Insights into Causal Link Prediction through Causal Lifting

Bruno Ribeiro
Purdue University

GraPFiCs Workshop
October, 19, 2023

References in this talk

PROCEEDINGS A

royalsocietypublishing.org/journal/rspa

Research

Check for updates

Cite this article: Cotta L, Bevilacqua B, Ahmed N, Ribeiro B. 2023 Causal lifting and link prediction. *Proc. R. Soc. A* **479**: 20230121.
https://doi.org/10.1098/rspa.2023.0121

Causal lifting and link prediction

Leonardo Cotta[1], Beatrice Bevilacqua[2], Nesreen Ahmed[3] and Bruno Ribeiro[2]

[1]Vector Institute, Toronto, Canada
[2]Purdue University, West Lafayette, IN, USA
[3]Intel Labs, Hillsboro, OR, USA

LC, 0000-0002-4751-7643

Bala Srinivasan, PhD
(Amazon)

Published as a conference paper at ICLR 2020

ON THE EQUIVALENCE BETWEEN POSITIONAL NODE EMBEDDINGS AND STRUCTURAL GRAPH REPRESENTATIONS

Balasubramaniam Srinivasan
Department of Computer Science
Purdue University
bsriniv@purdue.edu

Bruno Ribeiro
Department of Computer Science
Purdue University
ribeiro@cs.purdue.edu

Leonardo Cotta, PhD
(Vector Institute)

Beatrice Bevilacqua
(PhD student)

Relational Pooling for Graph Representations

Ryan L. Murphy [1] **Balasubramaniam Srinivasan** [2] **Vinayak Rao** [1] **Bruno Ribeiro** [2]

Ryan Murphy, PhD
(Financial Industry Regulatory Authority)

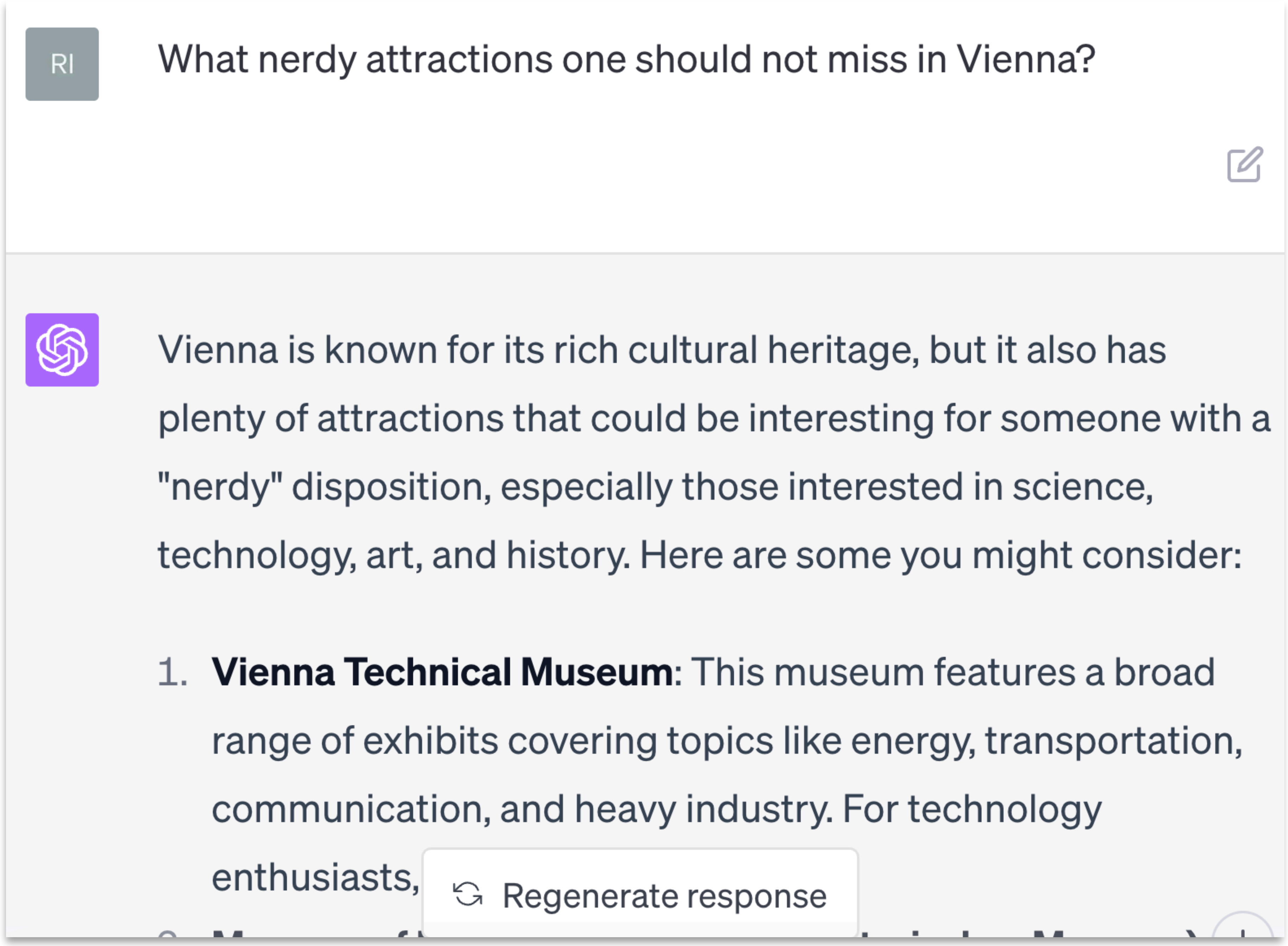
Vinayak Rao
Purdue

I ❤️ Symmetries

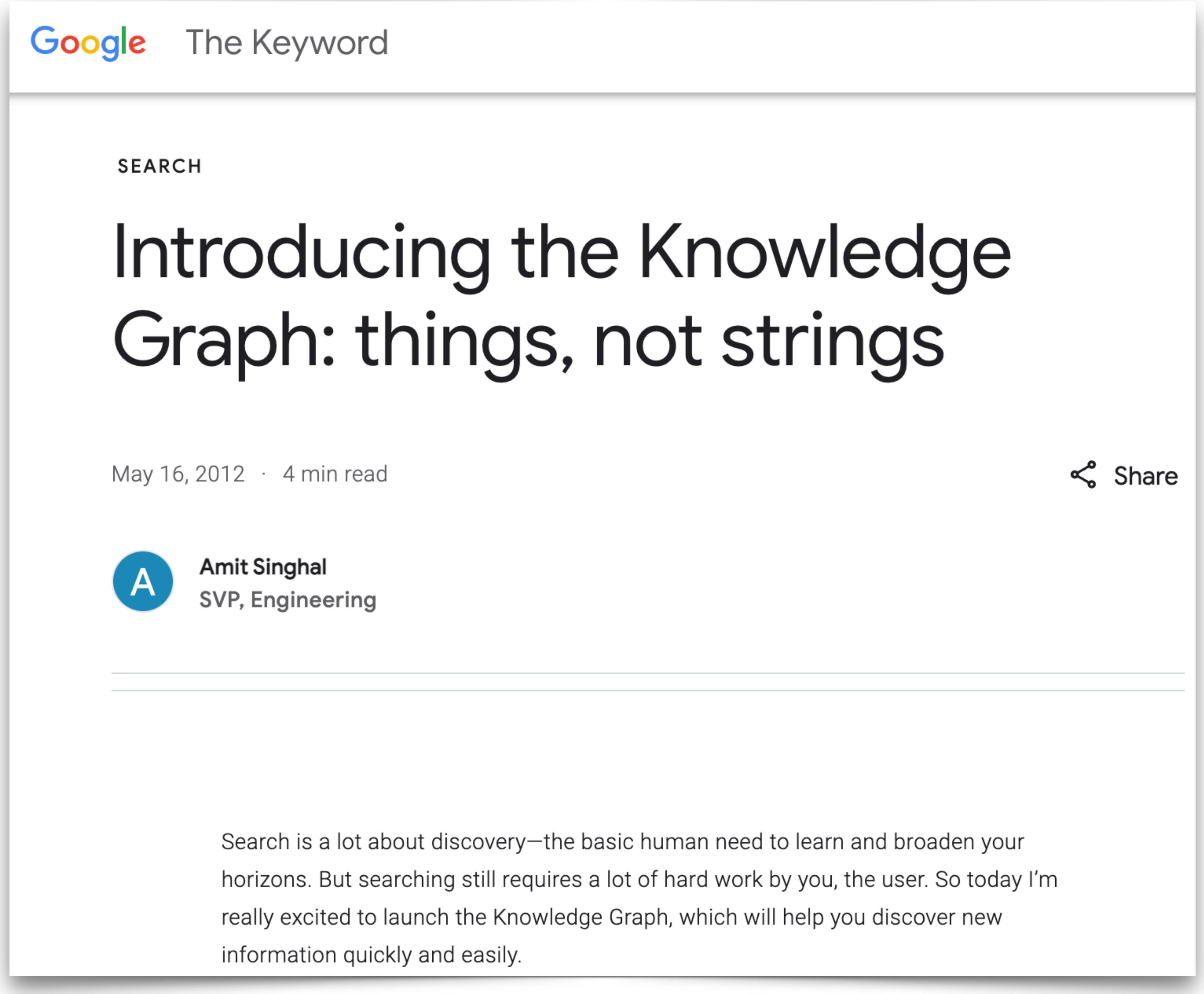
OpenAI vs Google

Knowledge search: **Strings** (text) or **things** (graphs)?

In **2022** OpenAI demoed ChatGPT, **“strings-only”** method.



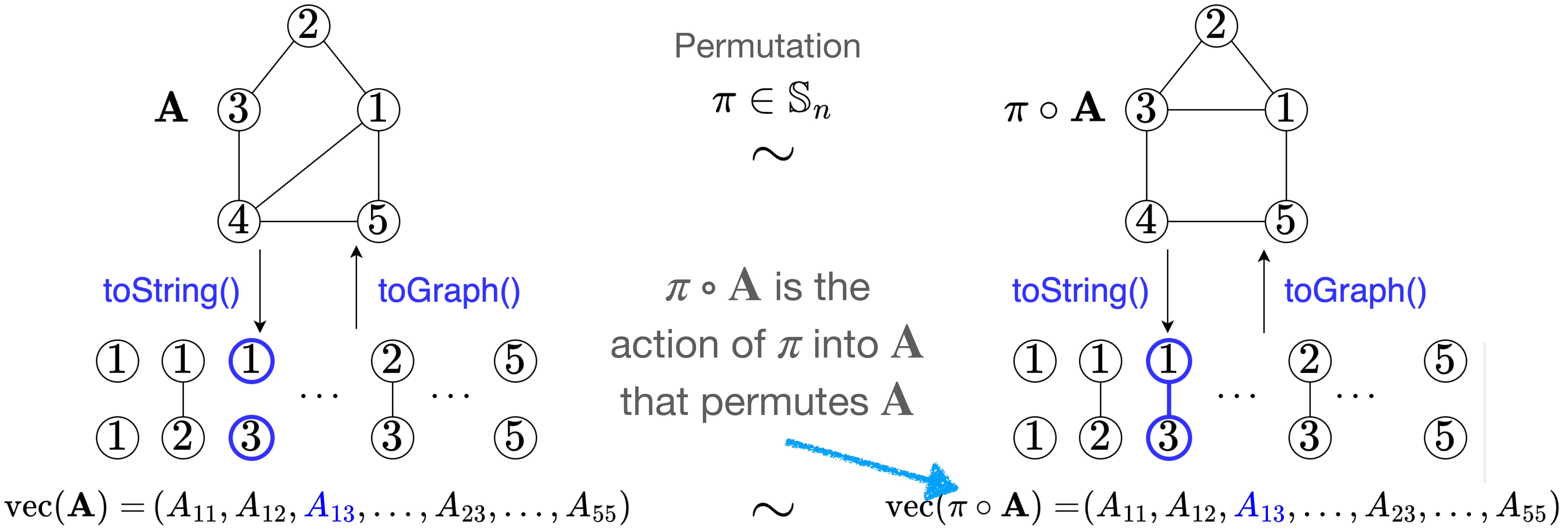
In **2012** Google declared *web search* as **“things, not strings”**.



Graph Learning: Graphs are "strings" + symmetries

- Is a graph a sequence (string) of edges?
- In **graph learning, we assume graphs *are* sequences of edges** with associated (permutation) **symmetries** since node ids are arbitrary [Murphy et al., 2019, Xu et al., 2019, Morris et al., 2019].
- In statistics this assumption is called *exchangeability*

Graph "string" isomorphism: Graphs with distinct "strings" can be the **same graph.**



Why are symmetries relevant?

ChatGPT fails at multi-hop reasoning [Dziri et al., 2023].

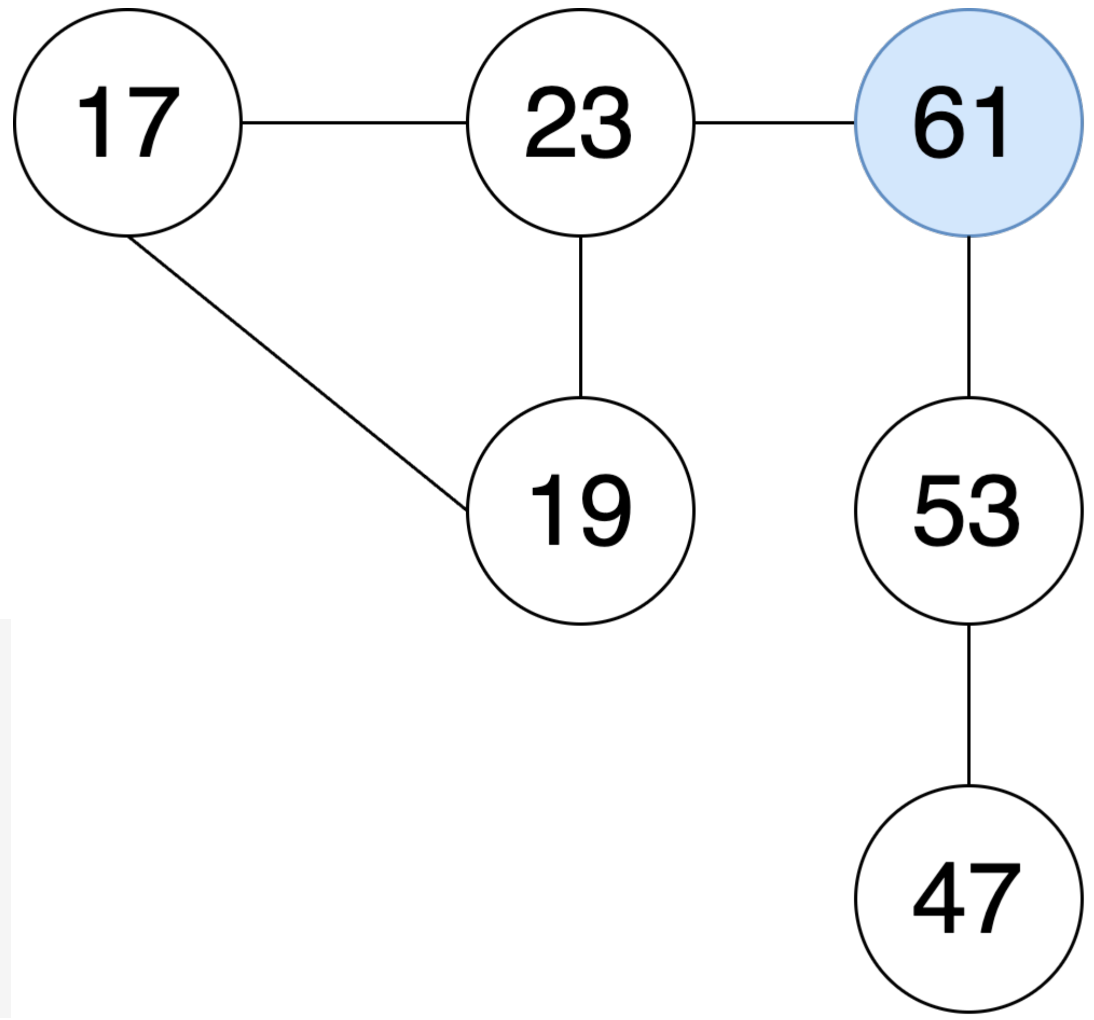
- Some tasks require symmetries but order-sensitive models give different answers based on the input order.

Q: What is the number of nodes that can reach node 61 in exactly two hops?

> In the given undirected graph, there are two nodes that are exactly 2-hops away from node 61. These nodes are 17 and 19.

But if we reorder the edges in the prompt the answer changes.

> In the given graph, there are three nodes that are exactly 2-hops away from node 61. These nodes are 17, 19, and 47.



- **ChatGPT's answers are sensitive to prompt order (order-sensitive model).**
- Differently, **models respecting symmetries must treat all paths identically.**

Defining Symmetries through Group Theory

A group G is a set together with a binary operation $\star$ such that:

- Closure holds i.e., $\forall a, b \in \mathcal{G},\ \ a \star b \in G$
- Associativity holds $(a \star b) \star c = a \star (b \star c)\ \ \forall a, b, c \in G$
- Identity element exists i.e., $\exists e \in \mathcal{G}$ s.t. $a \star e = e \star a = a\ \ \forall a \in G$
- **Inverse exists for every element** and $a \star a^{-1} = a^{-1} \star a = e\ \ \forall a \in G$

Credit: Bala Srinivasan

(Left) Group Actions

For a group G , binary operation $\star$, and with identity e, and a set X, a (left) group action is a function $\circ : G \times X \to X$, such that

- $e \circ x = x, \;\; \forall x \in X$
- $g \circ (h \circ x) = (g \star h) \circ x, \;\; \forall g, h \in \mathcal{G}, \;\; \forall x \in X$

Credit: Bala Srinivasan

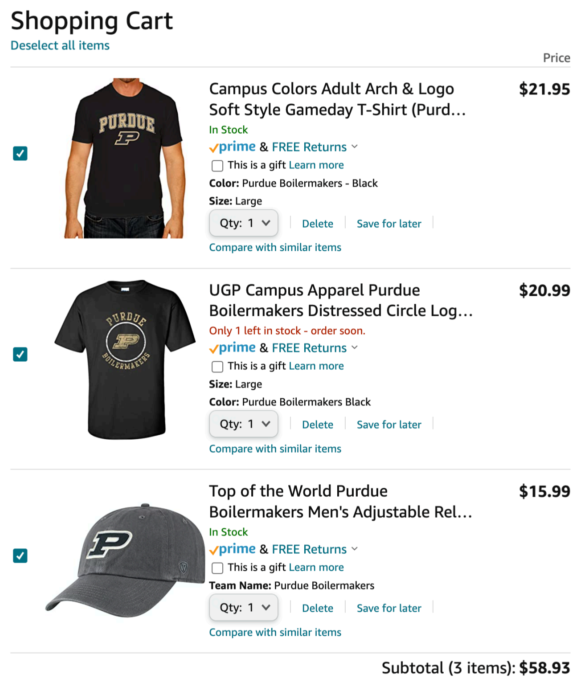
Examples of Transformation Groups

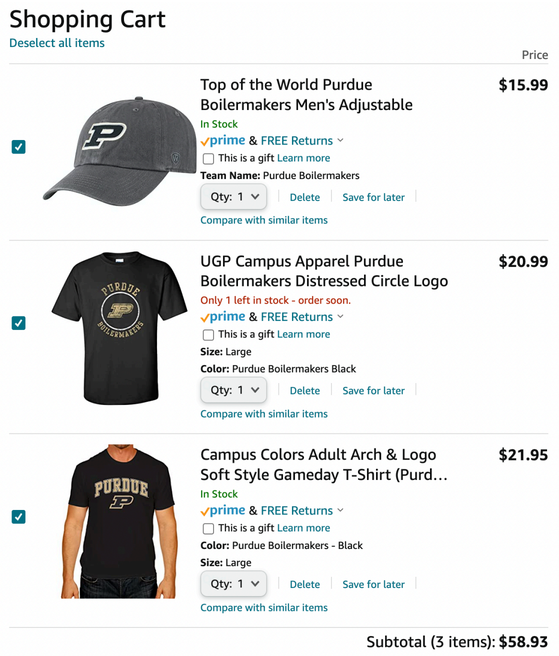
- **Permutation Group** ($\mathbb{S}_n$) - All $n!$ permutations of n objects

  E.g.: $\mathbb{S}_3 = \{(1,2,3), (2,3,1), (3,1,2), (1,3,2), (3,2,1), (2,1,3)\}$

  Action of π_1 on a sequence

  Example: $\pi_1 = (3,2,1)$

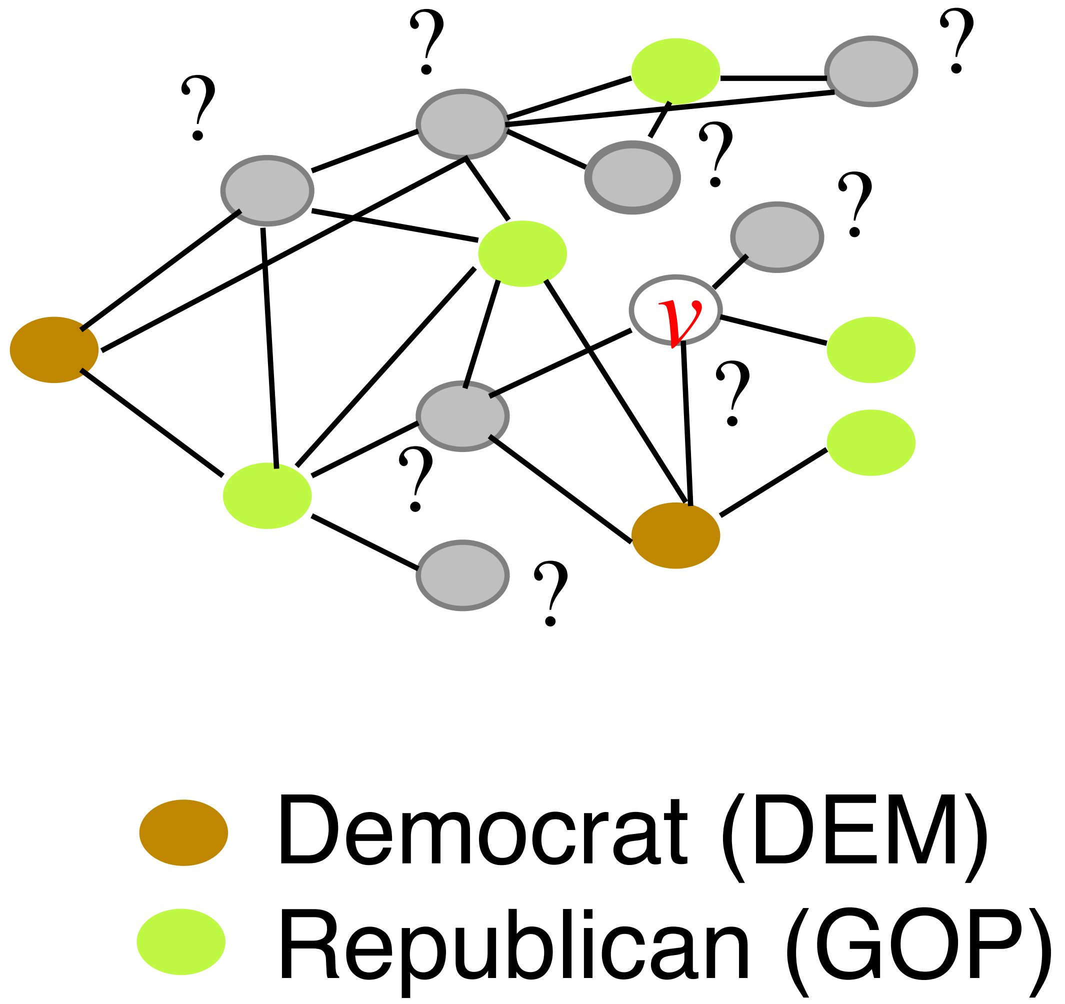
  $\pi_1 \circ \Big($ 

$\Big) =$ 

- **Special Orthogonal Group** ($\mathbf{SO}(3)$) - Orthogonal 3×3 matrices M, such that $\det(M) = 1$, and $M_1 \star M_2 = M_1 M_2$. **<u>(Rotation Group in 3D)</u>**

Example: Equivariant Embedding Function

- **Permutation Group** ($\mathbb{S}_n$): A Graph Neural Network, GNN($\mathbf{A}$), is a neural network that learns graph $\boldsymbol{A}$ embeddings, which are ***equivariant*** to $\pi \circ \mathbf{A}$, $\pi \in \mathbb{S}_n$

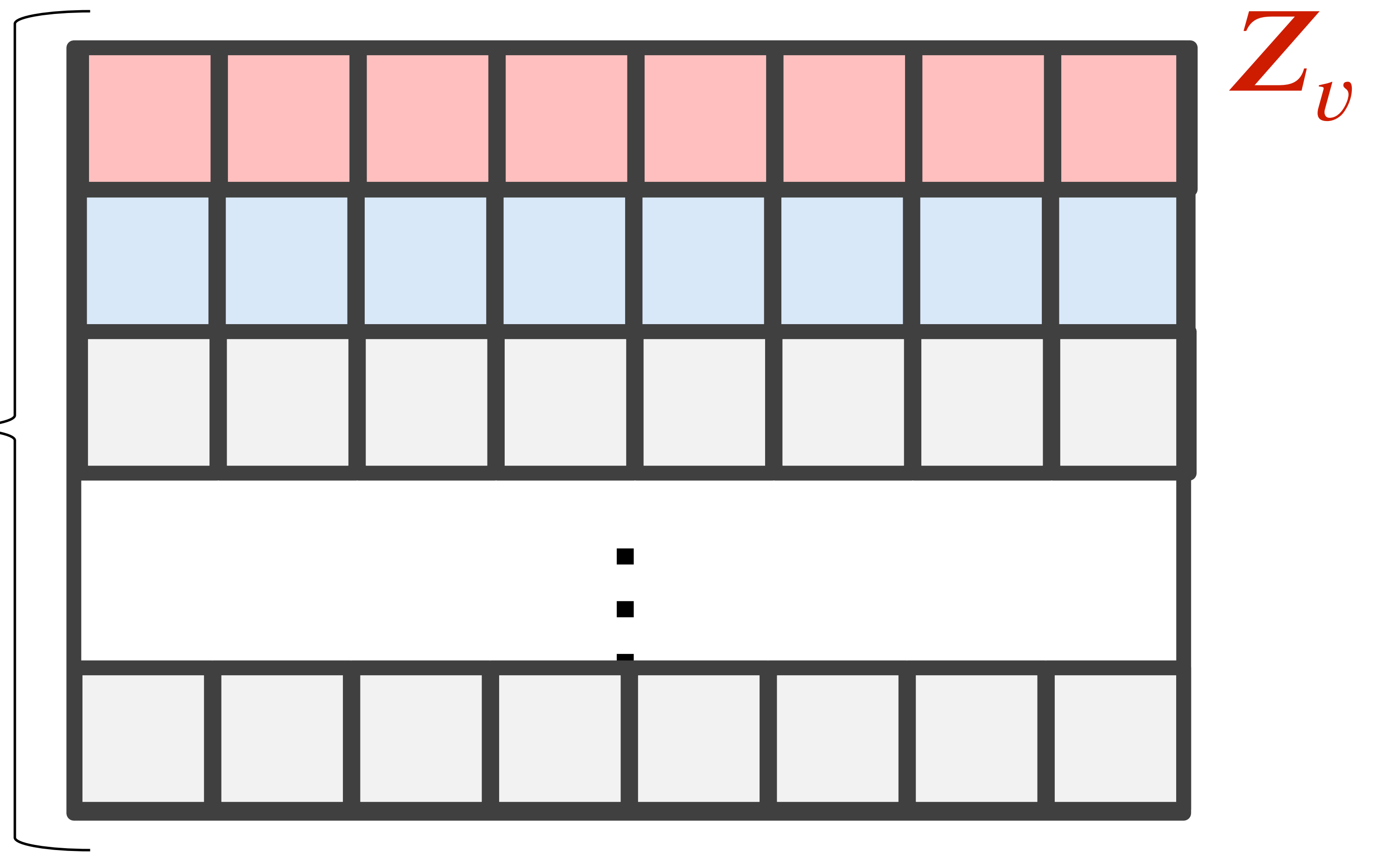


v

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

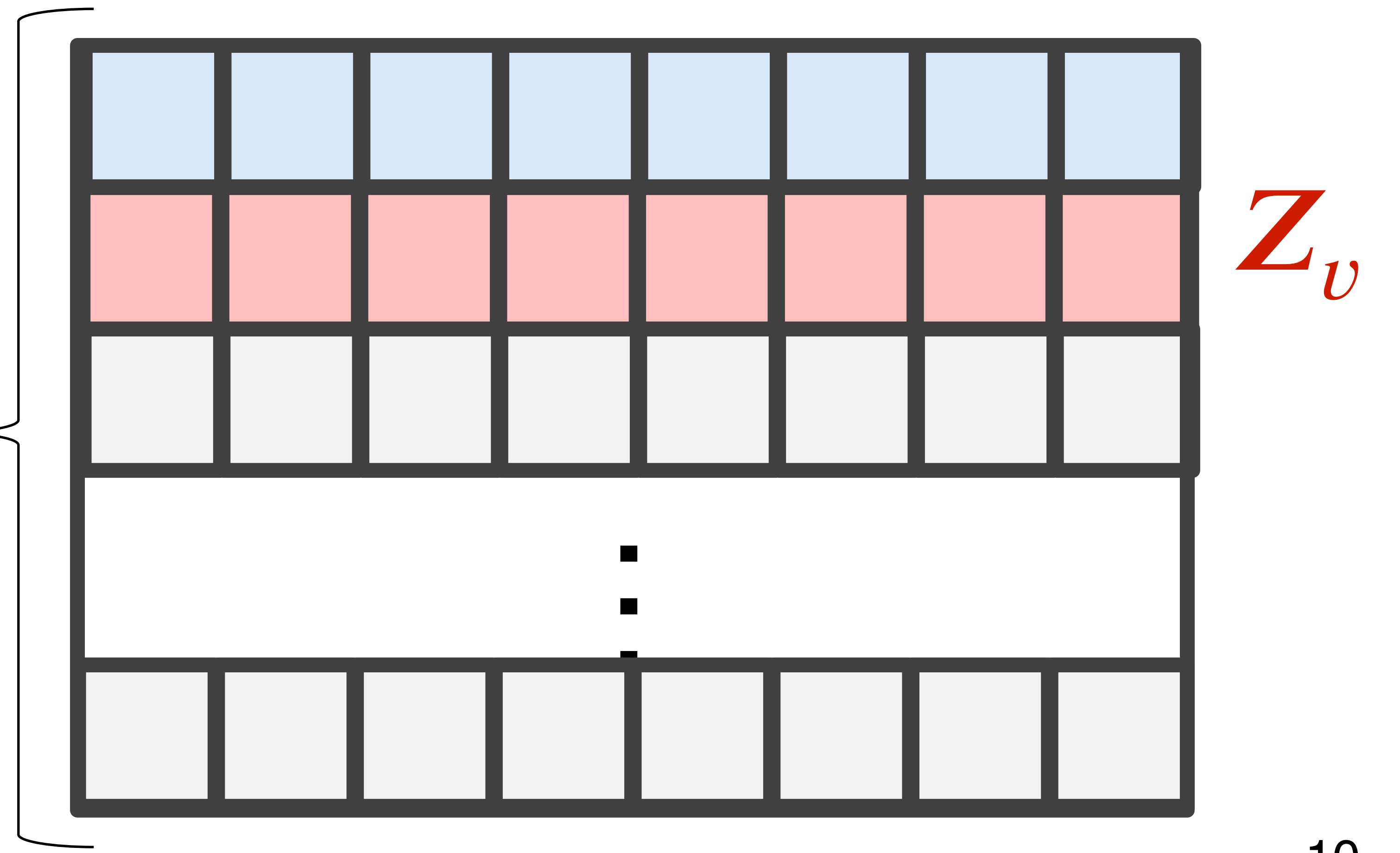
$f(\mathbf{A}) = \mathbf{Z}$

$\boldsymbol{Z} \in \mathbb{R}^{n \times d}$



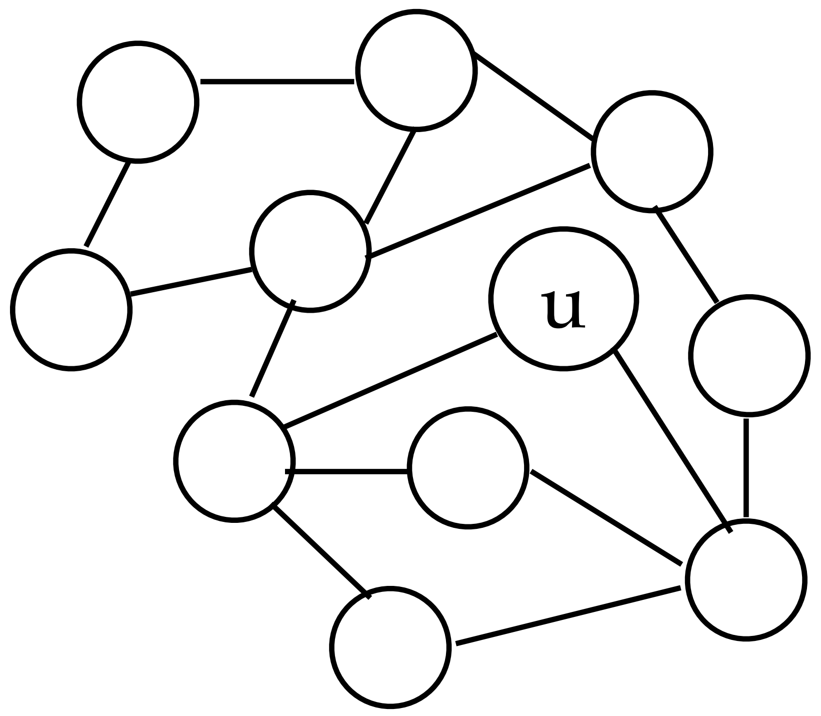
$$\pi \circ \mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$f(\pi \circ \mathbf{A}) = \pi \circ \mathbf{Z}$



Downstream Task — Node Classification

f — (Abstractly) A function that outputs node representations $f : \mathbb{R}^{n \times n \times (1+k+p)} \to \mathbb{R}^{n \times d}, \ \ d > 0$



G-invariant embedding $f(\mathbf{A}, u) = f(\pi \circ \mathbf{A}, \pi \circ u) \in \mathbb{R}^d$

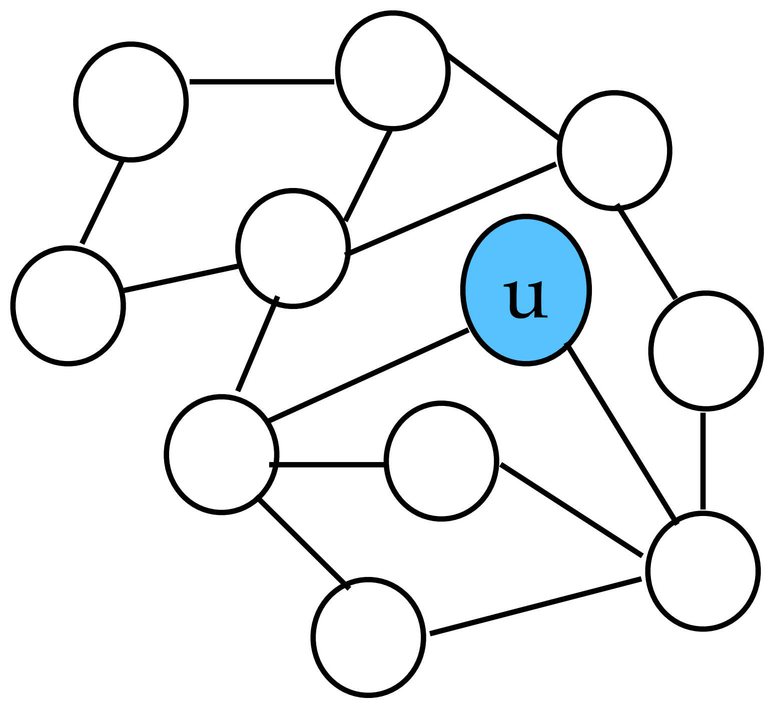
$f(\mathbf{A}, u) \in \mathbb{R}^d$

node u's embedding

Classifier

$g_\theta : \mathbb{R}^d \to \{1, \ldots, n_{\text{classes}}\}$

$\mathbf{A} \in \mathbb{R}^{n \times n \times (1+k+p)}$
simplified tensor notation of the graph



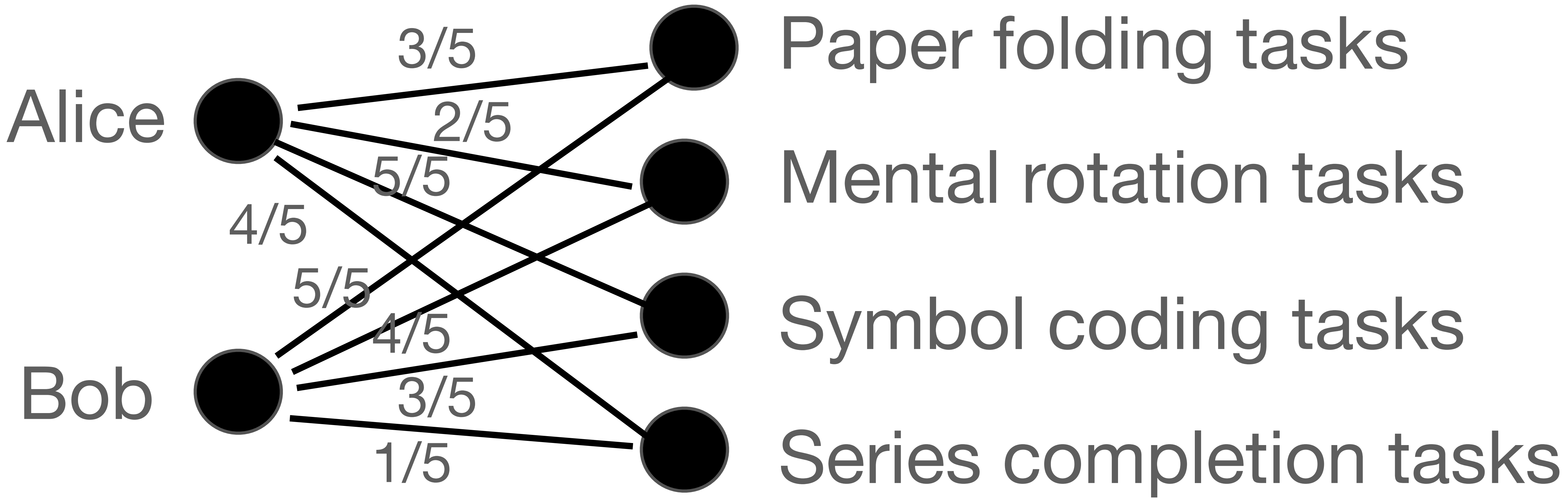
Node Classification
(Downstream Task)

Example:
Given a social network **A**, predict the types of ads to serve user **u**

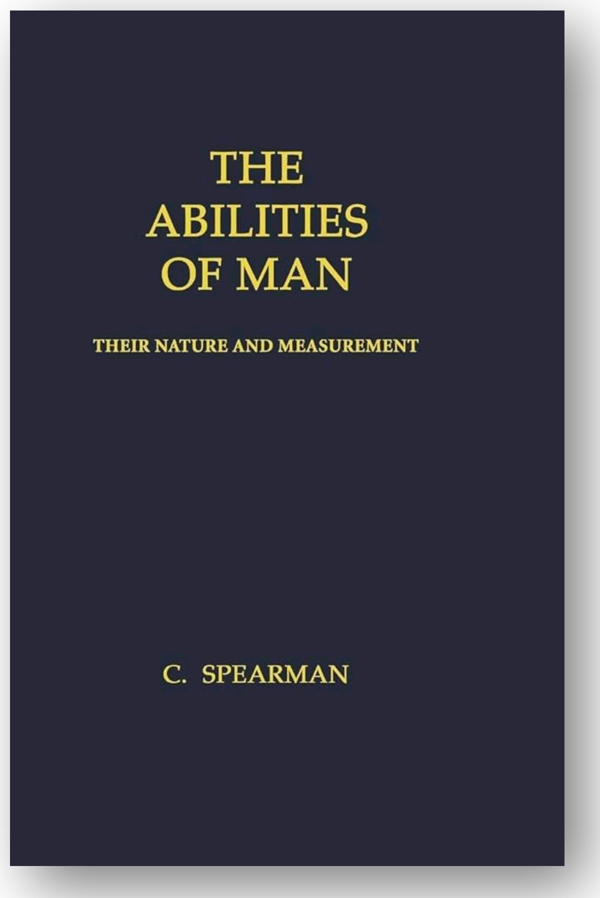
Causal Lifting: Origin Story

Causal interpretations of matrix factorization for path-dependent graphs…

- Matrix factorization derives from Spearman's *common factors of intelligence*
  - Conjectures that **latent factors of intelligence** manifest as abilities to perform tasks



- In 1914 Woolley and Fischer's observed that
  "*boys are [innately] enormously superior [to girls] at spatial relations*"



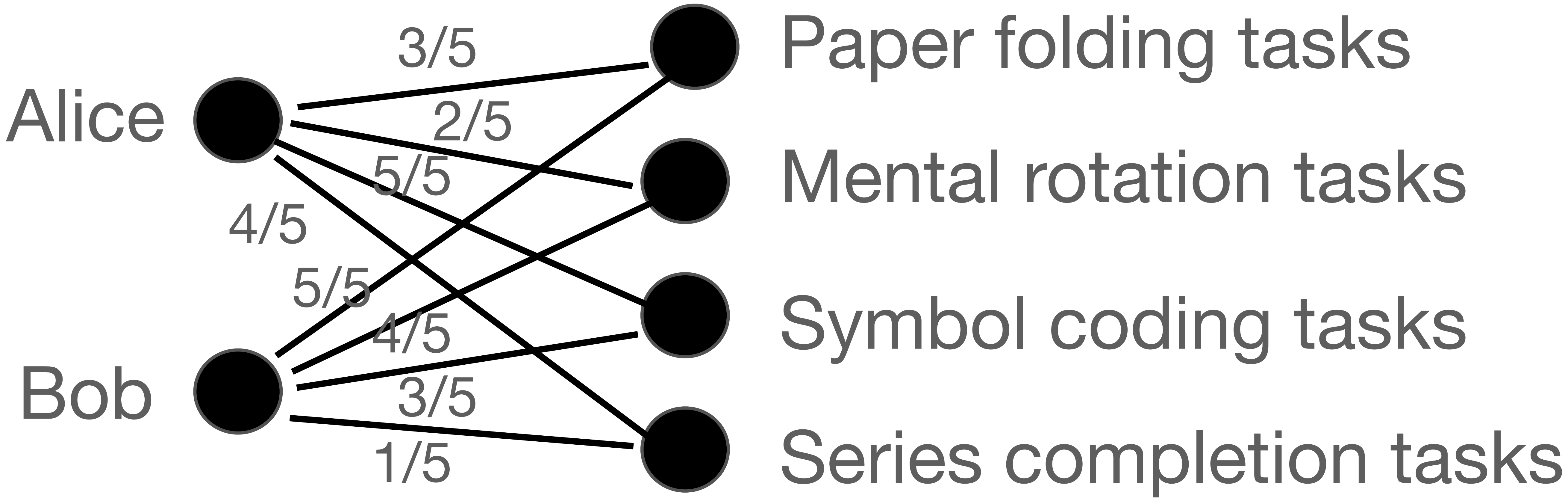
- But Spearman (1927) disagreed with the conclusion:
  "*evidence [of this difference being] innate [rather than acquired] is still dubious*".

Path-dependent link formation…

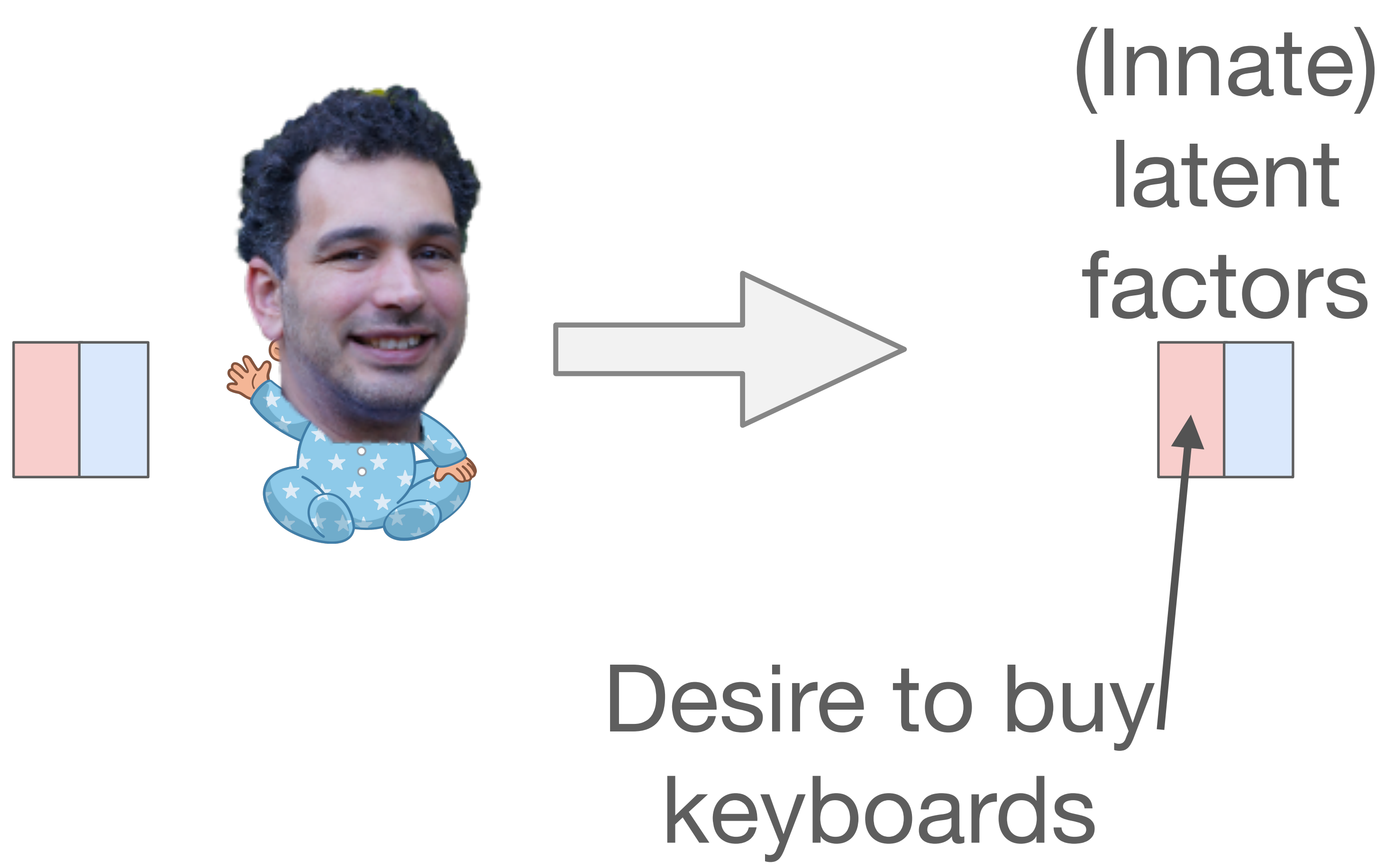
Wikipedia

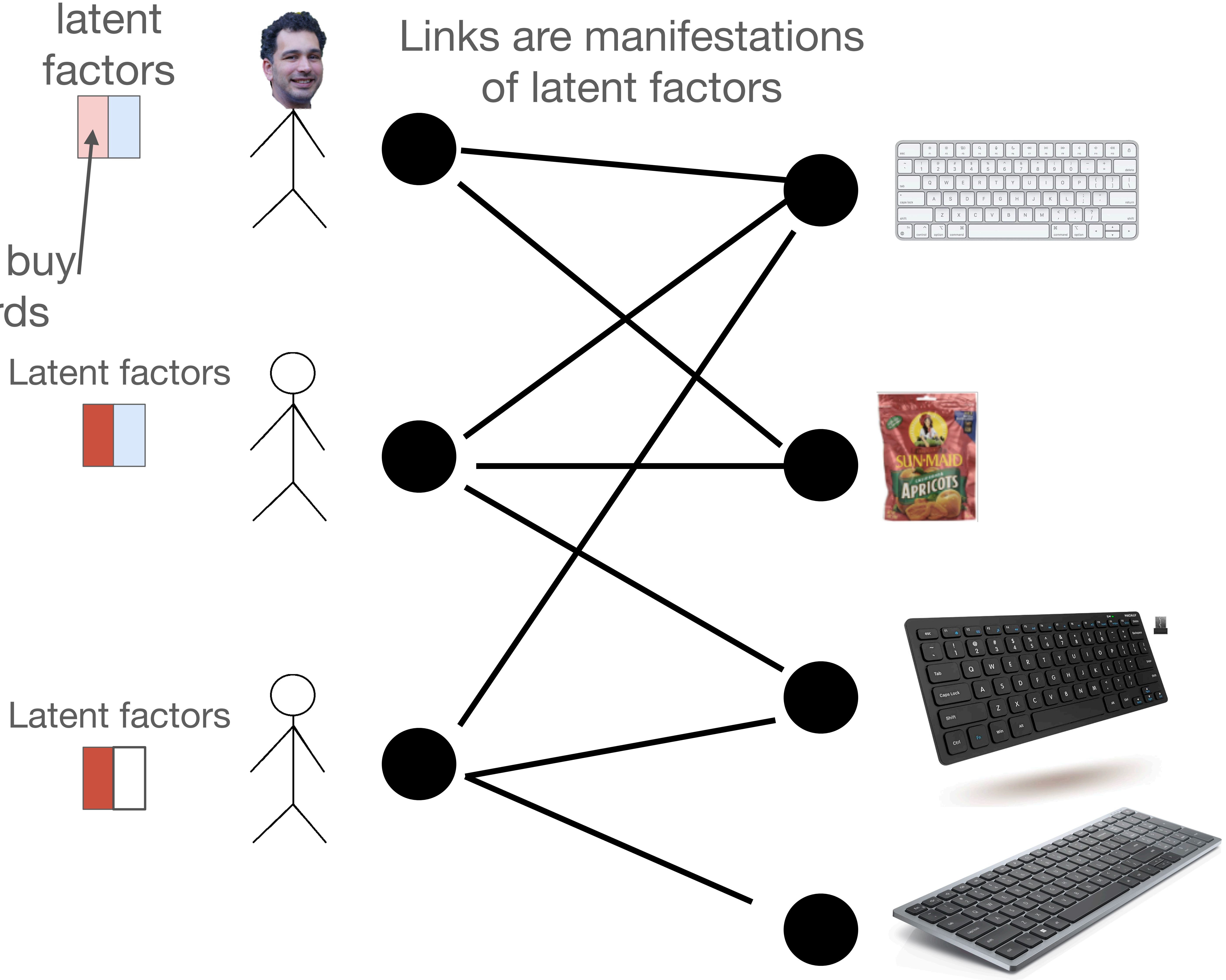
Had **Alice** practiced spatial tasks, she would had been better at them

i.e., path-dependent link weights:

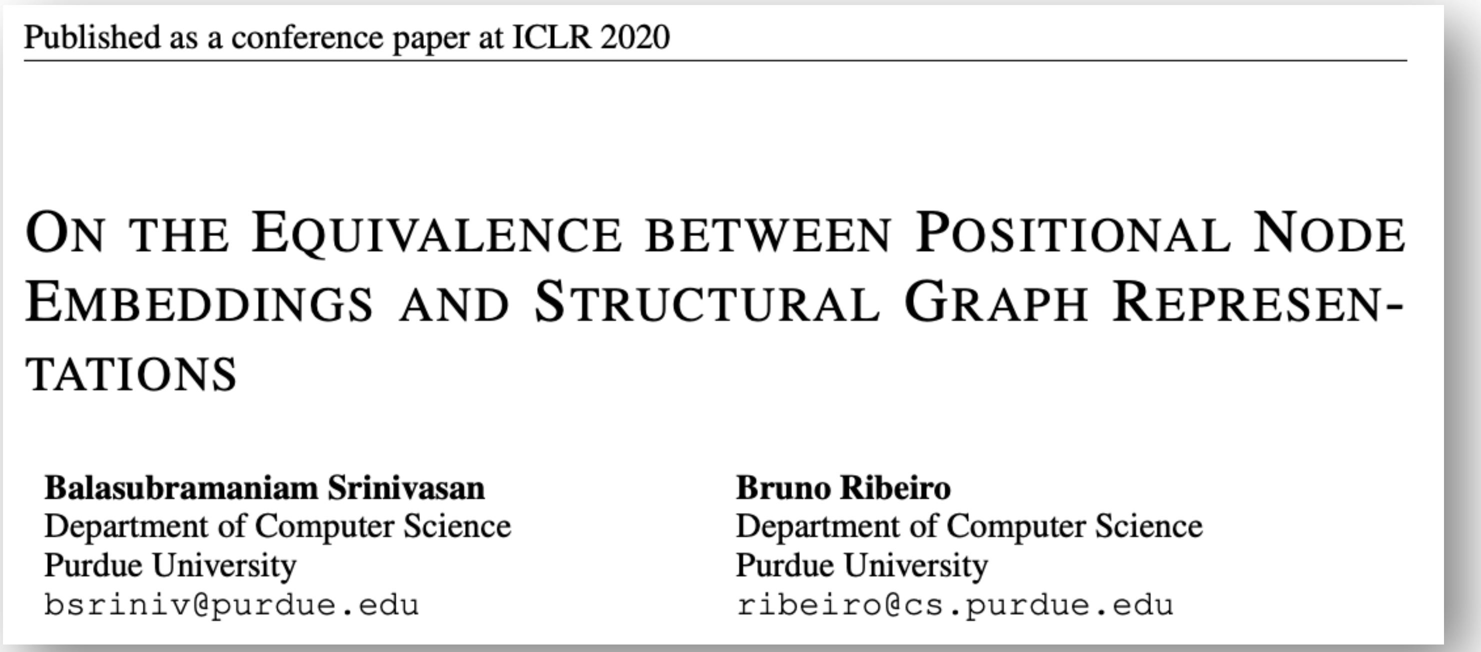


Matrix factorization: A graph formation story





Defines graph learning through symmetries



Published as a conference paper at ICLR 2020

ON THE EQUIVALENCE BETWEEN POSITIONAL NODE EMBEDDINGS AND STRUCTURAL GRAPH REPRESENTATIONS

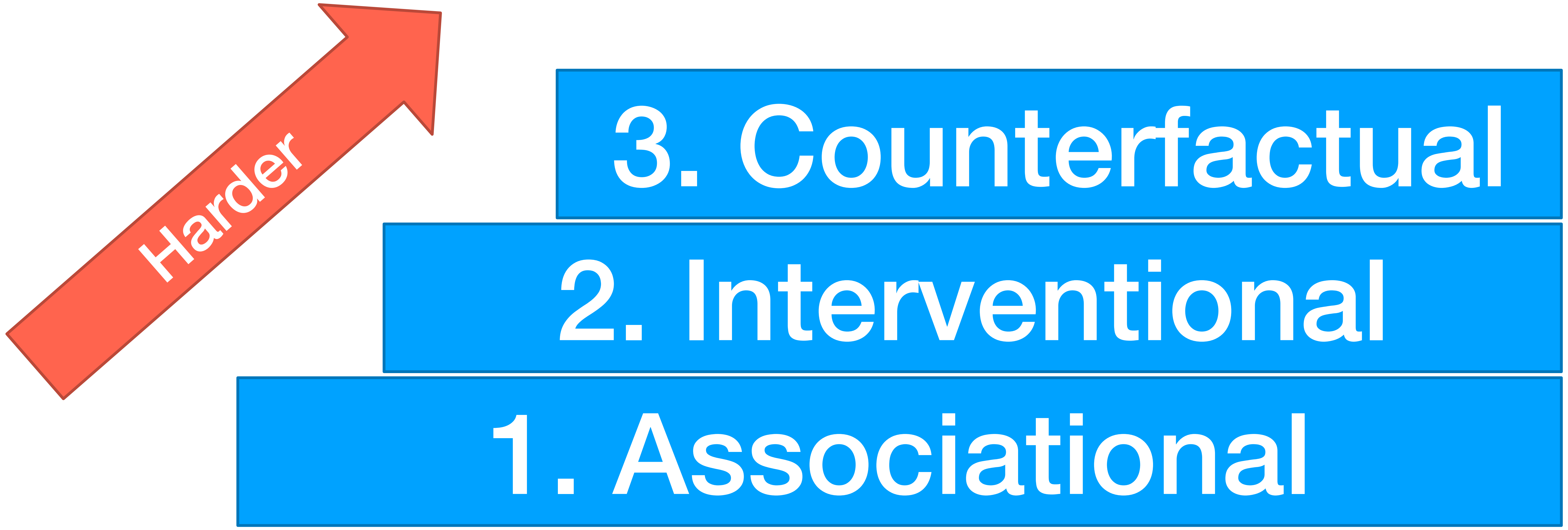
Balasubramaniam Srinivasan
Department of Computer Science
Purdue University
bsriniv@purdue.edu

Bruno Ribeiro
Department of Computer Science
Purdue University
ribeiro@cs.purdue.edu

End Origin Story

I ❤️ Causality

The 3 rungs of the ladder of causation



Rung 1: Associational

- Standard graph machine learning tasks

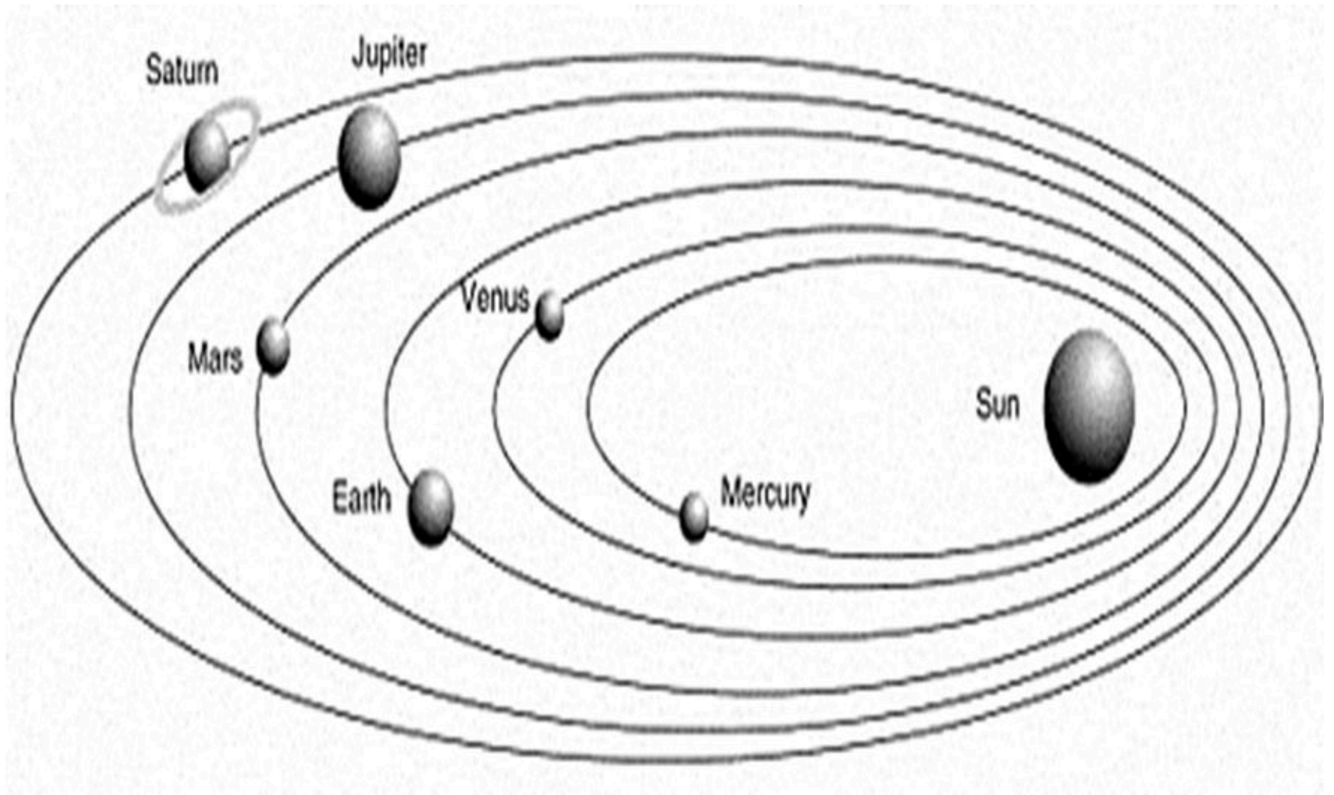
Assume $X \not\perp Y$

Task: Predict output Y from input X

Data: samples of (X,Y)

Most node & graph classification, link prediction tasks are associational

Kepler's Elliptical Orbits



socratic.org

Rung 2: Interventional

- Tasks where we must predict the effect of an intervention

Assume $X \not\perp\!\!\!\perp Y$

Task: Predict output Y from acting on input X

Data: samples of (Y,**do**(X=x))

Rung 2: Interventional (cont)

Imagine two hypothetical data generators for same $P(X, Y)$

$U_y, U_x \sim$ i.i.d. Uniform(0,1)

$$X := f_x(U_x) \qquad\qquad Y := f_y(U_x)$$
$$Y := f_y(X, U_Y) \quad = \quad X := f_x(Y, U_Y)$$

- $\mathrm{do}(X = x)$ changes f_x to a constant in data generation

$$X := x \qquad\qquad Y := f_y(U_x)$$
$$Y := f_y(X, U_Y) \quad \neq \quad X := x$$

Rung 3: Counterfactual

- Tasks where we must imagine the effect of an intervention at an event that has "already happened"

Assume $X \not\perp\!\!\!\perp Y$

Task: Predict output Y from acting on input X

Data: $Y(X = x) \,|\, X = x', Y = y'$ or $Y(X = x) \,|\, X = x'$

Rung 3: Counterfactual

Imagine two hypothetical data generators for same $P(X, Y)$

$U_y, U_x \sim$ i.i.d. Uniform$(0{,}1)$

$$X := f_x(U_x) \qquad\qquad \text{Y} := f_y(U_x)$$
$$Y := f_y(X, U_Y) \quad = \quad X := f_x(Y, U_Y)$$

- Now assume we know $X = x', Y = y'$

  This knowledge changes distribution of U_x and U_y

$$X := x \qquad\qquad \text{Y} := f_y\big(U_x \,|\, (X = x', Y = y')\big)$$
$$Y := f_y(X, U_Y \,|\, (X = x', Y = y')) \quad \neq \quad X := x$$

Some graph tasks are causal

Recommendations as treatments
Survey: (Joachims et al., AI Magazine 2021)

Link prediction for search & recommendations tends to be causal

$$Accept = y \mid do(Show\ recommendation = x)$$

PROCEEDINGS A

royalsocietypublishing.org/journal/rspa

Research

Cite this article: Cotta L, Bevilacqua B, Ahmed N, Ribeiro B. 2023 Causal lifting and link prediction. *Proc. R. Soc. A* **479**: 20230121. https://doi.org/10.1098/rspa.2023.0121

Causal lifting and link prediction

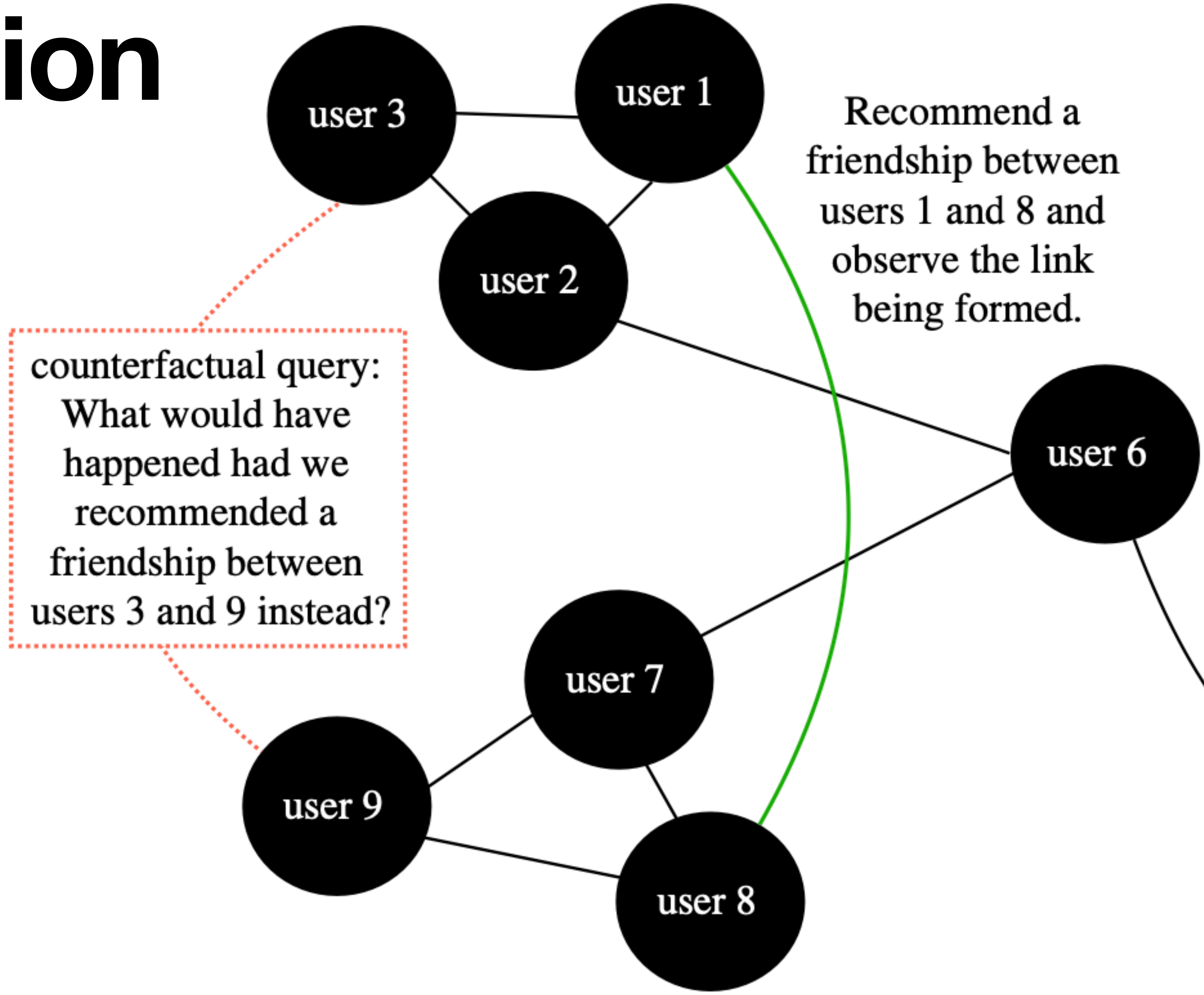
Leonardo Cotta[1], Beatrice Bevilacqua[2], Nesreen Ahmed[3] and Bruno Ribeiro[2]

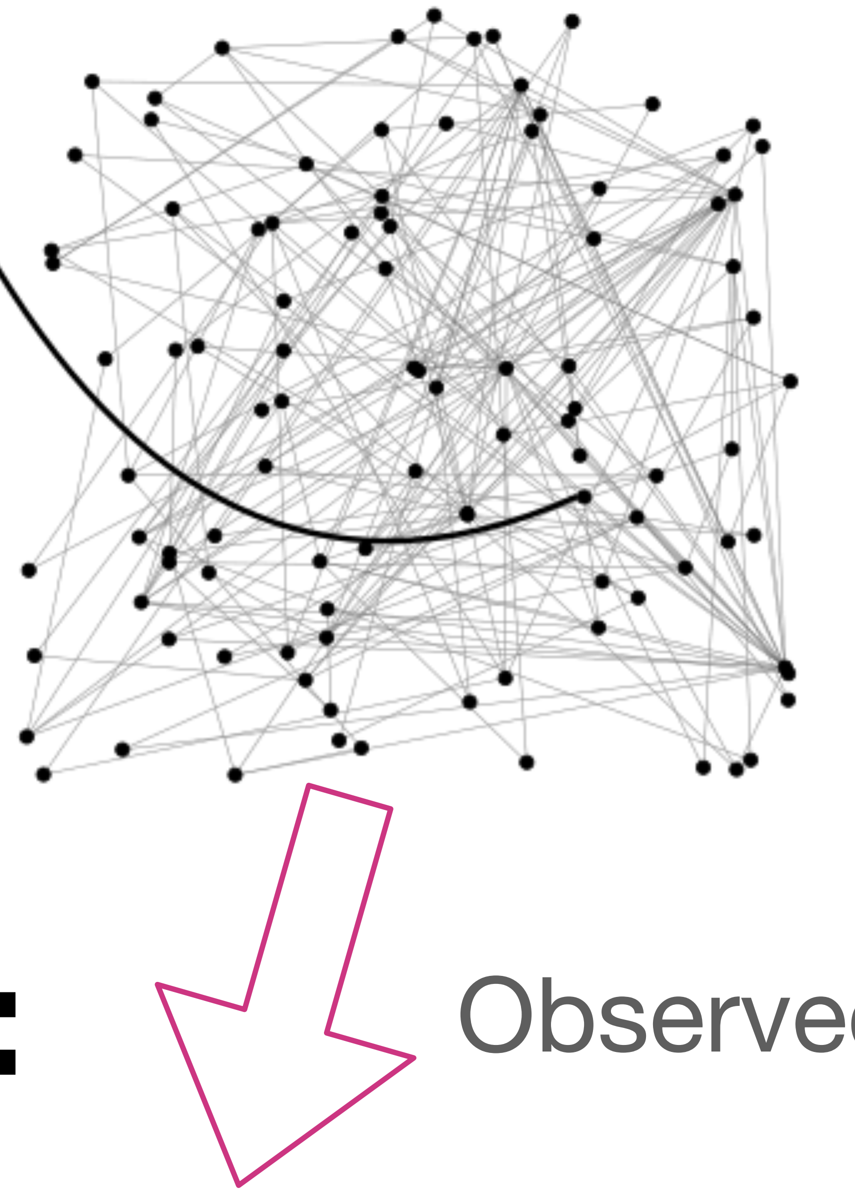
[1]Vector Institute, Toronto, Canada
[2]Purdue University, West Lafayette, IN, USA
[3]Intel Labs, Hillsboro, OR, USA

LC, 0000-0002-4751-7643

Causal Lifting: Causality ❤️ Symmetries

Observation



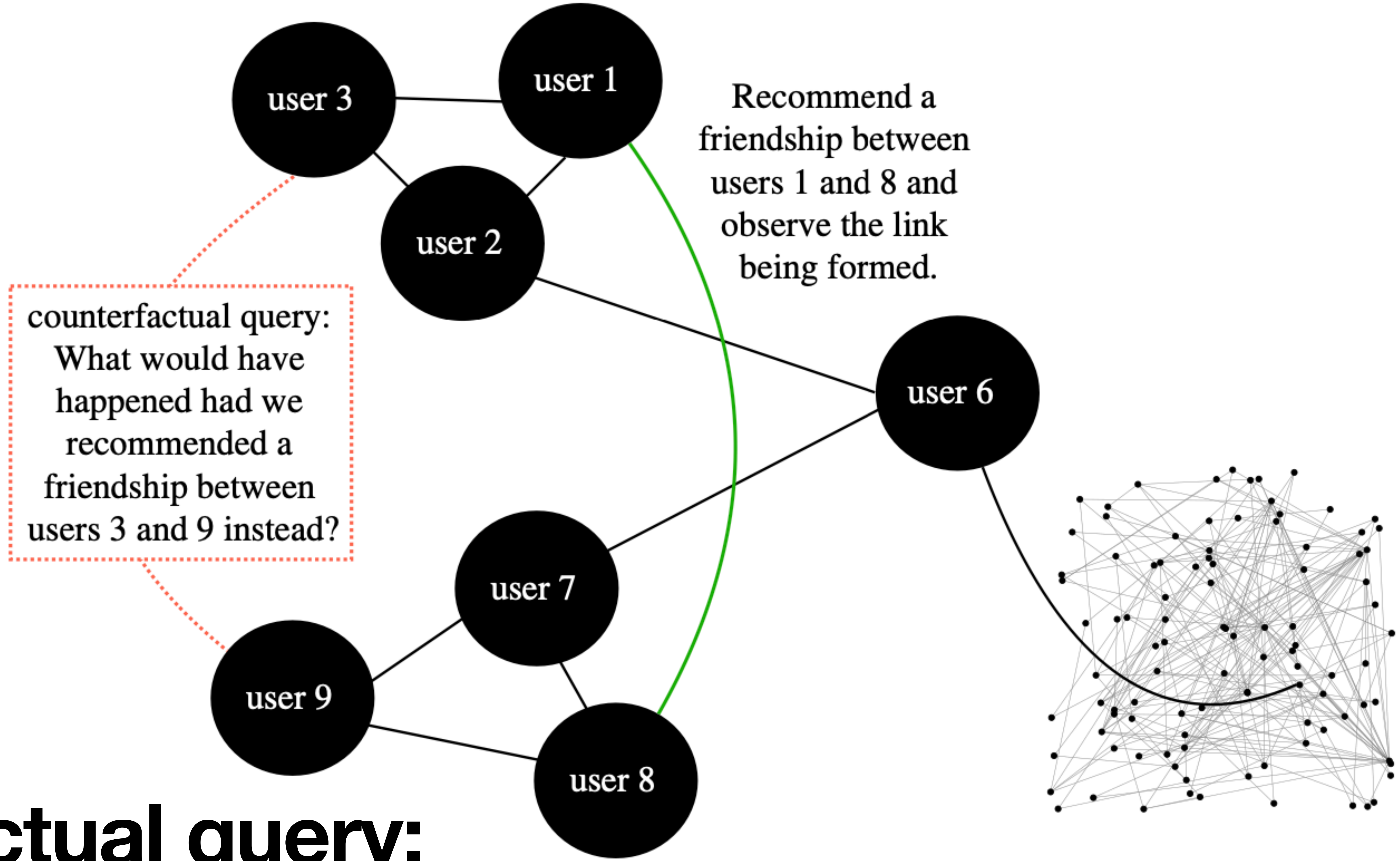


Observe an intervention (recommendation):

Observed graph at time t_0

$$\underbrace{Y_{IJ}^{(t_1)}}_{\text{outcome of probe in }(8,1)} := A^{(t_1)}_{\mathcal{E}^{(t_1)}} \underbrace{\left(\mathcal{E}^{(t_1)} = (8,1)\right)}_{\text{probe in }(8,1)} \mid G^{(t_0)},$$

Task

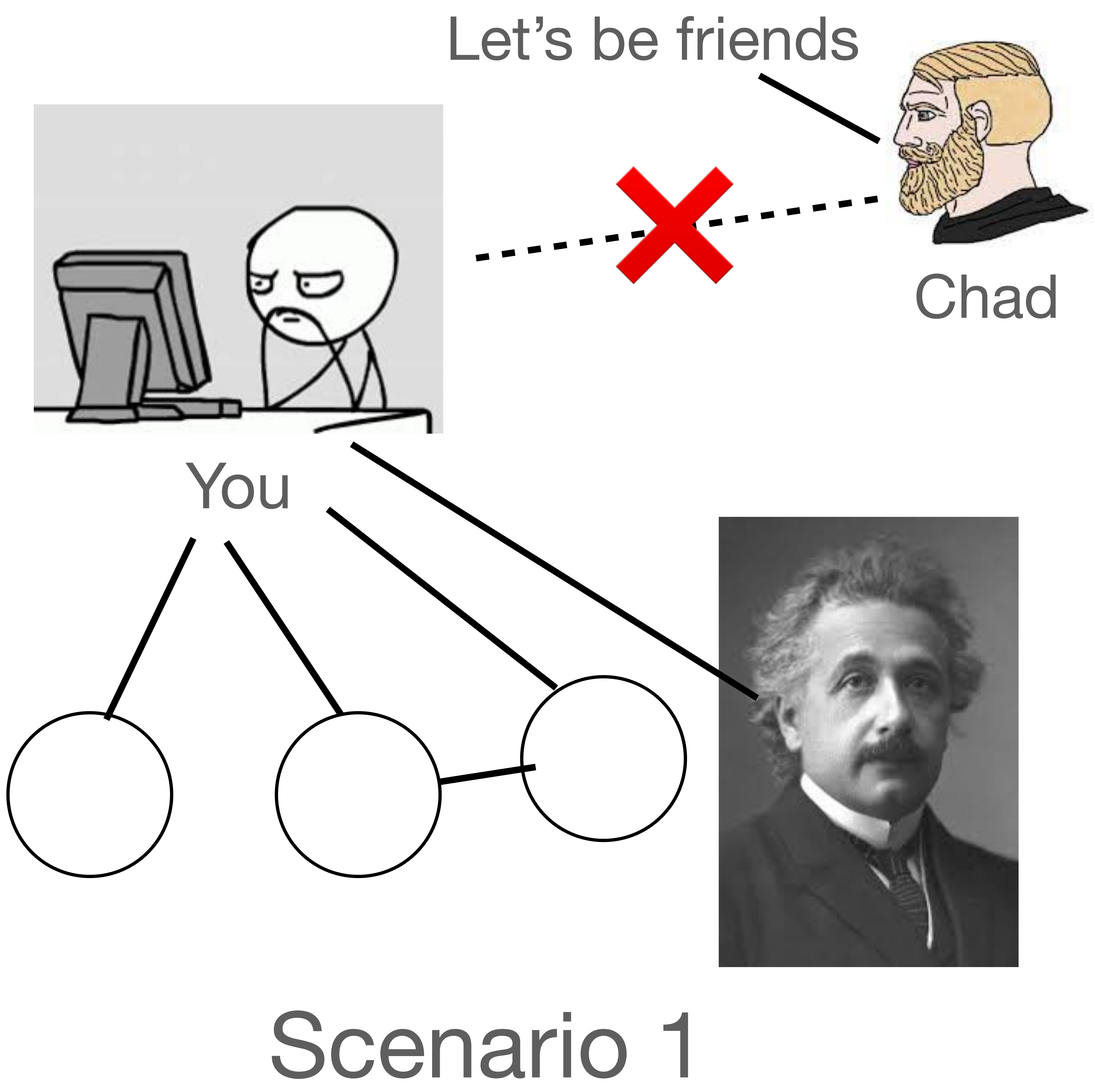


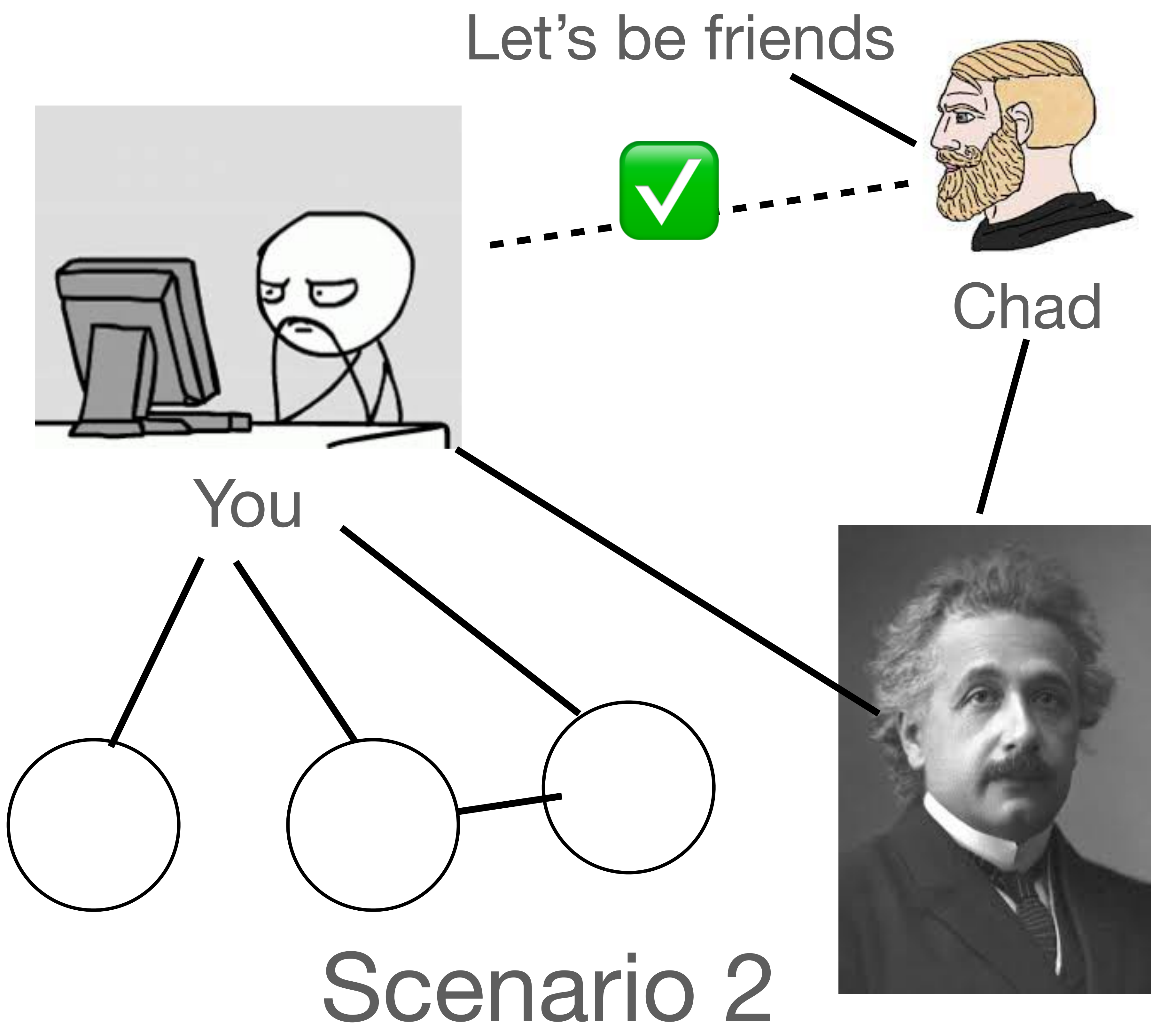
Counterfactual query:

$$P\Big(\underbrace{A^{(t_1)}_{\mathcal{E}^{(t_1)}}(\mathcal{E}^{(t_1)} = (3,9)\,)}_{\text{what would have happened had we probed in (3,9) instead?}} \mid A^{(t_1)}_{\mathcal{E}^{(t_1)}}(\mathcal{E}^{(t_1)} = (8,1)), G^{(t_0)}\Big) \equiv P\Big(Y^{(t_1)}_{39} \mid Y^{(t_1)}_{81}\Big)$$

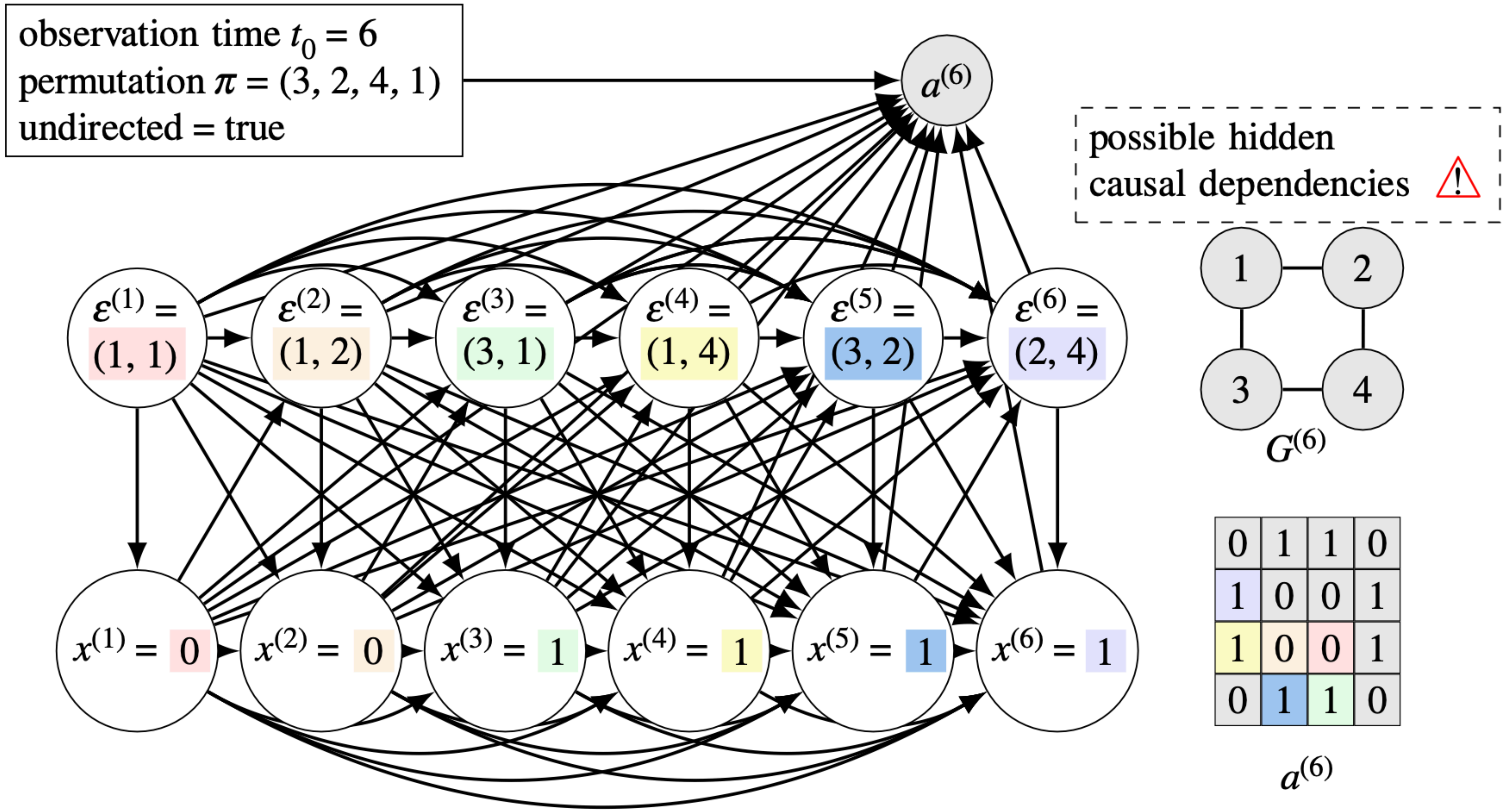
Challenge:

- Path-dependency in graph evolution
  - **Graph evolution may depend on current state of graph**





Universal Structural Causal Model for Path-dependent Graphs



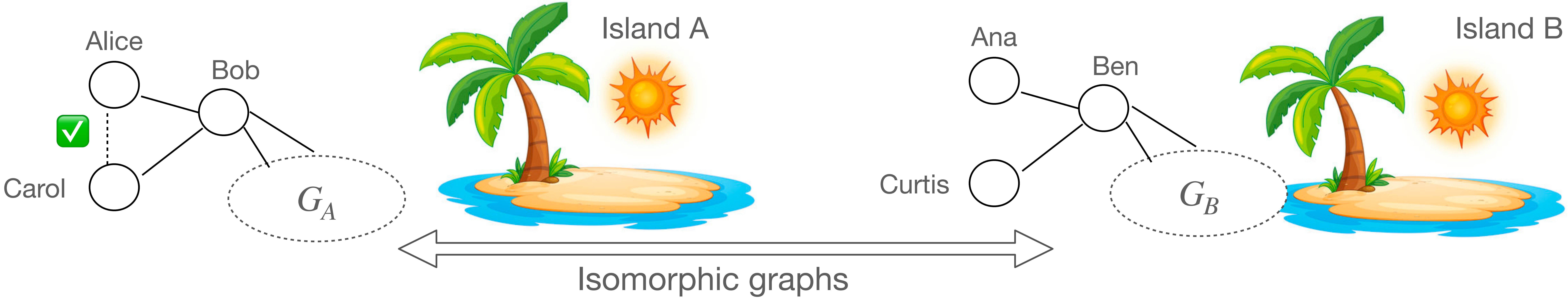
- Path-dependency = complex causal dependencies

Theorem 4.1 (Universality of our graph SCM). *Let $\mathbb{C}$ be the family of SCMs as defined in definition* 13 *and* 14 *in electronic supplementary material, appendix* C *and $\mathbb{A}$ be the domain of the entries of the adjacency matrices of the graphs generated by it. Then,*

(*i*) *For every SCM $\mathcal{C} \in \mathbb{C}$ at an arbitrary observation time $t_0 \geq 0$, $\mathcal{C}$ always generates observed graphs $G^{(t_0)}$ where $P(A^{(t_0)} = a) = P(A^{(t_0)} = a')$ for any two isomorphic graphs with adjacencies $a, a' \in \mathbb{A}$;*

(*ii*) *For all finite (jointly) exchangeable graph distributions $P(A^{(t_0)})$, if $\mathbb{A}$ is a countable set there exists an SCM $\mathcal{C} \in \mathbb{C}$ and an observation time $t_0 \geq 0$ that induces it.*

How Symmetries Can Help



- Consider two deserted islands
  - Assume the same **structural causal model** generated the two social networks
    - Also, for now, the social graphs of Islands A and B will be **isomorphic**
  - Assume we suggest **Alice to Carol** and she accepts ✅
  - In island B, we expect the suggestion of **Ana to Curtis** should have a similar outcome (in distribution)

Can graph structure can help with the counterfactual query?

*Given **Alice — Carol** was ✅,*
*what would have **happened** if we had probed **Ana — Curtis** instead?*

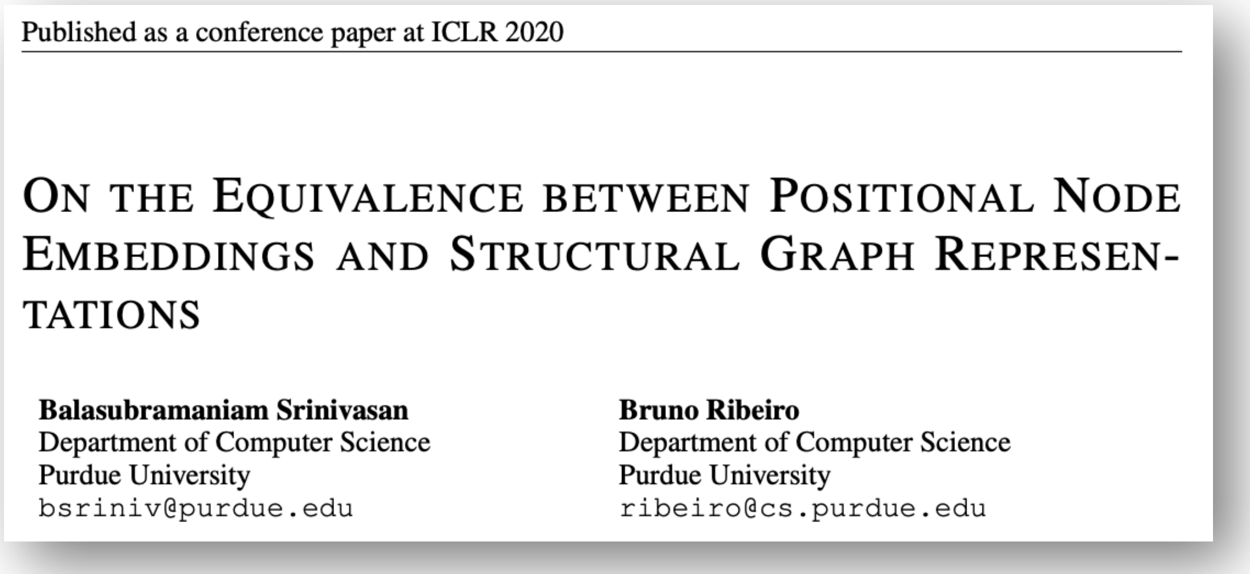
On Graph Learning

Positional node embeddings (e.g. matrix factorization)

Graph learning happens through graph embeddings

Positional node embeddings describe how nodes are positioned in the graph
- It does not preserve the symmetries in the graph

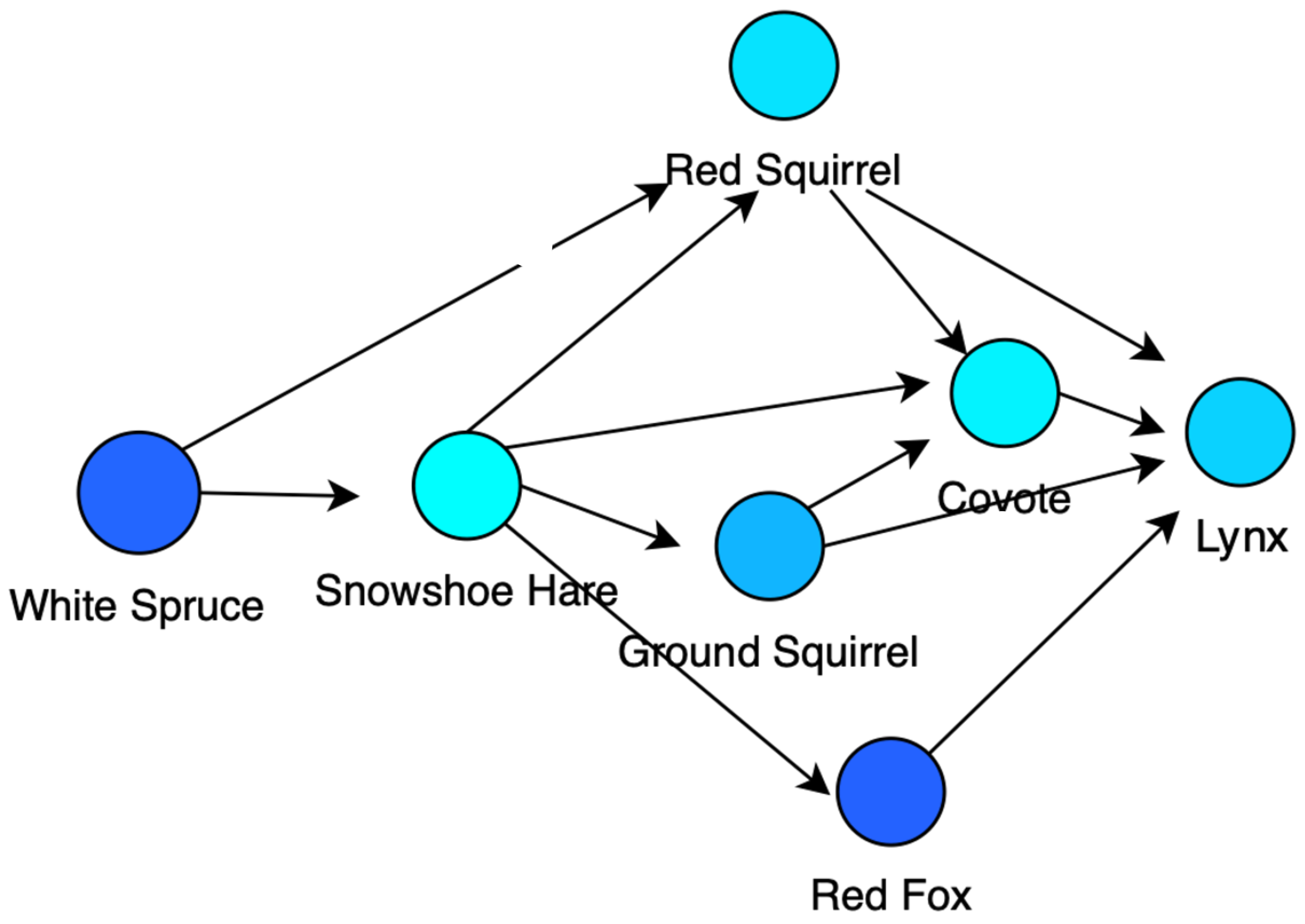
Consider positional node embeddings in the Arctic food web
- Similar colors = similar node embeddings

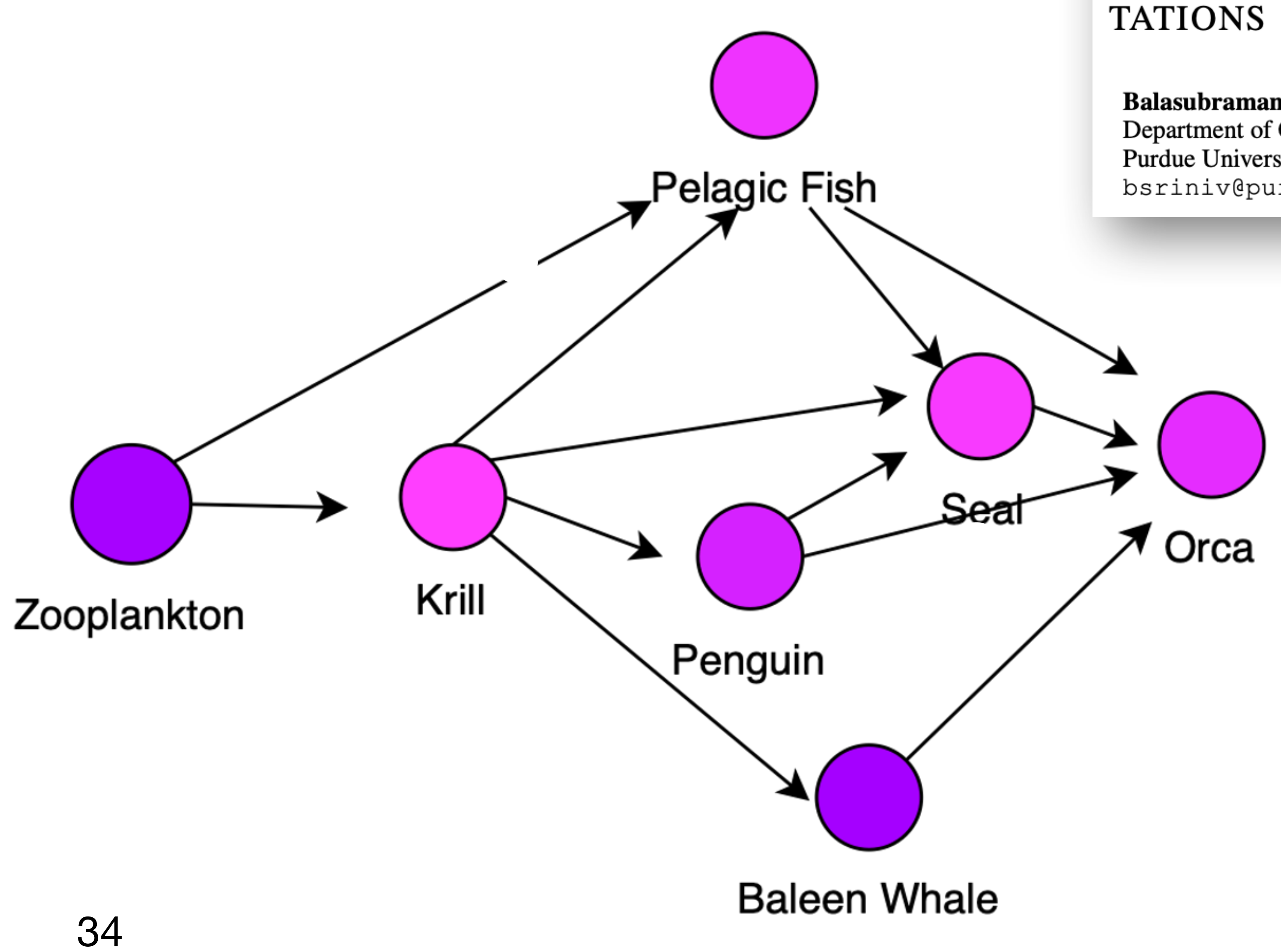

Published as a conference paper at ICLR 2020

On the Equivalence between Positional Node Embeddings and Structural Graph Representations

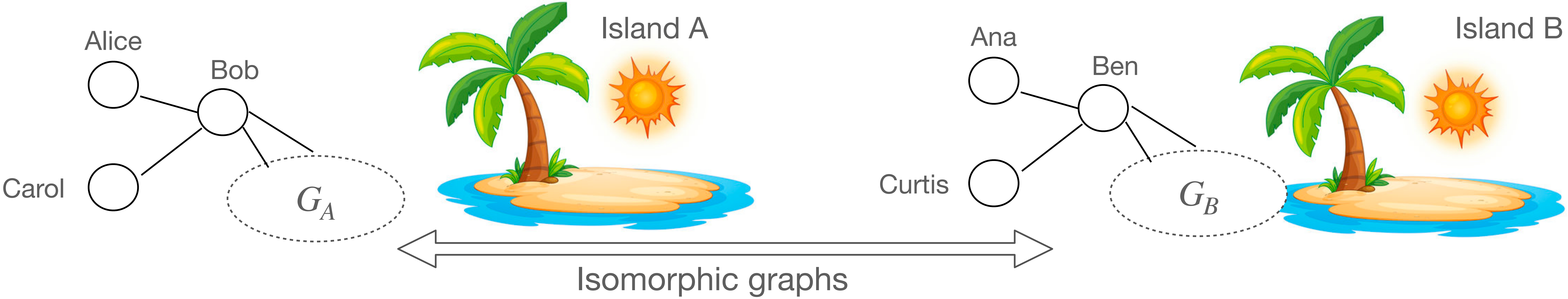
Balasubramaniam Srinivasan
Department of Computer Science
Purdue University
bsriniv@purdue.edu

Bruno Ribeiro
Department of Computer Science
Purdue University
ribeiro@cs.purdue.edu





Positional Embeddings and Causal Link Predictions



- The positional nature of node embeddings means whatever we learn for Alice and Carol won't transfer to Ana and Curtis

Causal Lifting

- Associational lifting: Let $\mathcal{G}$ be a group and $\circ$ is the left action of $\mathcal{G}$ onto $\text{supp}(X)$
  E.g., (Kimmig et al., 2014)

$$P(Y|X=x) = P(Y|X=g \circ x) \qquad g \in \mathcal{G}$$

- Definition 3.2 (Interventional lifting):

$$P(Y|\text{do}(X=x)) = P(Y|\text{do}(X=g \cdot x))$$

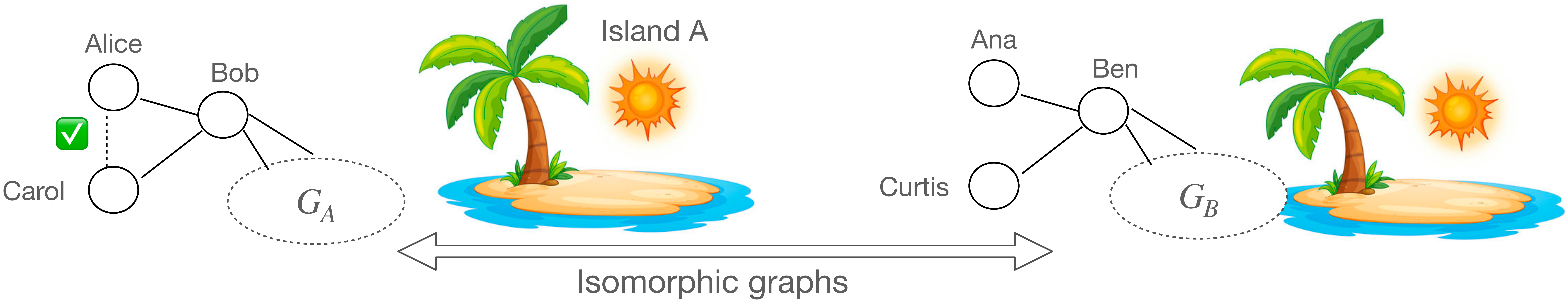
$$P(Y(X=x)) = P(Y(X=g \cdot x)) \qquad \text{Imbens notation}$$

- Definition 3.3 (Counterfactual lifting):

$$P(Y(X=x)|X=x') = P(Y(X=g \circ x)|X=x') \qquad \text{Imbens notation}$$

Causal Lifting on Graphs

- **The embeddings of two or more nodes must be structural, not depend on the ordering of the nodes**
  - **Structurally, Alice & Carol and isomorphic to Ana & Curtis before intervention**
    - **This is a structural symmetry**



What are the assumptions needed in the Structural Causal Model (SCM)

to get transferability-through-symmetry?

The Symmetry Assumptions in our SCM

Assumption 4.2 (Time gap ignorability (informal)). We say that our SCM satisfies time gap ignorability if the mechanism $f_{\mathcal{X}}^{(t_1)}$ is invariant to the SCM intermediate states between the time the intervention probe is performed t_0 and the instant before we see its effect in t_1.

Assumption 4.3 (Time exchangeability (informal)). We say that our SCM satisfies time exchangeability if the mechanism $f_{\mathcal{X}}^{(t_1)}$ is invariant to the order in which edges and non-edges have been generated.

Assumption 4.4 (Non-link ignorability (informal)). We say that our SCM satisfies non-link ignorability if the mechanism $f_{\mathcal{X}}^{(t_1)}$ is invariant to which pairs of nodes were generated as non-links or were not generated at all at time t_0.

Assumption 4.5 (Identifier exchangeability (informal)). We say that our SCM satisfies identifier exchangeability if the mechanism $f_{\mathcal{X}}^{(t_1)}$ is invariant to permutations of the node identifiers.

If Assumptions 1-4 hold, then…

- Theorem 4.6 (Invariances for interventional lifting in link prediction).
  - Under Assumptions 1-4 in our Universal SCM for path-dependent graphs, we can prove that causal lifting can be used to obtain an **equivalent SCM** using just the **observed graph** symmetries:
    - *Where $W_{O_{IJ}}$ is a variable shared by all nodes structurally identical to the pair IJ*

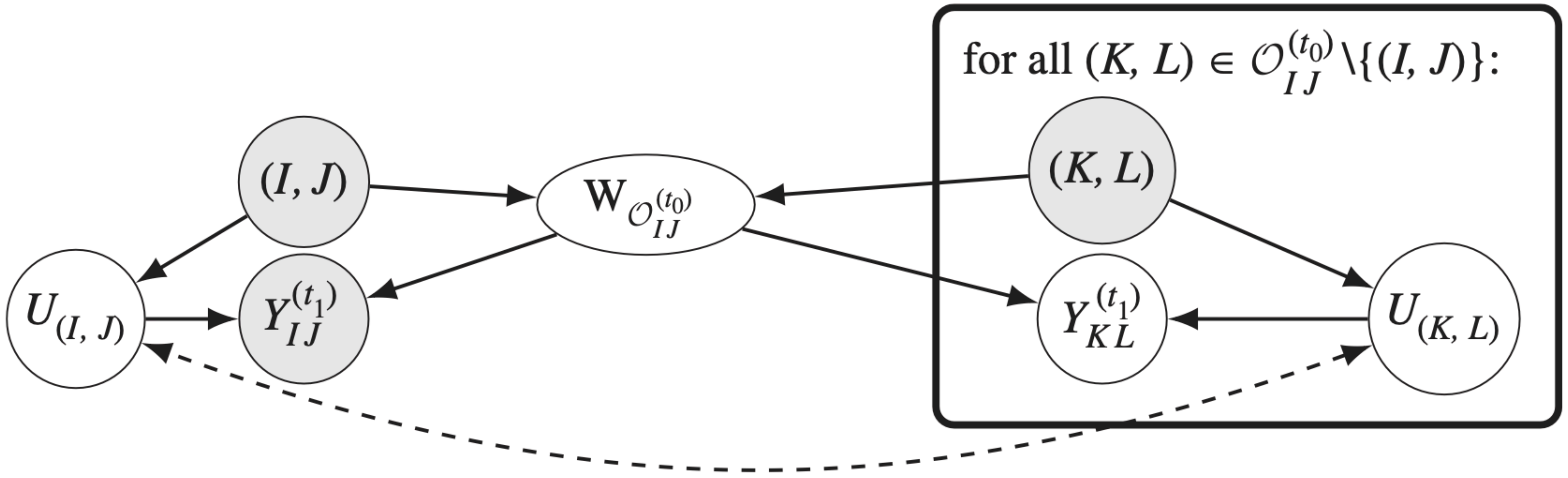


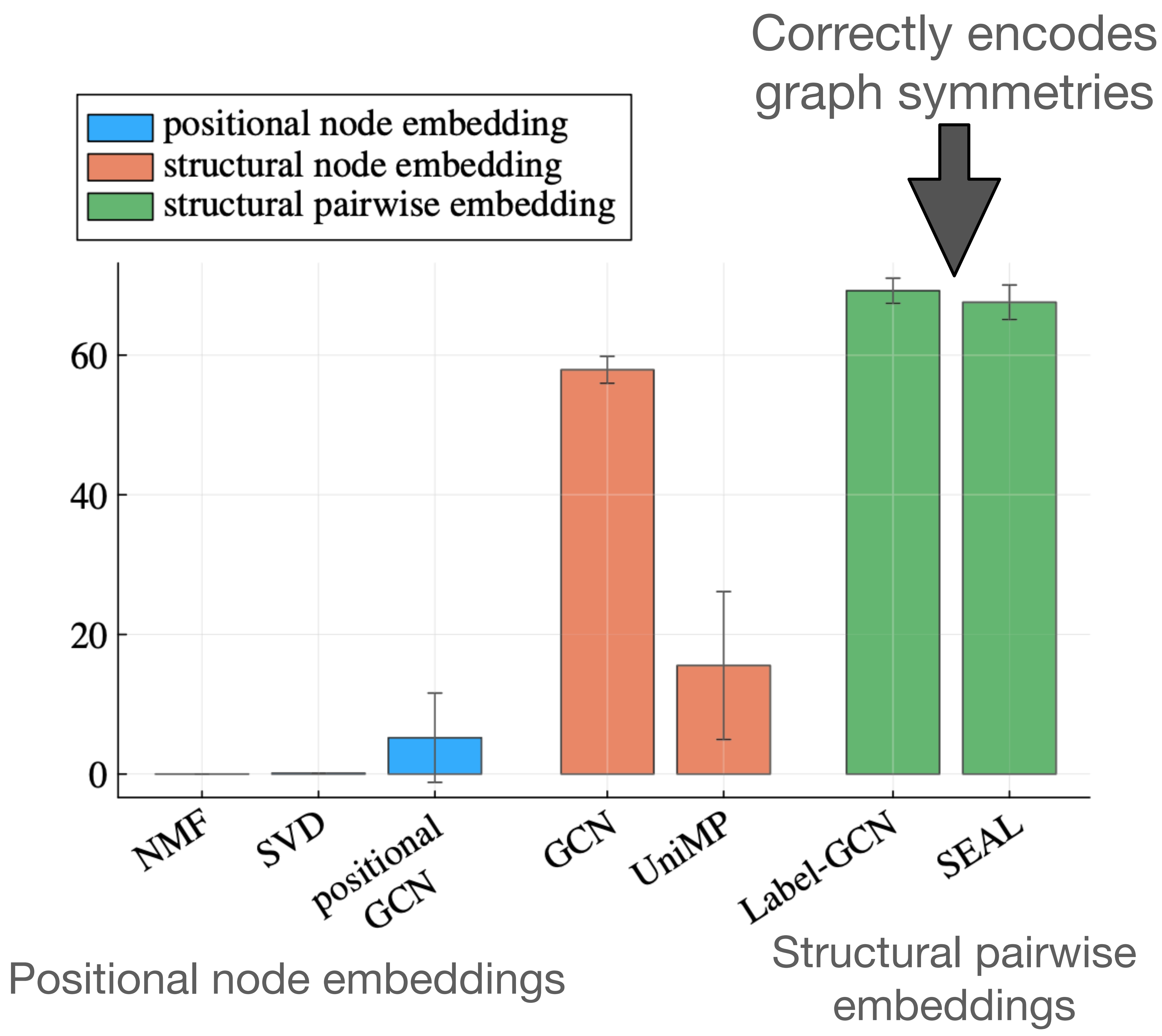
Figure 3. (Theorem 4.6(i)) Causal DAG of an equivalent data generating process of a probe in (i, j) (left) and in its orbit (right). As usual, we represent observed and unobserved variables with grey and white nodes, respectively.

Extra assumption (no spillover) needed for learning from multiple experiments

Future work: Redefine symmetries to account for spillover effects

Example:

- Recommendations for Amazon purchases
  - In training we consider the subgroup of male users in recommendations.
  - At test time, our counterfactual queries are about female users.



Take-home

- **Causal Lifting** allow us to use observed **symmetries** (invariances) to predict the outcome of **causal queries.**
  - It lifts the interventions in one part of the graph to predict on other parts of the graph
  - Relaxing Assumptions 1-4 create new symmetries in the SCM
    - Causal lifting still applies, but we may need extra data for the new symmetries

- Concept of **clustering** for **containing spillover** **changes** under **causal lifting**
  - Spillover also could be accounted for under an expanded notion of symmetries (Causal lifting can account for spillovers)

Thank you!