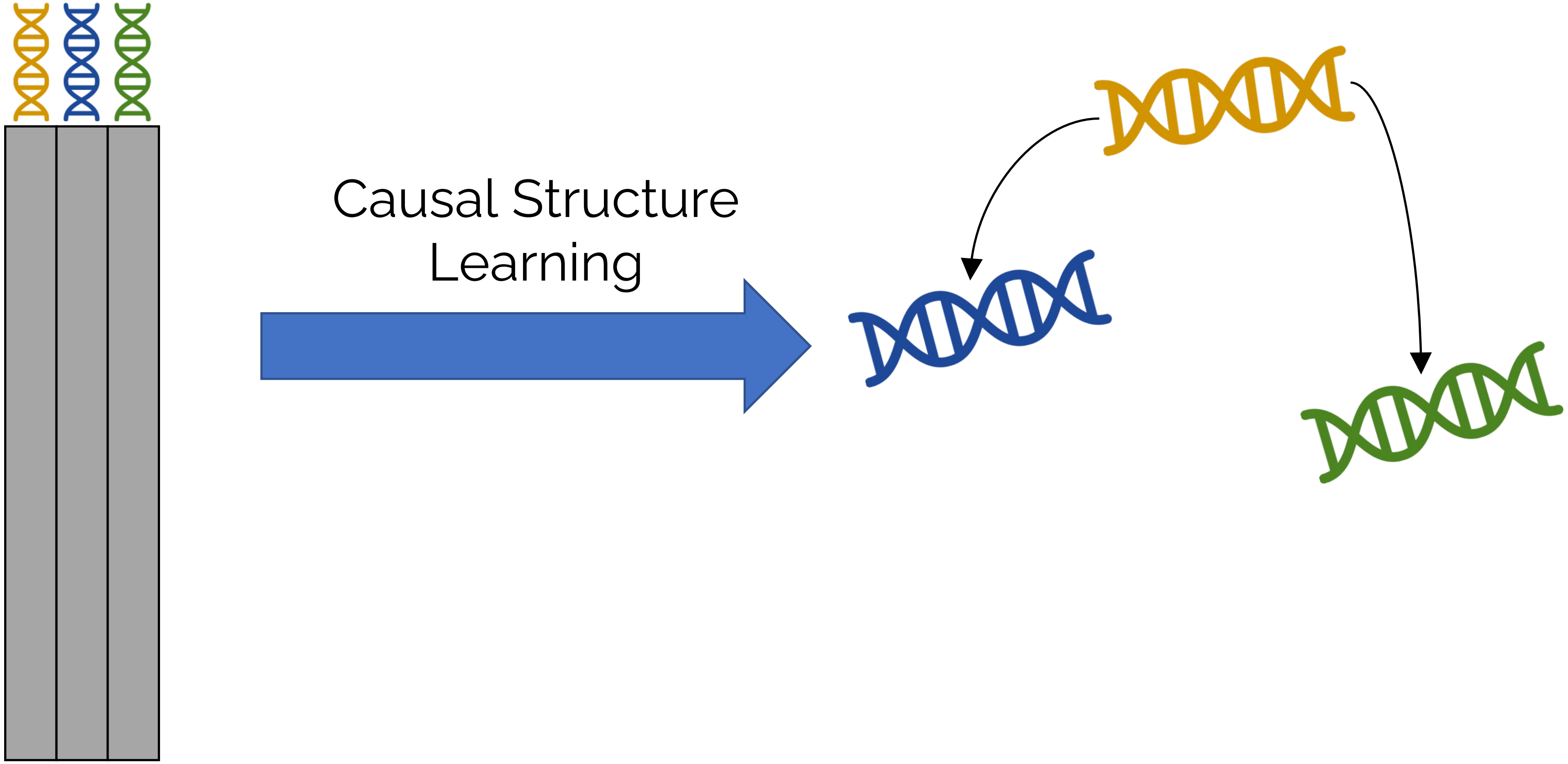
or: How hard is it to learn a causal model?

Chandler Squires

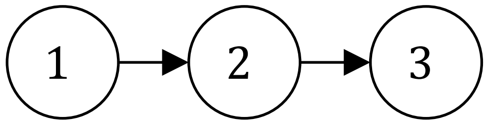
Causal Structure Learning (aka, Causal Discovery)

An inverse problem: given data, find the causal graph

A Canonical Application: Learning Gene Regulatory Networks

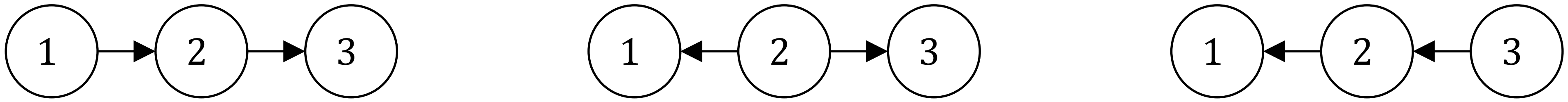


Observational data is (in general) insufficient to identify the true graph

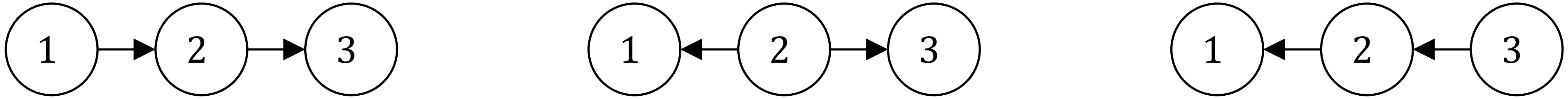


Observational data is (in general) insufficient to identify the true graph

Observational data is (in general) insufficient to identify the true graph



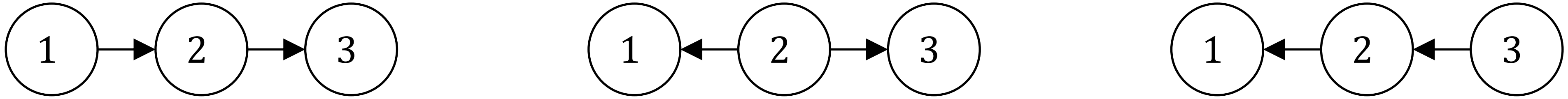
Observational data is (in general) insufficient to identify the true graph



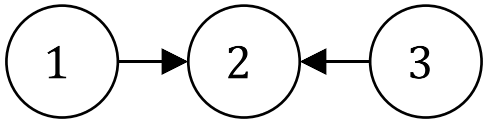
the **only**[1] Level 1 property (i.e., at the level of $\mathbb{P}(X_1, X_2, X_3)$) entailed by these graphs is $X_1 \perp\!\!\!\perp X_3 \mid X_2$.

[1]without further parametric assumptions

Observational data is (in general) insufficient to identify the true graph



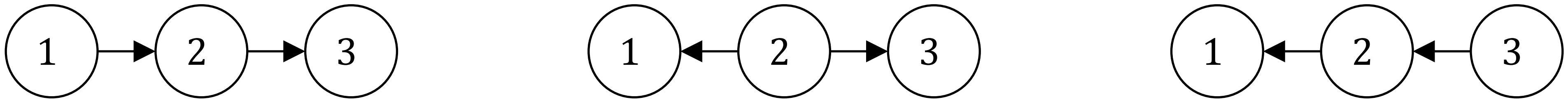
the **only**[1] Level 1 property (i.e., at the level of $\mathbb{P}(X_1, X_2, X_3)$) entailed by these graphs is $X_1 \perp\!\!\!\perp X_3 \mid X_2$.



However, this graph implies that $X_1 \perp\!\!\!\perp X_3$

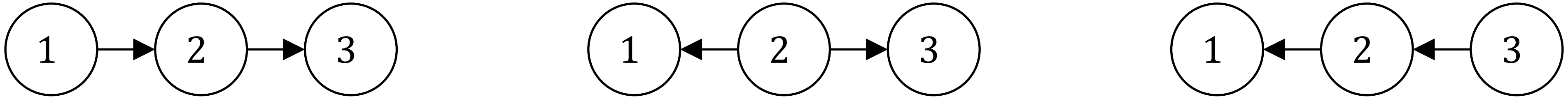
[1]without further parametric assumptions

What **can** we identify from observational data?



We call graphs which imply the same set of conditional independences **Markov equivalent.**

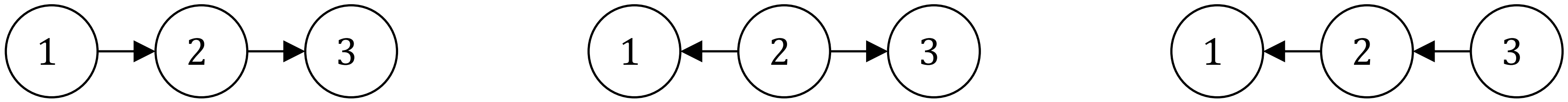
What **can** we identify from observational data?



We call graphs which imply the same set of conditional independences **Markov equivalent.**

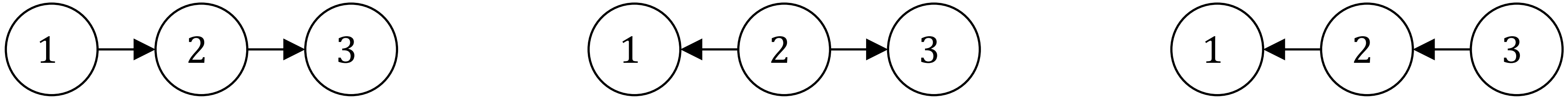
From observational data, we can only identify up to the **Markov equivalence class.**

What **can** we identify from observational data?



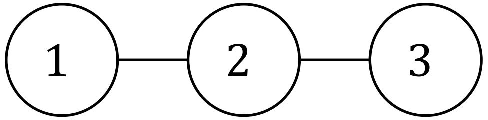
Two graphs are Markov equivalent if and only if they have the same adjacencies and **v-structures** (patterns of the form $i \rightarrow k \leftarrow j$ with i and j non-adjacent).

What **can** we identify from observational data?



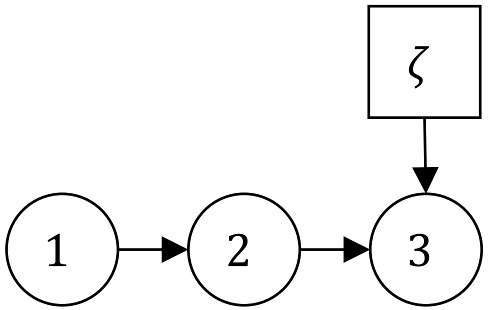
We represent the Markov equivalence class by the **essential graph**, which:

- Has the same adjacencies as the members of the Markov equivalence class
- Only has an edge directed if it is directed the same way across the entire Markov equivalence class.



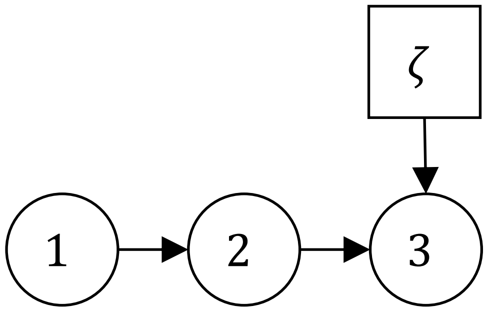
What do we learn from interventions?

Single-node interventions teach us the directions of edges adjacent to the intervened node

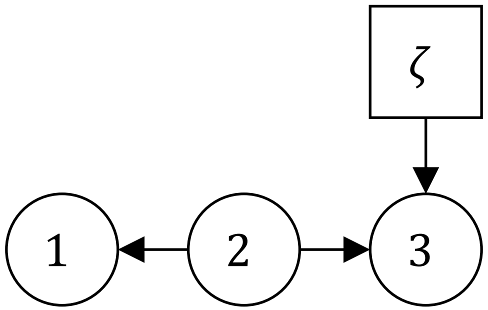


Only X_3 is affected

Single-node interventions teach us the directions of edges adjacent to the intervened node

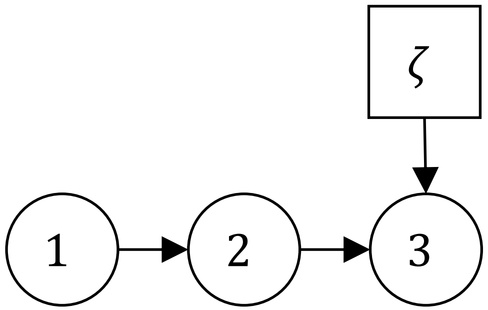


Only X_3 is affected

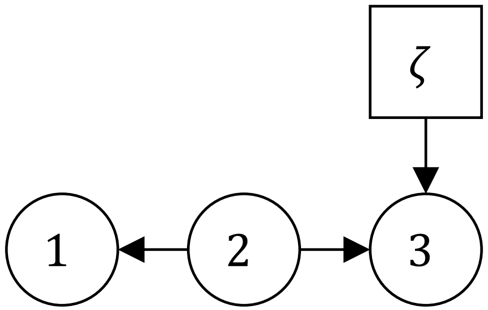


Only X_3 is affected

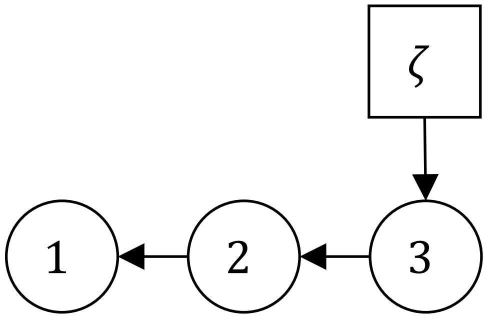
Single-node interventions teach us the directions of edges adjacent to the intervened node



Only X_3 is affected

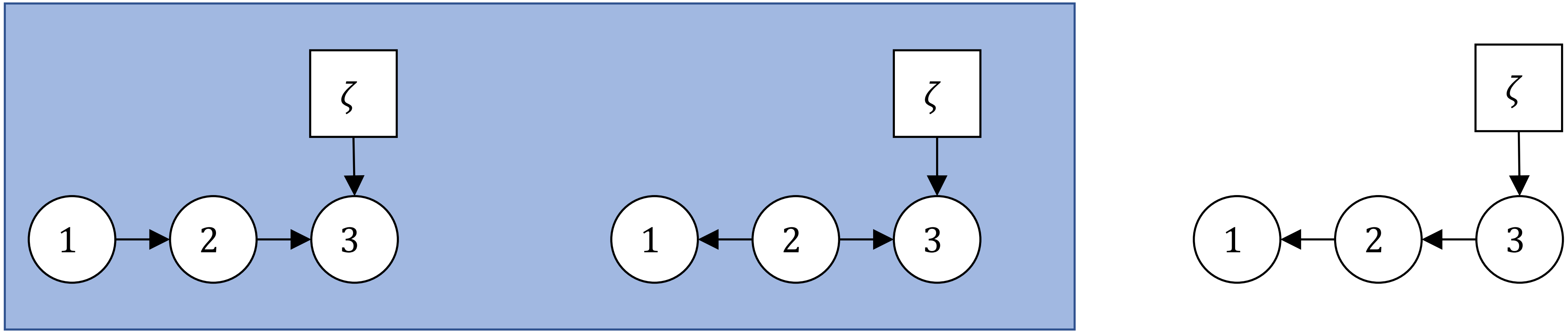


Only X_3 is affected

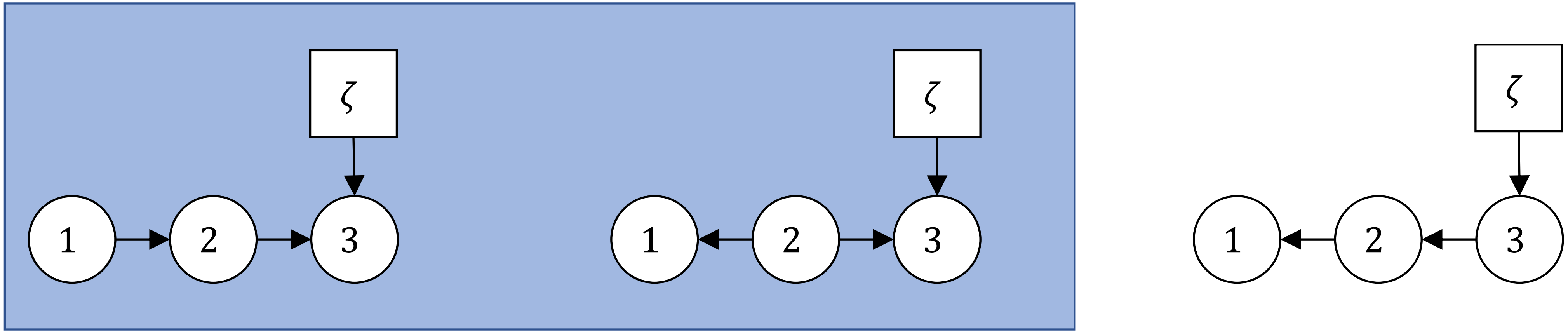


All three variables are affected

Interventional Markov equivalence



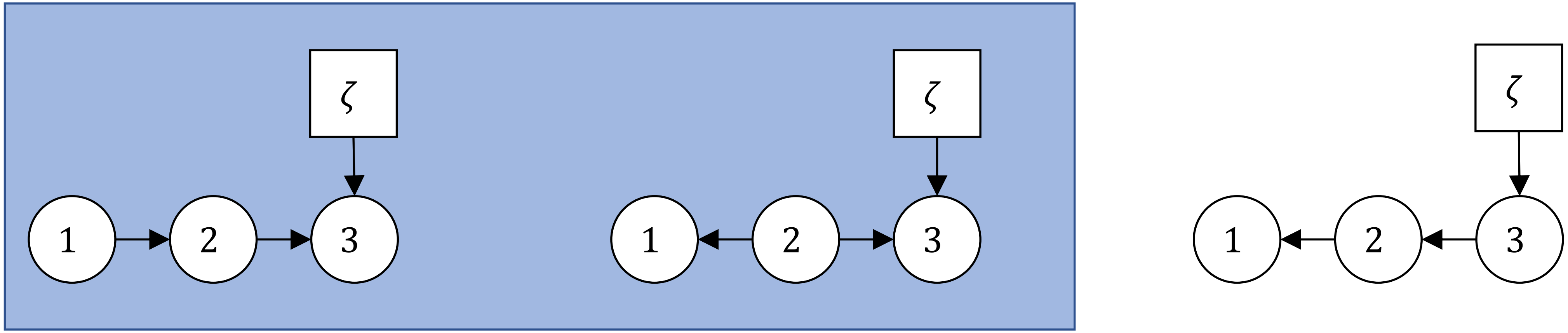
Interventional Markov equivalence



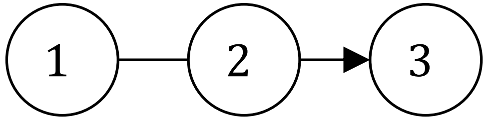
Denote the set of interventions as $\mathcal{J} = \{\{3\}\}$

We refer to the set of graphs entailing the same conditional independences and **conditional invariances** under $\mathcal{J}$ as the **$\mathcal{J}$-Markov equivalence class.**

Interventional Markov equivalence

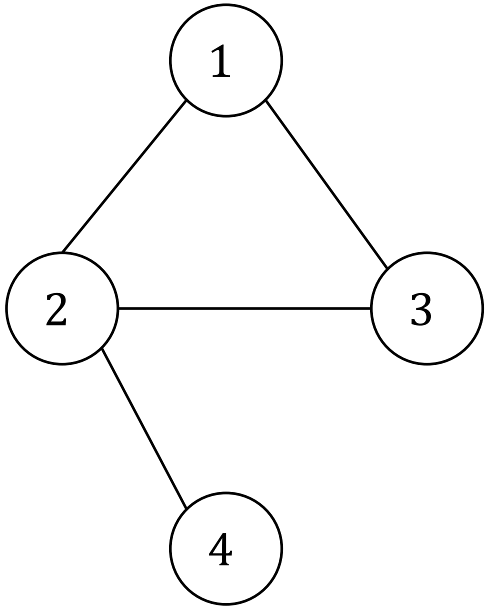


We represent the $\mathcal{I}$-Markov equivalence class using the **$\mathcal{I}$–essential graph**.

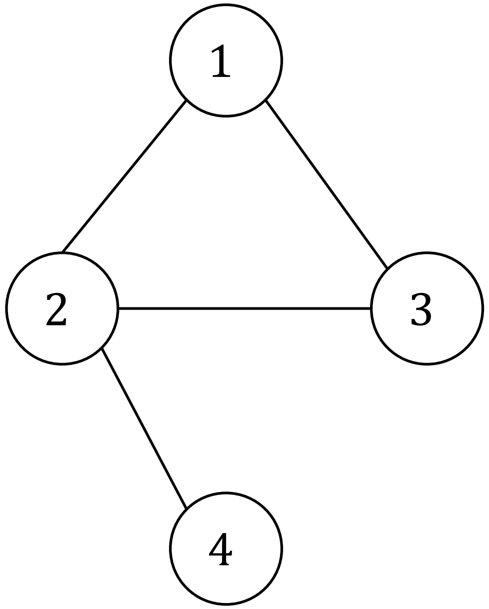


Do interventions teach us any other orientations?

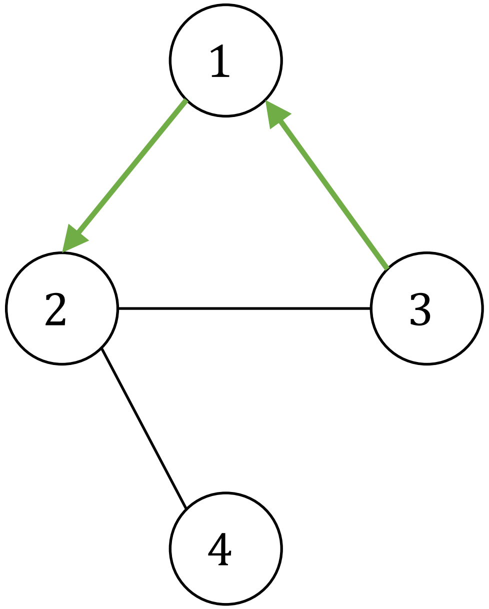
Yes: through the application of Meek's orientation rules



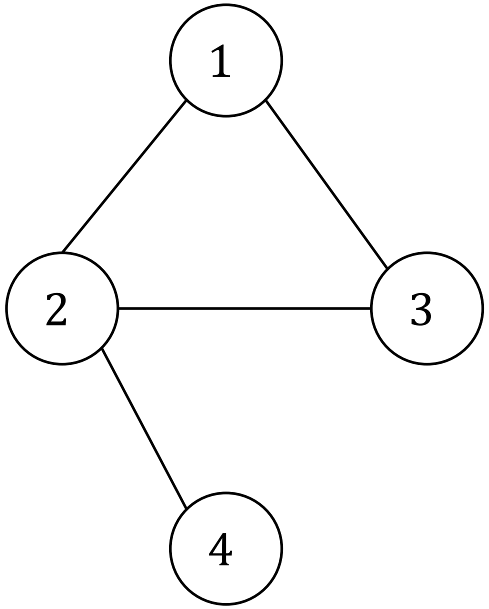
From observational data



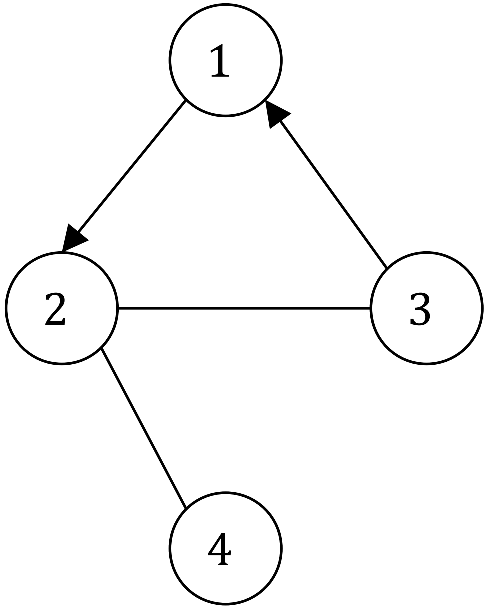
From observational data



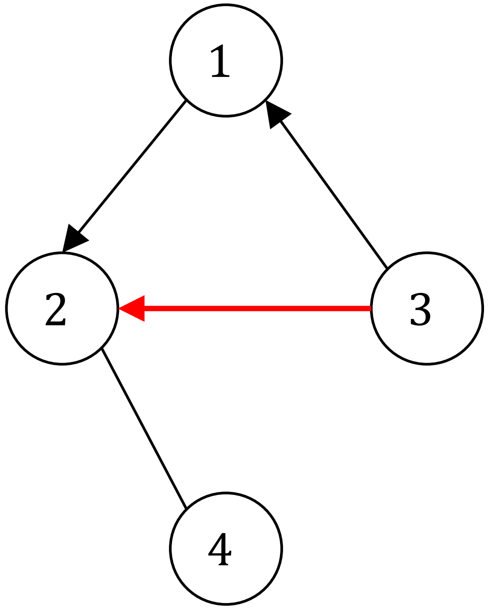
Adjacencies from an intervention on X_1



From observational data

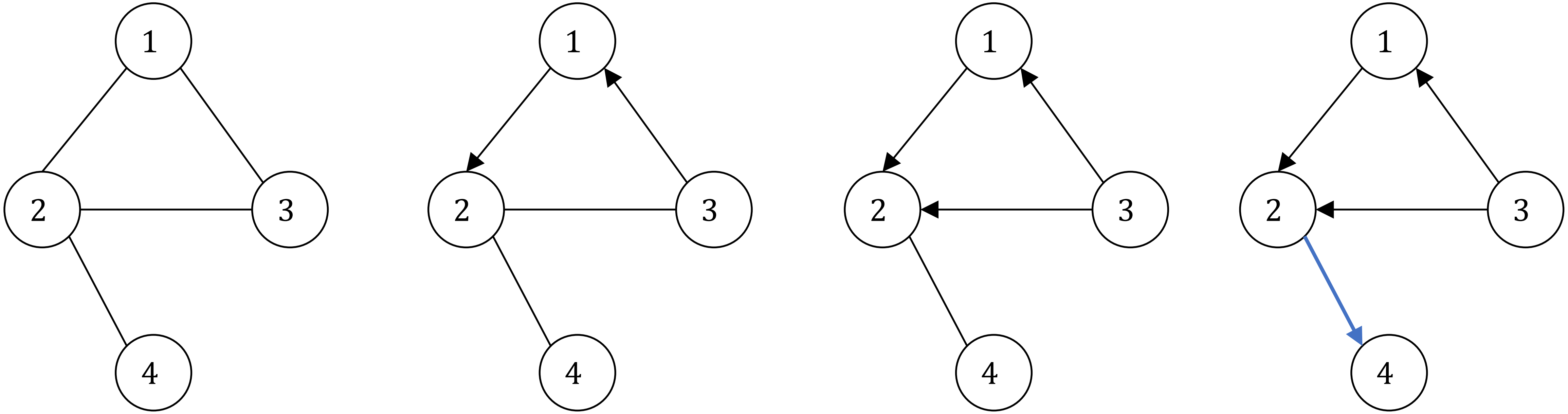


Adjacencies from an intervention on X_1



Avoid cycles

(Meek rule 2)

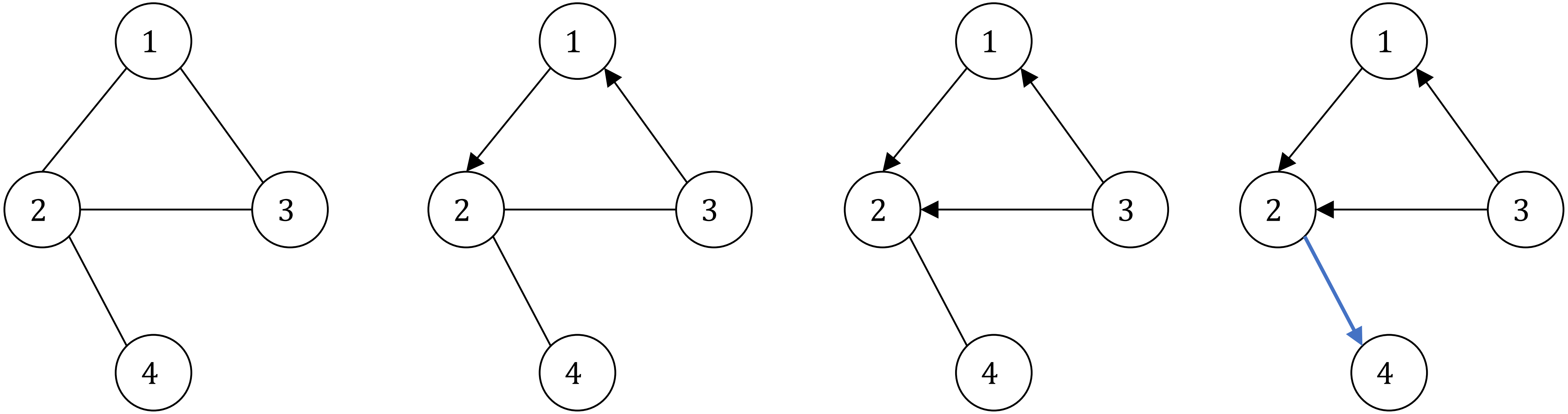


From observational data

Adjacencies from an intervention on X_1

Avoid cycles (Meek rule 2)

Avoid new v-structures (Meek rule 1)



From observational data

Adjacencies from an intervention on X_1

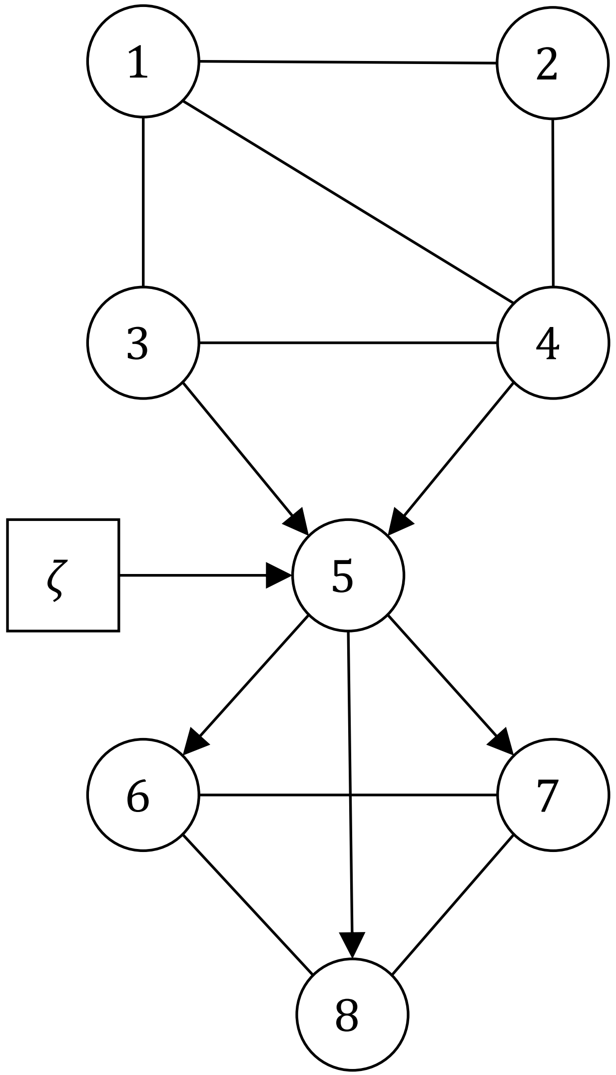
Avoid cycles (Meek rule 2)

Avoid new v-structures (Meek rule 1)

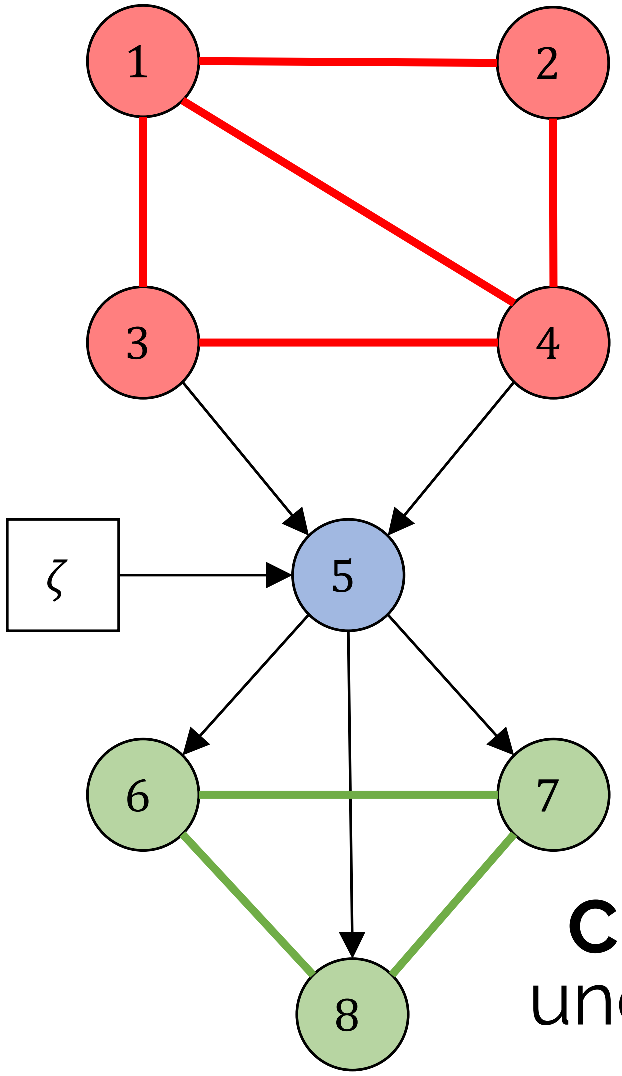
Note: There are two other Meek orientation rules; together these four rules are complete.

How many single-node interventions are required to fully orient a DAG?

$\mathcal{J}$-essential graphs decompose nicely

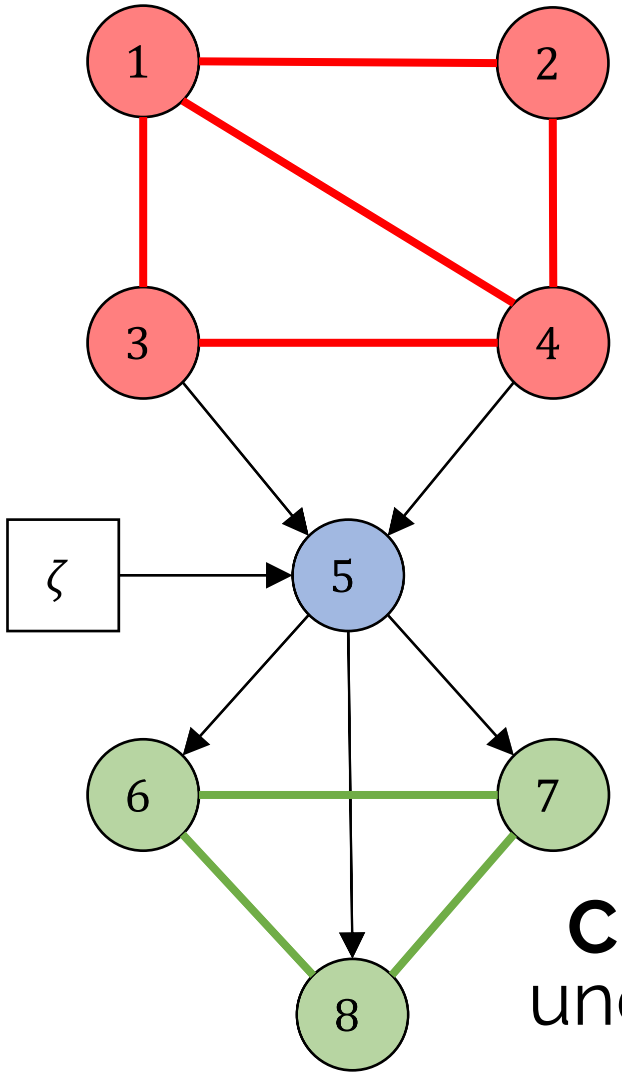


$\mathcal{J}$-essential graphs decompose nicely



Chain components:
undirected connected components

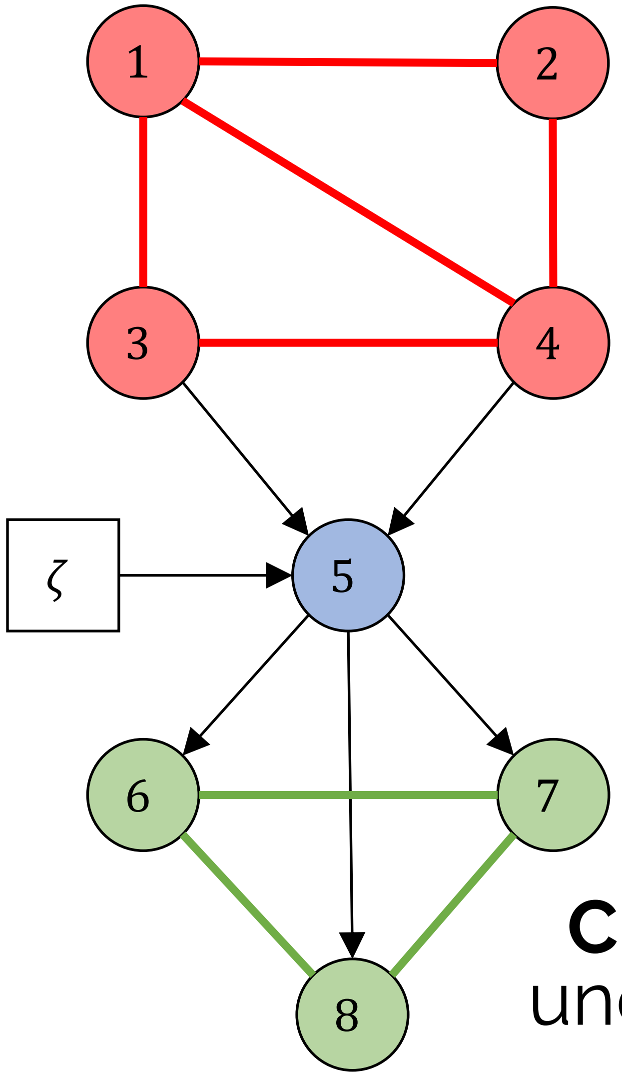
$\mathcal{J}$-essential graphs decompose nicely



- The chain components are **chordal.**
  - i.e., there are no cycles of length ≥ 4 which don't have a chord.

Chain components:
undirected connected components

$\mathcal{J}$-essential graphs decompose nicely



- The chain components are **chordal.**
  - i.e., there are no cycles of length ≥ 4 which don't have a chord.
- The orientations within separate chain components are **independent.**

Chain components:
undirected connected components

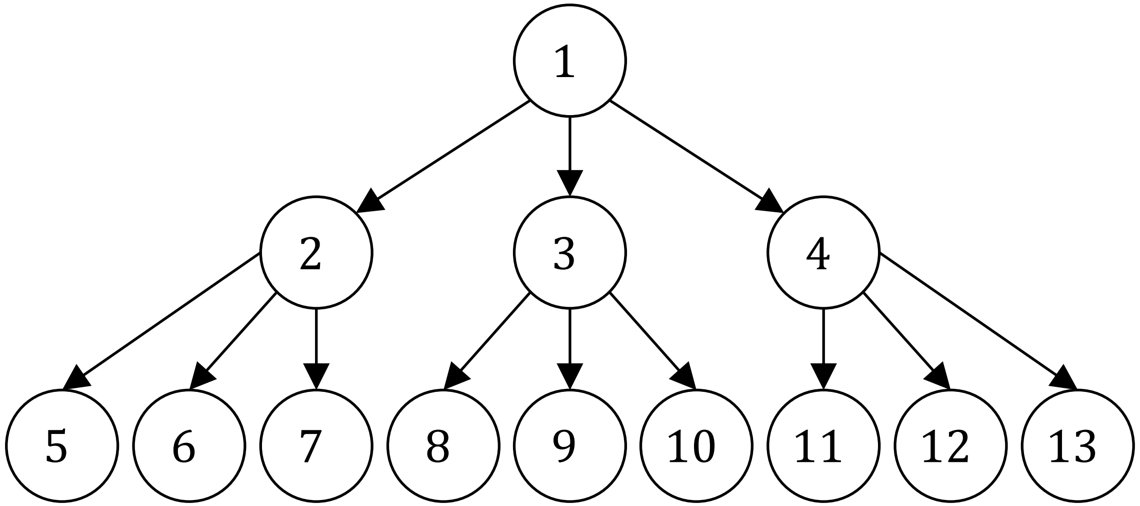
Verifying intervention sets

- Given a DAG G, we call a set $\mathcal{I}$ of interventions a **verifying intervention set (VIS)** if the $\mathcal{I}$-essential graph of G has no undirected edges.

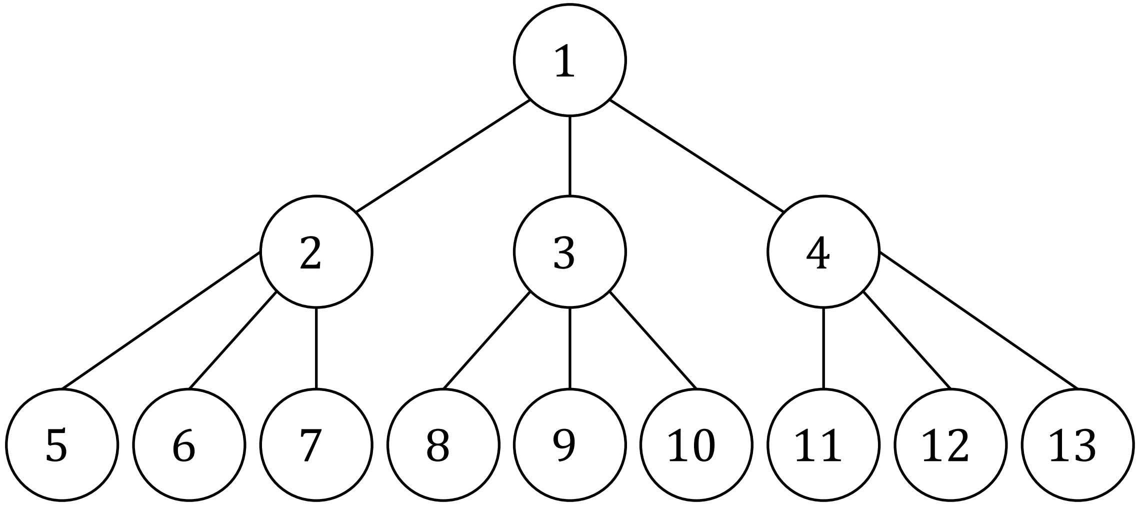
Verifying intervention sets

- Given a DAG G, we call a set $\mathcal{I}$ of interventions a **verifying intervention set (VIS)** if the $\mathcal{I}$-essential graph of G has no undirected edges.
- We call a set $\mathcal{I}$ of interventions a **minimal VIS (MVIS)** if it has the minimum cardinality of all VISes.

Special case: trees

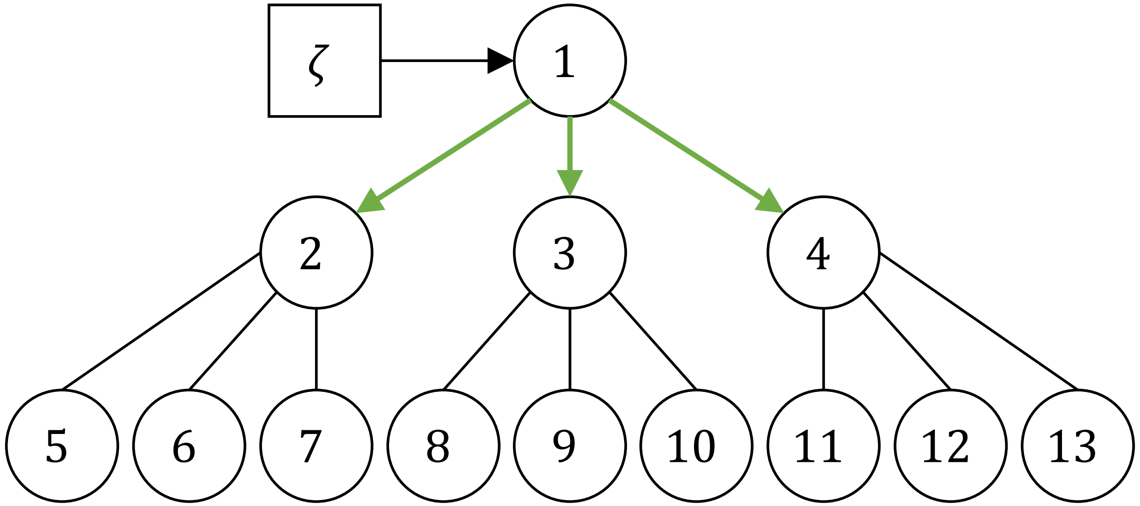


Special case: trees

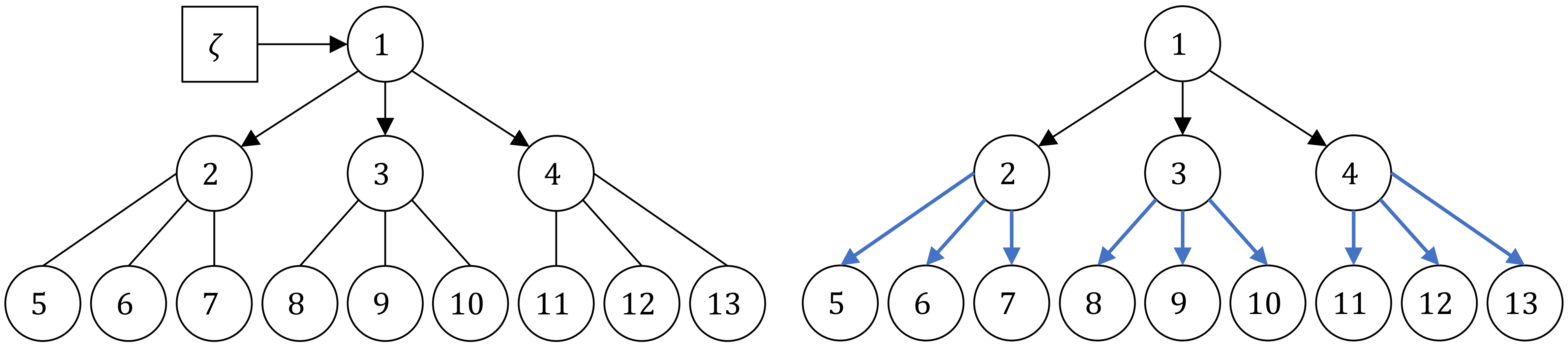


The essential graph

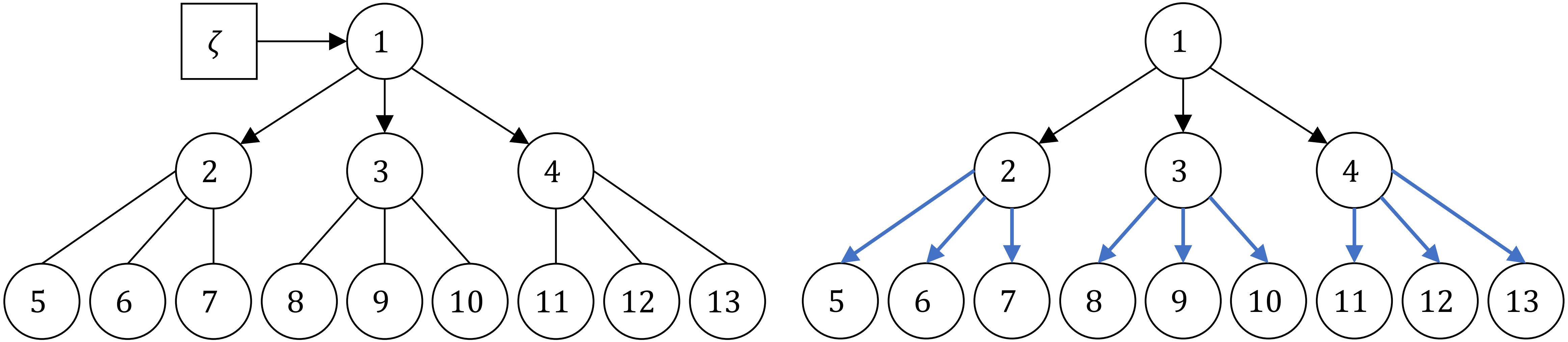
Special case: trees



Special case: trees

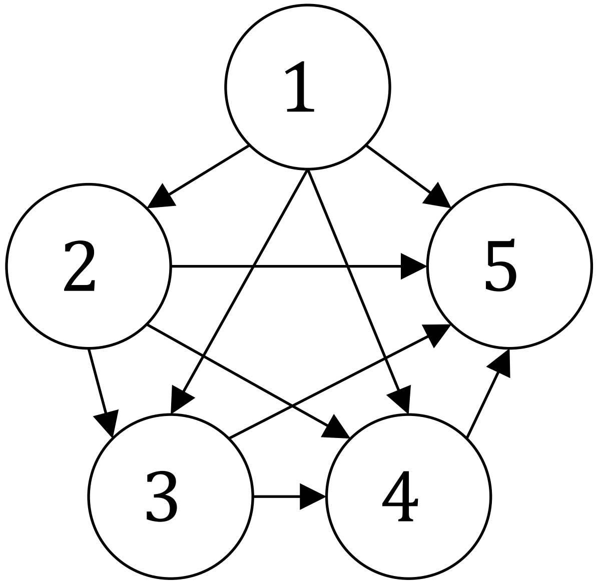


Special case: trees

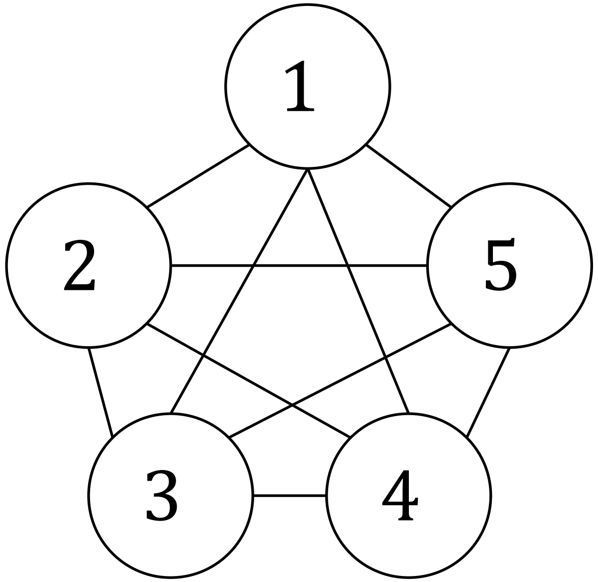


Only **one** intervention
is required

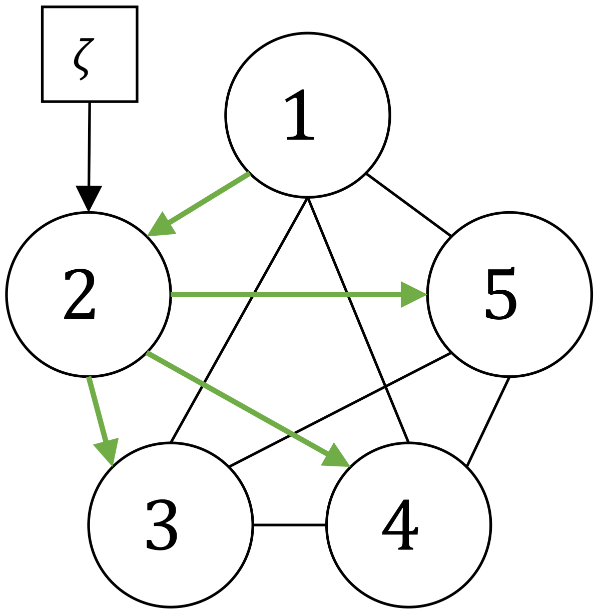
Special case: cliques



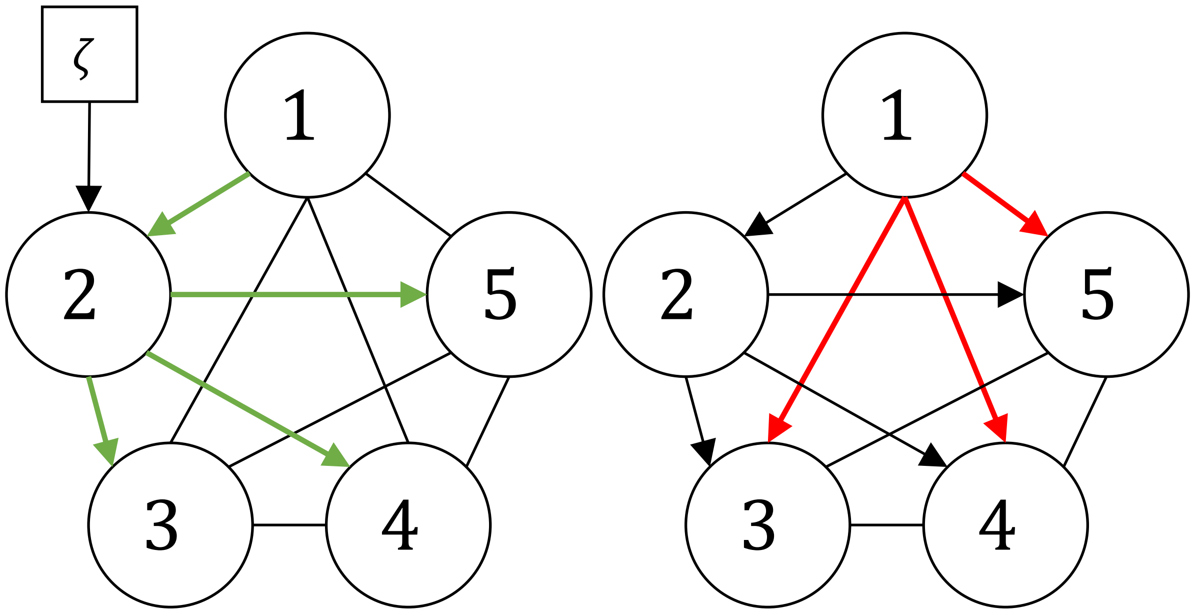
Special case: cliques



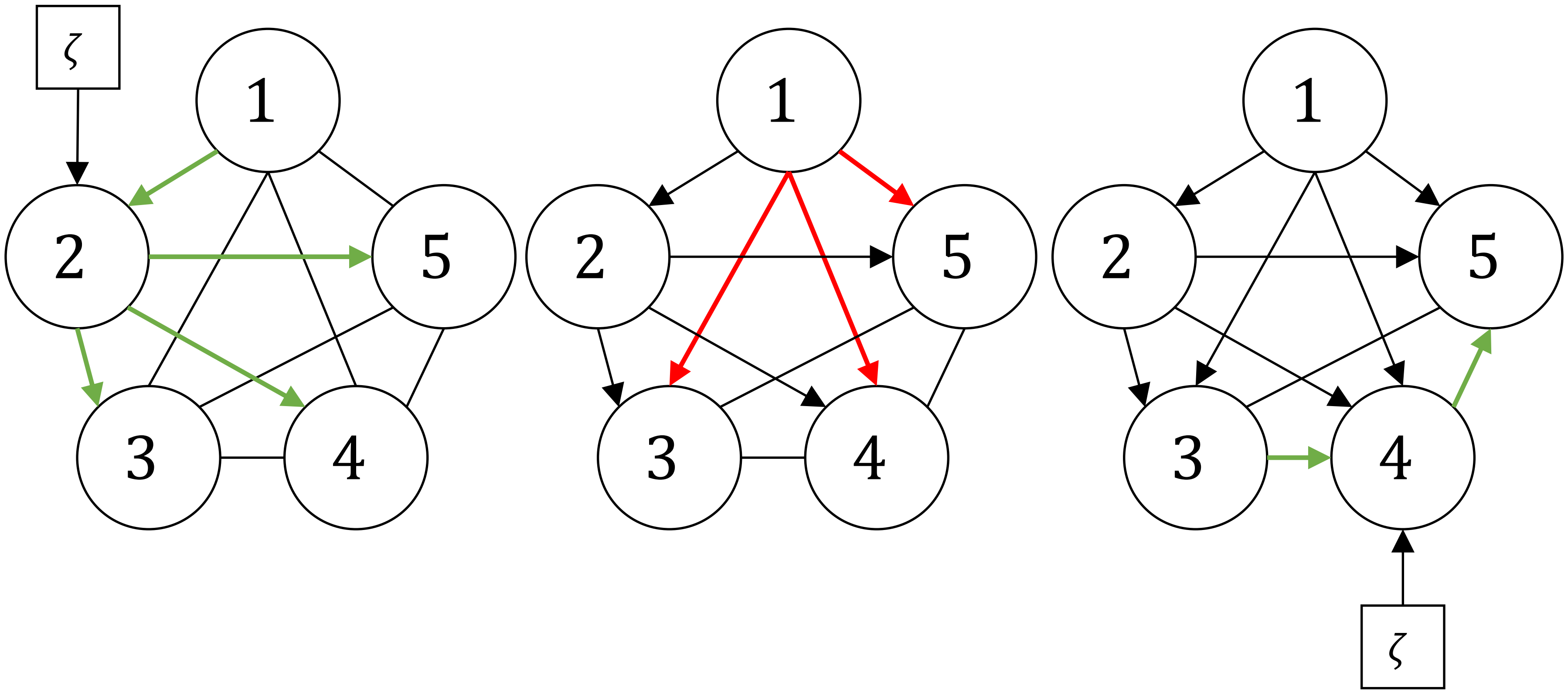
Special case: cliques



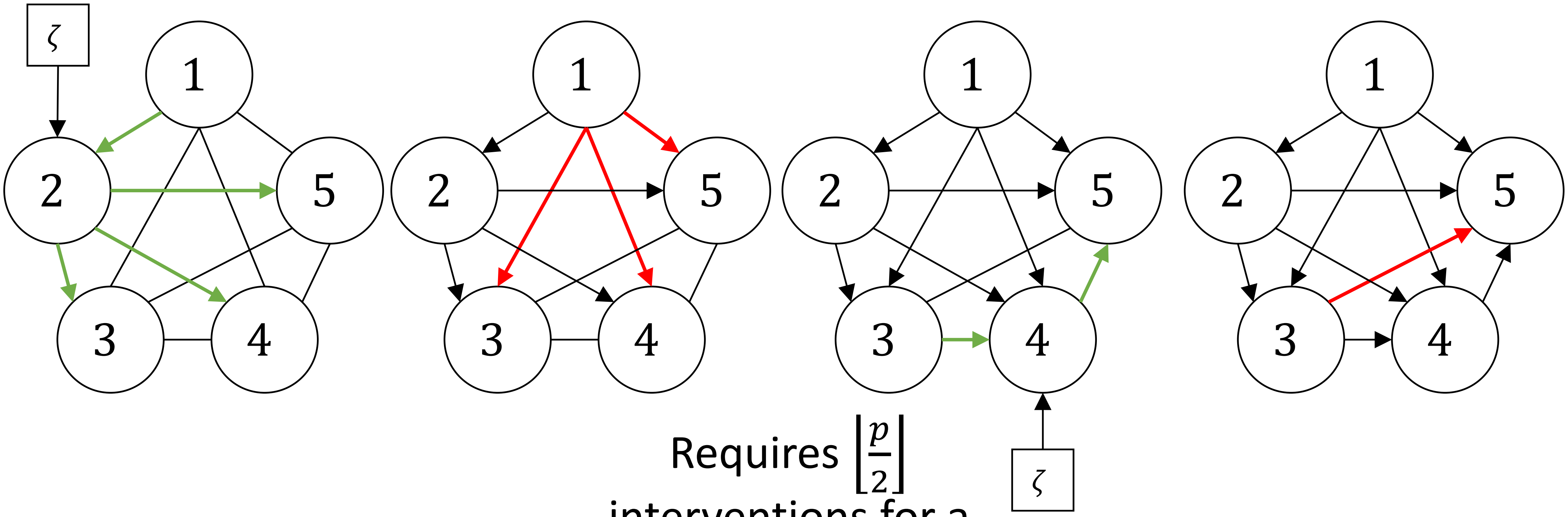
Special case: cliques



Special case: cliques



Special case: cliques



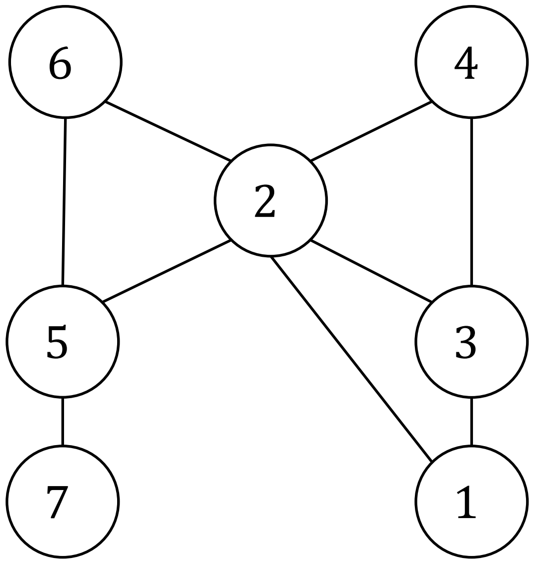
Requires $\left\lfloor \frac{p}{2} \right\rfloor$ interventions for a clique on p nodes

A general solution

Papers

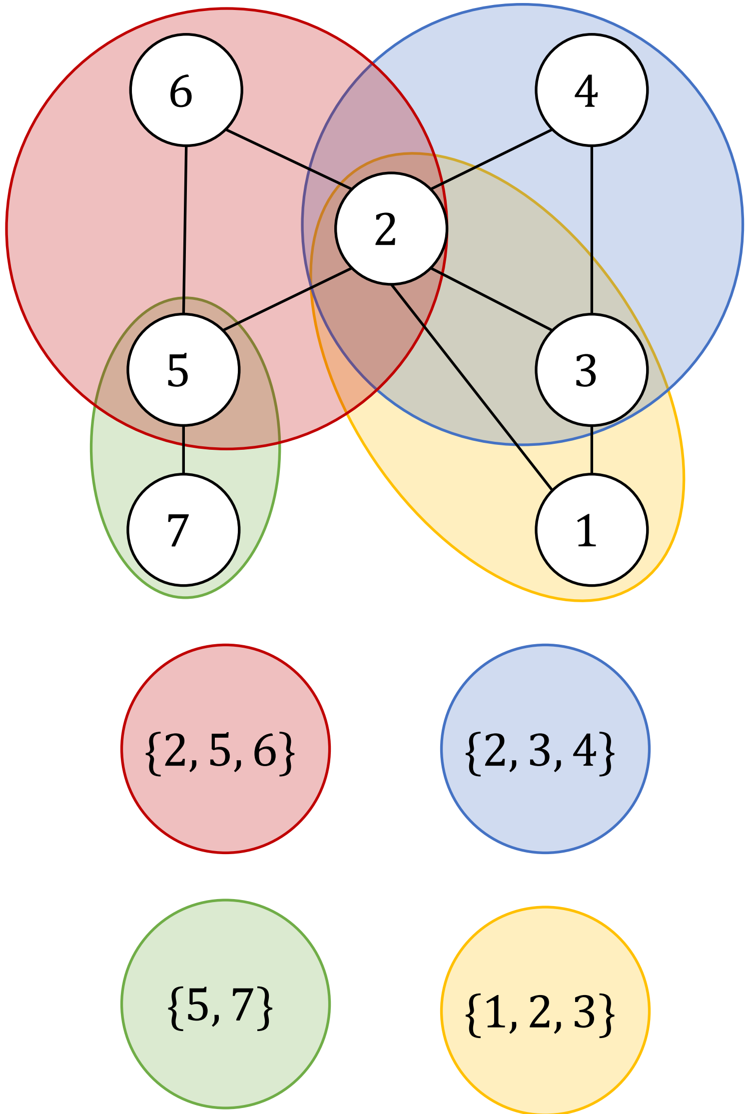
- **[SMG+]** *Active Structure Learning of Causal DAGs via Directed Clique Trees*, **Squires, C.,** Magliacane, S., Greenewald, K., Katz, D., Kocaoglu, M. and Shanmugam, K., 2020.

Clique trees of chordal graphs



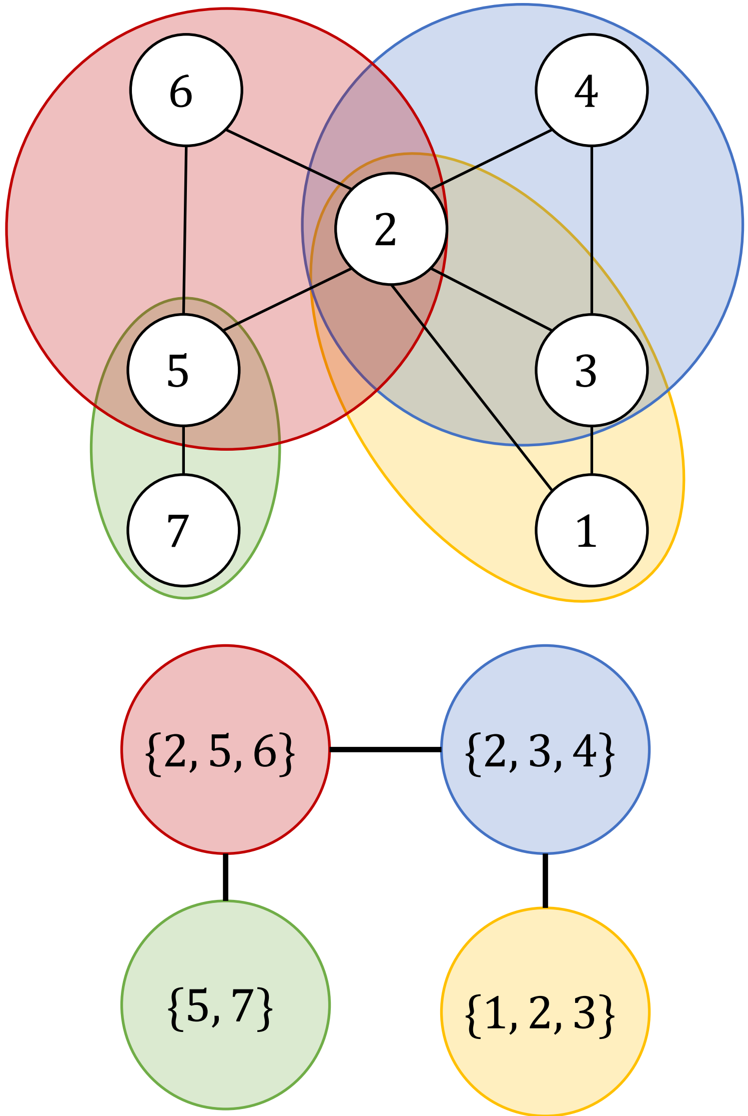
Clique trees of chordal graphs

- A new graph with:
  - Nodes being the **maximal cliques** of the original graph.



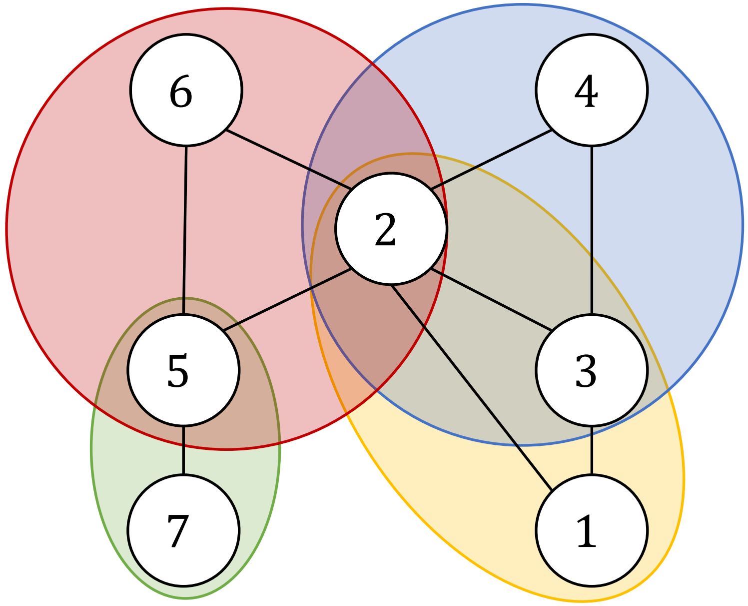
Clique trees of chordal graphs

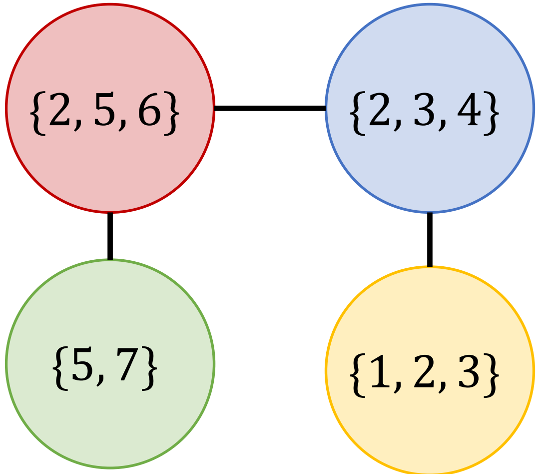
- A new graph with:
  - Nodes being the **maximal cliques** of the original graph.
  - Edges forming a tree such that the **running intersection property** holds.



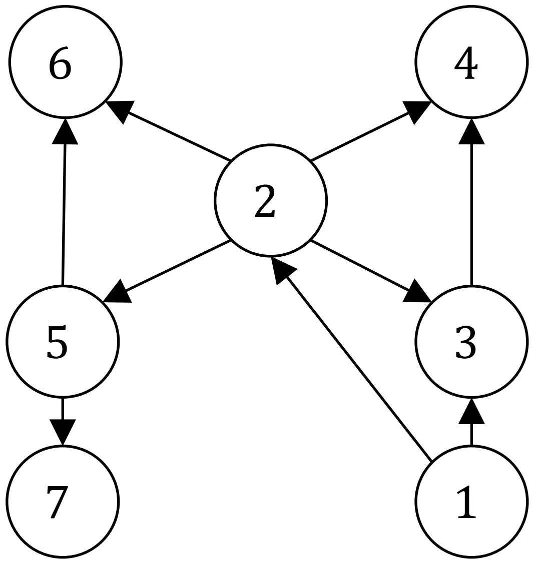
Clique trees of chordal graphs

- A new graph with:
  - Nodes being the **maximal cliques** of the original graph.
  - Edges forming a tree such that the **running intersection property** holds.
- The running intersection property: given two cliques C_1 and C_2, their intersection $C_1 \cap C_2$ is contained in all cliques along the path from C_1 to C_2.



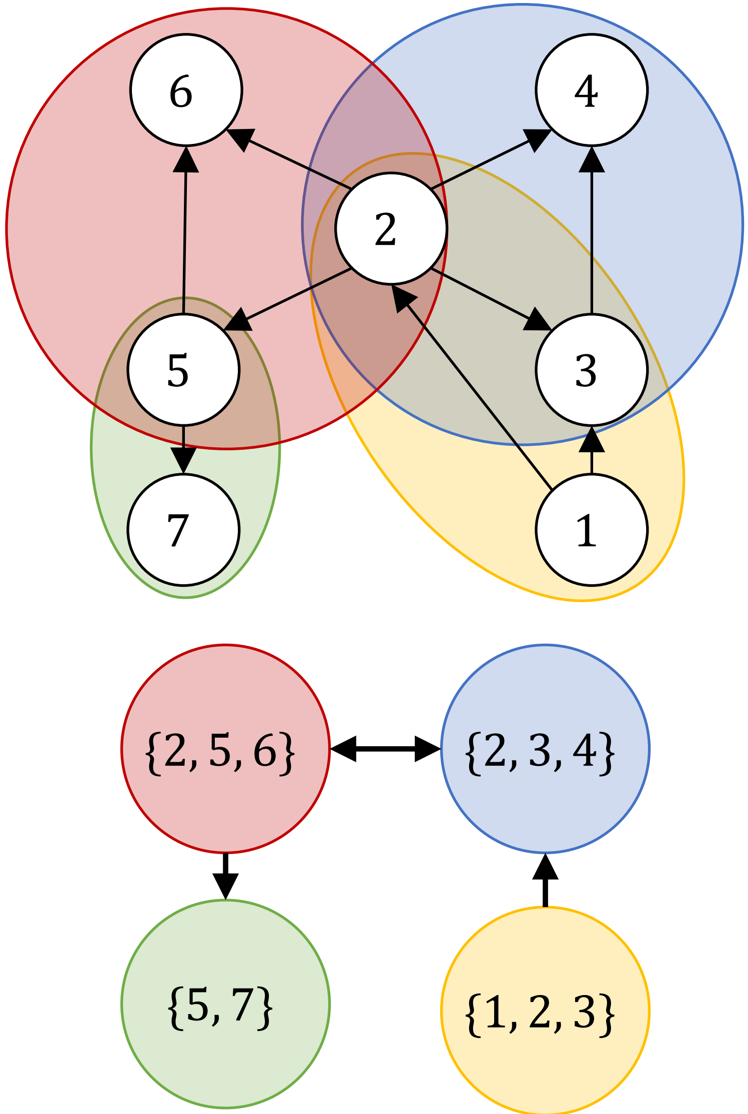


Directed Clique Trees (DCTs)



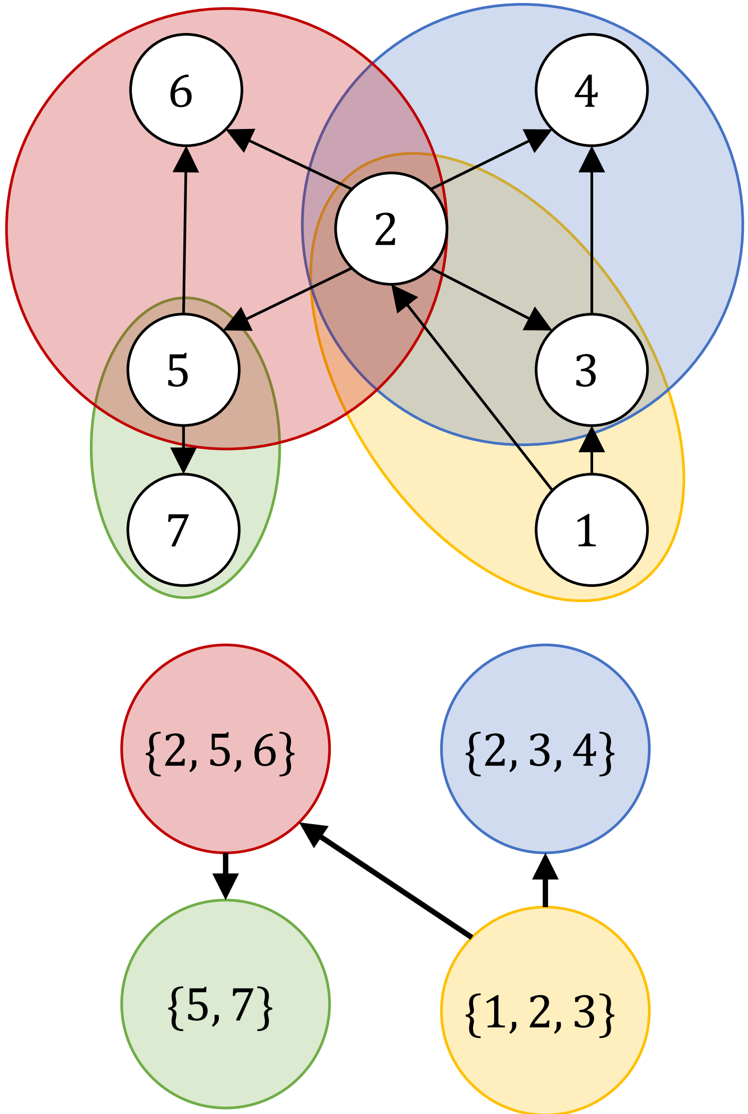
Directed Clique Trees (DCTs)

- Same vertices and adjacencies as a clique tree.
- Given adjacent cliques C_1 and C_2, add an arrowhead at C_1 if:
  - for all $v_{12} \in C_1 \cap C_2$ and $v_2 \in C_2 \setminus C_1$, we have $v_{12} \to v_2$



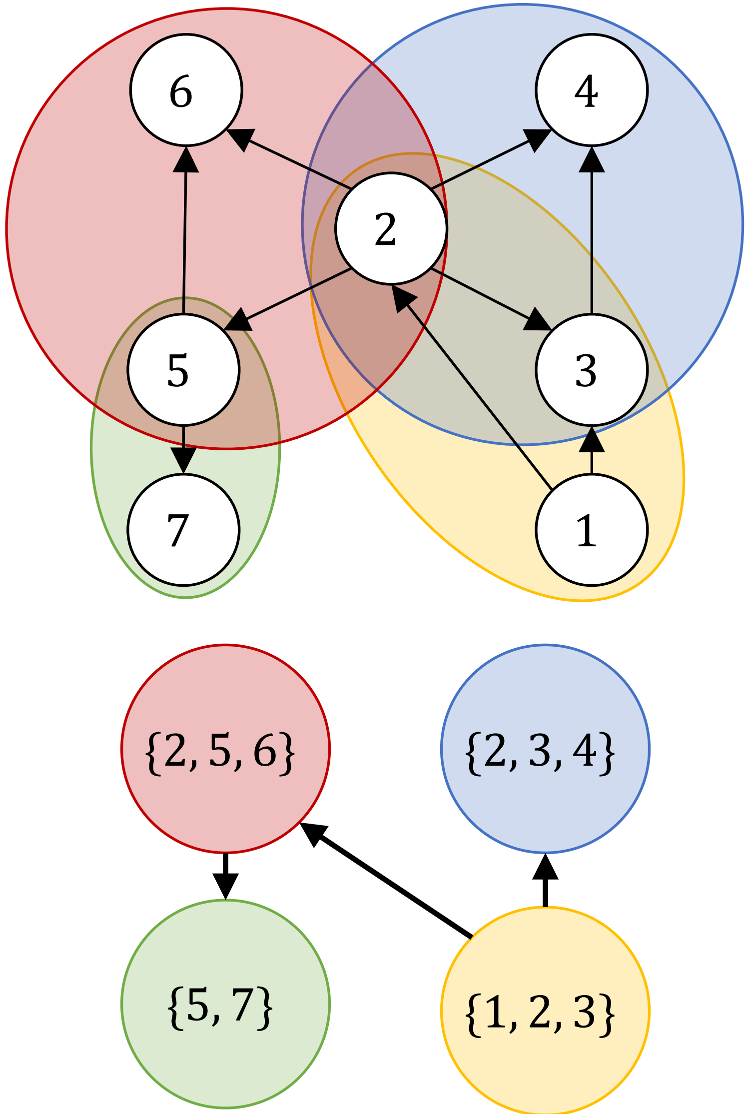
Directed Clique Trees (DCTs)

- There can be multiple DCTs for a single graph.



Directed Clique Trees (DCTs)

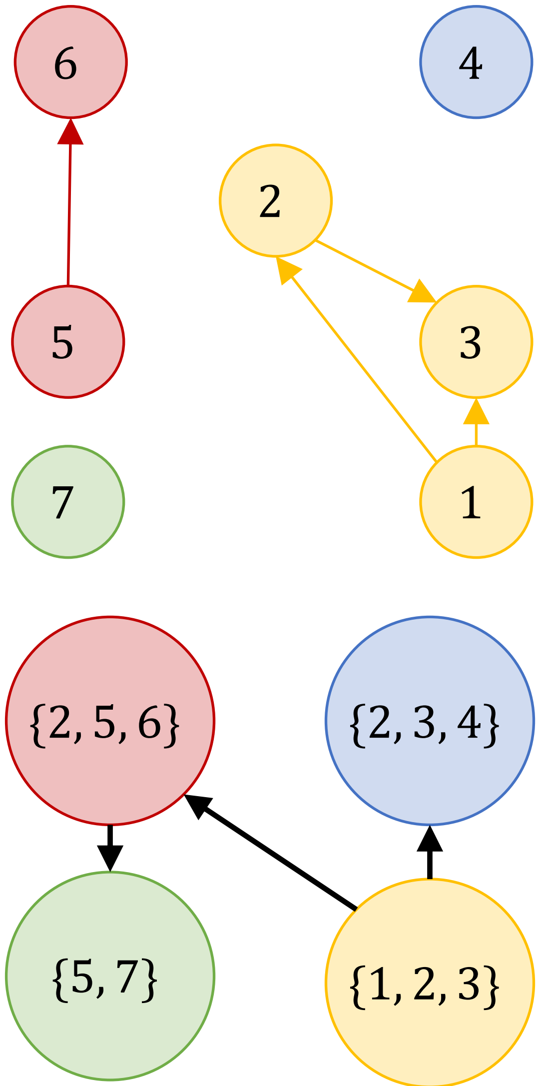
- There can be multiple DCTs for a single graph.
- We can always[1] find a DCT such that each clique has at most one parent (i.e., no colliders).



[1]Requires a **contraction operation** that joins cliques connected by bidirected edges.

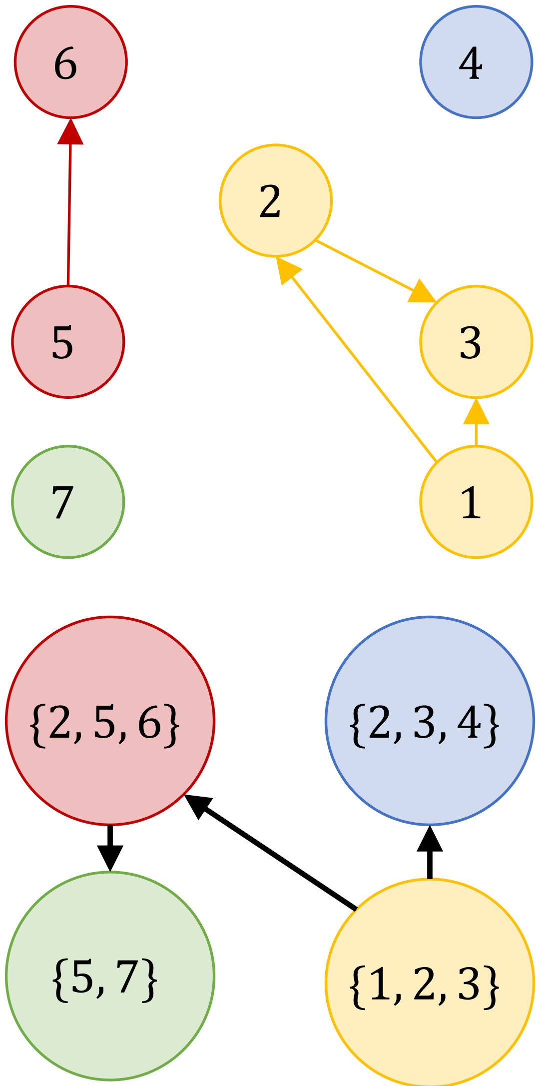
Residuals in DCTs

- The **residual** of a clique is the clique minus its parents.



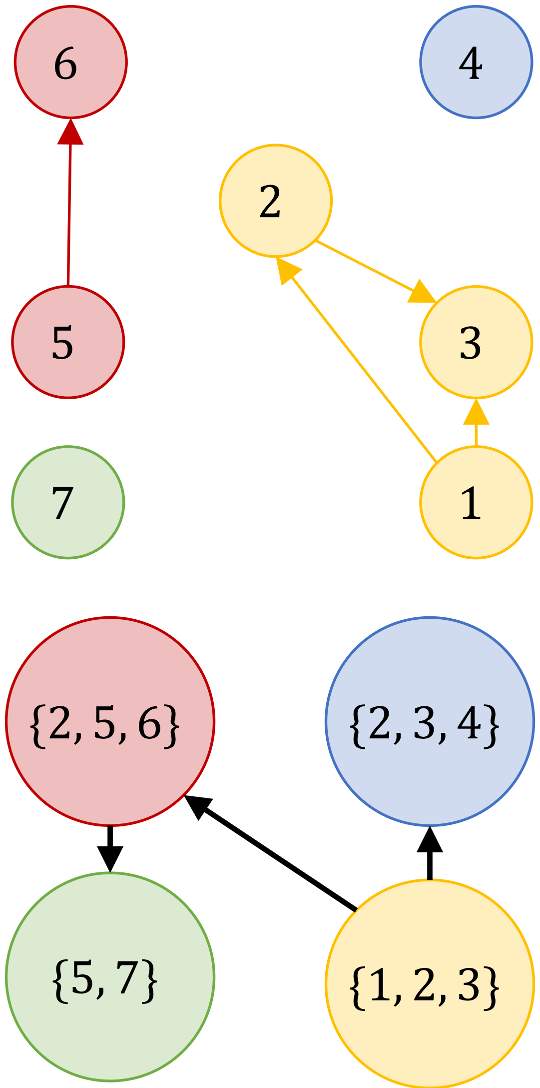
Residuals in DCTs

- The **residual** of a clique is the clique minus its parents.
- Orientations in one residual are independent of the orientations in other residuals.



Residuals in DCTs

- The **residual** of a clique is the clique minus its parents.
- Orientations in one residual are independent of the orientations in other residuals.
- Since the residuals are cliques (or other simple structures when we have to contract bidirected edges), we can efficiently compute the MVIS.



What else?

Instance-wise competitive bounds

- The **instance-wise competitive ratio (ic-ratio)** of an intervention policy on a DAG G is the (expected) number of interventions used by the policy, divided by the size of the MVIS.

Instance-wise competitive bounds

- The **instance-wise competitive ratio (ic-ratio)** of an intervention policy on a DAG G is the (expected) number of interventions used by the policy, divided by the size of the MVIS.
- We develop and analyze the **DCT-policy**, showing that the ic-ratio is logarithmic in the number of cliques for certain types of graphs.

Thanks!