Center for Causal Discovery

Day 2: Search

June 9, 2015

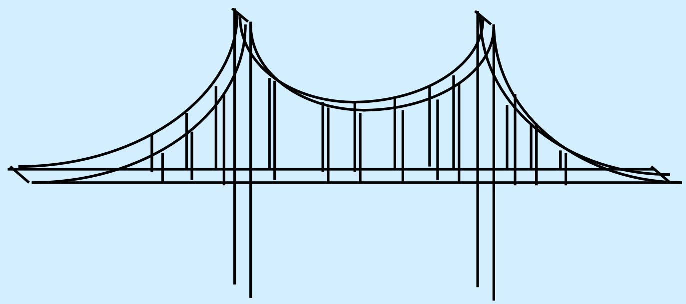
Carnegie Mellon University

Outline

Day 2: Search

1. Bridge Principles: Causation ←→ Probability
2. D-separation
3. Model Equivalence
4. Search Basics (PC, GES)
5. Latent Variable Model Search (FCI)
6. Examples

Causal Structure



Testable Statistical Predictions

Causal Graphs

$$X \rightarrow Y \rightarrow Z$$

e.g., Conditional Independence

$$X \perp\!\!\!\perp Z \mid Y$$

$$\forall x,y,z \;\; P(X = x, Z=z \mid Y=y) =$$

$$P(X = x \mid Y=y)\, P(Z=z \mid Y=y)$$

Bridge Principles:
Acyclic Causal Graph over V $\Rightarrow$ Constraints on P(V)

Weak Causal Markov Assumption

V_1,V_2 causally disconnected $\Rightarrow$ $V_1 \_||\_ V_2$

V_1,V_2 causally disconnected $\Leftrightarrow$

i. V_1 not a cause of V_2, and
ii. V_1 not an effect of V_2, and
iii. No common cause Z of V_1 and V_2

Bridge Principles:
Acyclic Causal Graph over V $\Rightarrow$ Constraints on P(V)

Weak Causal Markov Assumption

V_1, V_2 causally disconnected $\Rightarrow$ $V_1 \perp\!\!\!\perp V_2$

Determinism

(Structural Equations)

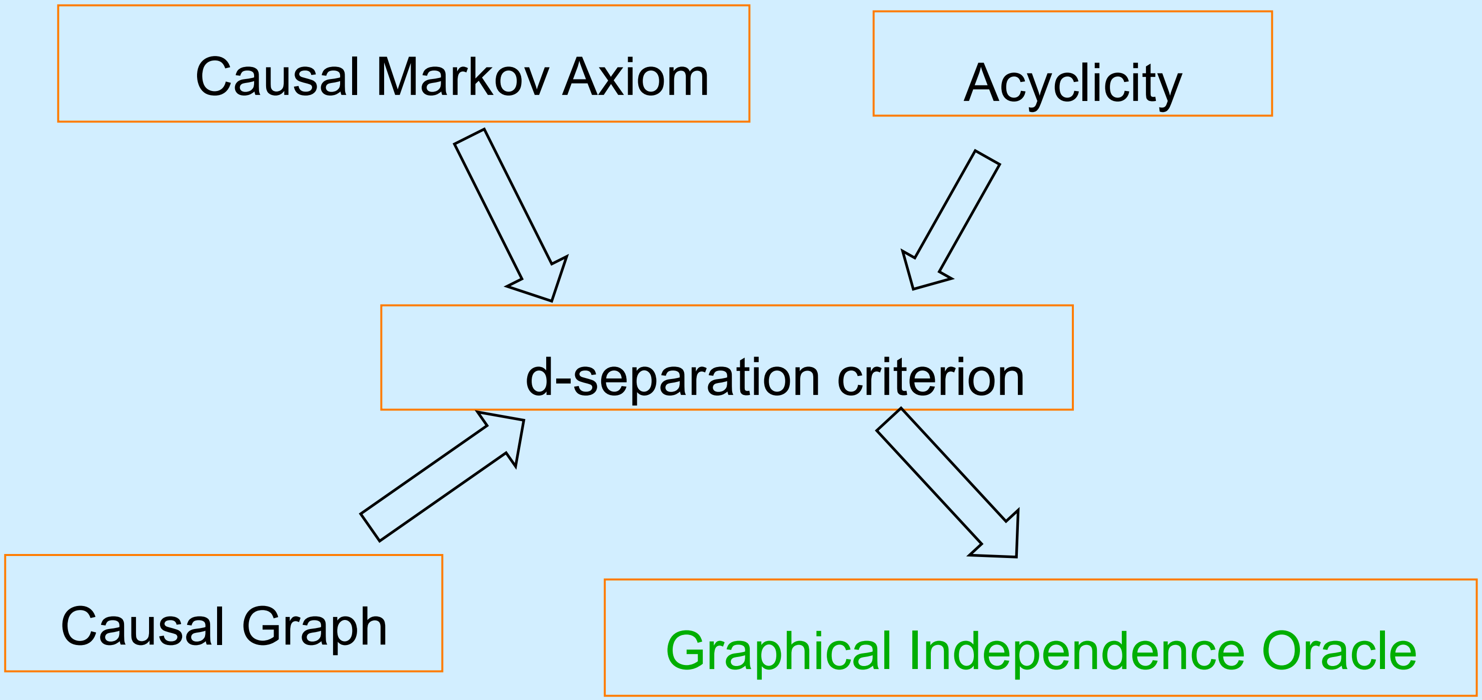
Causal Markov Axiom

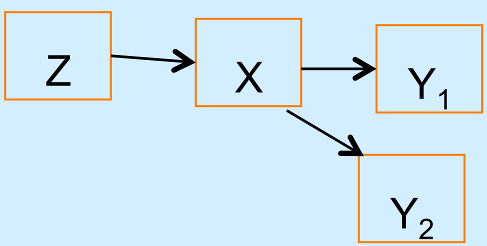
If G is a causal graph, and P a probability distribution over the variables in G, then in <G,P> satisfy the Markov Axiom iff:

every variable V is independent of its non-effects,

conditional on its immediate causes.

Bridge Principles:
Acyclic Causal Graph over V ⇒ Constraints on P(V)



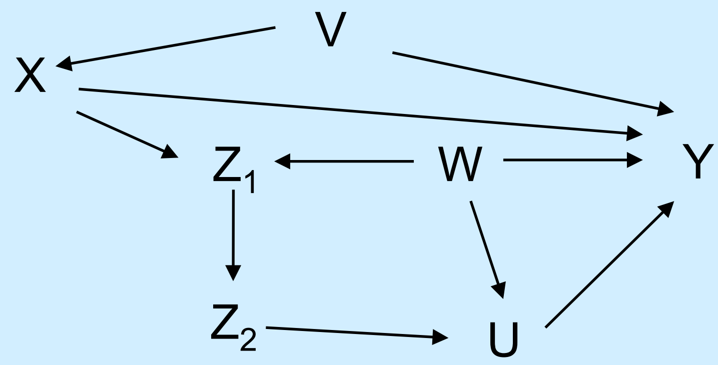


$Z \perp\!\!\!\perp Y_1 \mid X$	$Z \perp\!\!\!\perp Y_2 \mid X$
$Z \perp\!\!\!\perp Y_1 \mid X, Y_2$	$Z \perp\!\!\!\perp Y_2 \mid X, Y_1$
$Y_1 \perp\!\!\!\perp Y_2 \mid X$	$Y_1 \perp\!\!\!\perp Y_2 \mid X, Z$

D-separation

- Undirected Paths
- Colliders vs. Non-Colliders

D-separation: Undirected Paths

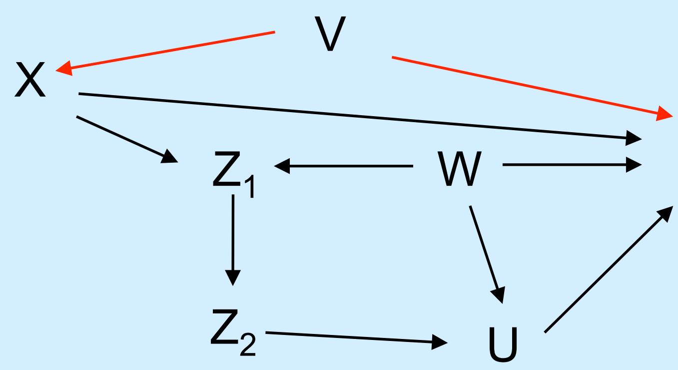


Undirected Path from X to Y:

- any sequence of edges beginning with X and ending at Y in which no edge repeats

Paths from X to Y:

D-separation: Undirected Paths



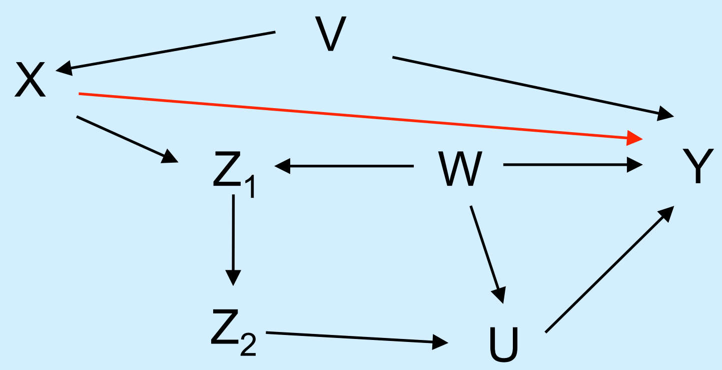
Undirected Path from X to Y:

- any sequence of edges beginning with X and ending at Y in which no edge repeats

Paths from X to Y:

1) X ← V → Y

D-separation: Undirected Paths



Undirected Path from X to Y:

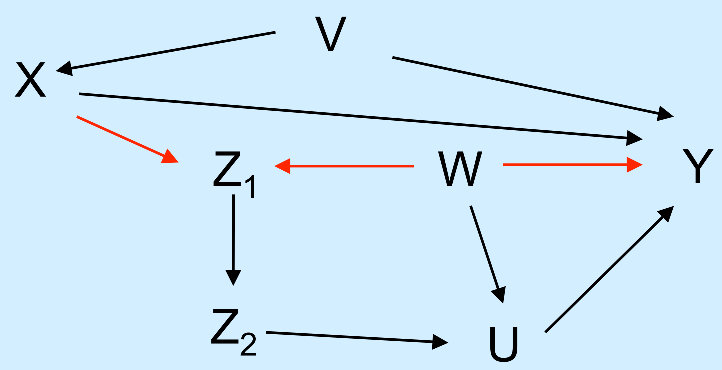
- any sequence of edges beginning with X and ending at Y in which no edge repeats

Paths from X to Y:

1) X ← V → Y

2) X → Y

D-separation: Undirected Paths



Undirected Path from X to Y:

- any sequence of edges beginning with X and ending at Y in which no edge repeats

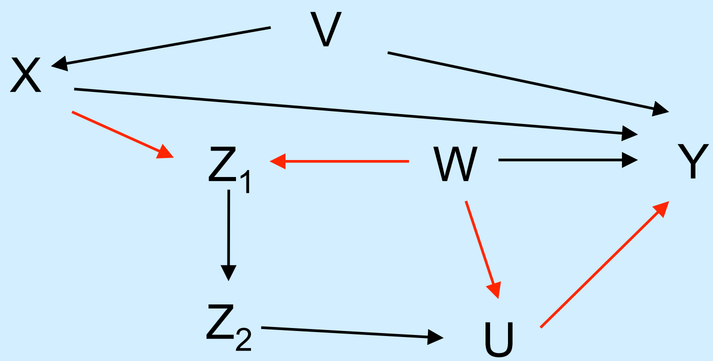
Paths from X to Y:

1) X ← V → Y

2) X → Y

3) X → Z1 ← W → Y

D-separation: Undirected Paths



Undirected Path from X to Y:

- any sequence of edges beginning with X and ending at Y in which no edge repeats

Paths from X to Y:

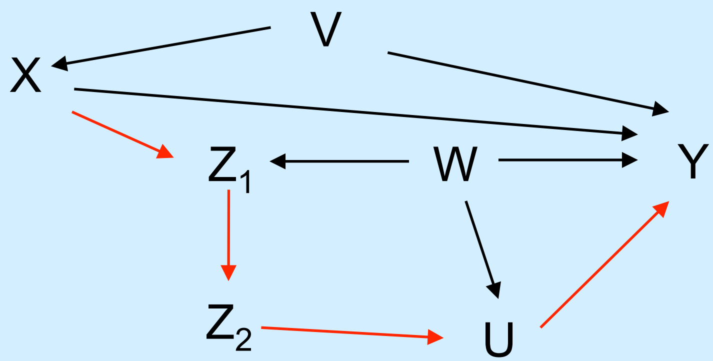
1) X ← V → Y

2) X → Y

3) X → Z1 ← W → Y

4) X → Z1 ← W → U → Y

D-separation: Undirected Paths



Undirected Path from X to Y:

- any sequence of edges beginning with X and ending at Y in which no edge repeats

Paths from X to Y:

1) X ← V → Y

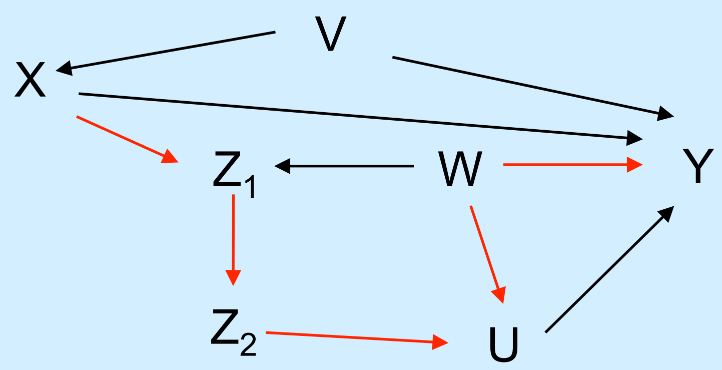
2) X → Y

3) X → Z1 ← W → Y

4) X → Z1 ← W → U → Y

5) X → Z1 → Z2 → U → Y

D-separation: Undirected Paths



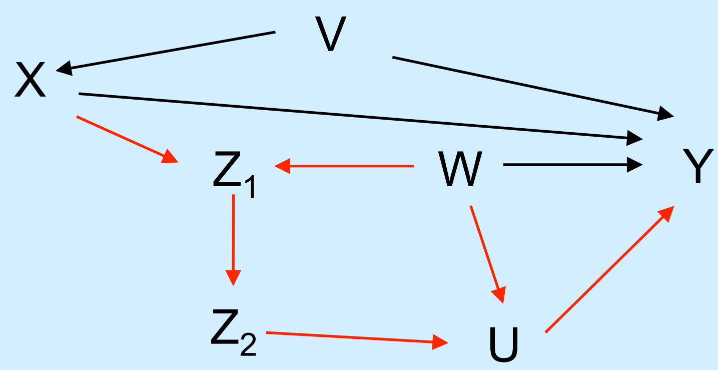
Undirected Path from X to Y:

- any sequence of edges beginning with X and ending at Y in which no edge repeats

Paths from X to Y:

1) X ← V → Y
2) X → Y
3) X → Z1 ← W → Y
4) X → Z1 ← W → U → Y
5) X → Z1 → Z2 → U → Y
6) X → Z1 → Z2 → U ← W → Y

D-separation: Undirected Paths



Undirected Path from X to Y:

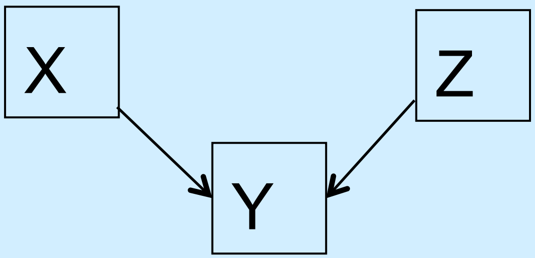
- any sequence of edges beginning with X and ending at Y in which no edge repeats

Illlegal Path from X to Y:

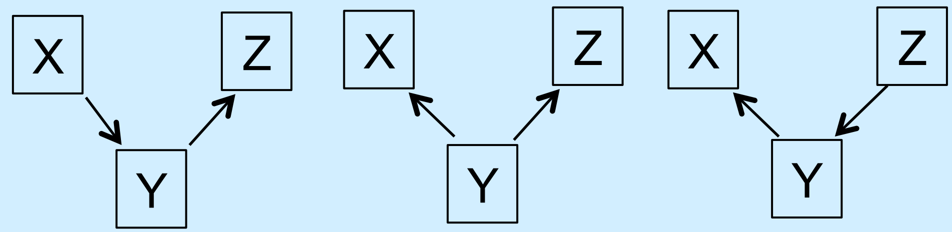
1) X ← $\boxed{\text{Z1} \rightarrow \text{Z2}}$ → U ← W → $\boxed{\text{Z1} \rightarrow \text{Z2}}$ → U → Y

Colliders

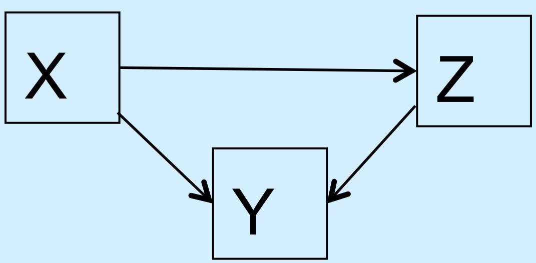
Y: *Collider*



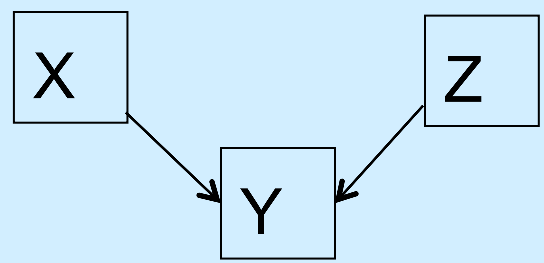
Y: *Non-Collider*



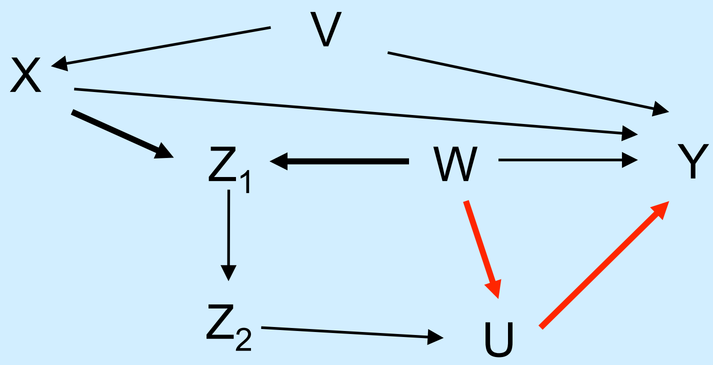
Shielded



Unshielded



A variable is or is not a collider *on a path*



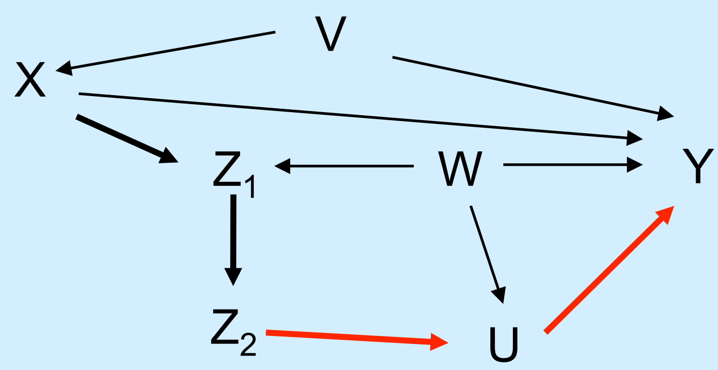
Variable: U

Paths from X to Y

X → Z1 ← W → U → Y

Paths on which U is a non-collider:

Colliders – *a variable on a path*



Variable: U

Paths from X to Y

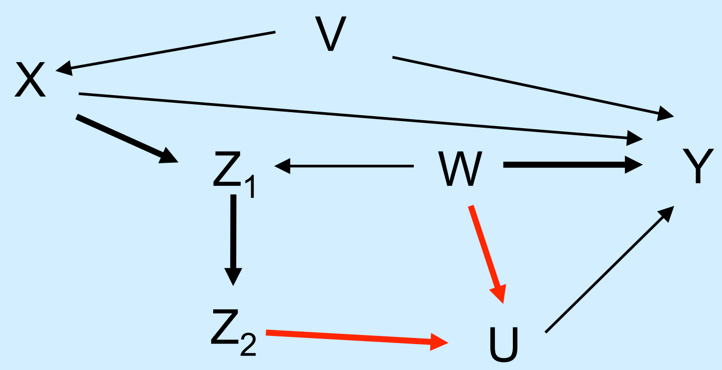
X → Z1 ← W → U → Y

Paths on which U is a non-collider:

X → Z1 → Z2 → U → Y

Path on which U is a collider:

Colliders – *a variable on a path*



Variable: U

Paths from X to Y

X → Z1 ← W → U → Y

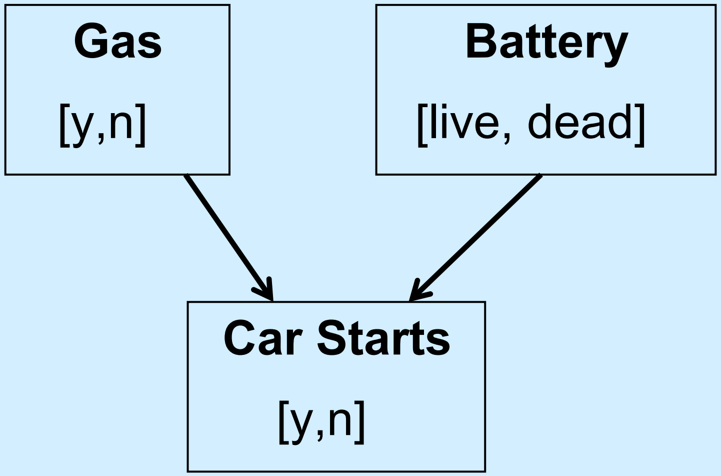
Paths on which U is a non-collider:

X → Z1 → Z2 → U → Y

Path on which U is a collider:

X → Z1 → Z2 → U ← W → Y

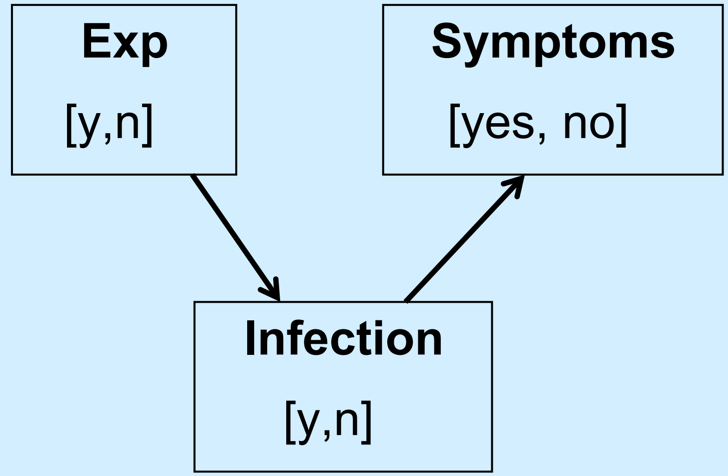
Conditioning on Colliders *induce* Association



Gas _||_ Battery

Gas ~~_||_~~ Battery | Car starts = no

Conditioning on Non-Colliders *screen-off* Association



Exp ~~_||_~~ Symptoms

Exp _||_ Symptoms | Infection

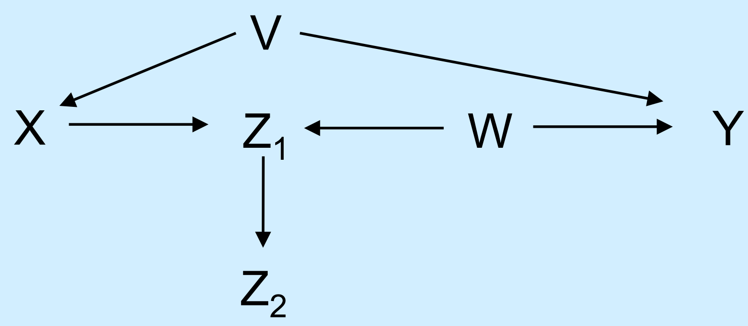
D-separation

X is *d-separated* from Y by **Z** in **G** iff
Every undirected path between X and Y in **G** is *inactive* relative to **Z**

An undirected path is *inactive* relative to **Z** iff
any node on the path is *inactive* relative to **Z**

A node N (on a path) is *inactive* relative to **Z** iff
a) N is a non-collider in Z, or
b) N is a collider that is not in Z, and has no descendant in Z

A node N (on a path) is *active* relative to **Z** iff
a) N is a non-collider not in Z, or
b) N is a collider that is in Z, or has a descendant in Z



X d-sep Y relative to **Z** = ∅ ?

Undirected Paths between X , Y:

1) $X \rightarrow Z_1 \leftarrow W \rightarrow Y$

2) $X \leftarrow V \rightarrow Y$

D-separation

X is *d-separated* from Y by **Z** in **G** iff
Every undirected path between X and Y in **G** is *inactive* relative to **Z**

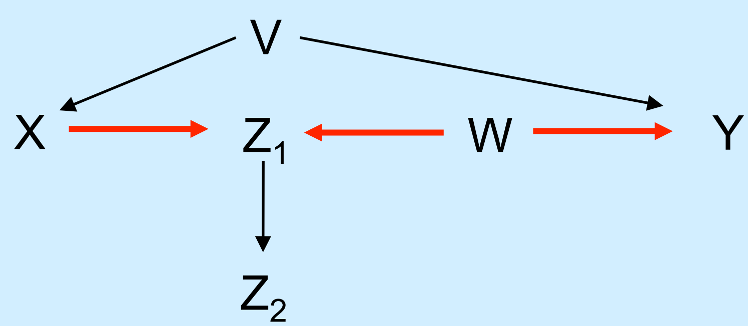
An undirected path is *inactive* relative to **Z** iff
any node on the path is *inactive* relative to **Z**

A node N (on a path) is *inactive* relative to **Z** iff

a) N is a non-collider in Z, or
b) N is a collider that is not in Z, and has no descendant in Z

A node N (on a path) is *active* relative to **Z** iff

a) N is a non-collider not in Z, or
b) N is a collider that is in Z, or has a descendant in Z



X d-sep Y relative to **Z** = ∅ ?

$X \rightarrow Z_1 \leftarrow W \rightarrow Y$ active? No

1) Z1 active? No
2) W active? Yes

D-separation

X is *d-separated* from Y by **Z** in **G** iff
Every undirected path between X and Y in **G** is *inactive* relative to **Z**

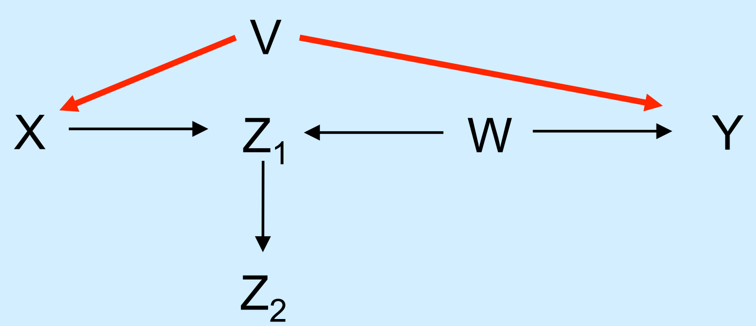
An undirected path is *inactive* relative to **Z** iff
any node on the path is *inactive* relative to **Z**

A node N (on a path) is *inactive* relative to **Z** iff

a) N is a non-collider in Z, or
b) N is a collider that is not in Z,
 and has no descendant in Z

A node N (on a path) is *active* relative to **Z** iff

a) N is a non-collider not in Z, or
b) N is a collider that is in Z,
 or has a descendant in Z



X d-sep Y relative to **Z** = ∅ ? No

X ← V → Y active? Yes

1) V active? Yes

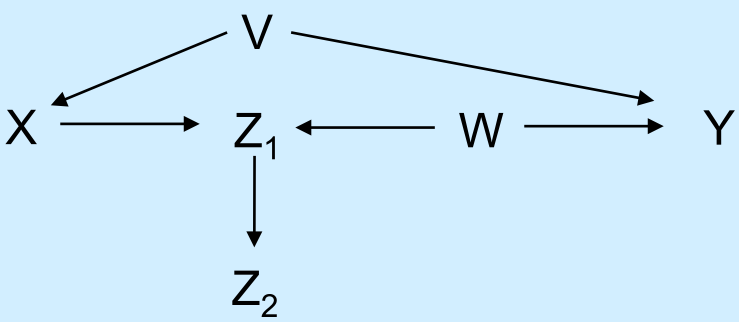
D-separation

X is *d-separated* from Y by **Z** in **G** iff
Every undirected path between X and Y in **G** is inactive relative to **Z**

An undirected path is inactive relative to **Z** iff
any node on the path is inactive relative to **Z**

A node N is inactive relative to **Z** iff
a) N is a non-collider in Z, or
b) N is a collider that is not in Z,
and has no descendant in Z

A node N (on a path) is *active* relative to **Z** iff
a) N is a non-collider not in Z, or
b) N is a collider that is in Z,
or has a descendant in Z



X d-sep Y relative to **Z** = $\{W, Z_2\}$?

Undirected Paths between X , Y:

1) $X \rightarrow Z_1 \leftarrow W \rightarrow Y$

2) $X \leftarrow V \rightarrow Y$

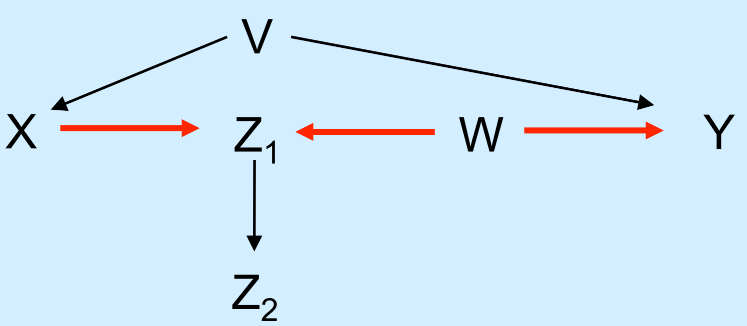
D-separation

X is *d-separated* from Y by **Z** in **G** iff
Every undirected path between X and Y in **G** is inactive relative to **Z**

An undirected path is inactive relative to **Z** iff
any node on the path is inactive relative to **Z**

A node N is inactive relative to **Z** iff
a) N is a non-collider in Z, or
b) N is a collider that is not in Z,
and has no descendant in Z

A node N (on a path) is *active* relative to **Z** iff
a) N is a non-collider not in Z, or
b) N is a collider that is in Z,
or has a descendant in Z



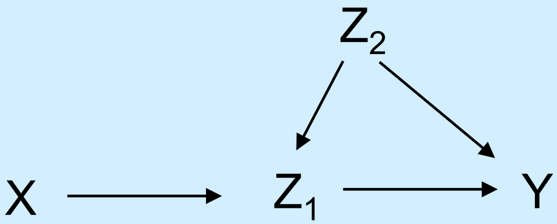
X d-sep Y relative to **Z** = {W, Z_2 } ? No

1) $X \rightarrow Z_1 \leftarrow W \rightarrow Y$

Z1 active? Yes

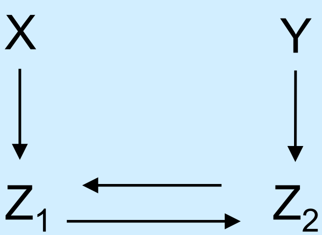
W active? No

D-separation



X d-sep Y given $\varnothing$? No

X d-sep Y given $\{Z_1\}$? No



X d-sep Z_2 given $\varnothing$? No

X d-sep Z_2 given $\{Z_1\}$? No

D-separation + Intervention:

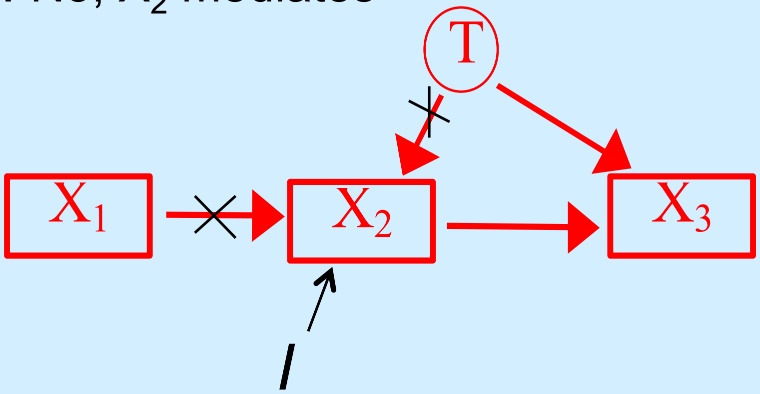
Statistical Control $\neq$ Experimental Control

Question: Does X_1 directly cause X_3?

How to find out?

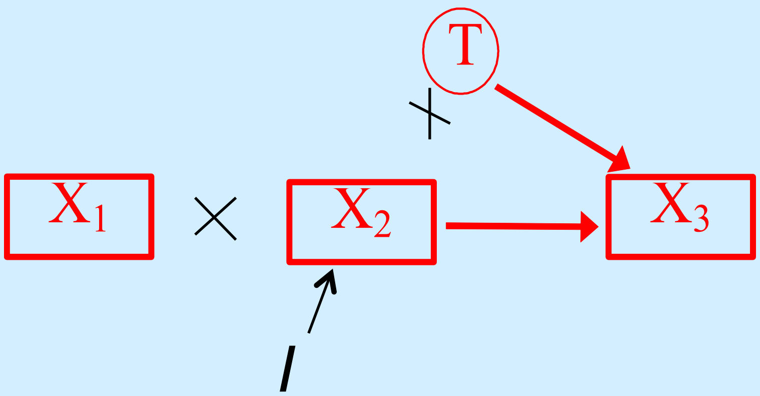
Truth: No, X_2 mediates

Experimentally control for X_2



D-separation + Intervention:

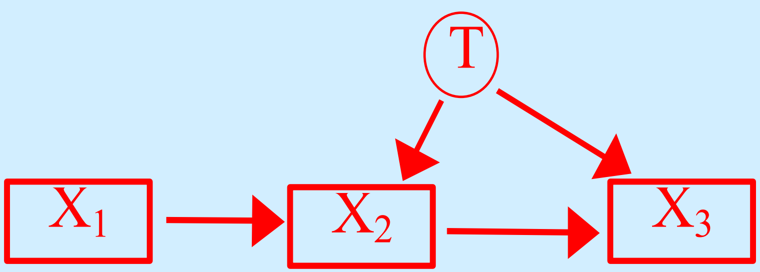
Statistical Control $\neq$ Experimental Control



Experimentally control for X_2

X_3 d-sep X_1 by {X_2set} ???

Yes: X_3 _||_ X_1 | X_2(set)



Statistically control for X_2

X_3 d-sep X_1 by {X_2} ???

No! X_3 ~~_||_~~ X_1 | X_2

Break

Equivalence Classes

Equivalence:

- Independence Equivalence: $M_1 \models (X \perp\!\!\!\perp Y \mid Z) \Leftrightarrow M_2 \models (X \perp\!\!\!\perp Y \mid Z)$
- Distribution Equivalence: $\forall \theta_1 \, \exists \theta_2 \; M_1(\theta_1) = M_2(\theta_2)$, and vice versa)

- Independence (d-separation equivalence)
  - DAGs : Patterns
  - PAGs : Partial Ancestral Graphs
  - Intervention Equivalence Classes
- Measurement Model Equivalence Classes
- Linear Non-Gaussian Model Equivalence Classes
- Etc.

d-separation/Independence Equivalence

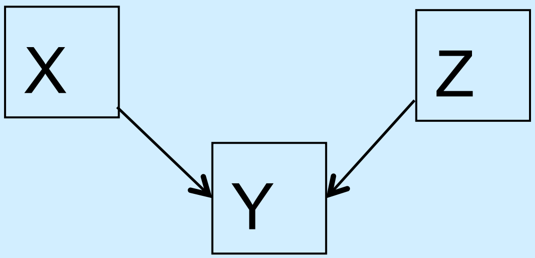
D-separation Equivalence Theorem (Verma and Pearl, 1988)

Two acyclic graphs over the same set of variables are d-separation equivalent iff they have:

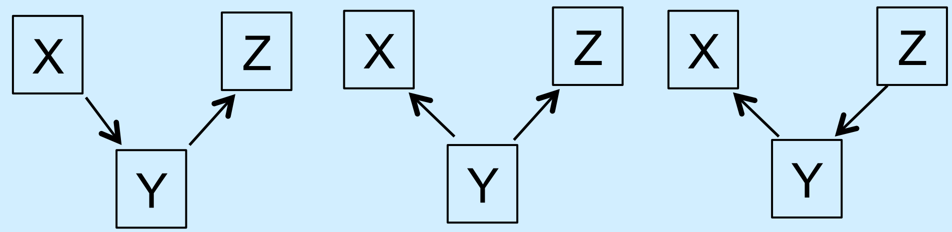
- the same adjacencies
- the same unshielded colliders

Colliders

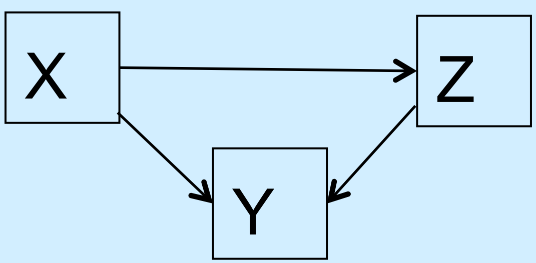
Y: *Collider*



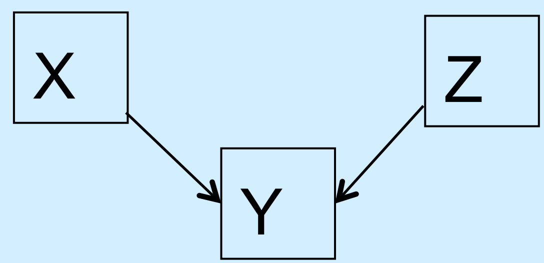
Y: *Non-Collider*



Shielded



Unshielded



d-separation/Independence Equivalence

D-separation Equivalence Theorem (Verma and Pearl, 1988)

Two acyclic graphs over the same set of variables are d-separation equivalent iff they have:

- the same adjacencies
- the same unshielded colliders

Exercises

1) Create a 4-variable DAG
2) Specify a 1-edge variant that is equivalent
3) Specify a 1-edge variant that is not
4) Show with IM and Estimators that you have succeeded in steps 2 and 3

Independence Equivalence Classes: Patterns & PAGs

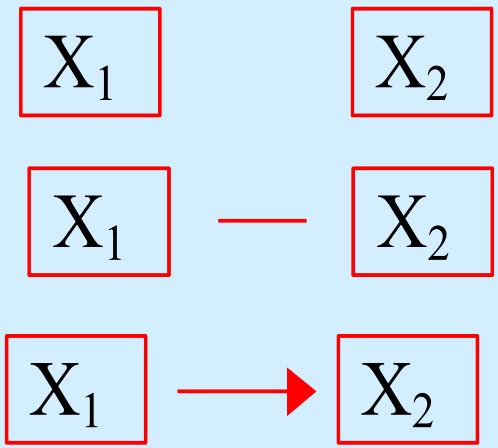
- Patterns (Verma and Pearl, 1990): graphical representation of d-separation equivalence class

(among models with no latent common causes)

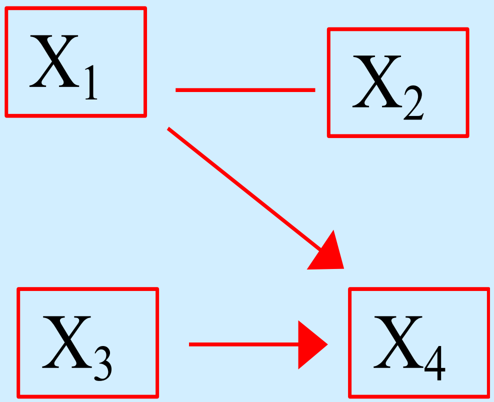
- PAGs: (Richardson 1994) graphical representation of a d-separation equivalence class that includes models with *latent common causes* and *sample selection bias* that are d-separation equivalent over a set of measured variables **X**

Patterns

Possible Edges



Example



Patterns: What the Edges Mean

$\boxed{X_1} \qquad \boxed{X_2}$

X_1 and X_2 are not adjacent in any member of the equivalence class

---

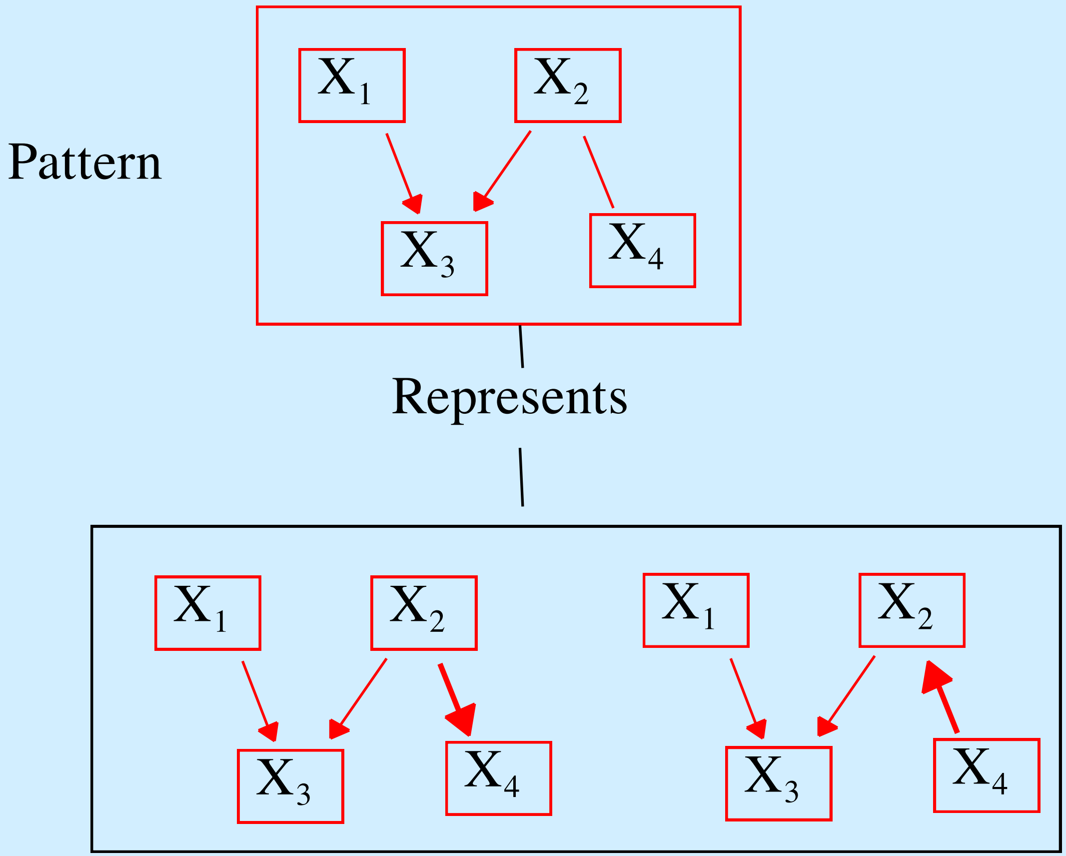
$\boxed{X_1} \longrightarrow \boxed{X_2}$

$X_1 \rightarrow X_2$ (X_1 is a cause of X_2) in every member of the equivalence class.

$\boxed{X_1} \; — \; \boxed{X_2}$

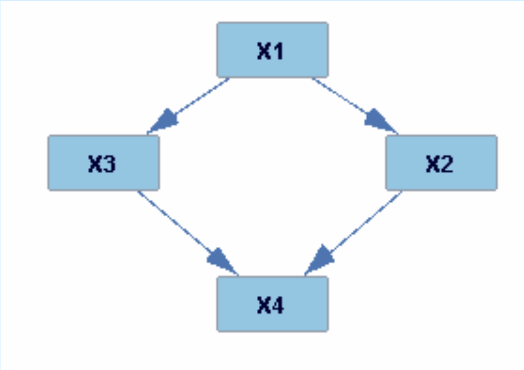
$X_1 \rightarrow X_2$ in some members of the equivalence class, and $X_2 \rightarrow X_1$ in others.

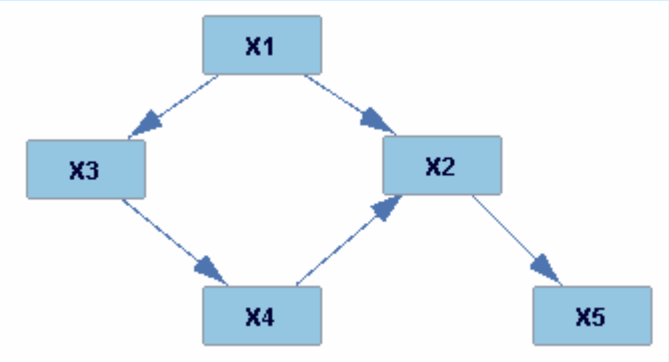
Patterns



Patterns

Specify all the causal graphs represented by the Pattern:

1) 

2) 

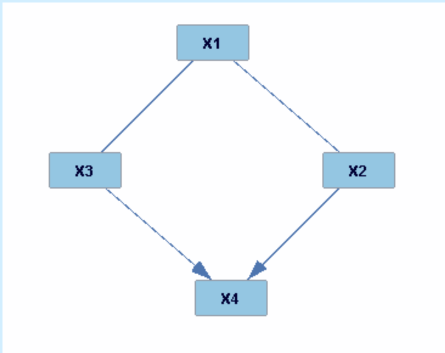
??

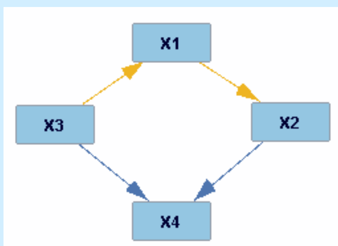
??

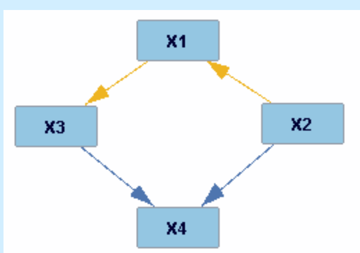
Patterns

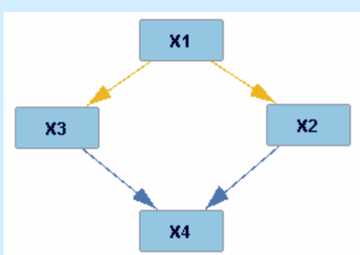
Specify all the causal graphs represented by the Pattern:

1)

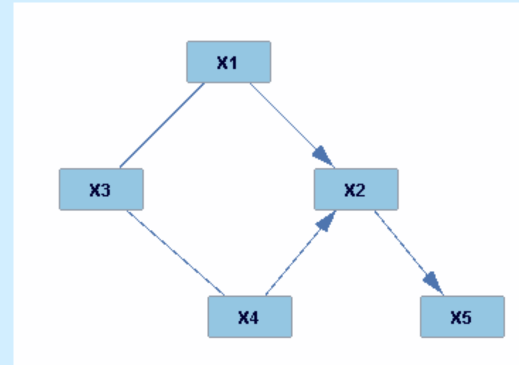


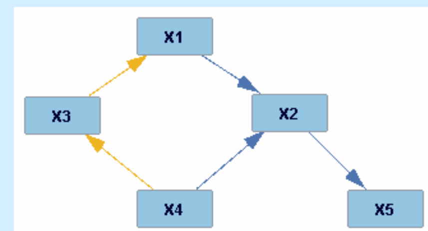


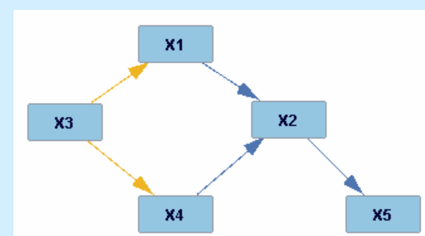


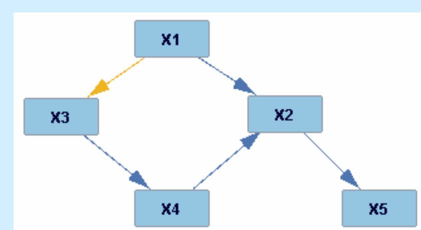


2)

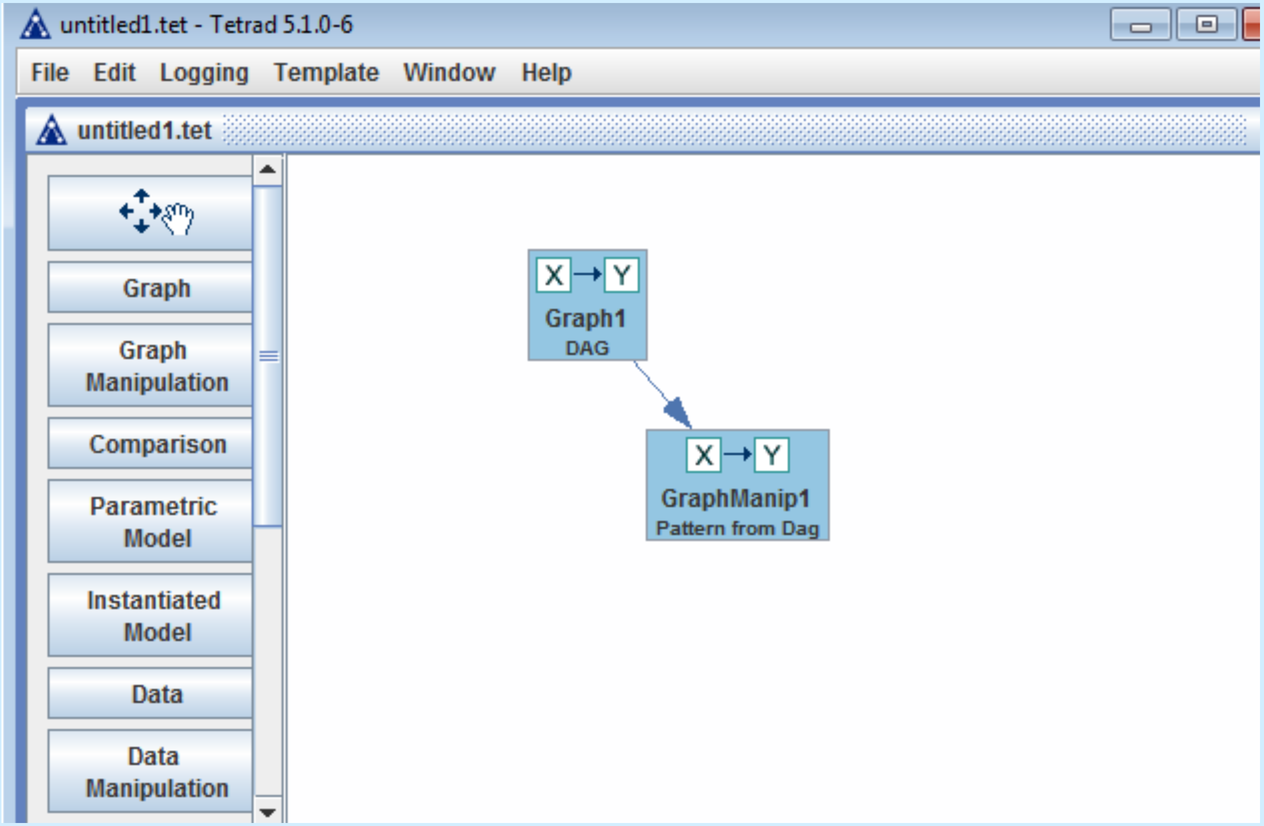








Tetrad Demo: Generating Patterns



Break

Causal Search Spaces are Large

- Directed Acyclic Graphs (between $2^{\binom{N}{2}}$ and $3^{\binom{N}{2}}$) …

  $\binom{N}{2}$ is $O(N^2)$
- Directed Graphs ($4^{\binom{N}{2}}$)
- Markov Equivalence Class of DAGs (patterns) : DAGs / 3.7
- Markov Equivalence Class of DAGs with confounders (roughly PAGs) ??
- Equivalence Class of "Linear Measurement Models" ??
- Equivalence Class of Directed Graphs with confounders
  - Relative to: Experimental Setup ***V*** = *{**Obs**, **Manip**}* *??*

Causal Search as a Method

Causal Knowledge
e.g.,
Markov Equivalence Class of Causal Graphs

*Experimental Setup(**V**)*

- ***V** = {**Obs**, **Manip**}*
- *P(**Manip**)*

Discovery Algorithm

$P_{Manip}(V)$

Statistical Inference

Data

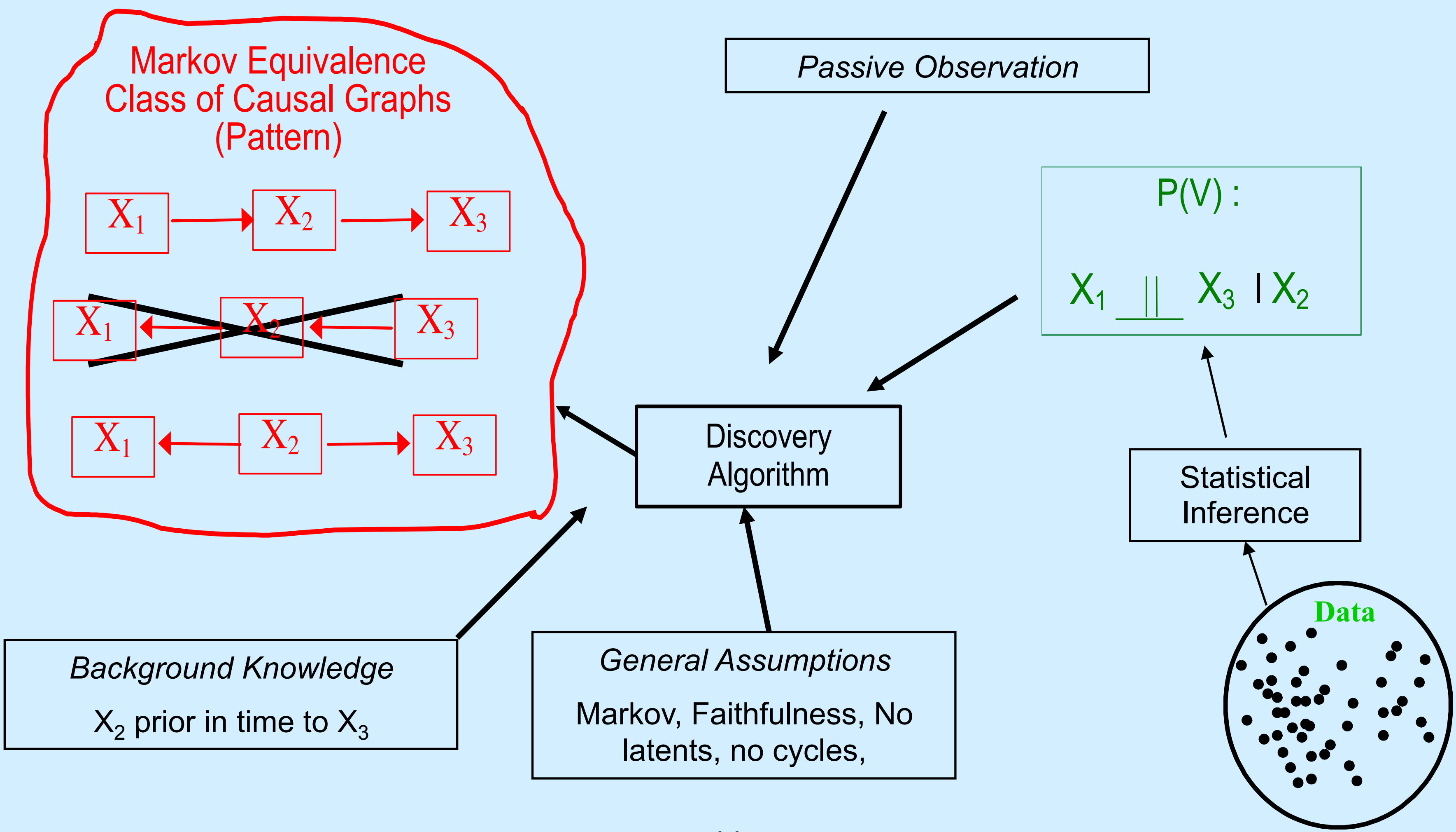
Background Knowledge

- Salary $\nrightarrow$ Gender
- Infection → Symptoms

General Assumptions

- Markov,
- Faithfulness
- Linearity
- Gaussianity
- Acyclicity

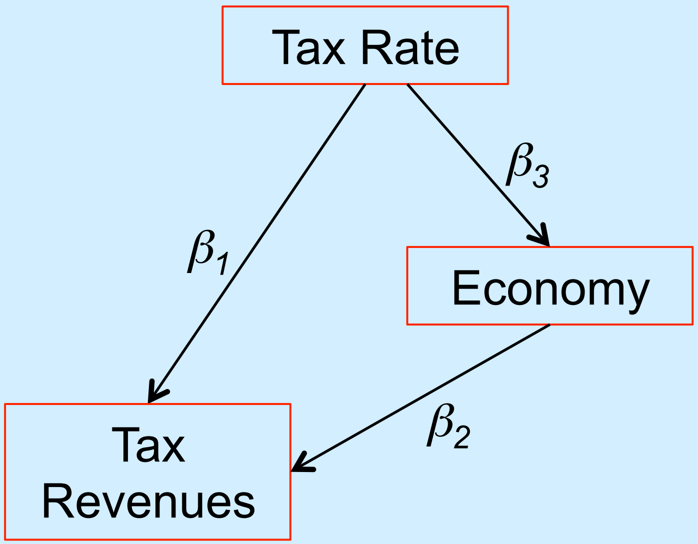
For Example



Markov Equivalence
Class of Causal Graphs
(Pattern)
X_1 → X_2 → X_3
X_1 ← X_2 ← X_3
X_1 ← X_2 → X_3
Passive Observation
P(V) :
$X_1 \perp\!\!\!\perp X_3 \mid X_2$
Discovery Algorithm
Statistical Inference
Data
Background Knowledge
X_2 prior in time to X_3
General Assumptions
Markov, Faithfulness, No latents, no cycles,

Faithfulness

Constraints on a probability distribution P generated by a causal structure G hold for all parameterizations of G.



Revenues $:= \beta_1 Rate + \beta_2 Economy + \varepsilon_{Rev}$

Economy $:= \beta_3 Rate + \varepsilon_{Econ}$

Faithfulness:

$$\beta_1 \neq -\beta_3\beta_2$$

$$\beta_2 \neq -\beta_3\beta_1$$

Faithfulness

Constraints on a probability distribution P generated by a causal structure G hold for all parameterizations of G.

All and only the constraints that hold in P(V) are entailed by the causal structure G(V), rather than lower dimensional surfaces in the parameter space.

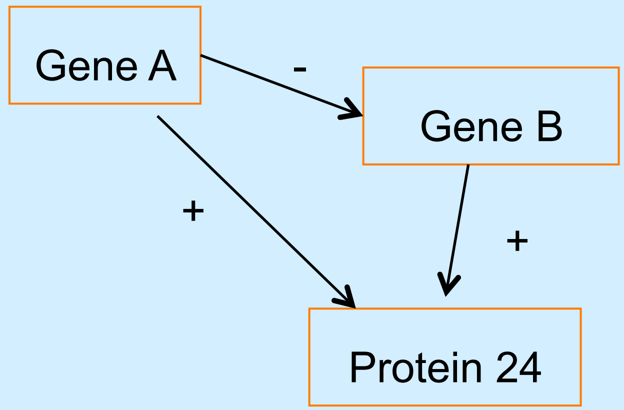
Causal Markov Axiom:

X and Y causally disconnected $\models$ X _||_ Y

Faithfulness:

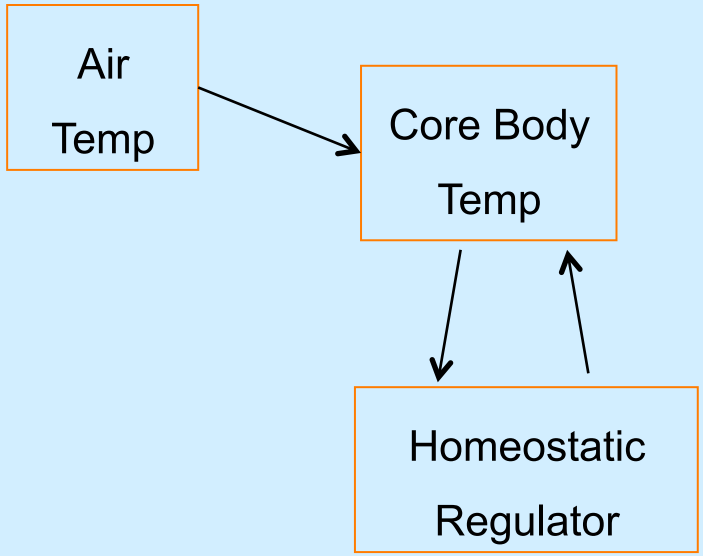
X and Y causally disconnected $=\!|$ X _||_ Y

Challenges to Faithfulness



By evolutionary design:

Gene A _||_ Protein 24



By evolutionary design:

Air temp _||_ Core Body Temp

Sampling Rate vs. Equilibration rate

Search Methods

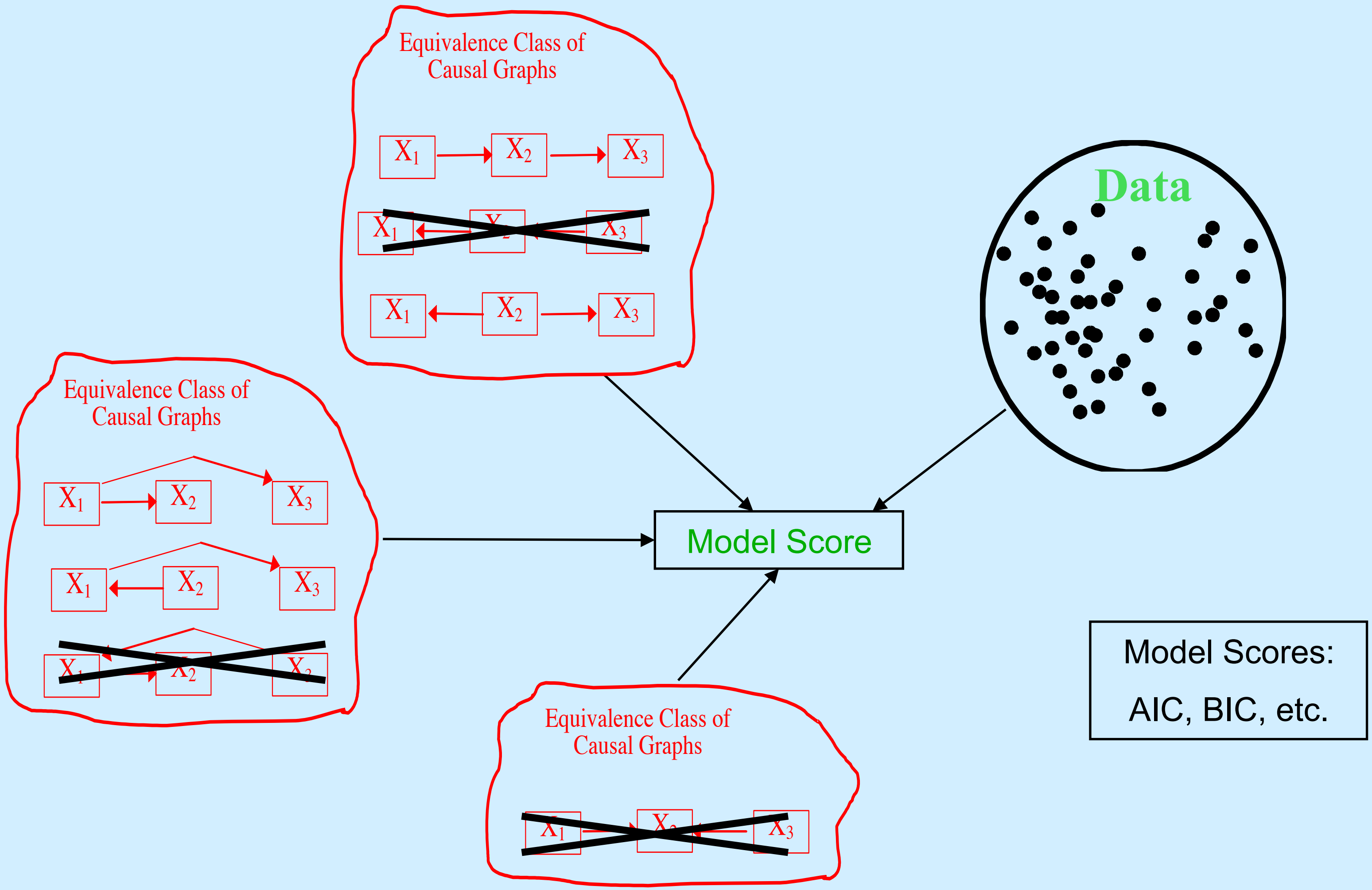
- Constraint Based Searches
  - PC, FCI
  - Pointwise, but not uniformly consistent

- Scoring Searches
  - Scores: BIC, AIC, etc.
  - Search: Hill Climb, Genetic Alg., Simulated Annealing
  - Difficult to extend to latent variable models
  - Meek and Chickering Greedy Equivalence Class (GES)
  - Pointwise, but not uniformly consistent

- Latent Variable Psychometric Model Search
  - BPC, MIMbuild, etc.

- Linear non-Gaussian models (Lingam)
- Models with cycles
- And more!!!

Score Based Search



Background Knowledge

e.g., X_2 prior in time to X_3