

# Causal Directed Acyclic Graphs

Kosuke Imai

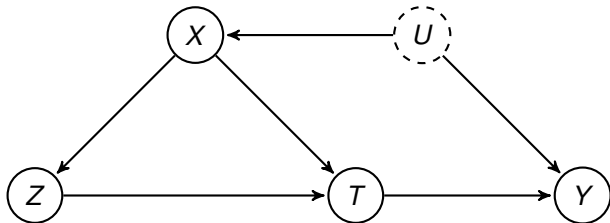
Harvard University

Spring 2021

# Elements of DAGs (Pearl, 2000. *Causality*. Cambridge UP)

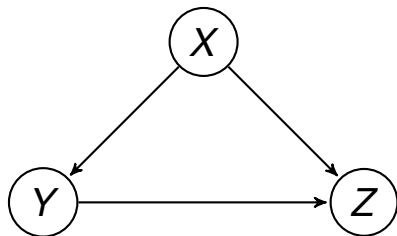
- $\mathcal{G} = (E, V)$

- 1  $V$ : nodes or vertices  $\rightsquigarrow$  variables (observed and unobserved)
- 2  $E$ : directed arrows  $\rightsquigarrow$  *possibly non-zero* direct causal effects



- **Acyclic**: no cycle, no simultaneity
- Encoded assumptions
  - Absence of variables: no other common causes (observed and unobserved) of any pair of variables
  - Absence of arrows: zero causal effect

## DAG Terminology

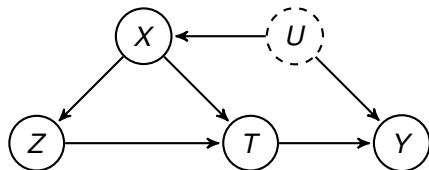


- chain:  $X \rightarrow Y \rightarrow Z$
- fork:  $Y \leftarrow X \rightarrow Z$
- inverted fork:  $X \rightarrow Z \leftarrow Y$

- Parents (Children): directly causing (caused by) a vertex  $i \rightarrow j$
- Ancestors (Descendants): directly or indirectly causing (caused by) a vertex  $i \rightarrow \dots \rightarrow j$
- **Path**: an acyclic sequence of adjacent nodes
  - causal path: all arrows pointing out of  $i$  and into  $j$
  - non-causal path: some arrows going against causal order
- **Collider**: a vertex on a path with two incoming arrows

# Nonparametric Structural Equation Models (NPSEM)

- Equivalence to the nonparametric structural equation models:



$$Y = f_1(T, U, \epsilon_1)$$

$$T = f_2(X, Z, \epsilon_2)$$

$$Z = f_3(X, \epsilon_3)$$

$$X = f_4(U, \epsilon_4)$$

- NPSEM allows for
  - causal interpretation
  - any functional form
  - any form of heterogeneous and interaction effects
  - LSEM as a special case
- Factorized likelihood function:

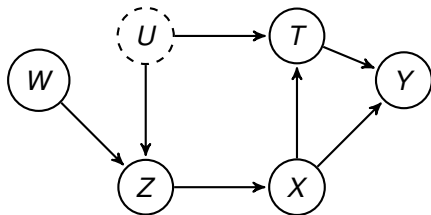
$$P(X_1, X_2, \dots, X_J) = \prod_{j=1}^J P(X_j \mid \text{pa}(X_j))$$

$$P(T, U, X, Y, Z) = P(Y \mid T, U)P(T \mid X, Z)P(Z \mid X)P(X \mid U)P(U)$$

# D-separation

- Does the **conditional independence** relation,  $A \perp\!\!\!\perp B \mid C$ , hold where  $A, B, C$  are sets of vertices?
  - 1 Identify all paths from any vertex in  $A$  to any vertex in  $B$
  - 2 Check if each path is **blocked**
  - 3 If all paths are blocked, then  $A$  is  **$d$ -separated** from  $B$  by  $C$
- Path is blocked if at least one of the following conditions holds
  - 1 it includes a noncollider vertex that is in  $C$
  - 2 it includes a collider that is not in  $C$  and no descendant of any collider is in  $C$
- If  $A$  and  $B$  are  $d$ -separated,  $A \perp\!\!\!\perp B \mid C$  holds
- If  $A$  and  $B$  are  **$d$ -connected** (i.e., not  $d$ -separated),  $A \not\perp\!\!\!\perp B \mid C$  in at least one distribution compatible with the DAG

## D-separation Example



- 1 Are  $W$  and  $Y$  marginally independent of each other?
  - blocked:  $W \rightarrow Z \leftarrow U \rightarrow T \rightarrow Y$     $W \rightarrow Z \leftarrow U \rightarrow T \leftarrow X \rightarrow Y$
  - unblocked:  $W \rightarrow Z \rightarrow X \rightarrow Y$     $W \rightarrow Z \rightarrow X \rightarrow T \rightarrow Y$
- 2 Can we make  $W$  and  $Y$  independent by conditioning?
  - block the unblocked paths without unblocking the blocked paths
  - conditioning on  $X$  would block the first path because  $X$  is a descendant of collider  $Z$
  - conditioning on  $T$  and/or  $Z$  would unblock some of the blocked paths because they are colliders

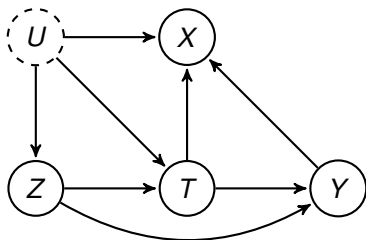
# Backdoor Criterion

- Can we nonparametrically identify the causal effect of  $T$  on  $Y$  given a set of variables  $X$ ?
- Backdoor criterion for  $X$ :
  - 1 No vertex in  $X$  is a decendent of  $T$  (no post-treatment bias), and
  - 2  $X$  blocks all paths between  $T$  and  $Y$  with an incoming arrow into  $T$  (backdoor paths)
- Idea: block all non-causal paths
- Estimation:

$$P(Y(t)) = \sum_x P(Y | T = t, X = x)P(X = x)$$

- **Confounder selection criterion** (VanderWeele and Shpitser. 2011. *Biometrics*)  
If there exist a set of observed covariates that meet the backdoor criterion, it is sufficient to condition on all observed pretreatment covariates that either cause treatment, outcome, or both.

## Example of Backdoor Criterion



- Can we identify the causal effect of  $T$  on  $Y$ ?
- Backdoor paths:
  - 1  $T \leftarrow Z \rightarrow Y$  and  $T \leftarrow U \rightarrow Z \rightarrow Y$
  - 2  $T \leftarrow Z \leftarrow U \rightarrow X \leftarrow Y$
  - 3  $T \leftarrow U \rightarrow X \leftarrow Y$
- Block (1) by conditioning on  $Z$ , which will not unblock (2) and (3)

$$P(Y(t)) = \sum_z P(Y | T = t, Z = z)P(Z = z)$$

# DAGitty (<http://dagitty.net/>)

Diagram style	Model	Examples	How to ...	Layout	Help	Causal effect identification
<input checked="" type="checkbox"/> classic <input type="checkbox"/> SEM-like						Adjustment (total effect) ▾ No adjustment is necessary to estimate the total effect of E on D.
<input checked="" type="checkbox"/> View mode <input checked="" type="radio"/> normal <input type="radio"/> moral graph <input type="radio"/> correlation graph						<input checked="" type="checkbox"/> Testable implications The model implies the following conditional independences: <ul style="list-style-type: none"><li>• <math>A \perp B</math></li><li>• <math>A \perp D \mid E</math></li><li>• <math>B \perp E</math></li><li>• <math>D \perp Z \mid A, B</math></li><li>• <math>D \perp Z \mid B, E</math></li><li>• <math>E \perp Z \mid A</math></li></ul> <a href="#">Export R code</a>
<input checked="" type="checkbox"/> Coloring <input checked="" type="checkbox"/> causal paths <input checked="" type="checkbox"/> biasing paths <input checked="" type="checkbox"/> ancestral structure						<input checked="" type="checkbox"/> Model code
<input checked="" type="checkbox"/> Effect analysis <input type="checkbox"/> atomic direct effects						A 1 0 -2.200, -1.520 B 1 0 1.400, -1.460 D 0 0 1.400, 1.621 E E 0 -2.200, 1.597 Z 1 0 -0.300, -0.082
<input checked="" type="checkbox"/> Legend						A E Z 0 -0.791, -1.045 B D Z 0 0.680, -0.496 E D
<input checked="" type="checkbox"/> Summary						
exposure(s) <b>E</b> outcome(s) <b>D</b> covariates <b>3</b> causal paths <b>1</b>						

```
graph TD; A((A)) --> Z((Z)); B((B)) --> Z; E((E)) --> D((D)); B --> D; A --> D; A --> Z; B --> Z; style A fill:#90EE90; style B fill:#6495ED; style E fill:#90EE90,stroke:#333,stroke-width:1px; style D fill:#6495ED,stroke:#333,stroke-width:1px; style Z fill:#A9A9A9; linkStyle 0 stroke:#008000,stroke-width:2px; linkStyle 1 stroke:#800000,stroke-width:2px;
```