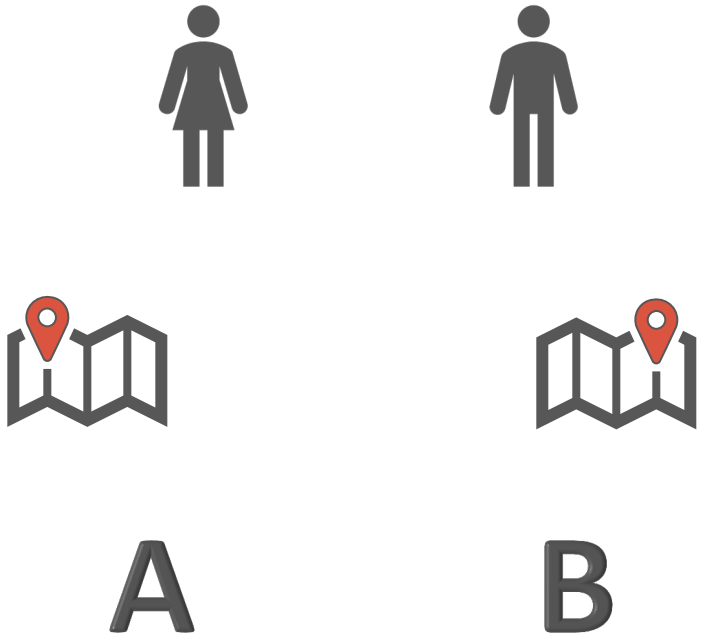
Directed Acyclic Graphs

an introduction

Louisa H. Smith
Kolokotrones Symposium
December 7, 2018

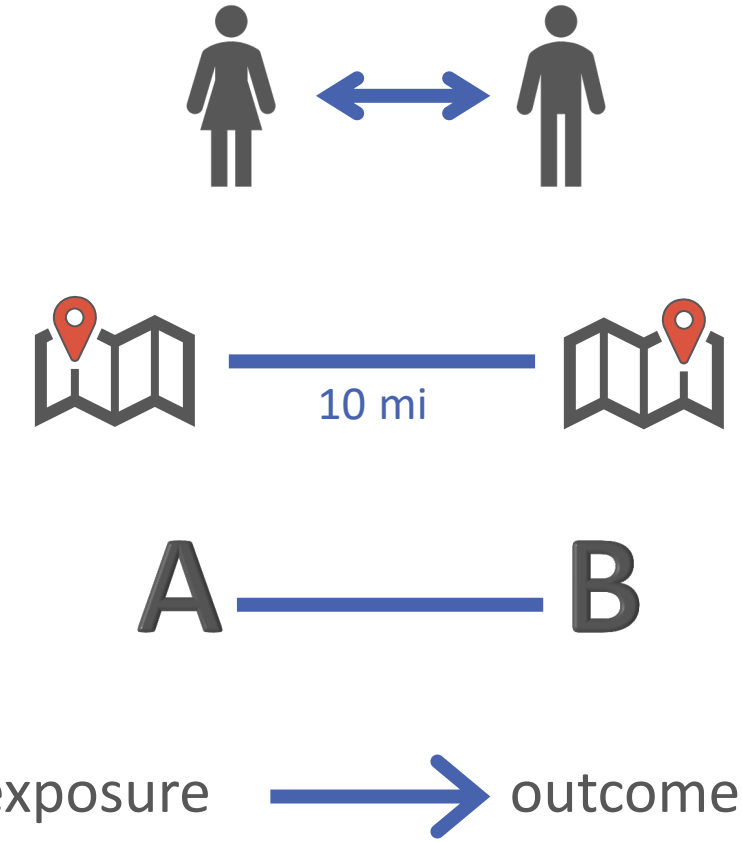
Components of graphs

Nodes



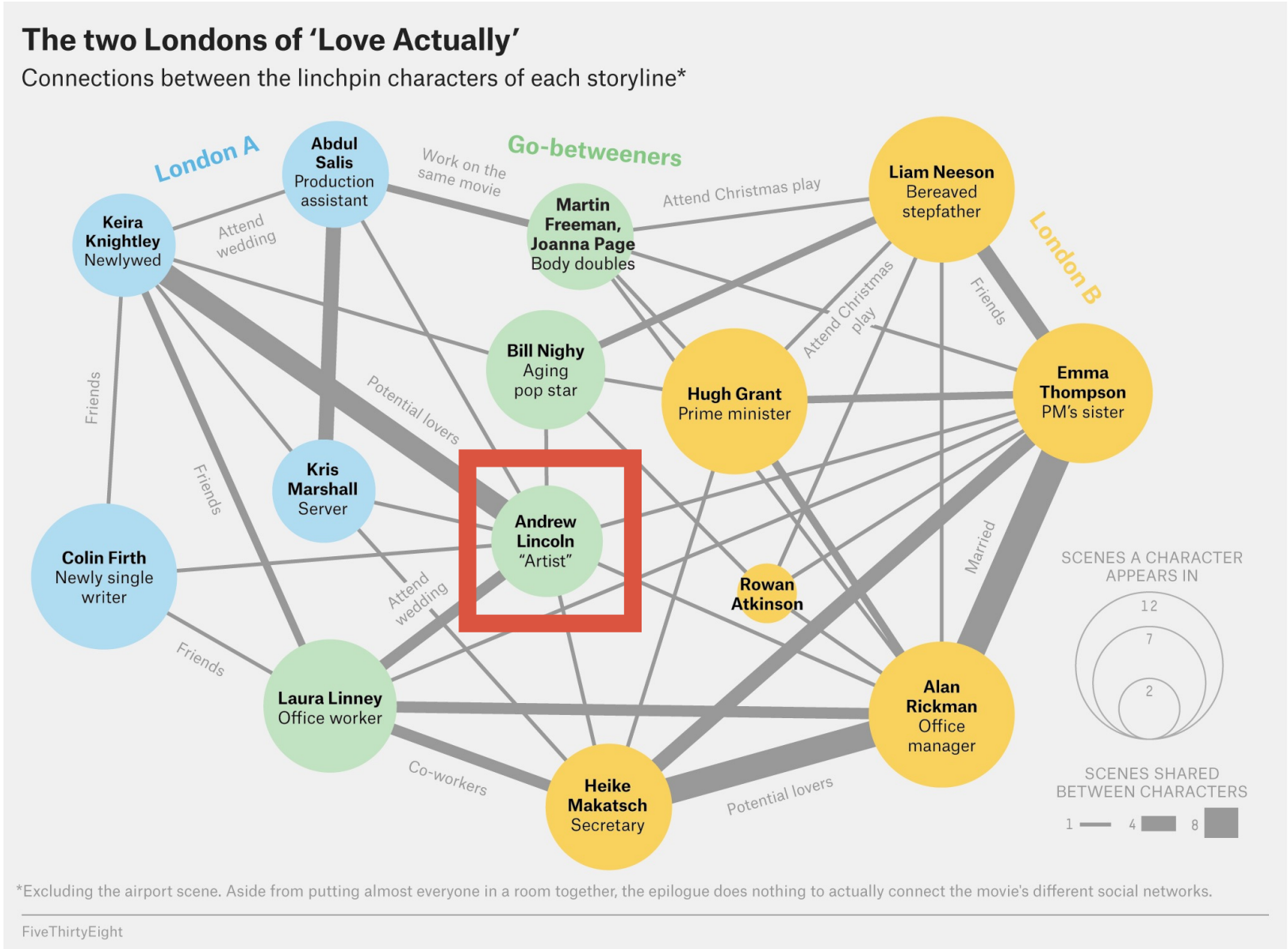
exposure outcome

Edges



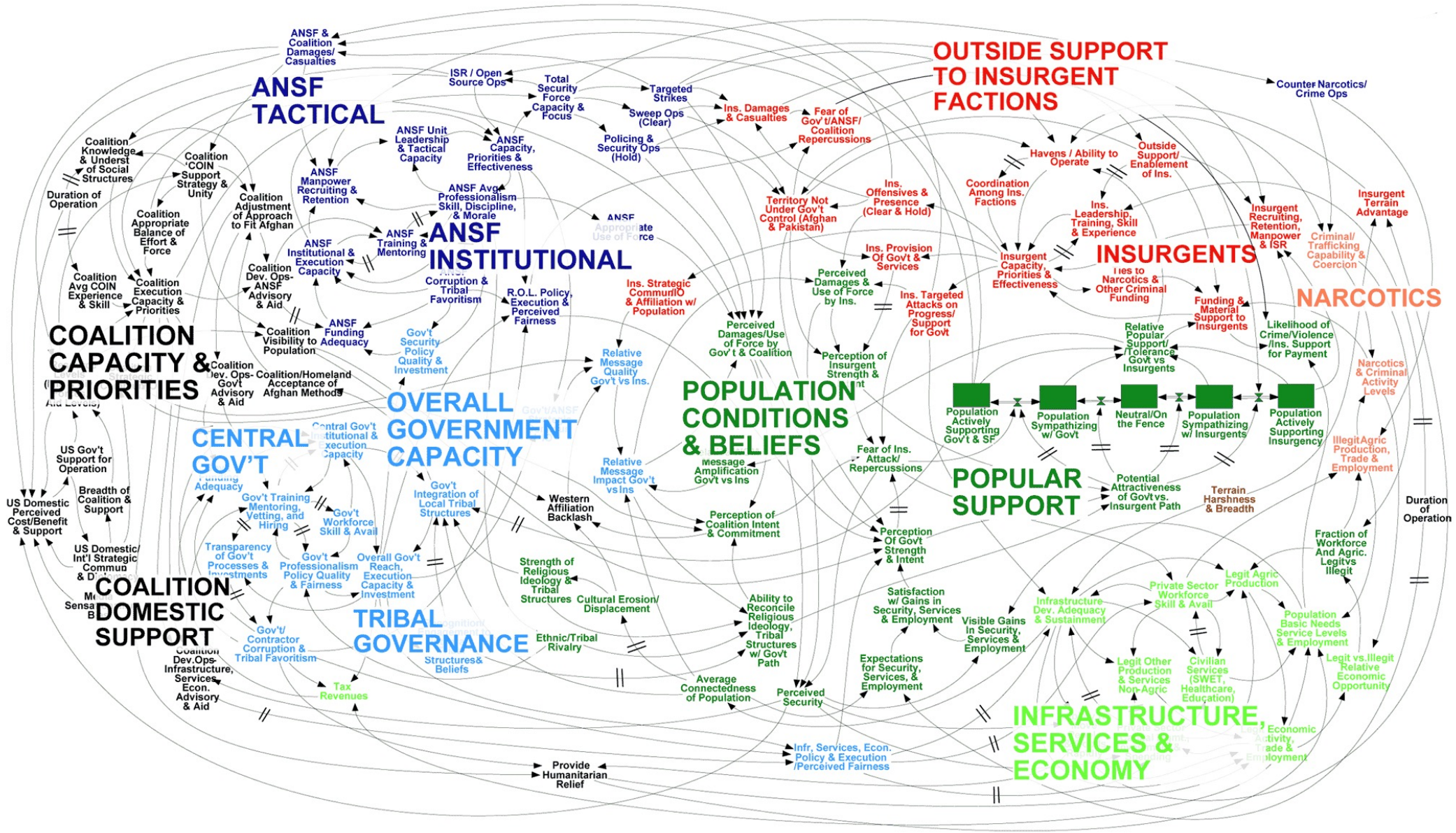
exposure → outcome

Love Actually graph



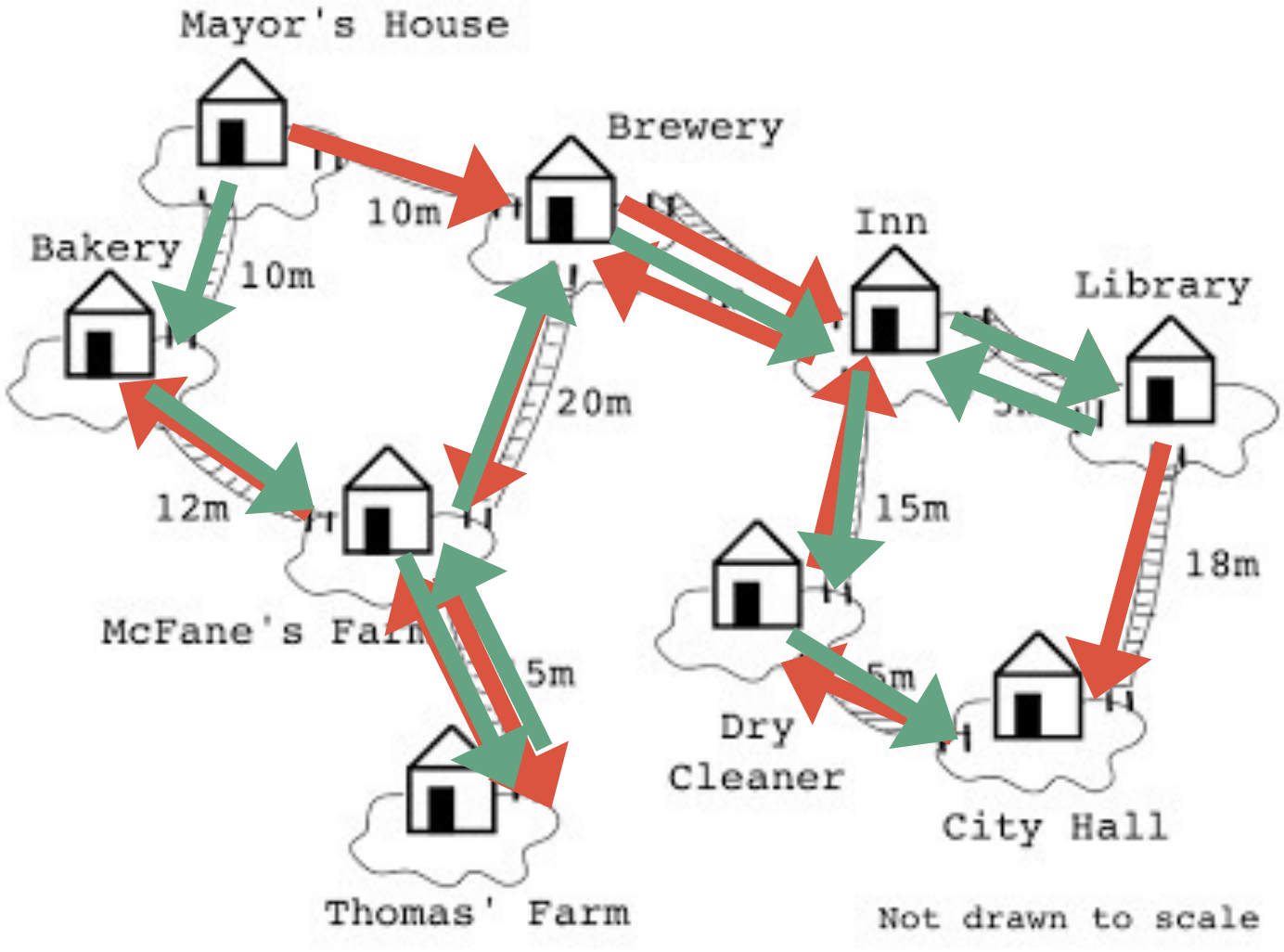
https://fivethirtyeight.com/features/the-definitive-analysis-of-love-actually-the-greatest-christmas-movie-of-our-time/

Afghanistan strategy graph



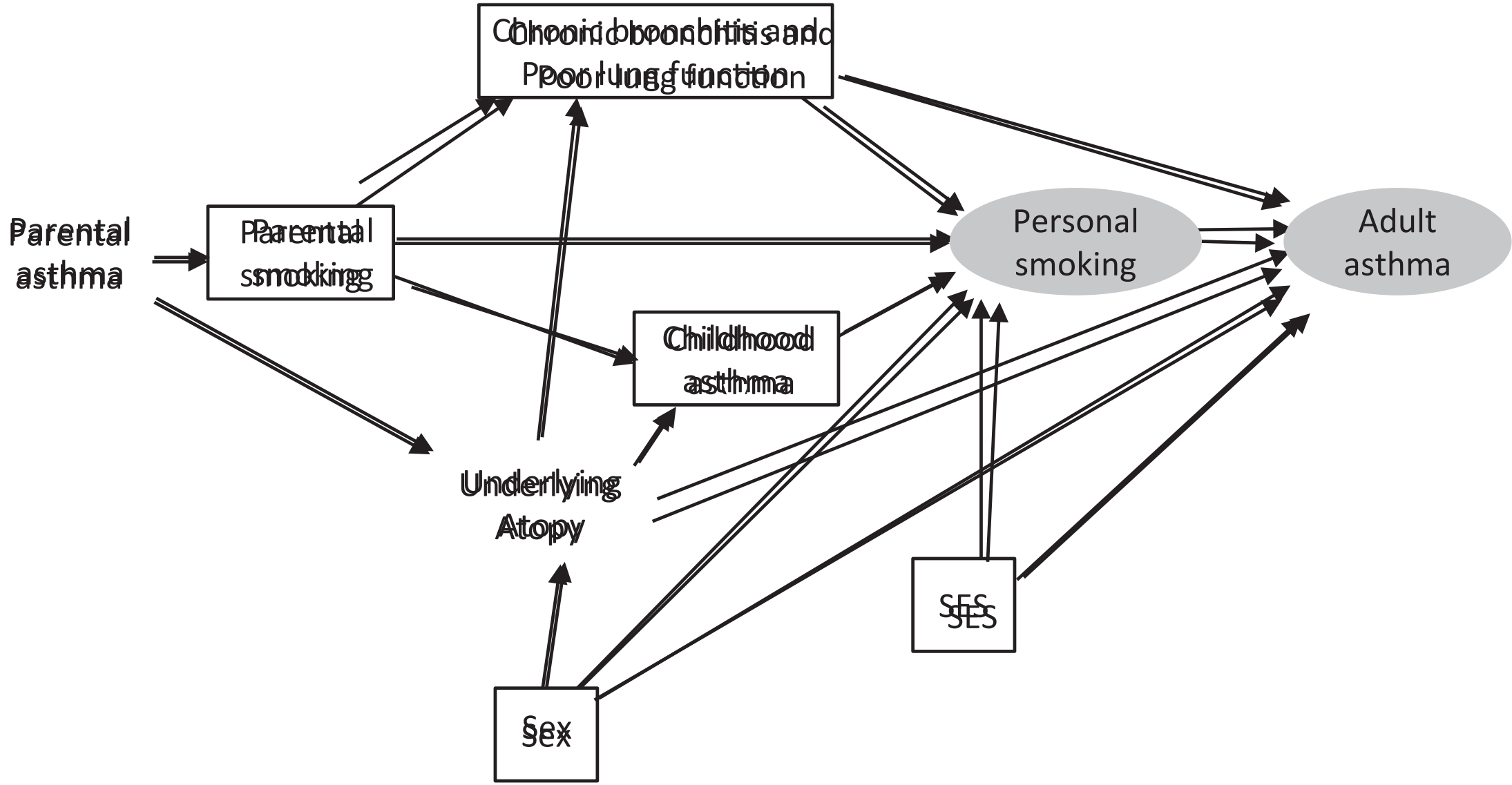
http://sdwise.com/2013/07/hey-new-york-times-a-causal-loop-diagram-is-not-a-powerpoint-fail/

Travelling salesman graph



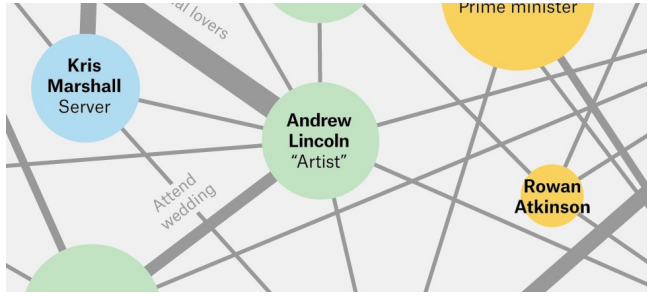
http://computationaltales.blogspot.com/2011/07/dijkstras-algorithm-on-scooters-part-4.html

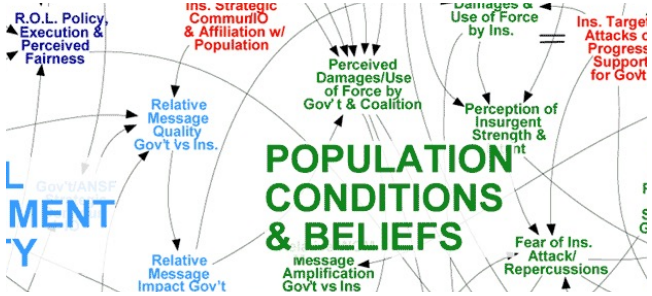
Smoking-asthma graph

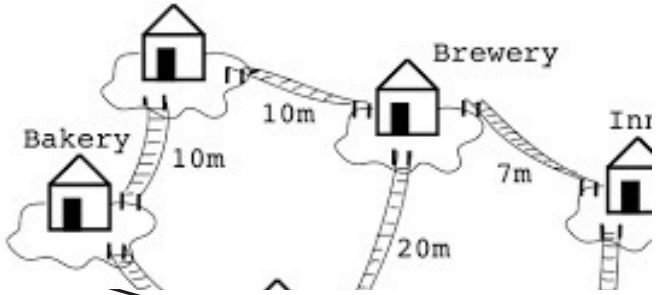


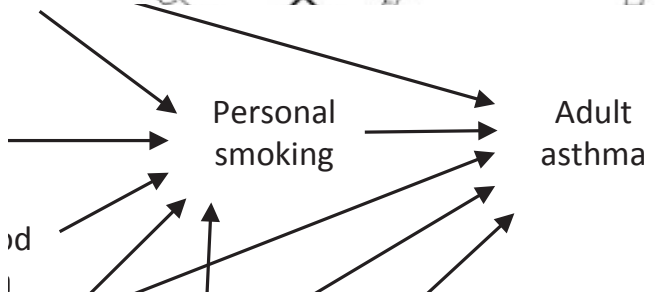
Williamson EJ, et al. Introduction to causal diagrams for confounder selection. Respirology. 2014;19(3):303–311.

Graphs help solve problems





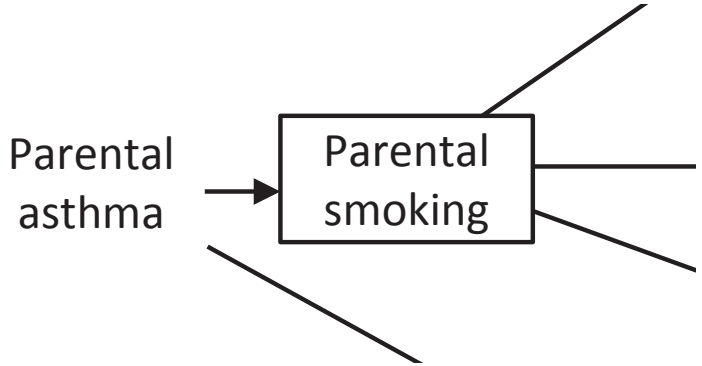




- Who is the central character tying together the various plots in the movie?
- What are the downstream effects of increasing resources in one sector vs. another?
- What is the shortest route connecting all the locations?
- How can we reduce (and avoid inducing) bias in our study?

Strategies that may reduce (or induce!) bias

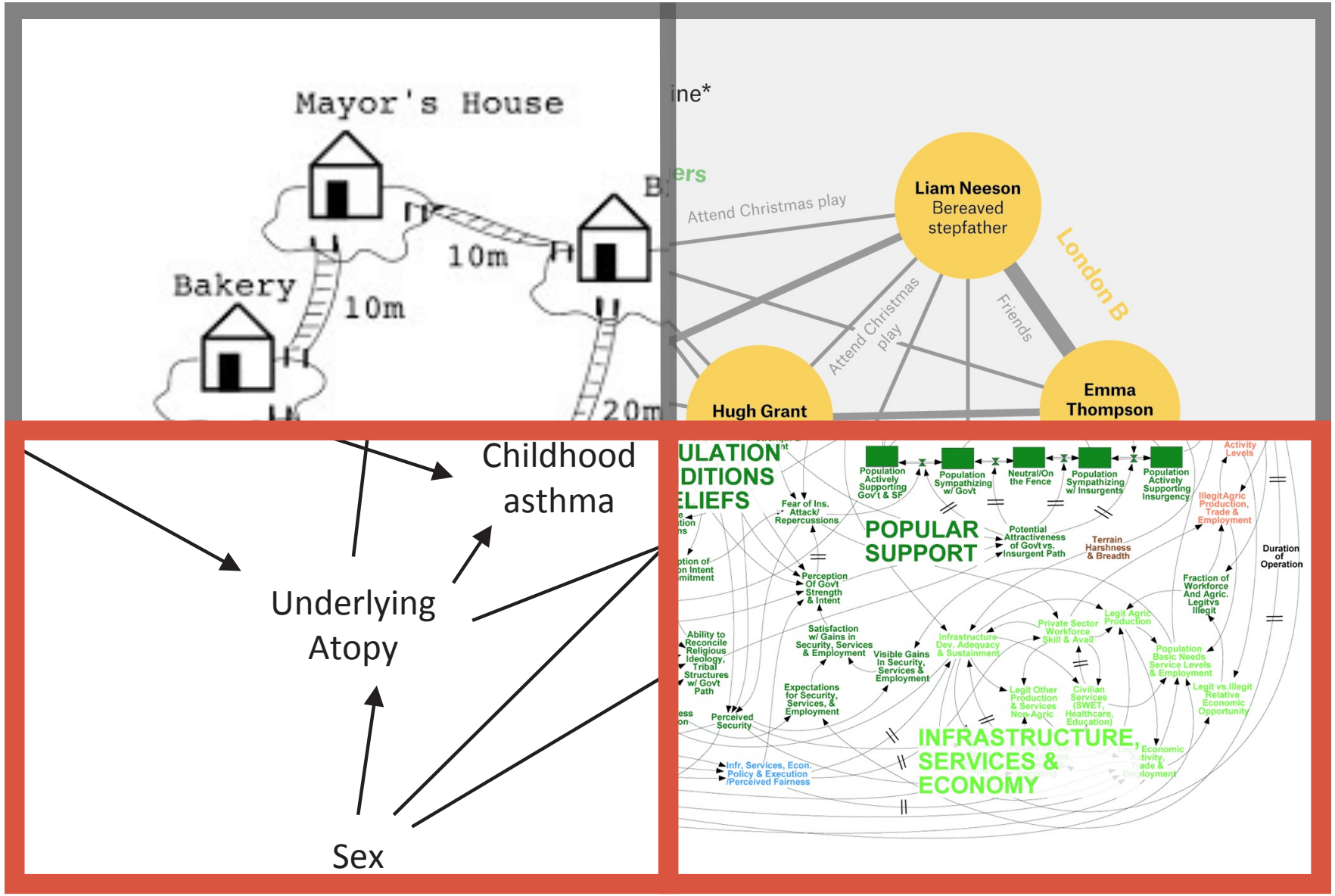
- Restrict a study's participants to those with certain values of a variable
  - e.g., recruit only people whose parents smoked
- Stratify by or statistically adjust for a variable
  - e.g., include parental smoking status (yes/no) in the model



We often indicate that we've conditioned on a variable (by restricting, adjusting, etc.) by drawing a box around that node

Directed Acyclic Graphs

Graph edges


Mayor's House
Bakery
10m
10m
20m
Attend Christmas play
Liam Neeson
Bereaved
stepfather
London B
Attend Christmas play
Friends
Hugh Grant
Emma
Thompson
Childhood
asthma
Underlying
Atopy
Sex
POPULAR
SUPPORT
INFRASTRUCTURE,
SERVICES &
ECONOMY

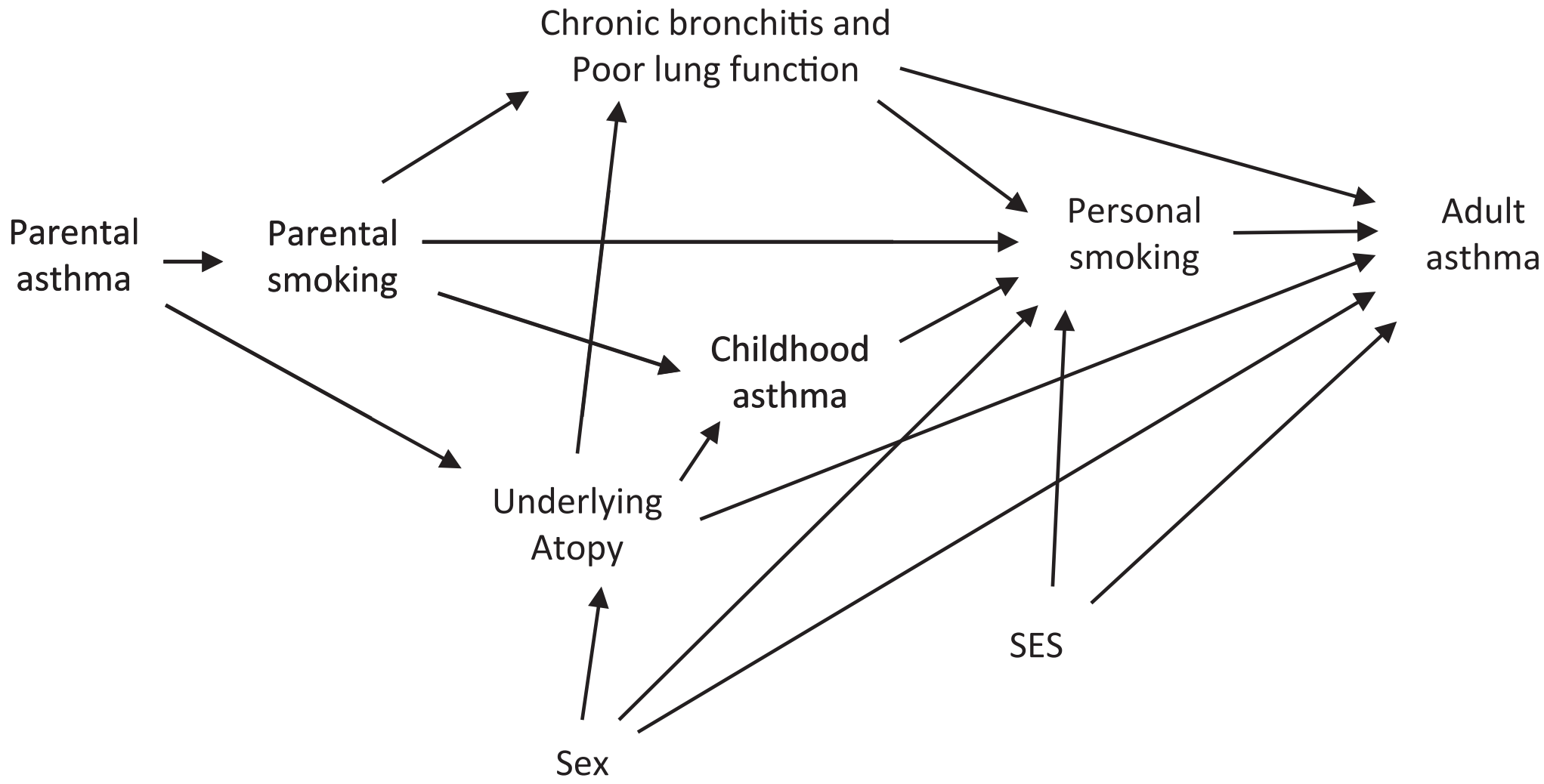
Edges are *directed*

A may cause B

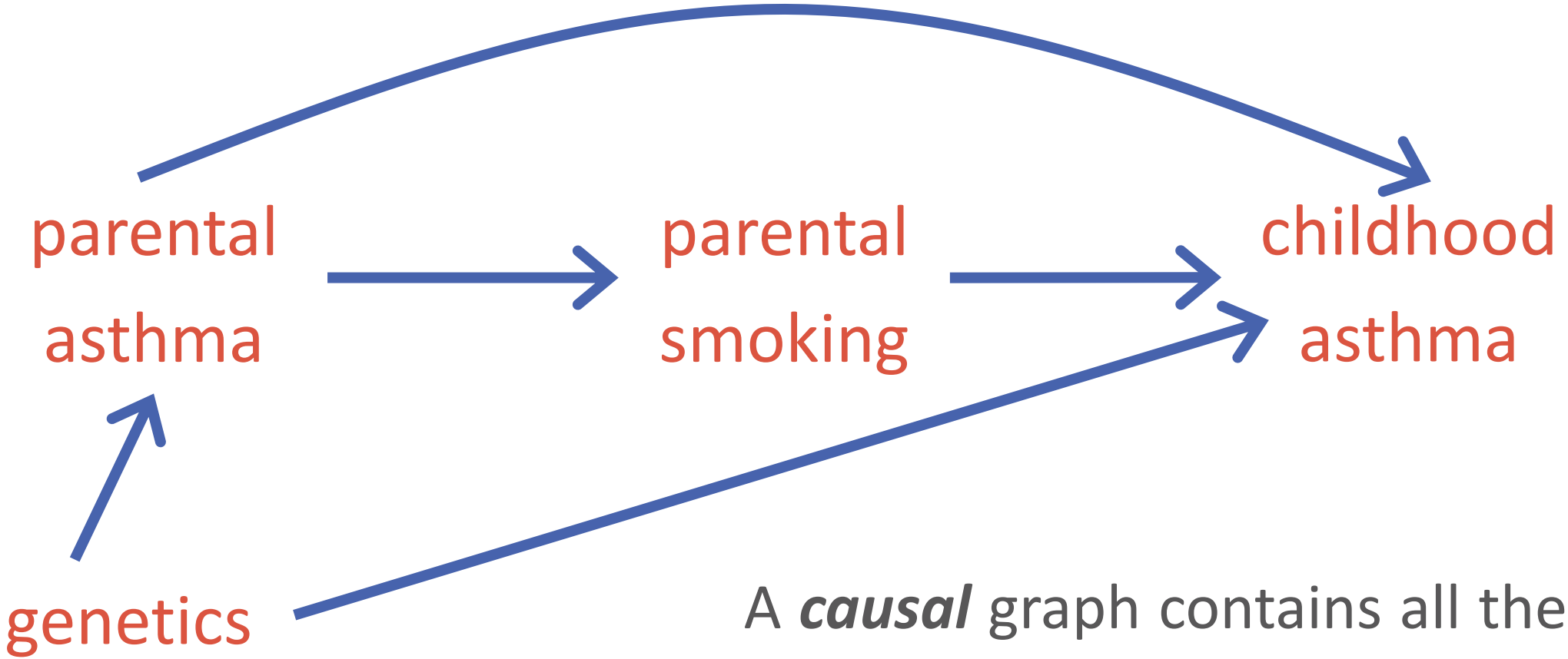
B may cause A

A doesn't cause B and B doesn't cause A

Smoking-asthma relationship assumptions



Smoking-asthma *causal* relationships

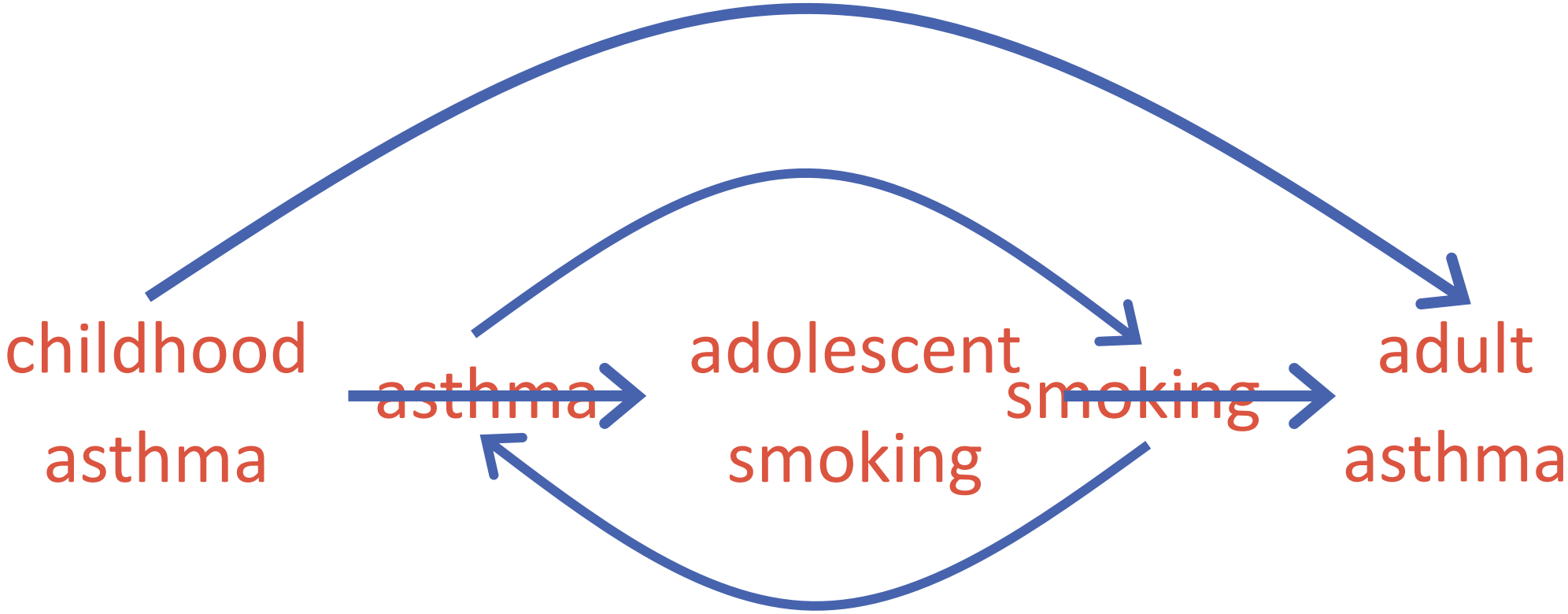


A ***causal*** graph contains all the common causes of the variables of interest
(even if unmeasured)

Directed Acyclic Graphs

Feedback loops

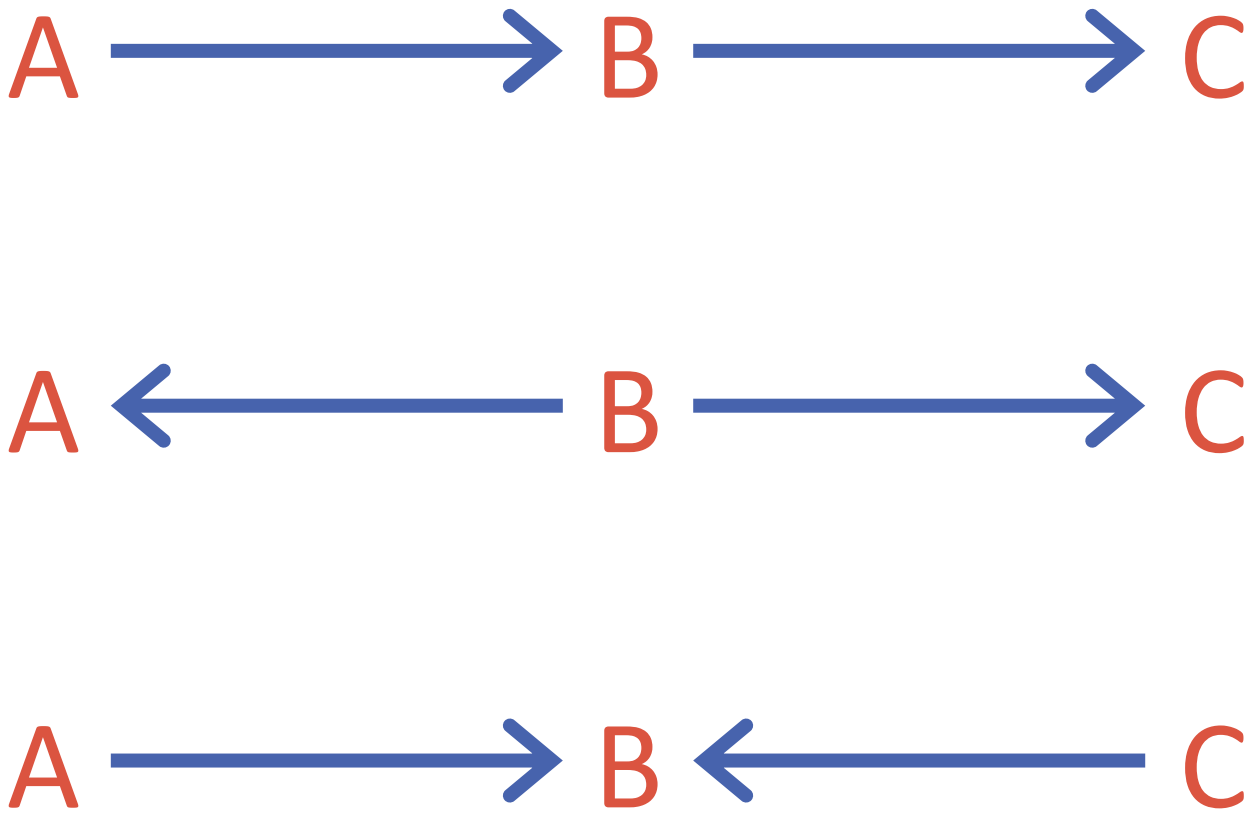
What do we do instead?



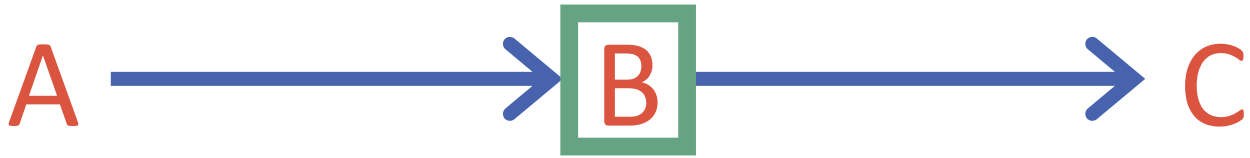
sequence of nodes through time

How do we read graphs?

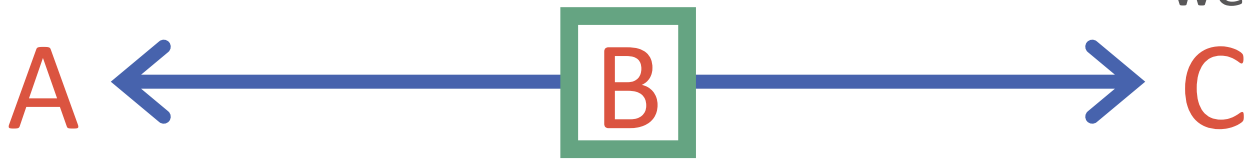
sequences of connected nodes form ***paths***



Paths may be *blocked*



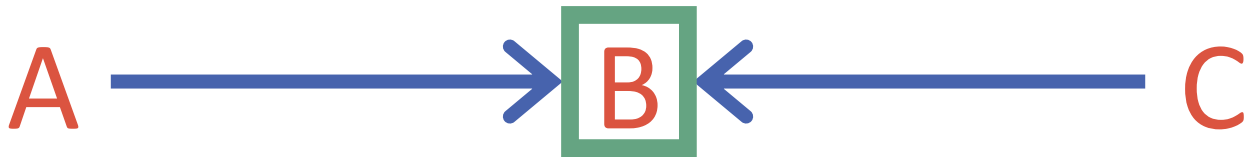
Recall that we often put boxes around nodes to show we've ***conditioned*** on those variables



A ***collider*** is a node with both arrowheads incoming

Or *unblocked*

Because the arrow into A goes the wrong way, this is a ***backdoor path***

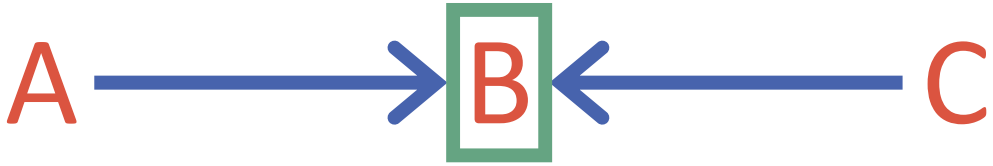


Conditioning on a collider ***unblocks*** the path

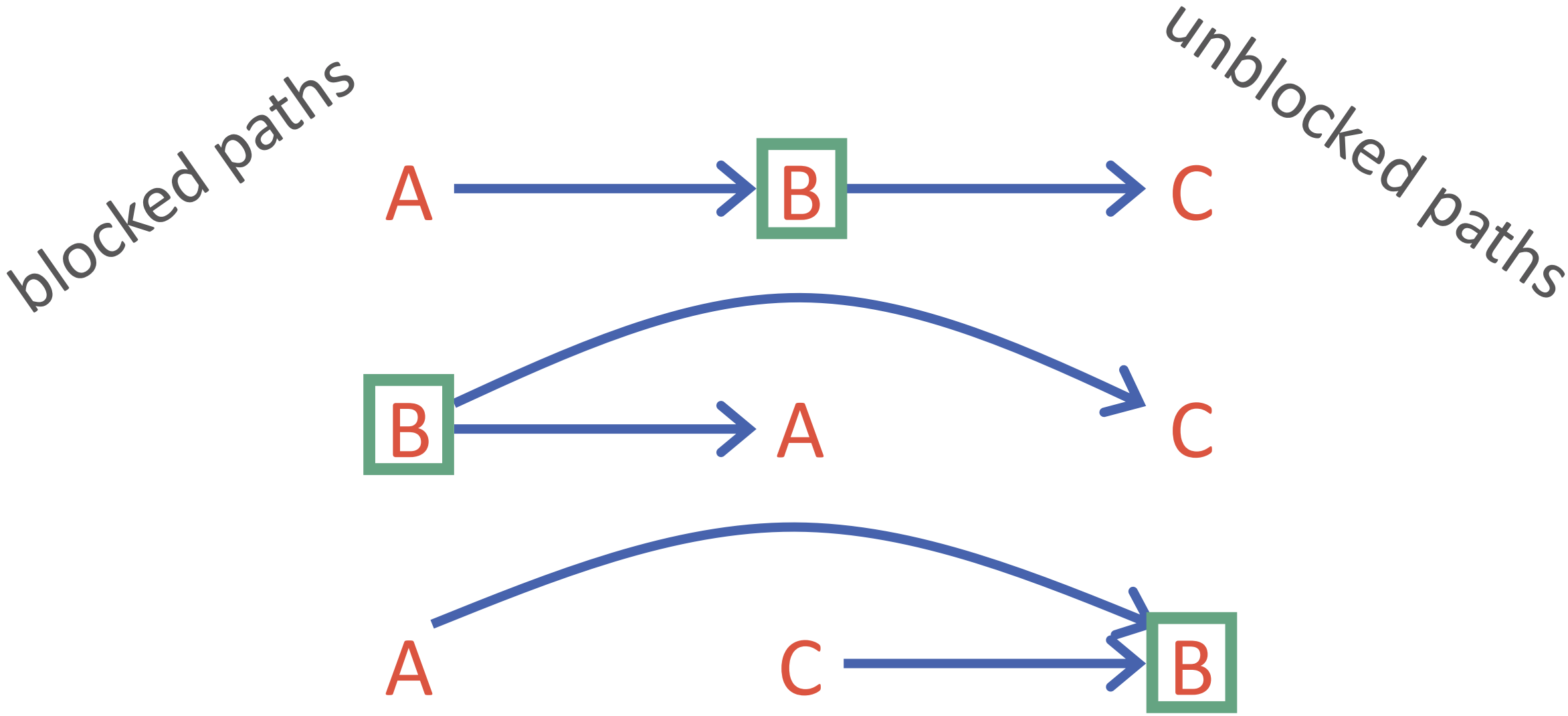
Unblocked paths mean variables are associated

- A and C are ***causally*** associated
  - We can estimate what will happen to C if we change A

- A and C are ***non-causally*** associated
  - Because they're associated in the data, we may think that A causes C
  - But this is just bias!



Graphs are often easier to read in *temporal order*



Directed Acyclic Graphs