Learning causal DAGs using adaptive interventions

Davin **C**hoo

24 Feb 2023
Computing Research Week - Open House 2023

This talk is based on joint work with
Arnab **B**hattacharyya, Themis **G**ouleakis, Kirankumar **S**hiragur

Important decisions in life…

- **What if** I ate Kaya Toast instead of Roti Prata for breakfast? Will I feel more satisfied?

https://www.sbs.com.au/food/recipes/kaya-toast-half-cooked-eggs-and-hainanese-coffee
https://danielfooddiary.com/2021/02/17/sinmingrotiprata/

Important decisions in life…

- **What if** I ate Kaya Toast instead of Roti Prata for breakfast? Will I feel more satisfied?

- **What if** I exercised more? Will I become fitter?

Important decisions in life…

- **What if** I ate Kaya Toast instead of Roti Prata for breakfast? Will I feel more satisfied?

- **What if** I exercised more? Will I become fitter?
- **What if** I went to University X instead of NUS? Will I be more successful?
- …

Important decisions in life…

- **What if** I ate Kaya Toast instead of Roti Prata for breakfast? Will I feel more satisfied?

- **What if** I exercised more? Will I become fitter?
- **What if** I went to University X instead of NUS? Will I be more successful? Not necessarily. We have great people here ☺
- …

David Hume
Philosopher

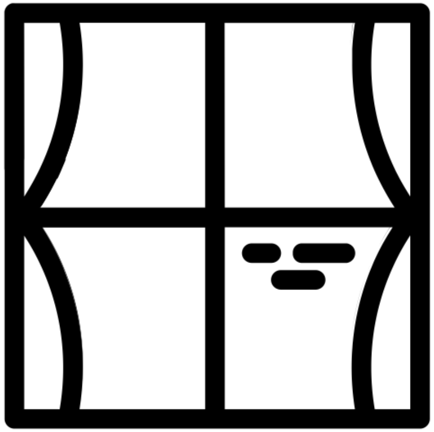
One of his philosophical ideas:

- To draw any causal conclusions from past experiences, one has to assume that the future resembles the past
- This assumption cannot be justified

https://en.wikipedia.org/wiki/David_Hume

David Hume
Philosopher

Yesterday



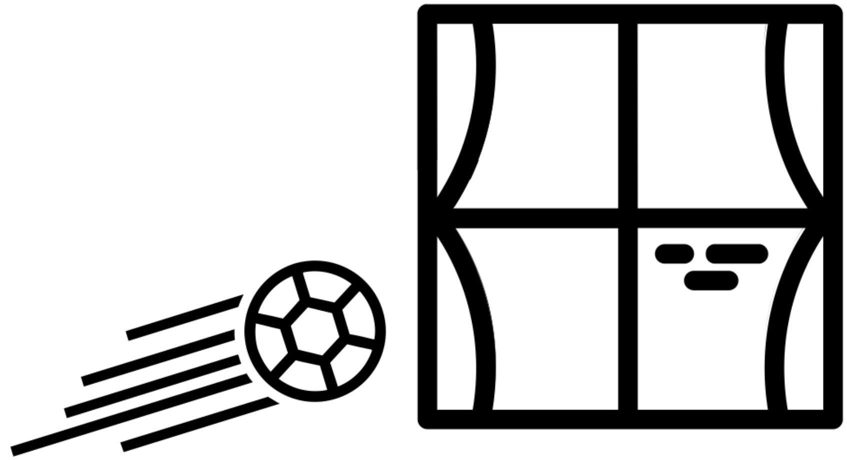
One of his philosophical ideas:

- To draw any causal conclusions from past experiences, one has to assume that the future resembles the past
- This assumption cannot be justified

https://thenounproject.com/icon/window-1586841/

David Hume
Philosopher

Yesterday



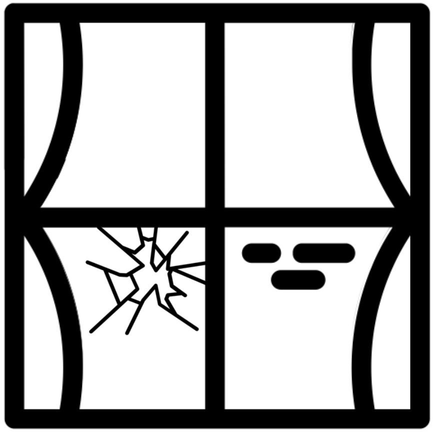
One of his philosophical ideas:

- To draw any causal conclusions from past experiences, one has to assume that the future resembles the past
- This assumption cannot be justified

https://thenounproject.com/icon/ball-1295289/

David Hume
Philosopher

Yesterday



One of his philosophical ideas:

- To draw any causal conclusions from past experiences, one has to assume that the future resembles the past
- This assumption cannot be justified

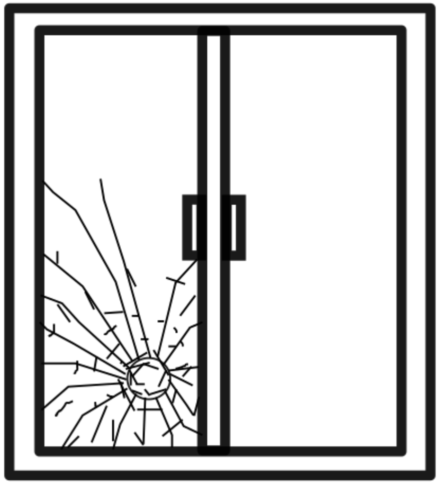
https://thenounproject.com/icon/broken-glass-762866/
https://thenounproject.com/icon/ball-5526599/

David Hume
Philosopher

One of his philosophical ideas:

- To draw any causal conclusions from past experiences, one has to assume that the future resembles the past
- This assumption cannot be justified

Today



Did the ball smash the window?

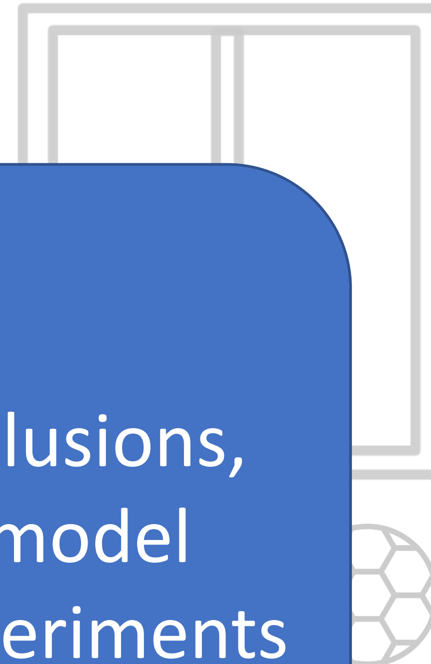
OR

Did the smashed window summon the ball?

https://thenounproject.com/icon/broken-window-419316/

Da
Ph

Today



To make useful causal conclusions, make useful/reasonable model assumptions or conduct experiments

One of

- To dra
past experiences, one has to assume that the future resembles the past
- This assumption cannot be justified

h the

OR

Did the smashed window summon the ball?

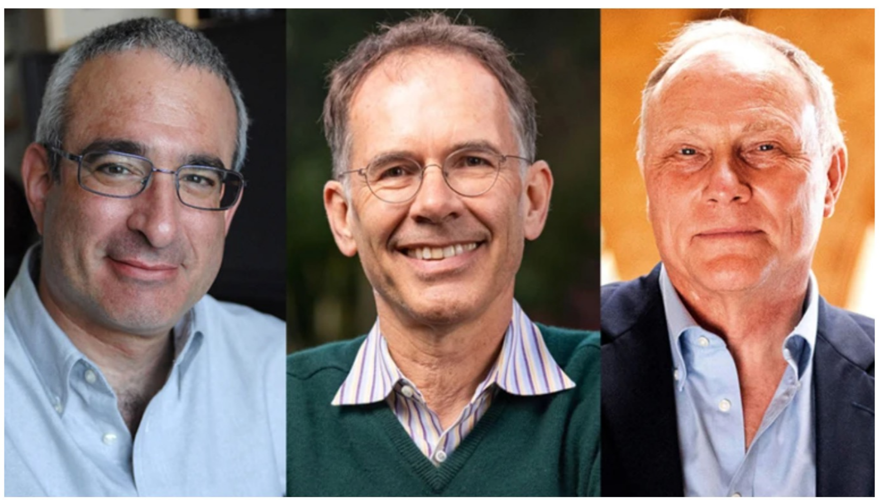
https://thenounproject.com/icon/broken-window-419316/

NEWS | 13 October 2021

Nobel-winning 'natural experiments' approach made economics more robust

Joshua Angrist, Guido Imbens and David Card share the prize for finding a way to identify cause and effect in social science.

Philip Ball



Left to Right: Joshua Angrist, Guido Imbens and David Card share the 2021 Nobel prize in economic sciences for work that has helped economics research undergo a 'credibility revolution'. Credit: MIT/EPA-EFE/Shutterstock, Andrew Brodhead/Stanford News Service/EPA-EFE/Shutterstock, Noah Berger/AP/Shutterstock

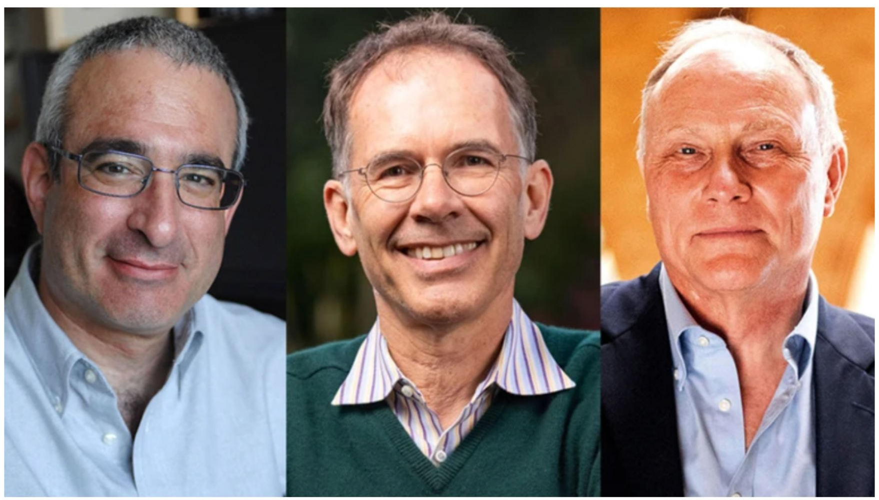
https://www.nature.com/articles/d41586-021-02799-7

NEWS | 13 October 2021

Nobel-winning 'natural experiments' approach made economics more robust

Joshua Angrist, Guido Imbens and David Card share the prize for finding a way to identify cause and effect in social science.

Philip Ball



Left to Right: Joshua Angrist, Guido Imbens and David Card share the 2021 Nobel prize in economic sciences for work that has helped economics research undergo a 'credibility revolution'. Credit: MIT/EPA-EFE/Shutterstock, Andrew Brodhead/Stanford News Service/EPA-EFE/Shutterstock, Noah Berger/AP/Shutterstock

JUDEA PEARL

United States – 2011

(2011 Turing award)

CITATION

For fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning.

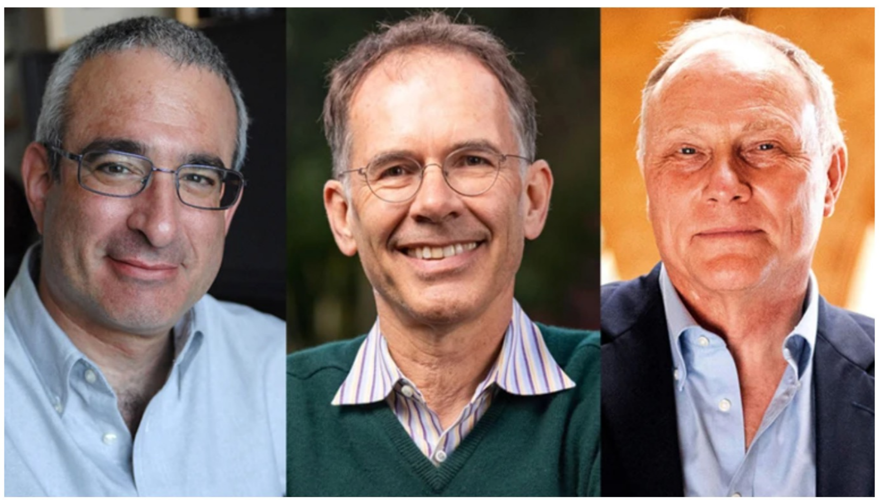
https://amturing.acm.org/award_winners/pearl_2658896.cfm

NEWS | 13 October 2021

Nobel-winning 'natural experiments' approach made economics more robust

Joshua Angrist, Guido Imbens and David Card share the prize for finding a way to identify cause and effect in social science.

Philip Ball



Left to Right: Joshua Angrist, Guido Imbens and David Card share the 2021 Nobel prize in economic sciences for work that has helped economics research undergo a 'credibility revolution'. Credit: MIT/EPA-EFE/Shutterstock, Andrew Brodhead/Stanford News Service/EPA-EFE/Shutterstock, Noah Berger/AP/Shutterstock

JUDEA PEARL

United States – 2011

(2011 Turing award)

CITATION

For fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning.

- Bayesian networks
  - Represent causal relationships as a directed acyclic graph (DAG)
- Do-calculus
  - A formalization of interventions
  - "What happens if we perform experiments on the causal graph?"

https://amturing.acm.org/award_winners/pearl_2658896.cfm

Modelling causal relations

"**We may regard the present state of the universe as the effect of its past and the cause of its future…**" – Pierre Simon Laplace, *A Philosophical Essay on Probabilities*, 1814

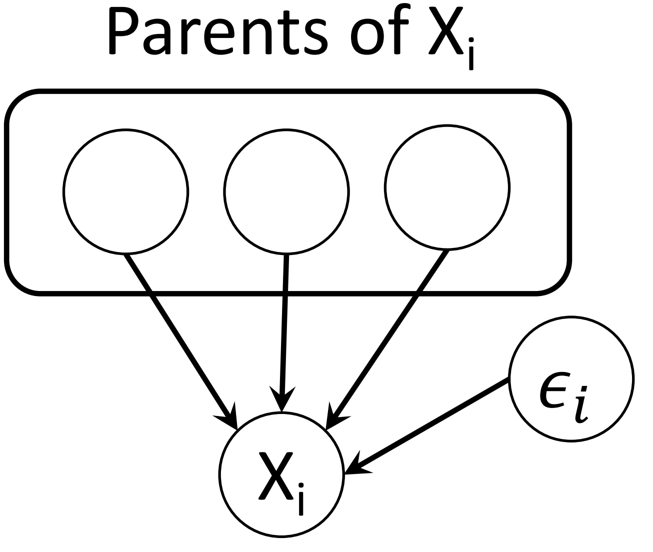
https://en.wikipedia.org/wiki/Laplace%27s_demon

Modelling causal relations

"**We may regard the present state of the universe as the effect of its past and the cause of its future…**" – Pierre Simon Laplace, *A Philosophical Essay on Probabilities*, 1814

$$X_i = f_i(pa_i, \epsilon_i)$$

The value of each variable X_i is function f_i of the values taken by its parents pa_i and some noise ϵ_i



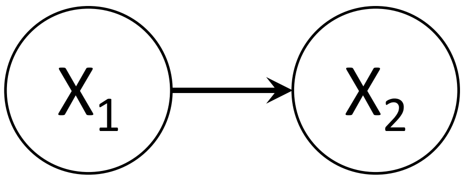
https://en.wikipedia.org/wiki/Laplace%27s_demon

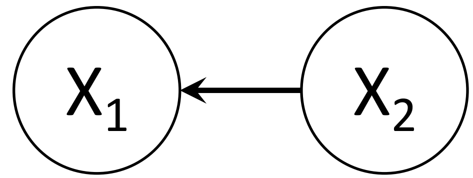
Which model generated this data?

X_1	-0.27	0.29	0.37	-0.09	0.34	0.33	0.30	-1.34	0.68
X_2	-0.10	1.65	0.47	1.92	2.04	1.67	0.11	-3.58	1.97

Which model generated this data?

X_1	-0.27	0.29	0.37	-0.09	0.34	0.33	0.30	-1.34	0.68
X_2	-0.10	1.65	0.47	1.92	2.04	1.67	0.11	-3.58	1.97





- $X_1 = \epsilon_1$
- $X_2 = a \cdot X_1 + \epsilon_2$

- $X_1 = b \cdot X_2 + \epsilon_3$
- $X_2 = \epsilon_4$

Simple linear relationship between variables

a and b are (hidden) positive constants

ϵ's are independent Gaussian terms with mean 0

Two equivalent causal models

X_1	-0.27	0.29	0.37	-0.09	0.34	0.33	0.30	-1.34	0.68
X_2	-0.10	1.65	0.47	1.92	2.04	1.67	0.11	-3.58	1.97

$$X_1 \rightarrow X_2 \qquad\qquad X_1 \leftarrow X_2$$

- $X_1 = \epsilon_1 \sim N(0,1)$
- $X_2 = X_1 + \epsilon_2 \sim N(0,2)$
- $\epsilon_1 \sim N(0,1)$
- $\epsilon_2 \sim N(0,1)$

- $X_1 = \frac{1}{2} \cdot X_2 + \epsilon_3 \sim N(0,1)$
- $X_2 = \epsilon_4 \sim N(0,2)$
- $\epsilon_3 \sim N\left(0, \frac{1}{2}\right)$
- $\epsilon_4 \sim N(0,2)$

Example from: https://youtu.be/rE6IMfSkOU0?t=849. See https://github.com/csquires/6.S091-causality for full course

Two equivalent causal models

X_1	-0.27	0.29	0.37	-0.09	0.34	0.33	0.30	-1.34	0.68
X_2	-0.10	1.65	0.47	1.92	2.04	1.67	0.11	-3.58	1.97

X_1 X_2

So what?
Who cares?

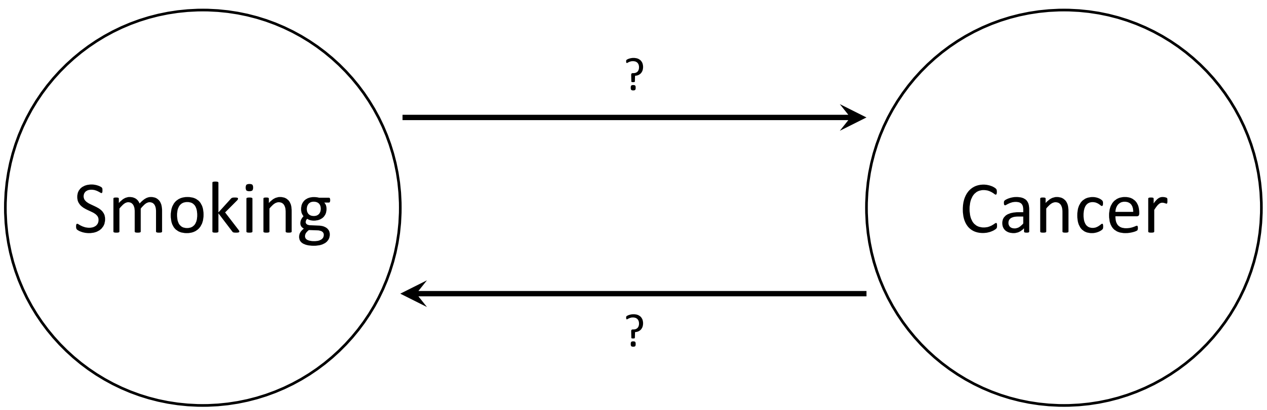
- $X_1 = \epsilon_1 \sim N(0,1)$
- $X_2 = X_1 + \epsilon_2 \sim N(0,2)$
- $\epsilon_1 \sim N(0,1)$
- $\epsilon_2 \sim N(0,1)$

- $X_1 = \frac{1}{2} \cdot X_2 + \epsilon_3 \sim N(0,1)$
- $X_2 = \epsilon_4 \sim N(0,2)$
- $\epsilon_3 \sim N\left(0, \frac{1}{2}\right)$
- $\epsilon_4 \sim N(0,2)$

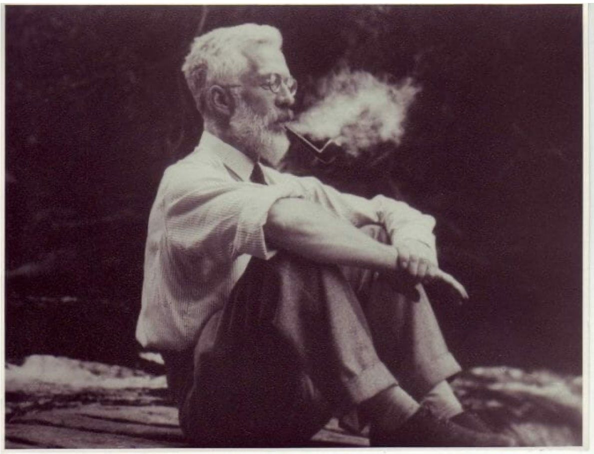
Example from: https://youtu.be/rE6IMfSkOU0?t=849. See https://github.com/csquires/6.S091-causality for full course

Smoking

Smoking	Yes	Yes	Yes	No	No	No	…
Cancer	No	Yes	Yes	No	No	Yes	…



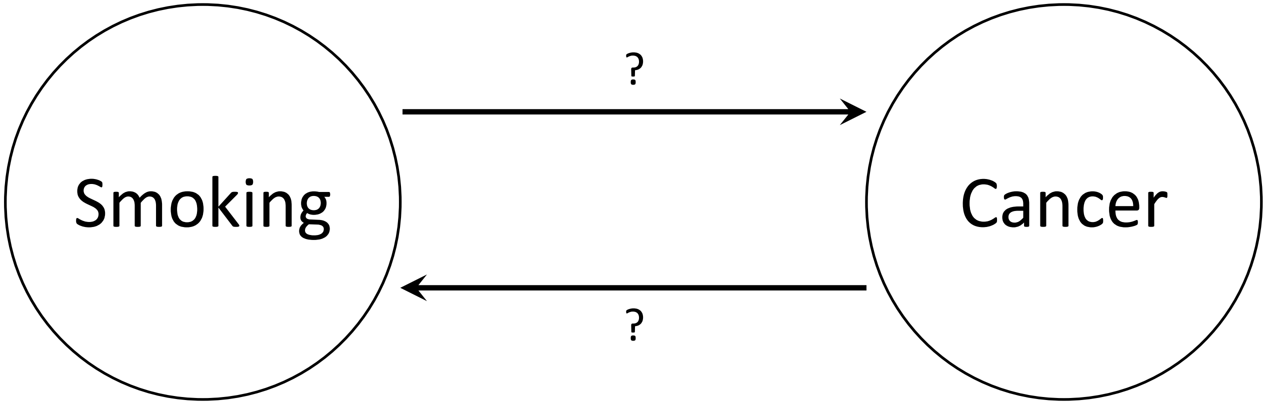
Smoking



Ronald Fisher

Fisher's letter to Nature, 1958:

"The curious associations with lung cancer found in relation to smoking habits do not, in the minds of some of us, lend themselves easily to the simple conclusion that the products of combustion reaching the surface of the bronchus induce, though after a long interval, the development of a cancer... **Such results suggest that an error has been made, of an old kind, in arguing from correlation to causation**..."

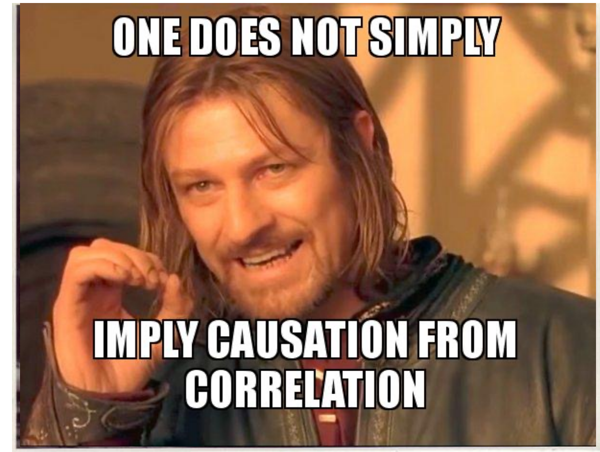


Fisher's letter to Nature, 1958: https://www.nature.com/articles/182596a0.pdf
Nice blog post: https://priceonomics.com/why-the-father-of-modern-statistics-didnt-believe/ (picture of Fisher from this link)

Smoking

Fisher’s letter to Nature, 1958:

“The curious associations with lung cancer found in relation to smoking habits do not, in the minds of some of us, lend themselves easily to the simple conclusion that the products of combustion reaching the surface of the bronchus induce, though after a long interval, the development of a cancer... **Such results suggest that an error has been made, of an old kind, in arguing from correlation to causation**...”



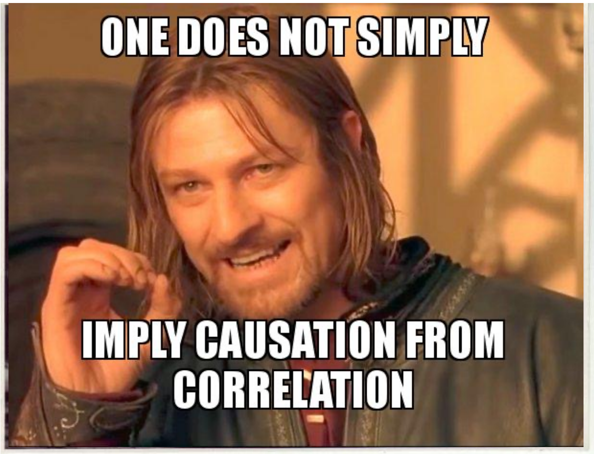
Ronald Fisher

From Google image search: https://makeameme.org/meme/one-does-not-5aed73

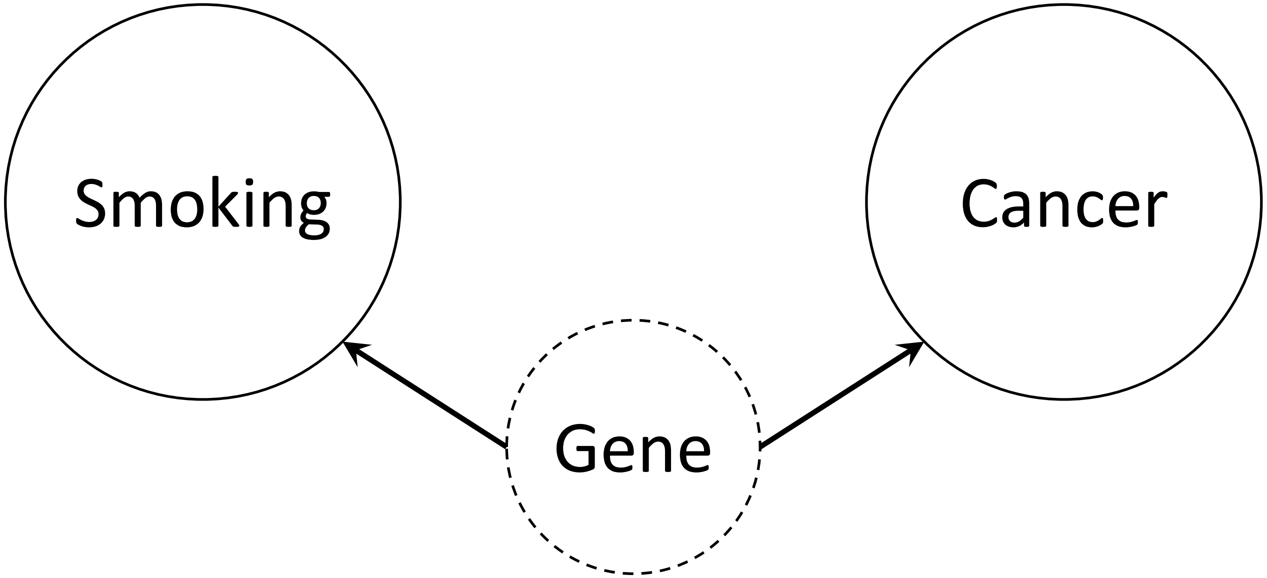
Smoking

Fisher's letter to Nature, 1958:

"… Such results suggest that an error has been made, of an old kind, in arguing from correlation to causation, and that the possibility should be explored that the different smoking classes, non-smokers, cigarette smokers, cigar smokers, pipe smokers, etc., have adopted their habits partly by reason of their personal temperaments and dispositions, and are not lightly to be assumed to be equivalent in their **genotypic composition**…"

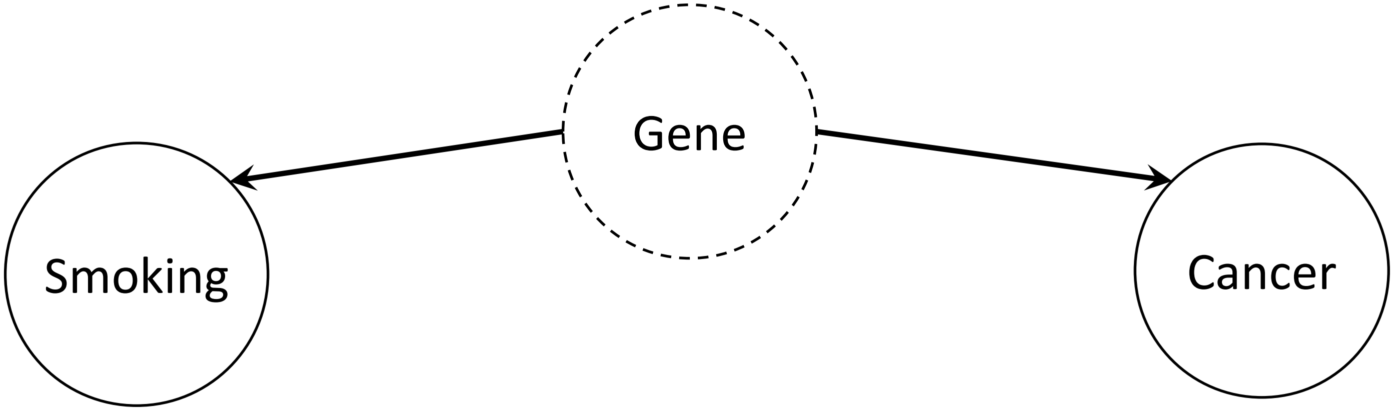


Ronald Fisher

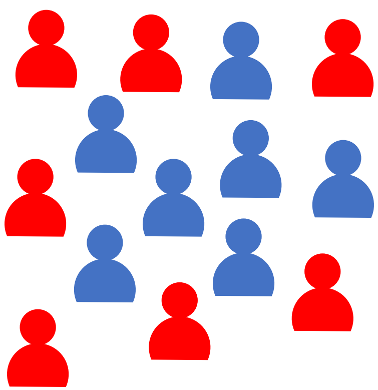


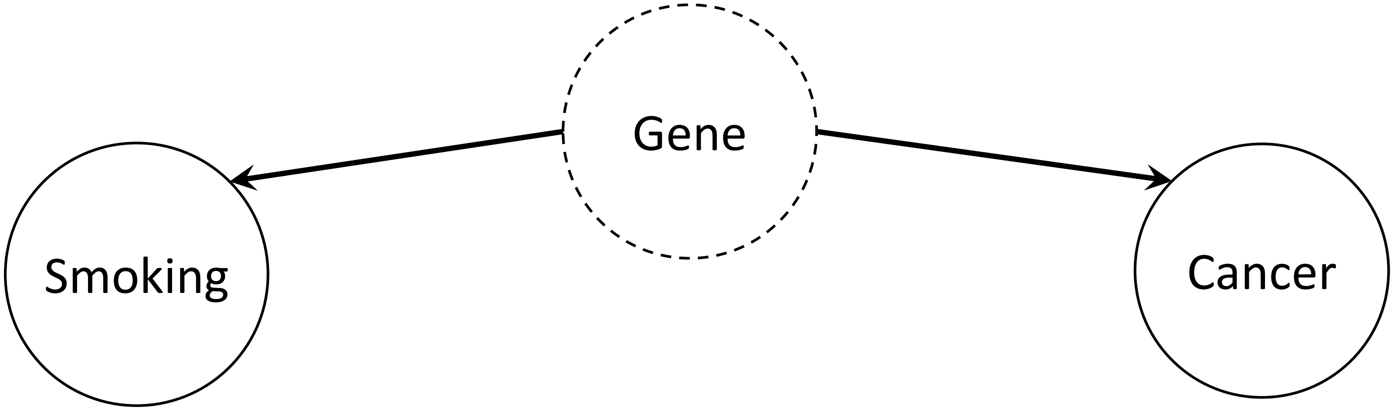
Maybe there's an unmeasured confounder?

From Google image search: https://makeameme.org/meme/one-does-not-5aed73

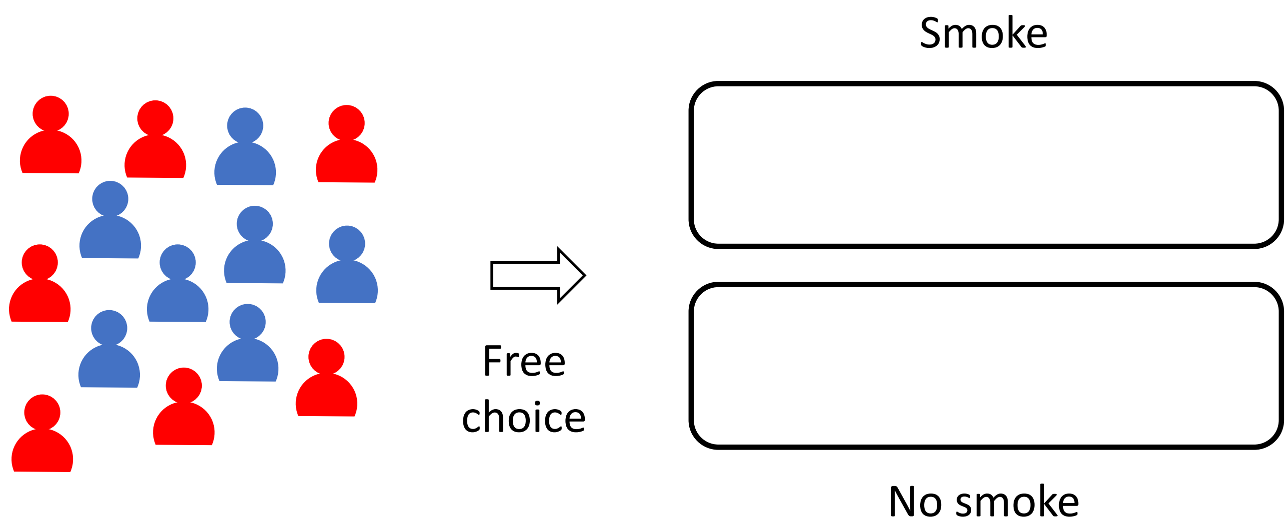


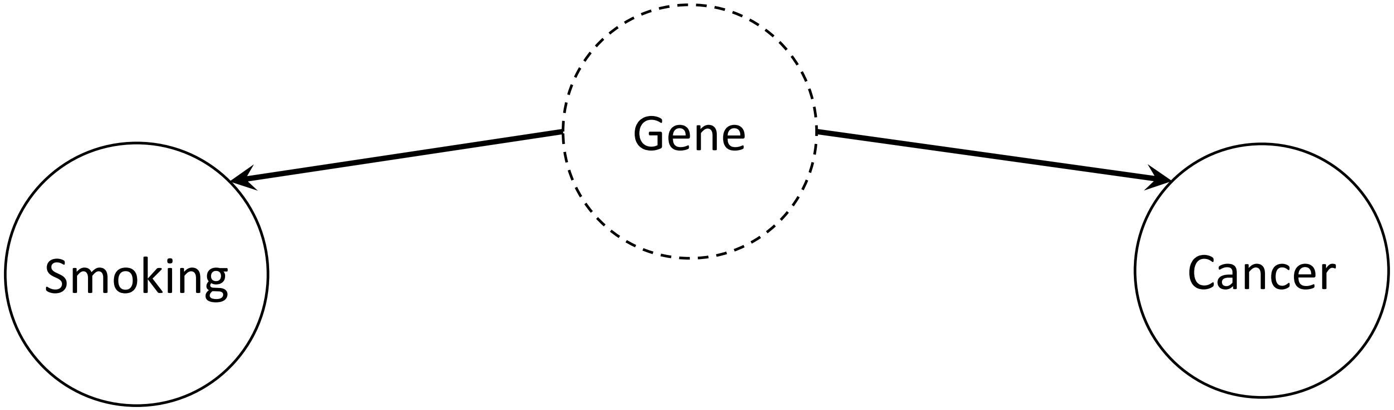
Hypothesis: There are two types of people in the world



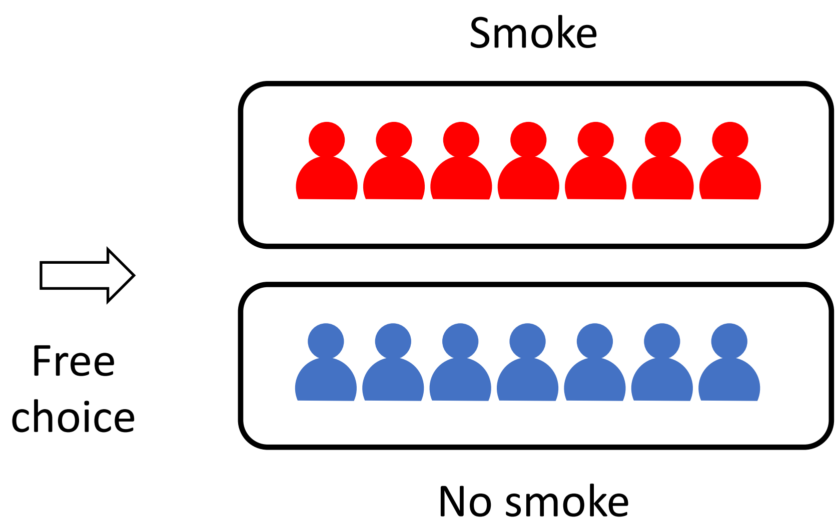


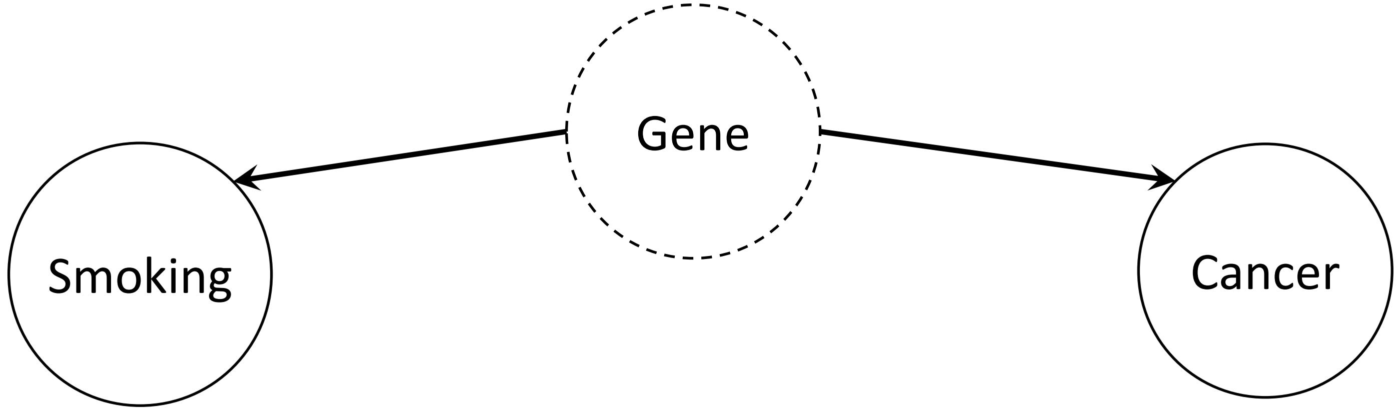
Hypothesis: There are two types of people in the world



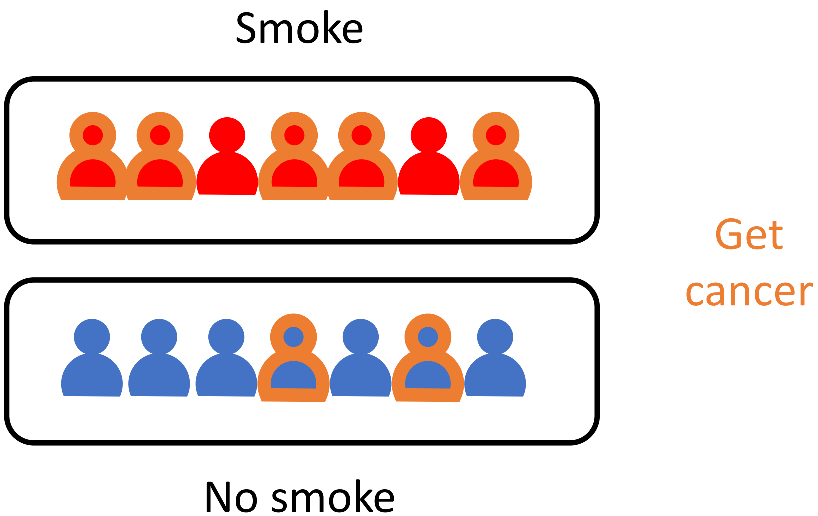


Hypothesis: There are two types of people in the world



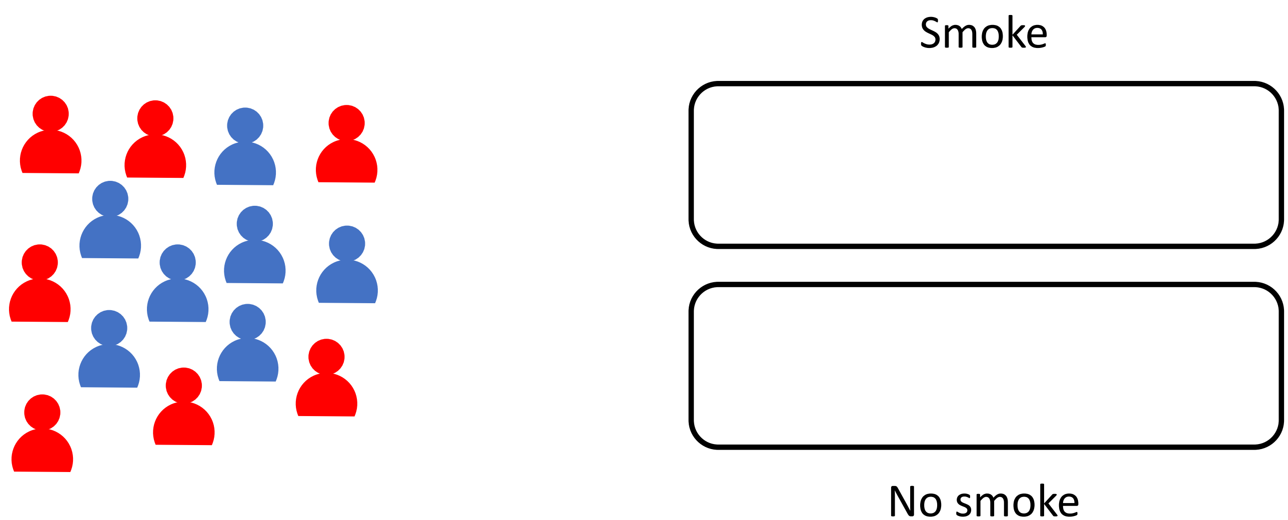


Hypothesis: There are two types of people in the world



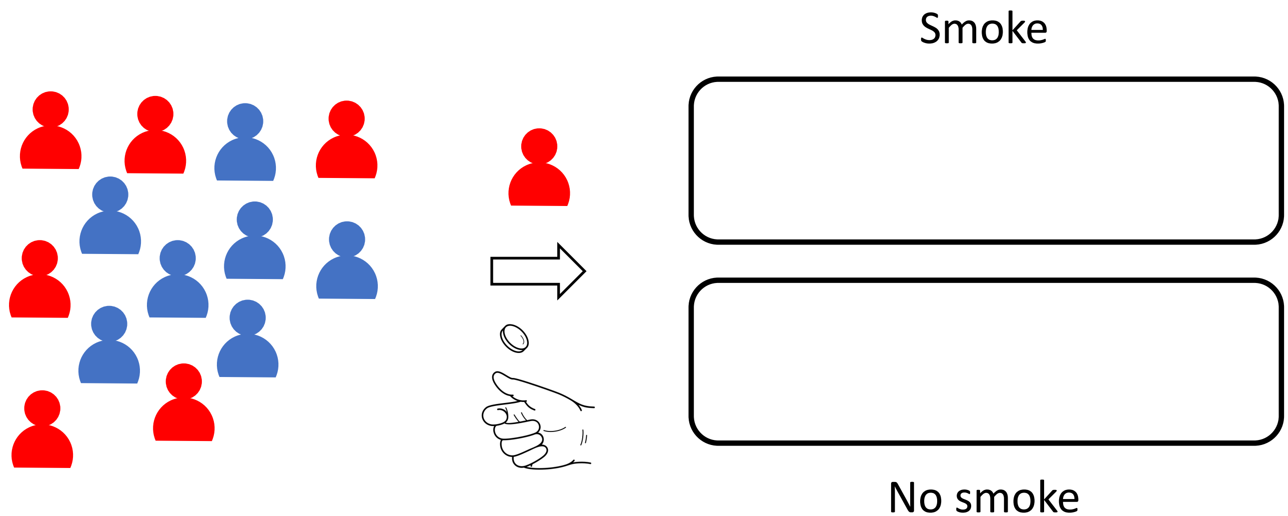
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



Randomized controlled trials

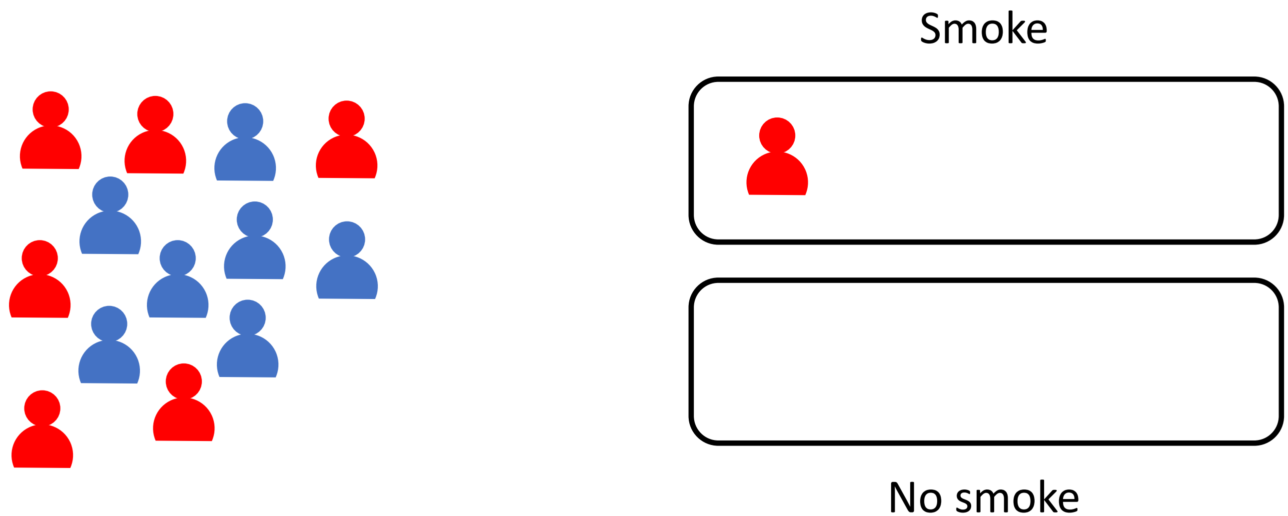
- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



https://thenounproject.com/icon/coin-flipping-2307580/

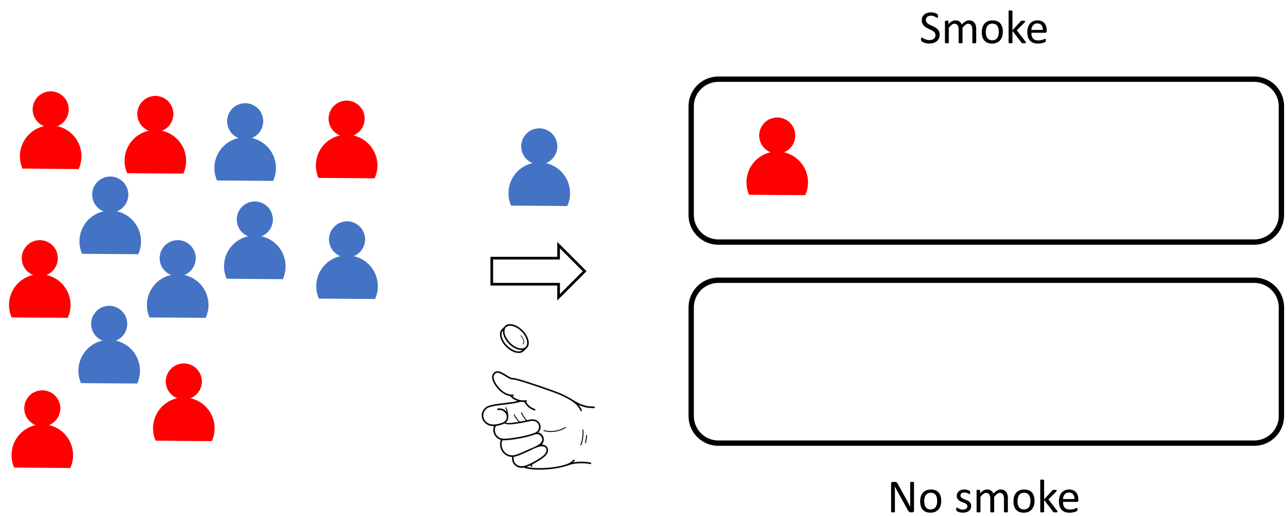
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



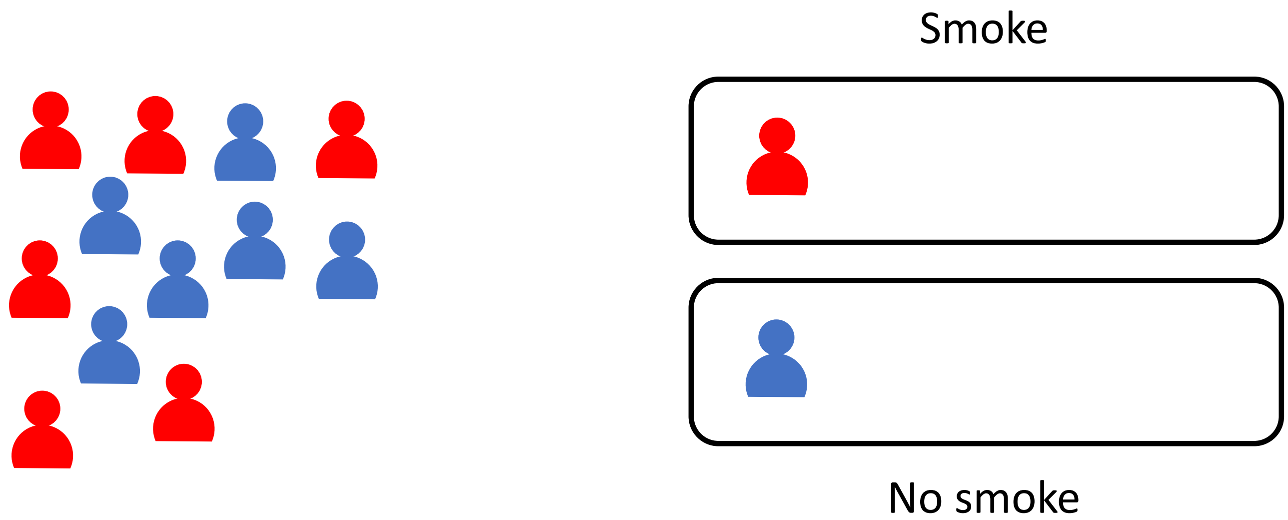
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



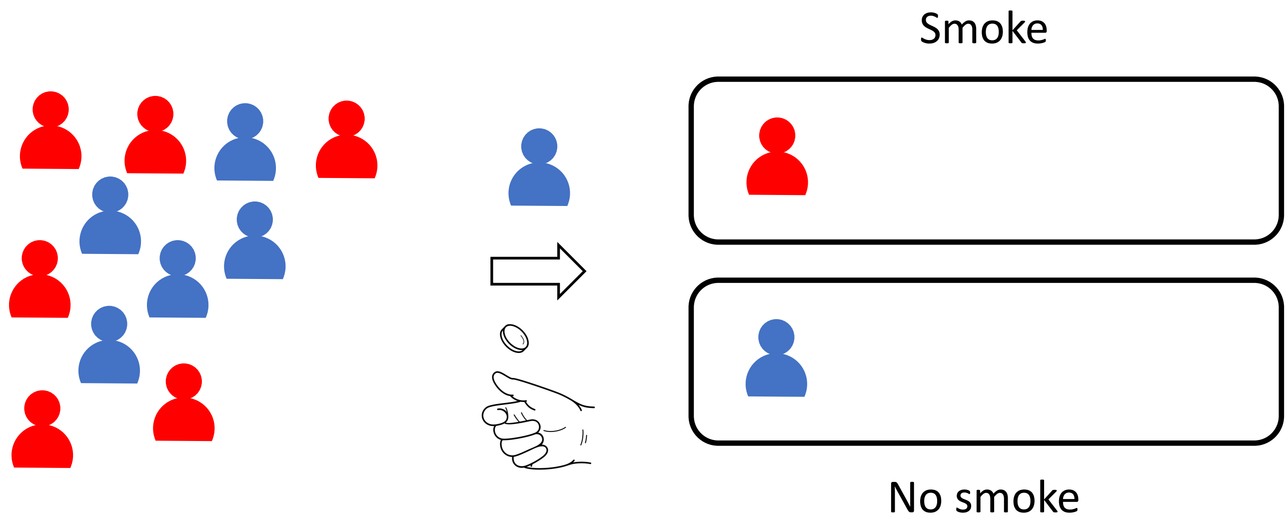
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



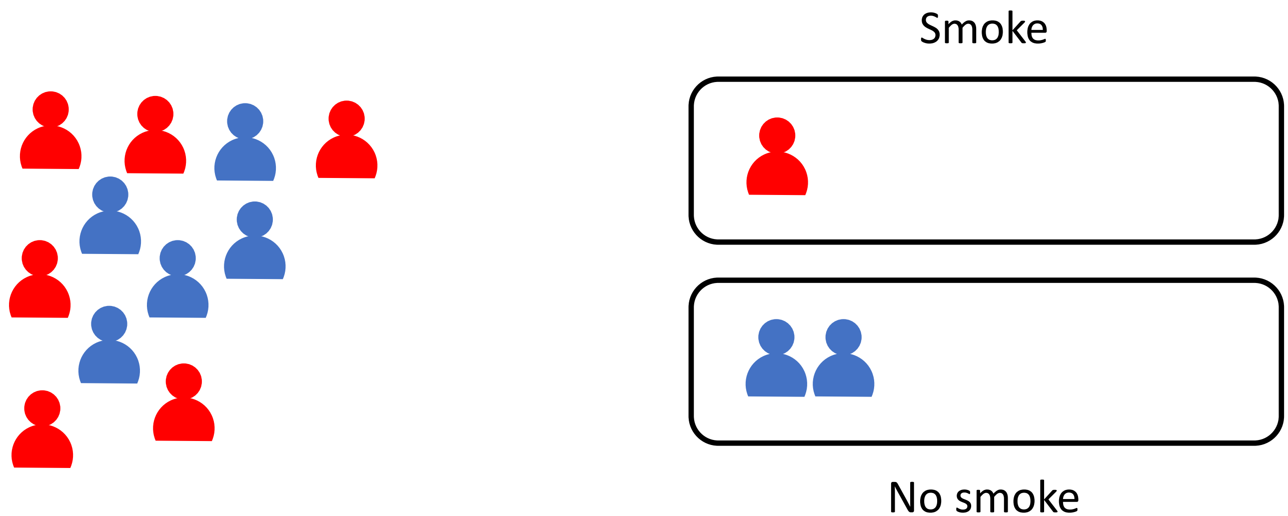
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



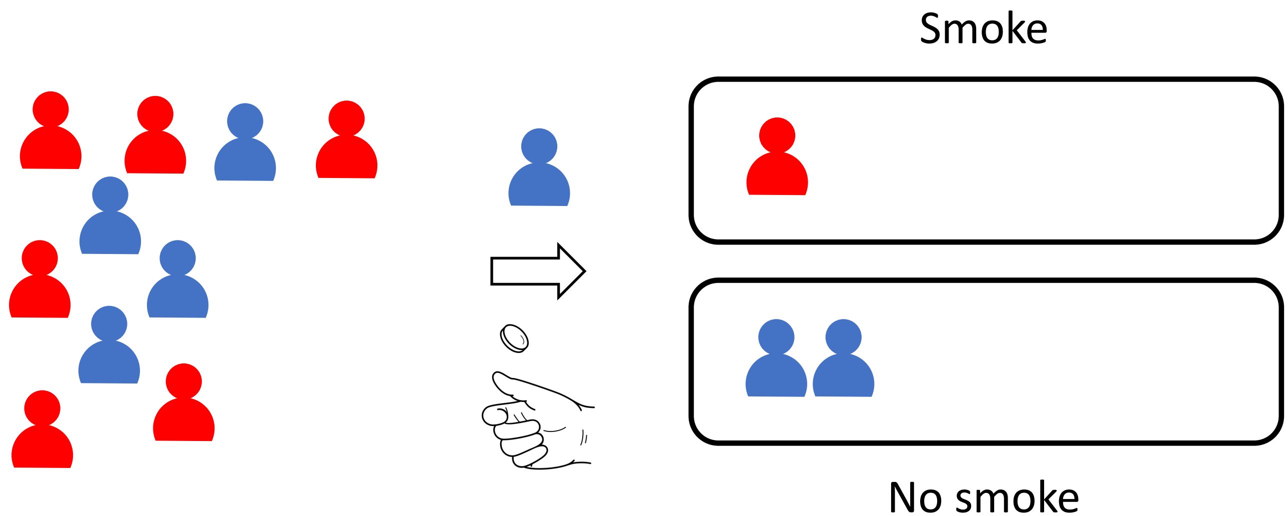
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs $\equiv$ Interventions in causality



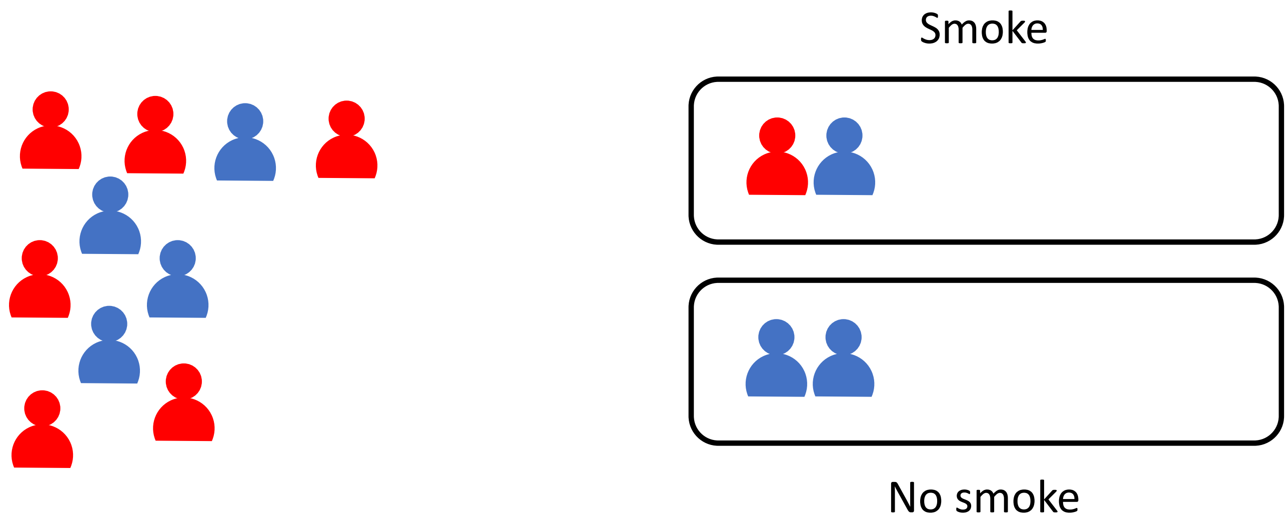
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



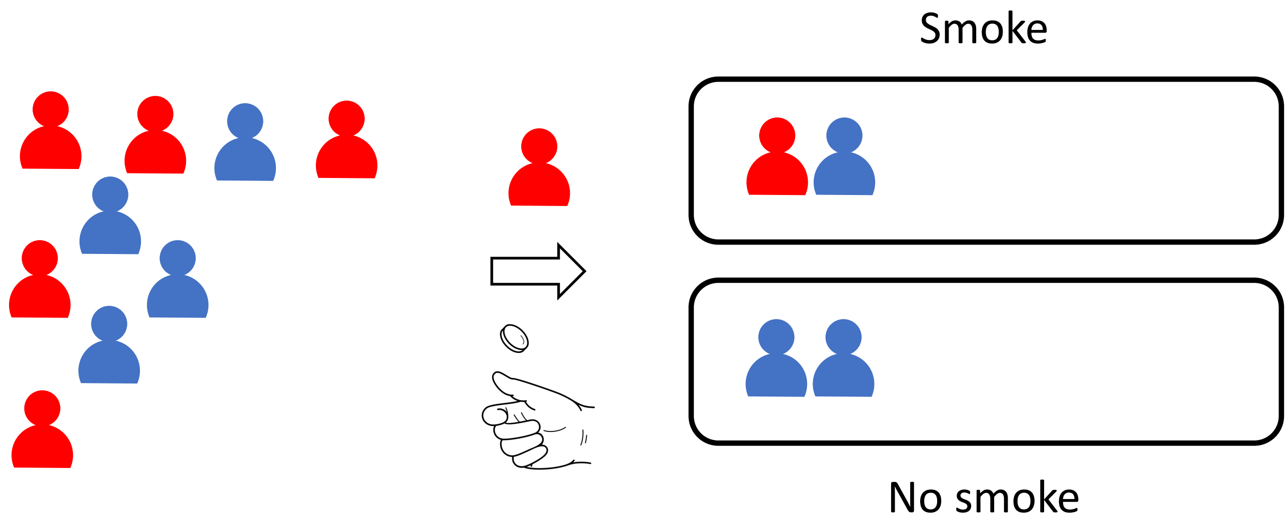
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



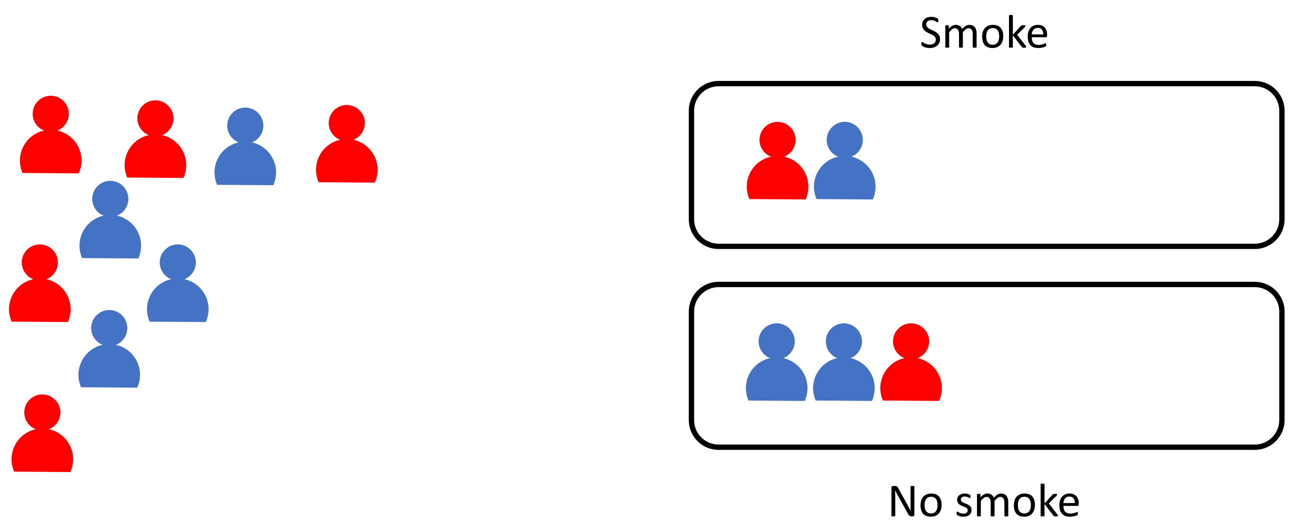
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



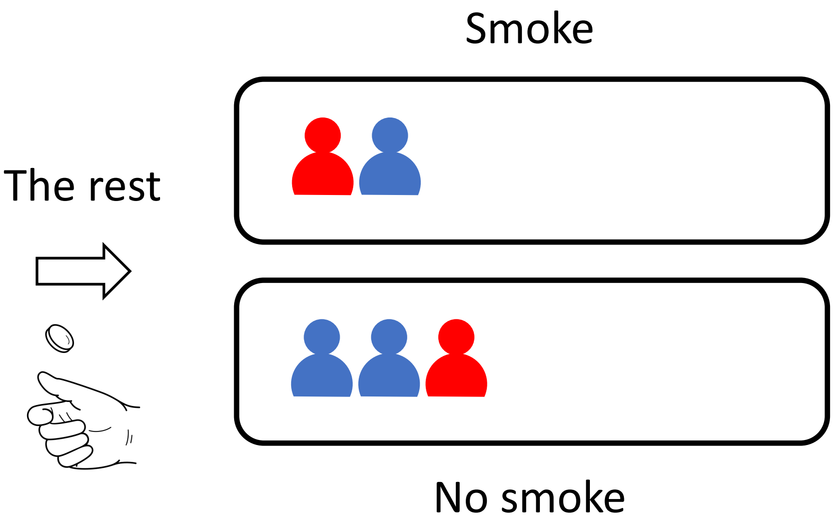
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



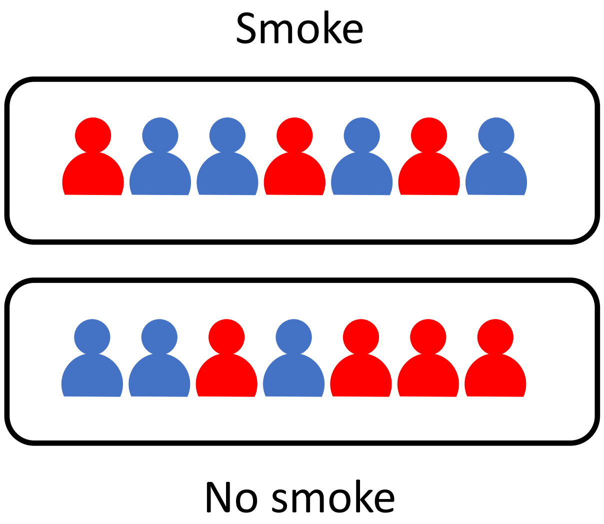
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



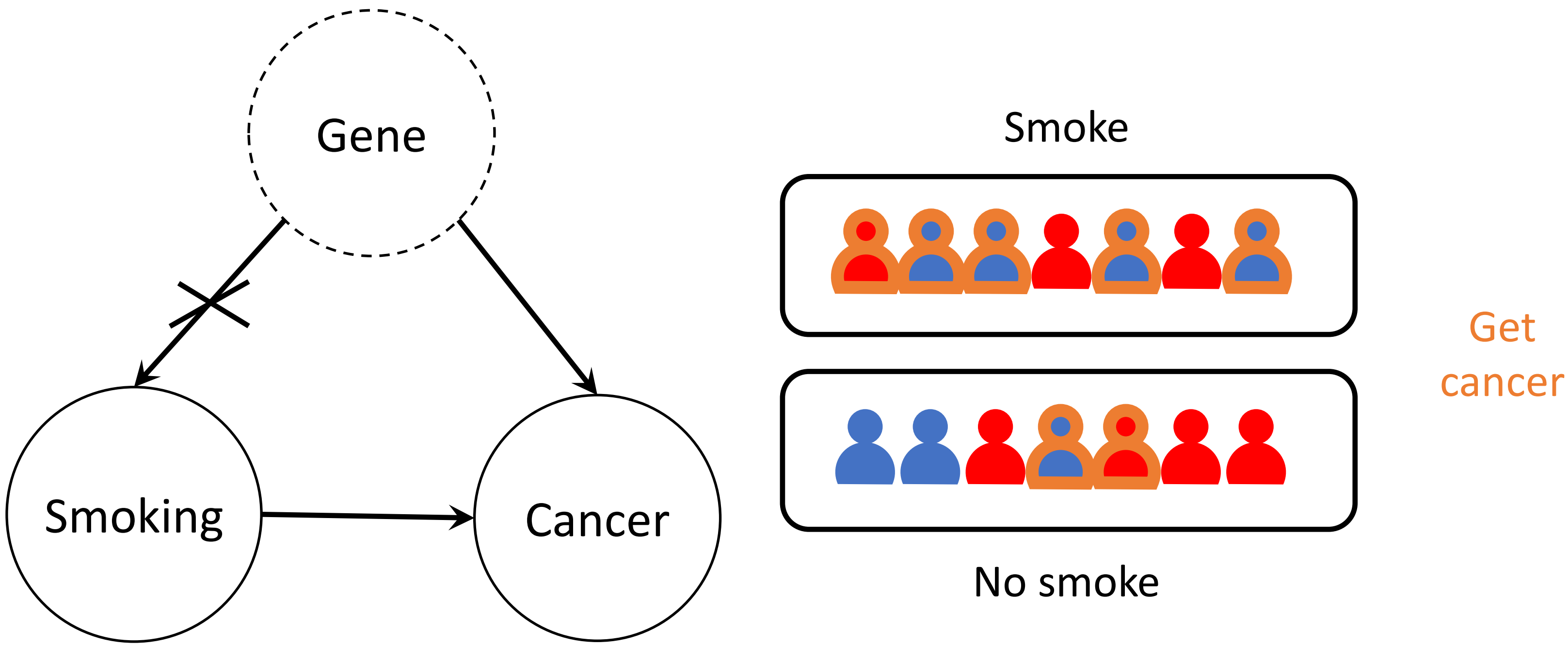
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



Randomized controlled trials

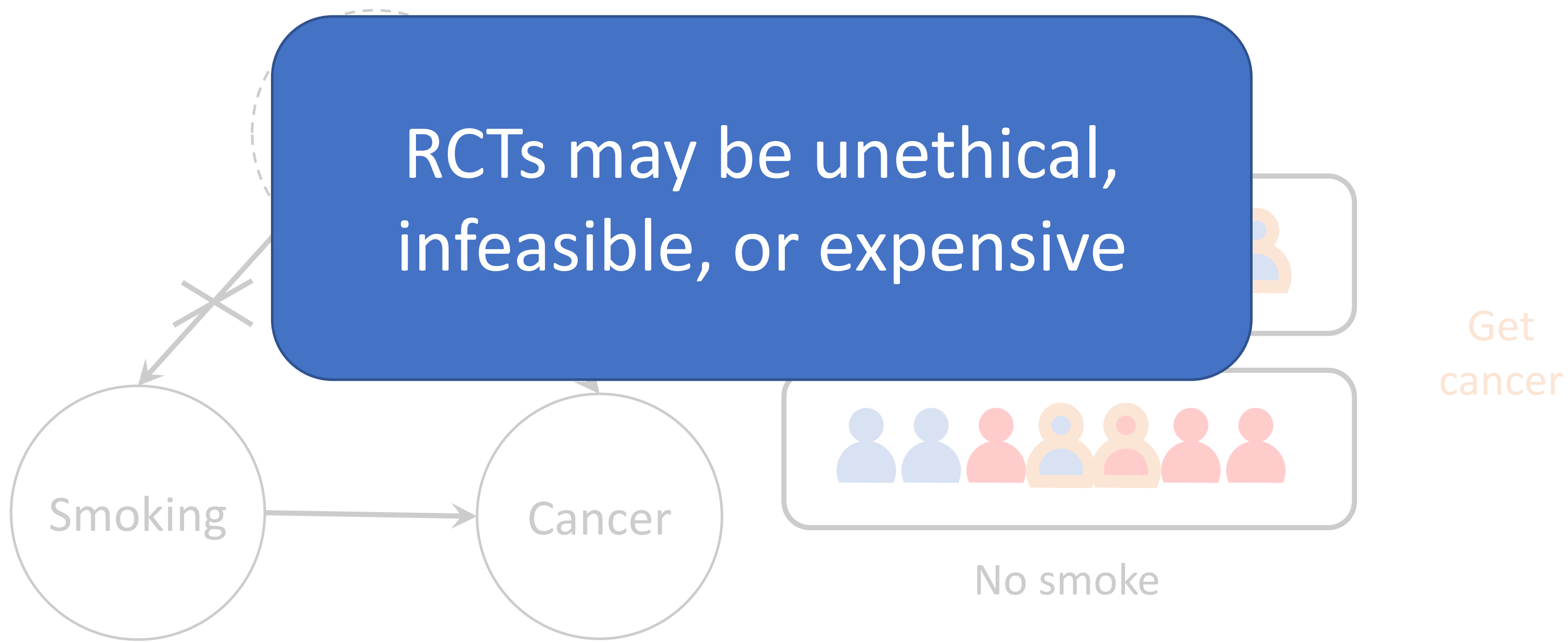
- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



RCT removed causal link from “gene” to “smoking”
If smoking and cancer still highly correlated, then smoking causes cancer

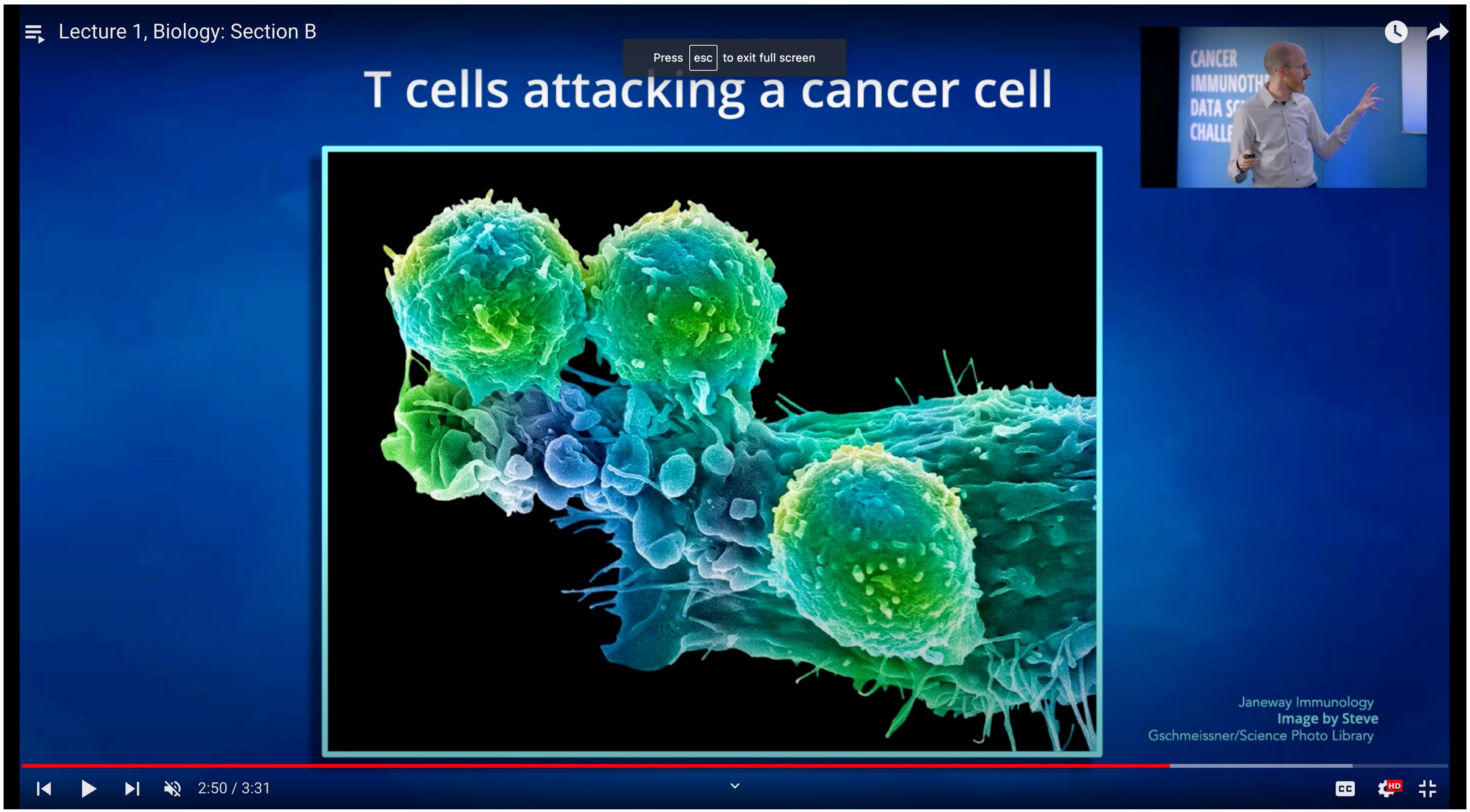
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality

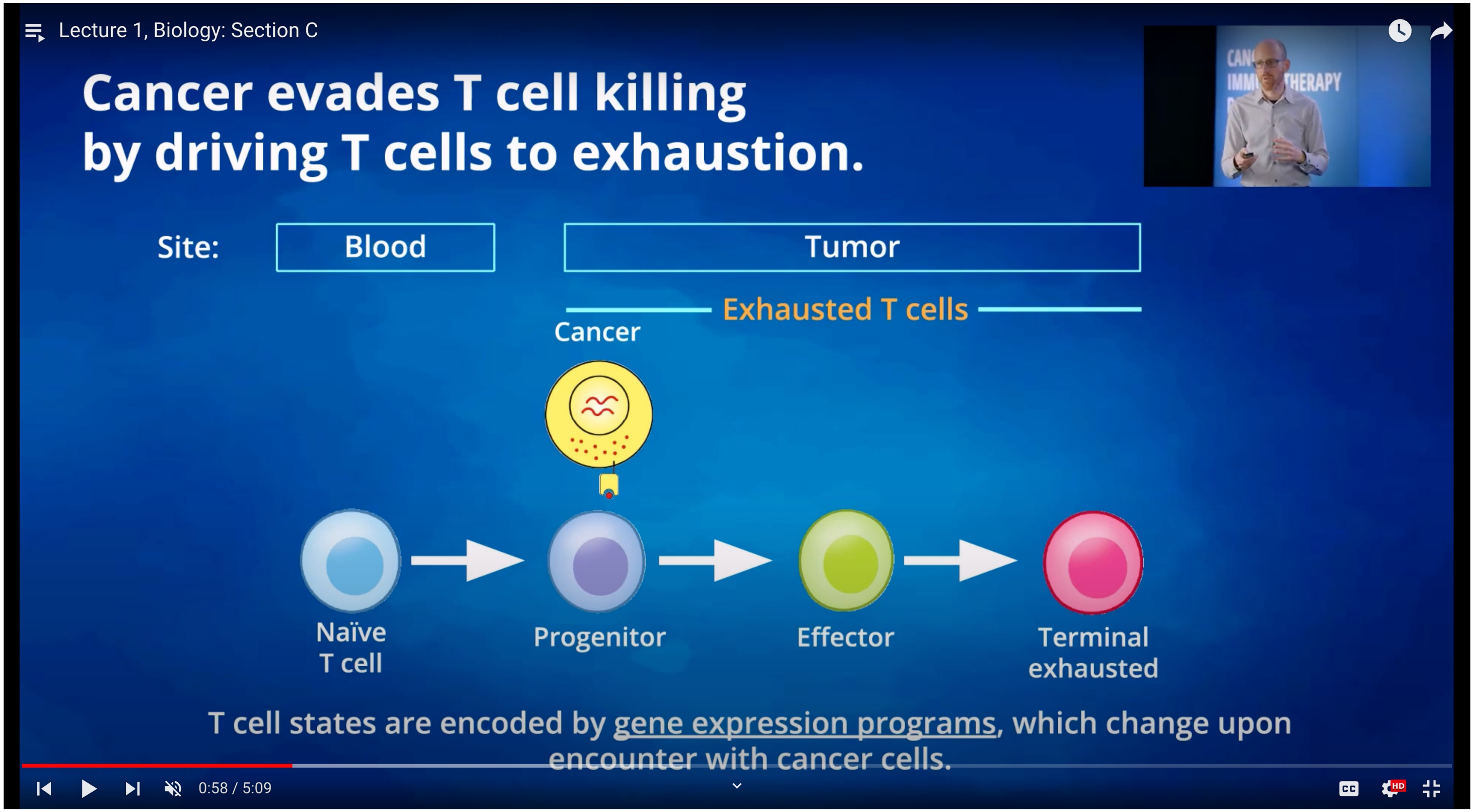


RCTs may be unethical, infeasible, or expensive

RCT removed causal link from “gene” to “smoking”
If smoking and cancer still highly correlated, then smoking causes cancer



https://www.topcoder.com/challenges/25f60820-2e69-444b-bc03-490686af2c87
https://www.youtube.com/playlist?list=PL7Ioc09GvxoqmZOFJUc6ni5WS1FgpiIeN

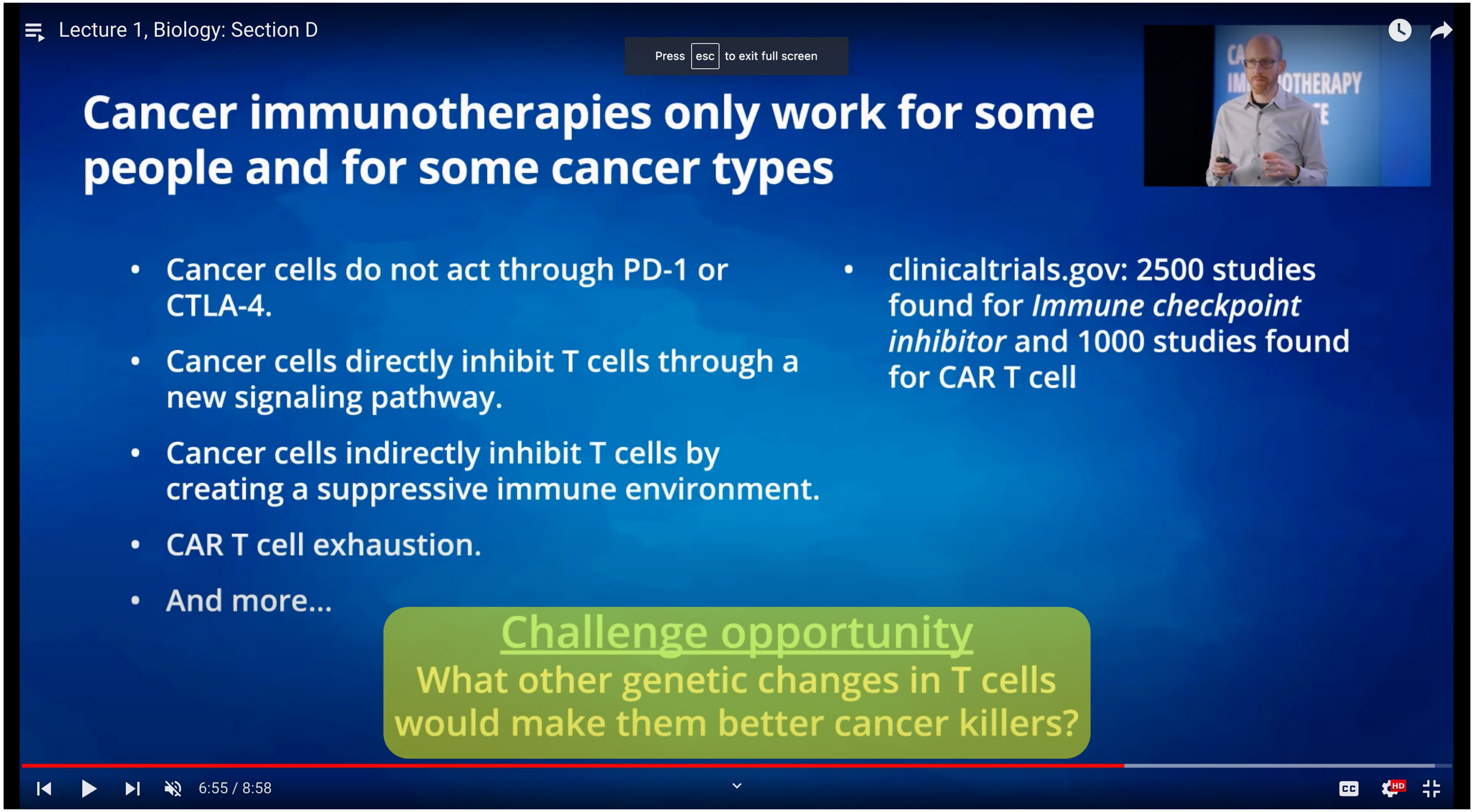


https://www.topcoder.com/challenges/25f60820-2e69-444b-bc03-490686af2c87
https://www.youtube.com/playlist?list=PL7Ioc09GvxoqmZOFJUc6ni5WS1FgpiIeN

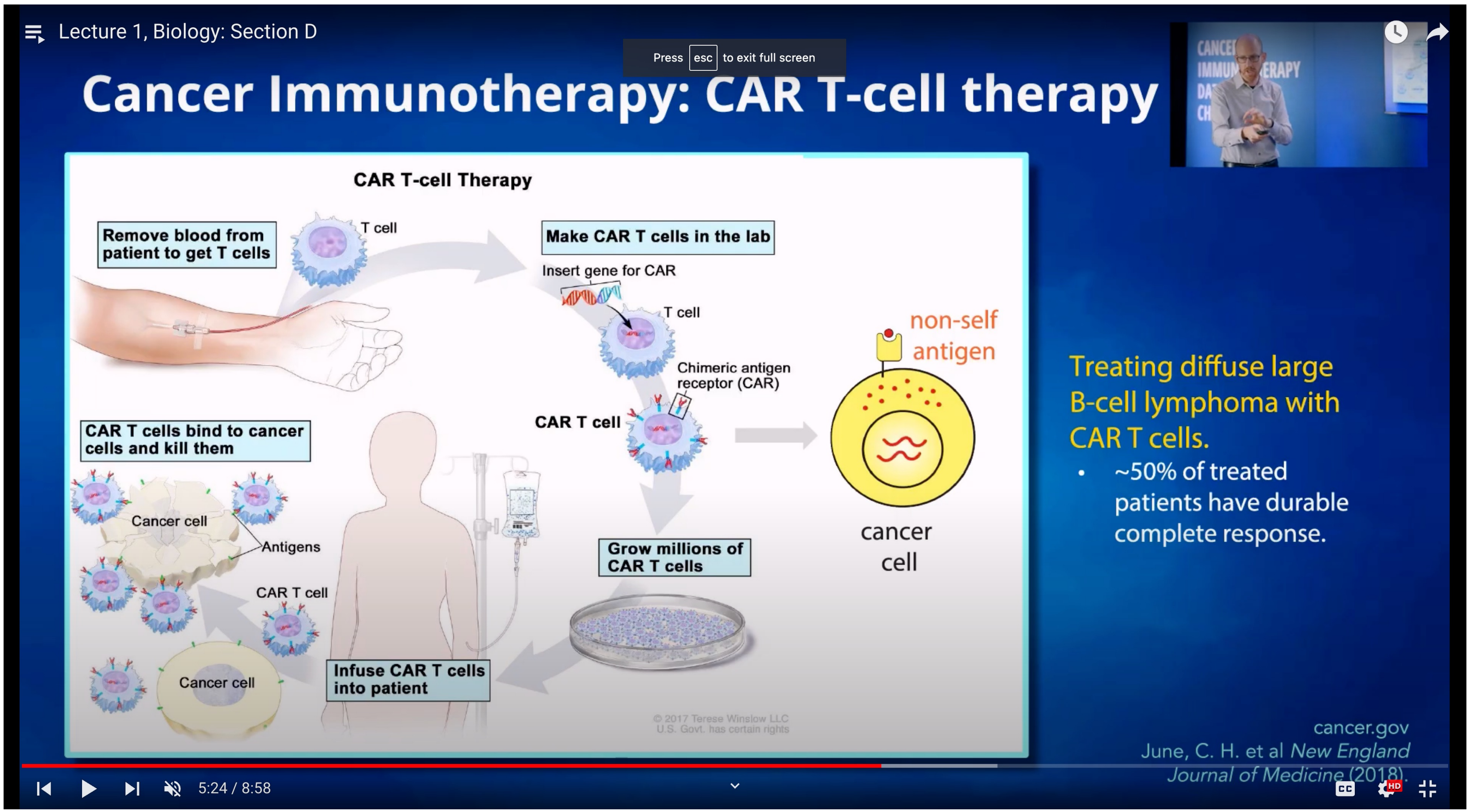
ERIC AND WENDY SCHMIDT CENTER AT BROAD INSTITUTE

CANCER IMMUNOTHERAPY DATA SCIENCE GRAND CHALLENGE

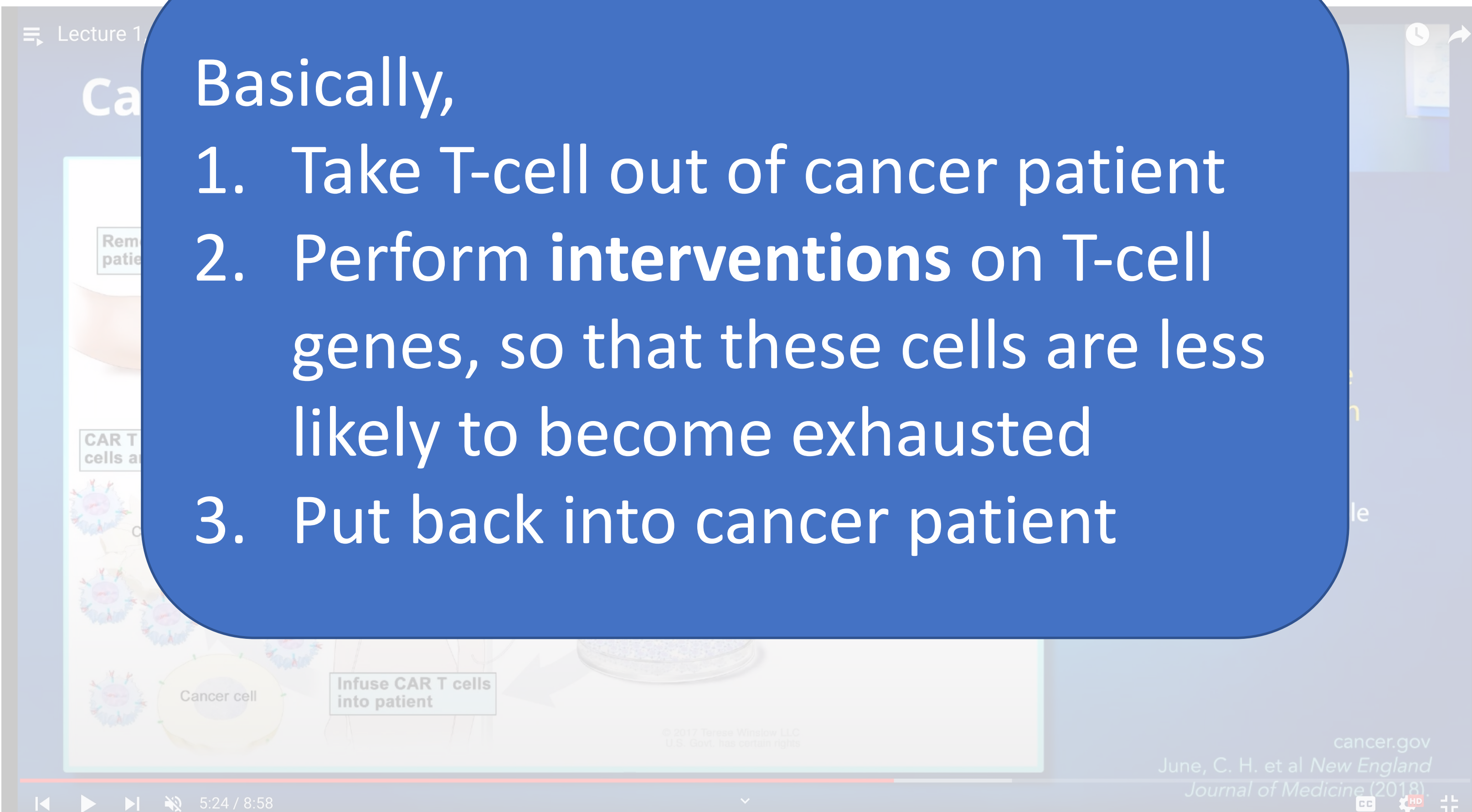
2023



https://www.topcoder.com/challenges/25f60820-2e69-444b-bc03-490686af2c87
https://www.youtube.com/playlist?list=PL7Ioc09GvxoqmZOFJUc6ni5WS1FgpiIeN



https://www.topcoder.com/challenges/25f60820-2e69-444b-bc03-490686af2c87
https://www.youtube.com/playlist?list=PL7Ioc09GvxoqmZOFJUc6ni5WS1FgpiIeN



Basically,

1. Take T-cell out of cancer patient
2. Perform **interventions** on T-cell genes, so that these cells are less likely to become exhausted
3. Put back into cancer patient

https://www.topcoder.com/challenges/25f60820-2e69-444b-bc03-490686af2c87
https://www.youtube.com/playlist?list=PL7loc09GvxoqmZOFJUc6ni5WS1FgpileN

NEWS | 07 October 2020

Pioneers of revolutionary CRISPR gene editing win chemistry Nobel

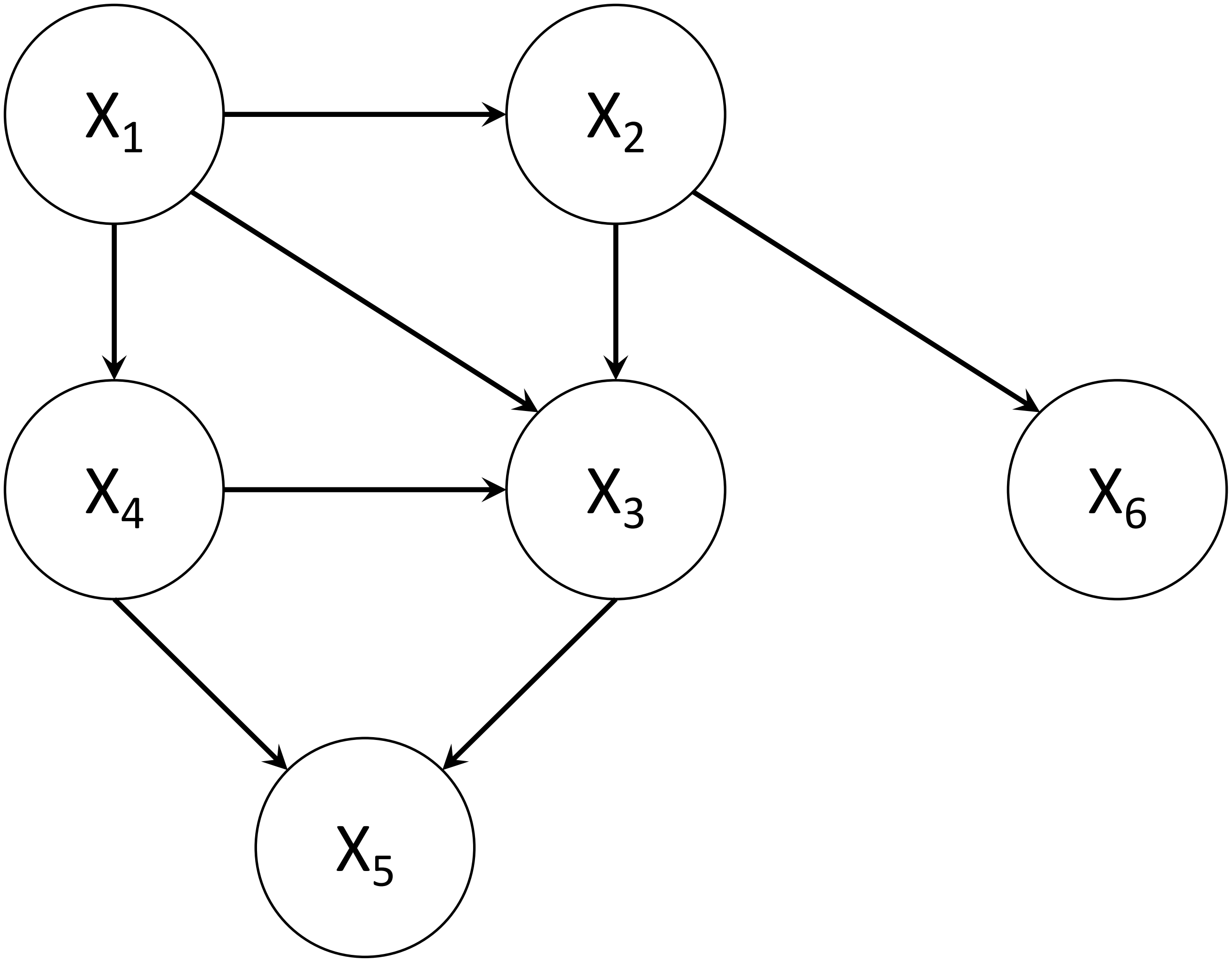
Emmanuelle Charpentier and Jennifer Doudna share the award for developing the precise genome-editing technology.

Heidi Ledford & Ewen Callaway

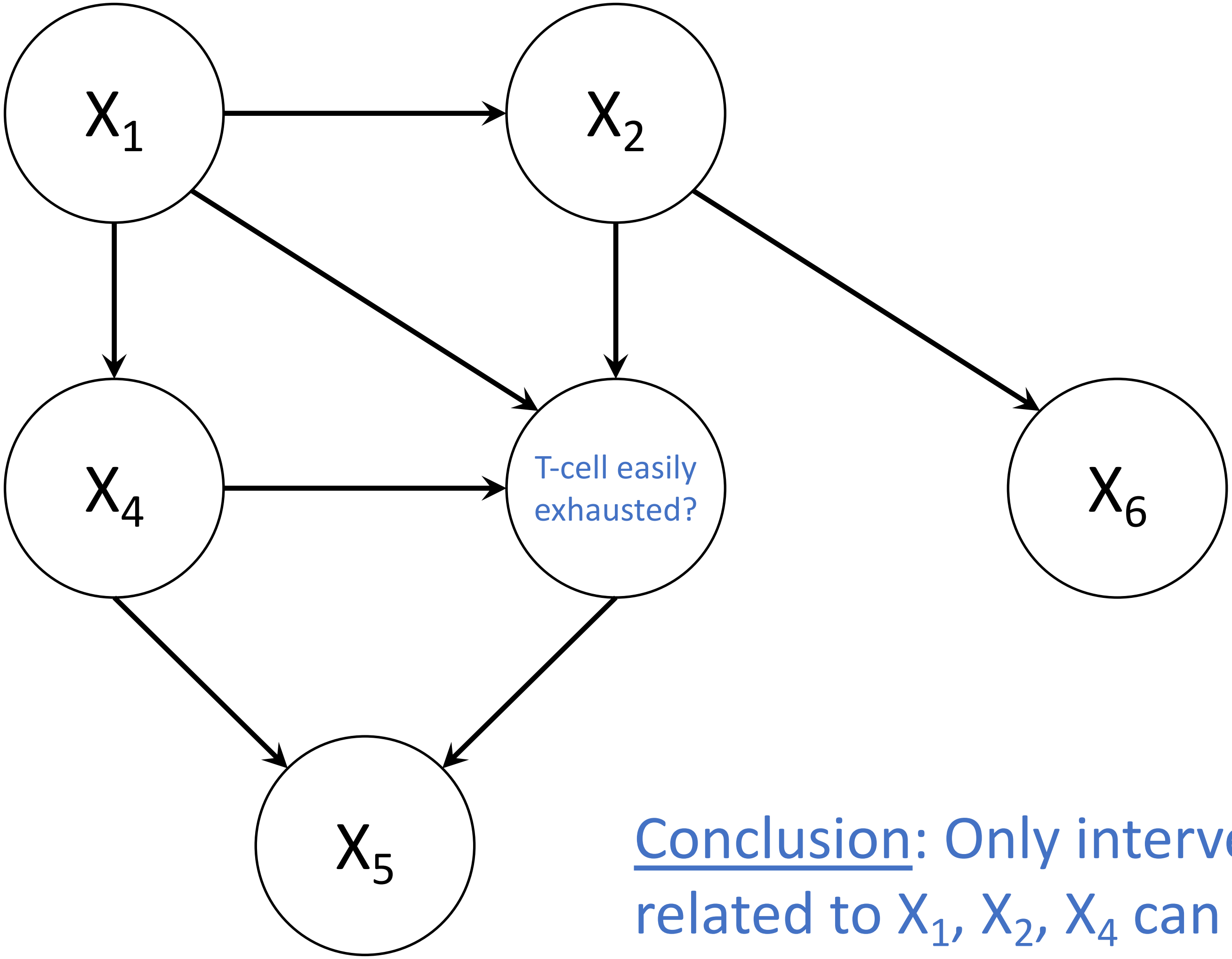
Jennifer Doudna and Emmanuelle Charpentier share the 2020 Nobel chemistry prize for their discovery of a game-changing gene-editing technique. Credit: Alexander Heinel/Picture Alliance/DPA

https://www.nature.com/articles/d41586-020-02765-9

Why structure learning

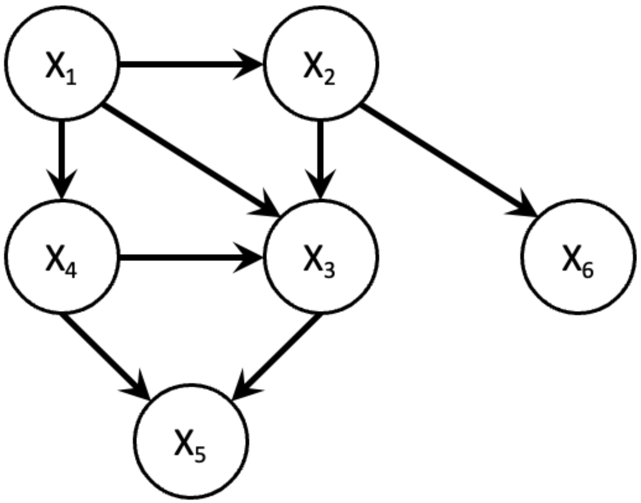


Why structure learning

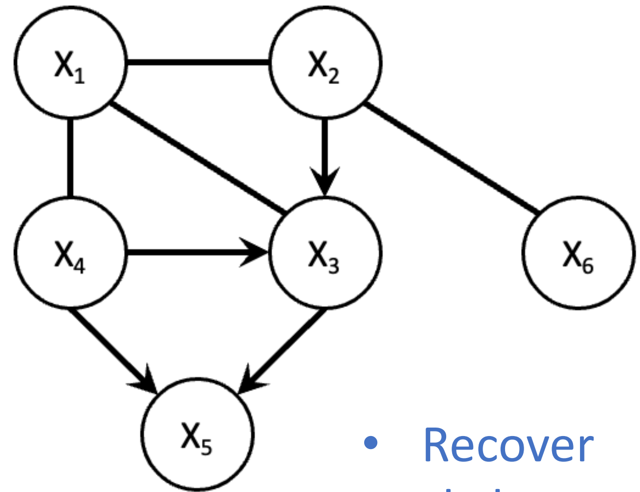


Conclusion: Only interventions related to X_1, X_2, X_4 can affect our objective of interest

Structure learning (simplified)



This represents an equivalence class of graphs

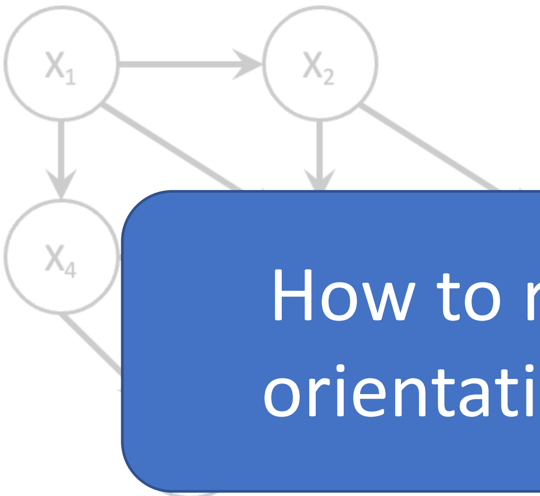


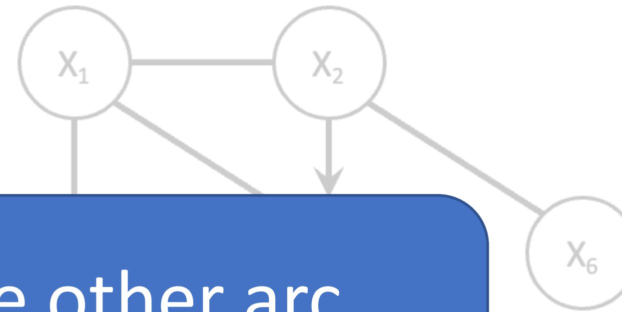
- Recover skeleton
- Orient *some* edges

Get samples

	X_1	X_2	X_3	X_4	X_5	X_6
Sample 1	0.22	0.04	0.84	0.48	0.98	0.82
Sample 2	0.87	0.17	0.61	0.67	0.67	0.23
Sample 3	0.55	0.54	0.67	0.86	0.93	0.23
…	…	…	…	…	…	…
Sample M	0.12	0.95	0.79	0.47	0.05	0.92

Structure learning (simplified)





How to recover all the other arc orientations? Use interventions!

Get samples

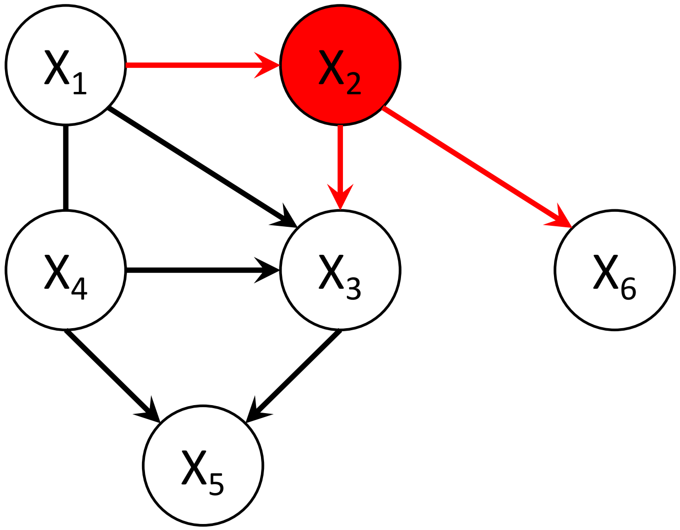
cover skeleton

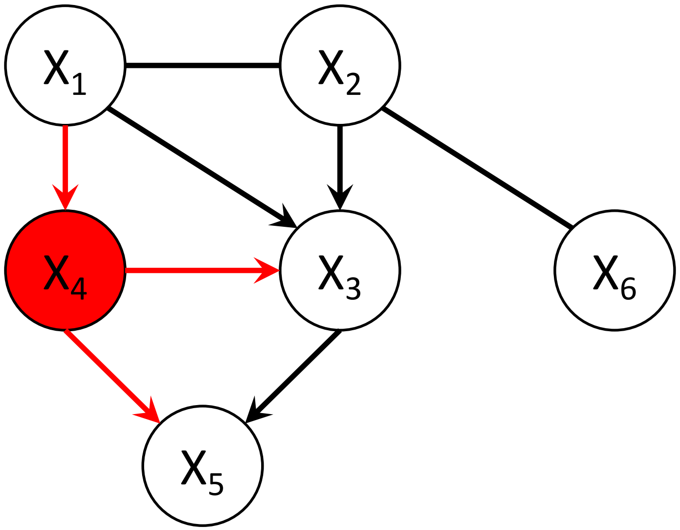
- Orient *some* edges

	X_1	X_2	X_3	X_4	X_5	X_6
Sample 1	0.22	0.04	0.84	0.48	0.98	0.82
Sample 2	0.87	0.17	0.61	0.67	0.67	0.23
Sample 3	0.55	0.54	0.67	0.86	0.93	0.23
…	…	…	…	…	…	…
Sample M	0.12	0.95	0.79	0.47	0.05	0.92

What do interventions give us?

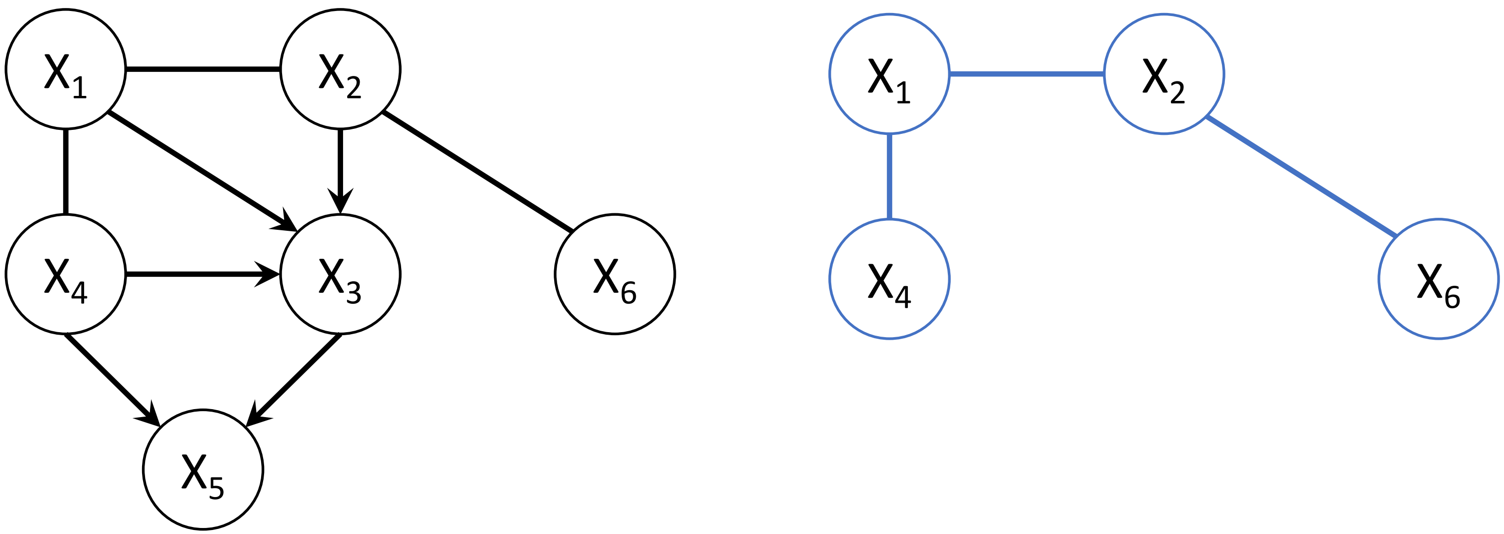
- When we intervene on a vertex, we recover the orientations of edges incident to the vertex





What do interventions give us?

- When we intervene on a vertex, we recover the orientations of edges incident to the vertex



- Naïve: Compute **minimum vertex cover** on subgraph induced by unoriented arcs

Meek rules [Meek 1995]

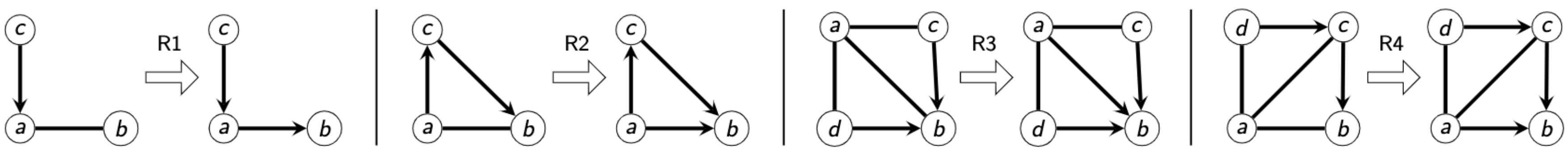
- **Sound** and **complete**
  (with respect to arc orientations with acyclic completion)

We won't miss out on any information

We won't wrongly orient arcs

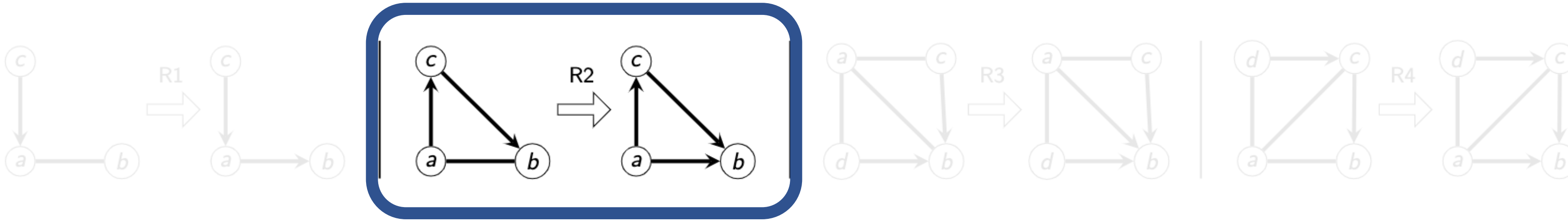
Meek rules [Meek 1995]

- **Sound** and **complete**
  (with respect to arc orientations with acyclic completion)



Meek rules [Meek 1995]

- **Sound** and **complete**
  (with respect to arc orientations with acyclic completion)

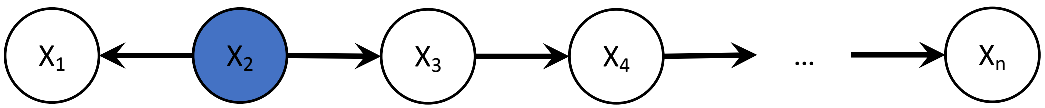


If $b \leftarrow a$, then cycle

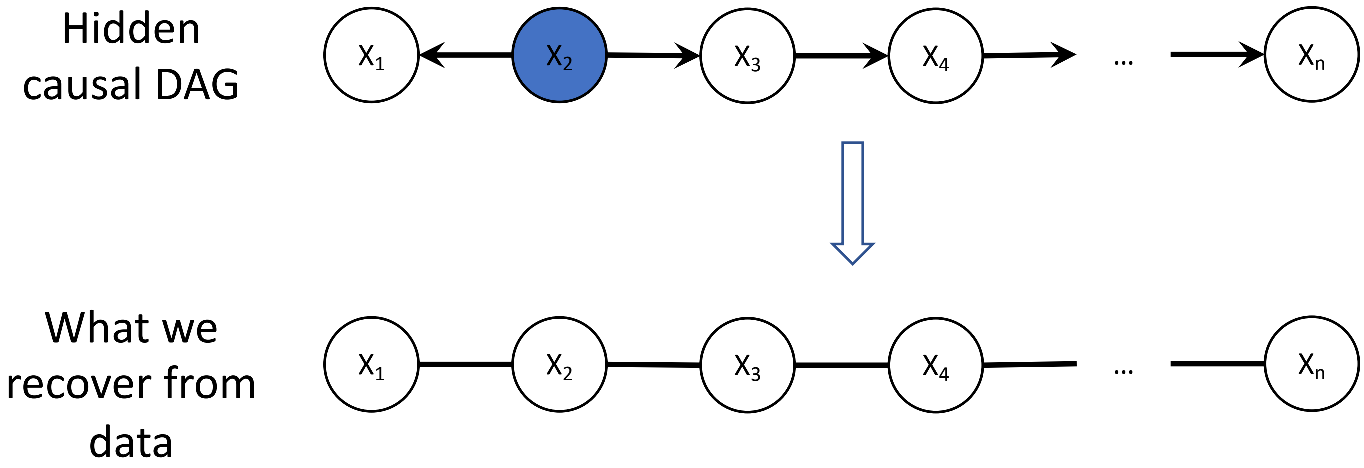
- Converge in polynomial time [Wienöbst, Bannach, Liśkiewicz 2021]

A simple causal graph example

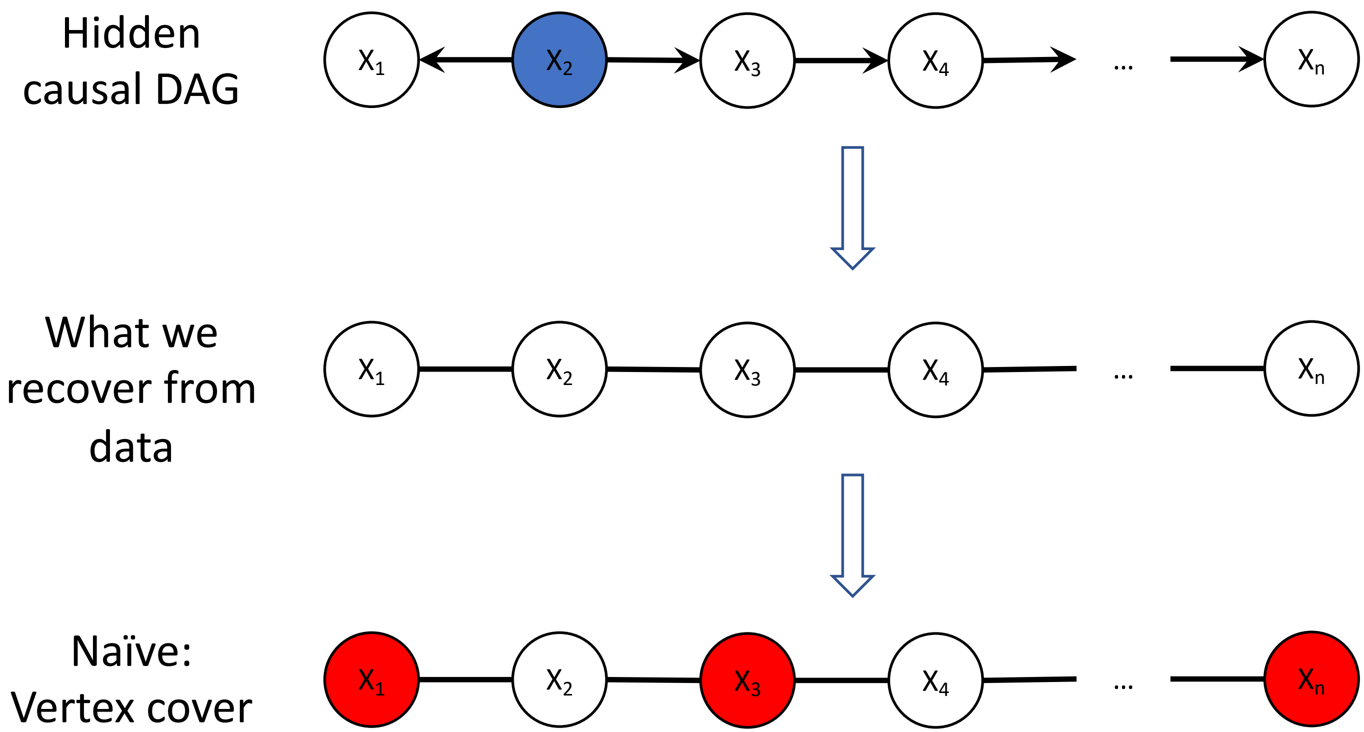
Hidden causal DAG



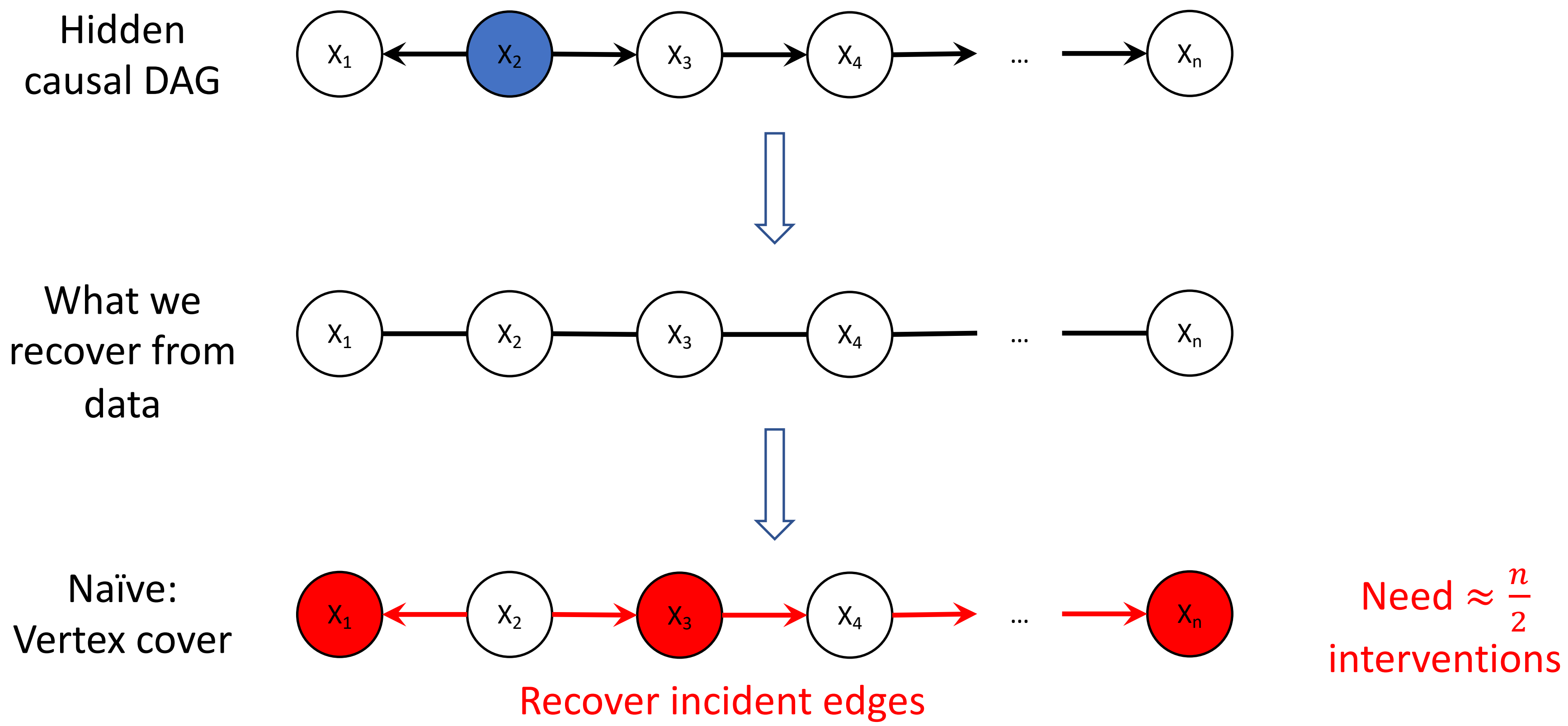
A simple causal graph example



A simple causal graph example

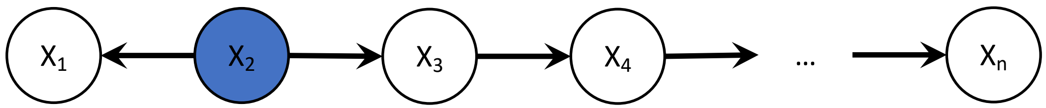


A simple causal graph example

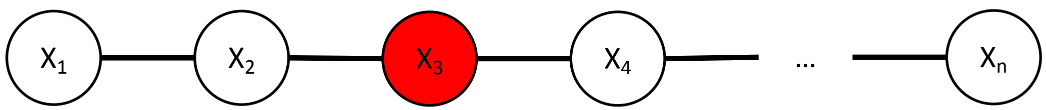


A simple causal graph example

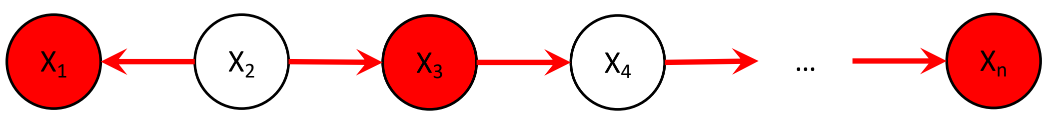
Hidden causal DAG



Suppose we intervene X_3



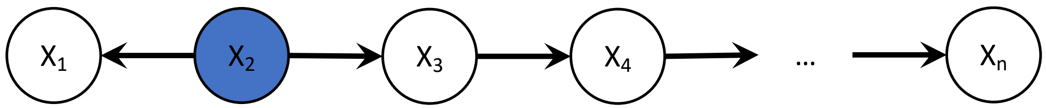
Naïve: Vertex cover



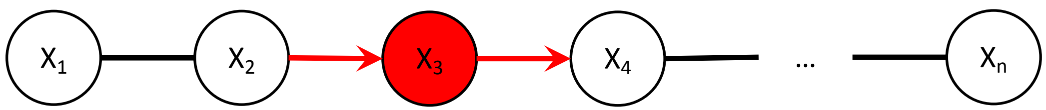
Need $\approx \frac{n}{2}$ interventions

A simple causal graph example

Hidden causal DAG

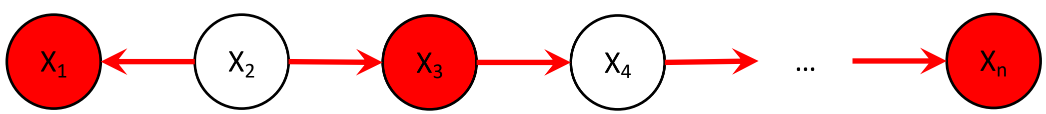


Suppose we intervene X_3



Recover incident edges

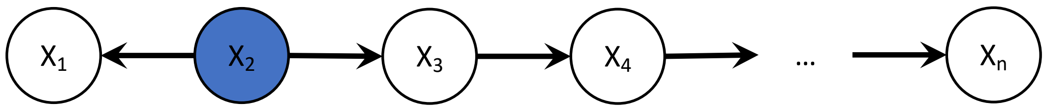
Naïve: Vertex cover



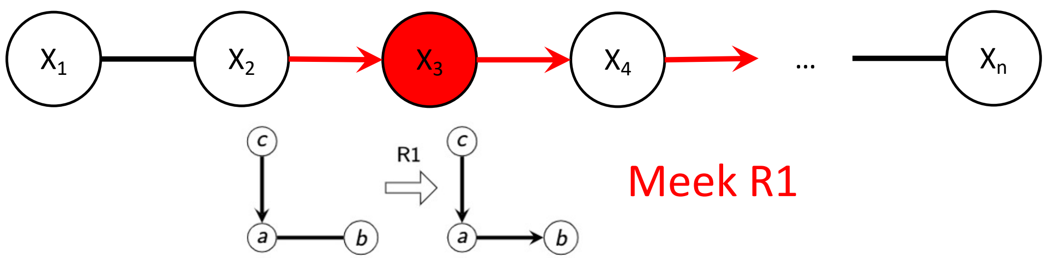
Need $\approx \frac{n}{2}$ interventions

A simple causal graph example

Hidden causal DAG

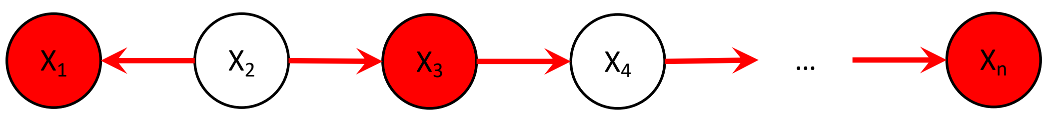


Suppose we intervene X_3



Meek R1

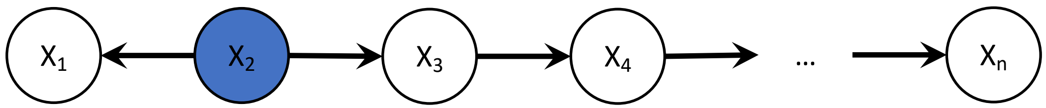
Naïve: Vertex cover



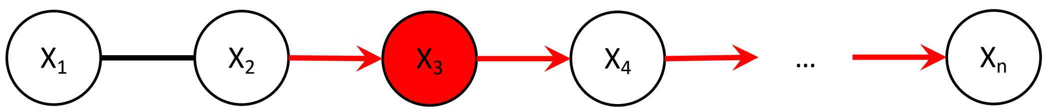
Need $\approx \frac{n}{2}$ interventions

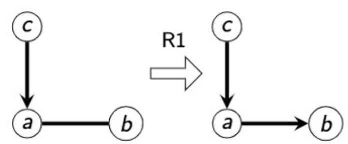
A simple causal graph example

Hidden causal DAG



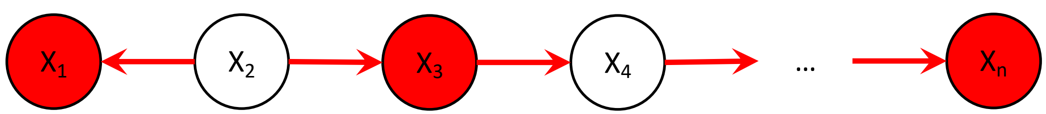
Suppose we intervene X_3





Meek R1

Naïve: Vertex cover



Need $\approx \frac{n}{2}$ interventions

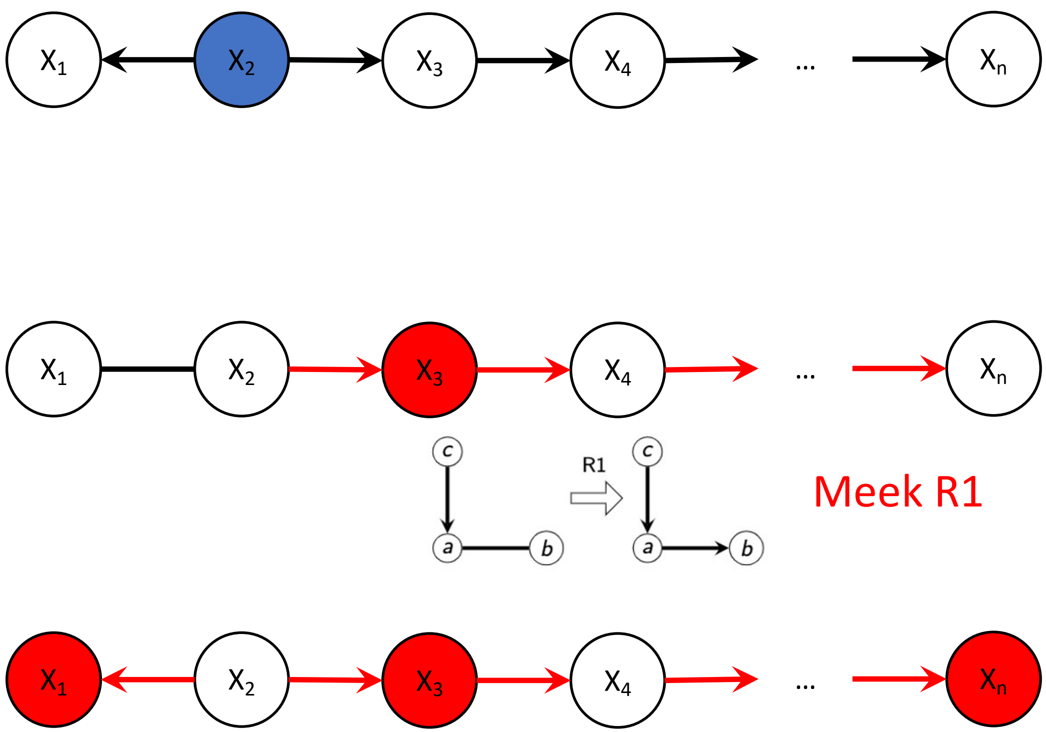
A si

Question: How many interventions do we need for this example?

Hidden causal DAG

Suppose we intervene X_3

Naïve: Vertex cover

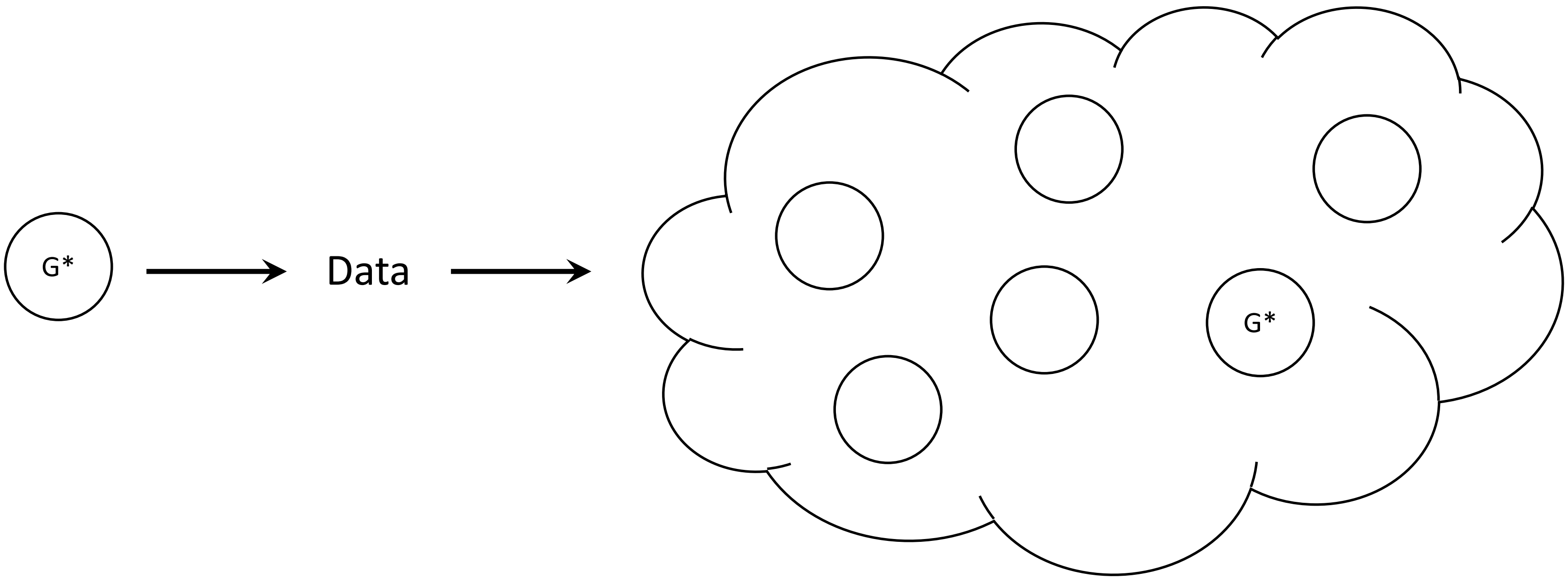


Meek R1

Need $\approx \frac{n}{2}$ interventions

Searching using adaptive interventions

Identify G*

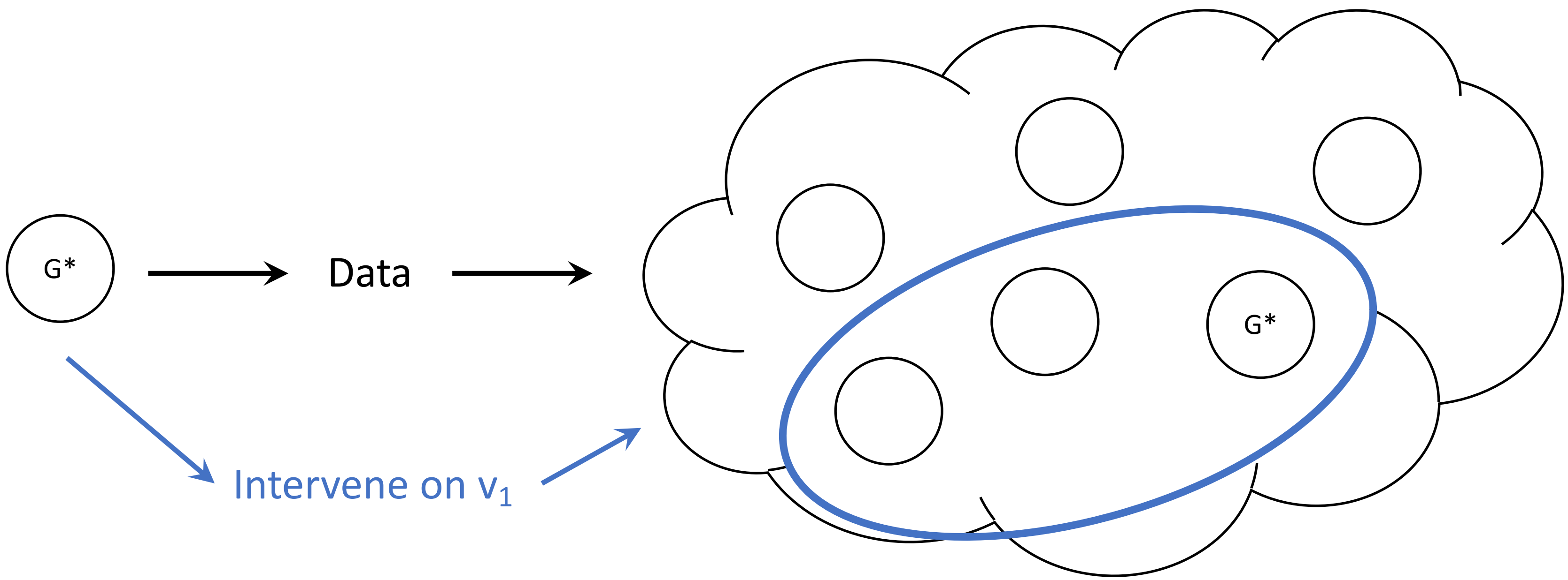


Equivalence class of causal graphs

Can be represented by a partially oriented causal graph

Searching using adaptive interventions

Identify G* using **interventions**

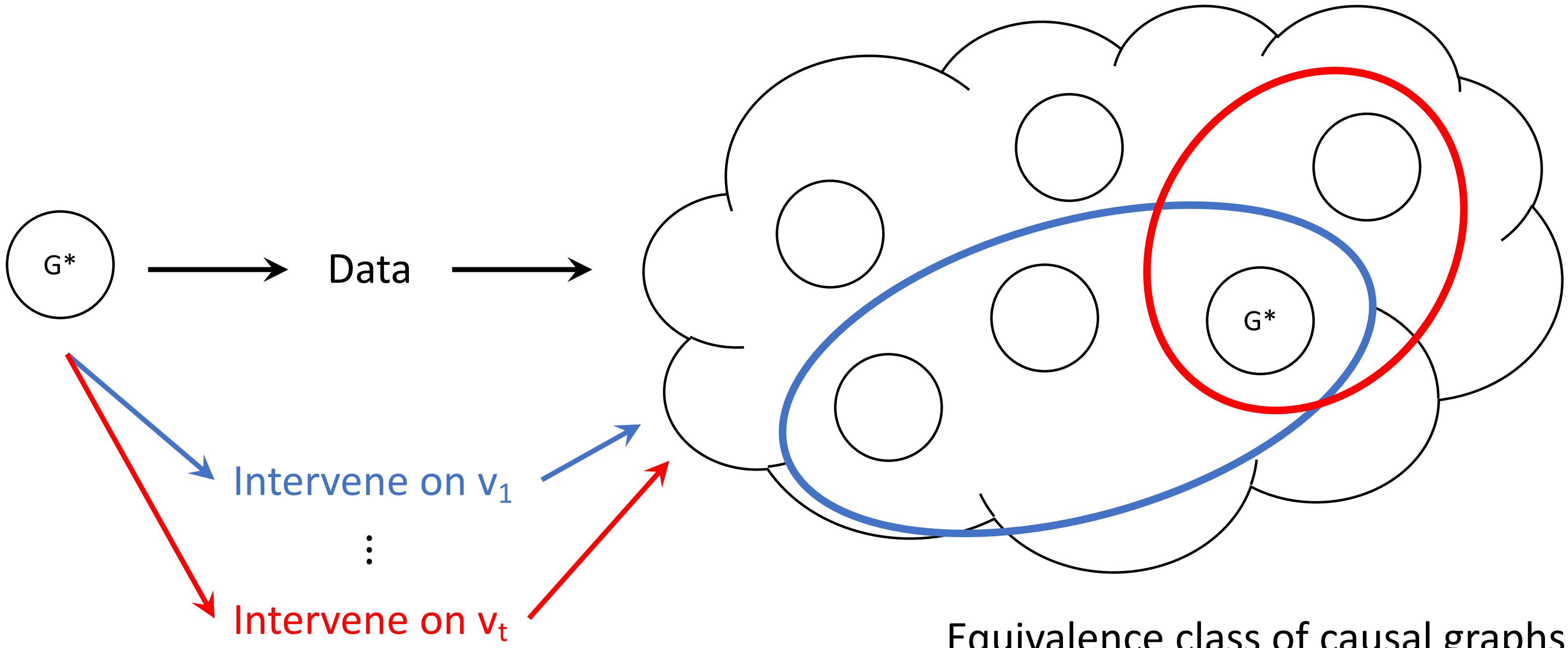


Equivalence class of causal graphs

Can be represented by a partially oriented causal graph

Searching using adaptive interventions

Identify G* using **interventions**

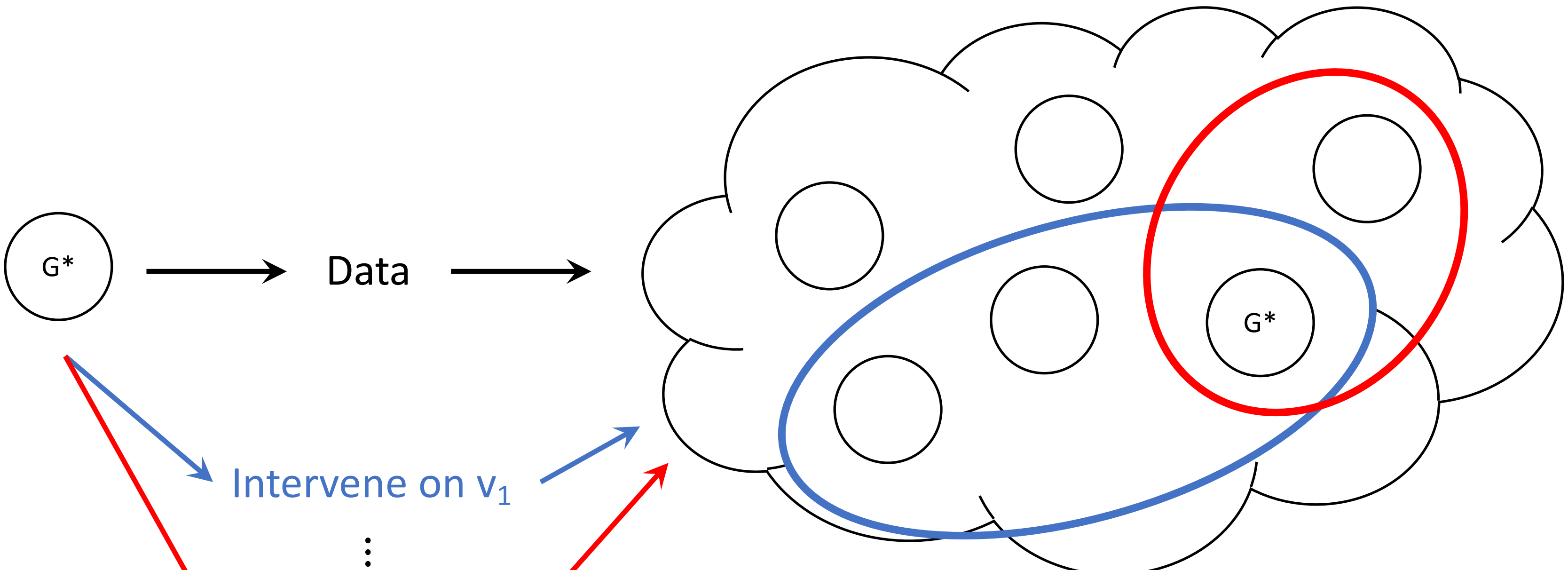


Equivalence class of causal graphs

Can be represented by a partially oriented causal graph

Searching using adaptive interventions

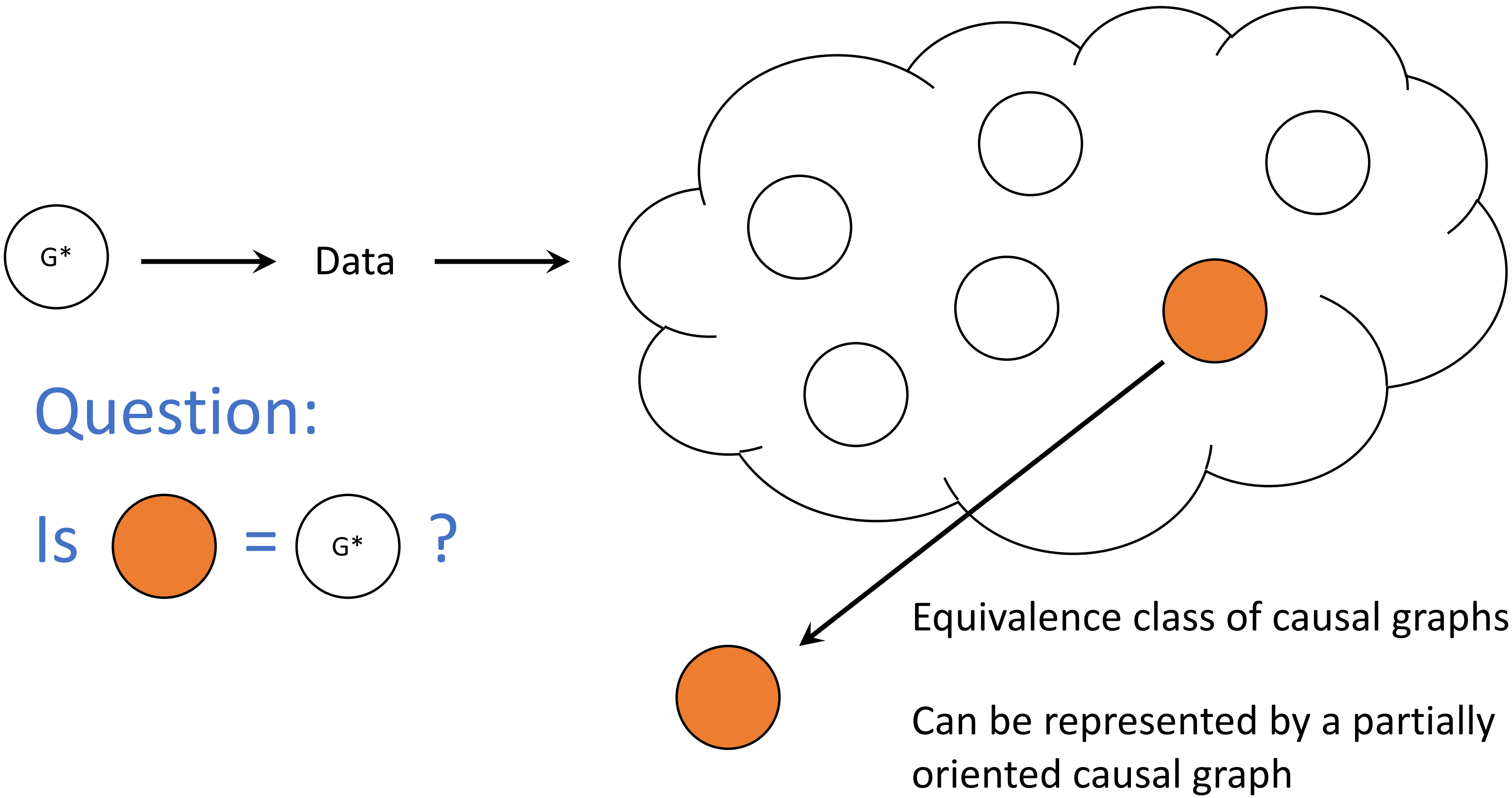
Identify G* using **as few interventions as possible** (minimize t)



Equivalence class of causal graphs

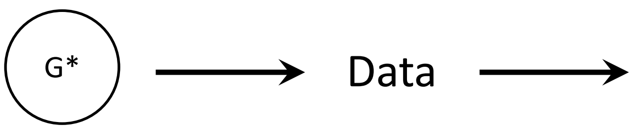
Can be represented by a partially oriented causal graph

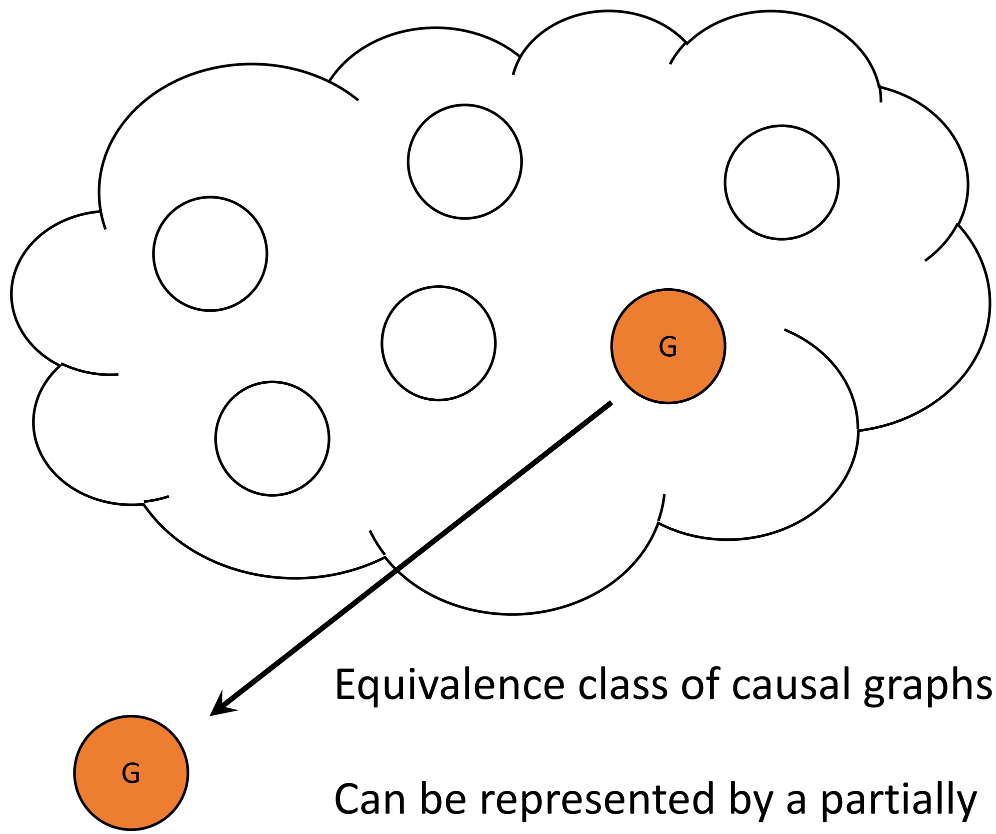
Verification: A simpler problem



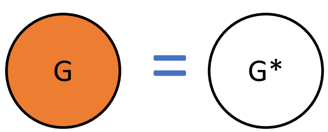
Verification: A simpler problem

Let $\nu(G)$ be the minimum number of interventions needed to answer this question





Question:

Is  ?

Equivalence class of causal graphs

Can be represented by a partially oriented causal graph

Verification: A simpler problem

- What was known

Maximal clique size

1. $\nu(\mathrm{G}) \geq \left\lfloor \frac{\omega(G)}{2} \right\rfloor$ [Squires, Magliacane, Greenewald, Katz, Kocaoglu, Shanmugam 2020]

Number of maximal cliques

2. $\left\lceil \frac{n-r}{2} \right\rceil \leq \nu(\mathrm{G}) \leq n - r$ [Porwal, Srivastava, Sinha 2022]

Verification: A simpler problem

- What was known

Maximal clique size

1. $\nu(\mathrm{G}) \geq \left\lfloor \frac{\omega(G)}{2} \right\rfloor$ [Squires, Magliacane, Greenewald, Katz, Kocaoglu, Shanmugam 2020]

Number of maximal cliques

2. $\left\lceil \frac{n-r}{2} \right\rceil \leq \nu(\mathrm{G}) \leq n - r$ [Porwal, Srivastava, Sinha 2022]

- What we show [Choo, Shiragur, Bhattacharyya 2022]
  - **Exact** characterization of $\nu(\mathrm{G})$ for any causal DAG G via a minimum vertex cover on an induced subgraph of G

Verification: A simpler problem

- What was known

Maximal clique size

1. $\nu(\mathrm{G}) \geq \left\lfloor \frac{\omega(G)}{2} \right\rfloor$ [Squires, Magliacane, Greenewald, Katz, Kocaoglu, Shanmugam 2020]

Number of maximal cliques

2. $\left\lceil \frac{n-r}{2} \right\rceil \leq \nu(\mathrm{G}) \leq n - r$ [Porwal, Srivastava, Sinha 2022]

- What we show [Choo, Shiragur, Bhattacharyya 2022]
  - **Exact** characterization of $\nu(\mathrm{G})$ for any causal DAG G via a minimum vertex cover on an induced subgraph of G
  - Proof idea: Induction on a topo ordering + Meek rules

Verification: A simpler problem

- What was known

Maximal clique size

1. $\nu(\mathrm{G}) \geq \left\lfloor \frac{\omega(G)}{2} \right\rfloor$ [Squires, Magliacane, Greenewald, Katz, Kocaoglu, Shanmugam 2020]

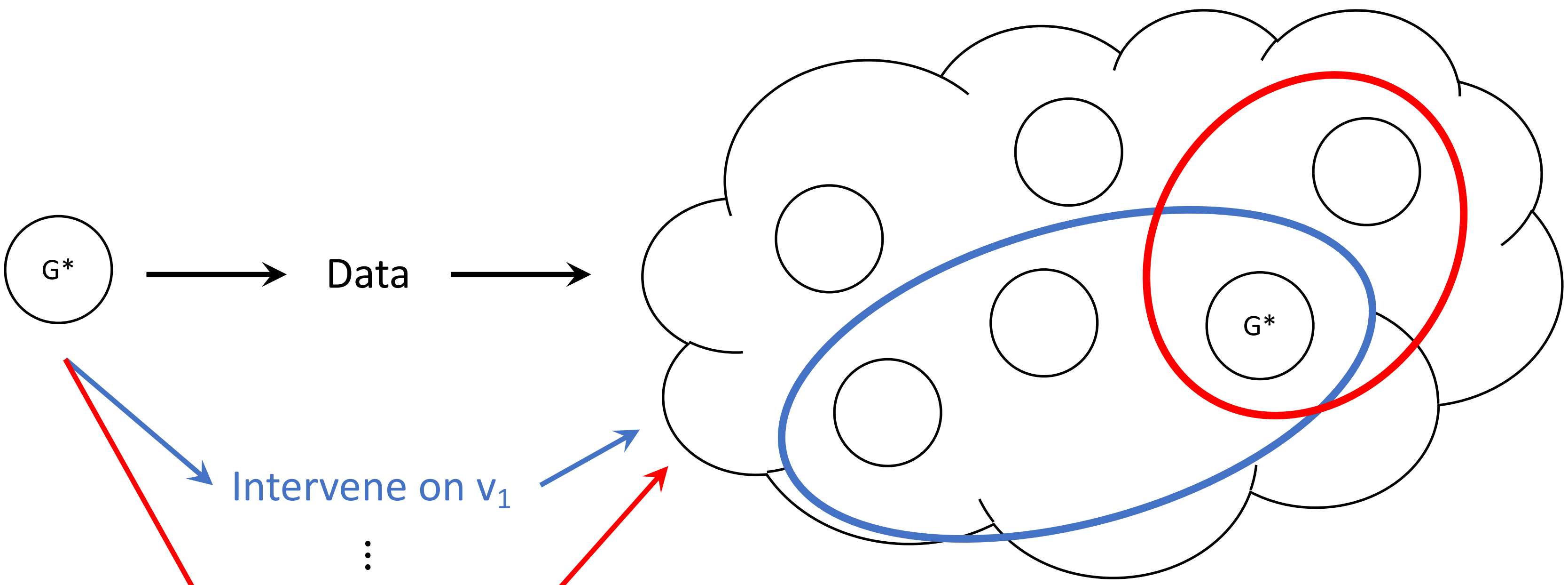
Number of maximal cliques

2. $\left\lceil \frac{n-r}{2} \right\rceil \leq \nu(\mathrm{G}) \leq n - r$ [Porwal, Srivastava, Sinha 2022]

- What we show [Choo, Shiragur, Bhattacharyya 2022]
  - **Exact** characterization of $\nu(\mathrm{G})$ for any causal DAG G via a minimum vertex cover on an induced subgraph of G
  - Proof idea: Induction on a topo ordering + Meek rules
  - Efficiently computable since this subgraph is a forest

Back to the search problem

Identify G* using **as few interventions as possible** (minimize t)



Equivalence class of causal graphs

Can be represented by a partially oriented causal graph

Two classes of interventions

- Non-adaptive
  - Given equivalence class, decide on a single fixed set of interventions that recovers *any possible causal DAG*
  - Need to intervene on a *G-separating system*
    [Kocaoglu, Dimakis, Vishwanath 2017]

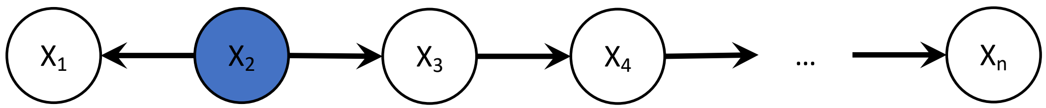
In this simplified talk, where we intervene on a single vertex per intervention, **this is just vertex cover**

Two classes of interventions

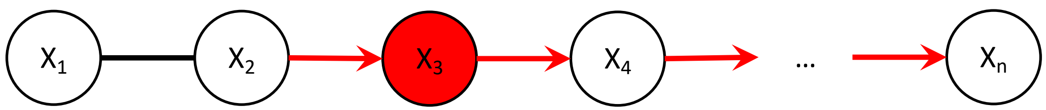
- Non-adaptive
  - Given equivalence class, decide on a single fixed set of interventions that recovers *any possible causal DAG*
  - Need to intervene on a *G-separating system*
    [Kocaoglu, Dimakis, Vishwanath 2017]
- Adaptive
  - Given equivalence class,
    - Decide on first intervention
    - See outcome
    - Decide on second intervention
    - See outcome
    - …

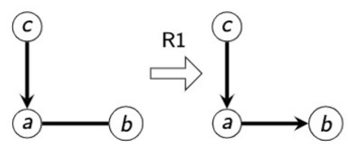
The power of adaptivity

Hidden causal DAG



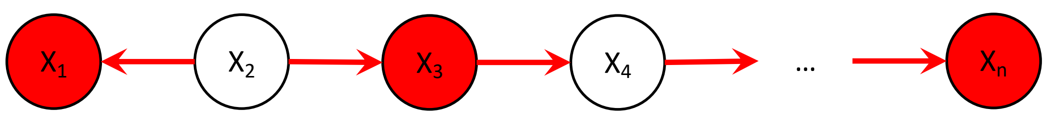
Suppose we intervene X_3





Meek R1

Naïve: Vertex cover



Need $\approx \frac{n}{2}$ interventions

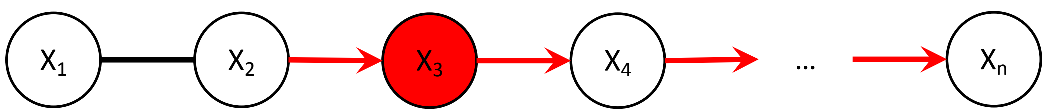
The power of adaptivity

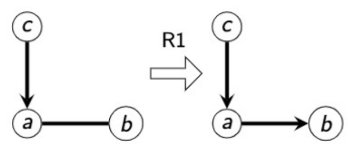
Hidd

causal

We can do something like binary search and only use $\mathcal{O}(\log n)$ interventions

Suppose we intervene X_3

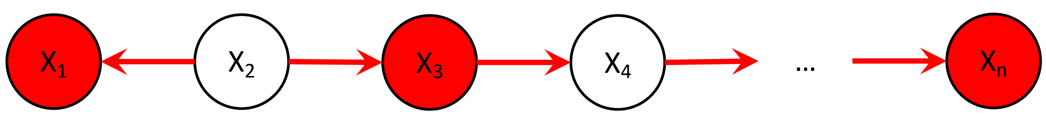




Meek R1

$\mathcal{O}(\log n)$ interventions suffice

Naïve: Vertex cover



Need $\approx \frac{n}{2}$ interventions

The adaptive search problem

- What we know
  - We know at least $\nu(G^*)$ is necessary
  - Prior works only have guarantees for special classes of graphs: cliques, trees, intersection incomparable, etc.

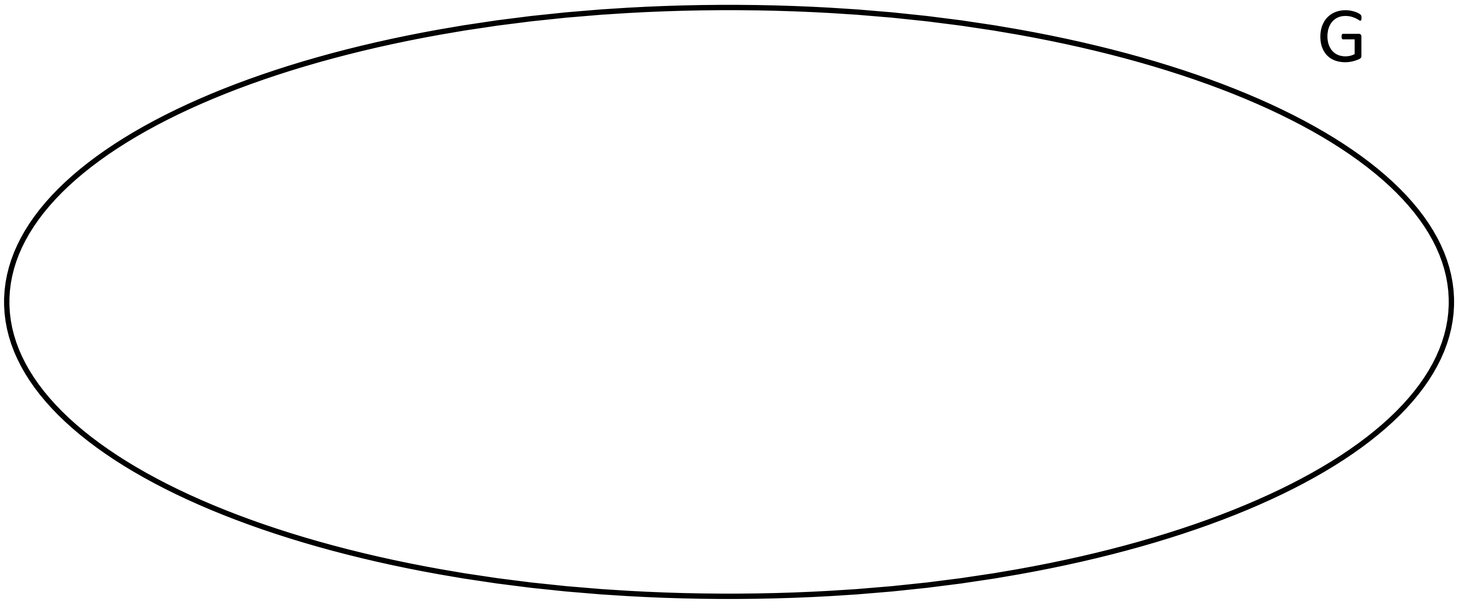
The adaptive search problem

- What we know
  - We know at least $\nu(G^*)$ is necessary
  - Prior works only have guarantees for special classes of graphs: cliques, trees, intersection incomparable, etc.
- What we show [Choo, Shiragur, Bhattacharyya 2022]
  - Punchline: $\mathcal{O}\left(\log n \cdot \nu(G^*)\right)$ interventions suffice
  - "Search is almost as easy as verification"

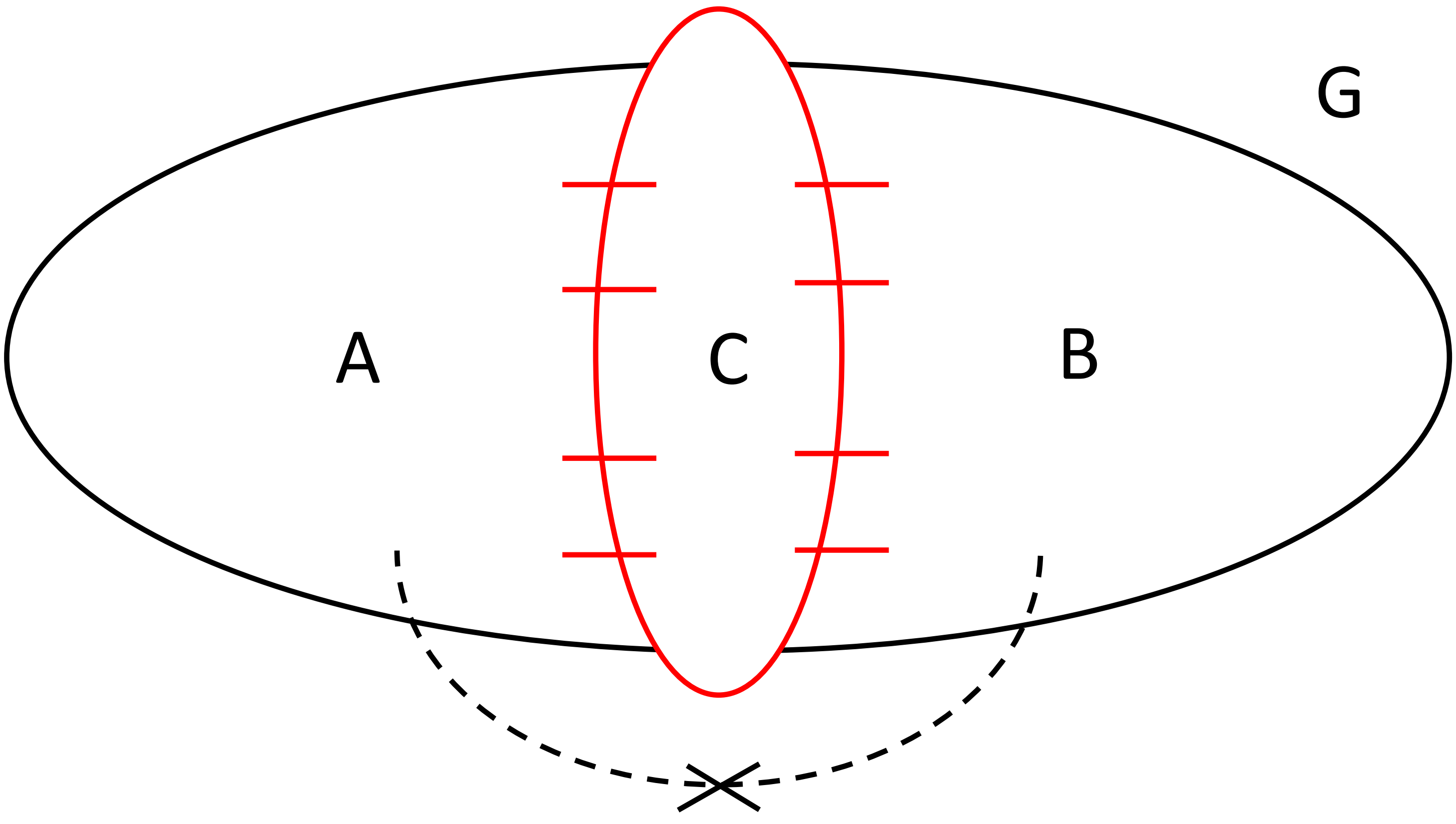
The adaptive search problem

- What we know
  - We know at least $\nu(G^*)$ is necessary
  - Prior works only have guarantees for special classes of graphs: cliques, trees, intersection incomparable, etc.
- What we show [Choo, Shiragur, Bhattacharyya 2022]
  - Punchline: $\mathcal{O}\left(\log n \cdot \nu(G^*)\right)$ interventions suffice
  - "Search is almost as easy as verification"
  - Algorithm does not even know what $\nu(G^*)$ is!
  - $\Omega(\log n)$ is unavoidable when causal graph is a directed path on n nodes

Key algorithmic idea: Graph separators



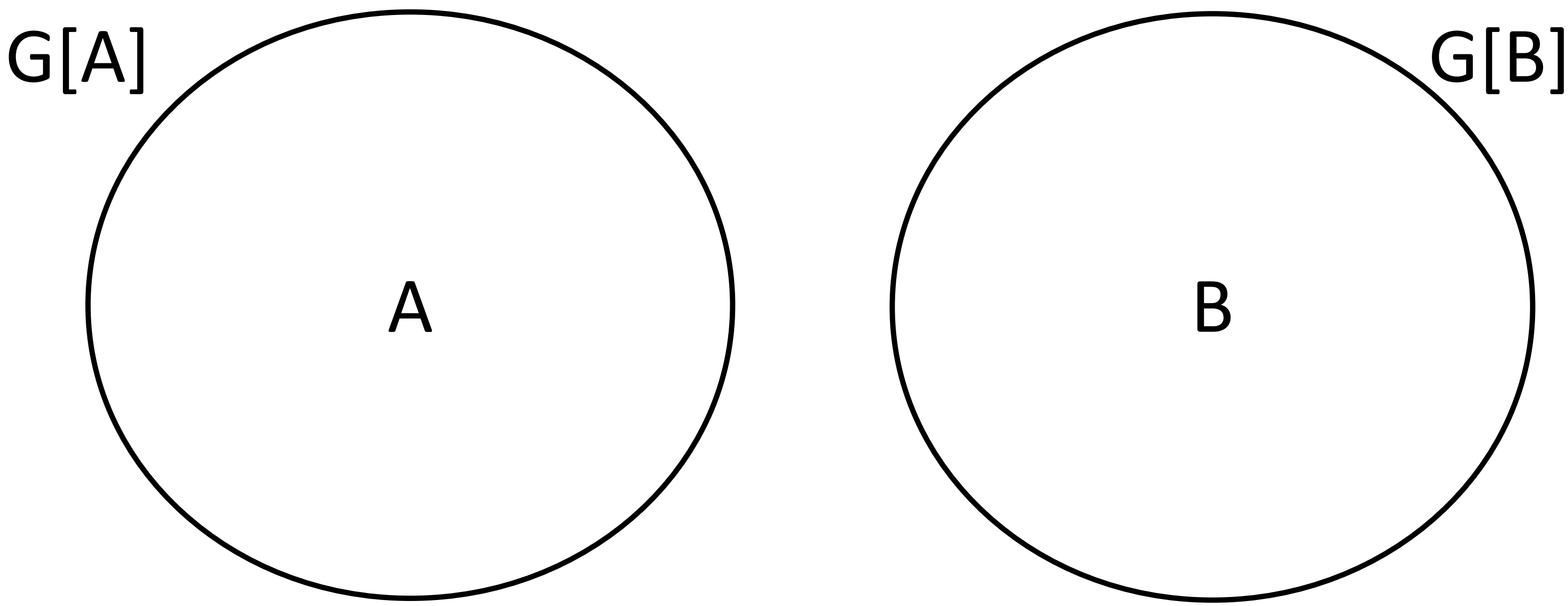
Key algorithmic idea: Graph separators



Partition vertex set V into A, B, C:

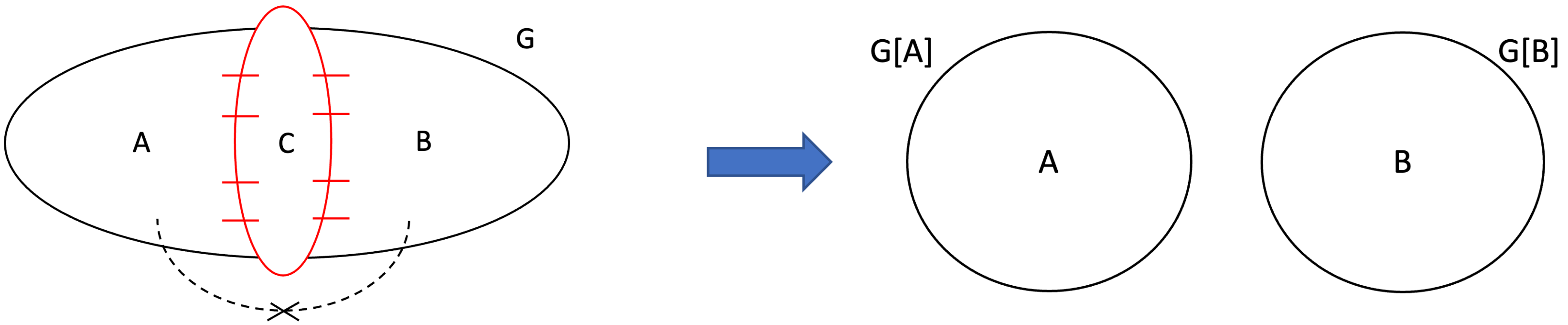
1. C separates A and B
2. $|A|, |B| \leq |V| / 2$

Key algorithmic idea: Graph separators



Recurse on smaller subgraphs of half the size

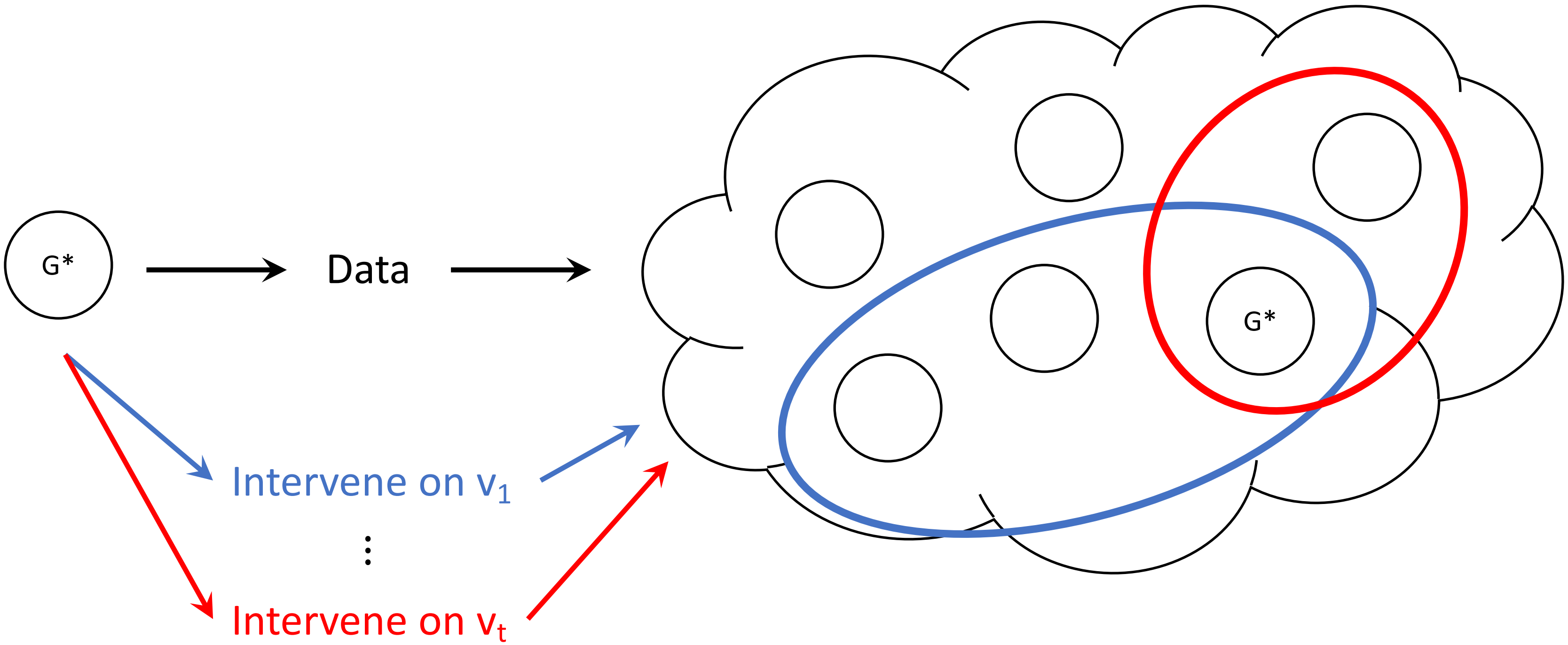
Key algorithmic idea: Graph separators



- Analysis:
  - $\mathcal{O}(\log n)$ rounds ← Chordal graph separator [Gilbert, Rose, Edenbrandt 1984]
  - $\mathcal{O}\big(\nu(G^*)\big)$ per round ← We prove new lower bound on $\nu(G^*)$

Problem setup

Identify G* using **as few interventions as possible** (minimize t)



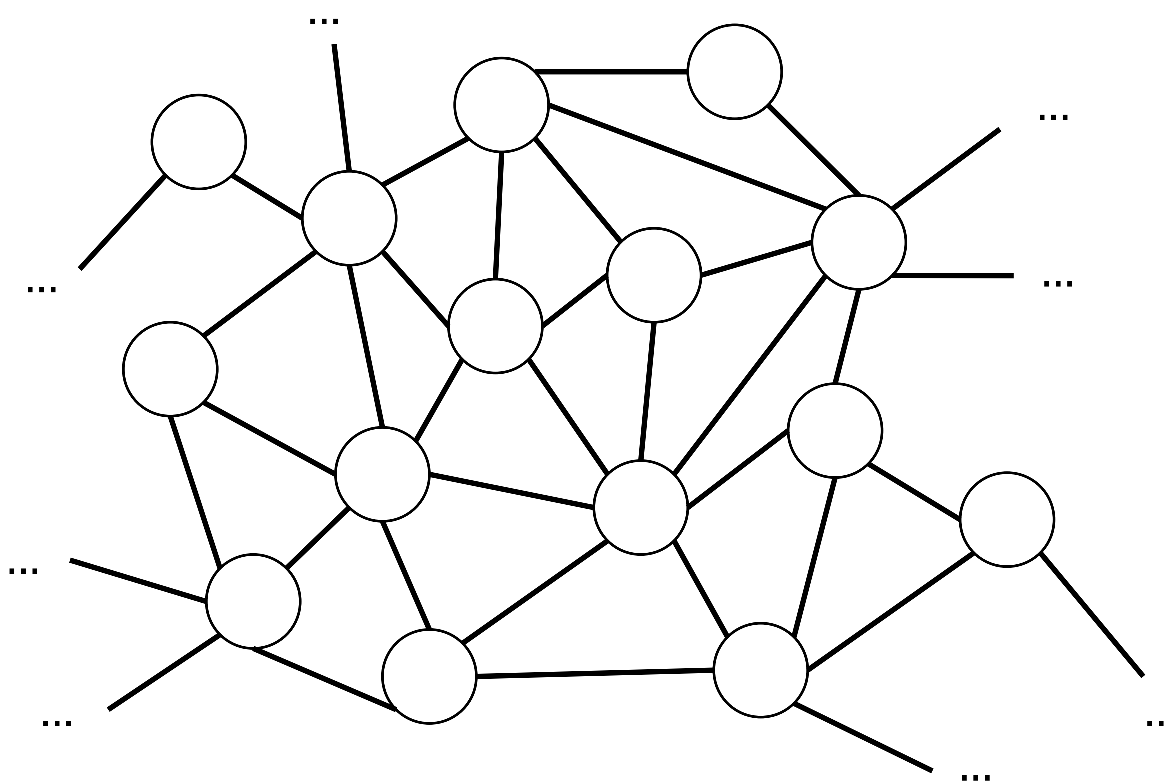
Verification: $\nu(G^*)$ = size of minimum vertex cover of covered edges [CSB 2022]
Search: $\mathcal{O}\big(\log n \cdot \nu(G^*)\big)$ interventions suffice [CSB 2022]

Natural extensions and questions

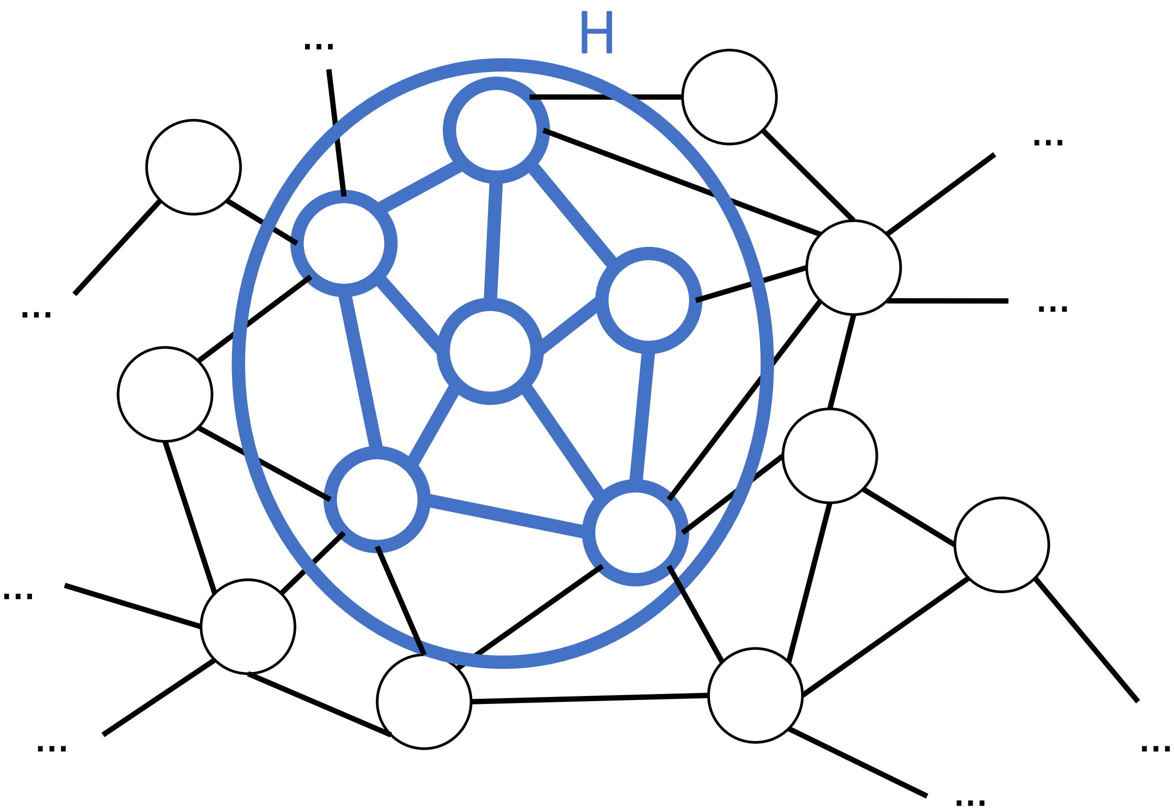
- What if the causal graph is HUGE?

Note: Some are in the papers mentioned earlier in this talk; some are solved and under submission. If you are interested, come talk to me to find out more ☺

What if causal graph is HUGE?



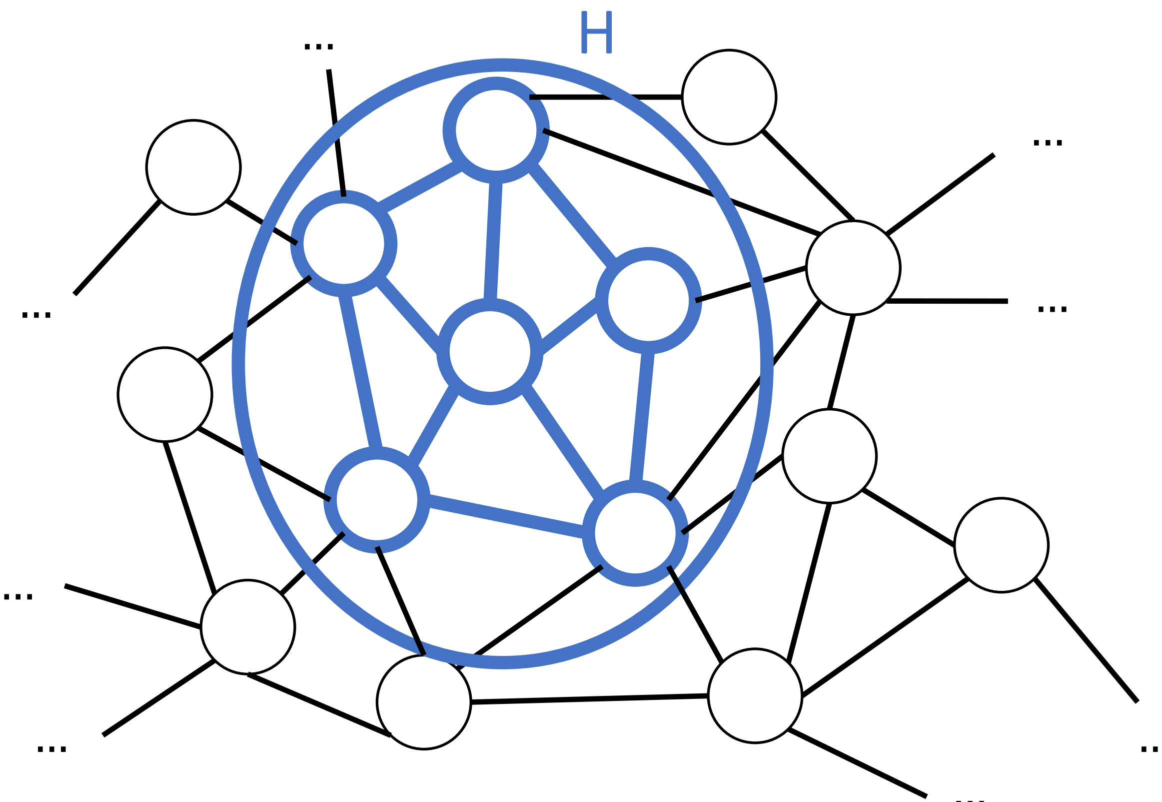
What if causal graph is HUGE?



Local causal discovery:

Only care about a small subgraph of the larger graph

What if causal graph is HUGE?



Local causal discovery:

Only care about a small subgraph of the larger graph

(Informal) Verification: Generalization of "DP on covered edge forest". [CS 2023]

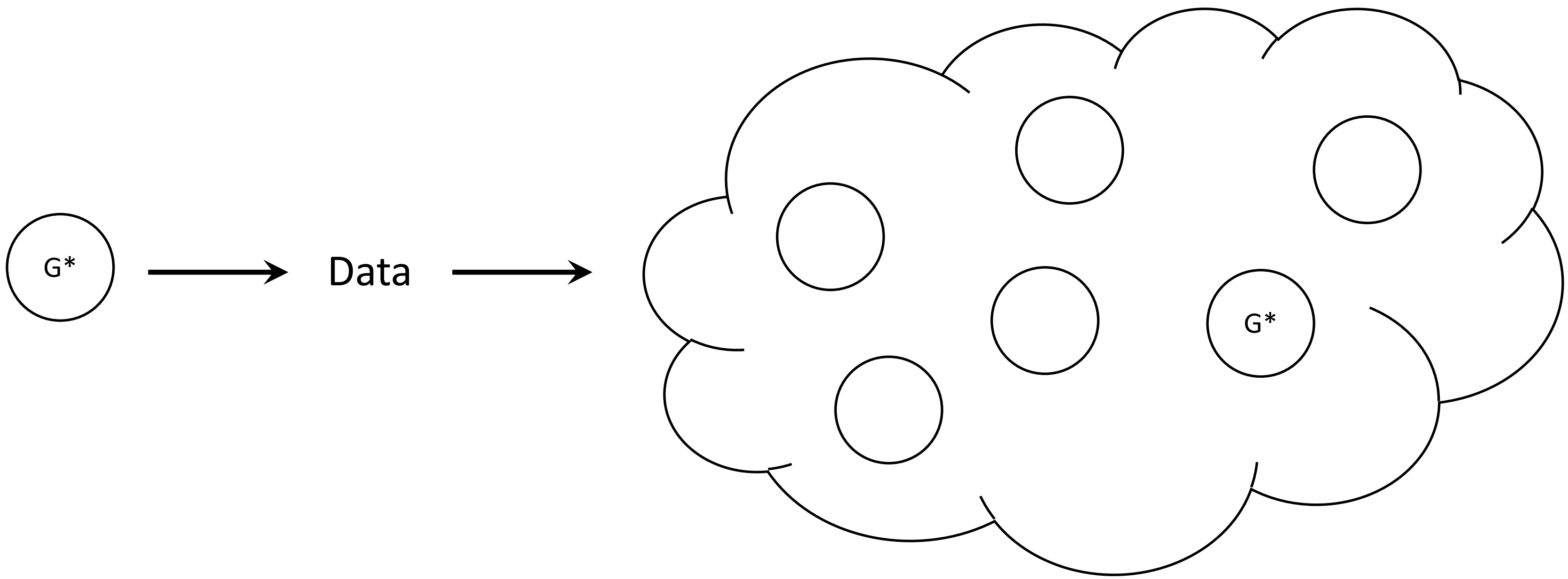
(Informal) Search: $\mathcal{O}\left(\log |H| \cdot \nu(G^*)\right)$ interventions suffices [CS 2023]

Natural extensions and questions

- What if the causal graph is HUGE?
- What if we consult domain experts for advice?

Note: Some are in the papers mentioned earlier in this talk; some are solved and under submission. If you are interested, come talk to me to find out more ☺

In many problem domains…



There are domain experts!

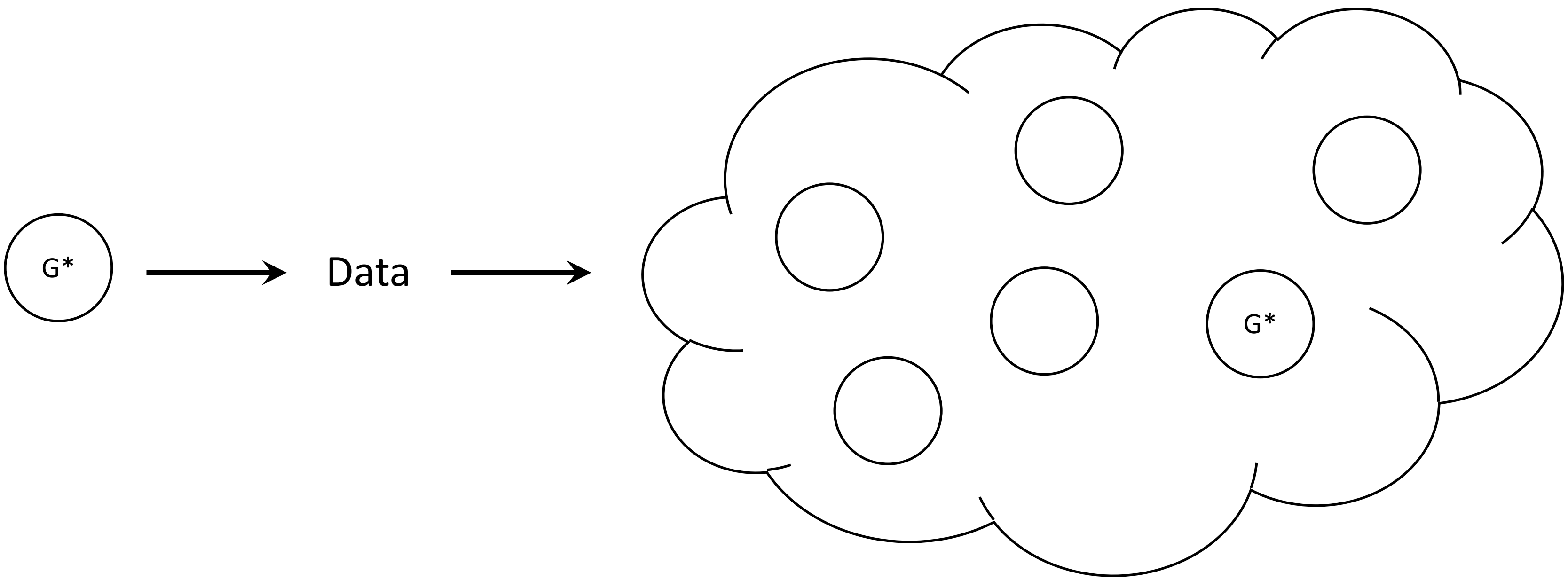


Image credit: https://thenounproject.com/icon/doctor-1285618/

There are domain experts!

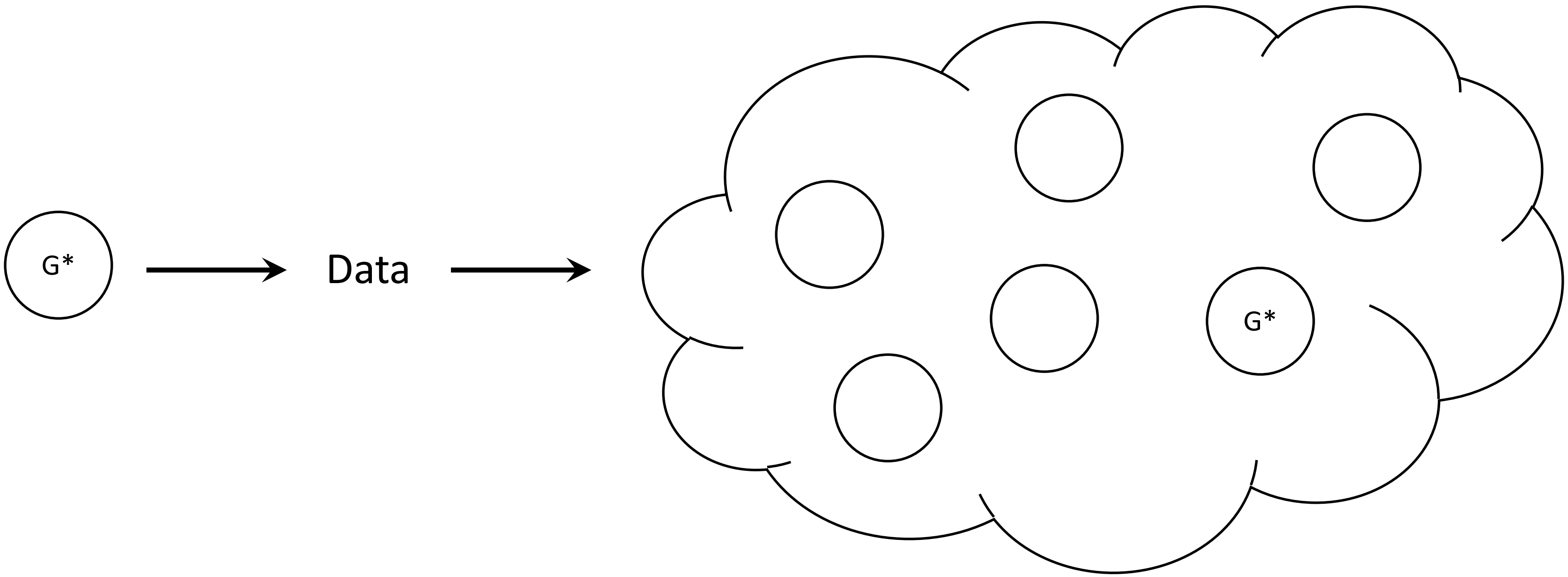
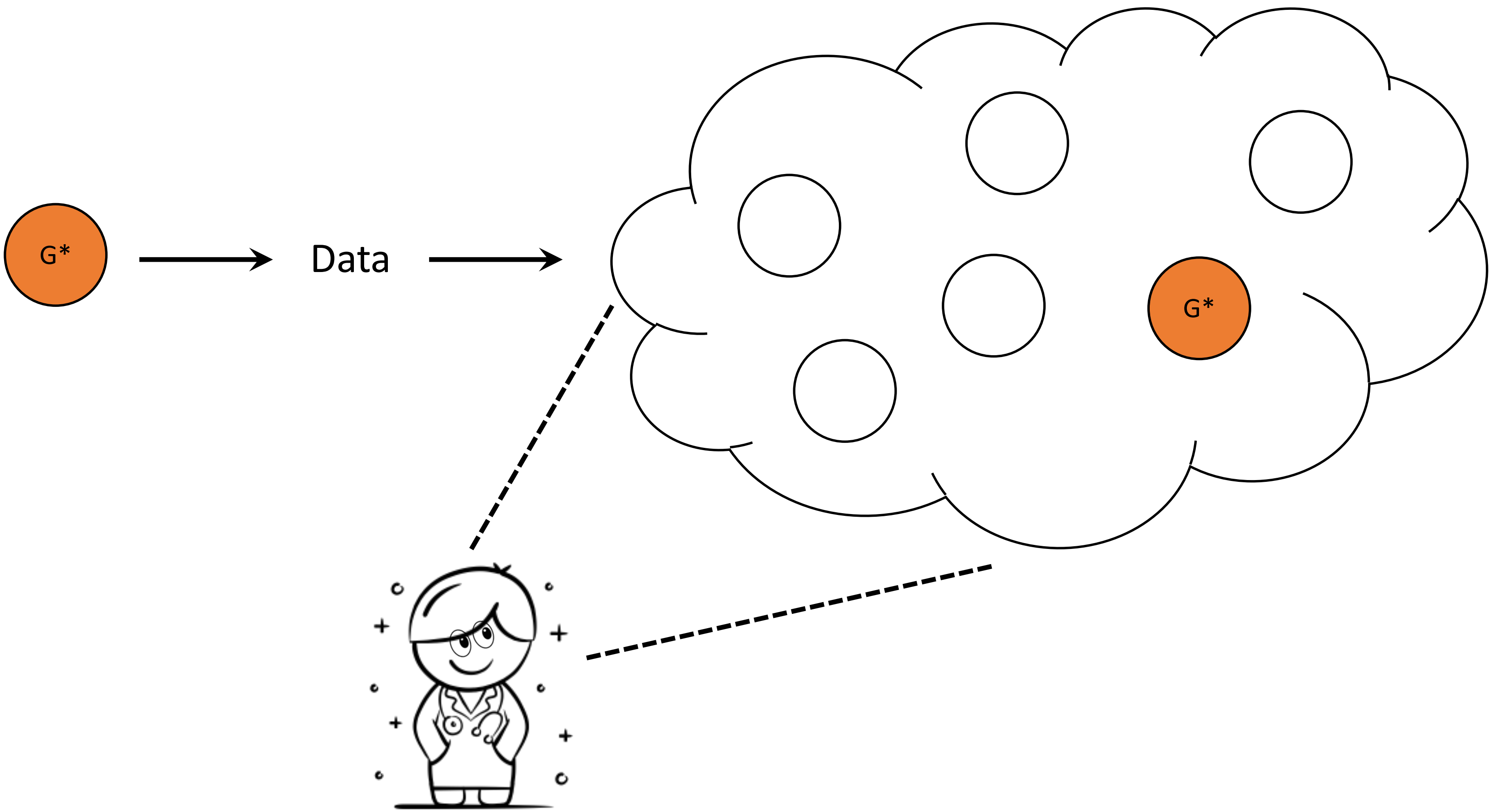
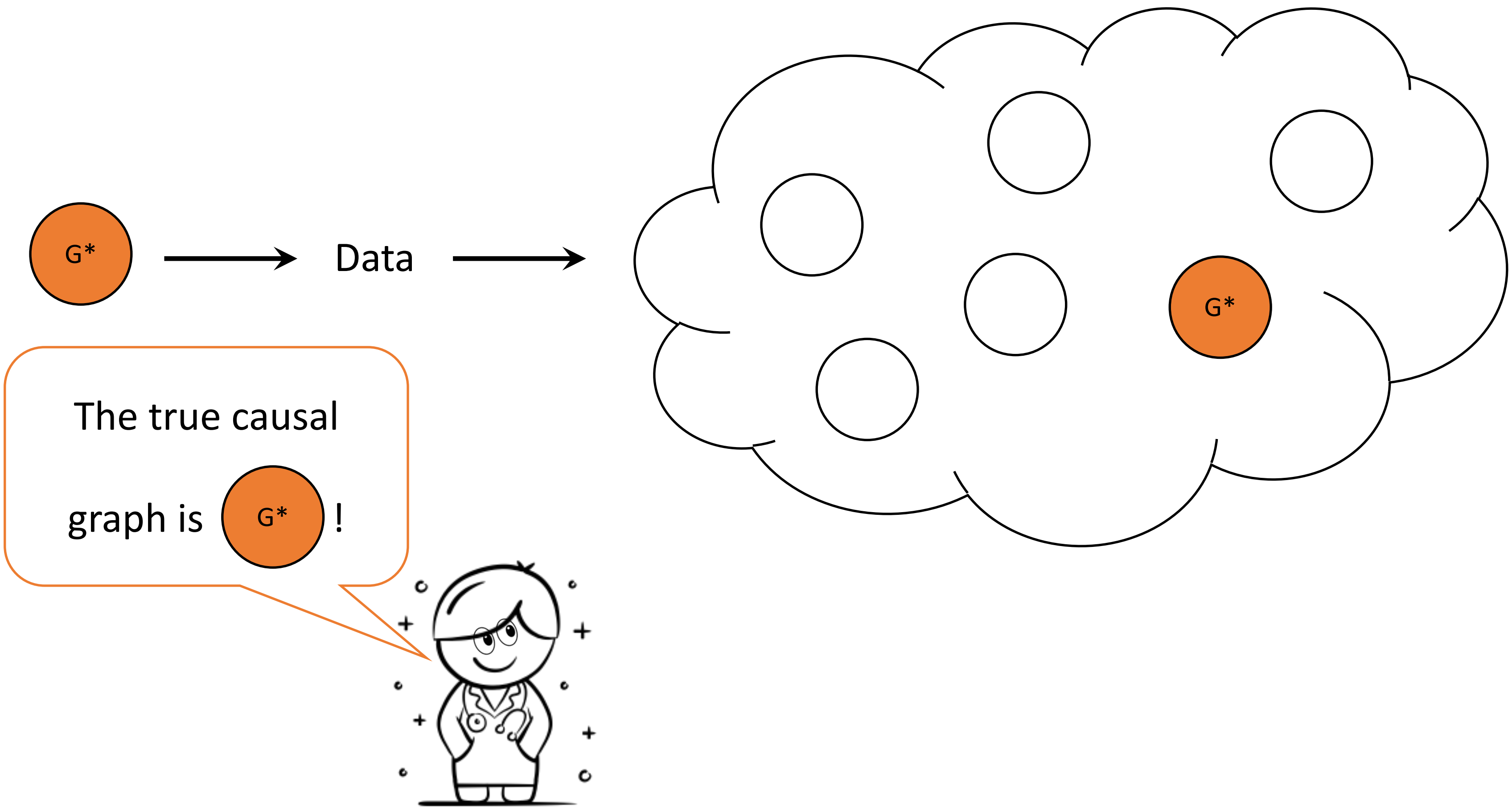


Image credit: https://thenounproject.com/icon/look-4968809/

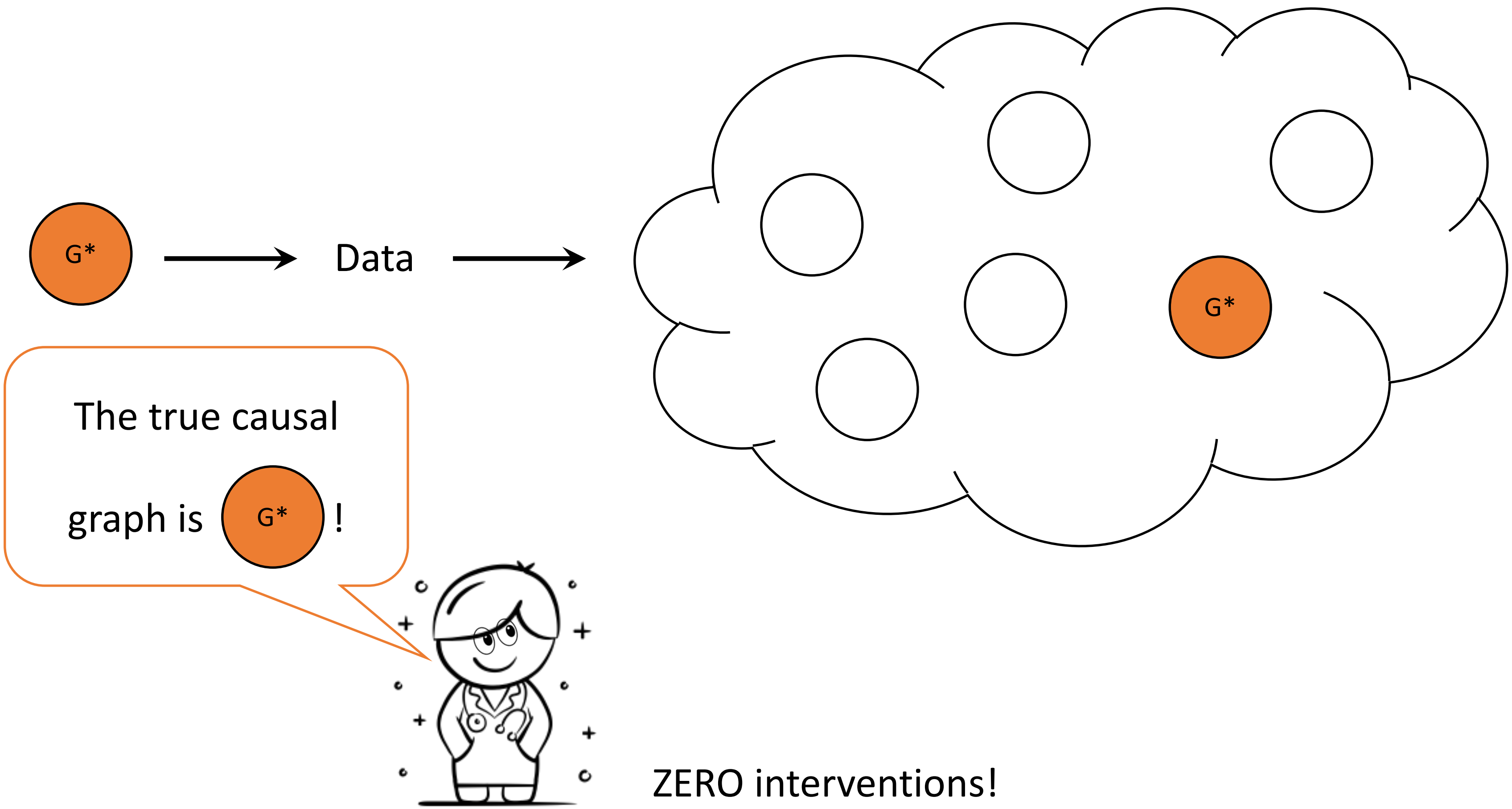
There are domain experts!



There are domain experts!

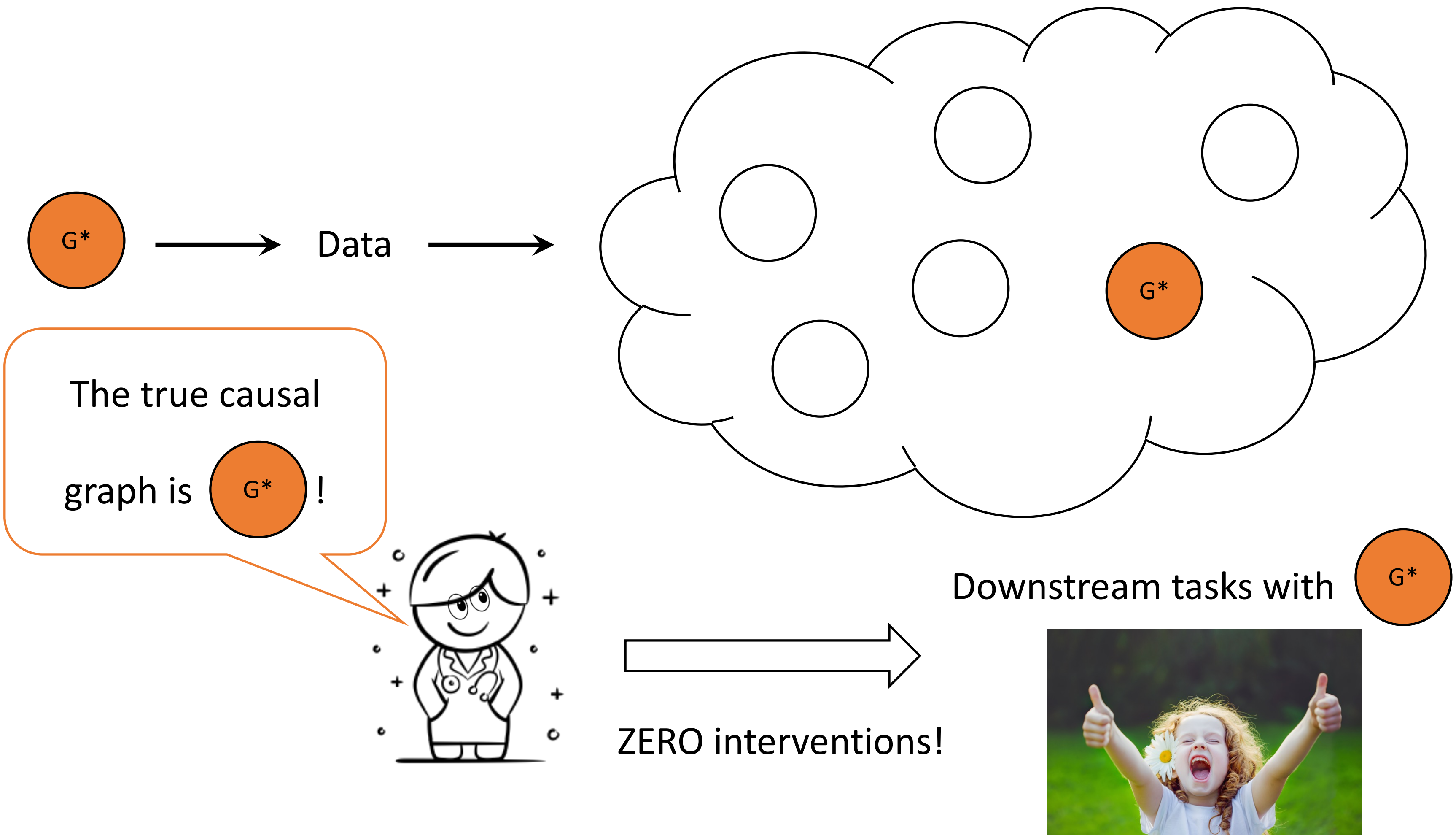


There are domain experts!

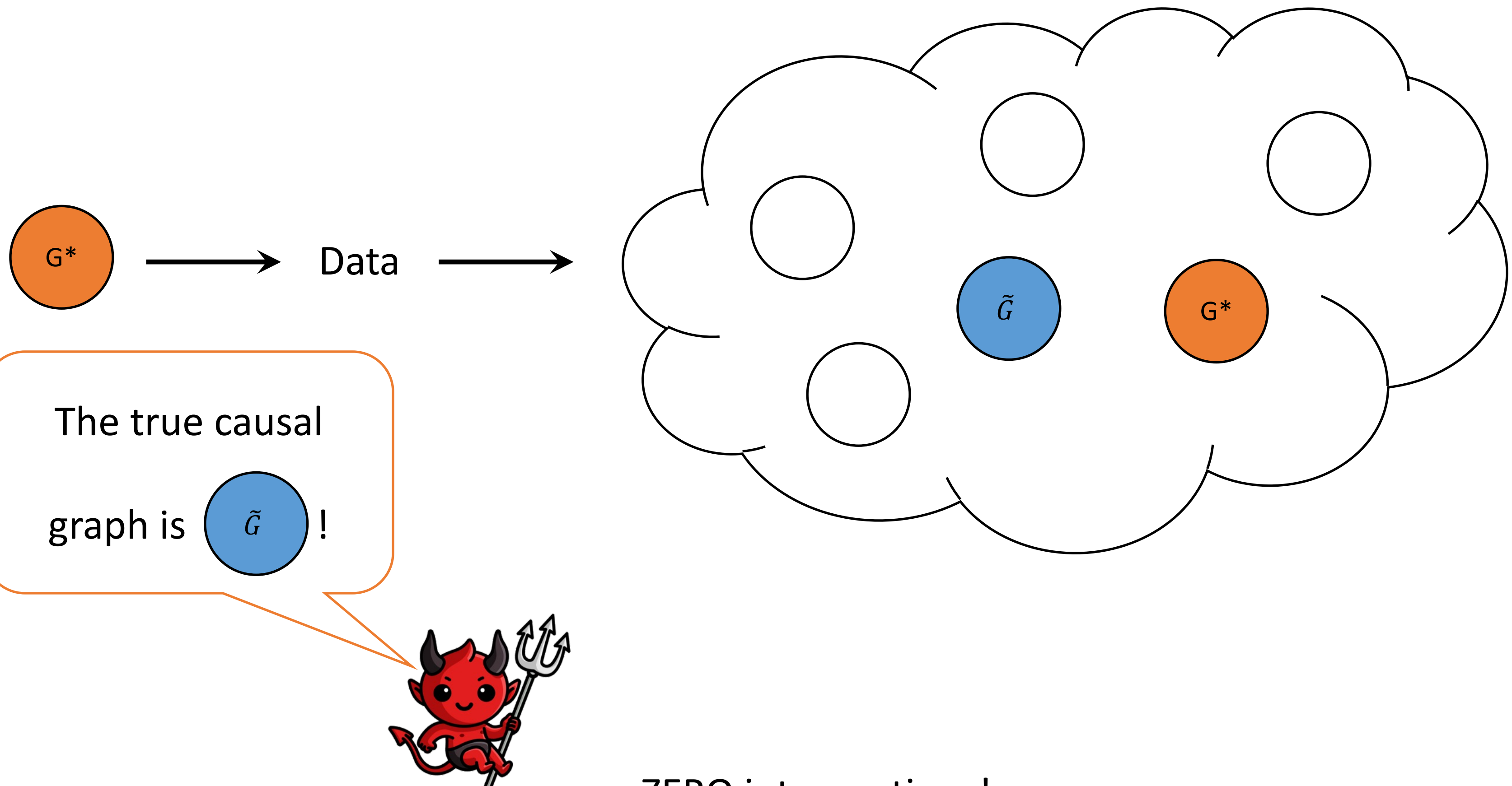


ZERO interventions!

There are domain experts!



But… experts can be wrong

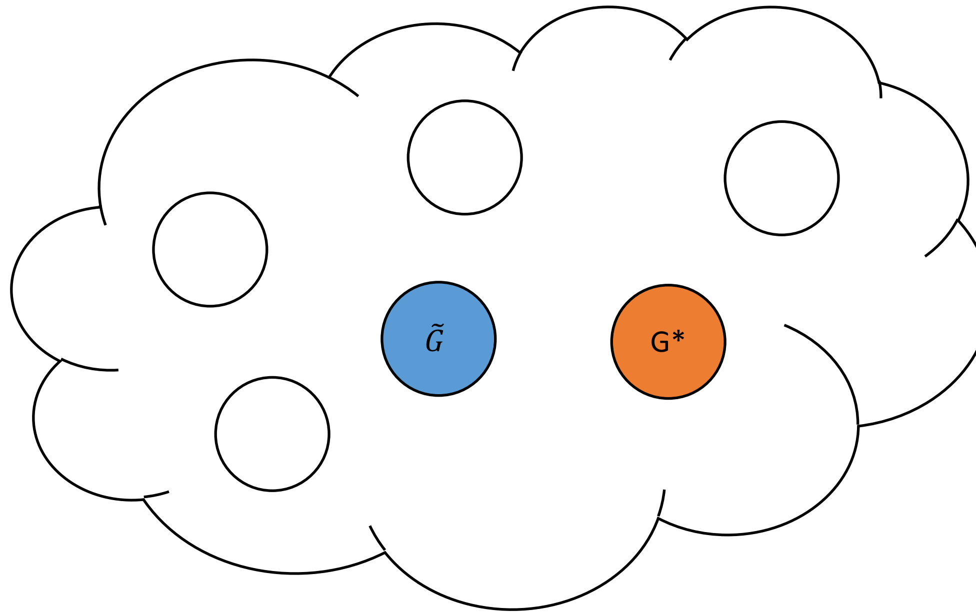


ZERO interventions!

Image credit: https://dribbble.com/shots/14489872-Devil

But… experts can be wrong

G^* ⟶ Data ⟶



The true causal graph is $\tilde{G}$!

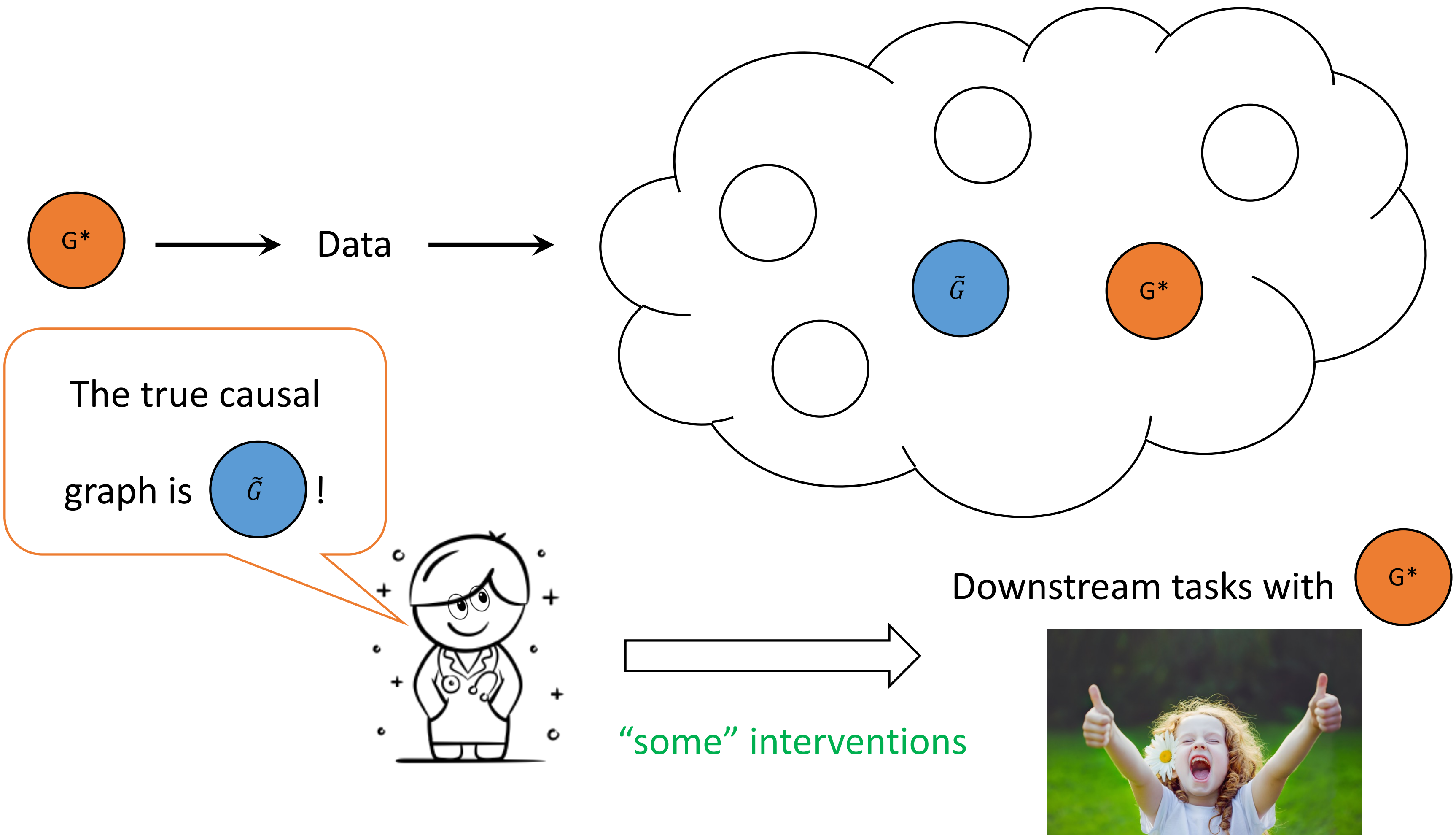
⟹

ZERO interventions!

Downstream tasks with $\tilde{G}$

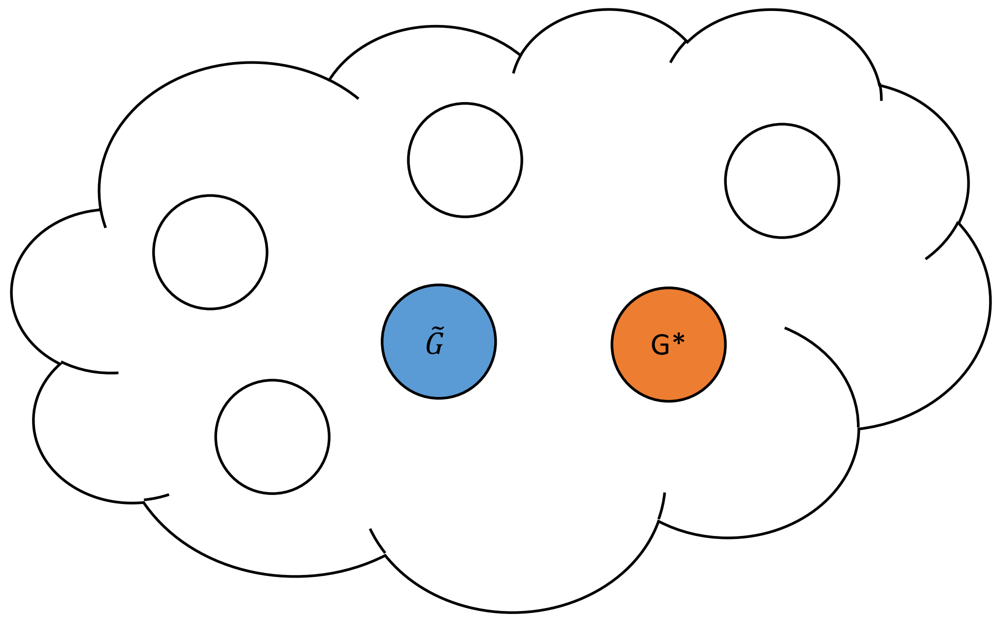
Image credit: https://dribbble.com/shots/3759014-Atomic-Illustrations/attachments/3759014-Atomic-Illustrations?mode=media

How to use imperfect advice?



How to use imperfect advice?

G* ⟶ Data ⟶



The true causal graph is $\tilde{G}$!

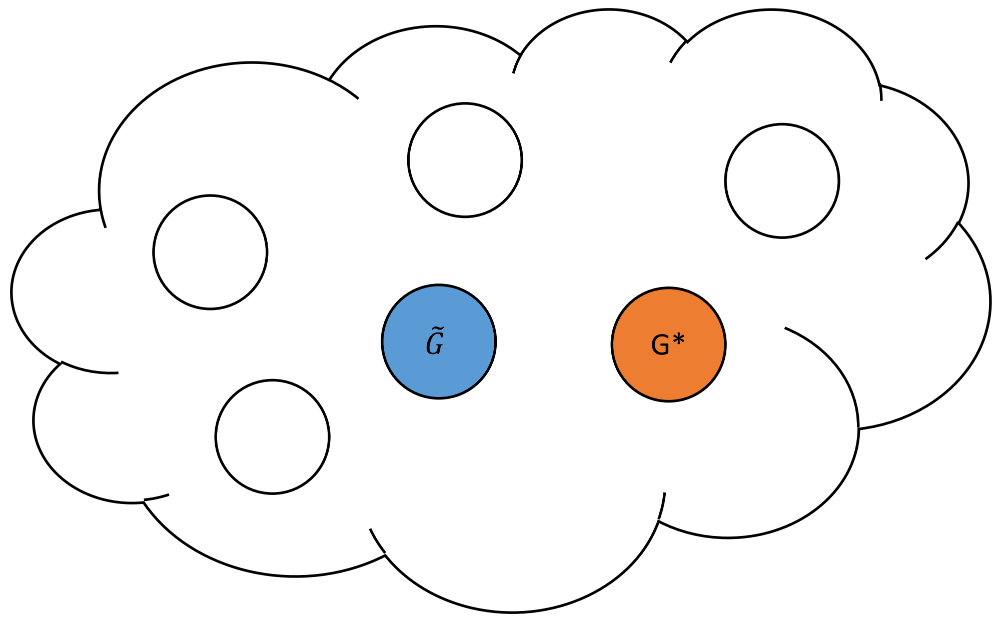
Downstream tasks with G*

"some" interventions

Advice search: $\mathcal{O}\left(\log \psi(G^*, \tilde{G}) \cdot \nu(G^*)\right)$ interventions [Choo, Gouleakis, Bhattacharyya 2023]

How to use imperfect advice?

G^* ⟶ Data ⟶



The true causal graph is $\tilde{G}$!

Downstream tasks with G^*

Quality of advice

$0 \leq \psi(G^*, \tilde{G}) \leq n$

(Good) (Bad)

"some" interventions

Advice search: $\mathcal{O}\left(\log \psi(G^*, \tilde{G}) \cdot \nu(G^*)\right)$ interventions

[Choo, Gouleakis, Bhattacharyya 2023]

Natural extensions and questions

- What if the causal graph is HUGE?
- What if we consult domain experts for advice?
- What if we have limited rounds of adaptivity?
- What if vertices have different interventional costs?
- What if we intervene >1 vertex per intervention?
- Can we weaken/remove some causal assumptions?

Note: Some are in the papers mentioned earlier in this talk; some are solved and under submission. If you are interested, come talk to me to find out more ☺

Some of our relevant papers

Choo, Shiragur, Bhattacharyya. **Verification and search algorithms for causal DAGs**. NeurIPS 2022.

Choo, Shiragur. **Subset verification and search algorithms for causal DAGs**. AISTATS 2023.

Choo, Gouleakis, Bhattacharyya. **Active causal structure learning with advice**. Submitted to ICML 2023. Under review.

Choo, Shiragur. **New metrics and search algorithms for weighted causal DAGs**. Submitted to ICML 2023. Under review.

Choo, Shiragur. **Adaptivity Complexity for Causal Graph Discovery**. Submitted to UAI 2023. Under review.

History about causal inference (5min)

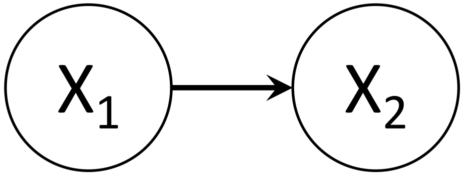
- David Hume (early 18th century)

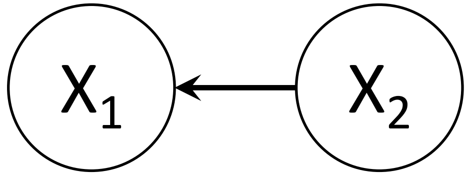
- Judea pearl
- Smoking, Fisher
- Etc
- Oil is bad. Then suddenly say oil is not bad, it’s sugar
- Correlation versus causation
- PICTURES
- Sell the past

Causal graphical models

- Application: Gene regulatory network
- Some story
- Why do we care
- Broad challenge
- One way to model: Graphical modelling
- Sell the future

Two equivalent causal models



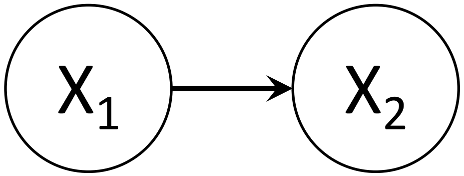


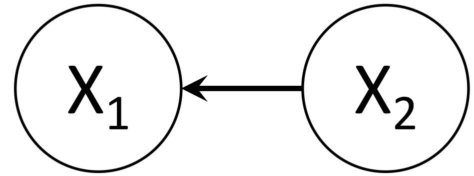
- $X_1 \leftarrow N(0,1)$
- $X_2 \leftarrow X_1$

- $X_2 \leftarrow N(0,1)$
- $X_1 \leftarrow X_2$

Data from both are fully characterized by covariance matrix $\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$

Two equivalent causal models





- $X_1 = \epsilon_1$
- $X_2 = a \cdot X_1 + \epsilon_2$
- $\epsilon_1 \sim N(0, 1)$
- $\epsilon_2 \sim N(0, 1)$

- $X_1 = \frac{a}{a^2+1} \cdot X_2 + \epsilon_1$
- $X_2 = \epsilon_2$
- $\epsilon_1 \sim N\left(0, \frac{1}{a^2+1}\right)$
- $\epsilon_2 \sim N(0, a^2 + 1)$

Data from both are fully characterized by covariance matrix $\begin{bmatrix} 1 & a \\ a & a^2 + 1 \end{bmatrix}$

Example from: https://youtu.be/rE6IMfSkOU0?t=849. See https://github.com/csquires/6.S091-causality for full course

Two equivalent causal models

- $X_1 =$
- $X_2 =$
- $\epsilon_1 \sim$
- $\epsilon_2 \sim$

1. Make more assumptions

OR

1. **Perform interventions**

Data fron

a

$a^2 + 1$

Example from: https://youtu.be/rE6IMfSkOU0?t=849. See https://github.com/csquires/6.S091-causality for full course

The problem of structure learning

- Broad challenge
- Debugging cloud micro-services
- Causal bandits

https://www.topcoder.com/challenges/25f60820-2e69-444b-bc03-490686af2c87

What is studied in causality?

- Judea Pearl
- Blah blah
- In this talk, we focus on the problem of structure learning

Why structure learning?

- Broad challenge
- Debugging cloud micro-services
- Causal bandits

Intervention example in gene regulatory network

- CRISPR, nobel prize
- Perturb-seq

- Pull up numbers
    - about interventions is costly
    - Way more observational data than interventions

Adaptivity

Contrast non-adaptive versus adaptive

e.g. finding a number in a sorted array

To explain why is it difficult

Say what people have tried

Say what we have

Keep the talk light-hearted

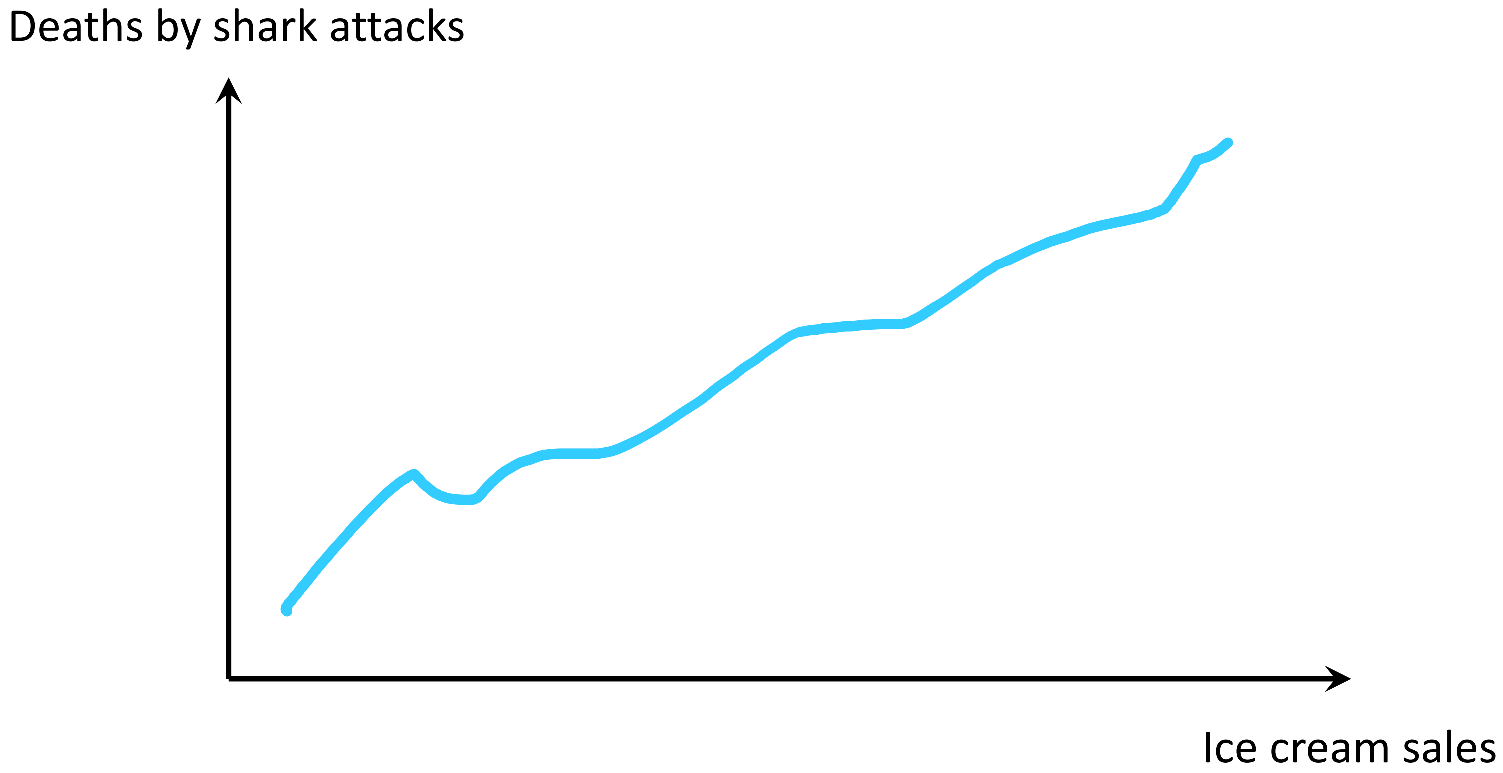
Increase humor in slides

Relate to open house / NUS / Singapore

Causal graphical models

- Talk only about search results
- Only say “compete against an algorithm that KNOWS the ground truth”. Don’t even define verification number
- Say that the problem becomes super hard if we allow non-uniform costs

Correlation is not causation



Correlation is not causation

Deaths by shark attacks

Ice cream sales

https://imgflip.com/memegenerator/28604674/Scared-Kid
https://imgflip.com/memegenerator/20892444/Mr-Burns

Suppose we are given some data and we want to discover causal relationships between them

	X_1	X_2	X_3	X_4	X_5	X_6
Sample 1	0.22	0.04	0.84	0.48	0.98	0.82
Sample 2	0.87	0.17	0.61	0.67	0.67	0.23
Sample 3	0.55	0.54	0.67	0.86	0.93	0.23
…	…	…	…	…	…	…
Sample M	0.12	0.95	0.79	0.47	0.05	0.92

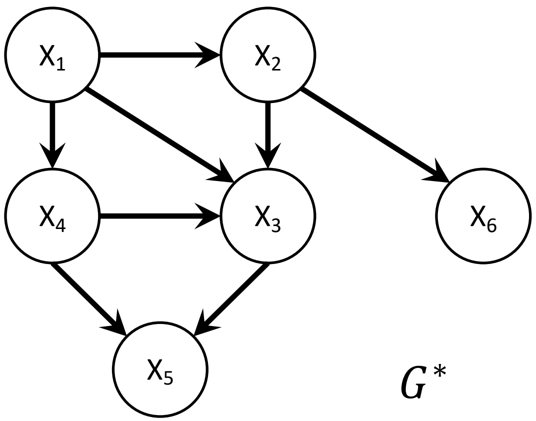
Suppose we are given some data and we want to discover causal relationships between them

	X_1	X_2	X_3	X_4	X_5	X_6
Sample 1	0.22	0.04	0.84	0.48	0.98	0.82
Sample 2	0.87	0.17	0.61	0.67	0.67	0.23
Sample 3	0.55	0.54	0.67	0.86	0.93	0.23
…	…	…	…	…	…	…
Sample M	0.12	0.95	0.79	0.47	0.05	0.92

Genetics	Gene 1	Gene 2	Gene 3	Gene 4	Gene 5	Gene 6
Finance	AAPL	GOOGL	MSFT	AMZN	META	TSLA
…	…	…	…	…	…	…
Health care	Diet	Exercise	Weight	Blood pressure	Blood glucose	Cholesterol levels

One possible way: use graphical modelling

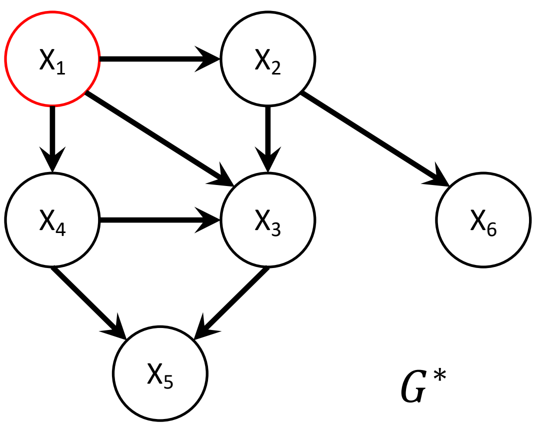
	X_1	X_2	X_3	X_4	X_5	X_6
Sample 1	0.22	0.04	0.84	0.48	0.98	0.82
Sample 2	0.87	0.17	0.61	0.67	0.67	0.23
Sample 3	0.55	0.54	0.67	0.86	0.93	0.23
…	…	…	…	…	…	…
Sample M	0.12	0.95	0.79	0.47	0.05	0.92



G^*

A directed acyclic graphs (DAG) representation

	X_1	X_2	X_3	X_4	X_5	X_6
Sample 1	0.22	0.04	0.84	0.48	0.98	0.82
Sample 2	0.87	0.17	0.61	0.67	0.67	0.23
Sample 3	0.55	0.54	0.67	0.86	0.93	0.23
...	...	...	...	...	...	...
Sample M	0.12	0.95	0.79	0.47	0.05	0.92



G^*

$$X_1 = f_1(\epsilon_1)$$

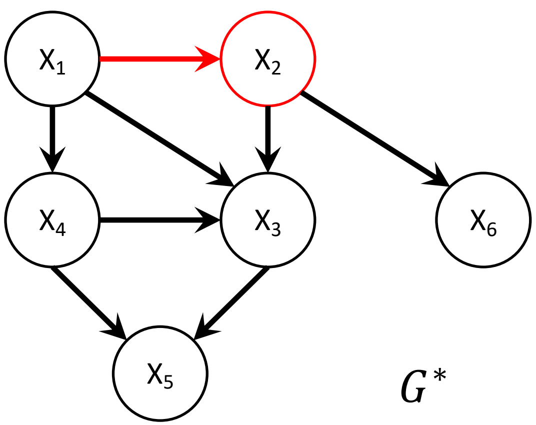
Structural equation model (SEM)

ϵ_1

noise

A directed acyclic graphs (DAG) representation

	X_1	X_2	X_3	X_4	X_5	X_6
Sample 1	0.22	0.04	0.84	0.48	0.98	0.82
Sample 2	0.87	0.17	0.61	0.67	0.67	0.23
Sample 3	0.55	0.54	0.67	0.86	0.93	0.23
…	…	…	…	…	…	…
Sample M	0.12	0.95	0.79	0.47	0.05	0.92



G^*

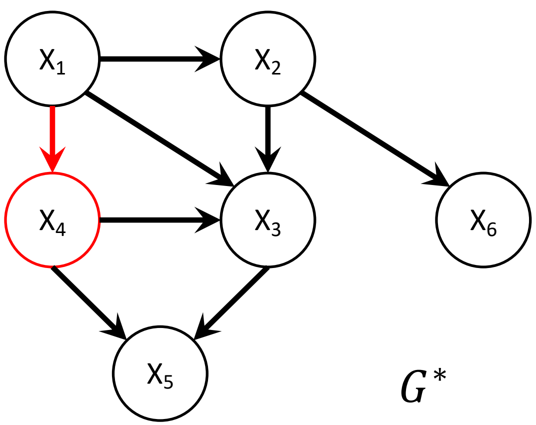
$$X_1 = f_1(\epsilon_1)$$
$$X_2 = f_2(X_1, \epsilon_2)$$

Structural equation model (SEM)

$\epsilon_1, \epsilon_2,$ independent noise

A directed acyclic graphs (DAG) representation

	X_1	X_2	X_3	X_4	X_5	X_6
Sample 1	0.22	0.04	0.84	0.48	0.98	0.82
Sample 2	0.87	0.17	0.61	0.67	0.67	0.23
Sample 3	0.55	0.54	0.67	0.86	0.93	0.23
...	...	...	...	...	...	...
Sample M	0.12	0.95	0.79	0.47	0.05	0.92



G^*

$$X_1 = f_1(\epsilon_1)$$
$$X_2 = f_2(X_1, \epsilon_2)$$

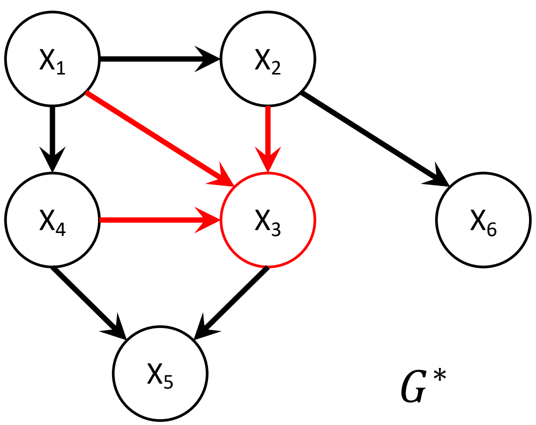
$$X_4 = f_4(X_1, \epsilon_4)$$

Structural equation model (SEM)

$\epsilon_1, \epsilon_2, \quad \epsilon_4$ independent noise

A directed acyclic graphs (DAG) representation

	X_1	X_2	X_3	X_4	X_5	X_6
Sample 1	0.22	0.04	0.84	0.48	0.98	0.82
Sample 2	0.87	0.17	0.61	0.67	0.67	0.23
Sample 3	0.55	0.54	0.67	0.86	0.93	0.23
…	…	…	…	…	…	…
Sample M	0.12	0.95	0.79	0.47	0.05	0.92



G^*

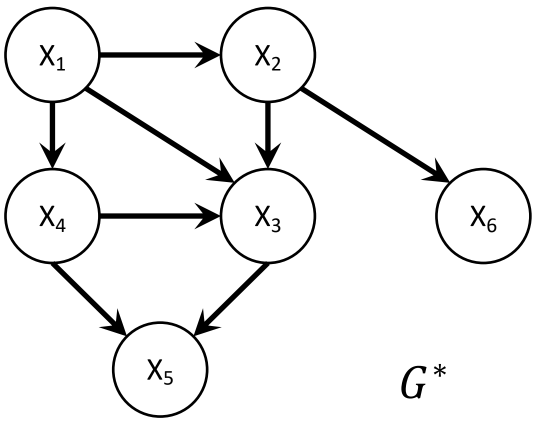
$$X_1 = f_1(\epsilon_1)$$
$$X_2 = f_2(X_1, \epsilon_2)$$
$$X_3 = f_3(X_1, X_2, X_4, \epsilon_3)$$
$$X_4 = f_4(X_1, \epsilon_4)$$

Structural equation model (SEM)

$\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4$ independent noise

A directed acyclic graphs (DAG) representation

	X_1	X_2	X_3	X_4	X_5	X_6
Sample 1	0.22	0.04	0.84	0.48	0.98	0.82
Sample 2	0.87	0.17	0.61	0.67	0.67	0.23
Sample 3	0.55	0.54	0.67	0.86	0.93	0.23
…	…	…	…	…	…	…
Sample M	0.12	0.95	0.79	0.47	0.05	0.92



G^*

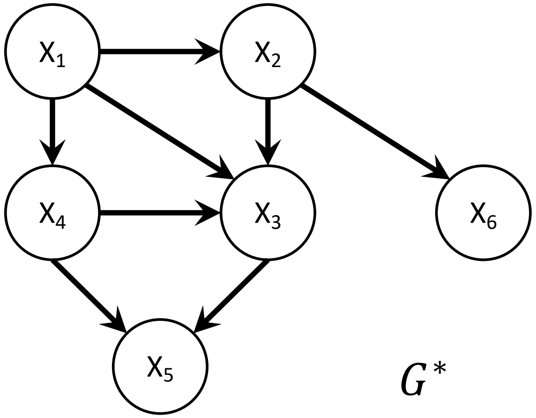
$$
\begin{aligned}
X_1 &= f_1(\epsilon_1) \\
X_2 &= f_2(X_1, \epsilon_2) \\
X_3 &= f_3(X_1, X_2, X_4, \epsilon_3) \\
X_4 &= f_4(X_1, \epsilon_4) \\
X_5 &= f_5(X_3, X_4, \epsilon_5) \\
X_6 &= f_6(X_2, \epsilon_6)
\end{aligned}
$$

$\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5, \epsilon_6$ independent noise

Structural equation model (SEM)

A directed acyclic graphs (DAG) representation

	X_1	X_2	X_3	X_4	X_5	X_6
Sample 1	0.22	0.04	0.84	0.48	0.98	0.82
Sample 2	0.87	0.17	0.61	0.67	0.67	0.23
Sample 3	0.55	0.54	0.67	0.86	0.93	0.23
…	…	…	…	…	…	…
Sample M	0.12	0.95	0.79	0.47	0.05	0.92



G^*

$$X_1 = f_1(\epsilon_1)$$
$$X_2 = f_2(X_1, \epsilon_2)$$
$$X_3 = f_3(X_1, X_2, X_4, \epsilon_3)$$
$$X_4 = f_4(X_1, \epsilon_4)$$
$$X_5 = f_5(X_3, X_4, \epsilon_5)$$
$$X_6 = f_6(X_2, \epsilon_6)$$

$\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5, \epsilon_6$ independent noise

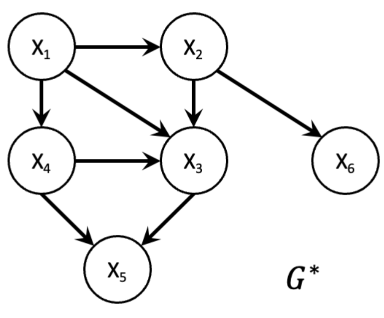
Structural equation model (SEM)

Using the Bayesian network, one can decompose the joint distribution as follows:

$$\Pr[X_1] \cdot \Pr[X_2 | X_1] \cdot \Pr[X_4 | X_1] \cdot \Pr[X_3 | X_1, X_2, X_4] \cdot \Pr[X_5 | X_3, X_4] \cdot \Pr[X_6 | X_2]$$

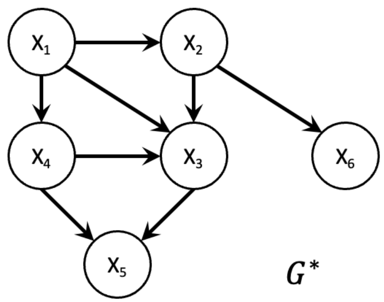
Conditional independence (CI) tests

- A standard way (under some causal assumptions) to recover graph structure from data is to perform CI tests
  - e.g. PC (Peter-Clark) algorithm [Spirtes, Glymour, Scheines, Heckerman 2000]



Conditional independence (CI) tests

- A standard way (under some causal assumptions) to recover graph structure from data is to perform CI tests
  - e.g. PC (Peter-Clark) algorithm [Spirtes, Glymour, Scheines, Heckerman 2000]

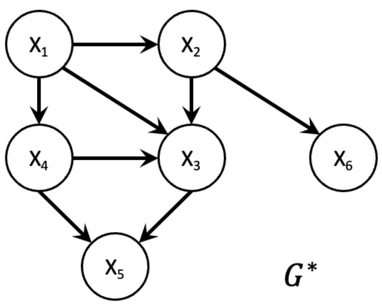


Get samples

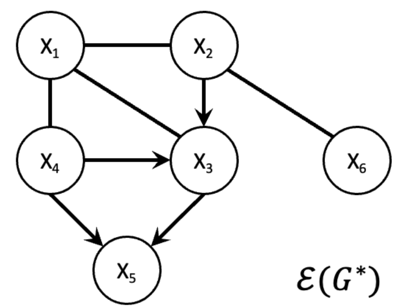
	X_1	X_2	X_3	X_4	X_5	X_6
Sample 1	0.22	0.04	0.84	0.48	0.98	0.82
Sample 2	0.87	0.17	0.61	0.67	0.67	0.23
Sample 3	0.55	0.54	0.67	0.86	0.93	0.23
…	…	…	…	…	…	…
Sample M	0.12	0.95	0.79	0.47	0.05	0.92

Conditional independence (CI) tests

- A standard way (under some causal assumptions) to recover graph structure from data is to perform CI tests
  - e.g. PC (Peter-Clark) algorithm [Spirtes, Glymour, Scheines, Heckerman 2000]



G^*



$\mathcal{E}(G^*)$

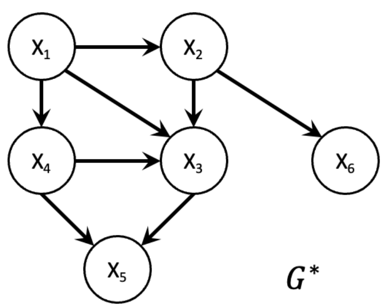
Get samples

	X_1	X_2	X_3	X_4	X_5	X_6
Sample 1	0.22	0.04	0.84	0.48	0.98	0.82
Sample 2	0.87	0.17	0.61	0.67	0.67	0.23
Sample 3	0.55	0.54	0.67	0.86	0.93	0.23
...	...	...	...	...	...	...
Sample M	0.12	0.95	0.79	0.47	0.05	0.92

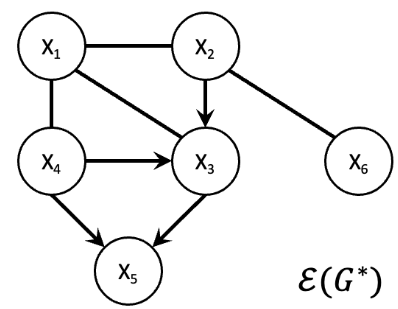
Do CI tests
- Recover skeleton
- Orient *some* edges

Conditional independence (CI) tests

- A standard way (under some causal assumptions) to recover graph structure from data is to perform CI tests
  - e.g. PC (Peter-Clark) algorithm [Spirtes, Glymour, Scheines, Heckerman 2000]



G^*



$\mathcal{E}(G^*)$

(Recover up to an equivalence class)

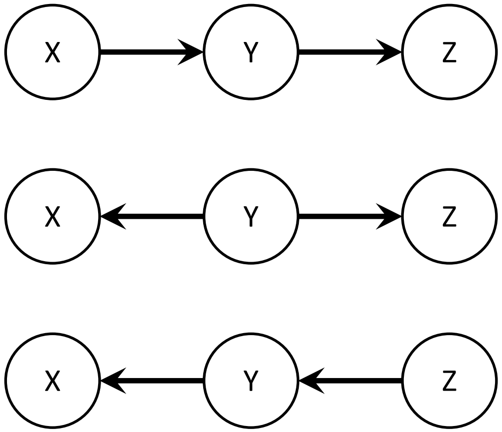
Get samples

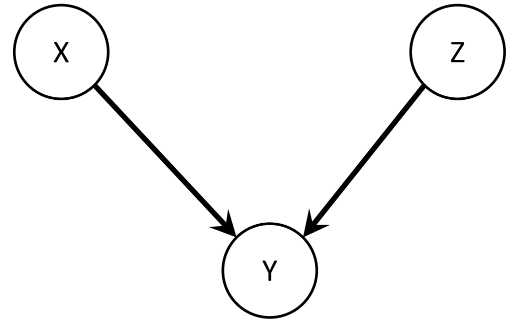
	X_1	X_2	X_3	X_4	X_5	X_6
Sample 1	0.22	0.04	0.84	0.48	0.98	0.82
Sample 2	0.87	0.17	0.61	0.67	0.67	0.23
Sample 3	0.55	0.54	0.67	0.86	0.93	0.23
...	...	...	...	...	...	...
Sample M	0.12	0.95	0.79	0.47	0.05	0.92

Do CI tests

- Recover skeleton
- Orient *some* edges

Unshielded colliders / v-structures

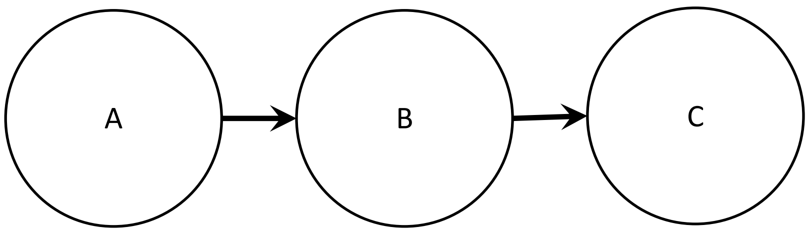




$X \not\perp\!\!\!\perp Y$
$X \not\perp\!\!\!\perp Z$
$Y \not\perp\!\!\!\perp Z$
$X \not\perp\!\!\!\perp Y \mid Z$
$X \perp\!\!\!\perp Z \mid Y$
$Y \not\perp\!\!\!\perp Z \mid X$

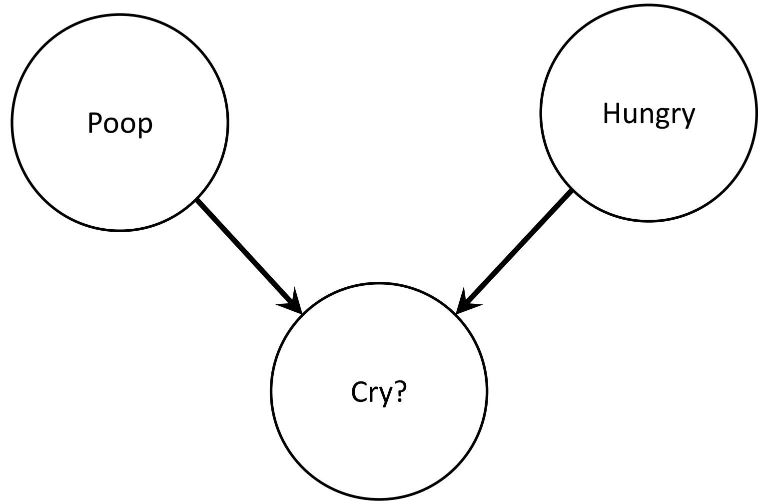
$X \not\perp\!\!\!\perp Y$
$X \perp\!\!\!\perp Z$
$Y \not\perp\!\!\!\perp Z$
$X \not\perp\!\!\!\perp Y \mid Z$
$X \not\perp\!\!\!\perp Z \mid Y$
$Y \not\perp\!\!\!\perp Z \mid X$

Example



$\not\perp\!\!\!\perp$

$\perp\!\!\!\perp$

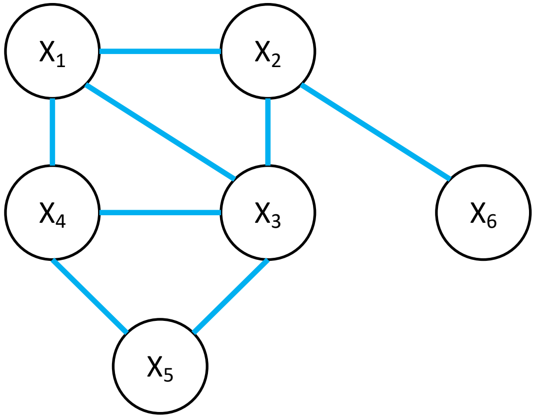


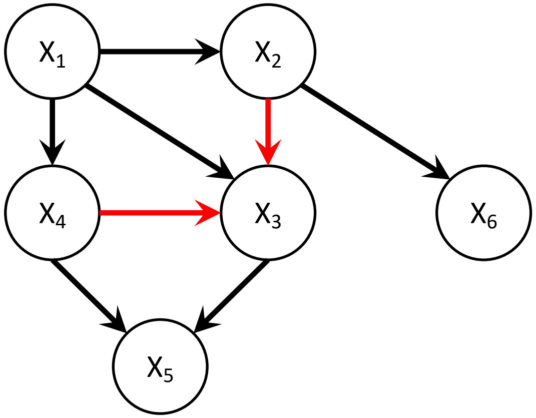
$\perp\!\!\!\perp$

$\not\perp\!\!\!\perp$

Markov Equivalence Class (MEC)

- Two DAGs are Markov equivalent if they encode the same CI relations
- skeleton and v-structures of DAG G^* earlier





- Theorem [Verma, Pearl 1990; Andersson, Madigan, Perlman 1997]

  G and G' are Markov equivalent **if and only if**

  1) G and G' have the same skeleton
  2) G and G' have the same v-structures

- For any DAG G^*, we use $[G^*]$ to denote its MEC

Essential graphs $\mathcal{E}(G^*)$

- Used to graphically represent a MEC $[\mathrm{G}^*]$
- DAGs in same MEC have the same essential graph

Essential graphs $\mathcal{E}(G^*)$

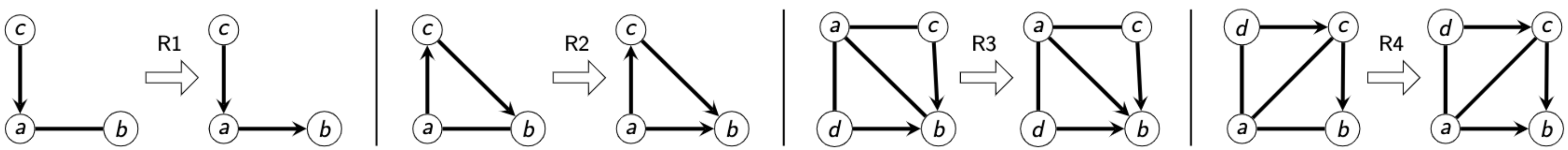
- Used to graphically represent a MEC $[\mathrm{G}^*]$
- DAGs in same MEC have the same essential graph
- Partially oriented DAG
  - $X \sim Y$ is oriented as $X \to Y$ if **all** DAGs in the MEC agree
  - $X \sim Y$ is unoriented arc if there **exists** disagreement
    - $\exists G_1, G_2 \in [G^*]$ in MEC such that $X \to Y$ in G_1 and $X \leftarrow Y$ in G_2.

Essential graphs $\mathcal{E}(G^*)$

- Used to graphically represent a MEC $[\mathrm{G}^*]$
- DAGs in same MEC have the same essential graph
- Partially oriented DAG
  - $X \sim Y$ is oriented as $X \to Y$ if **all** DAGs in the MEC agree
  - $X \sim Y$ is unoriented arc if there **exists** disagreement
    - $\exists G_1, G_2 \in [G^*]$ in MEC such that $X \to Y$ in G_1 and $X \leftarrow Y$ in G_2.
- How to compute essential graph $\mathcal{E}(G^*)$ of G^*?
  1. Look at skeleton of G^*
  2. Orient v-structures in G^*
  3. Apply Meek rules [Meek 1995]

Meek rules [Meek 1995]

- **Sound** and **complete**
  (with respect to arc orientations with acyclic completion)



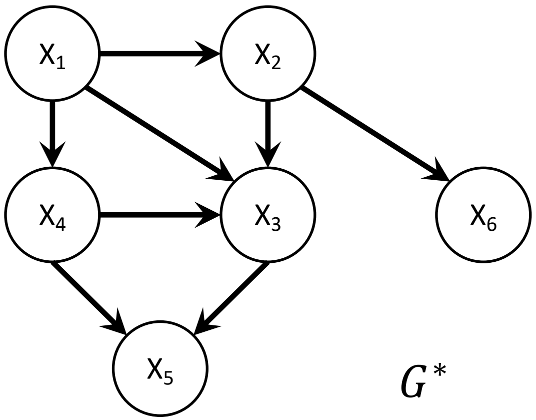
If $b \leftarrow a$,
then v-structure

If $b \leftarrow a$,
then cycle

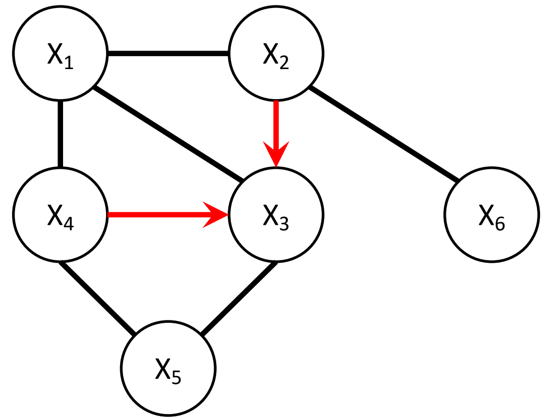
If $b \leftarrow a$, then unoriented arcs would have been oriented **in the same way** in all DAGs within the MEC (via R2)

- Converge in polynomial time [Wienöbst, Bannach, Liśkiewicz 2021]

Essential graph example

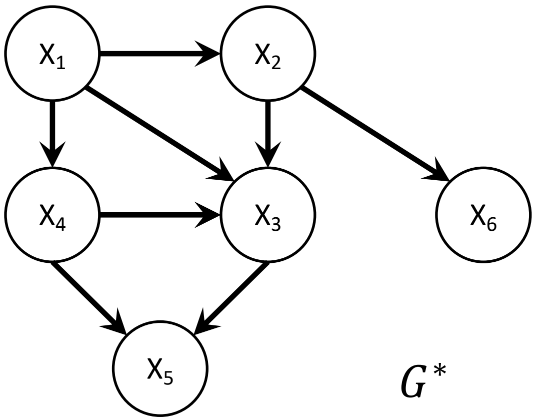


G^*

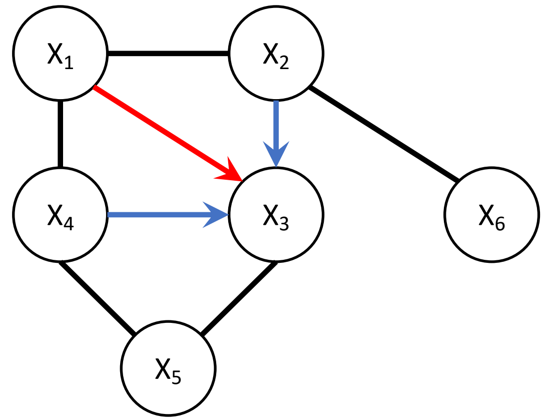


- Use CI tests: recover skeleton and v-structures

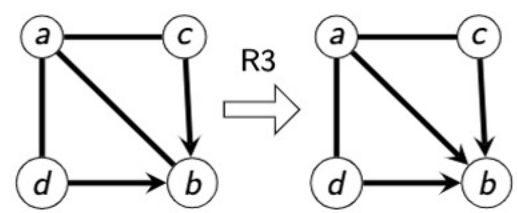
Essential graph example



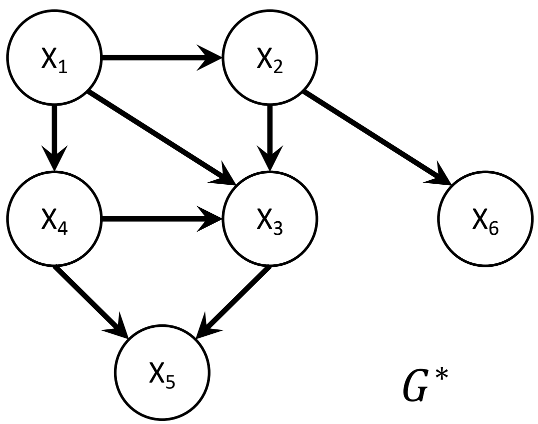
G^*

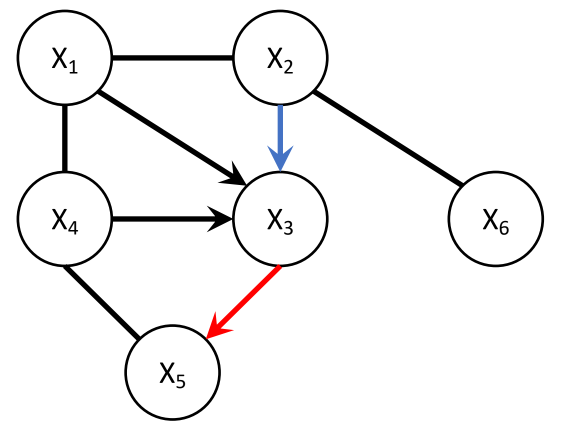


- Use CI tests: recover skeleton and v-structures
- Meek R3

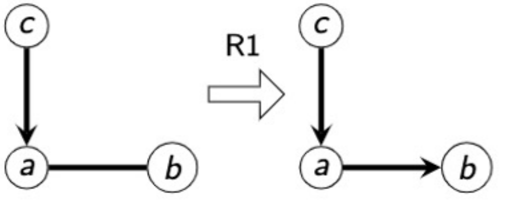


Essential graph example

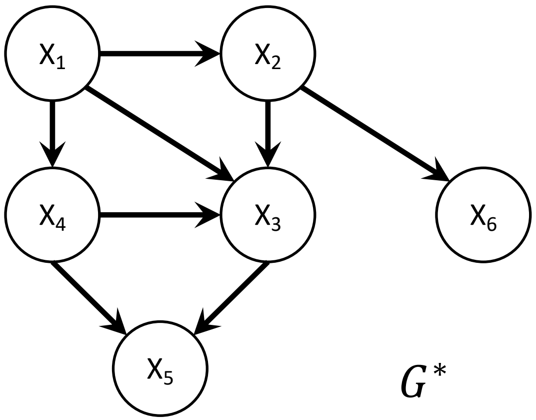




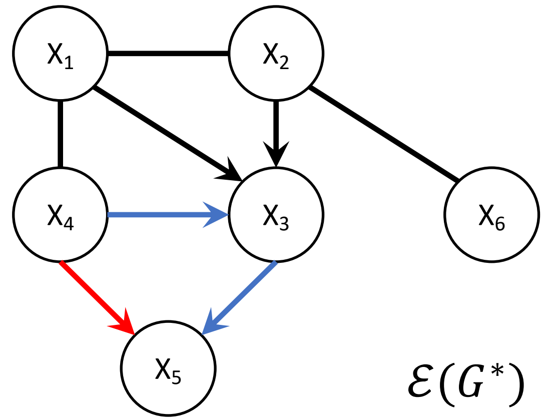
- Use CI tests: recover skeleton and v-structures
- Meek R3
- Meek R1



Essential graph example

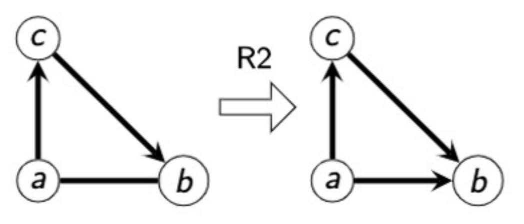


G^*

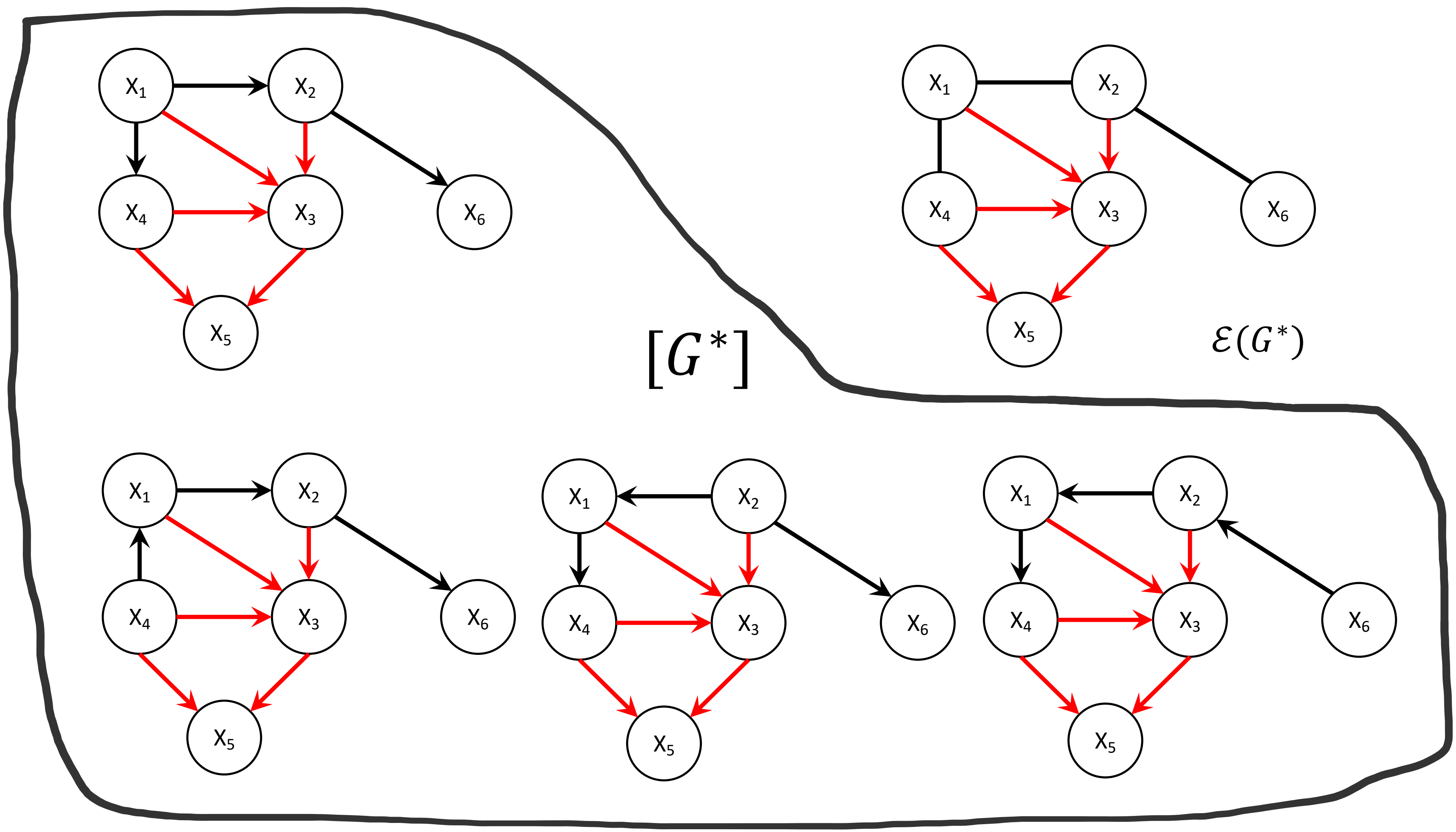


$\mathcal{E}(G^*)$

- Use CI tests: recover skeleton and v-structures
- Meek R3
- Meek R1
- Meek R2



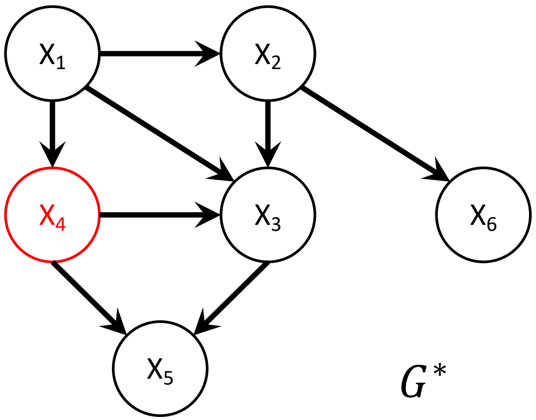
Essential graph example



For this talk…

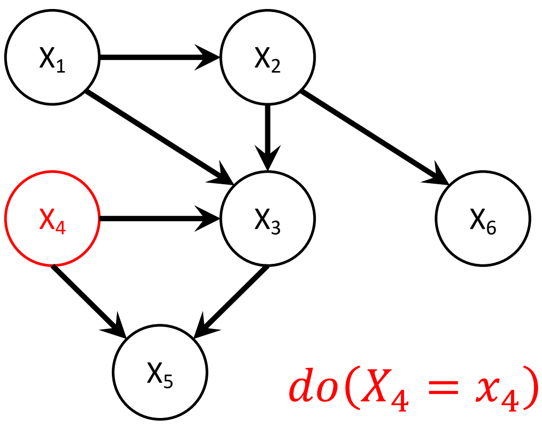
- Some standard causal assumptions
  - Causal sufficiency
  - Faithfulness
  - Can identify conditional independencies from data
- Simplifying assumptions for this talk
  - Atomic intervention: One vertex per intervention
  - Each vertex has unit cost
- Objective
  - Minimize total number of vertices intervened

Hard interventions



G^*

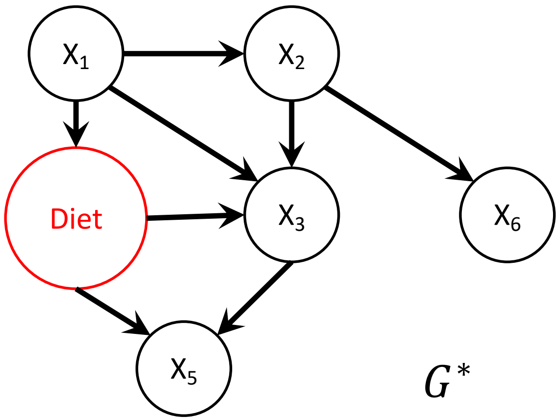
$$X_1 = f_1(\epsilon_1)$$
$$X_2 = f_2(X_1, \epsilon_2)$$
$$X_3 = f_3(X_1, X_2, X_4, \epsilon_3)$$
$$X_4 = f_4(X_1, \epsilon_4)$$
$$X_5 = f_5(X_3, X_4, \epsilon_5)$$
$$X_6 = f_6(X_2, \epsilon_6)$$
$\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5, \epsilon_6$ independent noise



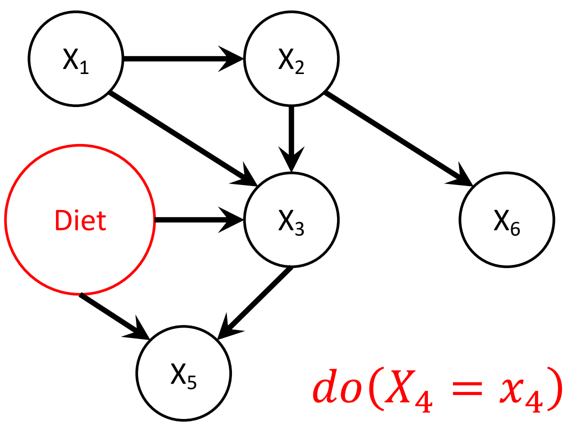
$do(X_4 = x_4)$

$$X_1 = f_1(\epsilon_1)$$
$$X_2 = f_2(X_1, \epsilon_2)$$
$$X_3 = f_3(X_1, X_2, X_4, \epsilon_3)$$
$$X_4 = \text{intervened value } x_4$$
$$X_5 = f_5(X_3, X_4, \epsilon_5)$$
$$X_6 = f_6(X_2, \epsilon_6)$$
$\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5, \epsilon_6$ independent noise

Hard interventions



$X_1 = f_1(\epsilon_1)$
$X_2 = f_2(X_1, \epsilon_2)$
$X_3 = f_3(X_1, X_2, X_4, \epsilon_3)$
$X_4 = f_4(X_1, \epsilon_4)$
$X_5 = f_5(X_3, X_4, \epsilon_5)$
$X_6 = f_6(X_2, \epsilon_6)$
$\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5, \epsilon_6$ independent noise



$X_1 = f_1(\epsilon_1)$
$X_2 = f_2(X_1, \epsilon_2)$
$X_3 = f_3(X_1, X_2, X_4, \epsilon_3)$
$X_4 =$ Eat Z apples a day
$X_5 = f_5(X_3, X_4, \epsilon_5)$
$X_6 = f_6(X_2, \epsilon_6)$
$\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5, \epsilon_6$ independent noise

What can we learn from data and hard interventions?

- Essential
- Skeleton
- V-structures
- Meek rules
- Property about chordal graph

Meek rules [Meek 1995] (Example)

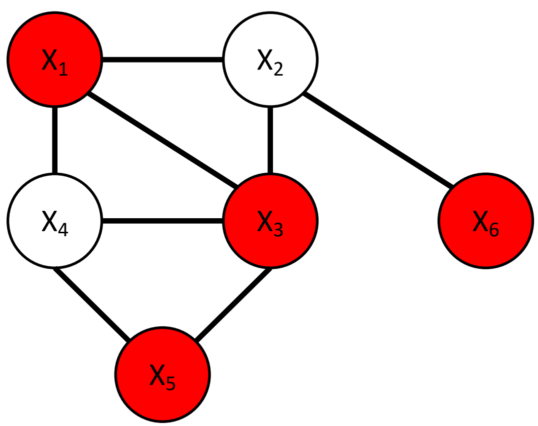
- Look at path. Explain R1
- Look at clique. Explain R2

Two classes of interventions

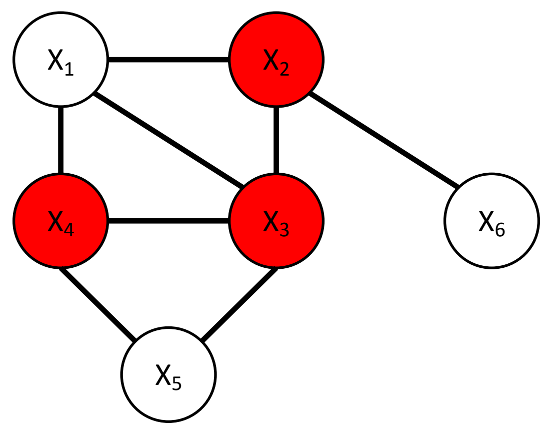
- Non-adaptive
  - Given MEC $[G^*]$, decide on a single fixed set of interventions that recovers *any possible* $G^* \in [G^*]$
  - Need to intervene on a *G-separating system*
    [Kocaoglu, Dimakis, Vishwanath 2017]
- Adaptive
  - Given MEC $[G^*]$,
    - Decide on first intervention
    - See outcome
    - Decide on second intervention
    - See outcome
    - …

G-separating system [Kocaoglu, Dimakis, Vishwanath 2017]

- Fix an undirected graph $G = (V, E)$
- A subset $\mathcal{J} \subseteq 2^V$ is a called a G-separating system if
  - For every edge $\{u, v\} \in E$, $\exists$ intervention $\mathrm{I} \in \mathcal{J}$ such that either $(u \in I \wedge v \notin I)$ or $(u \notin I \wedge v \in I)$
- For atomic interventions $\Rightarrow$ vertex cover



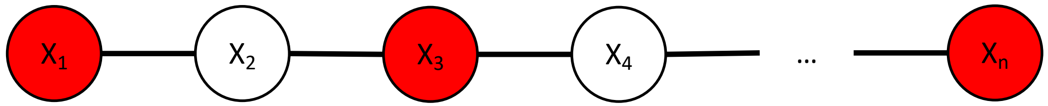
$|\mathcal{J}| = 4$



$|\mathcal{J}| = 3$

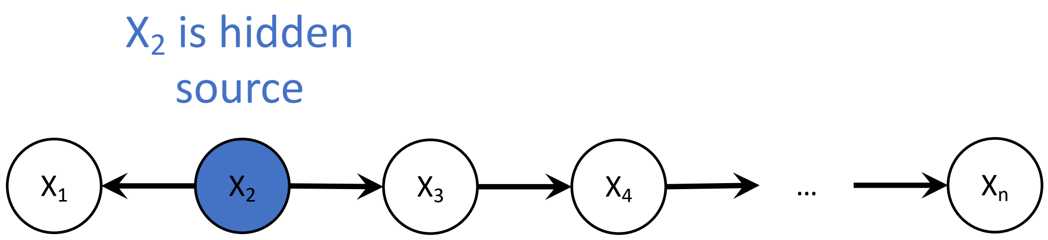
Power of adaptivity

- Path essential graph
  - n possible DAGs (pick a source node and orient away)
  - G-separating system needs $\geq \left\lfloor \frac{n}{2} \right\rfloor \in \Omega(n)$ vertices



Power of adaptivity

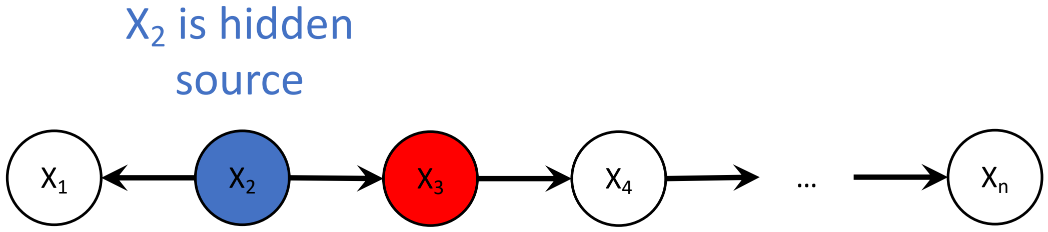
- Path essential graph
  - n possible DAGs (pick a source node and orient away)
  - G-separating system needs $\geq \left\lfloor \frac{n}{2} \right\rfloor \in \Omega(n)$ vertices



- Meanwhile, adaptive search can act like binary search!
  i.e. only $\mathcal{O}(\log n)$ interventions required

Power of adaptivity

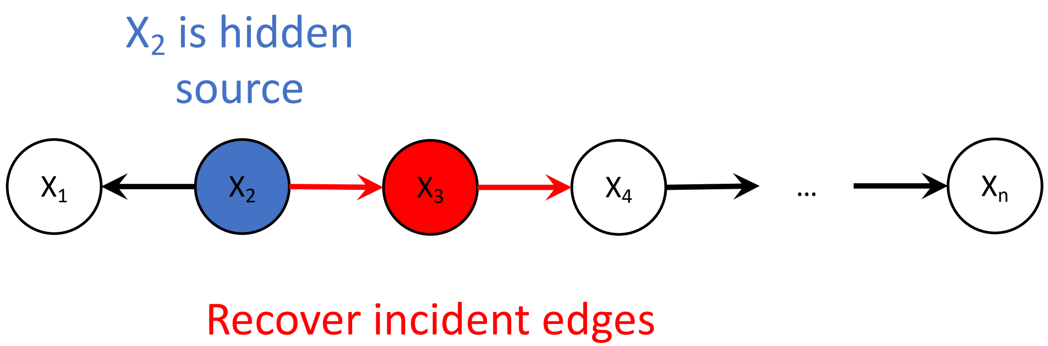
- Path essential graph
  - n possible DAGs (pick a source node and orient away)
  - G-separating system needs $\geq \left\lfloor \frac{n}{2} \right\rfloor \in \Omega(n)$ vertices



- Meanwhile, adaptive search can act like binary search!
  i.e. only $\mathcal{O}(\log n)$ interventions required

Power of adaptivity

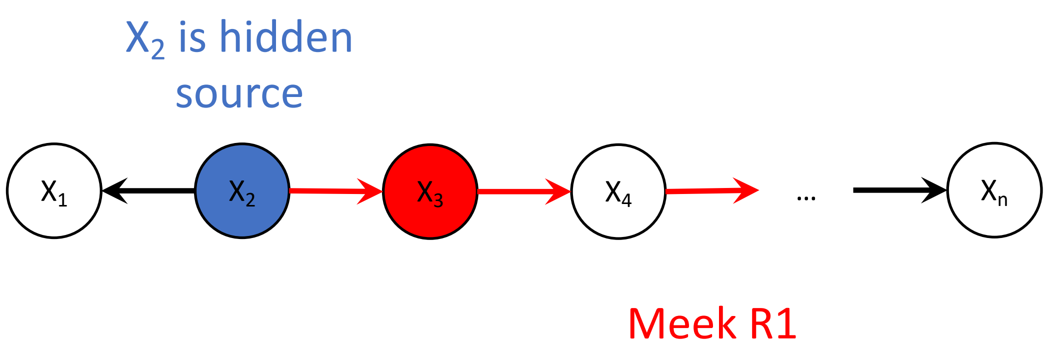
- Path essential graph
  - n possible DAGs (pick a source node and orient away)
  - G-separating system needs $\geq \left\lfloor \frac{n}{2} \right\rfloor \in \Omega(n)$ vertices



  - Meanwhile, adaptive search can act like binary search!
    i.e. only $\mathcal{O}(\log n)$ interventions required

Power of adaptivity

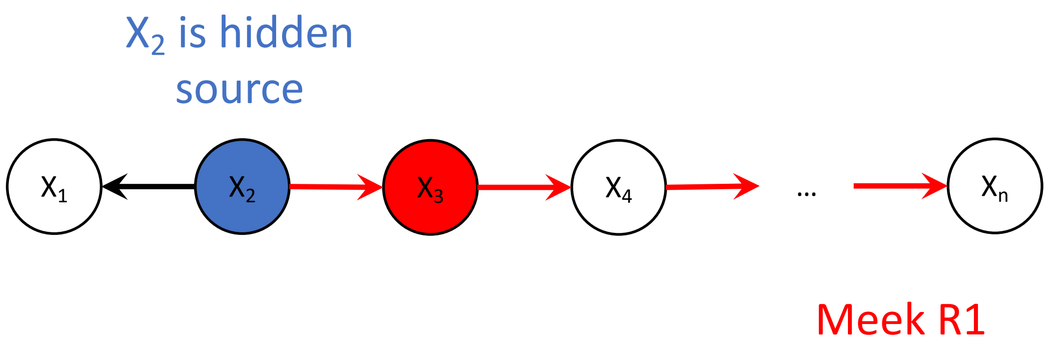
- Path essential graph
  - n possible DAGs (pick a source node and orient away)
  - G-separating system needs $\geq \left\lfloor \frac{n}{2} \right\rfloor \in \Omega(n)$ vertices



  - Meanwhile, adaptive search can act like binary search!
    i.e. only $\mathcal{O}(\log n)$ interventions required

Power of adaptivity

- Path essential graph
  - n possible DAGs (pick a source node and orient away)
  - G-separating system needs $\geq \left\lfloor \frac{n}{2} \right\rfloor \in \Omega(n)$ vertices

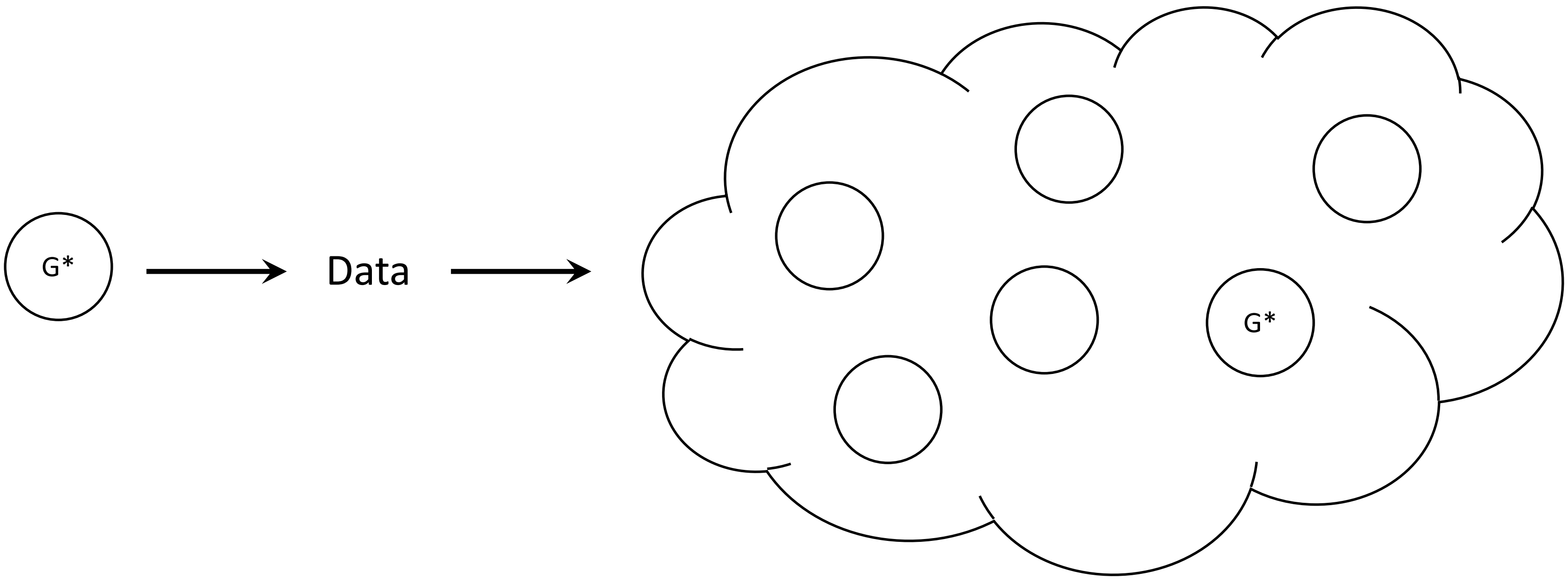


  - Meanwhile, adaptive search can act like binary search!
    i.e. only $\mathcal{O}(\log n)$ interventions required

Problem setup

Identify G*

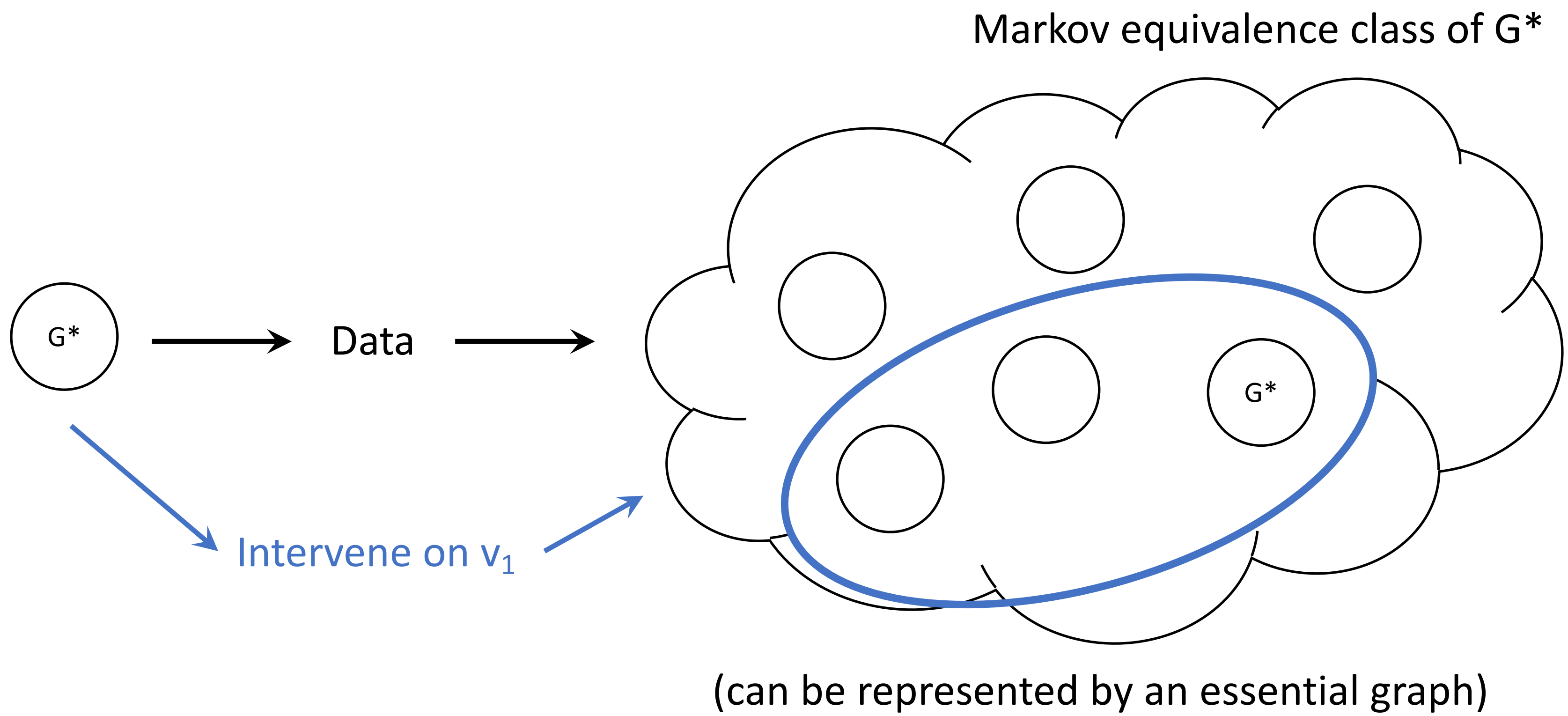
Markov equivalence class of G*



(can be represented by an essential graph)

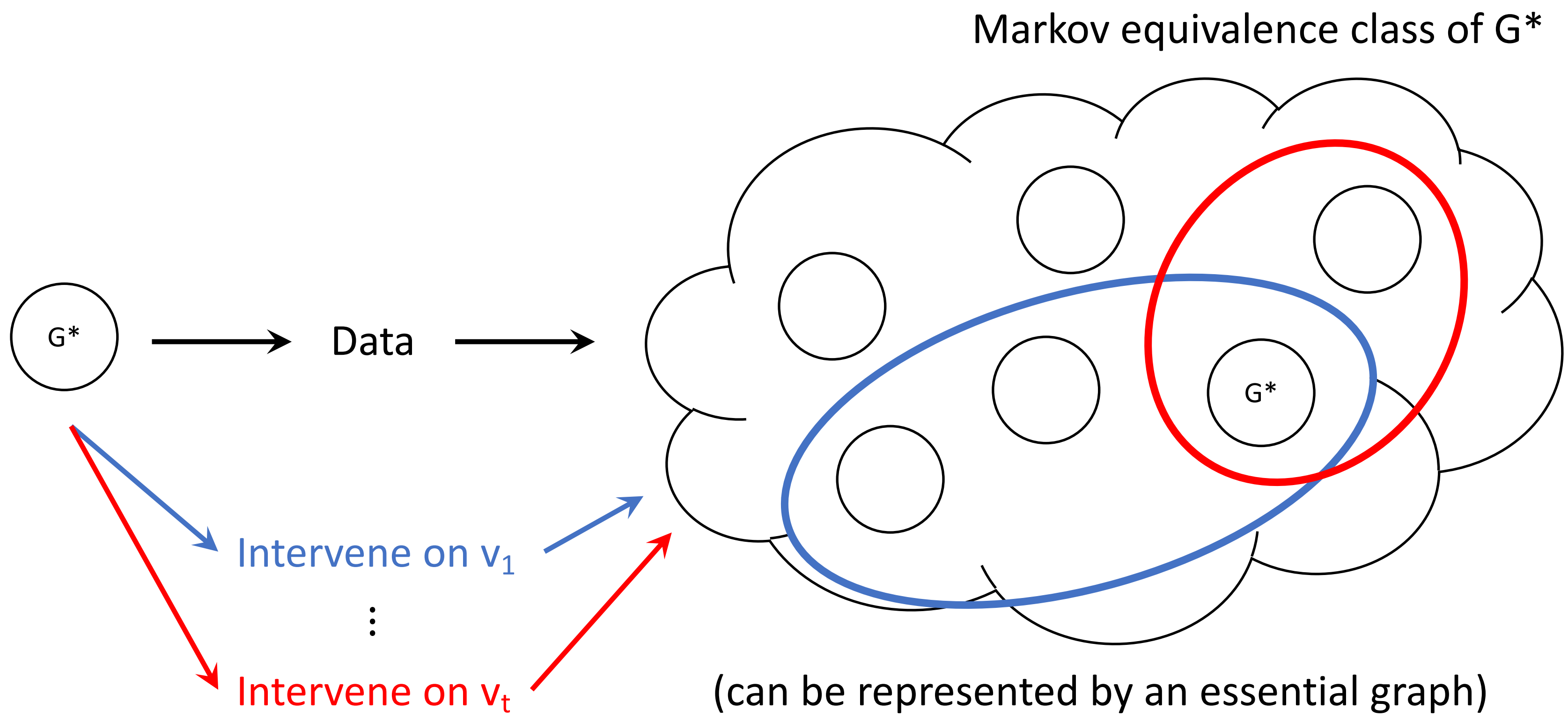
Problem setup

Identify G* using **interventions**



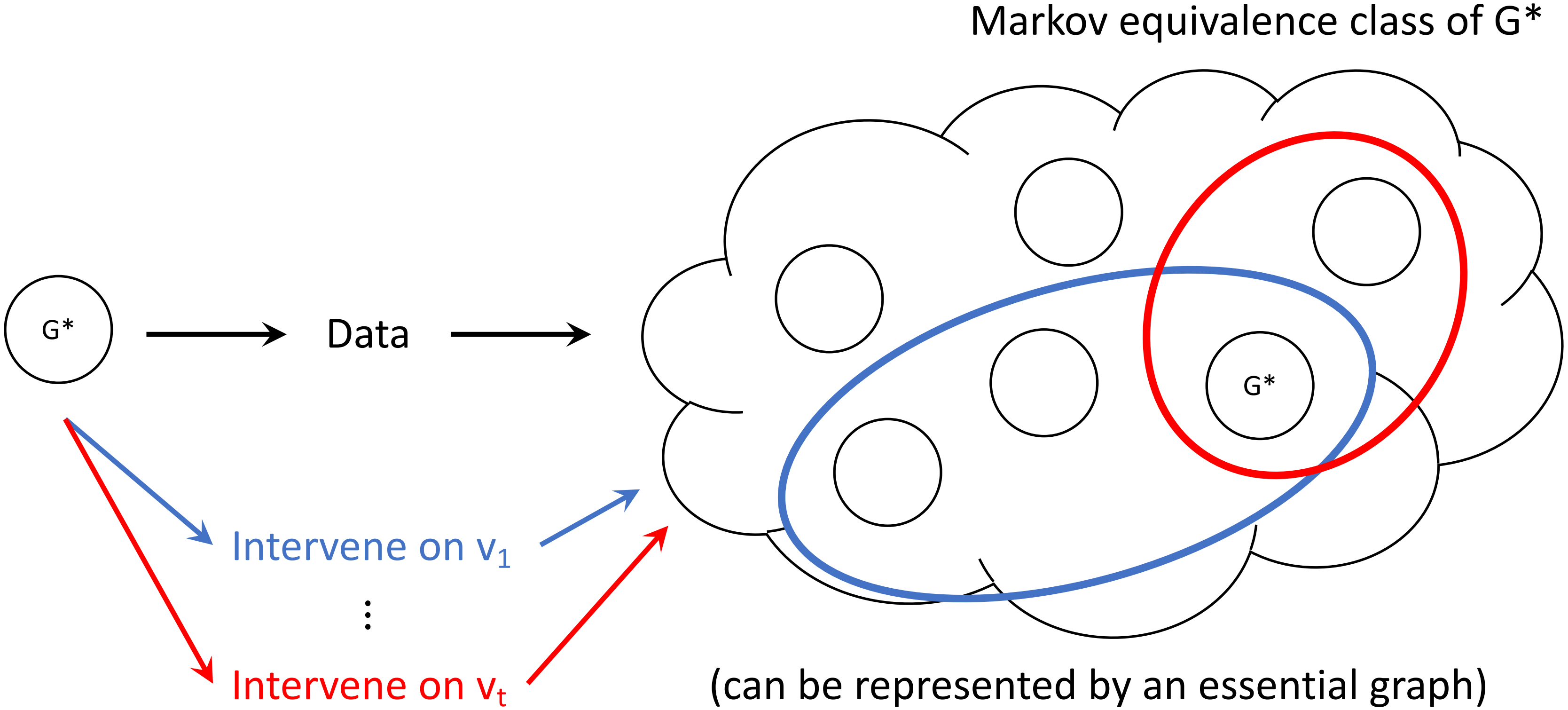
Problem setup

Identify G* using **interventions**



Problem setup

Identify G* using **as few interventions as possible** (minimize t)



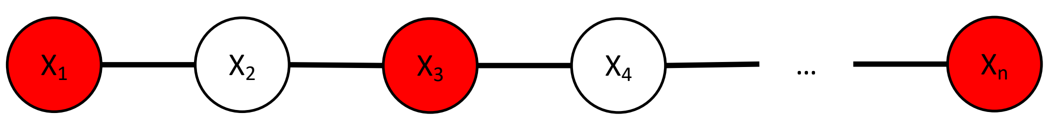
(can be represented by an essential graph)

The verification problem

- Given MEC $[G^*]$ and some $\mathrm{G} \in [G^*]$, check whether $\mathrm{G} = G^*$ using interventions
  - Denote the minimum number required by $\nu(\mathrm{G})$
  - $\nu(G^*)$ is **lower bound** for **searching** for G^* within $[G^*]$

The verification problem

- Given MEC $[G^*]$ and some $\mathrm{G} \in [G^*]$, check whether $\mathrm{G} = G^*$ using interventions
  - Denote the minimum number required by $\nu(\mathrm{G})$
  - $\nu(G^*)$ is **lower bound** for **searching** for G^* within $[G^*]$
- Trivial solution
  - Compute minimum vertex cover on all unoriented arcs of the essential graph $\mathcal{E}(G) = \mathcal{E}(G^*)$
  - Check if revealed orientations agree with G
  - Worst case: $\Omega(n)$ interventions, e.g. on a line



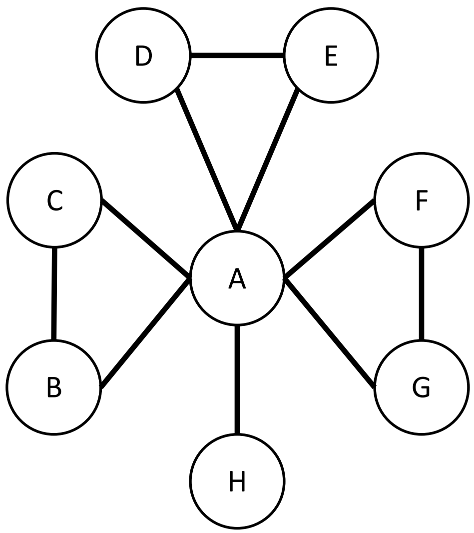
What was known

Maximal clique size

1. $\nu(\mathrm{G}) \geq \left\lfloor \frac{\omega(G)}{2} \right\rfloor$ [Squires, Magliacane, Greenewald, Katz, Kocaoglu, Shanmugam 2020]

Number of maximal cliques

2. $\left\lceil \frac{n-r}{2} \right\rceil \leq \nu(\mathrm{G}) \leq n - r$ [Porwal, Srivastava, Sinha 2022]



MEC $[G^*]$

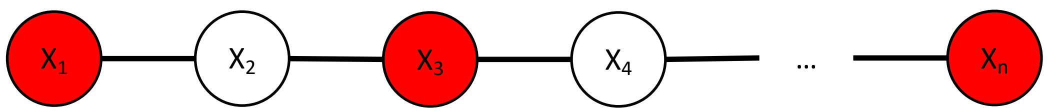
n = 8, $\omega(G)$ = 3, r = 4

1. $1 \leq \nu(\mathrm{G})$
2. $2 \leq \nu(\mathrm{G}) \leq 4$

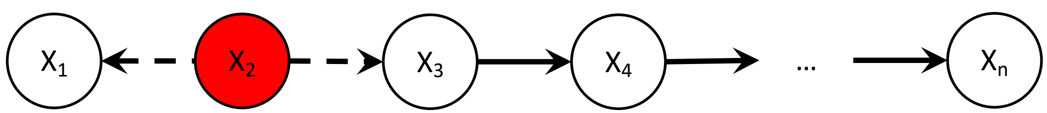
Characterization via covered edges

- A set $\mathcal{I} \subseteq V$ is a verifying set for DAG $\mathrm{G} = (\mathrm{V}, \mathrm{E})$ if and only if $\mathcal{I}$ is a **minimum vertex cover** of the **covered edges** [Chickering 1995] of G
  - $u \sim v$ is covered edge if they share same parents

Naïve:



Our characterization:



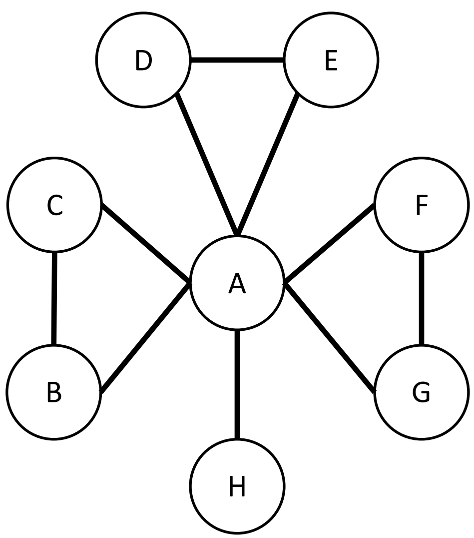
X_2 is source in G

Comparison

Maximal clique size

Number of maximal cliques

1. $\nu(\mathrm{G}) \geq \left\lfloor \frac{\omega(G)}{2} \right\rfloor$ [SMG+20]
2. $\left\lceil \frac{n-r}{2} \right\rceil \leq \nu(\mathrm{G}) \leq n - r$ [PSS22]



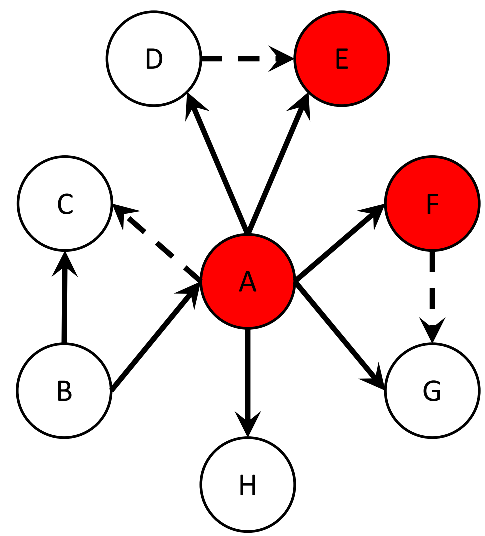
MEC $[G^*]$

n = 8, $\omega(G)$ = 3, r = 4

1. $1 \leq \nu(\mathrm{G})$
2. $2 \leq \nu(\mathrm{G}) \leq 4$

We can compute **exact** $\nu(\mathrm{G})$ for any given $\mathrm{G} \in [G^*]$

In fact, $\nu(\mathrm{G}) \in \{3,4\}$ for any $\mathrm{G} \in [G^*]$



One possible DAG from $[G^*]$

Efficient computation

- Wait… minimum vertex cover is NP-hard in general!

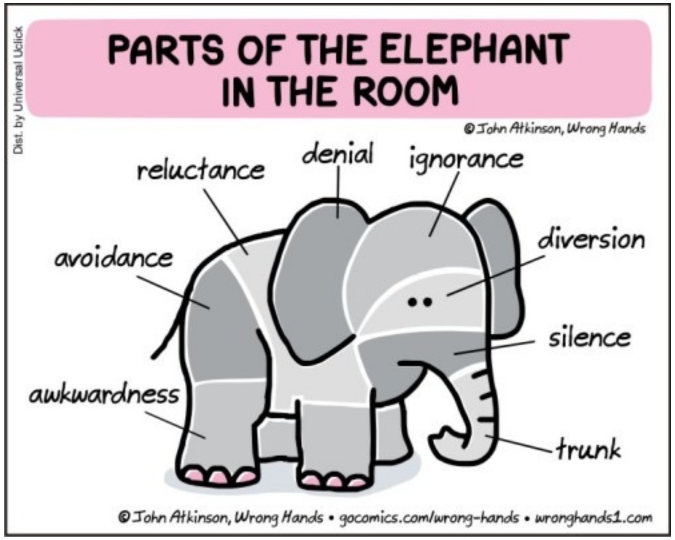
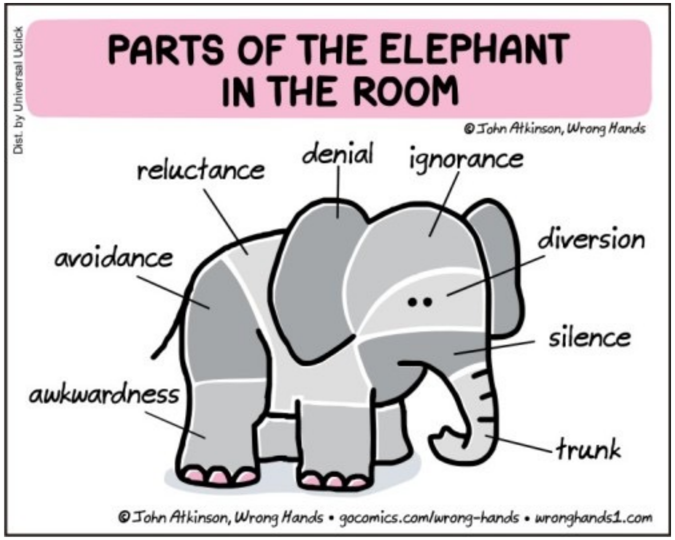


Image credit: https://wronghands1.com/2018/01/12/parts-of-the-elephant-in-the-room/

Efficient computation

- Wait… minimum vertex cover is NP-hard in general!



- Claim: Covered edges induce a forest
- Implication: $\nu(\mathrm{G})$ can be computed **exactly** via DP

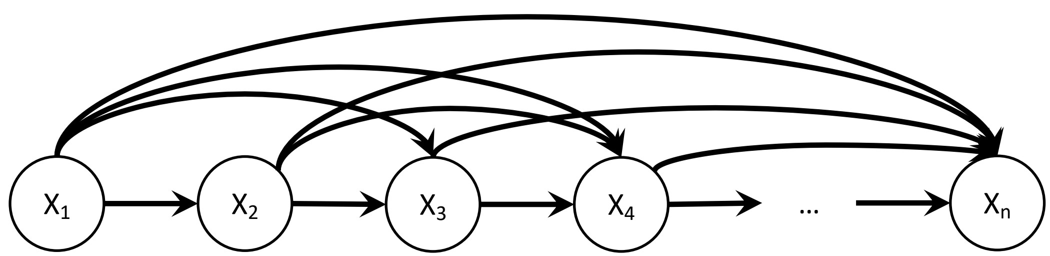
Image credit: https://wronghands1.com/2018/01/12/parts-of-the-elephant-in-the-room/

Through the lens of covered edges

- Covered edges cannot have both endpoints as sink of any maximal clique, so $\nu(\mathrm{G}) \leq n - r$

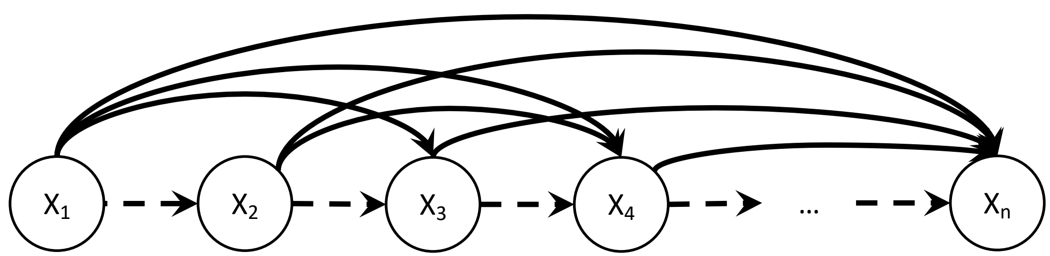
Through the lens of covered edges

- Covered edges cannot have both endpoints as sink of any maximal clique, so $\nu(\mathrm{G}) \leq n - r$
- G is a clique ⇒ Prior work: $\nu(\mathrm{G}) = \left\lfloor \frac{n}{2} \right\rfloor$



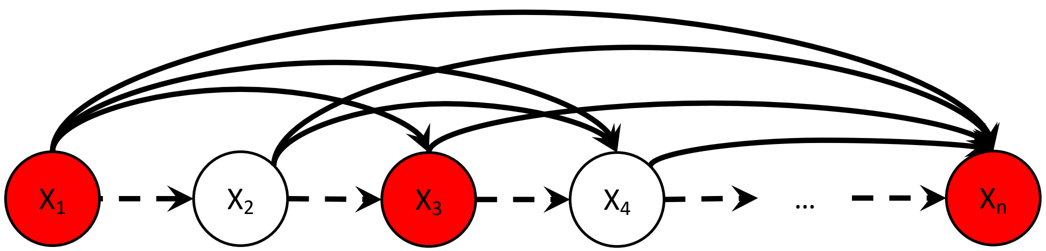
Through the lens of covered edges

- Covered edges cannot have both endpoints as sink of any maximal clique, so $\nu(\mathrm{G}) \leq n - r$
- G is a clique $\Rightarrow$ Prior work: $\nu(\mathrm{G}) = \left\lfloor \frac{n}{2} \right\rfloor$



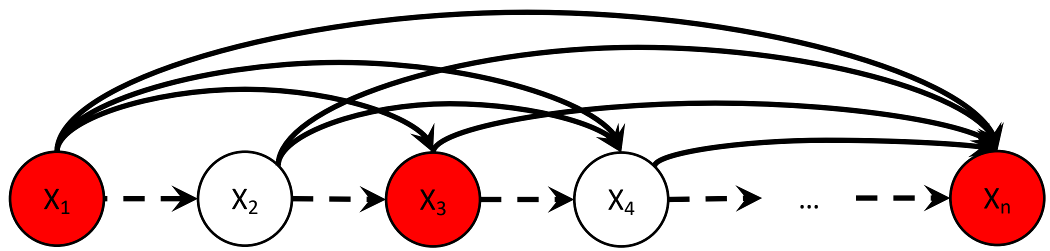
Through the lens of covered edges

- Covered edges cannot have both endpoints as sink of any maximal clique, so $\nu(\mathrm{G}) \leq n - r$
- G is a clique $\Rightarrow$ Prior work: $\nu(\mathrm{G}) = \left\lfloor \frac{n}{2} \right\rfloor$

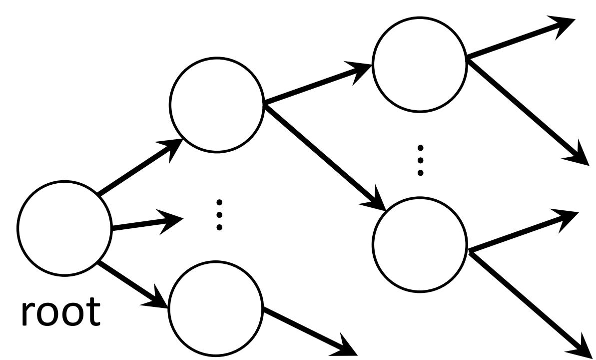


Through the lens of covered edges

- Covered edges cannot have both endpoints as sink of any maximal clique, so $\nu(\mathrm{G}) \leq n - r$
- G is a clique $\Rightarrow$ Prior work: $\nu(\mathrm{G}) = \left\lfloor \frac{n}{2} \right\rfloor$

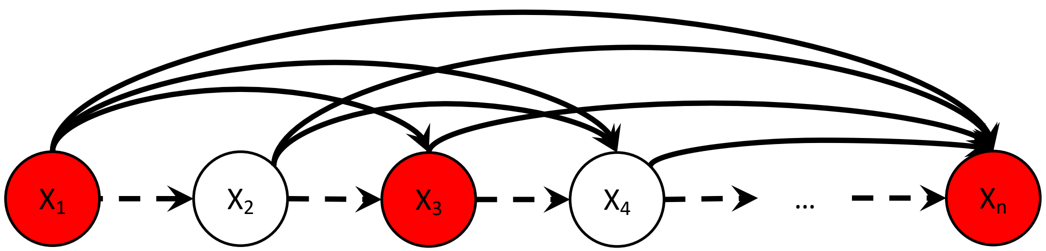


- G is a tree $\Rightarrow$
  Prior work: $\nu(\mathrm{G}) = 1$

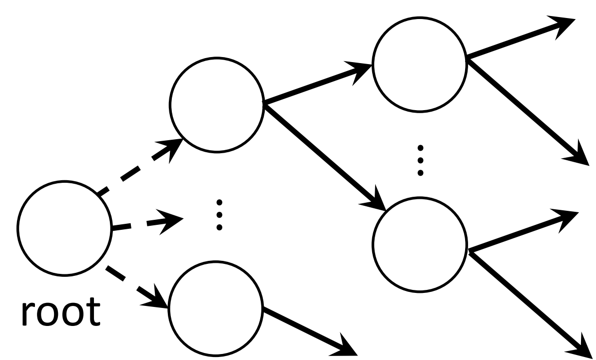


Through the lens of covered edges

- Covered edges cannot have both endpoints as sink of any maximal clique, so $\nu(\mathrm{G}) \leq n - r$
- G is a clique $\Rightarrow$ Prior work: $\nu(\mathrm{G}) = \left\lfloor \frac{n}{2} \right\rfloor$

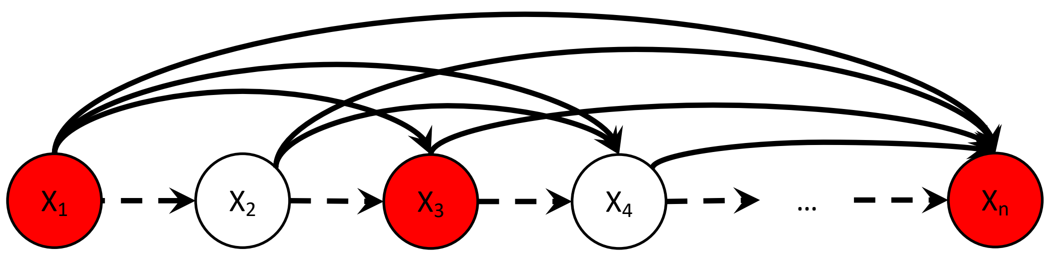


- G is a tree $\Rightarrow$
  Prior work: $\nu(\mathrm{G}) = 1$

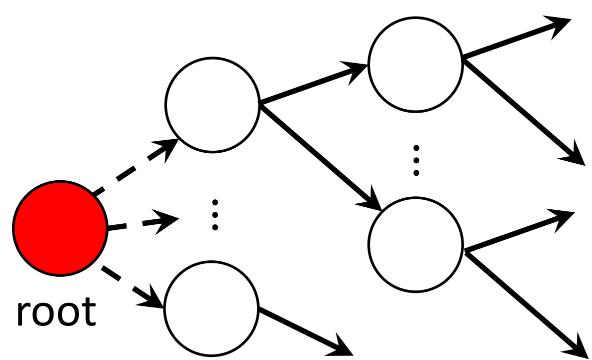


Through the lens of covered edges

- Covered edges cannot have both endpoints as sink of any maximal clique, so $\nu(\mathrm{G}) \leq n - r$
- G is a clique ⇒ Prior work: $\nu(\mathrm{G}) = \left\lfloor \frac{n}{2} \right\rfloor$



- G is a tree ⇒
  Prior work: $\nu(\mathrm{G}) = 1$



Through the lens of covered edges

- For non-adaptive interventions, we must intervene on a G-separating system

Through the lens of covered edges

- For non-adaptive interventions, we must intervene on a G-separating system
  - Two graphs have the same MEC $[G^*]$ **if and only if** there is a sequence of covered edge reversals that transform between them [Chickering 1995]

Through the lens of covered edges

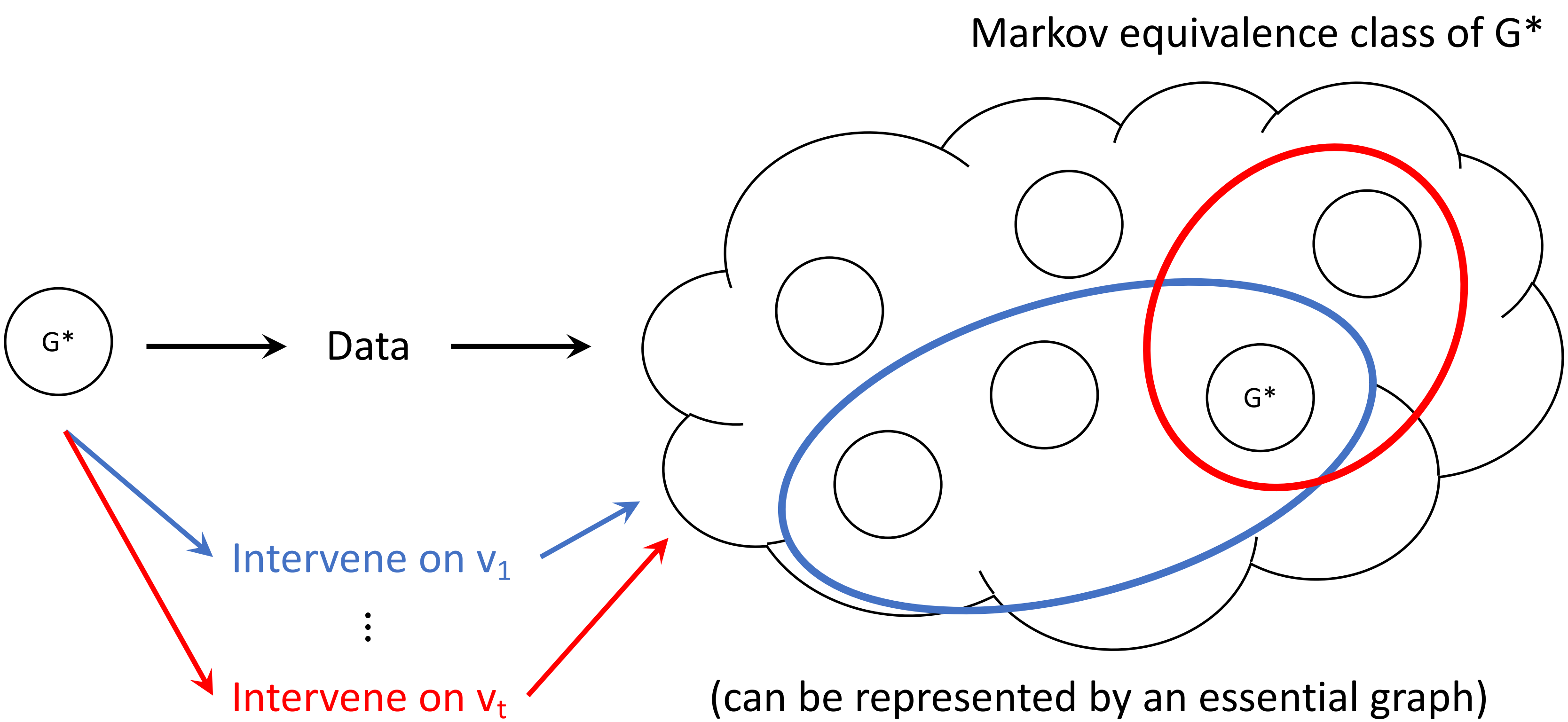
- For non-adaptive interventions, we must intervene on a G-separating system
  - Two graphs have the same MEC $[G^*]$ **if and only if** there is a sequence of covered edge reversals that transform between them [Chickering 1995]
  - Unoriented in $[G^*] \Rightarrow$ Covered edge in *some* $\mathrm{G} \in [G^*]$

Through the lens of covered edges

- For non-adaptive interventions, we must intervene on a G-separating system
  - Two graphs have the same MEC $[G^*]$ **if and only if** there is a sequence of covered edge reversals that transform between them [Chickering 1995]
  - Unoriented in $\mathcal{E}(G^*) \Rightarrow$ Covered edge in *some* $\mathrm{G} \in [G^*]$
  - So, "non-adaptive must cut all unoriented in $\mathcal{E}(G^*)$", i.e. a G-separating system

Problem setup

Identify G* using **as few interventions as possible** (minimize t)



The search problem

- Given MEC $[G^*]$ and recover G^* using interventions
  - We know at least $\nu(G^*)$ is necessary
  - Prior works only have guarantees for special classes of graphs: cliques, trees, intersection incomparable, etc.

The search problem

- Given MEC $[G^*]$ and recover G^* using interventions
  - We know at least $\nu(G^*)$ is necessary
  - Prior works only have guarantees for special classes of graphs: cliques, trees, intersection incomparable, etc.
- Punchline: $\mathcal{O}\big(\log n \cdot \nu(G^*)\big)$ interventions suffice
  - "Search is almost as easy as verification"

The search problem

- Given MEC $[G^*]$ and recover G^* using interventions
  - We know at least $\nu(G^*)$ is necessary
  - Prior works only have guarantees for special classes of graphs: cliques, trees, intersection incomparable, etc.
- Punchline: $\mathcal{O}\left(\log n \cdot \nu(G^*)\right)$ interventions suffice
  - “Search is almost as easy as verification”
  - Algorithm does not even know what $\nu(G^*)$ is!

The search problem

- Given MEC $[G^*]$ and recover G^* using interventions
  - We know at least $\nu(G^*)$ is necessary
  - Prior works only have guarantees for special classes of graphs: cliques, trees, intersection incomparable, etc.
- Punchline: $\mathcal{O}\big(\log n \cdot \nu(G^*)\big)$ interventions suffice
  - “Search is almost as easy as verification”
  - Algorithm does not even know what $\nu(G^*)$ is!
  - $\Omega(\log n)$ is unavoidable when $[G^*]$ is a path on n nodes
    - $\nu(G^*) = 1$
    - “Cannot do better than binary search”

The adaptive search algorithm

- Intervene and remove oriented arcs ⇒ Chordal graph. Handle each connected component [Hauser, Bühlmann 2012, 2014]
- For any chordal graph G, one can compute in polynomial time a clique separator C [Gilbert, Rose, Edenbrandt 1984]

The adaptive search algorithm

- Intervene and remove oriented arcs ⇒ Chordal graph. Handle each connected component [Hauser, Bühlmann 2012, 2014]
- For any chordal graph G, one can compute in polynomial time a clique separator C [Gilbert, Rose, Edenbrandt 1984]
  - $|A|, |B| \leq \frac{|G|}{2}$

The adaptive search algorithm

- Intervene and remove oriented arcs $\Rightarrow$ Chordal graph. Handle each connected component [Hauser, Bühlmann 2012, 2014]
- For any chordal graph G, one can compute in polynomial time a clique separator C [Gilbert, Rose, Edenbrandt 1984]
  - $|A|, |B| \leq \frac{|G|}{2}$; C is a clique, i.e. $|C| \leq \omega(G)$

The adaptive search algorithm

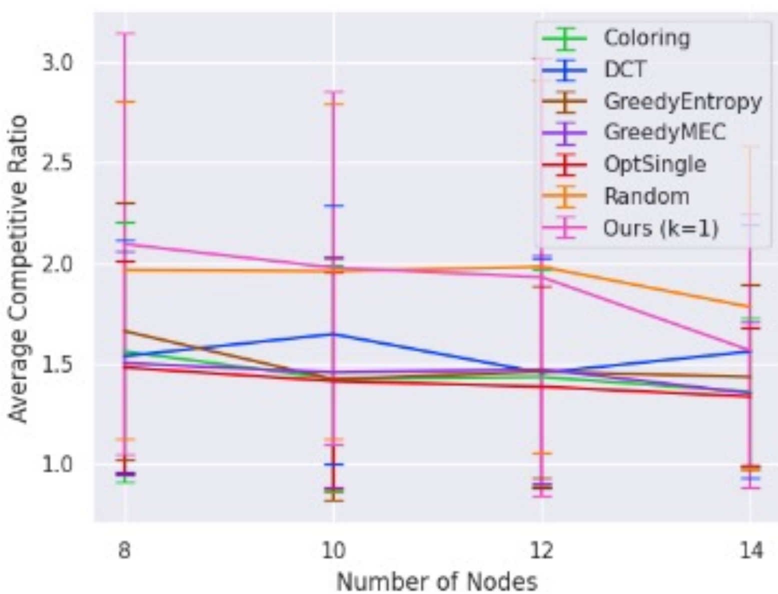
- Intervene and remove oriented arcs $\Rightarrow$ Chordal graph. Handle each connected component [Hauser, Bühlmann 2012, 2014]
- For any chordal graph G, one can compute in polynomial time a clique separator C [Gilbert, Rose, Edenbrandt 1984]
  - $|A|, |B| \leq \frac{|G|}{2}$; C is a clique, i.e. $|C| \leq \omega(G)$
- Algorithm: Find clique separator C_H in each component H; Intervene on all nodes in C_H's; Recurse

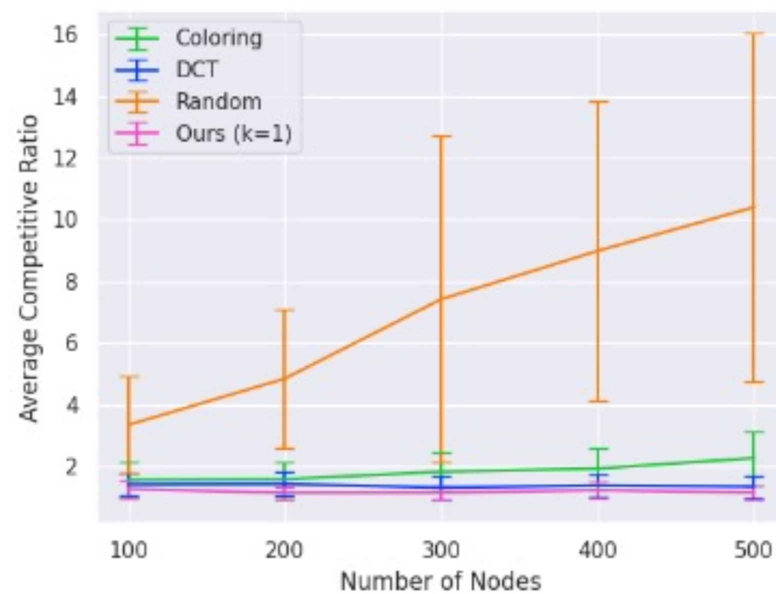
The adaptive search algorithm

- Intervene and remove oriented arcs ⇒ Chordal graph. Handle each connected component [Hauser, Bühlmann 2012, 2014]
- For any chordal graph G, one can compute in polynomial time a clique separator C [Gilbert, Rose, Edenbrandt 1984]
  - $|A|, |B| \leq \frac{|G|}{2}$; C is a clique, i.e. $|C| \leq \omega(G)$
- Algorithm: Find clique separator C_H in each component H; Intervene on all nodes in C_H's; Recurse
- Analysis:
  - $\mathcal{O}(\log n)$ rounds suffices ← [Gilbert, Rose, Edenbrandt 1984]
  - $\mathcal{O}\left(\nu(G^*)\right)$ per round ← We prove new lower bound on $\nu(G^*)$

The adaptive search algorithm

- Qualitatively, our algorithm is competitive with state-of-the-art adaptive search algorithms
  - We run $\sim 10\times$ faster in some experiments

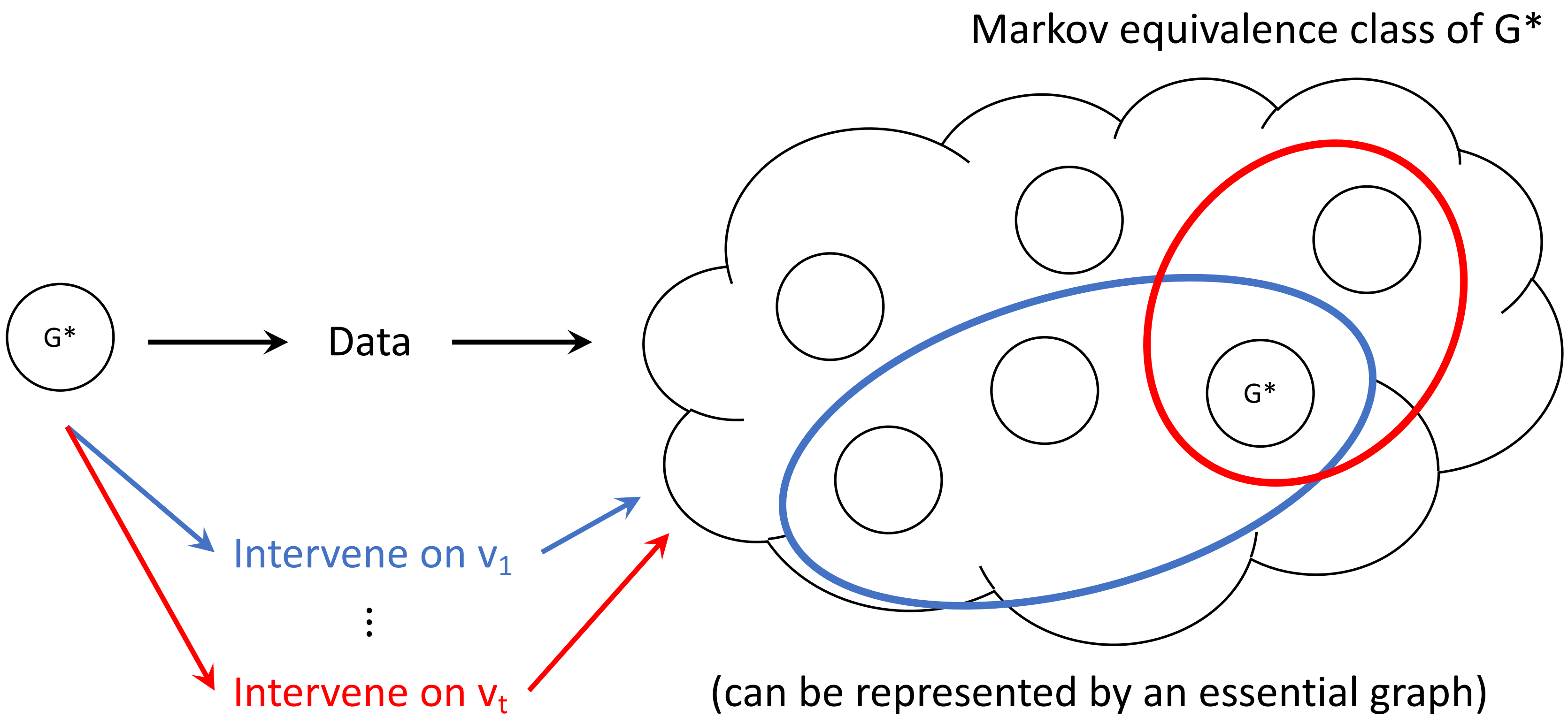




Implementation and more experimental results: https://github.com/cxjdavin/verification-and-search-algorithms-for-causal-DAGs

Problem setup

Identify G* using **as few interventions as possible** (minimize t)



Verification: $\nu(G^*)$ = size of minimum vertex cover of covered edges [CSB22]
Search: $\mathcal{O}\left(\log n \cdot \nu(G^*)\right)$ interventions suffice [CSB22]

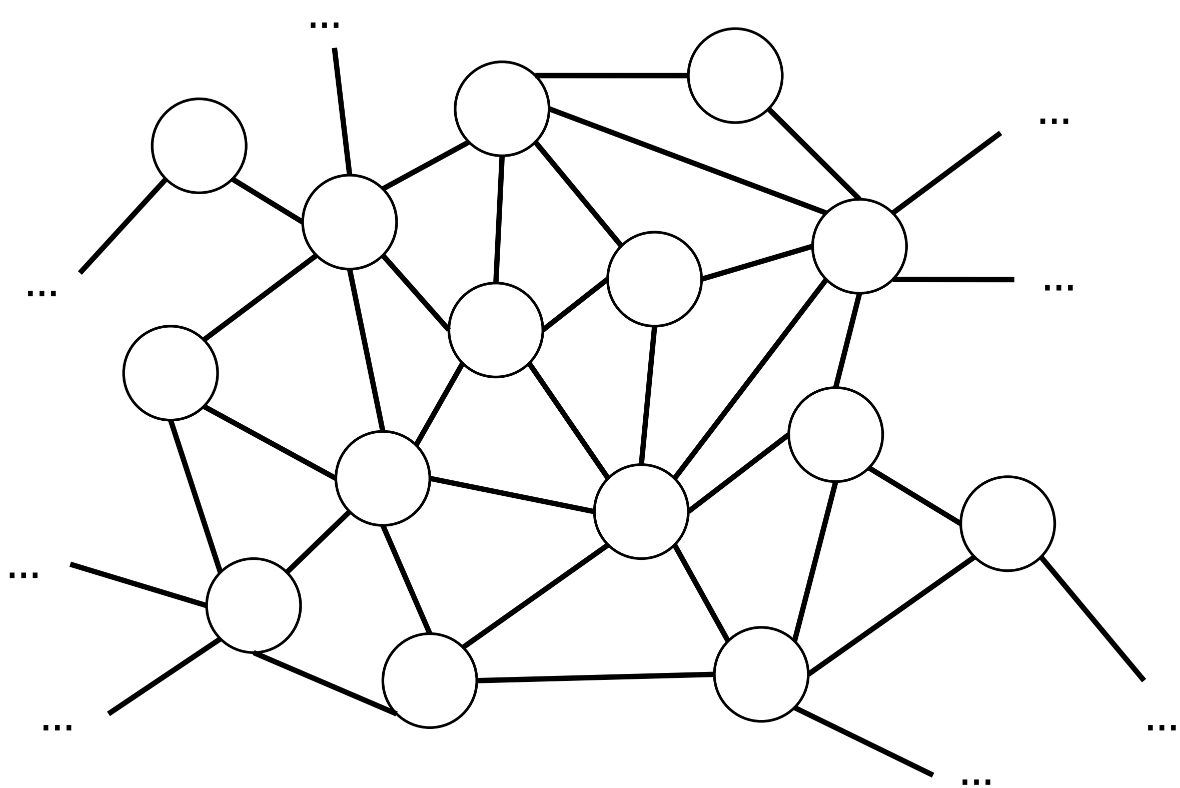
But wait, there’s more!

Other extensions / questions

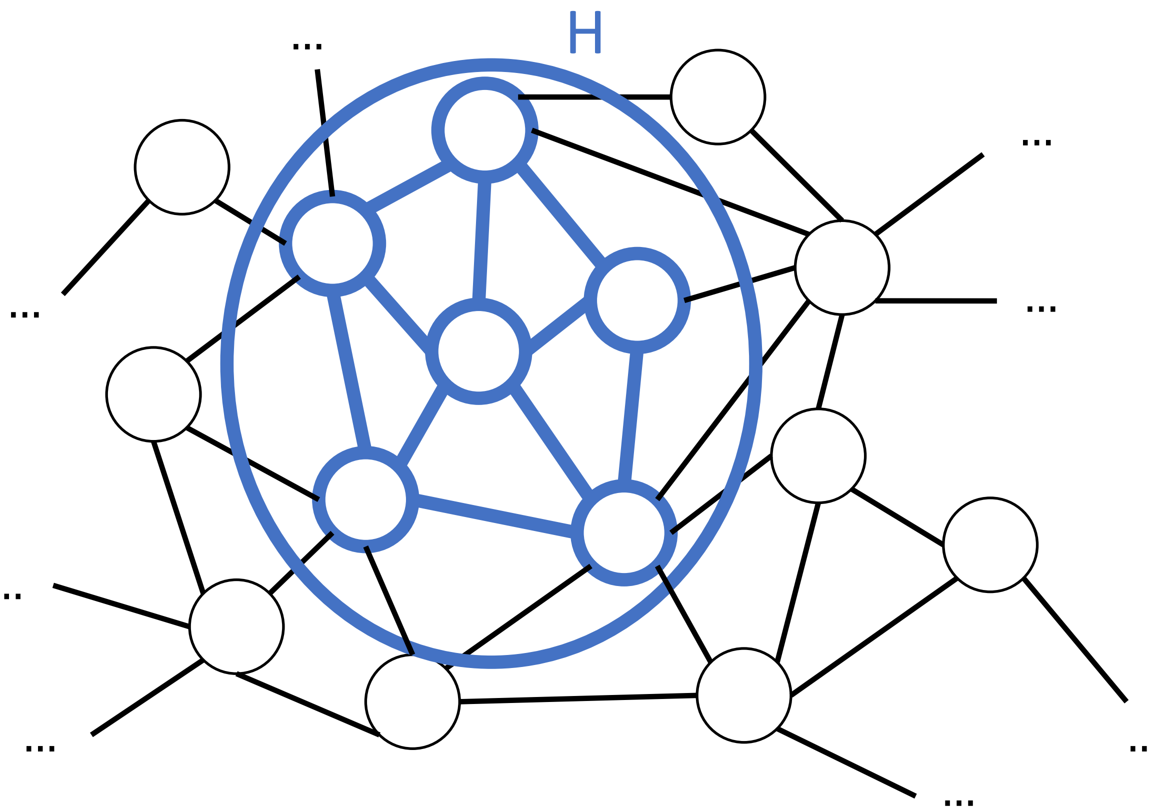
- What if the causal graph is HUGE?
- What if we consult domain experts for advice?
- What if we intervene >1 vertex per intervention?
  - Bounded size interventions
- What if vertices have different interventional costs?
  - Additive cost $\Rightarrow$ cost of intervention is cost of all vertices in it
- What if we have limited rounds of adaptivity?
  - At most r rounds, for $r < \log n$
- What if we have finite samples?
  - May incur error in conditional independence checks
- Can we weaken/remove the causal assumptions?
  - What if we don't have hard interventions? Soft/unknown interventions, etc
  - What if there are hidden confounders?
  - What if we don't have faithfulness?

Note: Some are in the papers mentioned earlier in this talk; some are solved and under submission. If you are interested, come talk to me to find out more ☺

What if causal graph is HUGE?



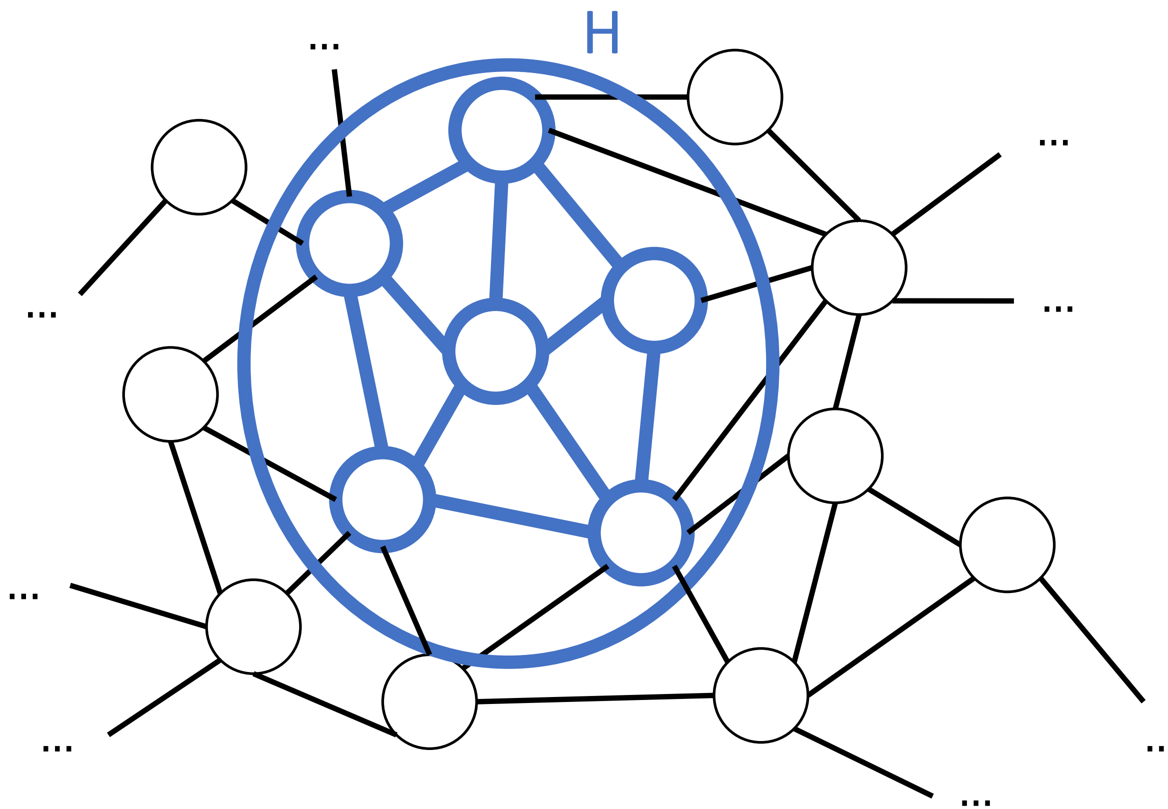
What if causal graph is HUGE?



Local causal discovery:

Only care about a small subgraph of the larger graph

What if causal graph is HUGE?



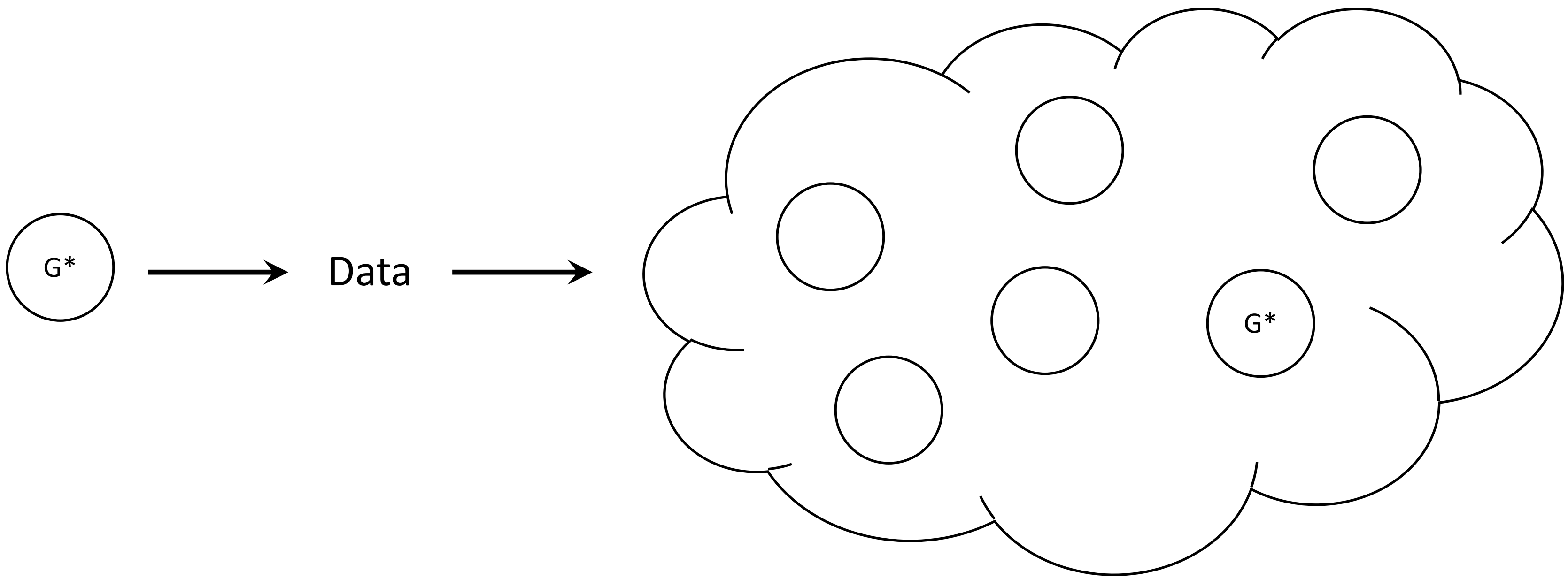
Local causal discovery:

Only care about a small subgraph of the larger graph

(Informal) Verification: Generalization of "DP on covered edge forest" [CS23]

(Informal) Search: $\mathcal{O}\left(\log |H| \cdot \nu(G^*)\right)$ interventions suffices [CS23]

In many problem domains…



There are domain experts!

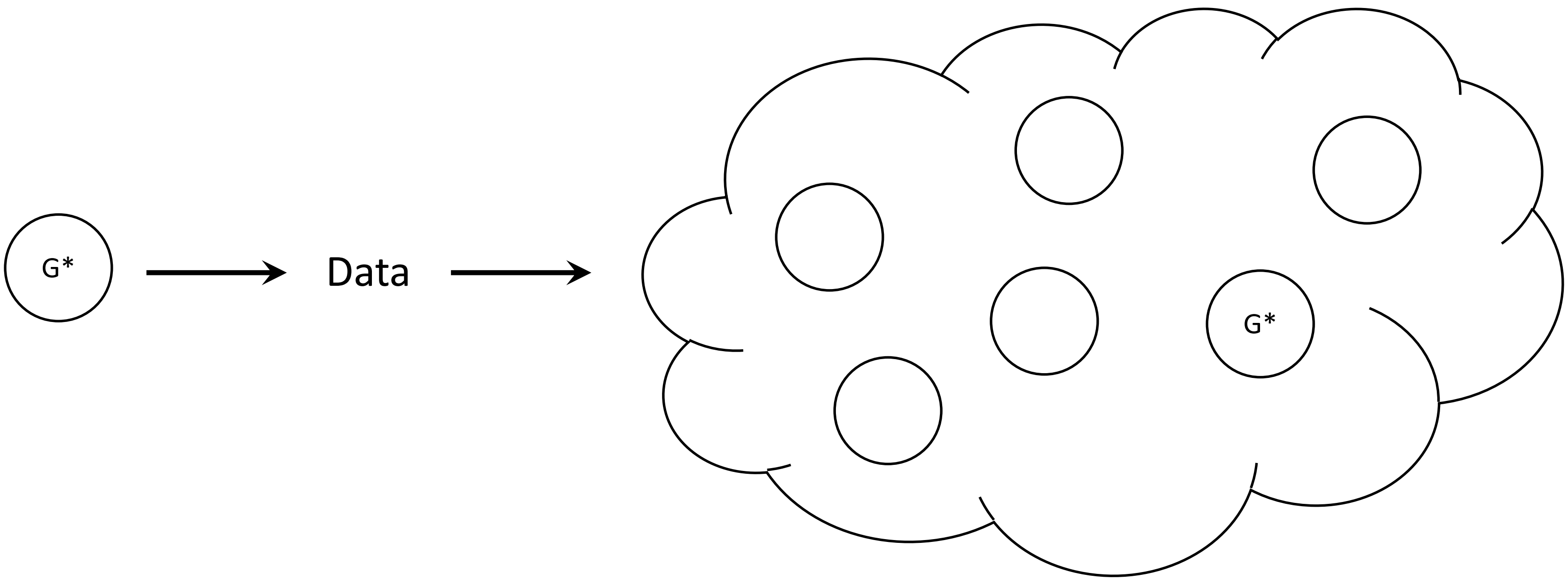
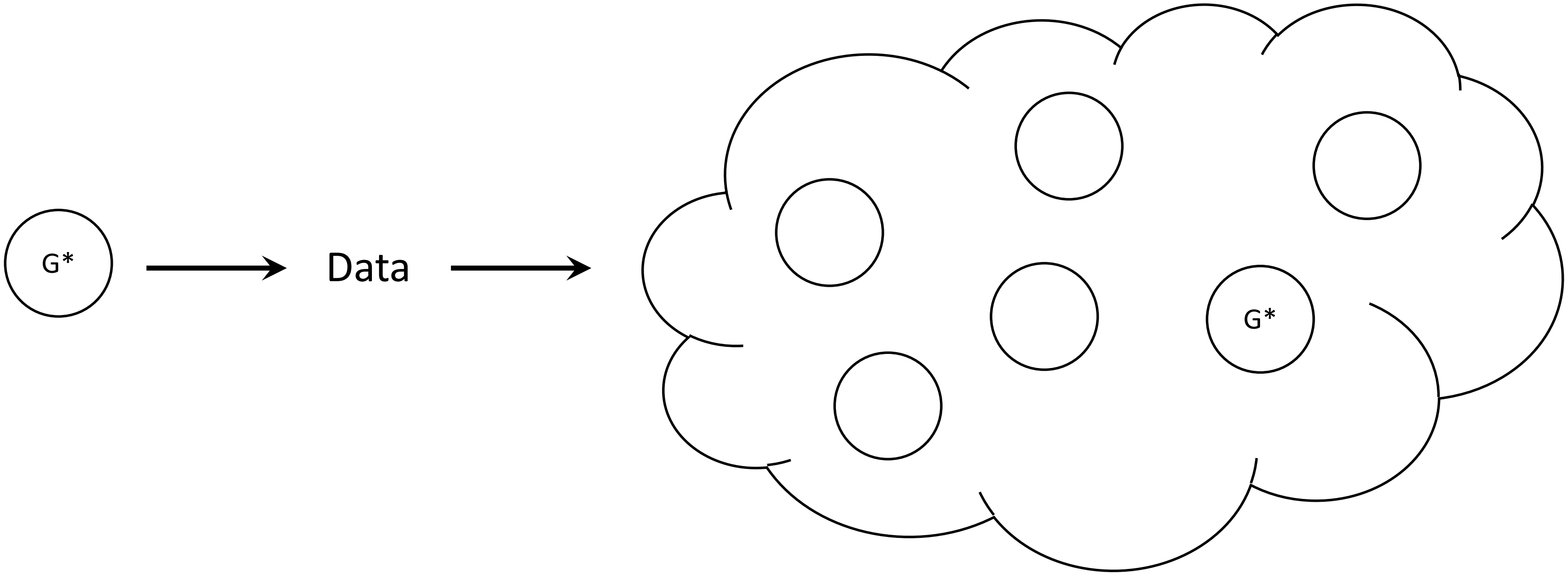


Image credit: https://thenounproject.com/icon/doctor-1285618/

There are domain experts!



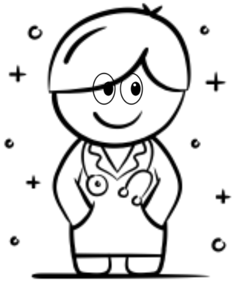
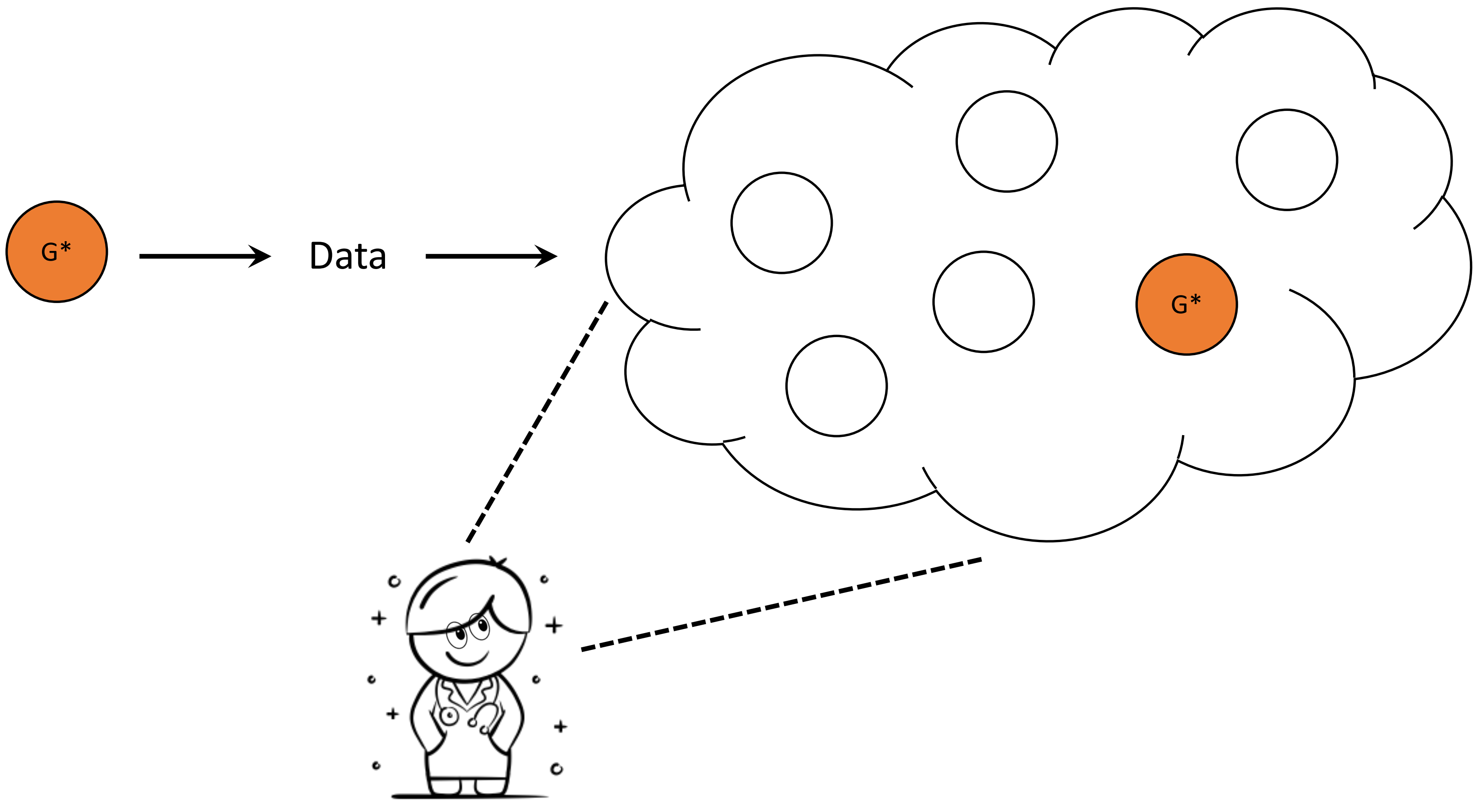
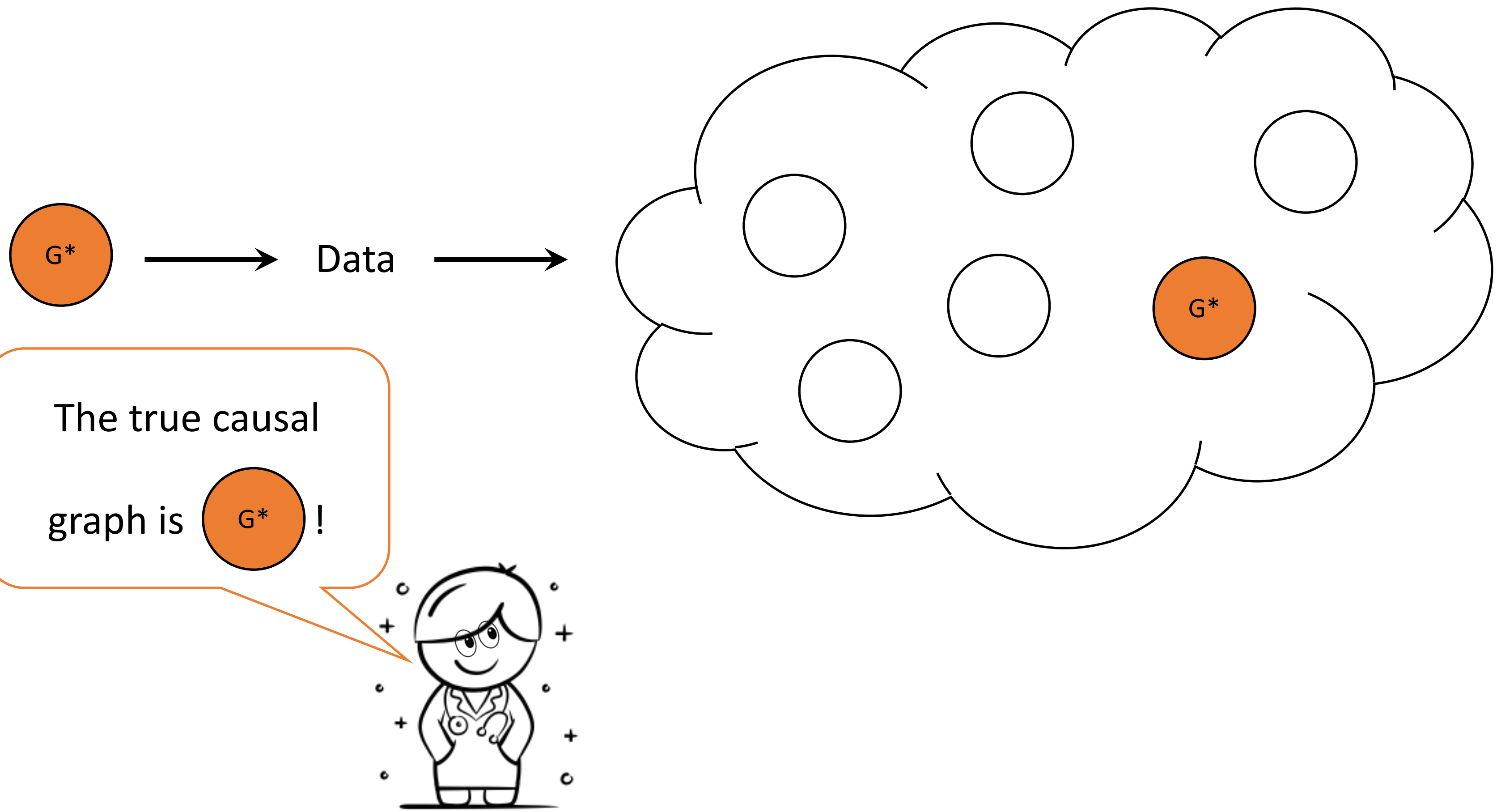


Image credit: https://thenounproject.com/icon/look-4968809/

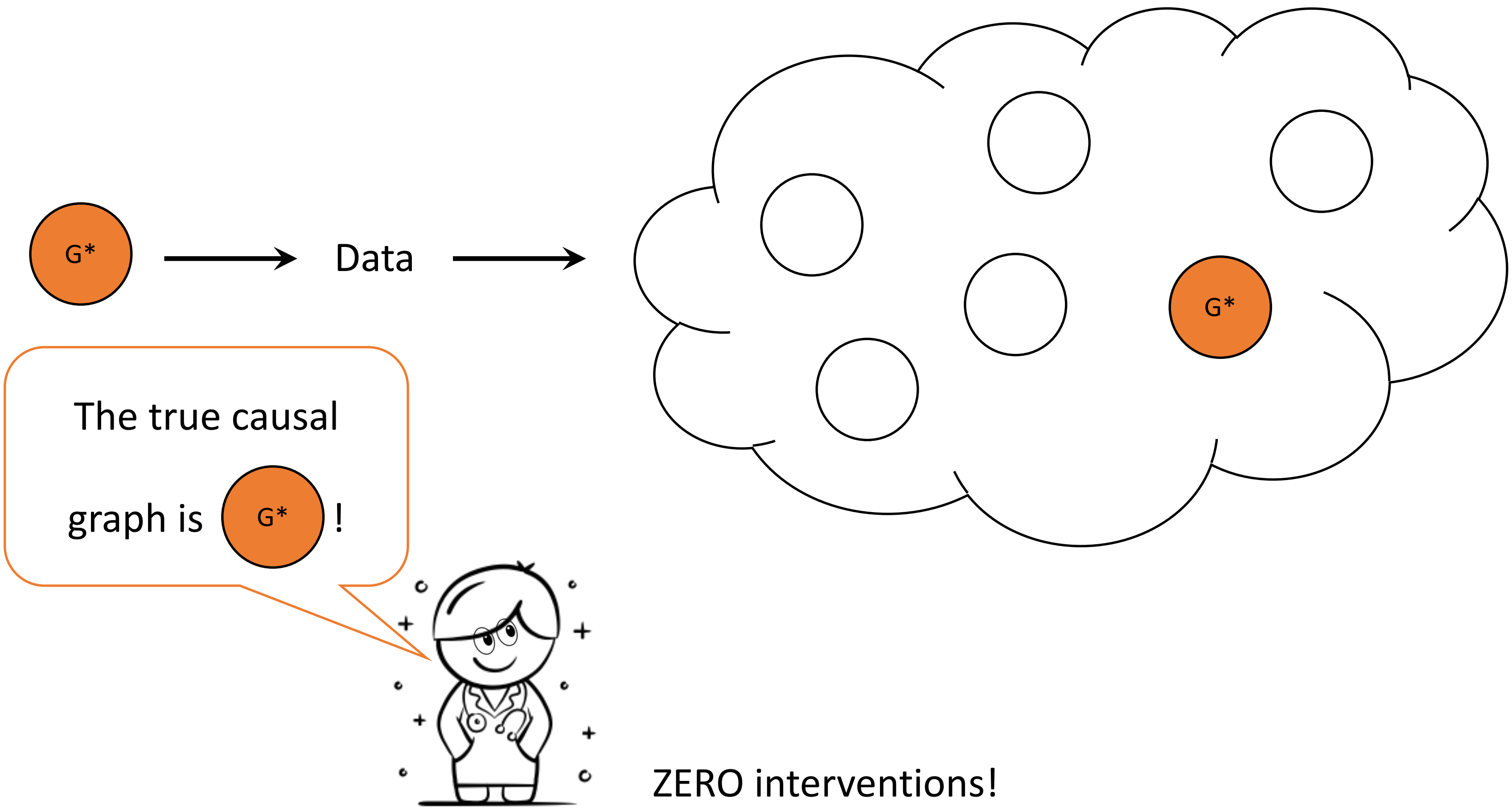
There are domain experts!



There are domain experts!

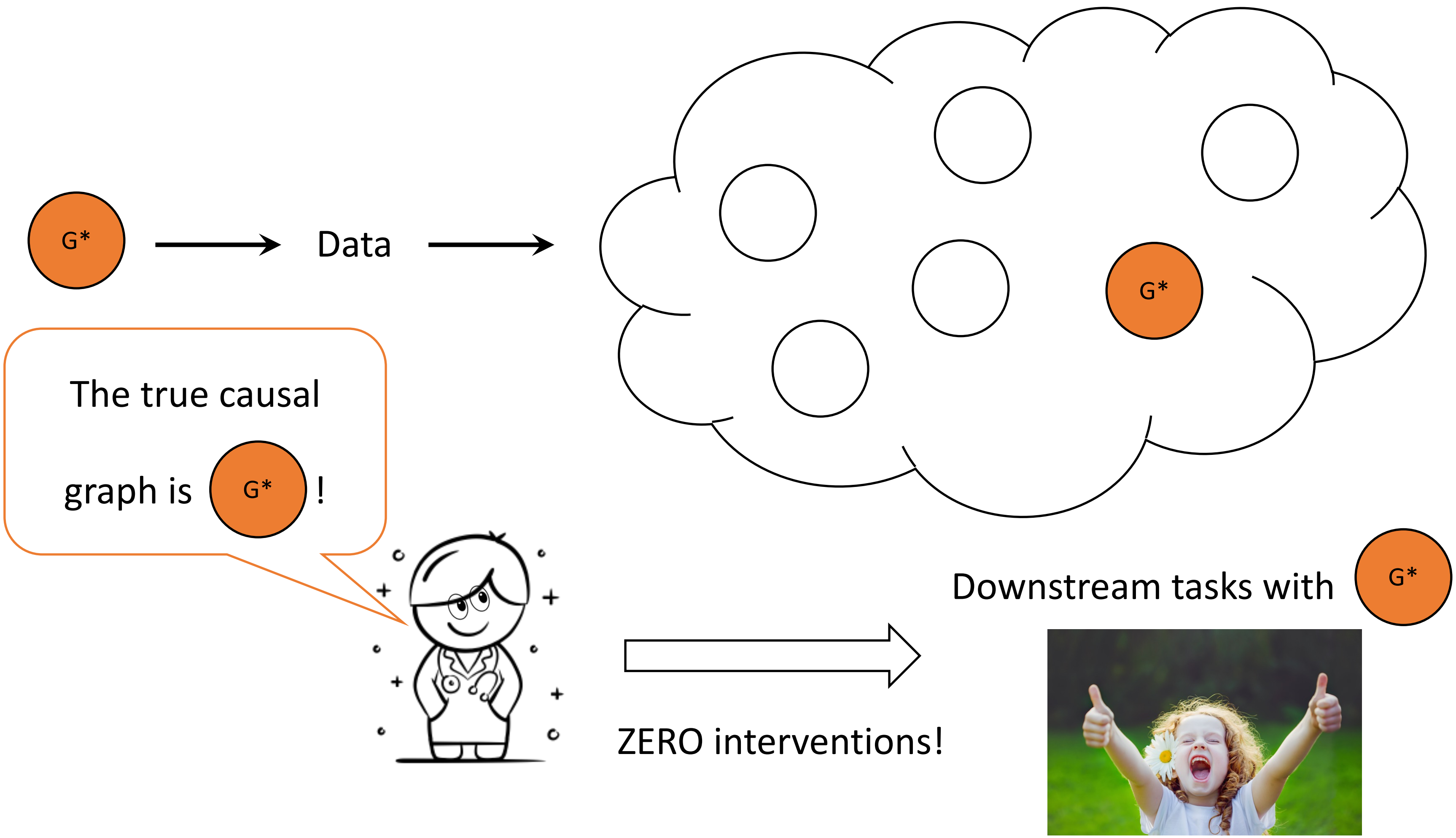


There are domain experts!

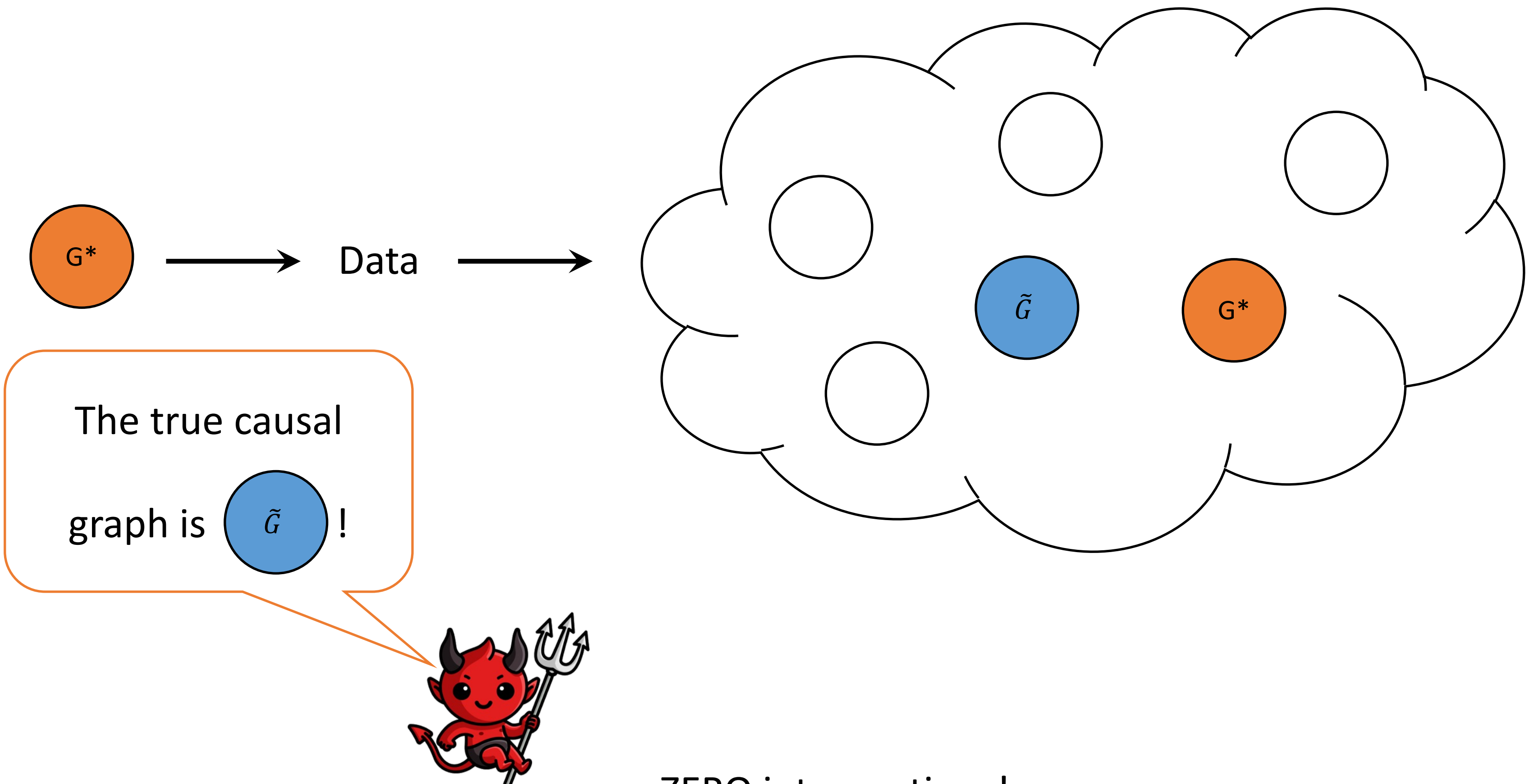


ZERO interventions!

There are domain experts!



But… experts can be wrong



ZERO interventions!

Image credit: https://dribbble.com/shots/14489872-Devil

But… experts can be wrong

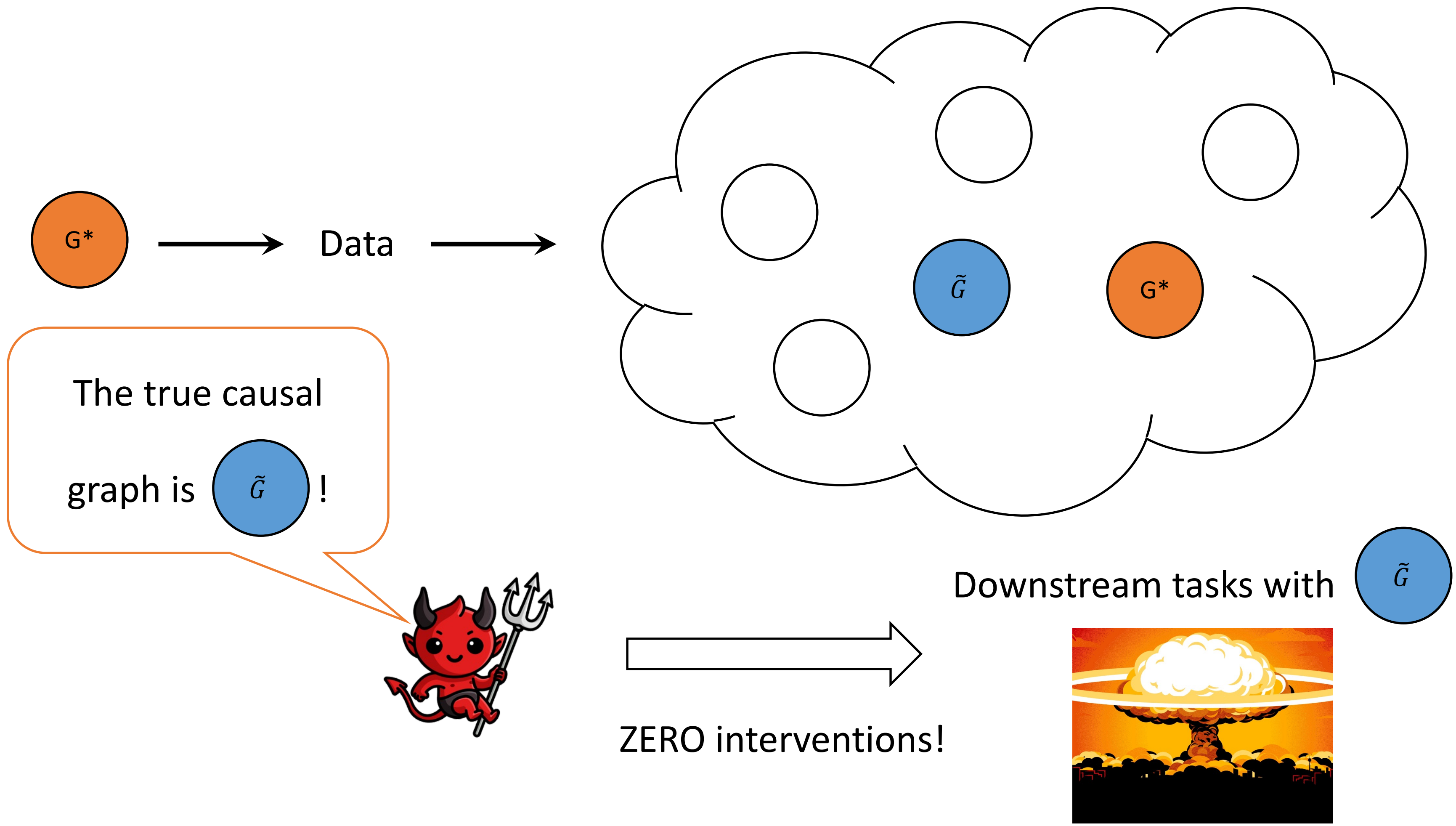
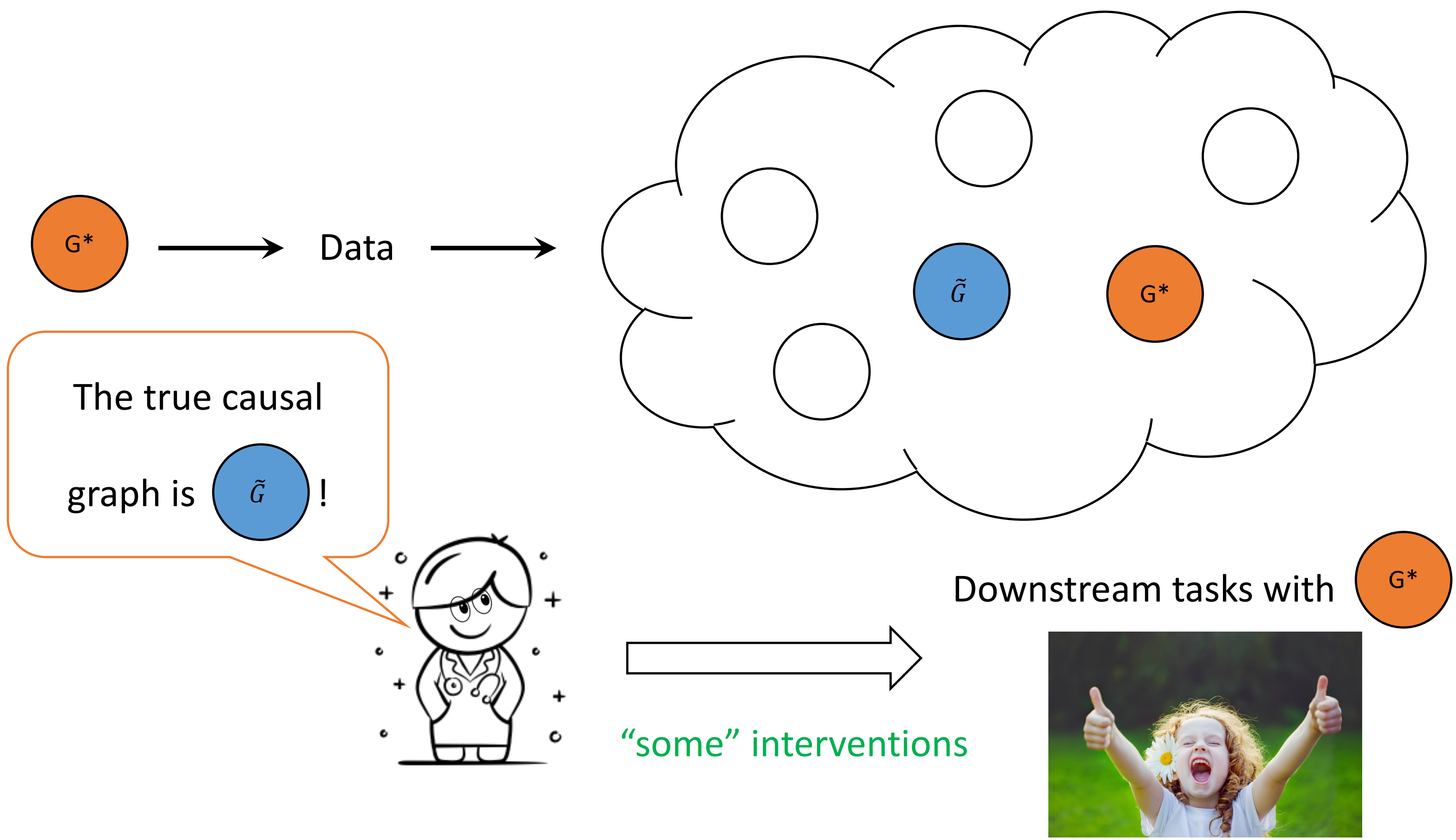
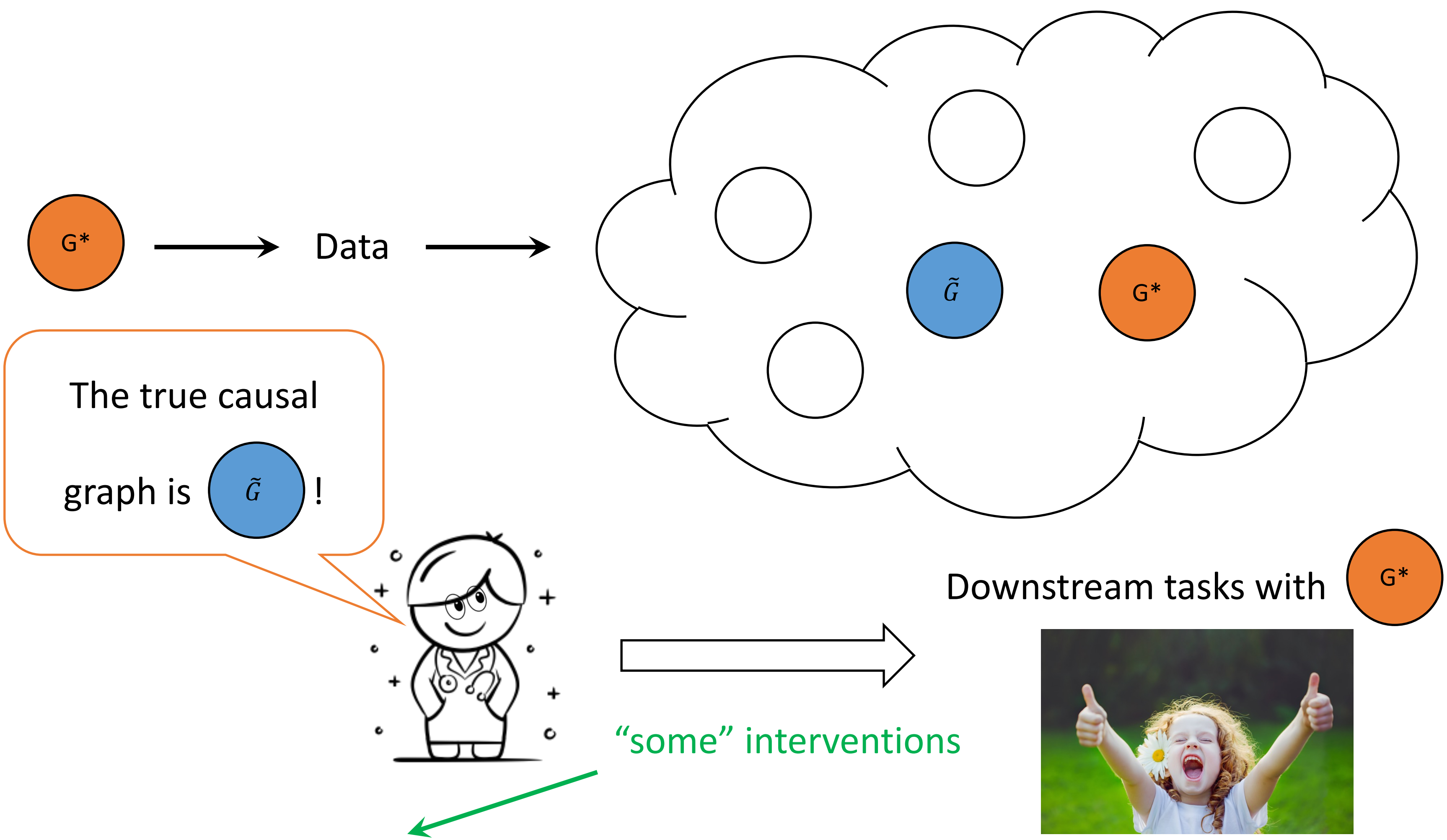


Image credit: https://dribbble.com/shots/3759014-Atomic-Illustrations/attachments/3759014-Atomic-Illustrations?mode=media

But… experts can be wrong



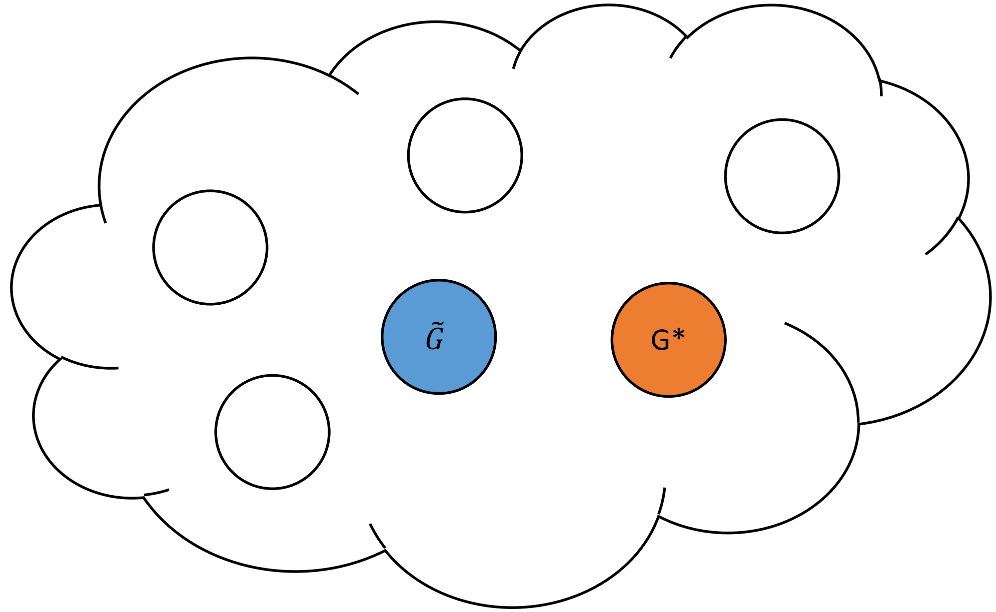
But... experts can be wrong



Advice search: $\mathcal{O}\left(\log \psi(G^*, \tilde{G}) \cdot \nu(G^*)\right)$ interventions [CGB23]

But… experts can be wrong

G* → Data →



The true causal graph is $\tilde{G}$!

Downstream tasks with G*

"some" interventions

Quality of advice $\tilde{G}$

$0 \leq \psi(G^*, \tilde{G}) \leq n$

Advice search: $\mathcal{O}\left(\log \psi(G^*, \tilde{G}) \cdot \nu(G^*)\right)$ interventions [CGB23]