

# Identificación usando DAGs

Ricardo Pasquini

Universidad Austral

5/06/2023

# Intro

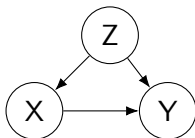
- ▶ Vamos a analizar la utilización de un DAG para inferir causalidad a partir de datos observados (no-experimentales).
- ▶ Nuestro punto de partida es el de siempre: en una situación de datos observacionales, diferenciar a un resultado en relación a una cierta política o tratamiento da como resultado algo que no necesariamente es el efecto de la política. Esa diferencia no solo devuelve el efecto en sí, sino que también podría encubrir diferencias preexistentes.
- ▶ La novedad aquí será explicitar cuáles son los factores que sesgan esta diferencia, y analizar qué método es necesario para recuperar una estimación libre de sesgos.

# ¿Qué es un DAG?

- ▶ Pearl y Mackenzie (2018) proponen utilizar a los gráficos causales para identificar cómo realizar un análisis que permita recuperar un efecto causal.
- ▶ Un gráfico permitirá identificar qué controles deberían agregarse al análisis y qué controles no.
- ▶ El gráfico directo acíclico, nos permite hacer una representación de cómo los constructos teóricos interactúan causalmente.
- ▶ Podemos también pensar a cada nodo como una variable y las flechas representan cómo se ven causalmente afectadas.
- ▶ El grafo debe incluir todas las relaciones causales que suponemos existen, y debieramos armarlo en base a nuestra experiencia y conocimiento de la literatura.

## Ejemplo de un DAG

- ▶ Tomemos el siguiente típico ejemplo de existente de un *confounder*:
- ▶ El supuesto es que  $Z$  tiene un efecto sobre  $X$  y sobre  $Y$ .
- ▶ La flecha entre  $X$  e  $Y$  representa al efecto que quisiéramos identificar.
- ▶ Esta situación se corresponde con una típica situación donde la correlación de  $X$  con  $Y$  no es necesariamente indicativa del efecto causal, porque  $Z$  sesga esa relación: nótese que el efecto de  $Z$  podría hacernos ver un efecto aún en un caso de que no lo haya.
- ▶ Sabemos que resolvemos el problema si controlamos por  $Z$ .



# Sistematizando el análisis de un DAG

- ▶ La propuesta de identificación de Pearl, consiste en identificar todos los posibles senderos desde  $X$  a  $Y$ . Un sendero es un camino a través de los nodos conectados sin tener en cuenta la direccionalidad de las flechas.
- ▶ La idea es que los posibles senderos pueden introducir confusión en la relación, y por esta razón nuestro análisis estará enfocado en "bloquear" esos senderos. Esos senderos se los llama "back-door paths".
- ▶ Por ejemplo, en el caso anterior, tenemos dos senderos: i) el sendero de interés  $X \Rightarrow Y$  y ii) el sendero que genera la confusión:  $X \Leftarrow Z \Rightarrow Y$ .

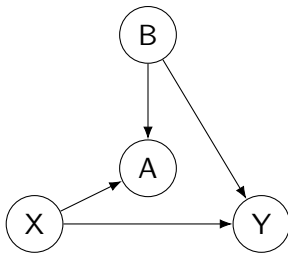
# Analizando el DAG

La direccionalidad de las flechas en un sendero tiene implicancias importantes. Podemos distinguir distintos casos de acuerdo a como llegan las flechas a un nodo intermedio.

- ▶ Una *cadena* tiene la forma  $X \Rightarrow Z \Rightarrow Y$ . Controlar por  $Z$  previene que detectemos una relación entre  $X$  e  $Y$ . Por lo tanto no quisieramos controlar por  $Z$  cuando hay una cadena.
- ▶ En una *bifurcación*  $X \Leftarrow Z \Rightarrow Y$  controlar por  $Z$  previene que  $X$  e  $Y$  mantengan una relación espúrea generada por  $Z$ . Por lo tanto quisieramos controlar por  $Z$  de manera evitar una relación espúrea.
- ▶ En un *colisionador*  $X \Rightarrow Z \Leftarrow Y$ ,  $X$  e  $Y$  son independientes, pero si se empieza a controlar  $Z$ , entonces se observará una relación entre  $X$  e  $Y$ . Por lo tanto, no quisieramos controlar por un colisionador.

# Juegos para De-confundir

## Juego 2



# Juegos para De-confundir

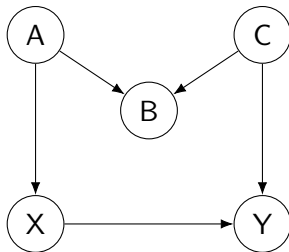
## Juego 2

- ▶ No es necesario controlar por nada.
- ▶ Controlar por A sería una mala idea, ya que habilitaría el sendero  $X \Rightarrow A \Leftarrow B \Rightarrow Y$ .
- ▶ Notar que controlar por A y B simultáneamente resolvería el problema, pero no era necesario controlar por nada en primer lugar.



# Juegos para De-confundir

## Juego 3 "M-Bias"



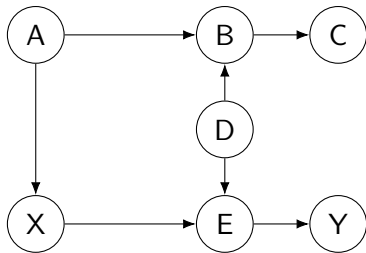
# Juegos para De-confundir

## Juego 3 "M-Bias"

- ▶ No es buena idea controlar por B
- ▶ Si lo hago, entonces también tengo que controlar por A y C

# Juegos para De-confundir

## Juego 4



# Juegos para De-confundir

## Juego 4

- ▶ No es necesario incorporar controles pues B es un colisionador.
- ▶ Controlar por B es una mala idea.
- ▶ Si controlamos por B, tambien tendremos que controlar por A para bloquear nuevamente el sendero.
- ▶ Controlar por E es una mala idea, ya que vamos a capturar el efecto que X tiene sobre Y.