CAUSAL INFERENCE FROM RELATIONAL DATA

Presenters:
Elena Zheleva, University of Illinois Chicago @elenadata
David Arbour, Adobe Research @darbour26

UIC
Adobe

AAAI 2022 Tutorial
February 23, 2022
https://netcause.github.io

TUTORIAL LOGISTICS

Website: https://netcause.github.io

- All materials, slides & references
- Our contact information

You can ask David and Elena questions during the tutorial over chat

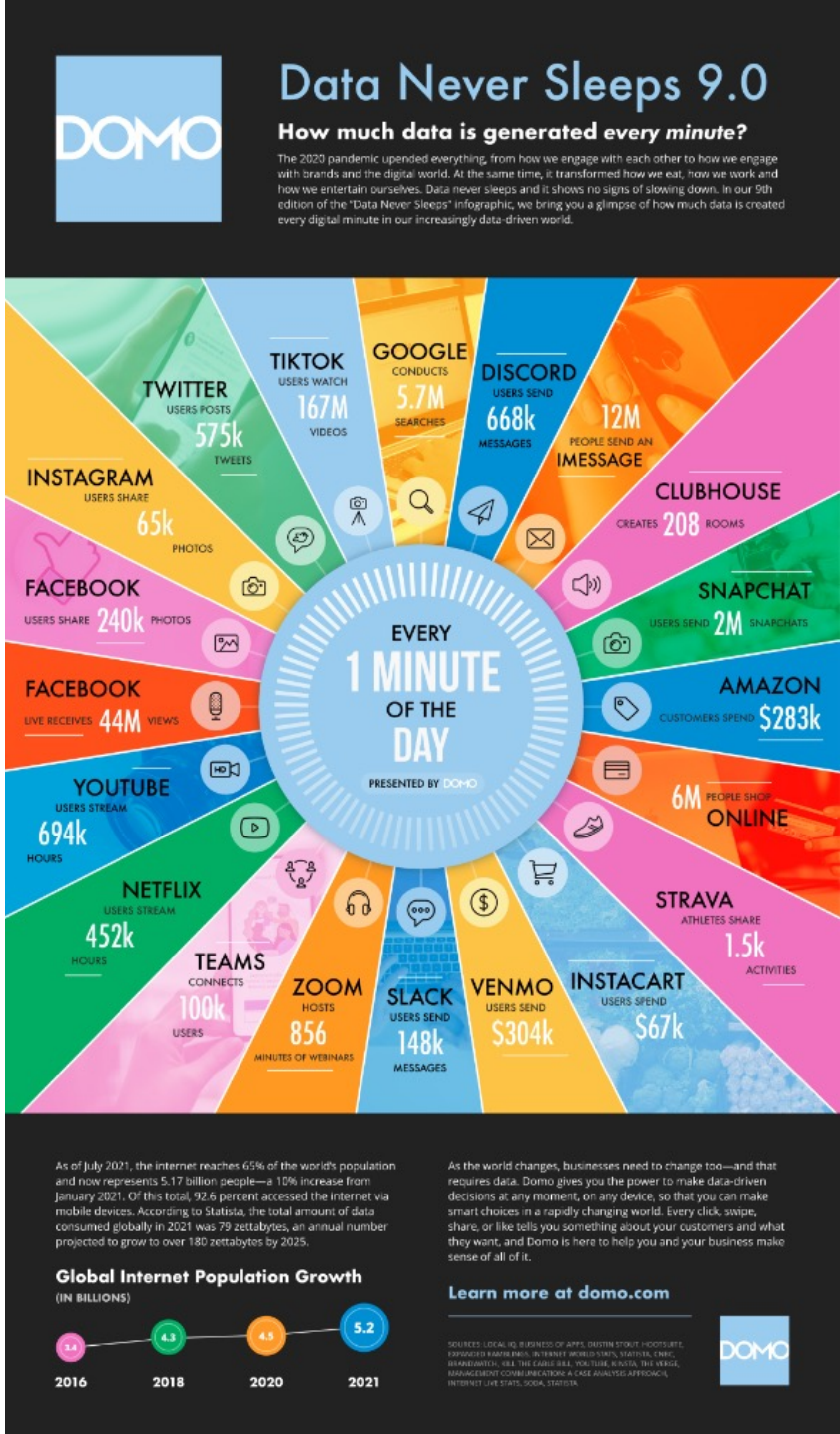
There will be a short break half-way through the tutorial

Note: the tutorial uses images from the papers it covers

CAUSAL INFERENCE

Causal inference is the study of how actions, interventions, or treatments affect outcomes of interest

Increasing interest in studying social phenomena and extracting causal insights from large amounts of "found" data




What messages in online support groups
cause people to feel more empathy?

Can social media interactions make users more “hateful” and why?


79
94
143
376
23
273
17
462
What social interventions can facilitate
the viral spread of a product?

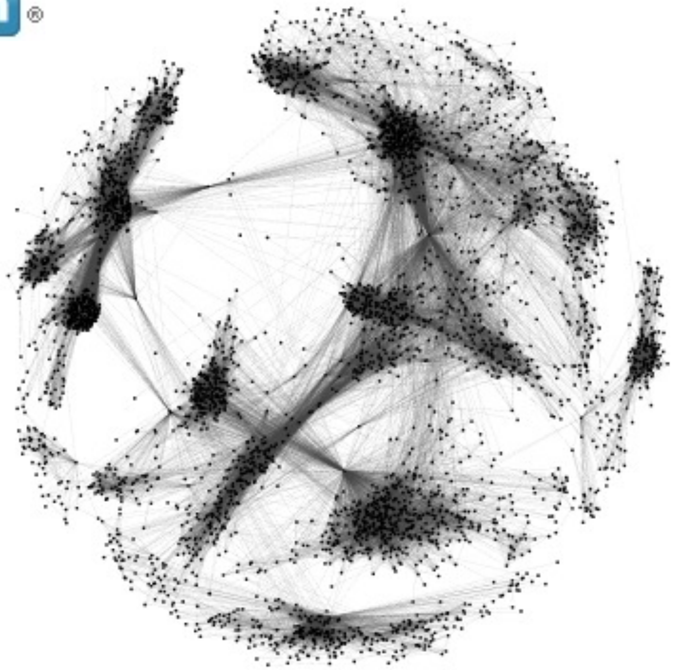
CAUSAL INFERENCE AND INTERFERENCE

Common among these questions:

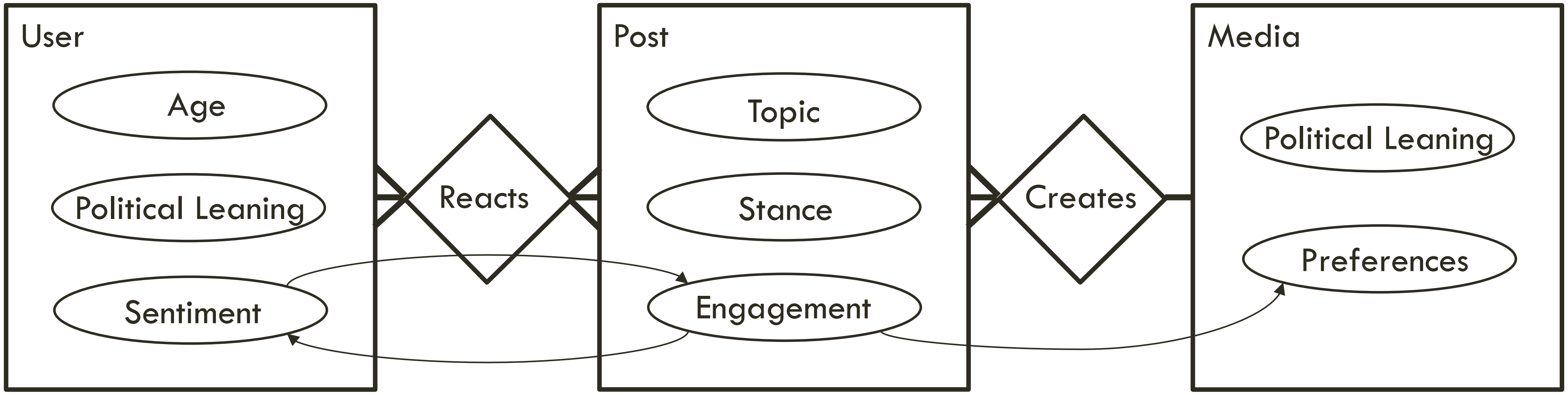
1) They are concerned with causes and effects
2) There is data from digital platforms that may help with answering them
3) Interference: the actions of one user can affect the actions of others

When and how can we answer causal questions of interest while accounting for interference?

INTERFERENCE

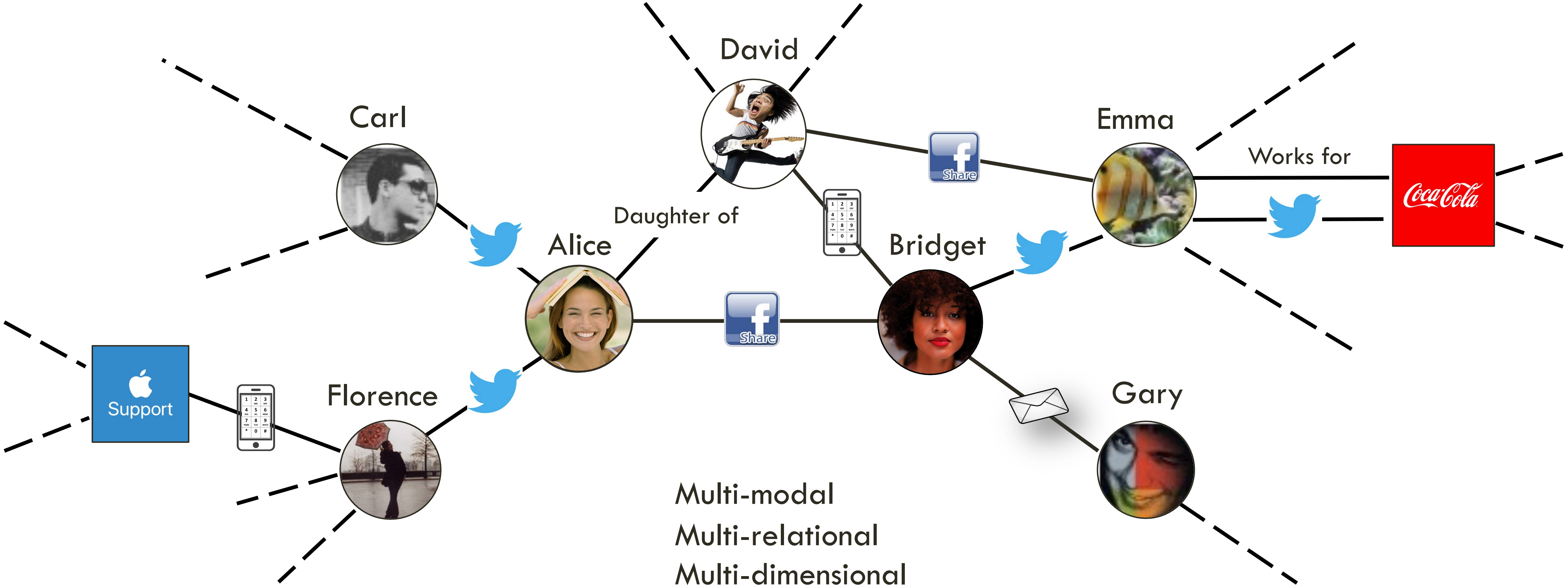


RELATIONAL DATA



Real-world data is rarely flat!

HETEROGENEOUS NETWORKS



Multi-modal
Multi-relational
Multi-dimensional

TUTORIAL OUTLINE

Background

- Motivation
- Causal inference 101
- Causal effects in networks

Interventions and network experiment design

1/3 of tutorial

Counterfactuals & causal effects in observational data

- Representation, identification, estimation
  - Block representation

  --- 10-minute BREAK ---

  - Representation challenges
  - Chain and segregated graphs
  - Multi-relational data and abstract ground graphs
- Discovery

2/3 of tutorial

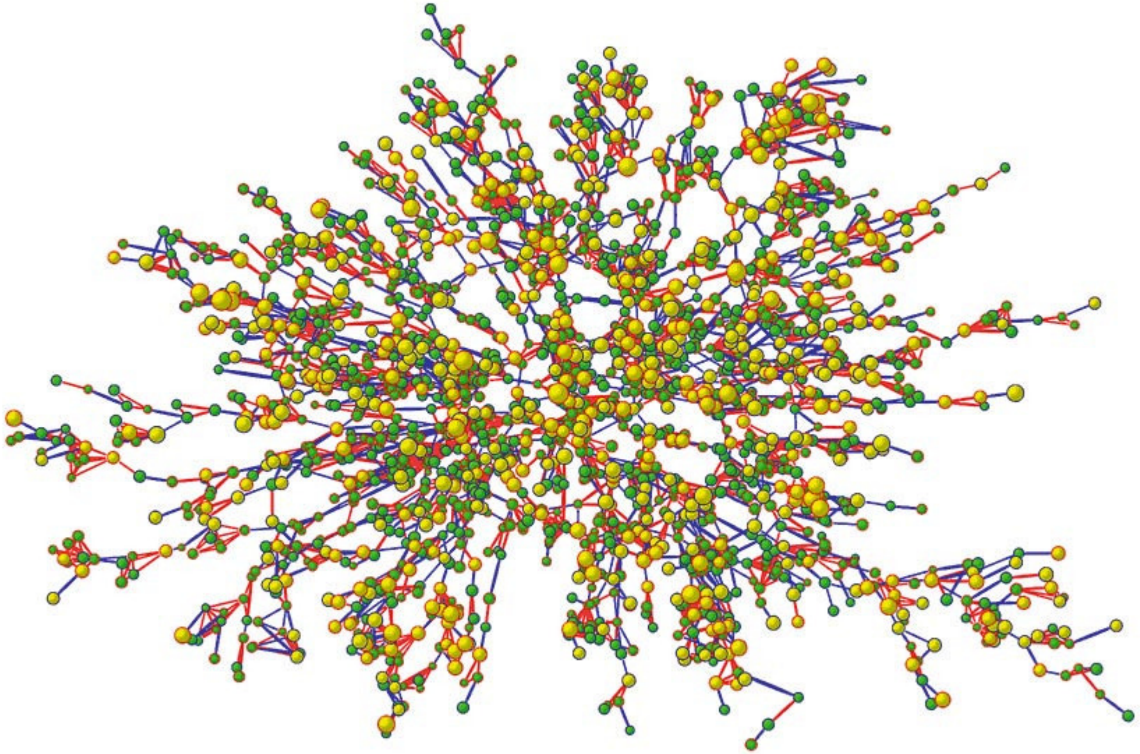
EXAMPLE: SPREAD OF OBESITY

Analyzed person-to-person spread of obesity

"A person's chances of becoming obese increased by 57% if he or she had a friend who became obese in a given interval"

Similar studies on spread of smoking and happiness

These studies may suffer from spurious associations due to network dependence**



Christakis & Fowler. *The Spread of Obesity in a Large Social Network Over 32 Years*. New England Journal of Medicine. 2007.
**Lee & Ogburn. *Network Dependence Can Lead to Spurious Associations and Invalid Inference*. Journal of American Statistical Association. 2020.

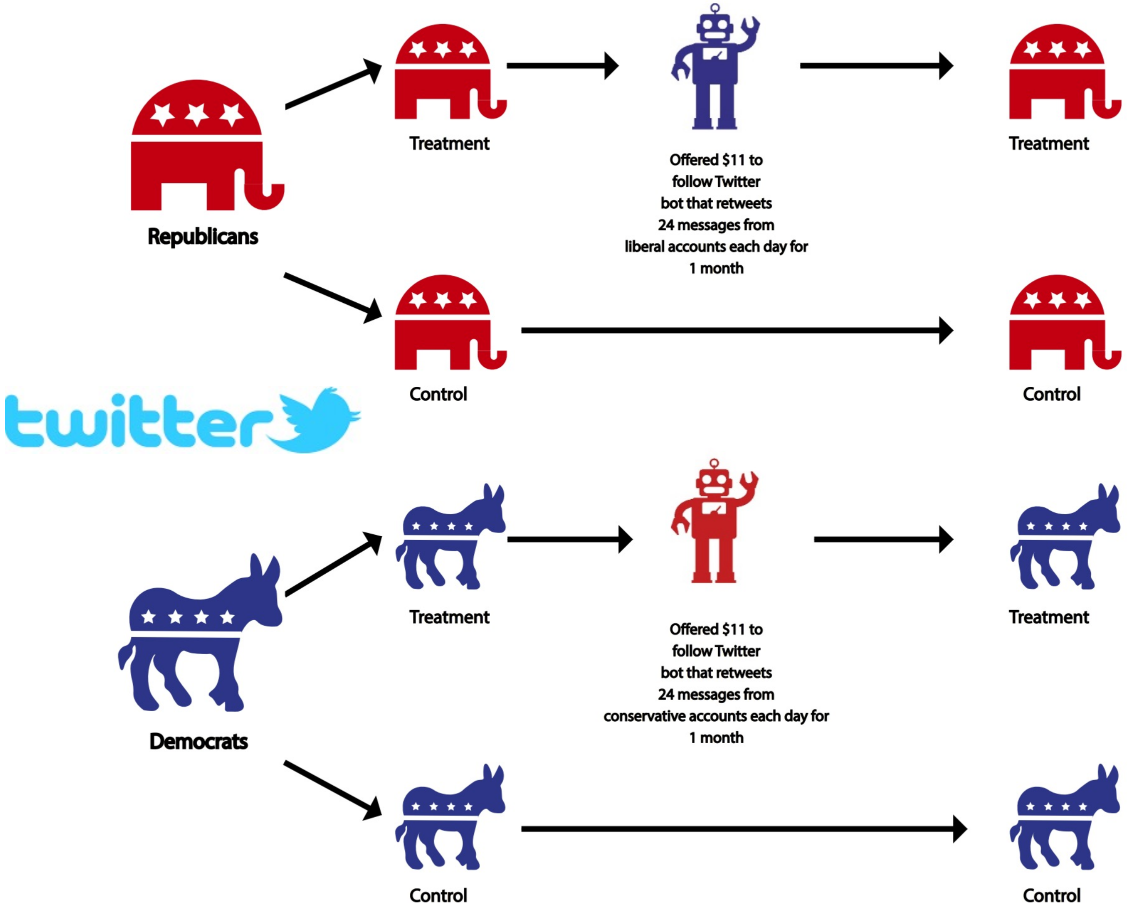
EXAMPLE: SOCIAL MEDIA AND POLARIZATION

Expose people to opposite views => get along better?

Block randomization at level of party attachment and interest in current events

Answered questions before and after 1 month of following bot of opposite view

Republicans became significantly more conservative and Democrats slightly more liberal



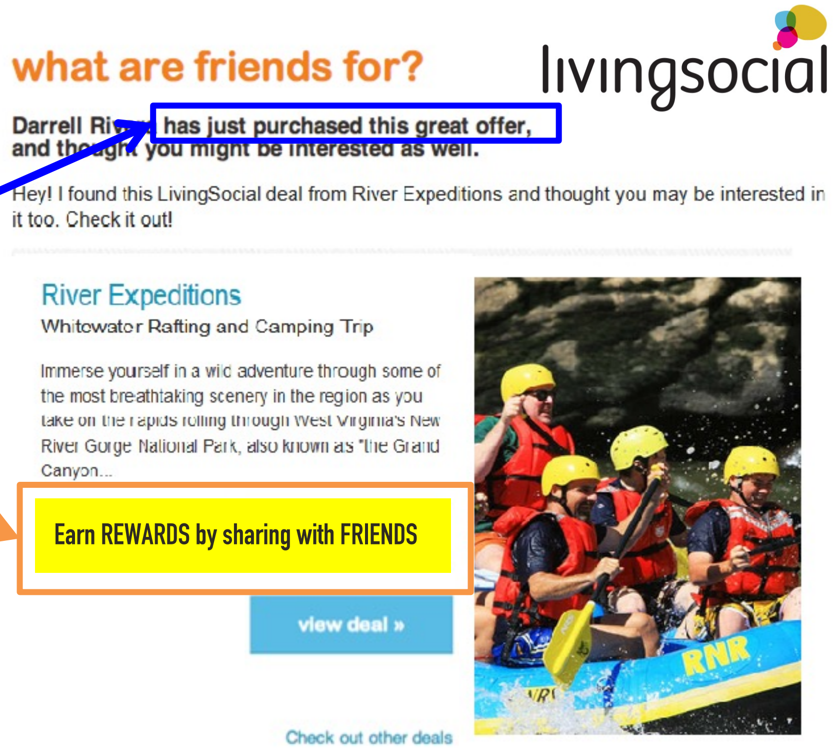
Bail, Argyle, Brown, Bumpus, Chen, Hunzaker, Lee, Mann, Merhout, Volfovsky. *Exposure to opposing views on social media can increase political polarization.* PNAS 2018.

EXAMPLE: VIRAL MARKETING

Customers can choose:

1. Product to share with friends
2. Share recipient

Company can vary the rest of the message



	Added info	Referred purchases	Follow-up referrals
Endorsement effect	Sharer purchase	15% lift	No effect
Incentive effect	Referral incentive	No effect	65% lift
	Both	No effect	No effect

T. Sun, S. Viswanathan, E. Zheleva. *Creating Social Contagion through Firm-mediated Message Design: Evidence from A Randomized Field Experiment.* Management Science 2021.

HOMOPHILY VS. CONTAGION

CAUSAL INFERENCE 101

RELATED TUTORIALS

Shalit & Sontag. Causal Inference for Observational Studies. ICML 2016

- https://shalit.net.technion.ac.il/homepage/causal-inference-tutorial-icml-2016/

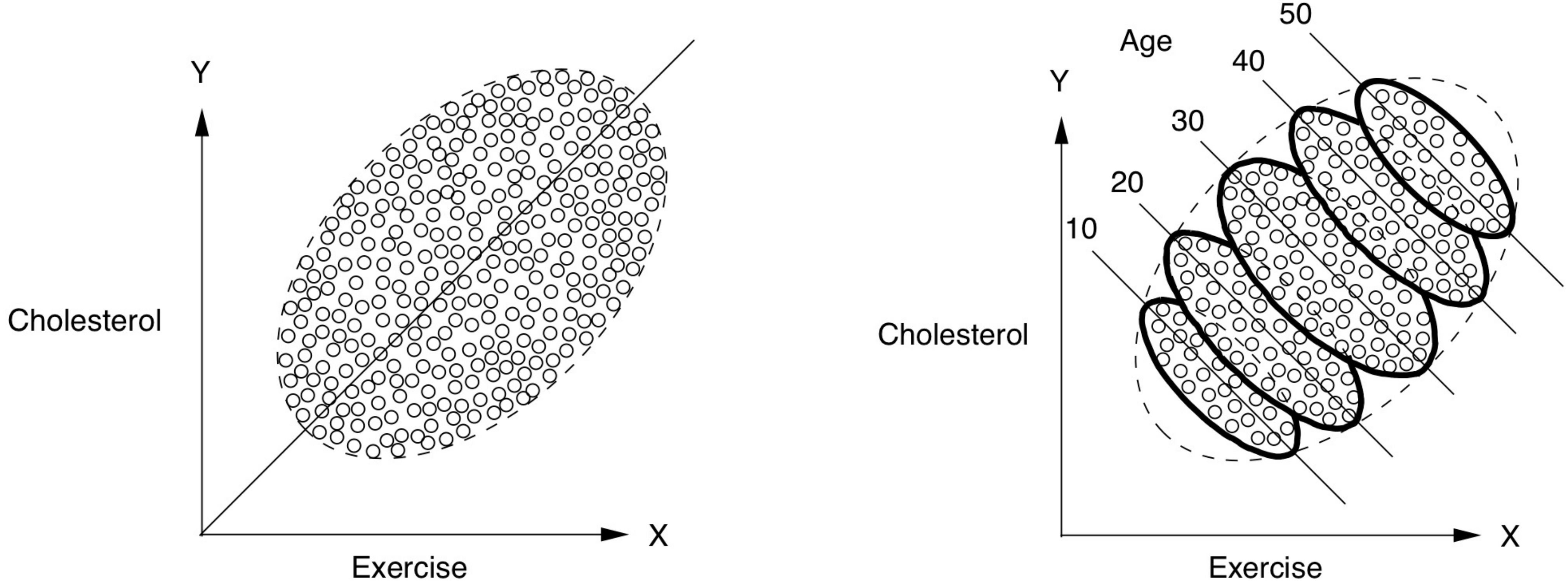
Kiciman, Sharma. Causal Inference and Counterfactual Reasoning. KDD 2018.

- https://causalinference.gitlab.io/kdd-tutorial/

Zheleva, Arbour. Causal Inference from Network Data. KDD 2021.

- https://netcause.github.io

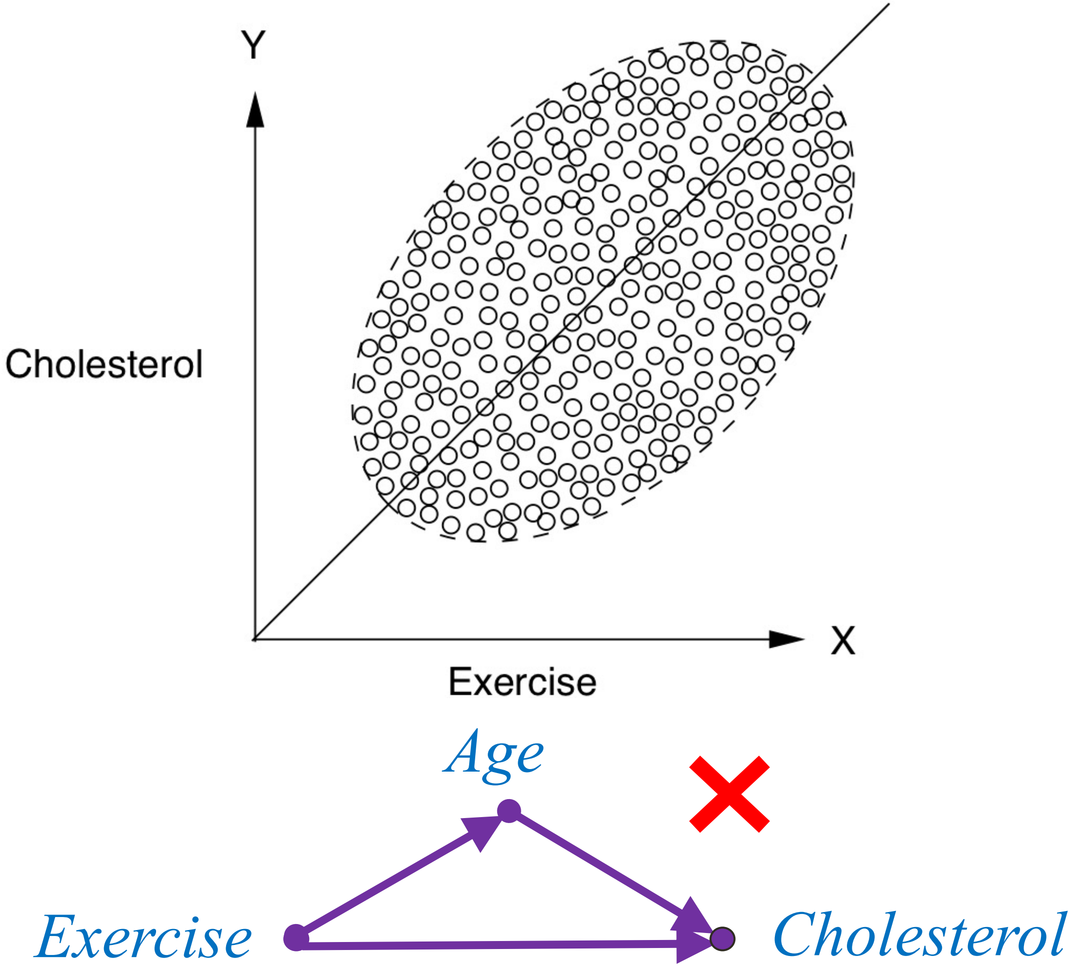
SIMPSON'S PARADOX

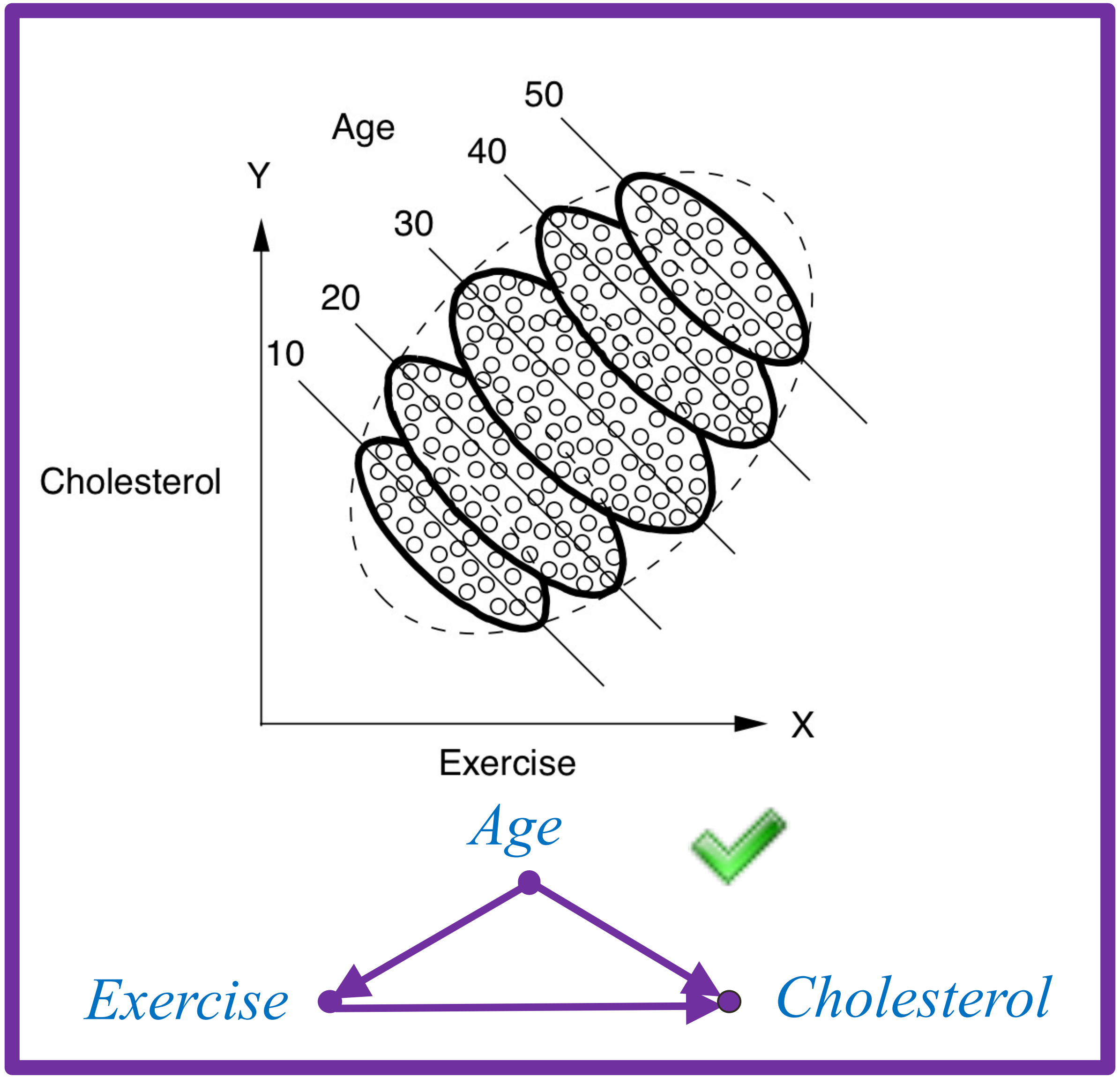


Same data can have different causal explanations!

Example by Judea Pearl.

SIMPSON'S PARADOX





POTENTIAL OUTCOMES AND COUNTERFACTUALS

Treatment (Z): something administered to experimental units; a cause of interest (e.g., *received vaccine or not*)

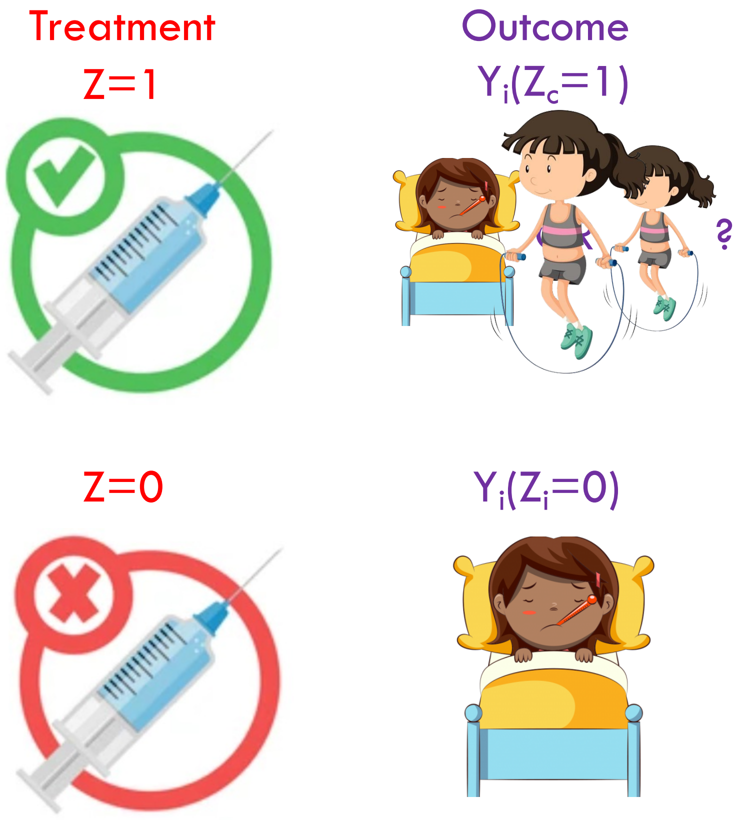
Potential outcome: the outcome $Y_i(z)$ that would be realized if an individual i received a specific treatment z (e.g., *got sick or not*)

Counterfactual: the outcome $Y_i(z_c)$ that would have been realized had an individual had a different treatment z_c than the observed z_i

Individual causal effect: $Y_i(Z=1)–Y_i(Z=0) = Y_i(1)–Y_i(0)$

Fundamental law of causal inference: $Y_i(0)$ can never be observed at the same time as $Y_i(1)$ and the causal effect cannot be measured

How do we estimate causal effects then?



COMMON CAUSAL ESTIMANDS

Individual effects are hard to estimate. Instead:

Under certain assumptions

Average treatment effect (ATE)

$$E[Y(1) - Y(0)] \cong \frac{1}{n}\sum_{i=1}^{n}(Y_i(1) - Y_i(0)) \cong \frac{1}{n}\sum_{i=1}^{n}(\overbrace{Y_i(1)Z_i}^{\text{Factual if in treatment}} - \overbrace{Y_i(0)(1 - Z_i)}^{\text{Factual if in control}})$$

Conditional average treatment effect (CATE)

$$E[Y(1) - Y(0)|\boldsymbol{X} = \boldsymbol{x}]$$

i	Z	Y(Z$_1$)	Y(Z$_0$)	Sex	Education
1		1	?	F	High School
2		?	0	F	Bachelors
3		?	1	M	High School
…					
n		1	?	M	Masters

$\boldsymbol{X}$

COMMON ASSUMPTIONS

Consistency: $Y_i(z_i) = y_i$ when $Z = z_i$

Positivity/overlap: a unit could have received any treatment $P(Z_i = z|\boldsymbol{X} = x_i) > 0, \forall z, x_i$

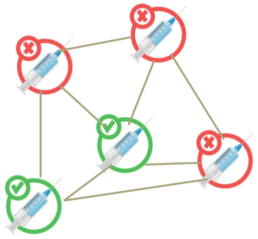
No unmeasured confounders/Ignorability/Exchangeability: $(Y(0), Y(1)) \perp Z|\boldsymbol{X}$

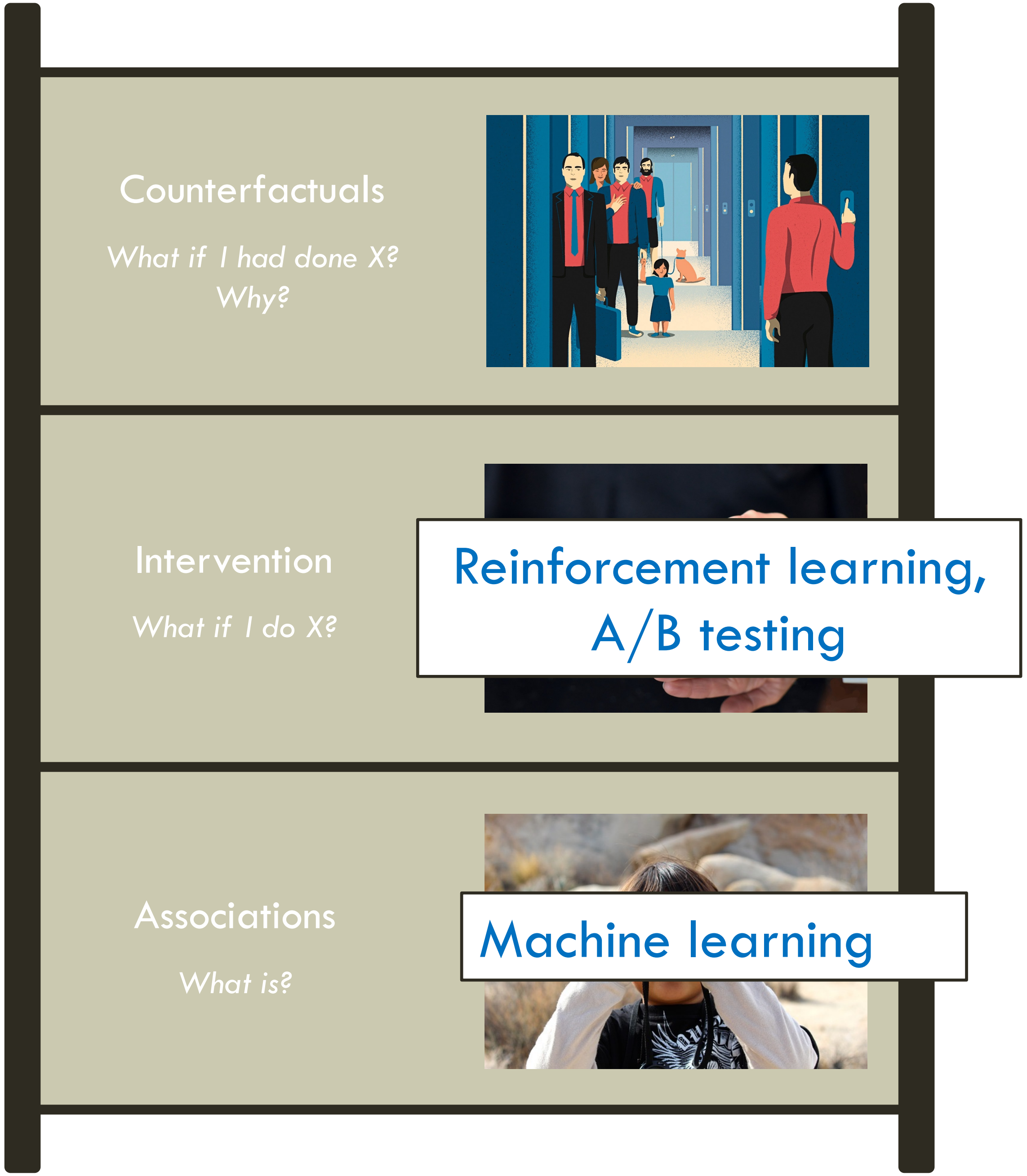
~~Stable unit treatment value assumption (SUTVA): $Y_i(\mathbf{z}) = Y_i(z_i)$, the outcome of unit i depends only on the treatment it receives and not on the treatment other units receive~~

- ~~This is violated in the presence of interference~~

Interference assumption: $Y_i(\mathbf{z}) = Y_i(z_i;\mathbf{z}_{Ni})$, a unit's response can be affected by the treatment it receives and by the treatments received by its neighbors/peers

- E.g., whether someone gets sick depends on the vaccination status of peers





LADDER OF CAUSATION*

Associations: $P(y|z)$ [Level 1]

- Example question: Is working in academia (z) correlated with happiness (y)?

Interventions: $P(y|do(z), x)$ [Level 2]

- Example: If Alice takes a job in industry, would she be happier than taking one in academia?
- Treatment z, outcome y, context x

Counterfactuals: $P(y_z|z',y')$ [Level 3]

- Example: If Alice stayed in industry (z), would Alice have been happier, given that she took a job in academia (z')?

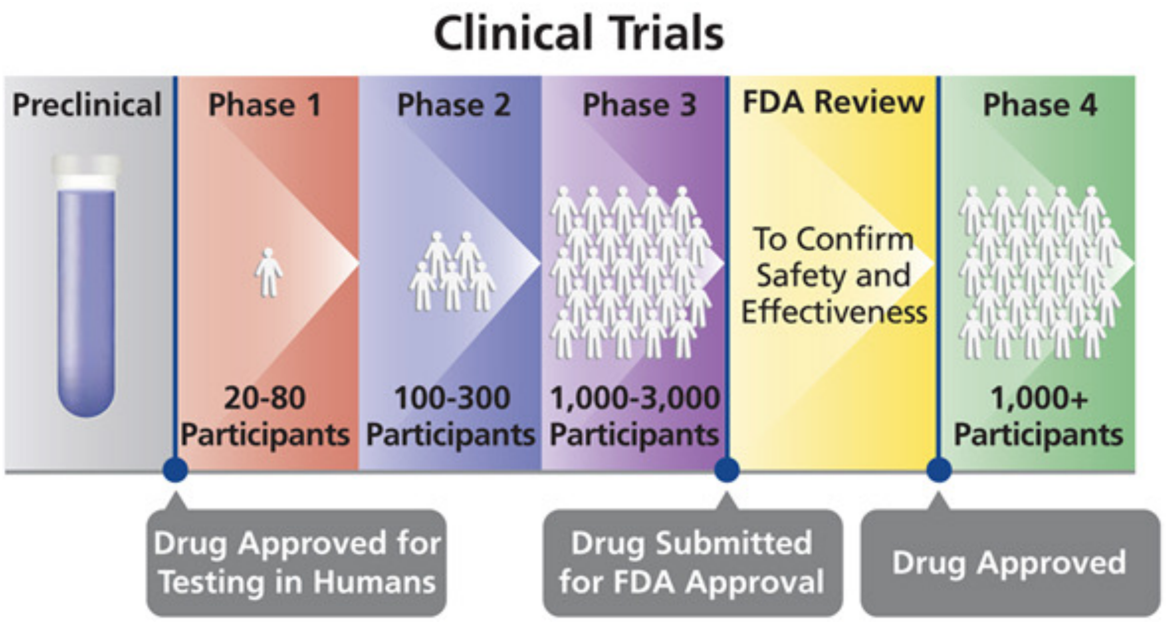
Counterfactual queries require different tools from associational ones!

Questions from level j can be answered if you have information from a higher level but not the other way around

*J. Pearl. *The seven tools of causal inference, with reflections on machine learning*. Communications of the ACM 2019.

INTERVENTIONS

- Randomized controlled trials required for drug approval by FDA
  - A random group given the drug is compared to a random group given the placebo

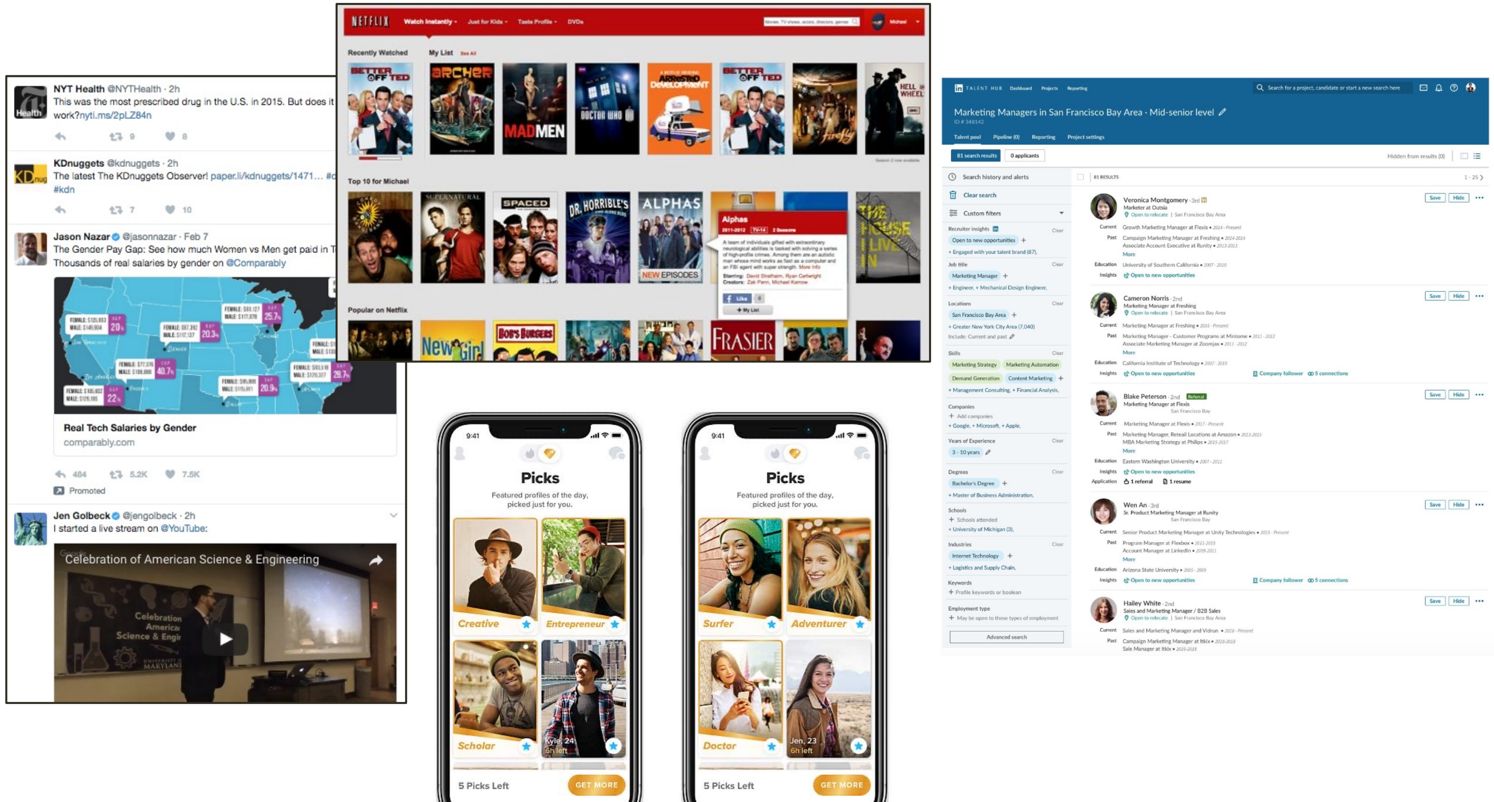


https://www.fda.gov/drugs/development-approval-process-drugs

WHICH RECOMMENDATION ALGORITHM IS BETTER?

A/B testing = controlled experiment = randomized controlled trials

- Best scientific design for establishing causality between a change and user behavior
- Is the outcome better on average for people “treated with” version A or version B?



$$ATE = E[Y(Z_1)] - E[Y(Z_0)]$$

INTERVENTIONS NOT ALWAYS POSSIBLE

Ethical concerns

Too expensive

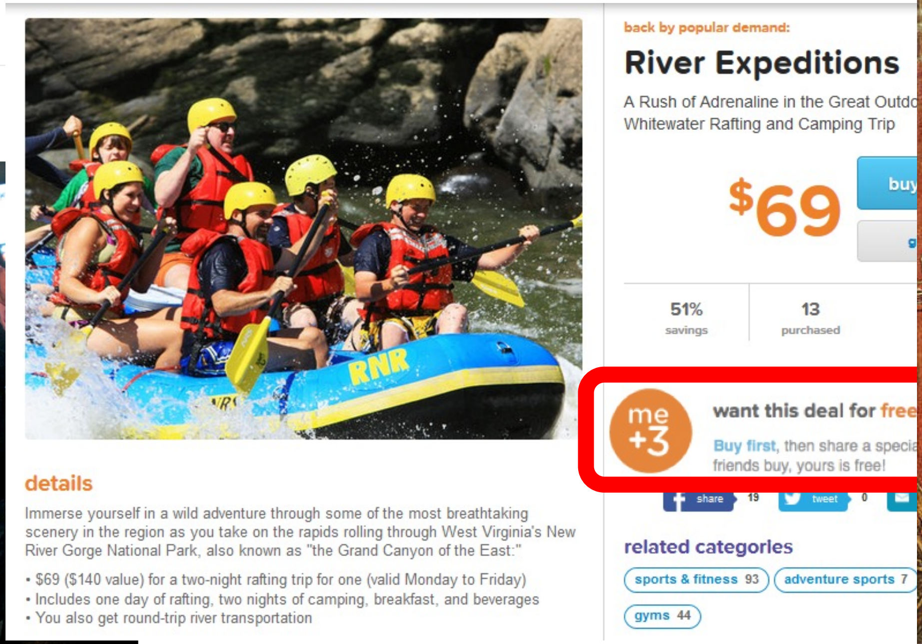
Immutable characteristics

The New York Times

OKCupid Plays With Love in User Experiments

137

Mingling at an event in Manhattan sponsored by OKCupid, which on Monday published the results of three experiments. Yana Paskova for The New York Times



STRUCTURAL CAUSAL MODELS (SCM)

SCM describes how nature assigns values to variables of interest

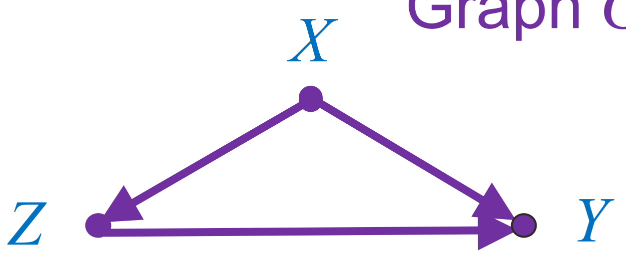
- Variables: U (exogenous) and V (endogenous)
- Functions: assign each variable in V a value based on other variables
  - Direct cause: X is direct cause of Y if X is in the function assigning Y
  - Cause: X is a cause of Y if it is a direct cause of Y or of any cause of Y
- Graphical causal model: nodes represent variables, edges represent functional dependences
  - Also referred to as graph or graphical model or causal diagram
  - Allows us to reason about exchangeability through d-separation

Do-calculus: Provides rules for estimating causal effects from observational data when identification possible, given an SCM

- Works even when some variables are latent

$X=f_X(U_X)$; $Z=f_Z(X,U_Z)$; $Y=f_Y(X,Z,U_Y)$

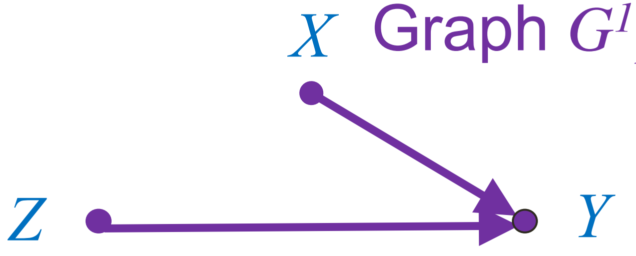
Graph G^1



$$P(Y=y|do(Z=z))=?$$

Causal model under intervention

Graph $G^1{}_M$



Pearl. *Causality: Models, Reasoning and Inference*. 2009.

BACKDOOR CRITERION

A common rule for deriving a valid causal estimand and an estimate from observational data whenever possible

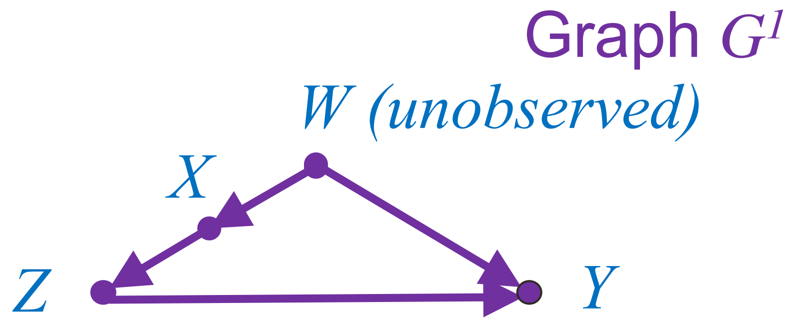
Given an ordered pair of variables (Z, Y) in a directed acyclic graph G, a set of variables X satisfies the backdoor criterion relative to (Z, Y) if no node in X is a descendant of Z, and X blocks every path between Z and Y that contains an arrow into Z. (X d-separates Z and Y on these paths)

$$P(Y = y|do(Z = z) = \sum_x P(Y = y|Z = z, X = x)P(X = x)$$

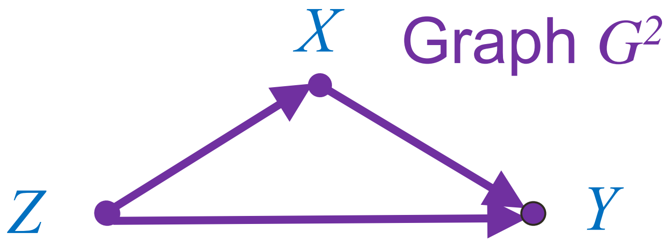
$$= \sum_x \frac{P(Y = y, Z = z, X = x)}{\boxed{P(Z = z|X = x)}}$$

propensity score

The adjustment formula is "controlling" for X



$$P(Y = y|do(Z = z)) = \sum_x P(Y = y|Z = z, X = x)P(X = x)$$

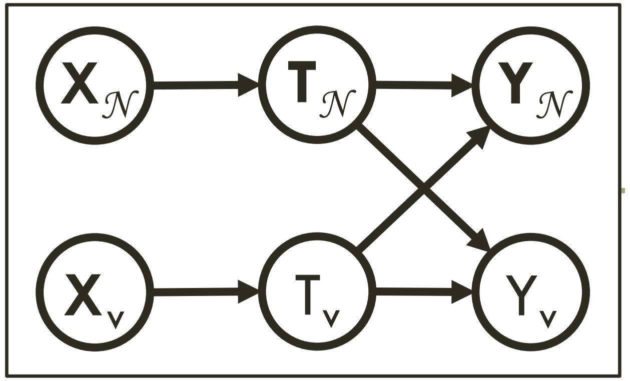


$$P(Y = y|do(Z = z)) = P(Y = y|Z = z)$$

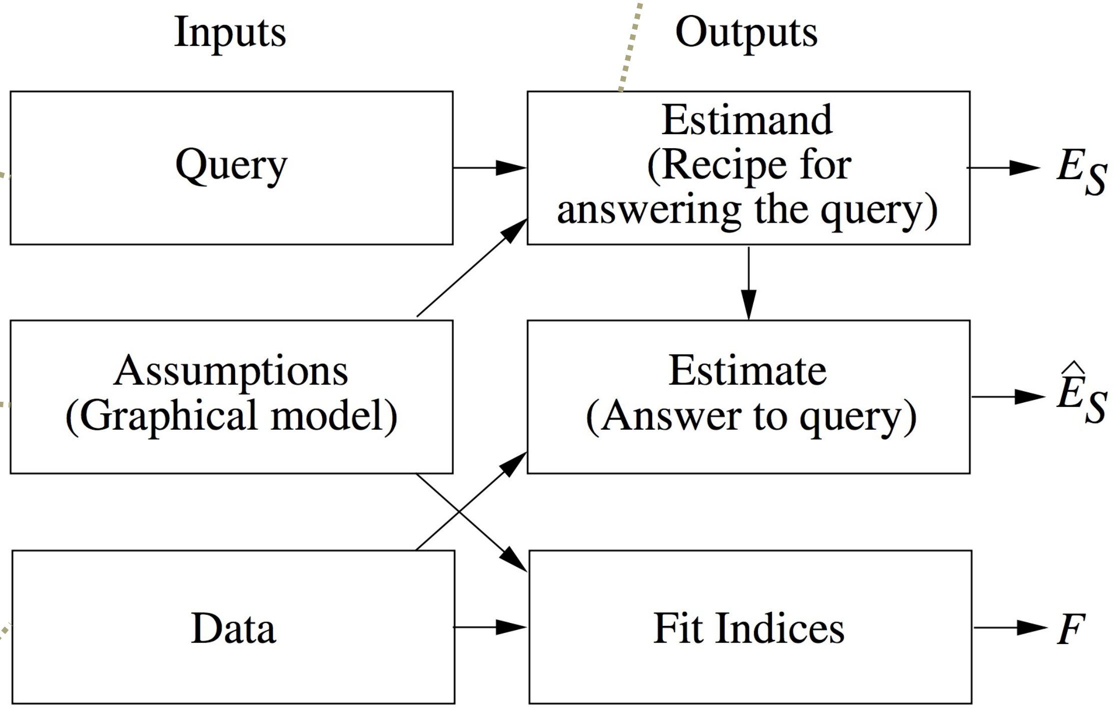
CAUSAL INFERENCE ENGINE

Based on do-calculus rules
e.g., $P(y_v|do(t_v)) = P(y_v|t_v)$

If Mia *read Flo's tweets*, would she have *vaccinated* herself?



Name	Age	Gender	Race	VaccineView	…	Vaccinated
Mia	50	F	Asian	?	…	No
Flo	34	F	Black	?	…	Yes
LotusOak	42	F	White	Yes	…	No



Pearl. *The Seven Tools of Causal Inference with Reflections on Machine Learning*. 2019.

- Motivation
- Causal inference 101
- **Causal effects in networks**
- Interventions and network experiment design
- Counterfactuals & causal effects in observational data
    - Representation, identification, estimation
        - Blocks
        - Representation challenges
        - Chain and segregated graphs
        - Multi-relational data and abstract ground graphs
    - Discovery

CAUSAL EFFECTS IN NETWORKS

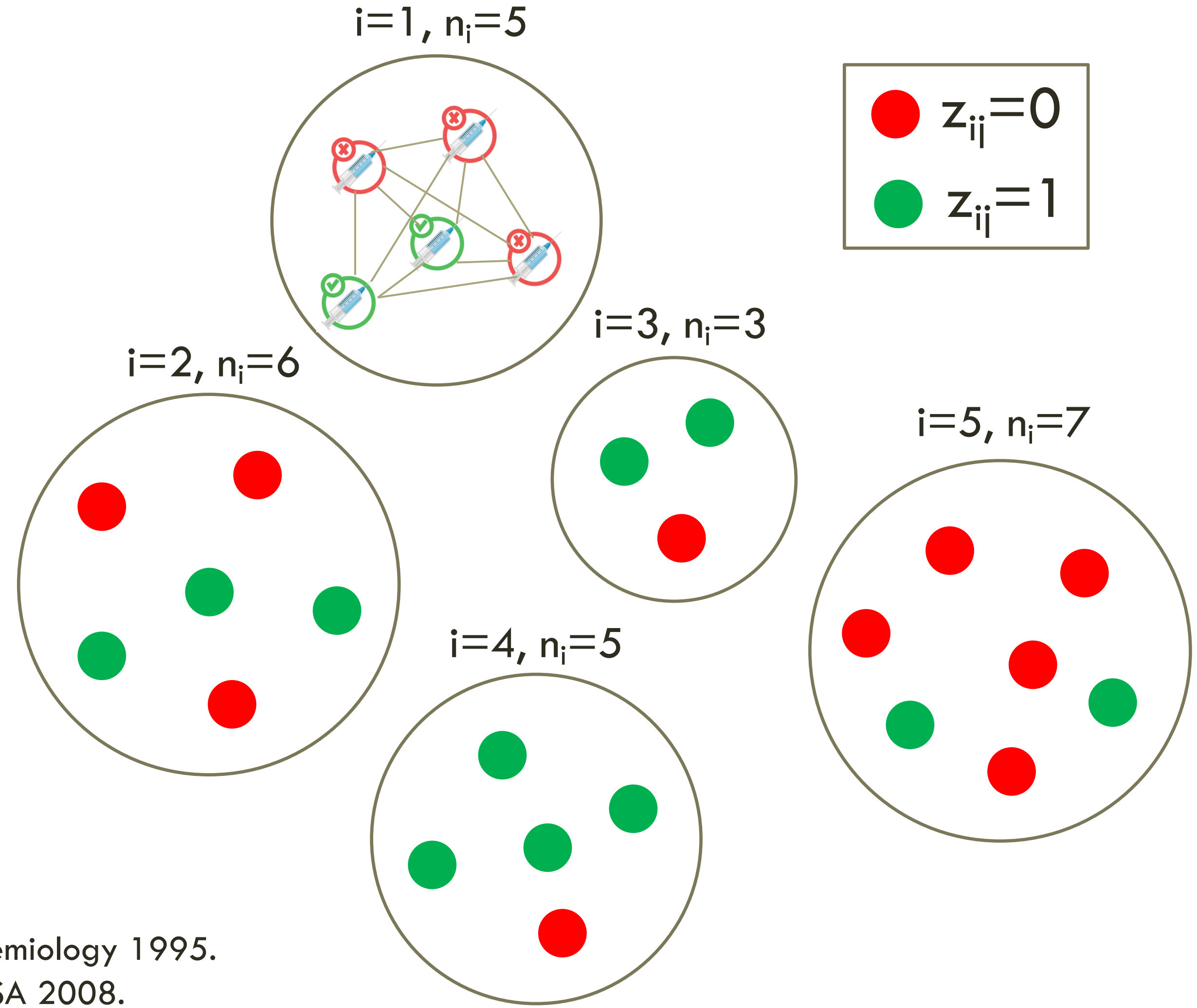
CAUSAL ESTIMANDS UNDER INTERFERENCE

Start with simplifying assumptions:

Multiple non-overlapping groups

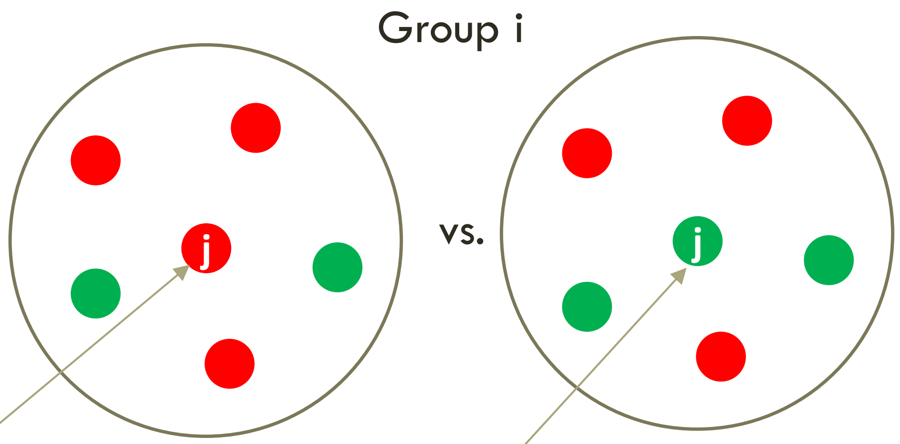
Partial interference: interference occurs within but not across groups

Treatment assignment within each group has treatment regime $P(Z=1)= \psi$



Halloran, Struchiner. *Causal inference in infectious diseases*. Epidemiology 1995.
Hudgens, Halloran. *Toward causal inference with interference*. JASA 2008.

DIRECT CAUSAL EFFECT



Individual Direct Causal Effect (DCE): the difference in outcome due to the treatment alone

- *e.g., effect of getting vaccinated on getting sick*

$$CE^{D}_{ij}\left(\mathbf{z}_{i(j)}\right) \equiv Y_{ij}\left(\mathbf{z}_{i(j)}, z_{ij}=0\right) - Y_{ij}\left(\mathbf{z}_{i(j)}, z_{ij}=1\right)$$

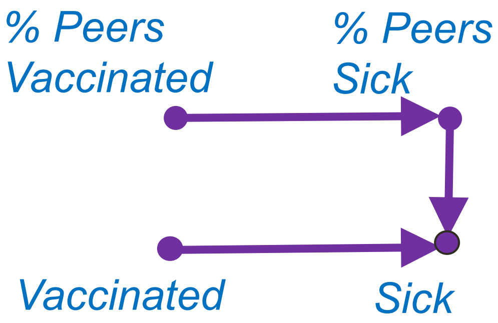
$\mathbf{z}_{i(j)}$: treatment assignment of units in j's group i

z_{ij}: treatment assignment of unit j in group i

Individual Avg. DCE: difference of expected values of the marginal distributions under treatment regime ψ of group i $\overline{CE}^{D}_{ij}(\psi) \equiv \overline{Y}_{ij}(0;\psi) - \overline{Y}_{ij}(1;\psi)$

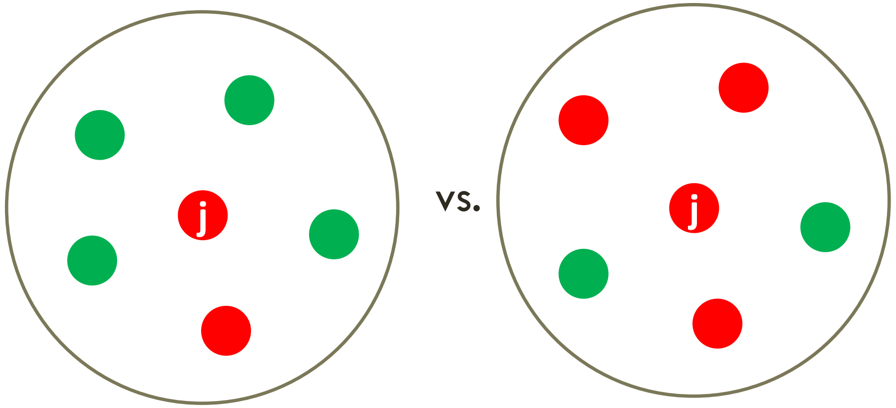
Group Avg. DCE: $\overline{CE}^{D}_{i}(\psi) \equiv \overline{Y}_{i}(0;\psi) - \overline{Y}_{i}(1;\psi) = \sum_{j=1}^{n_i} \overline{CE}^{D}_{ij}(\psi)/n_i$

Population Avg. DCE: $\overline{CE}^{D}(\psi) \equiv \overline{Y}(0;\psi) - \overline{Y}(1;\psi) = \sum_{i=1}^{N} \overline{CE}^{D}_{i}(\psi)/N$



Halloran, Struchiner. *Causal inference in infectious diseases*. Epidemiology 1995.
Hudgens, Halloran. *Toward causal inference with interference*. JASA 2008.

INDIRECT/PEER EFFECT



Individual indirect causal effect (ICE): the effect of the treatment received by others in the group on an individual outcome

- e.g., *effect of % vaccinated people on getting sick*

$\mathbf{z}_{i(j)}$: treatment assignment of unit i's neighbors (group j)

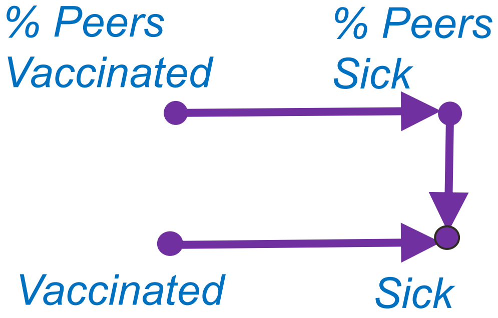
z_{ij}: treatment assignment of unit i

$$CE^{I}_{ij}\left(\mathbf{z}_{i(j)}, \mathbf{z}'_{i(j)}\right) \equiv Y_i\left(\mathbf{z}_{i(j)}, z_{ij} = 0\right) - Y_i\left(\mathbf{z}'_{i(j)}, z'_{ij} = 0\right)$$

Individual Avg. ICE: difference of expected values of the marginal distributions under two different treatment regimes ψ and ϕ of group i $\overline{CE}^{I}_{ij}(\phi, \psi) \equiv \overline{Y}_{ij}(0; \phi) - \overline{Y}_{ij}(0; \psi)$

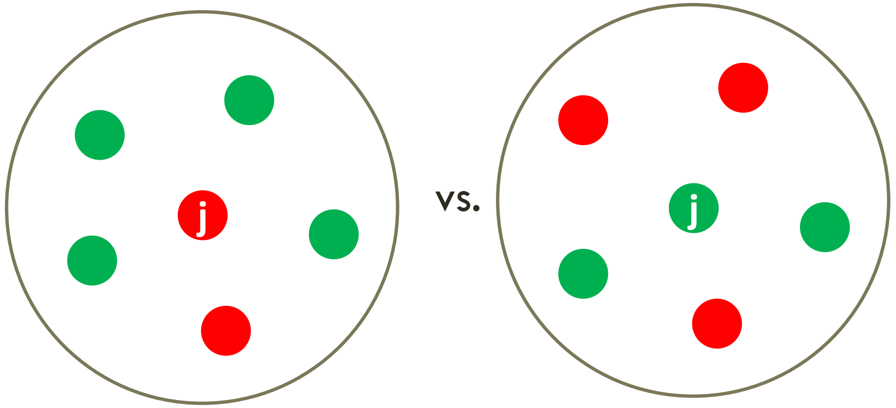
Group Avg. ICE: $\overline{CE}^{I}_{i}(\phi, \psi) \equiv \overline{Y}_{i}(0; \phi) - \overline{Y}_{i}(0; \psi) = \sum_{j=1}^{n_i} \overline{CE}^{I}_{ij}(\phi, \psi)/n_i$

Population Avg. ICE: $\overline{CE}^{I}(\phi, \psi) \equiv \overline{Y}(0; \phi) - \overline{Y}(0; \psi) = \sum_{i=1}^{N} \overline{CE}^{I}_{i}(\phi, \psi)/N$



Halloran, Struchiner. *Causal inference in infectious diseases.* Epidemiology 1995.
Hudgens, Halloran. *Toward causal inference with interference.* JASA 2008.

TOTAL EFFECT



Individual total causal effect (TCE): both direct and indirect effect of treatment assignment

- e.g., *effect of % vaccinated people and getting vaccinated on getting sick*

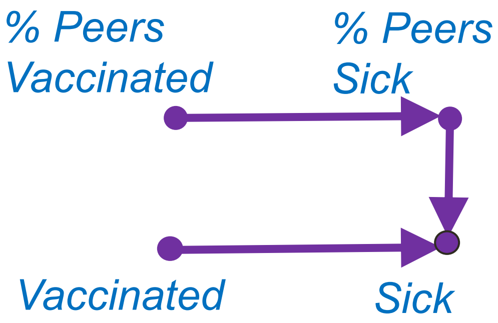
$$CE_{ij}^{T}\left(\mathbf{z}_{i(j)}, \mathbf{z}'_{i(j)}\right) \equiv Y_{ij}\left(\mathbf{z}_{i(j)}, z_{ij}=0\right) - Y_{ij}\left(\mathbf{z}'_{i(j)}, z'_{ij}=1\right)$$

Individual Avg. TCE: difference of expected values of the marginal distributions under two different treatment regimes $0;\psi$ and $1;\phi$ of group i $\overline{CE}_{ij}^{T}(\phi,\psi) \equiv \overline{Y}_{ij}(0;\phi) - \overline{Y}_{ij}(1;\psi)$

Group Avg. TCE: $\overline{CE}_{i}^{T}(\phi,\psi) \equiv \overline{Y}_{i}(0;\phi) - \overline{Y}_{i}(1;\psi) = \sum_{j=1}^{n_i} \overline{CE}_{ij}^{T}(\phi,\psi)/n_i$

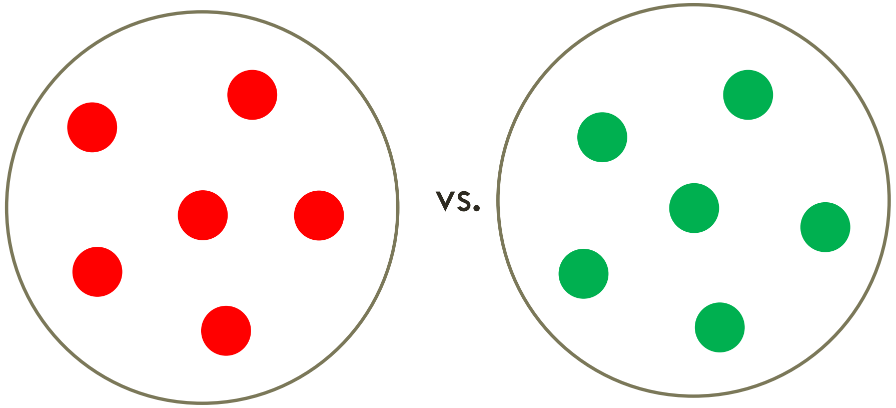
Population Avg. ICE:

$$\overline{CE}^{T}(\phi,\psi) \equiv \overline{Y}(0;\phi) - \overline{Y}(1;\psi) = \sum_{i=1}^{N} \overline{CE}_{i}^{T}(\phi,\psi)/N$$



Halloran, Struchiner. *Causal inference in infectious diseases*. Epidemiology 1995.
Hudgens, Halloran. *Toward causal inference with interference*. JASA 2008.

TOTAL EFFECT: ALTERNATIVE ESTIMAND

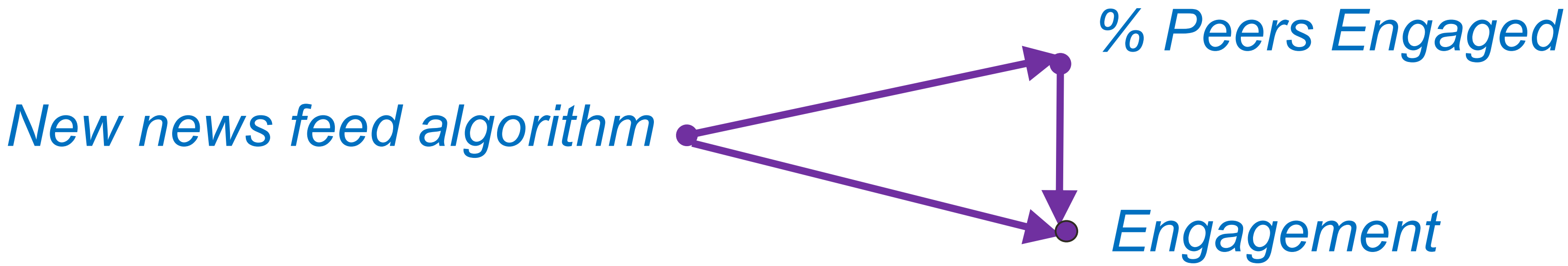


Total treatment effect (TTE): both direct and indirect effect of treatment assignment

- e.g., *effect of vaccinating everyone*

$$TTE = \frac{1}{N} \sum_{v_i \in V} (v_i.Y(\mathbf{Z_1}) - v_i.Y(\mathbf{Z_0}))$$

Applications: recommender systems



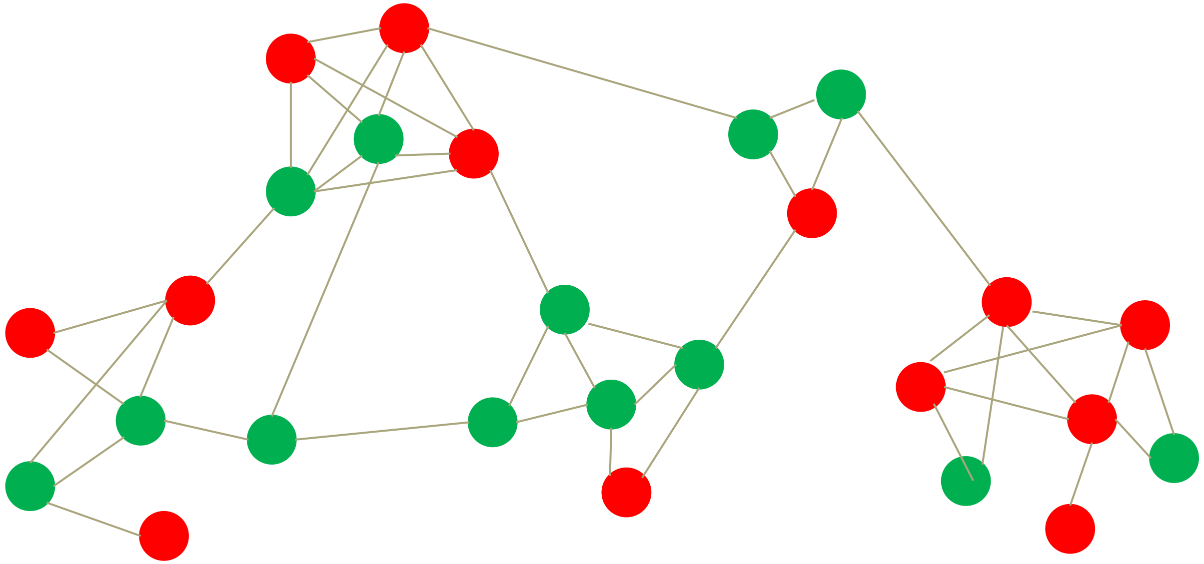
Ugander, Karrer, Backstrom, Kleinberg. *Graph cluster randomization: Network exposure to multiple universes.* KDD 2013.

INTERVENTIONS AND NETWORK EXPERIMENT DESIGN

RANDOMIZATION IN NETWORKS

Network experiment design:

Design for randomized controlled trials that take into consideration interactions and potential interference between units of interest



Randomization at the node level

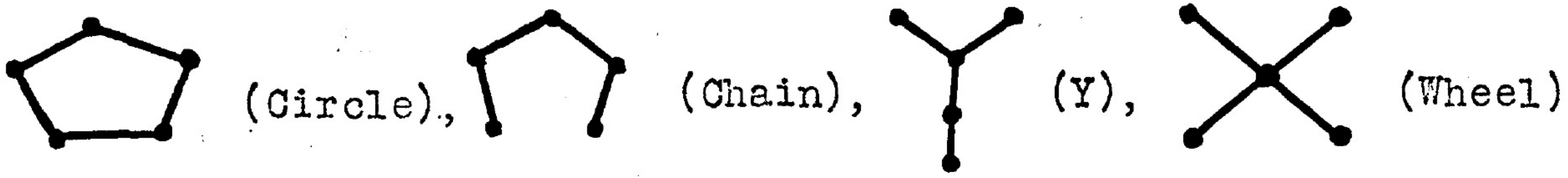
- High variance of estimators
- Need additional assumptions

The choice of randomization design depends on the causal effect of interest!

NETWORK EXPERIMENT DESIGN

Early network experiments in 1940s were performed in labs at a small scale

Leavitt: solve a data collation task using only one of four randomly assigned communication patterns

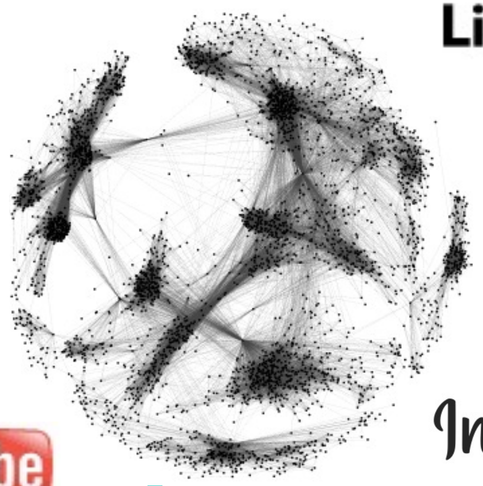


“The Circle was erratic, active (message-wise), unorganized, and leaderless, but satisfying to its members. The Wheel was less erratic, required few messages, was well organized, and had a definite leader, but was less satisfying to most of its members”

H. Leavitt. *Some effects of certain communication patterns on group performance*. The Journal of Abnormal and Social Psychology, 46(1): 38. 1951.

NETWORK EXPERIMENT DESIGN

Network experiments nowadays are often large-scale and use digital platforms with millions of users



Design?

Can peers influence voter turnout? [Bond et al. 2012]

Can product endorsements from friends increase ad clicks? [Bakshy et al. 2012]

Can emotional states be transferred via contagion? [Kramer et al 2014]

S. Aral. *Networked Experiments*. The Oxford Handbook of the Economics of Networks. 2016.

TWO-STAGE RANDOMIZATION DESIGN UNDER PARTIAL INTERFERENCE

Two-stage randomization

1. Assign groups to treatment and control with prob. ν
2. For each group i:

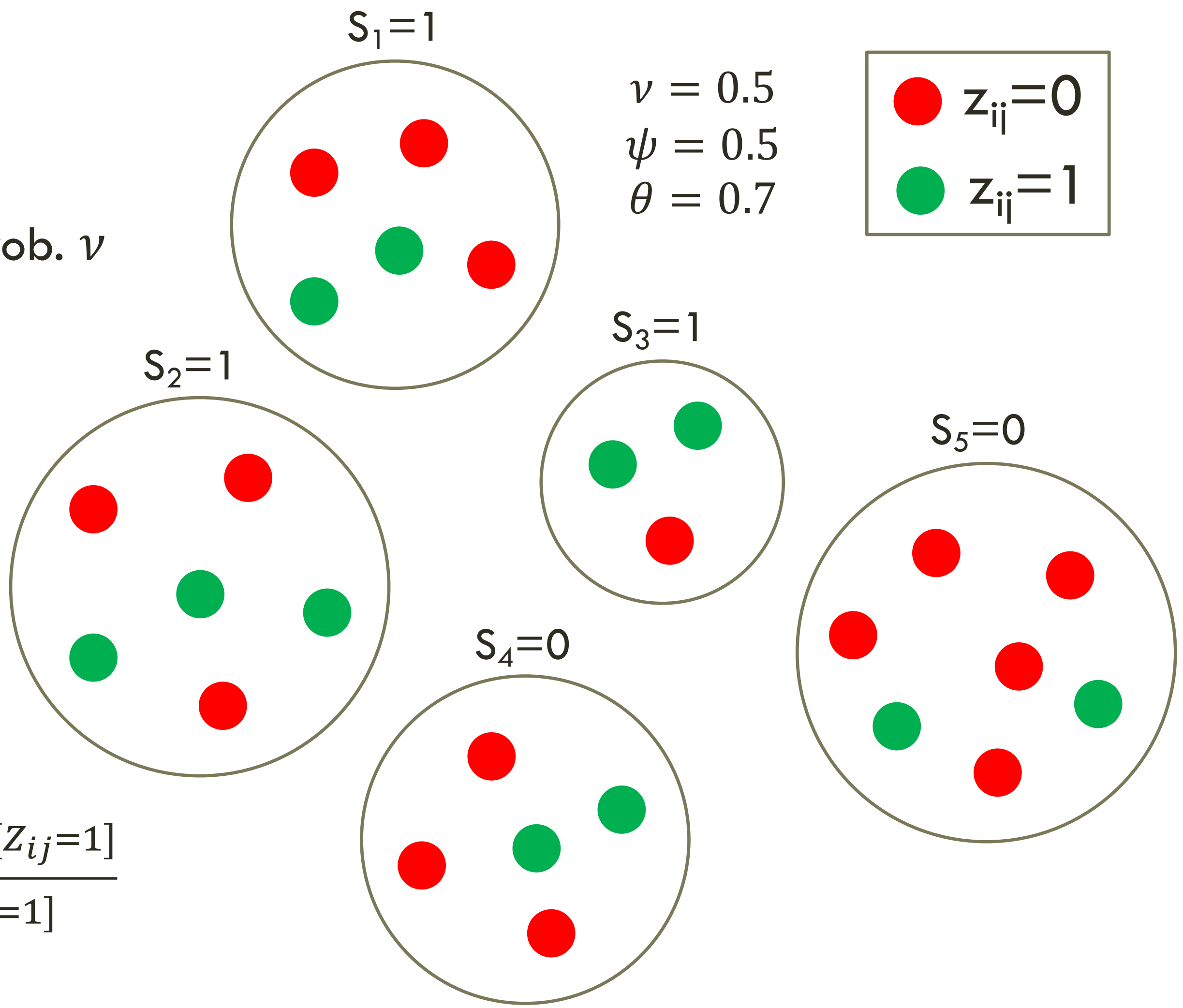
   If group in treatment (S_i=1), assign each unit to treatment with probability ψ

   Else group in control (S_i=0), assign each unit to treatment with probability θ

   E.g., Group Average Direct Causal Effect estimator

Estimand $$\overline{CE}_i^D(\psi) = \frac{1}{n_i}\sum_{j=1}^{n_i}\left(\bar{Y}_{ij}(0,\psi) - \bar{Y}_{ij}(1,\psi)\right)$$

Unbiased estimator $$\widehat{CE}_i^D(\psi) = \frac{\sum_{j=1}^{n_i} Y_{ij}(\mathbf{Z}_i) I[Z_{ij}=0]}{\sum_{j=1}^{n_i} I[Z_{ij}=0]} - \frac{\sum_{j=1}^{n_i} Y_{ij}(\mathbf{Z}_i) I[Z_{ij}=1]}{\sum_{j=1}^{n_i} I[Z_{ij}=1]}$$



Hudgens, Halloran. *Toward causal inference with interference.* JASA 2008.

INSULATED NEIGHBOR RANDOMIZATION DESIGN FOR K-LEVEL PEER EFFECT ESTIMATION

A potential outcome is defined based on the treatment assignment of neighbors

K-level treatment: a node is *k-exposed to peer influence effects* if exactly k of its neighbors are treated

Estimand for k-level peer effects:

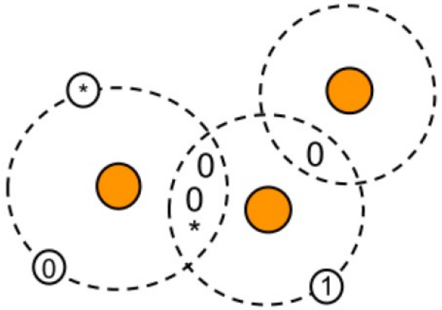
Outcome when k neighbors are treated but ego is not

Outcome when neither ego nor neighbors are treated

$$\delta_k \equiv \frac{1}{|V_k|} \sum_{i \in V_k} \left[\binom{n_i}{k}^{-1} \sum_{\mathbf{z} \in \mathbf{Z}(\mathcal{N}_i; k)} Y_i(0, \mathbf{z}) - Y_i(\mathbf{0}) \right]$$

V_k: nodes with $\geq$ k neighbors

possible combinations with exactly k treated neighbors



INR Design: nodes from V_k are sequentially assigned to either be k-exposed or 0-exposed

- Estimator bias depends on network topology and whether shared neighbors are as influential as non-shared ones

P. Toulis, E. Kao. *Estimation of Causal Peer Influence Effects*. ICML 2013.

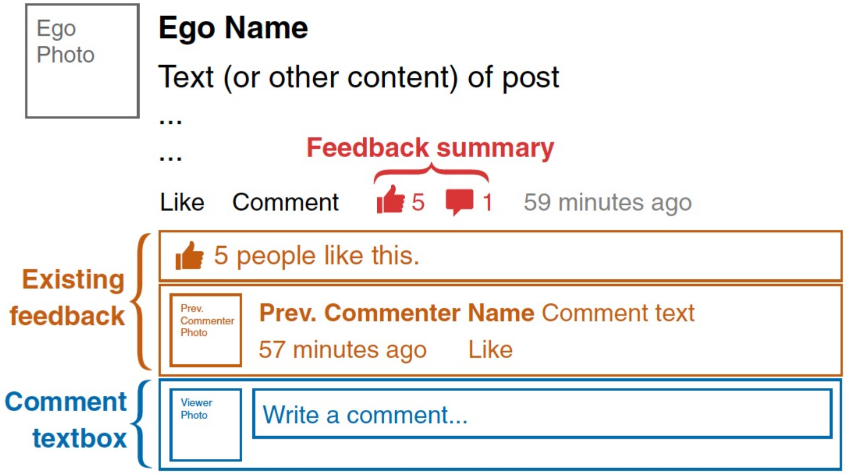
MECHANISM AND ENCOURAGEMENT DESIGNS FOR PEER EFFECT ESTIMATION

Randomizing peer behavior is not always realistic

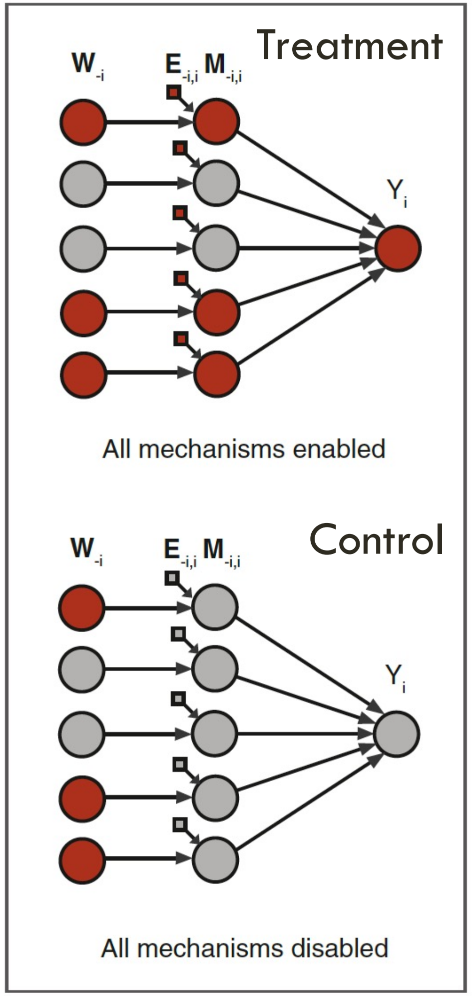
Mechanism designs: modulate the mechanism by which information about peer behavior is transmitted

Encouragement designs: measure peer effects of behaviors not directly controlled by the experimenter

Goal: Estimate effects of receiving feedback on how many posts egos make and how much feedback they give on others' posts



Mechanism design

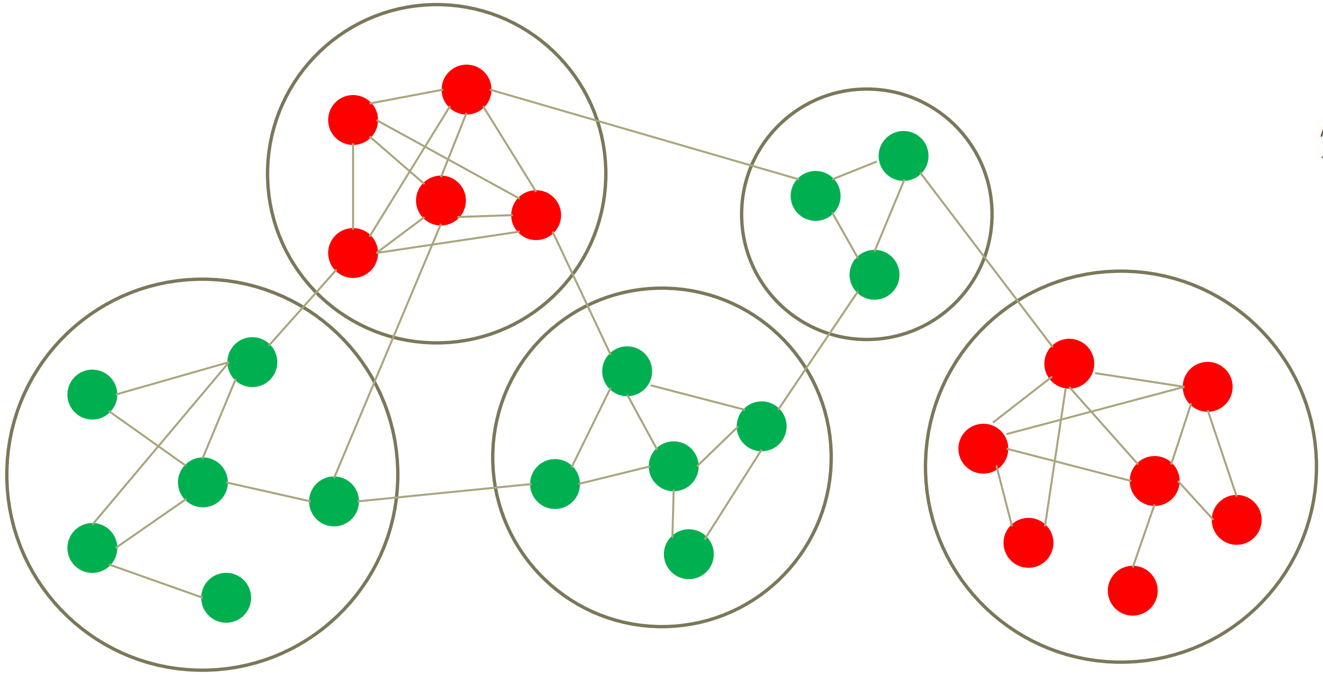


D. Eckles, R. Kizilcec, E. Bakshy. *Estimating peer effects in networks with peer encouragement designs.* PNAS 2016.

CLUSTER-BASED RANDOMIZATION DESIGNS FOR TOTAL TREATMENT EFFECT ESTIMATION

Design for estimating total treatment effect

- Assumes partial interference: interference can occur within clusters but not across clusters
- Minimizes spillover between treatment and control



Estimand of interest:

$$TTE = \frac{1}{N} \sum_{v_i \in V} (v_i.Y(\mathbf{Z_1}) - v_i.Y(\mathbf{Z_0}))$$

Horvitz-Thompson Estimator:

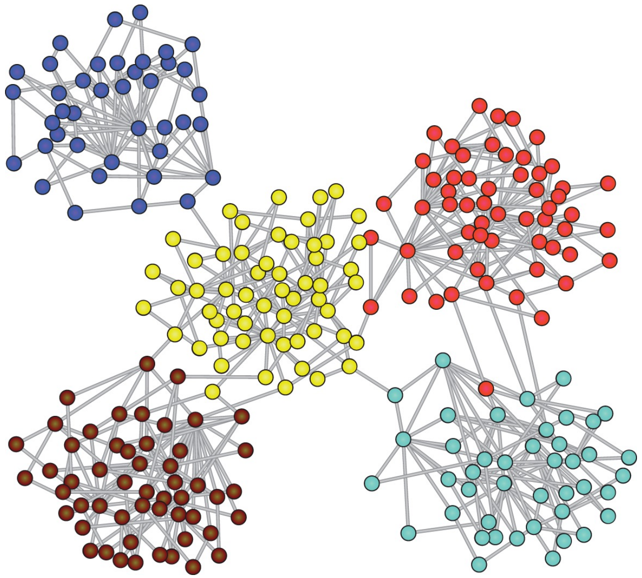
$$\hat{\tau}(Z) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{Y_i(Z)\mathbf{1}[Z \in \sigma_i^1]}{\Pr(Z \in \sigma_i^1)} - \frac{Y_i(Z)\mathbf{1}[Z \in \sigma_i^0]}{\Pr(Z \in \sigma_i^0)} \right)$$

Ugander, Karrer, Backstrom, Kleinberg. *Graph cluster randomization: Network exposure to multiple universes.* KDD 2013.

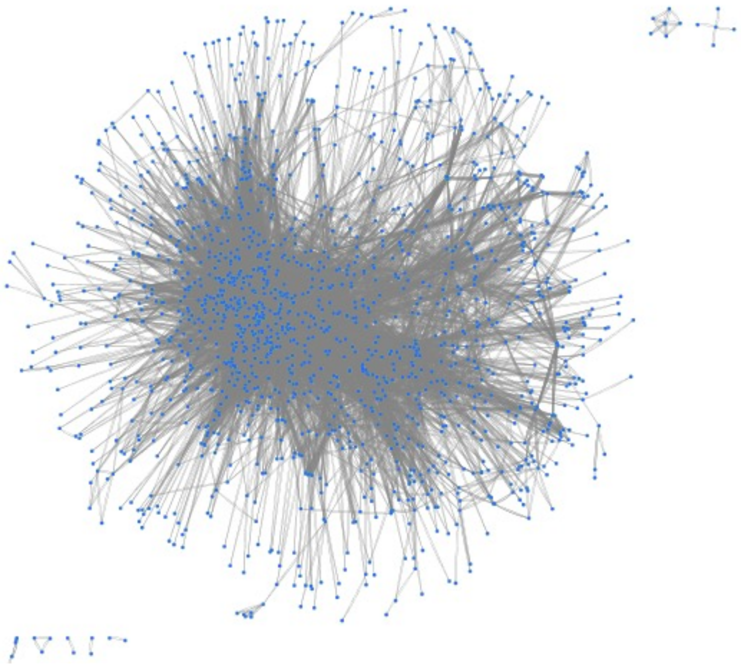
CHALLENGES WITH CLUSTER-BASED RANDOMIZATION

Challenge 1*: It can be hard to separate a real-world network into treatment and control clusters without leaving a lot of edges across

- E.g., LinkedIn graph clustering has 65-79% of inter-cluster edges**



Ideal network



Online social networks

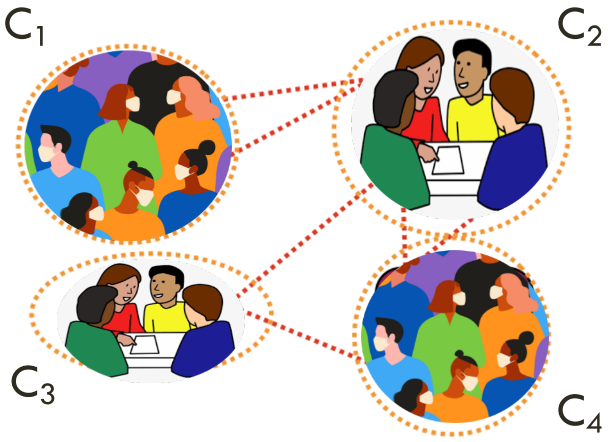
*Z. Fatemi, E. Zheleva. *Minimizing interference and selection bias in network experiment design*. ICWSM 2020.

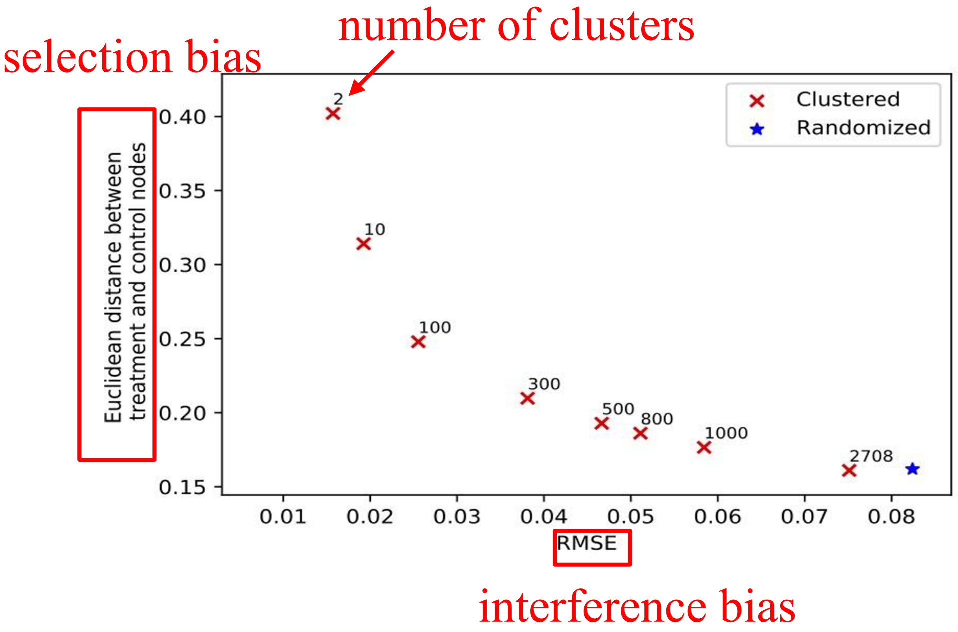
**Saveski, Pouget-Abadie, Saint-Jacques, Duan, Ghosh, Xu, Airoldi. *Detecting network effects: Randomizing over randomized experiments*. KDD 2017.

CHALLENGES WITH CLUSTER-BASED RANDOMIZATION

Challenge 2: **Treatment and control clusters can have different covariate distributions**

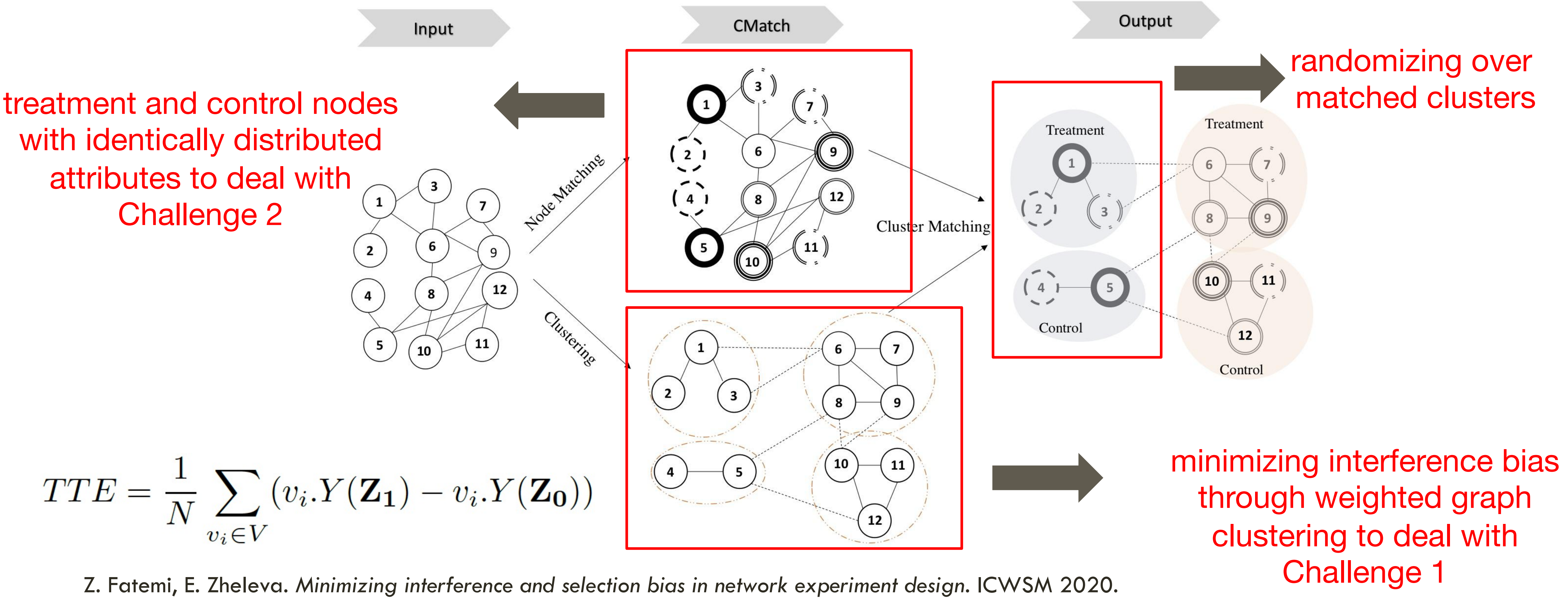
- Tradeoff between interference and selection bias based on number of clusters





Z. Fatemi, E. Zheleva. *Minimizing interference and selection bias in network experiment design.* ICWSM 2020.

CMATCH: CLUSTER-BASED RANDOMIZATION WITH CLUSTER MATCHING ON A WEIGHTED GRAPH



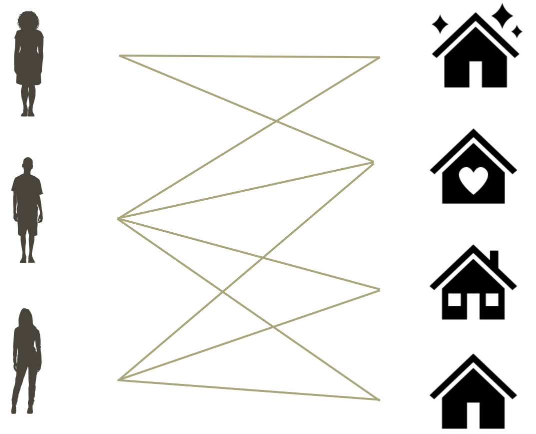
$$TTE = \frac{1}{N} \sum_{v_i \in V} (v_i.Y(\mathbf{Z_1}) - v_i.Y(\mathbf{Z_0}))$$

Z. Fatemi, E. Zheleva. *Minimizing interference and selection bias in network experiment design.* ICWSM 2020.
Stuart. *Matching methods for causal inference: a review and look forward.* Stat. Science 2010.

TWO-SIDED RANDOMIZATION FOR BIPARTITE GRAPH EXPERIMENTS

Two-sided markets

Customers Listings



Lower bias than customer randomization or listing randomization alone

Bias goes to zero as relative demand goes to zero or infinity

	Z_L=0	Z_L=1
Z_C=0	Control 1	Control 2
Z_C=1	Control 3	Treatment

Interference due to competition:

- Making one listing more attractive makes others less attractive
- Making one customer more likely to book reduces supply for other customers

R. Johari, H. Li, I. Liskovic, G. Weintraub. *Experimental design in two-sided platforms: An analysis of bias.* Arxiv 2020.

P. Bajari, B. Burdick, G. Imbens, J. McQueen, T. Richardson, I. Rosen. *Multiple randomization designs for interference.* ASSA Annual Meeting 2020.

COUNTERFACTUALS & CAUSAL EFFECTS IN OBSERVATIONAL NETWORK DATA

Blocks

REPRESENTATION: GRAPHICAL MODELS

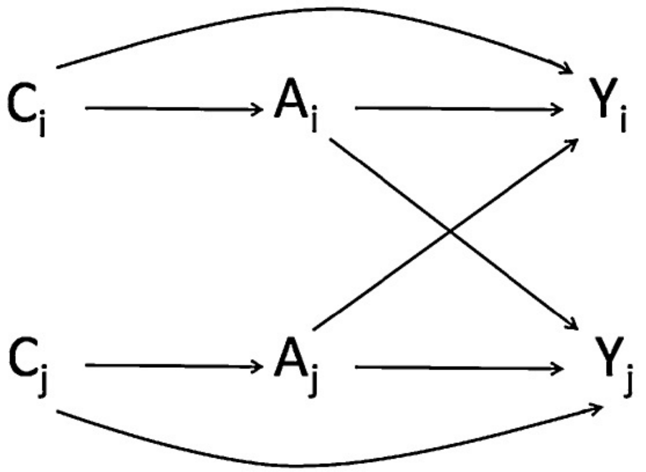
Blocks

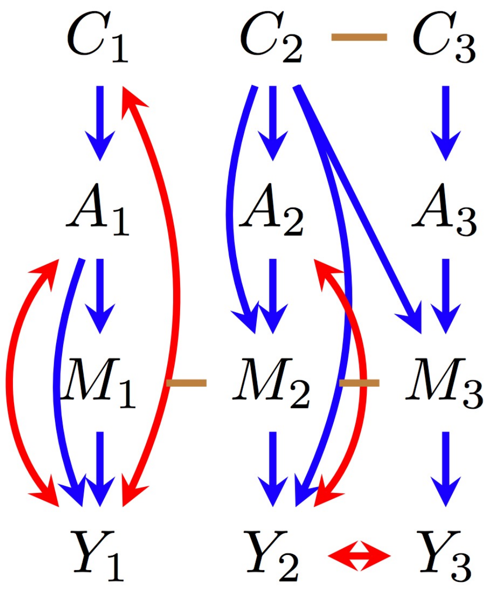
Chain and segregated graphs

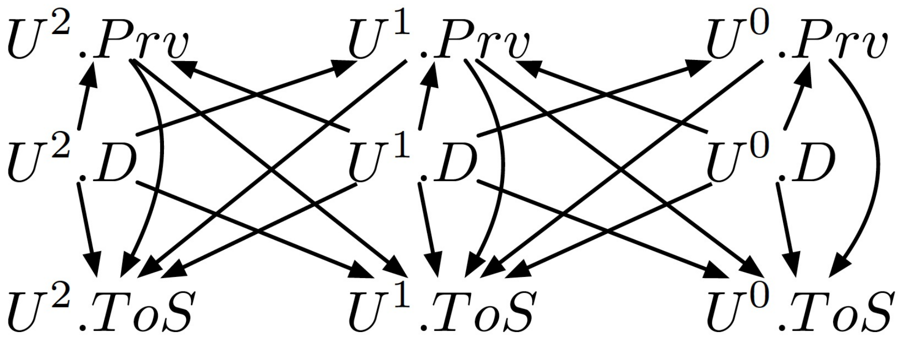
Abstract ground graphs

Assume partial interference

Can model more complex interference





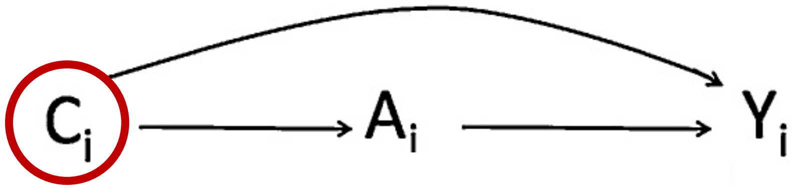


C-covariates A-treatment Y-outcome

BLOCKS FOR DIRECT INTERFERENCE

Blocks: repeatable patterns of interference

Direct interference: treatments of peers/neighbors affect ego's outcome



Exchangeability holds and the effect of **A** on Y_i is identifiable:
C_i blocks the backdoor paths* from A_i to Y_i and from A_j to Y_i

$$P(Y_i = y | do(A_i = a_i, A_j = a_j)) =$$
$$\sum_{c_i} P(Y_i = y | A_i = a_i, A_j = a_j, C_i = c_i) P(C_i = c_i)$$

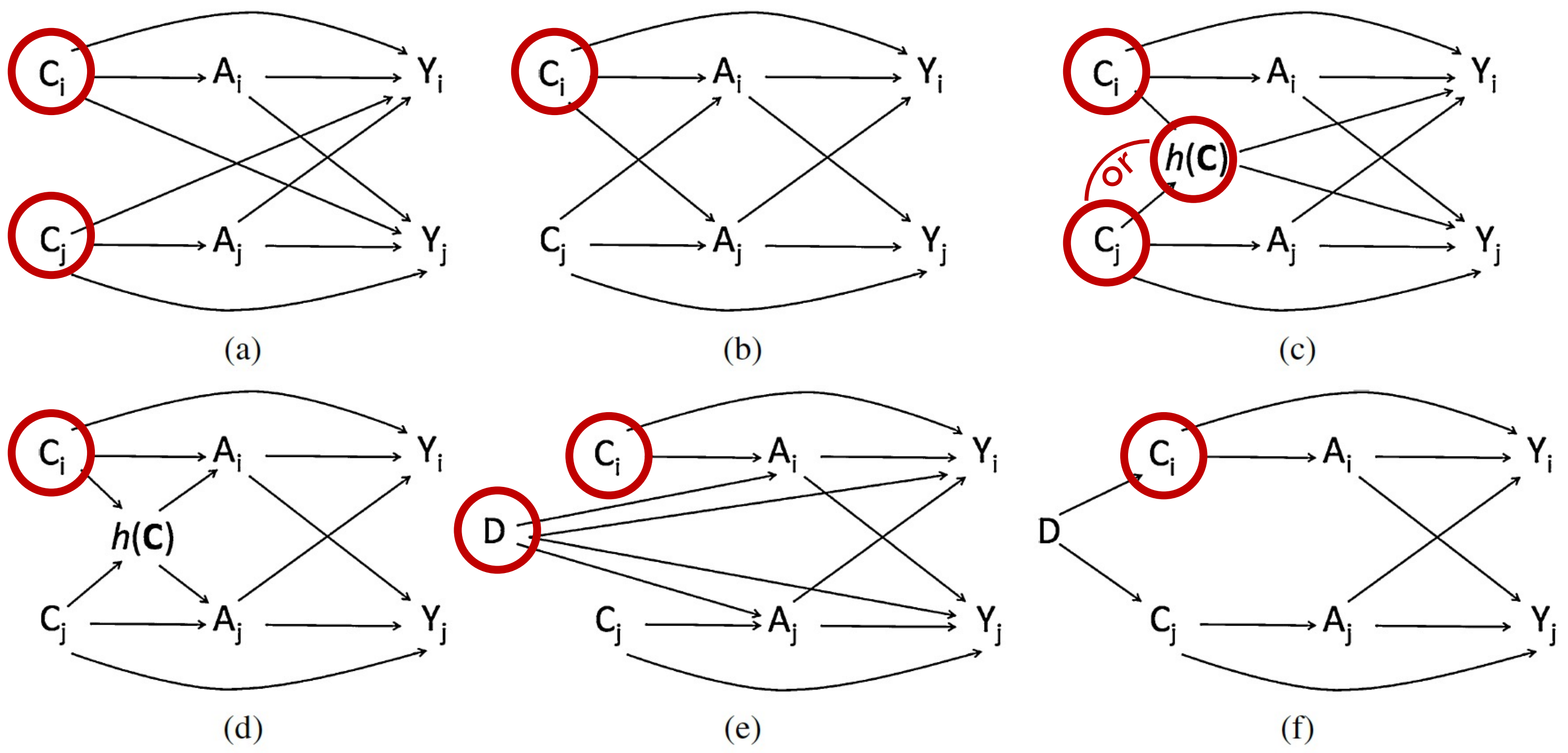
*A set of variables C satisfies the backdoor criterion relative to (A, Y) if no node in C is a descendant of A, and C blocks every path between A and Y that contains an arrow into A

C-covariates A-treatment Y-outcome

Ogburn, VanderWeele. *Causal Diagrams for Interference*. Statistical Science 2014.

BLOCKS FOR DIRECT INTERFERENCE

Identification of $E[Y_i|do(\boldsymbol{A}=\boldsymbol{a}_1)] - E[Y_i|do(\boldsymbol{A}=\boldsymbol{a}_2)]$ depends on the causal graph (domain knowledge) and which variables are available in the data

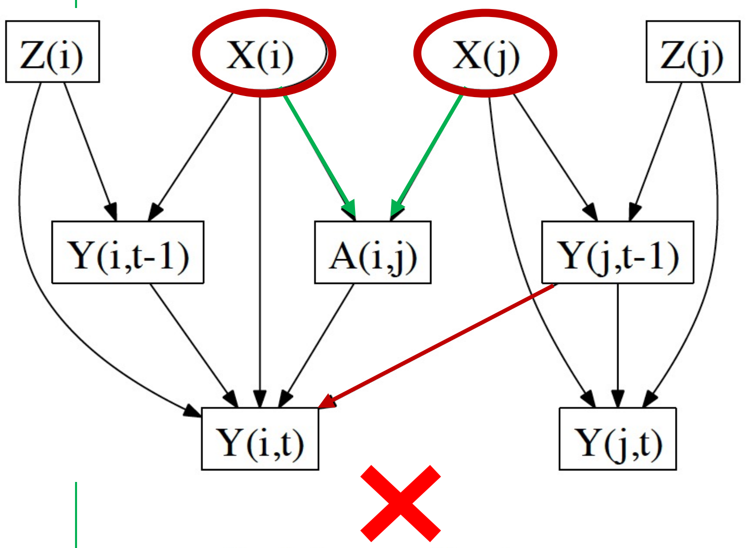


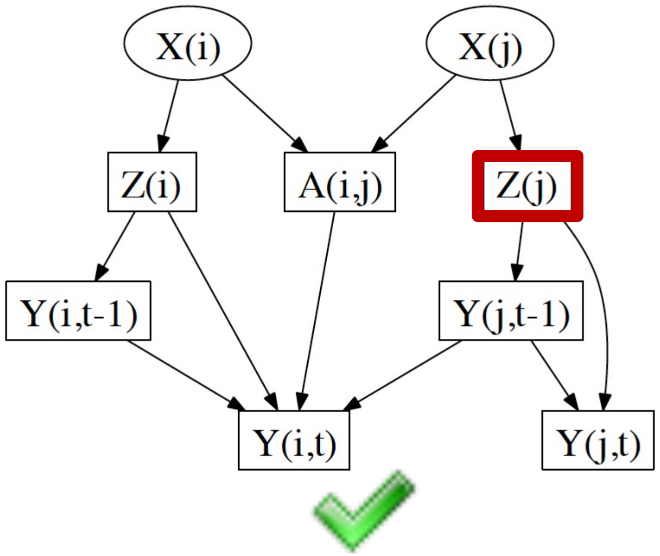
*A set of variables C satisfies the backdoor criterion relative to (A, Y) if no node in C is a descendant of A, and C blocks every path between A and Y that contains an arrow into A

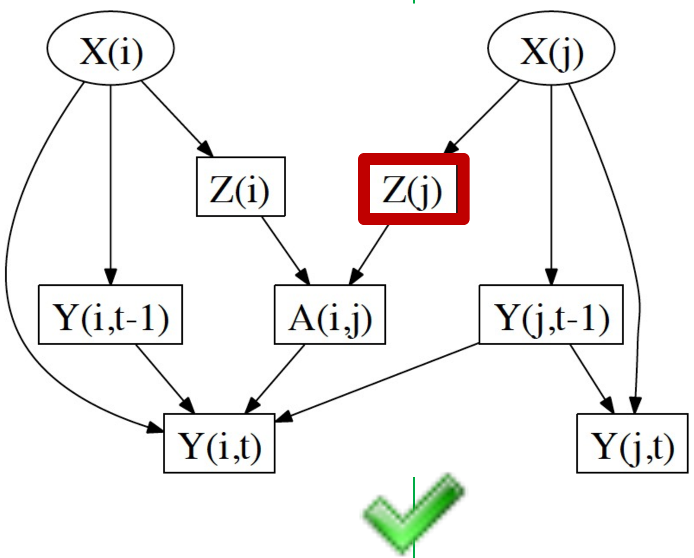
Ogburn, VanderWeele. *Causal Diagrams for Interference*. Statistical Science 2014.

IDENTIFYING CONTAGION

Contagion $E[Y_{i,t}|do(Y_{j,t-1} = y_1)] - E[Y_{i,t}|do(Y_{j,t-1} = y_1)]$ may not be identifiable due to latent homophily







Symbol	Meaning
i, j	Individuals
Z	Observed Traits
X	Latent Traits
Y	Observed Outcomes
A	Network Tie

Shalizi & Thomas. *Homophily and Contagion Are Generically Confounded in Observational Social Network Studies.* Sociological Methods & Research 2011.

EXAMPLE: LINEAR THRESHOLD MODEL

Linear threshold model (LTM)

- Model of information diffusion in social networks
- If a **proportion** of an individual's friends that have **activated** (e.g., adopted a product) are above a **threshold θ**, then that individual will **activate**

Heterogeneous peer effects

$$\rho(\mathbf{x}) = E[Y(I^t = i^t) - Y(I^t = i^{t'}) \mid \mathbf{X} = \mathbf{x}, Z = z]$$

- Identifiable in LTM according to SCM

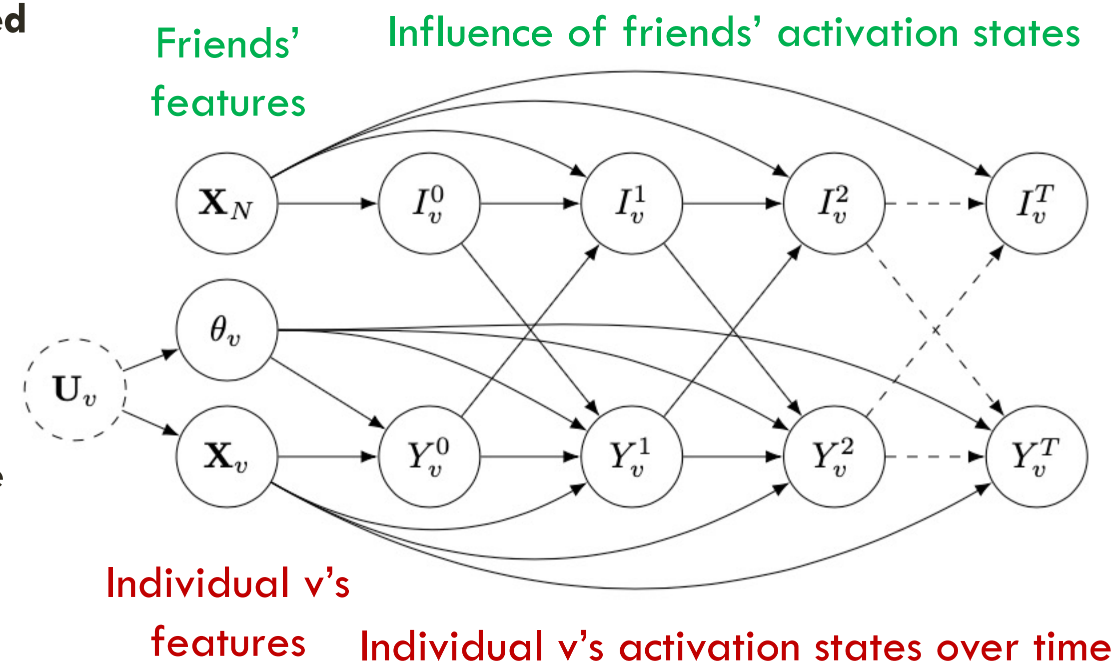
Individual-level threshold estimation

- Find minimum threshold that would cause a node to activate

$$\mathcal{P}(\mathbf{x}) = E[\mathcal{Y}(I^t \geq \hat{\theta}) - \mathcal{Y}(I^t < \hat{\theta}) \mid \mathbf{X} = \mathbf{x}]$$

- Two models: trigger-based causal trees and ST-learner

$$I_v^t = \sum_{u \in N(v)} w_{uv} Y_u^{t-1}$$



Tran & Zheleva. *Heterogeneous Peer Effects in the Linear Threshold Model*. AAAI 2022.

CAUSAL INFERENCE FROM RELATIONAL DATA

(10-MINUTE BREAK)

Presenters:
Elena Zheleva, University of Illinois Chicago @elenadata
David Arbour, Adobe Research @darbour26

AAAI 2022 Tutorial
February 23, 2022
https://netcause.github.io

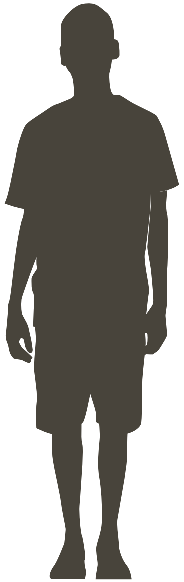
COUNTERFACTUALS & CAUSAL EFFECTS IN OBSERVATIONAL NETWORK DATA

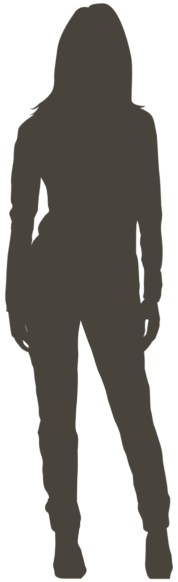
Representation Challenges


How do we get people to vote?


I Voted Early
Good job! Were the lines longs?

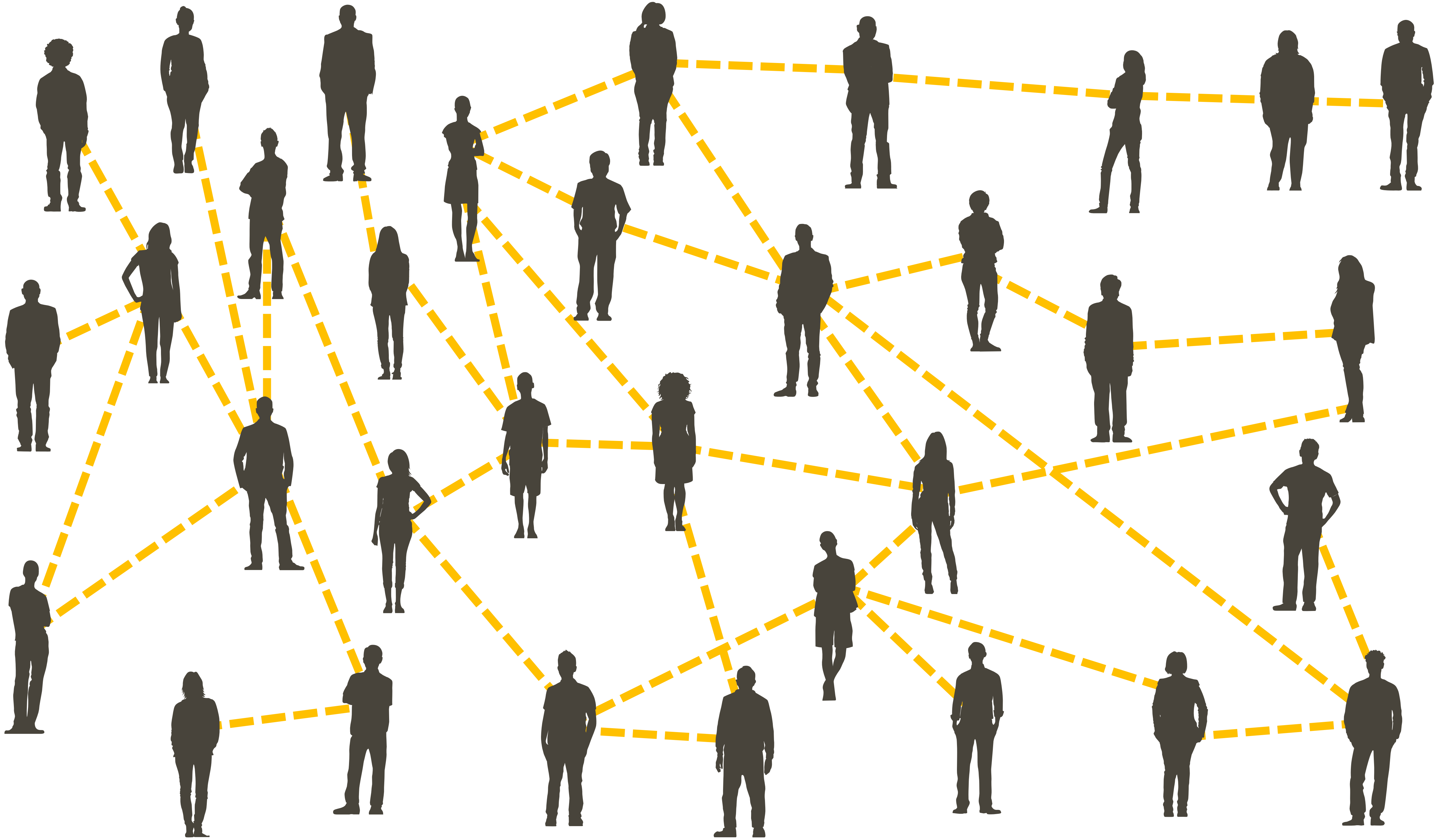
WHAT'S THE EFFECT?



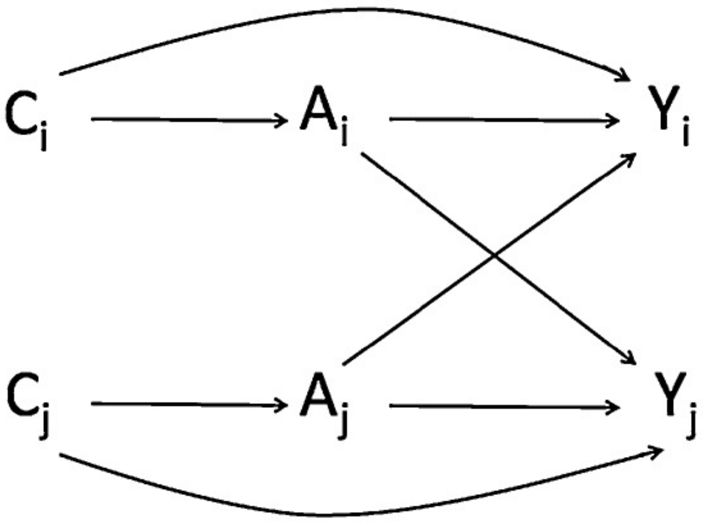


Last Name	First Name	DOB	State	Voted (Y/N)

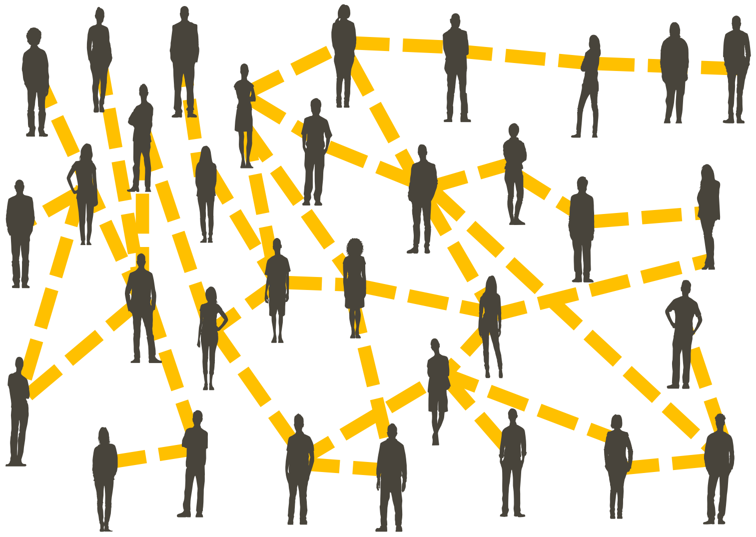
OBSERVED DATA



CHALLENGES

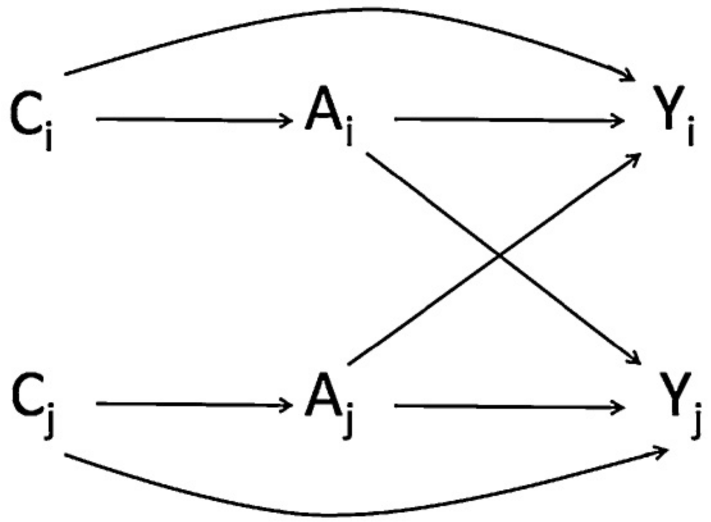


Causal



Network

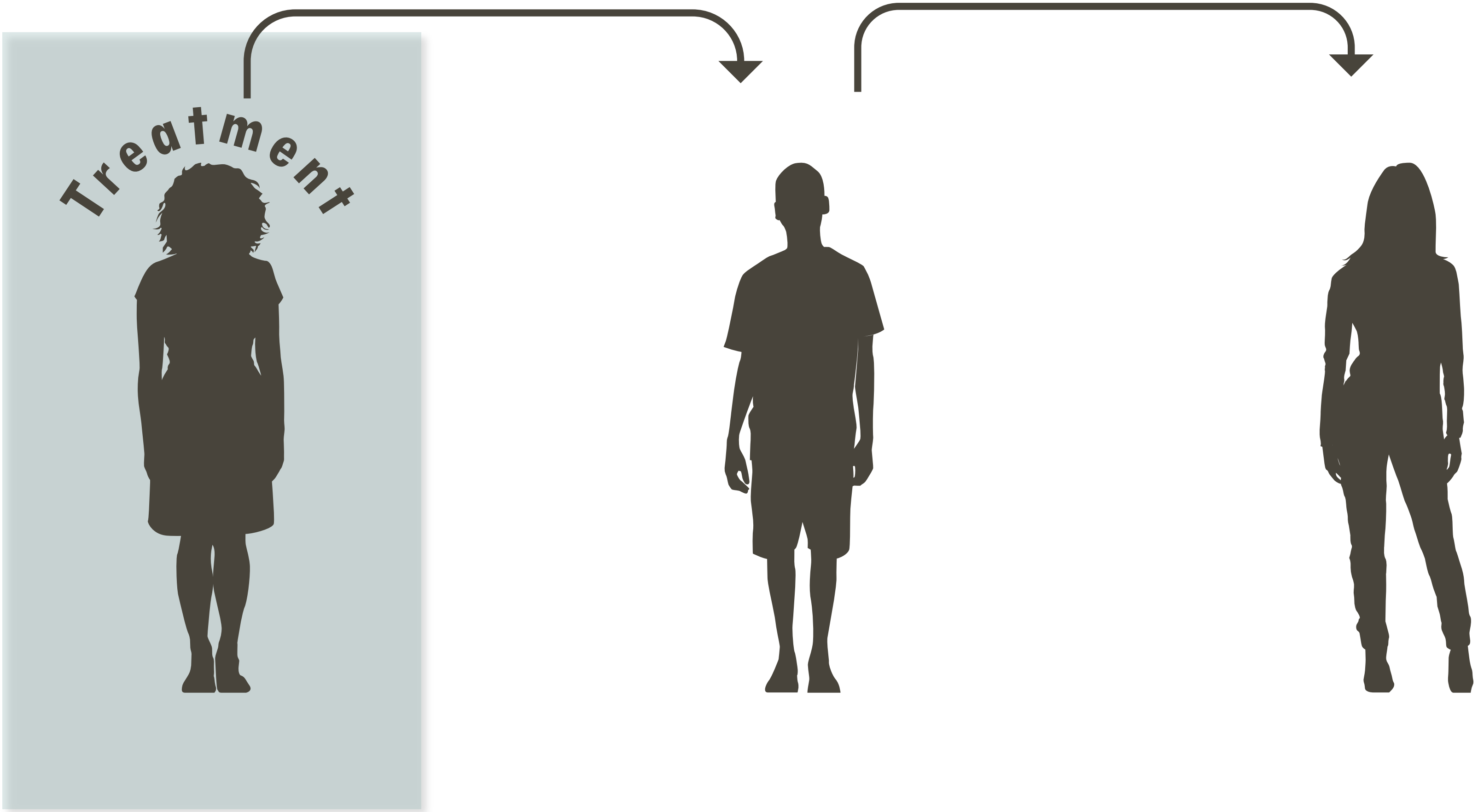
CASUAL CHALLENGES

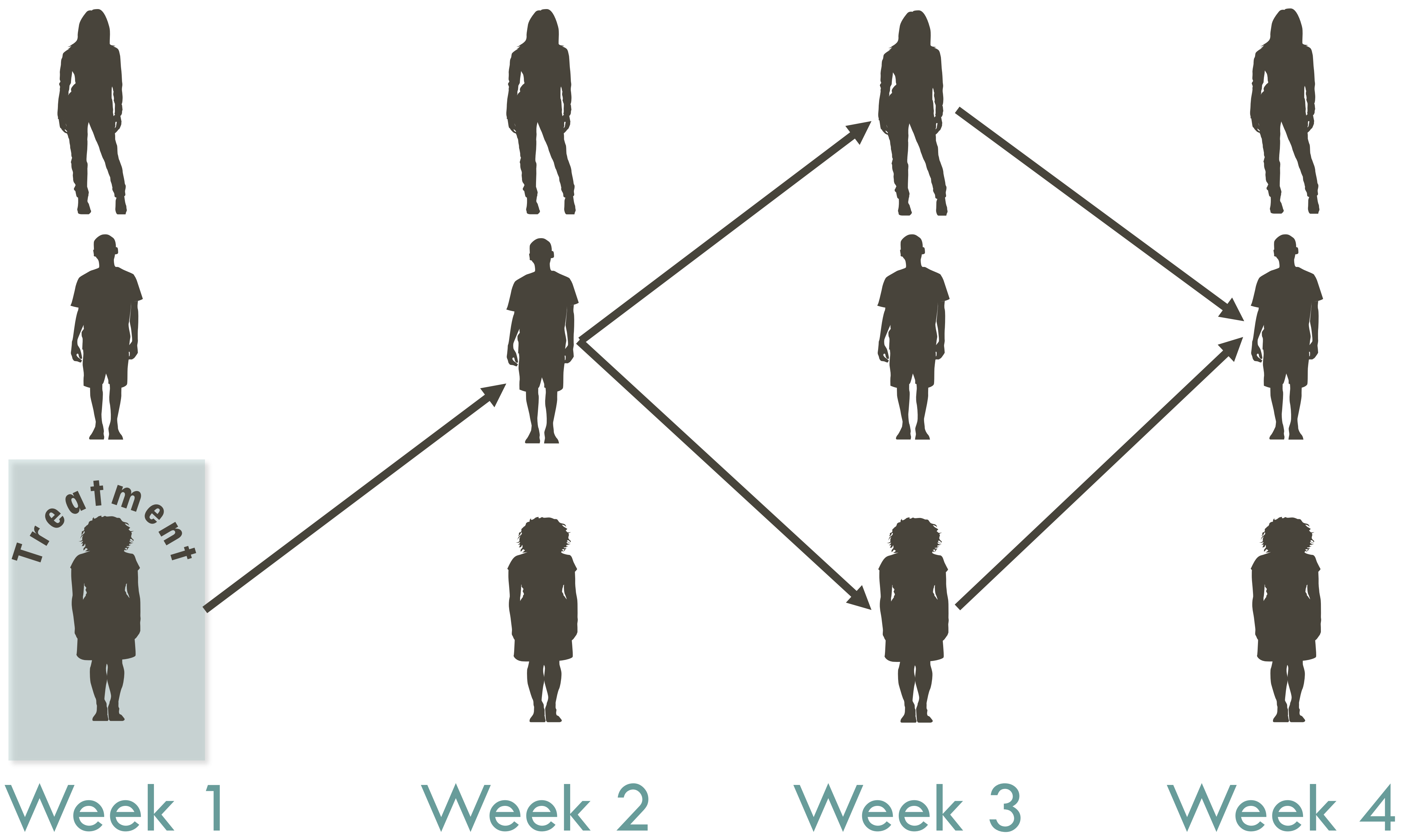


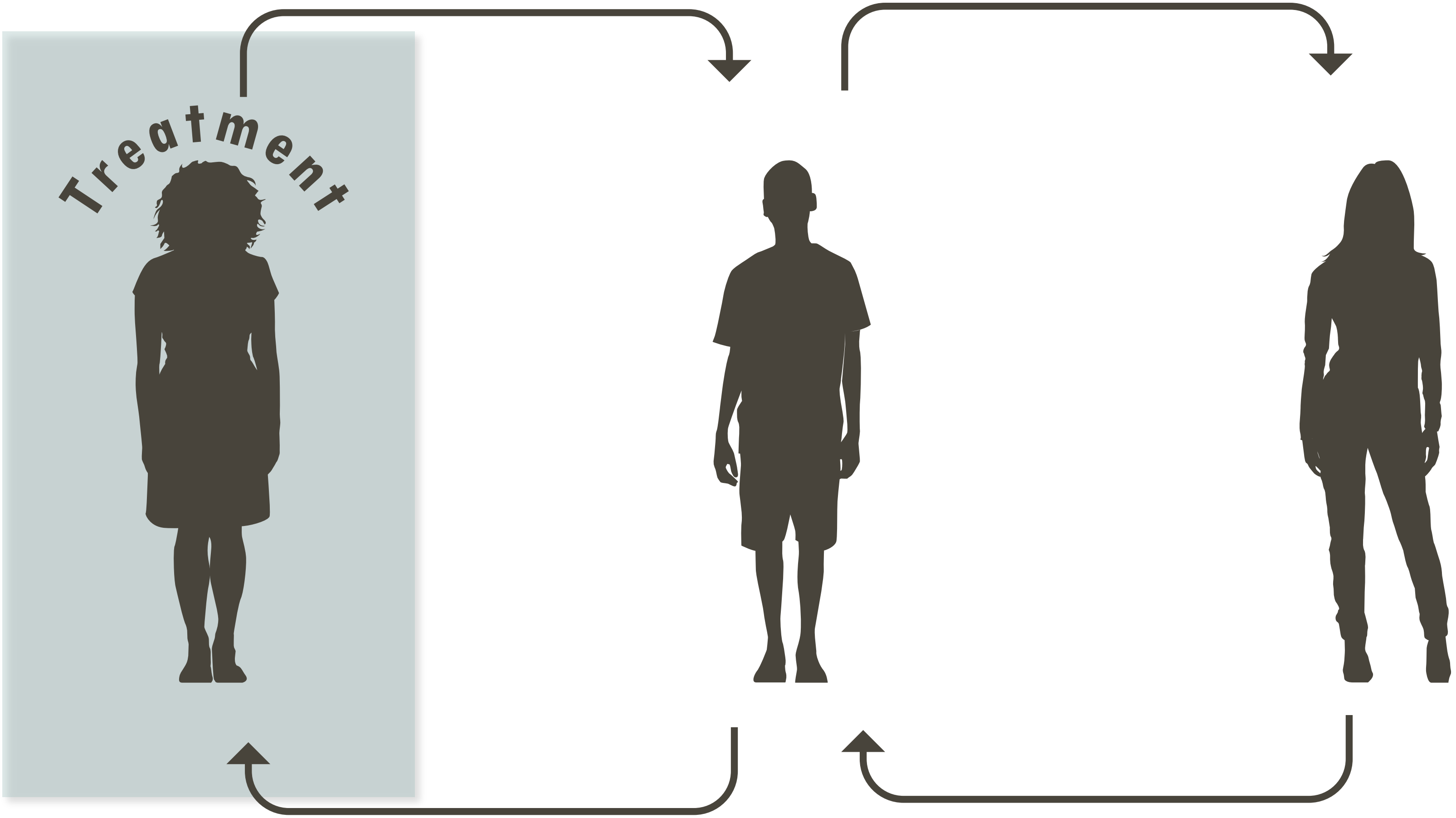
1. Feedback

2. Set Valued Counterfactuals

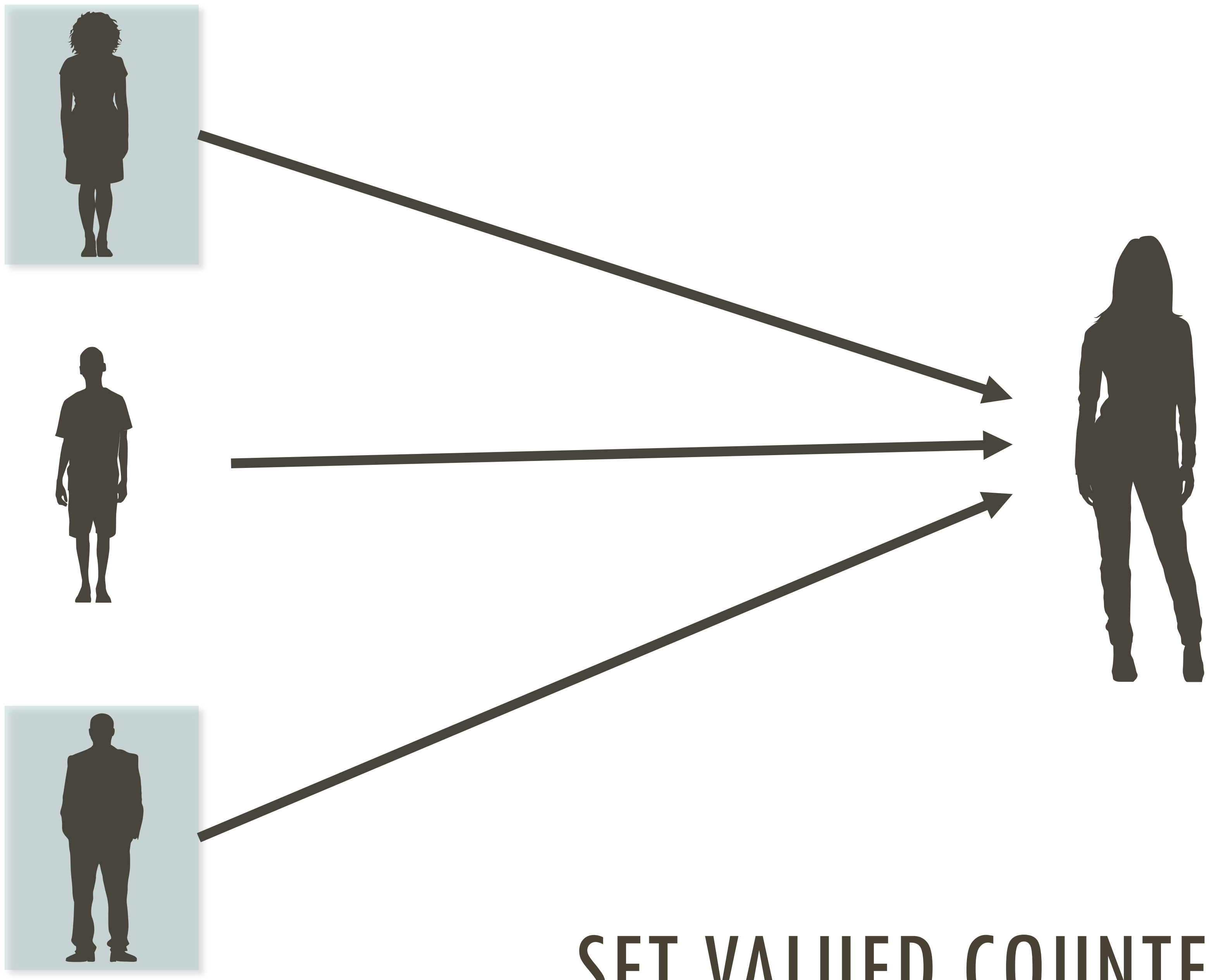
FEEDBACK



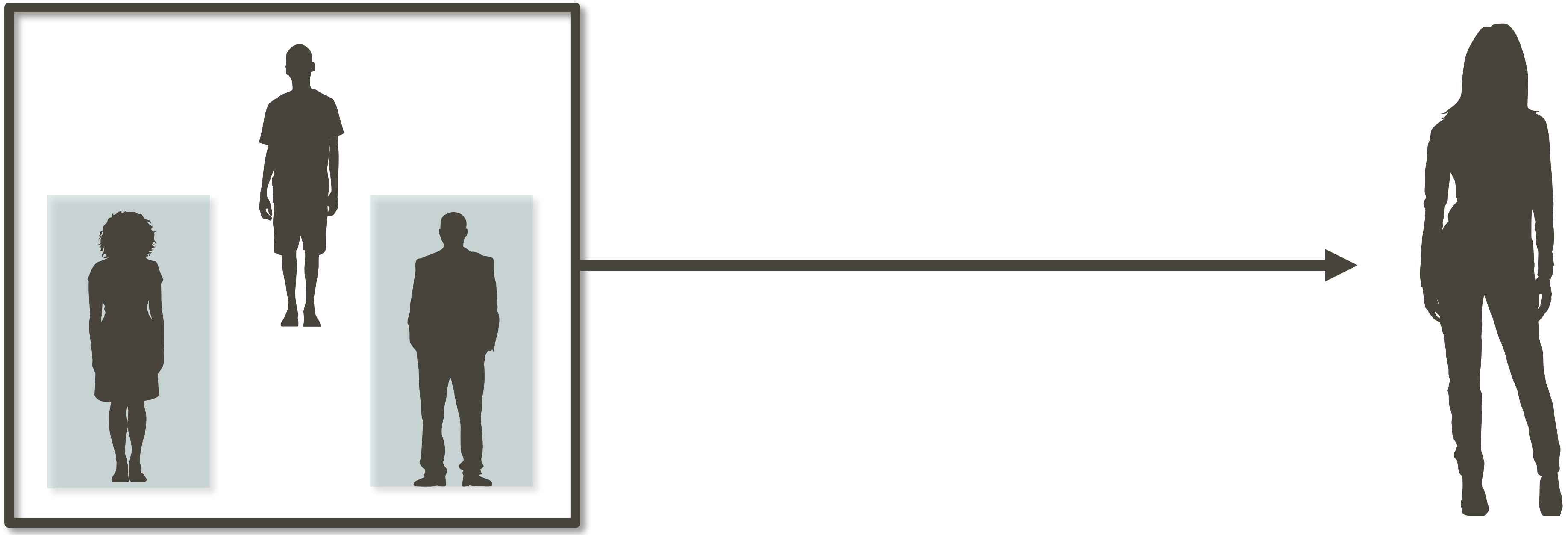

Treatment
Week 1
Week 2
Week 3
Week 4



Pooled Data

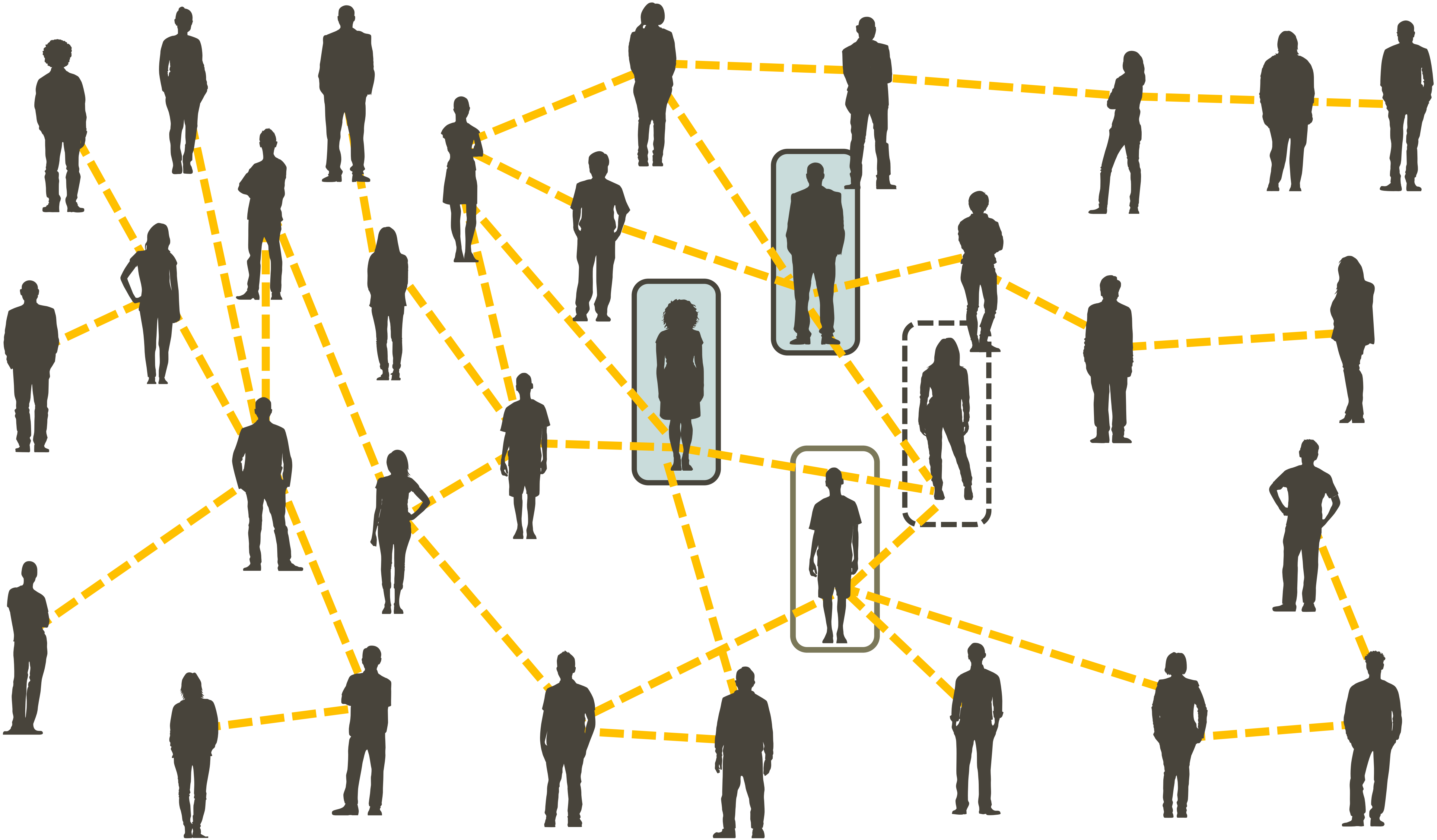


SET VALUED COUNTERFACTUALS

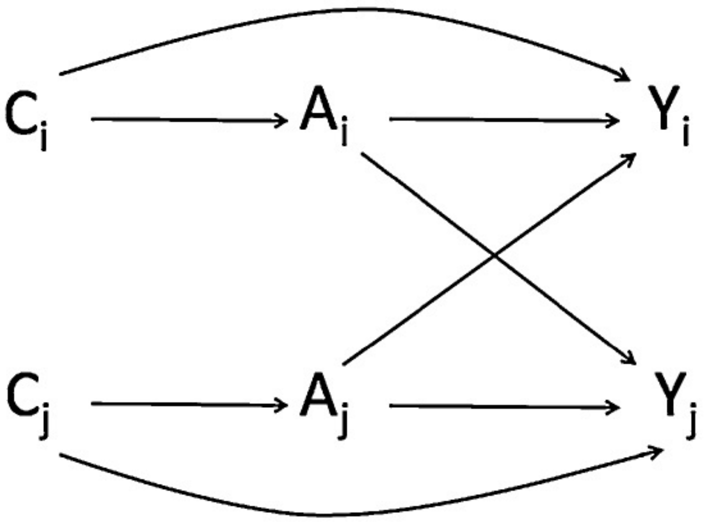


2/3 Treated

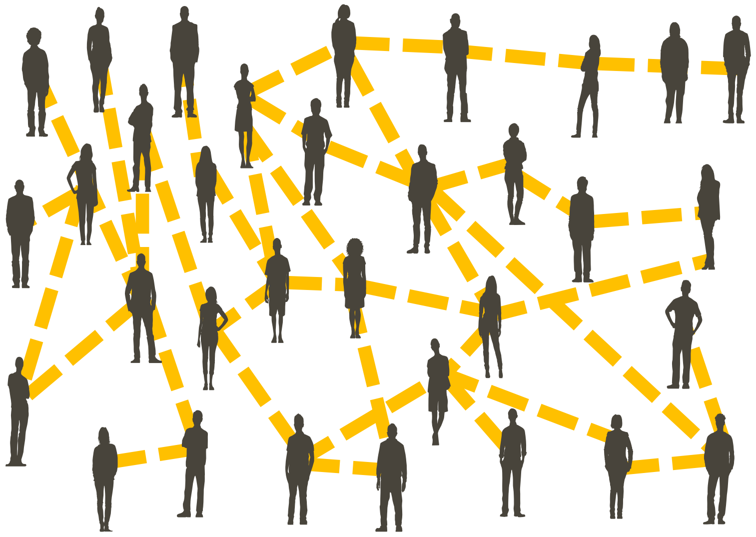
SET VALUED COUNTERFACTUALS



CHALLENGES



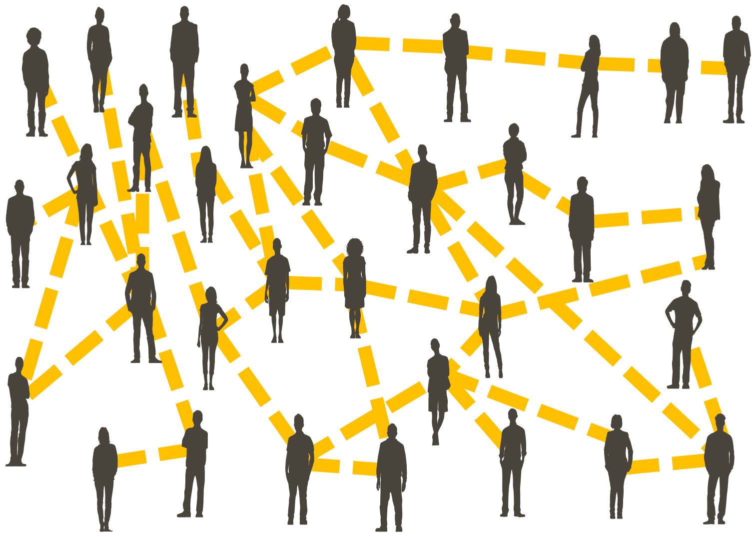
Causal



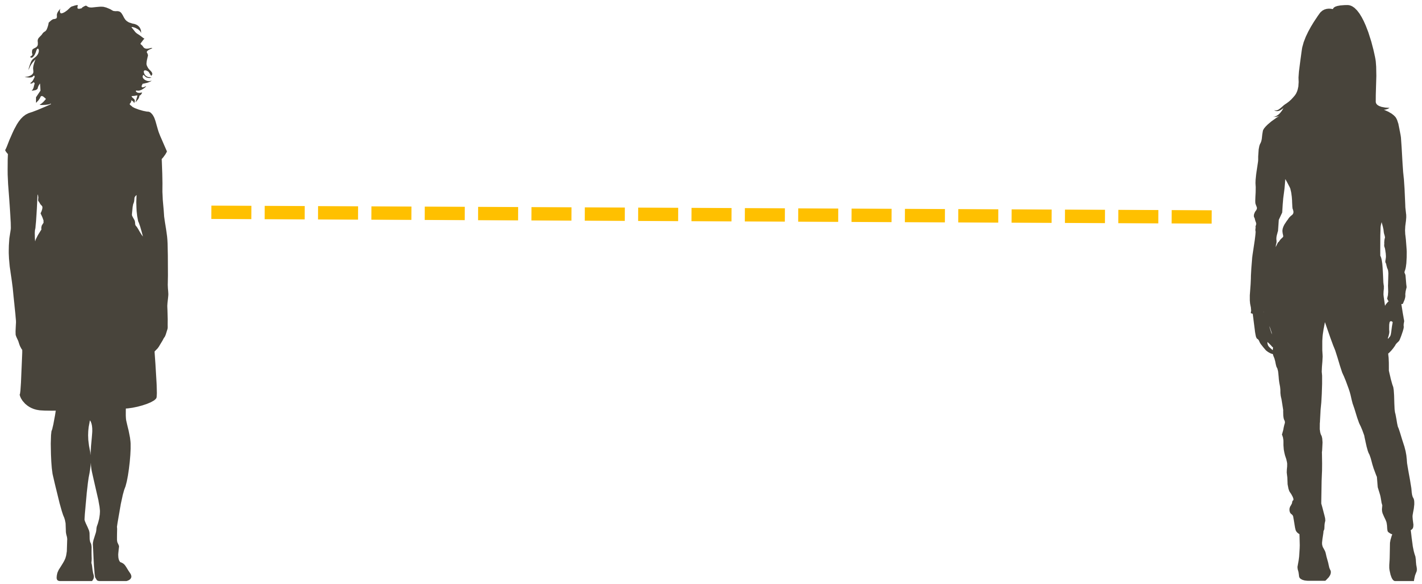
Network

NETWORK CHALLENGES

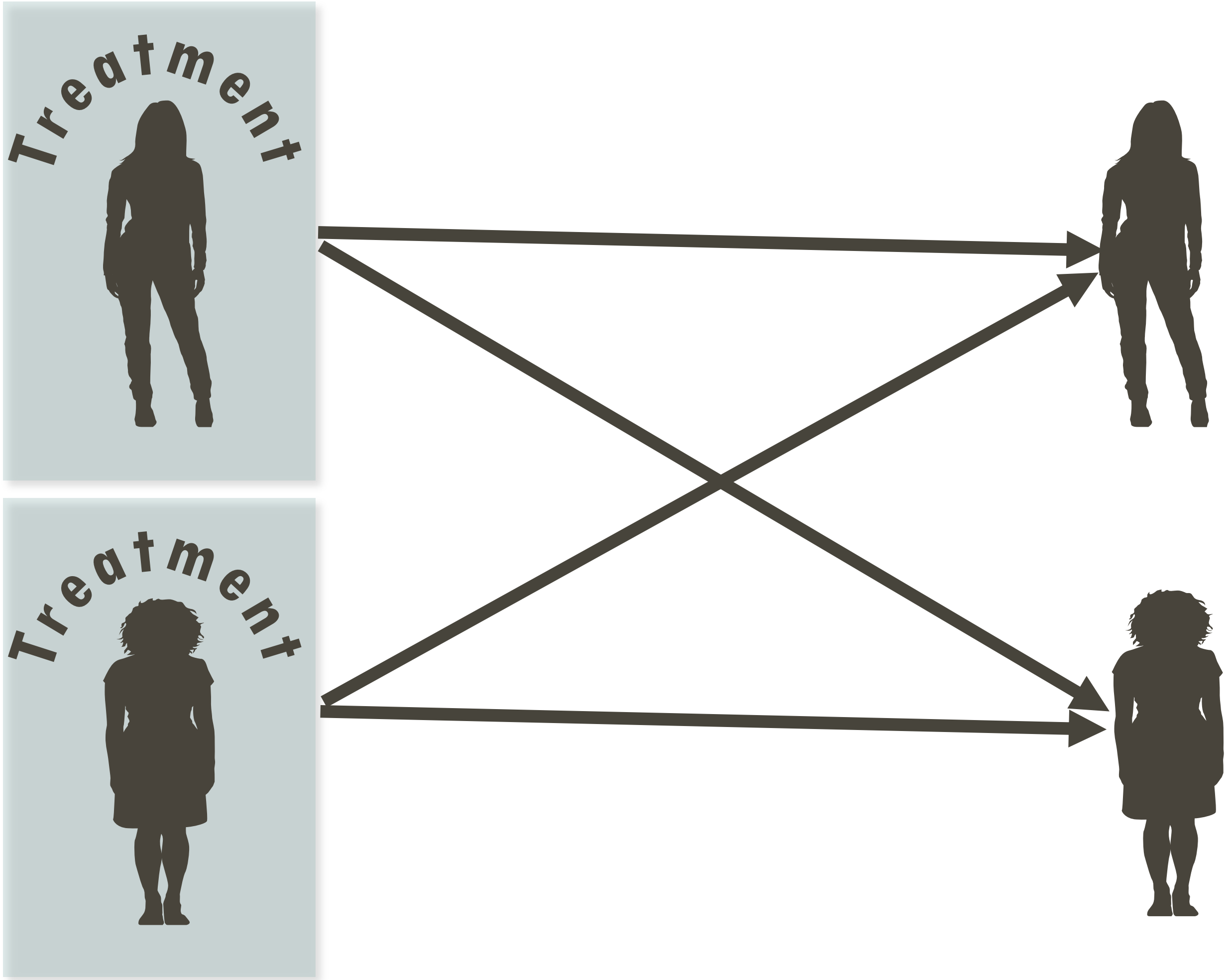
1. Directed / Undirected Edges
2. Multiple Entities & Relationships
3. Unobserved / Partially Observed



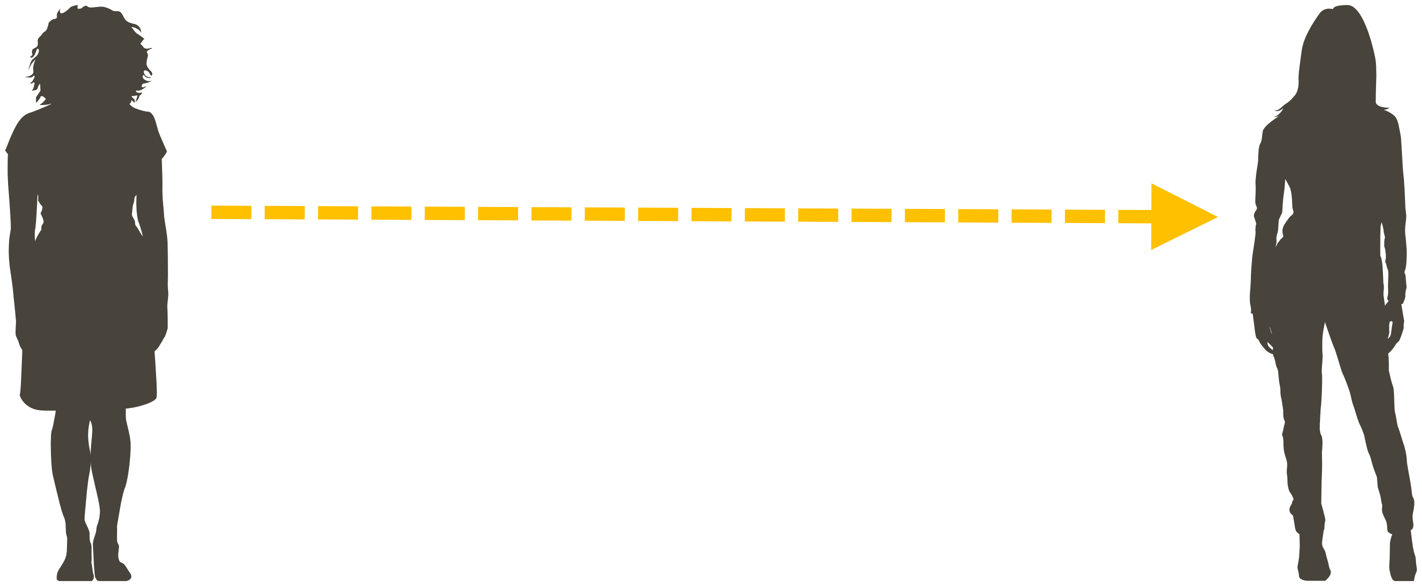
UNDIRECTED RELATIONSHIPS



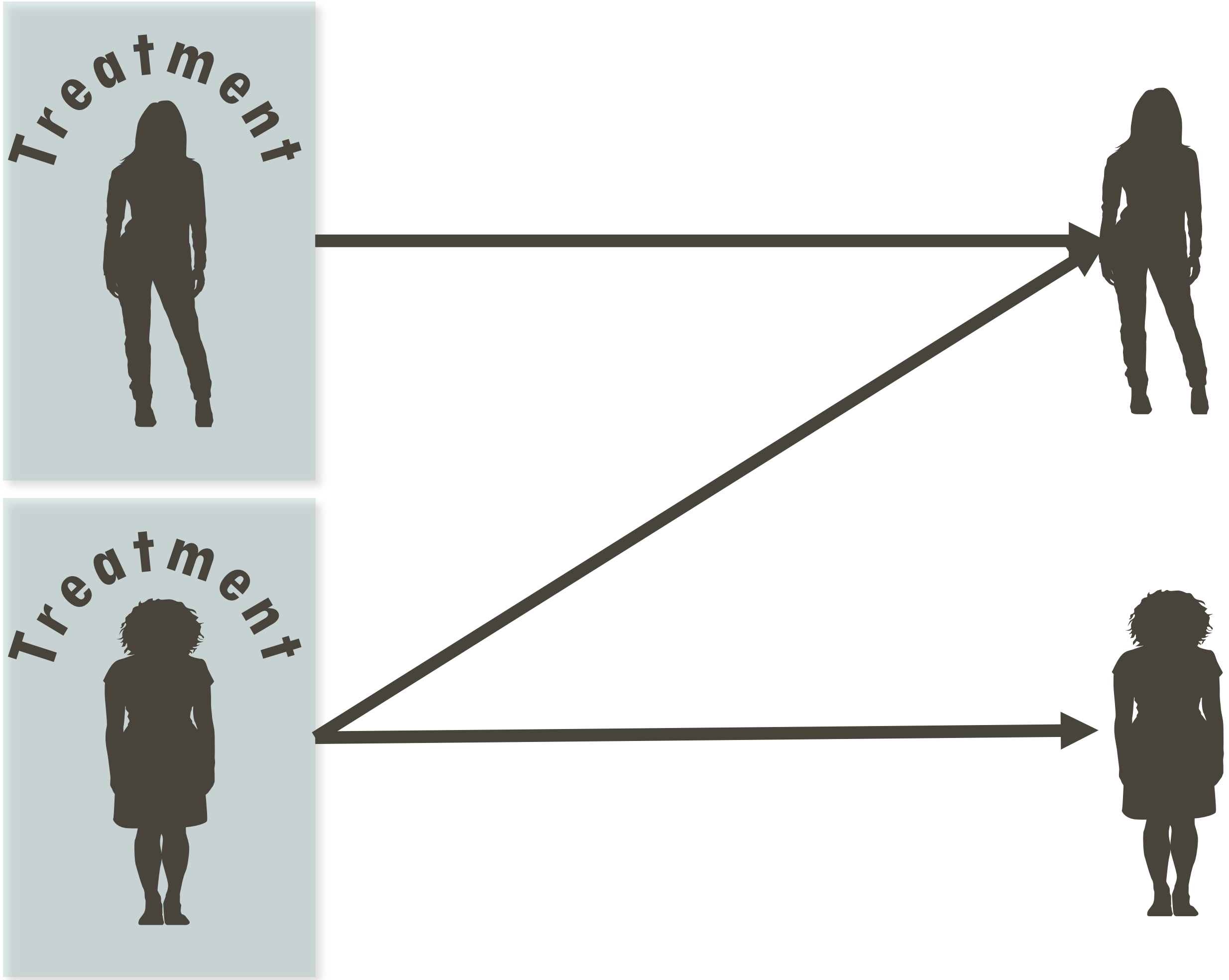
UNDIRECTED RELATIONSHIPS



DIRECTED RELATIONSHIPS



DIRECTED RELATIONSHIPS

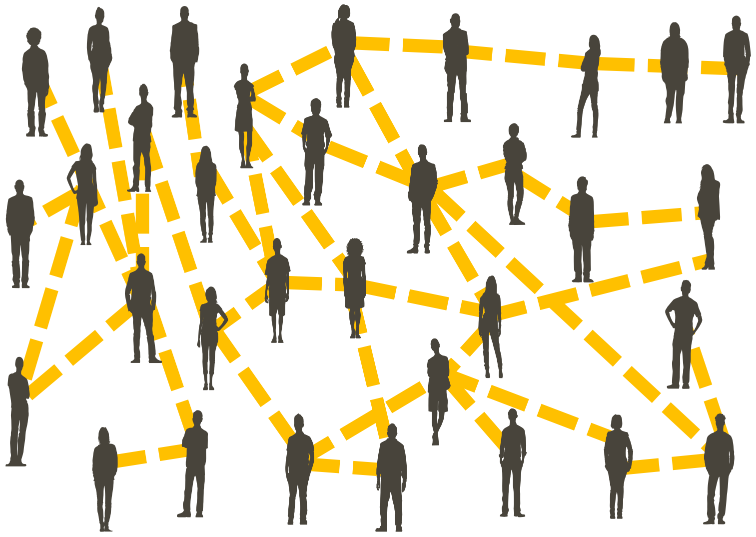


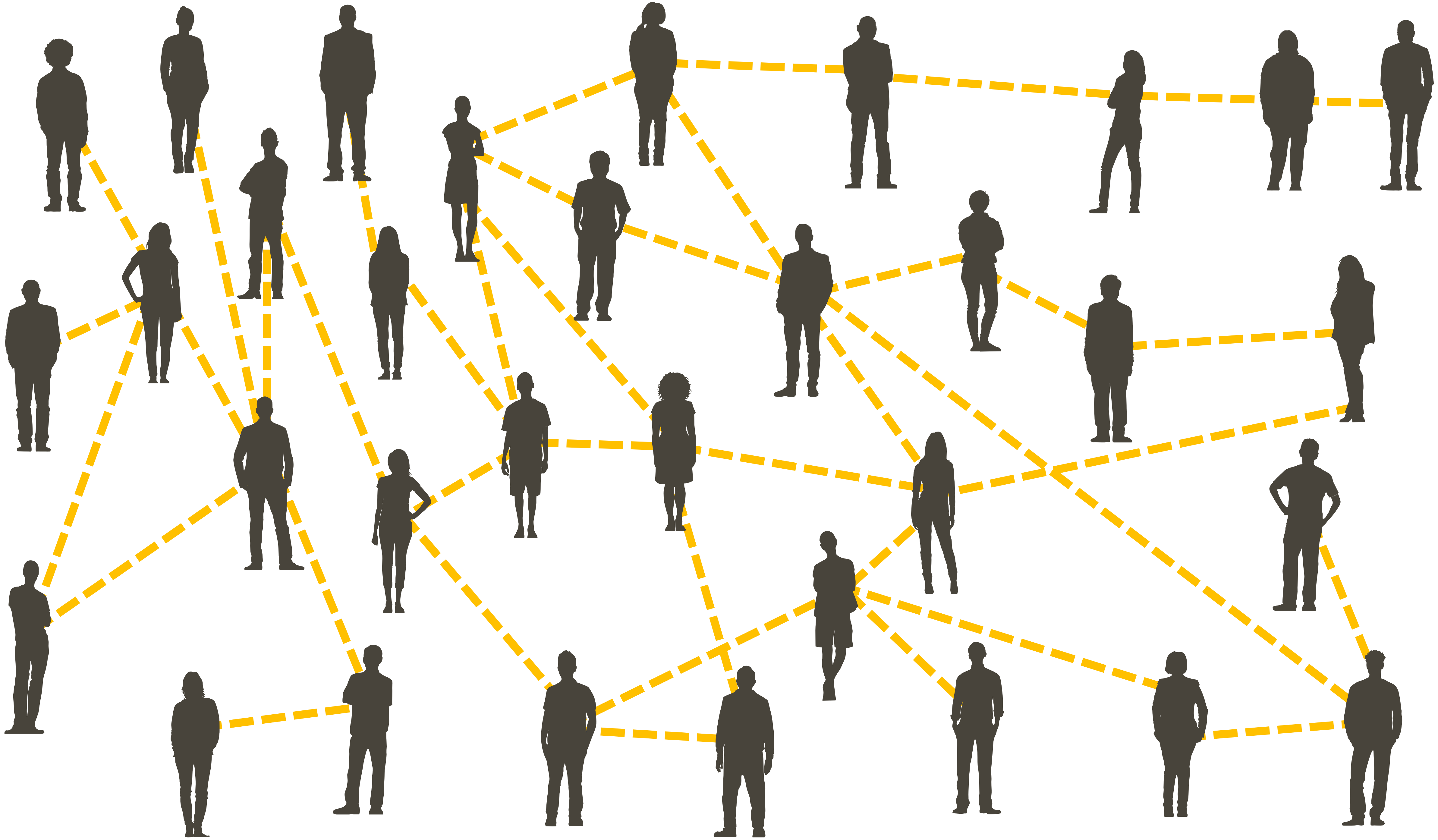
NETWORK CHALLENGES

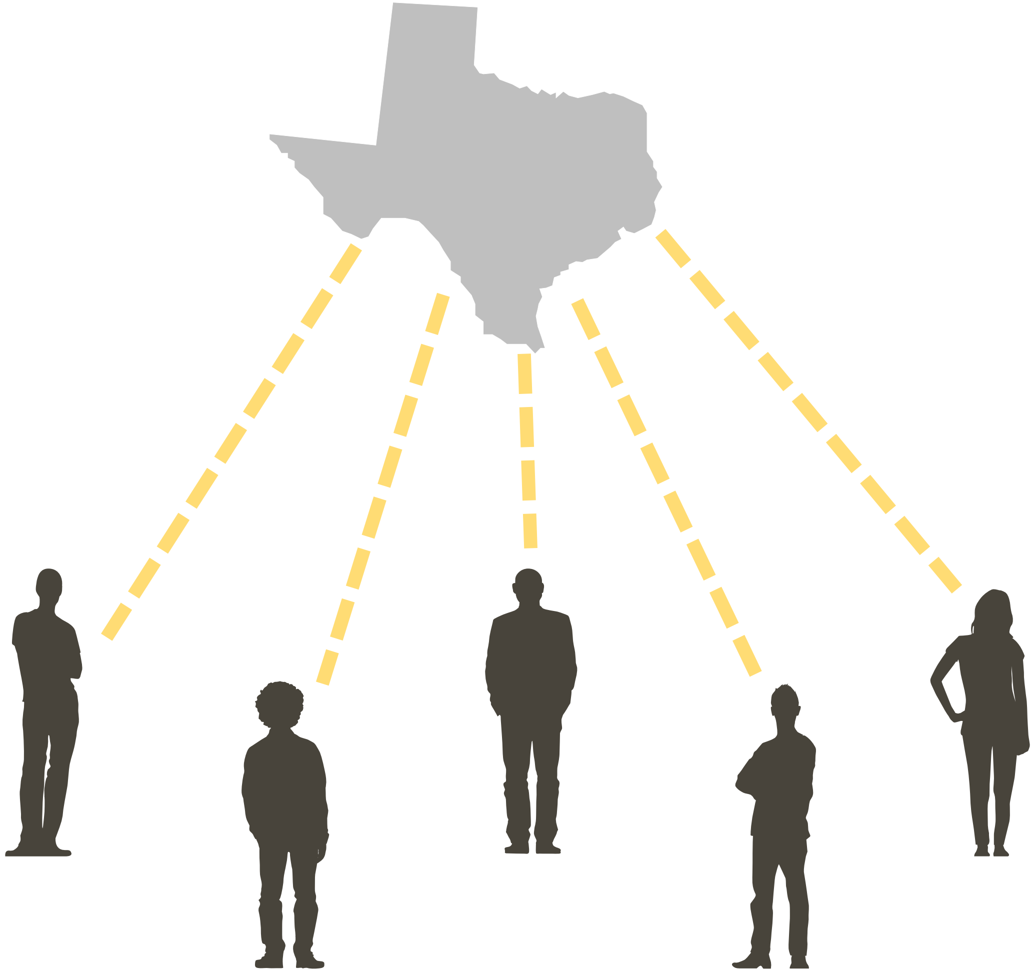
1. Directed / Undirected Edges

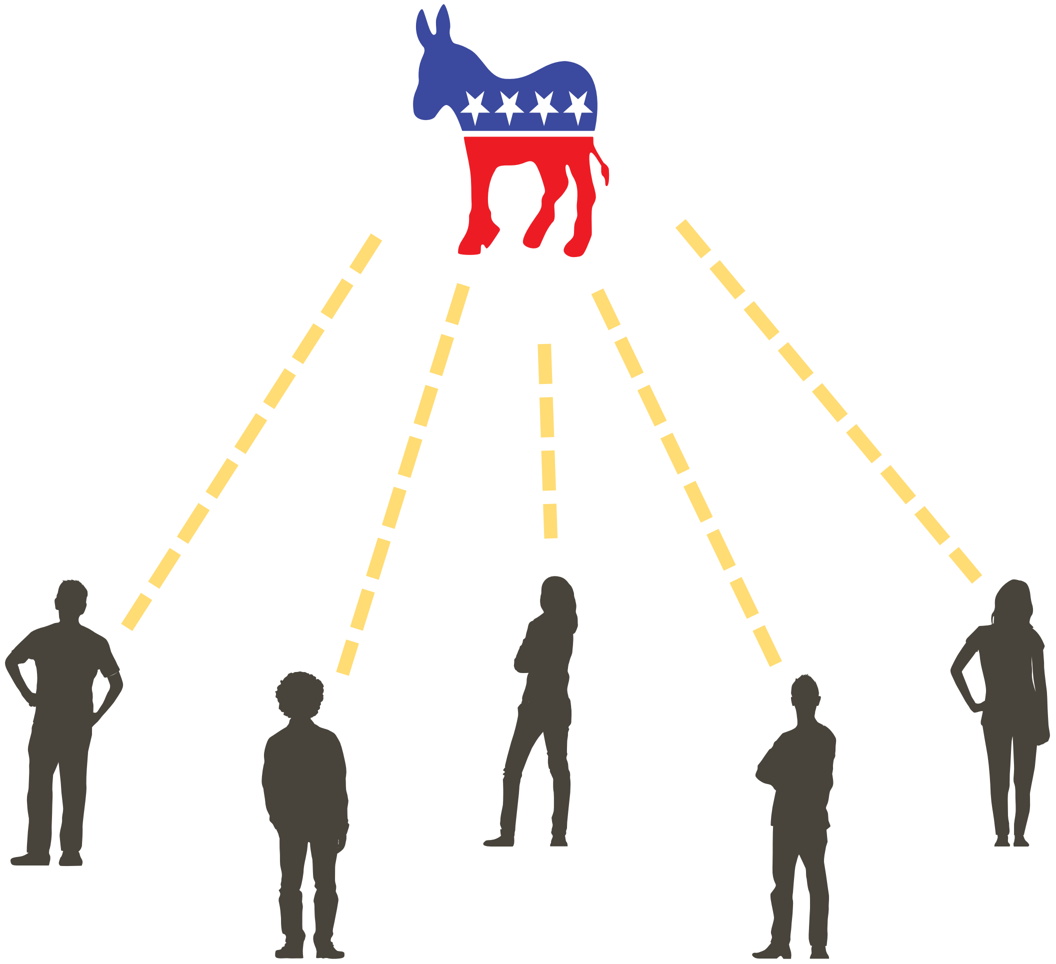
2. Multiple Entities & Relationships

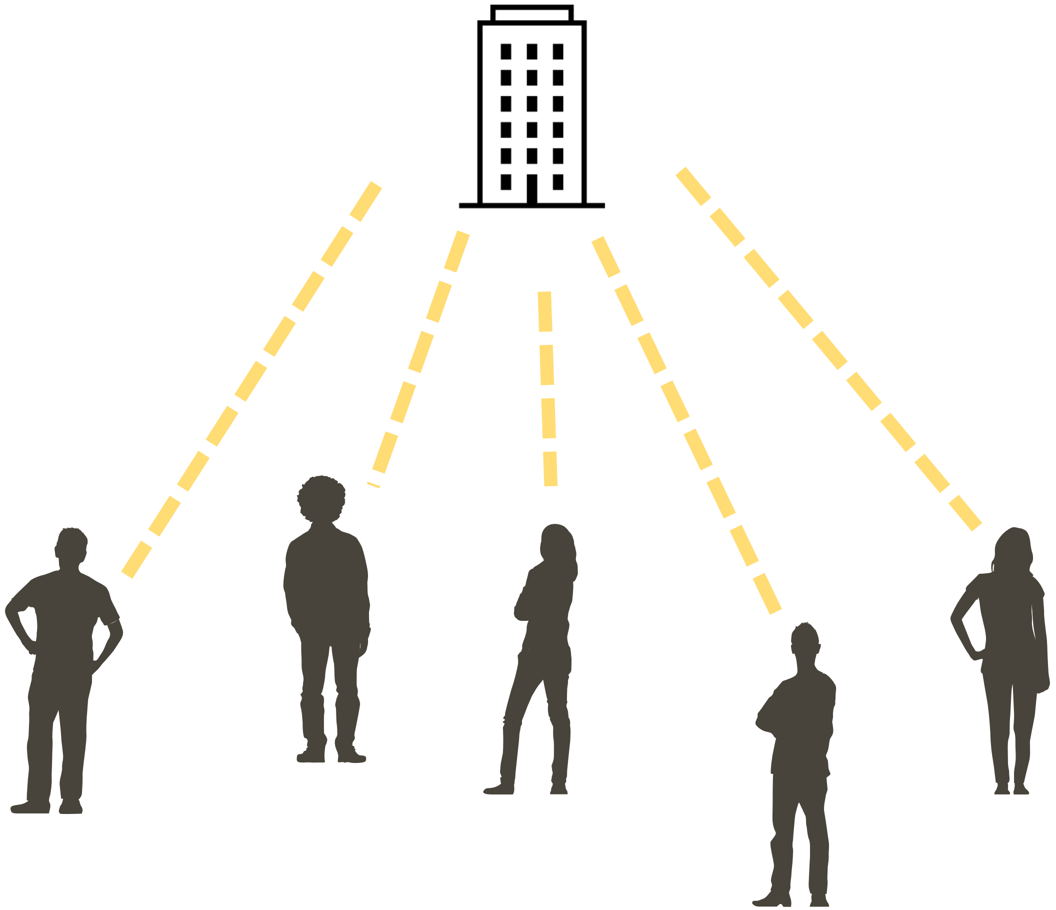
3. Unobserved / Partially Observed

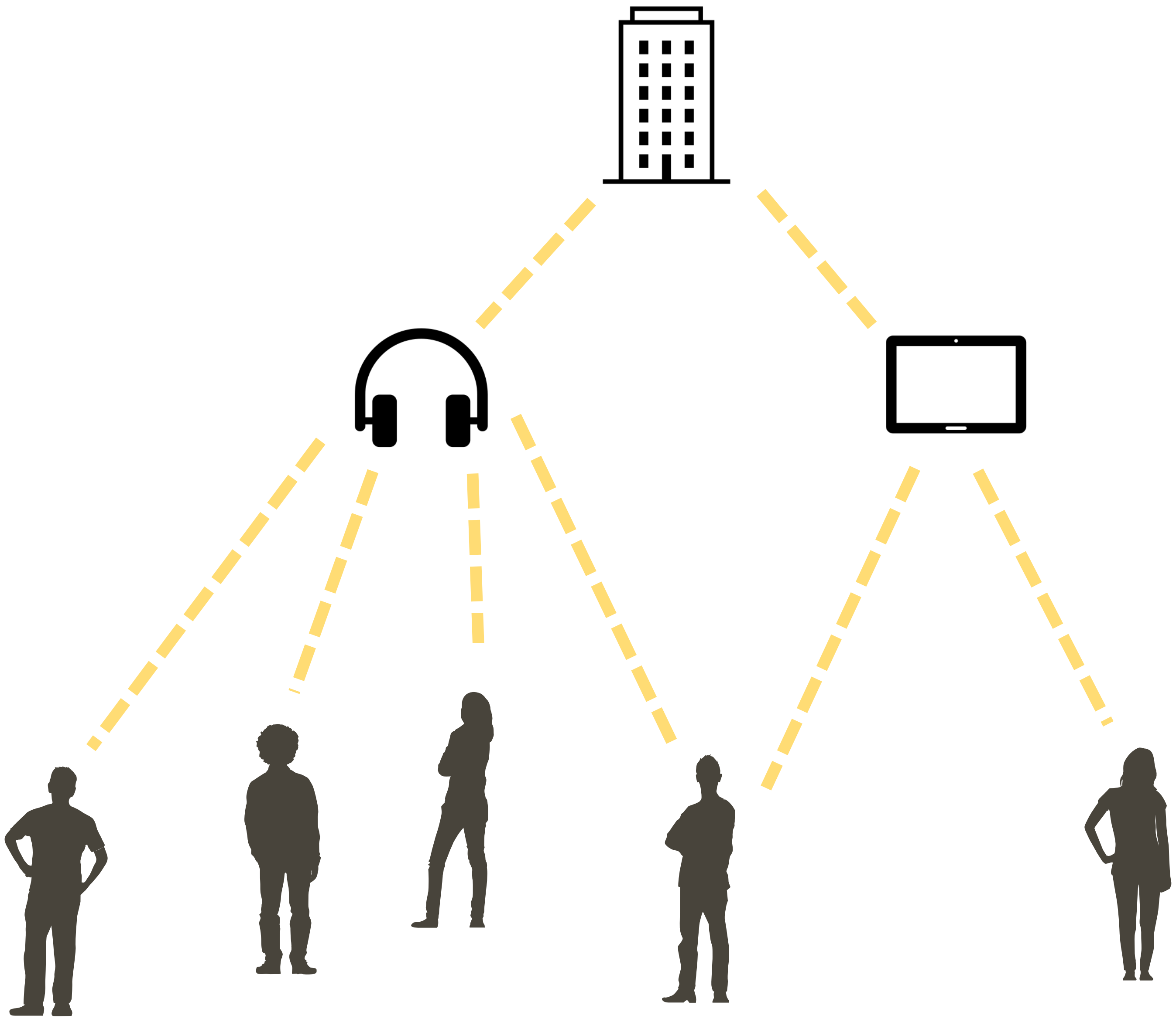










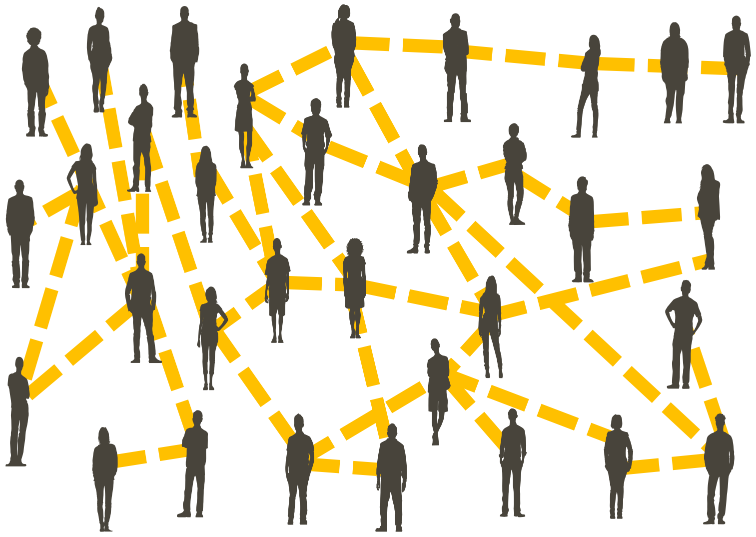


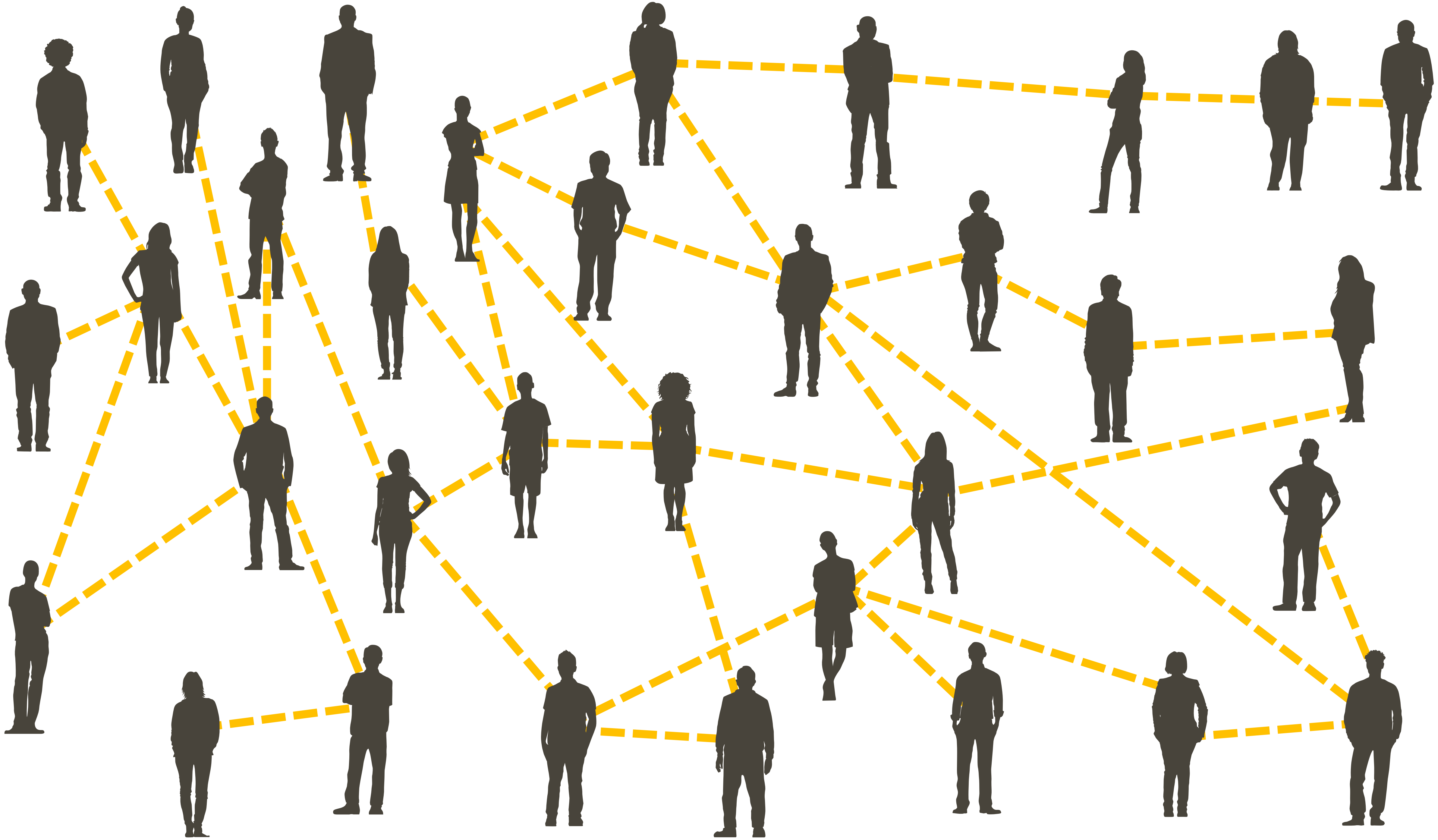
NETWORK CHALLENGES

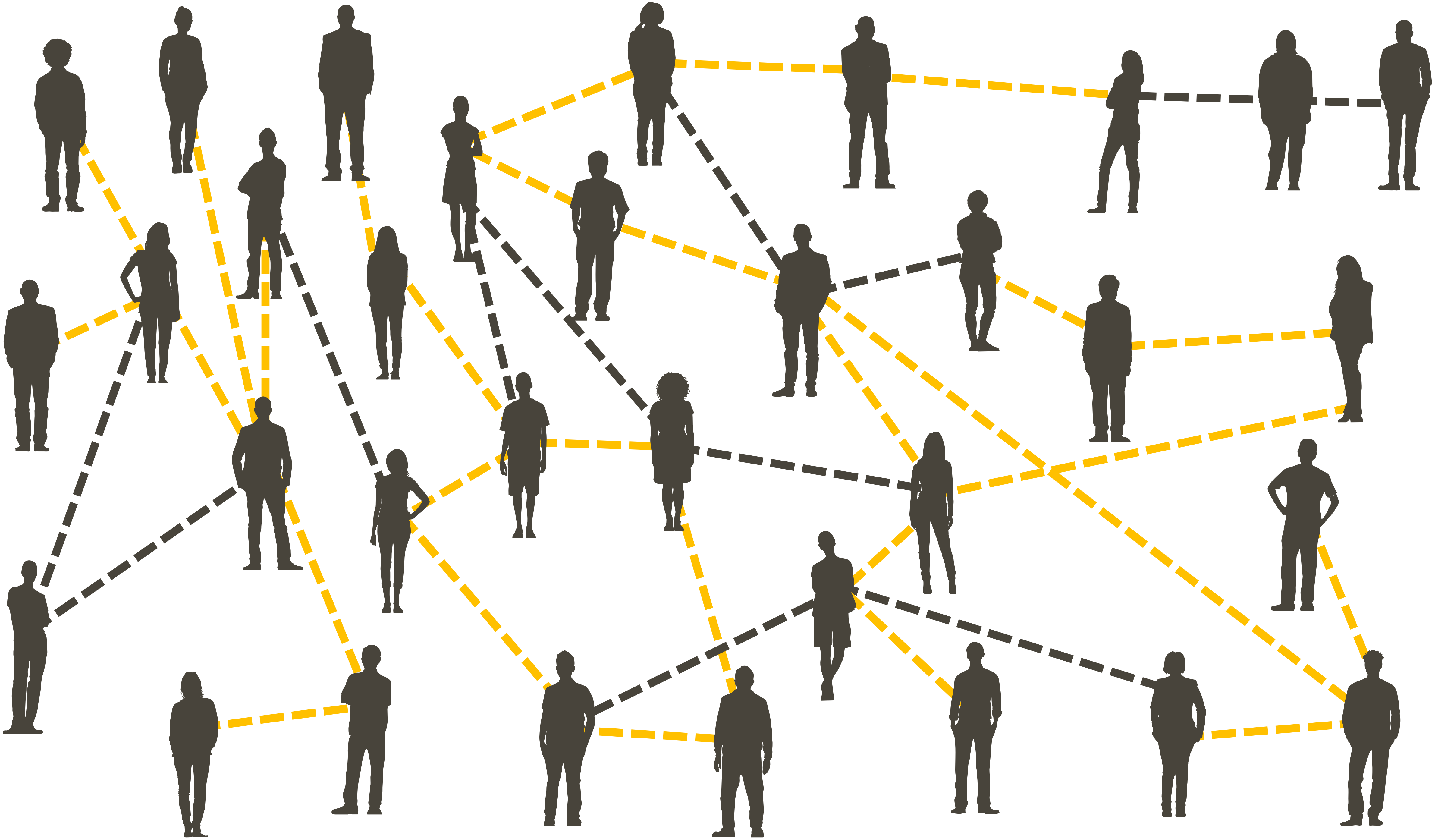
1. Directed / Undirected Edges

2. Multiple Entities & Relationships

3. Unobserved / Partially Observed





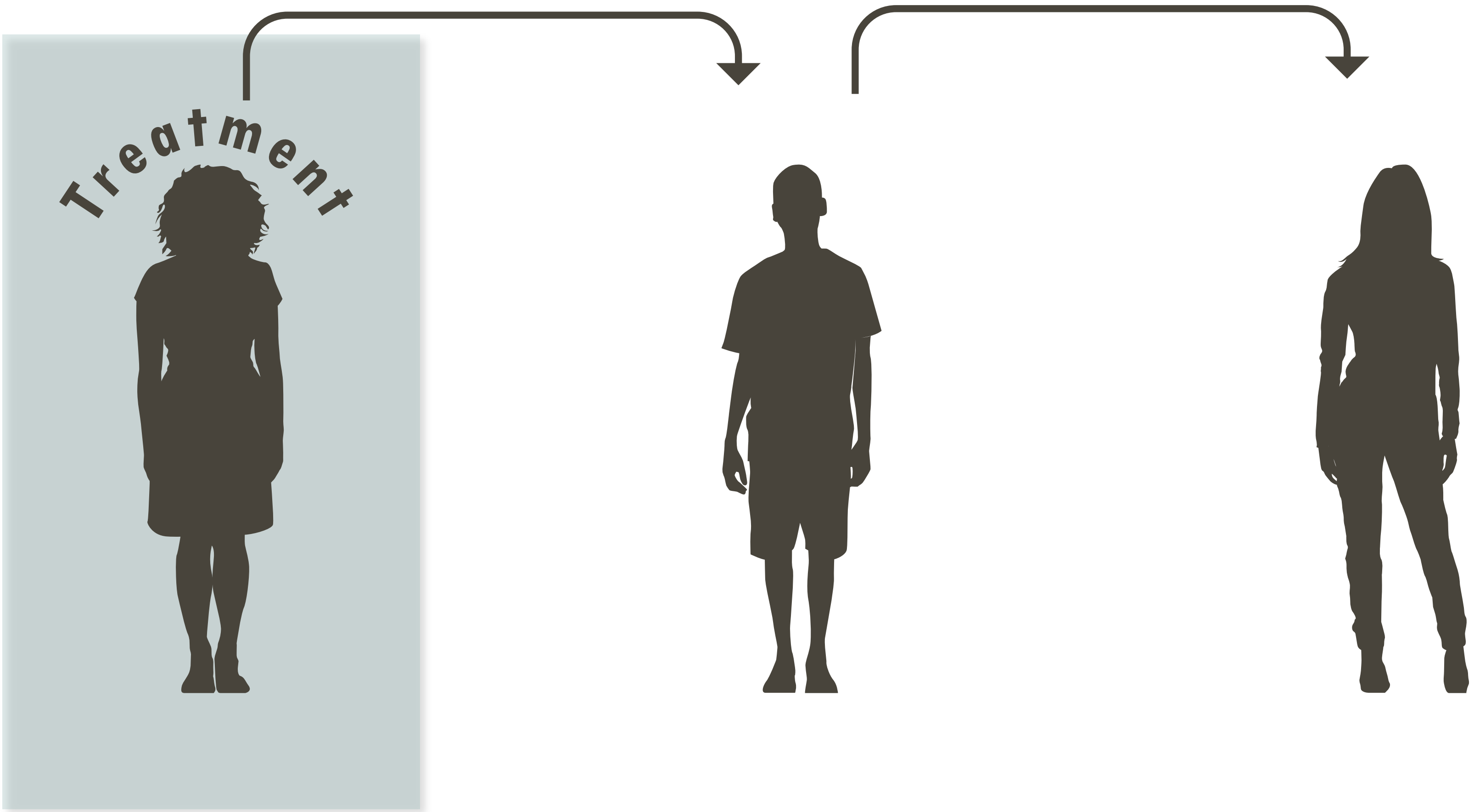


	Directed & Undirected Edges	Multiple Entities and Relationships	Partially Observed Networks
Chain Graphs	✓		✓ in discovery
Aggregate Ground Graphs	✓	✓	

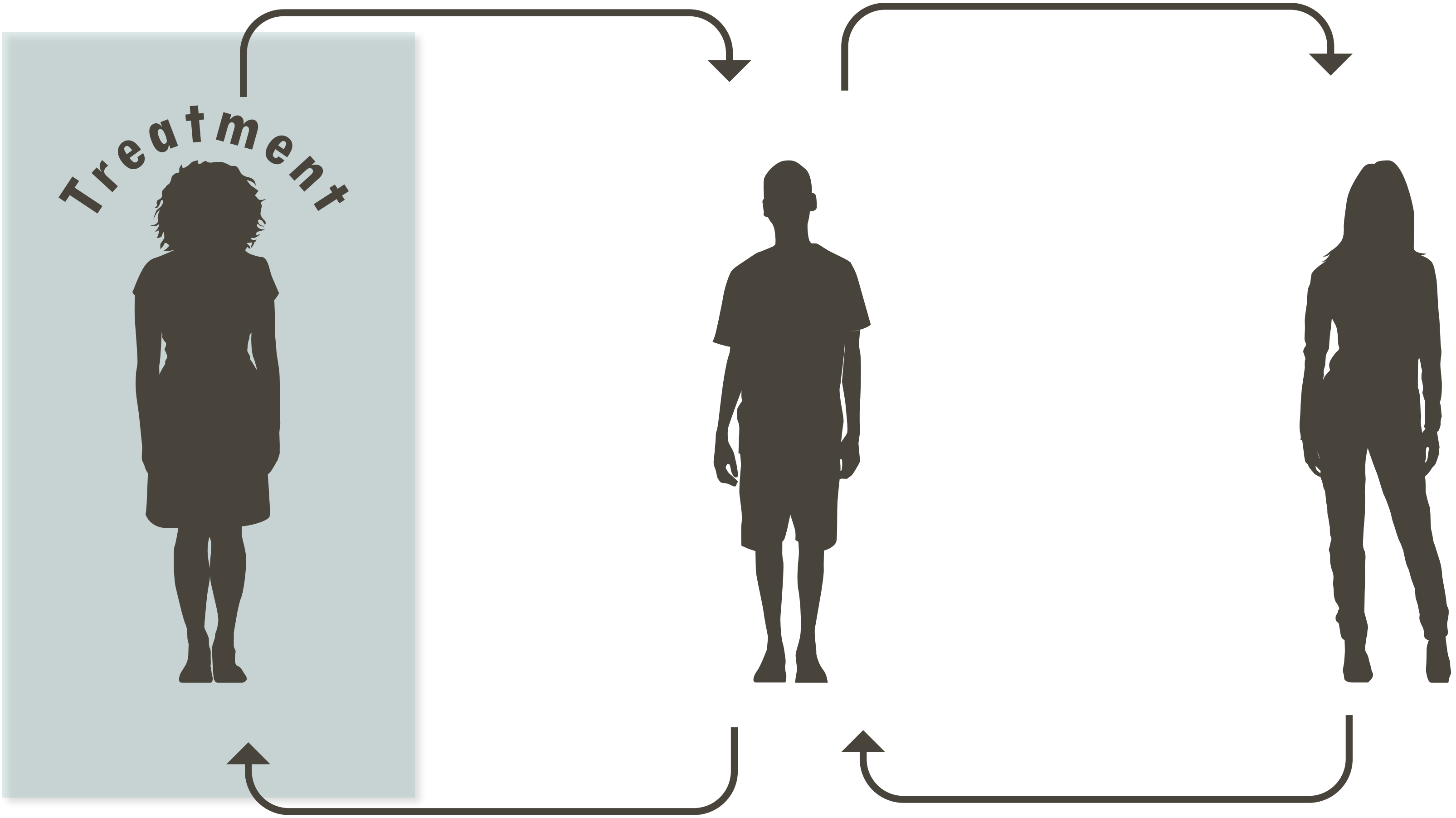
COUNTERFACTUALS & CAUSAL EFFECTS IN OBSERVATIONAL NETWORK DATA

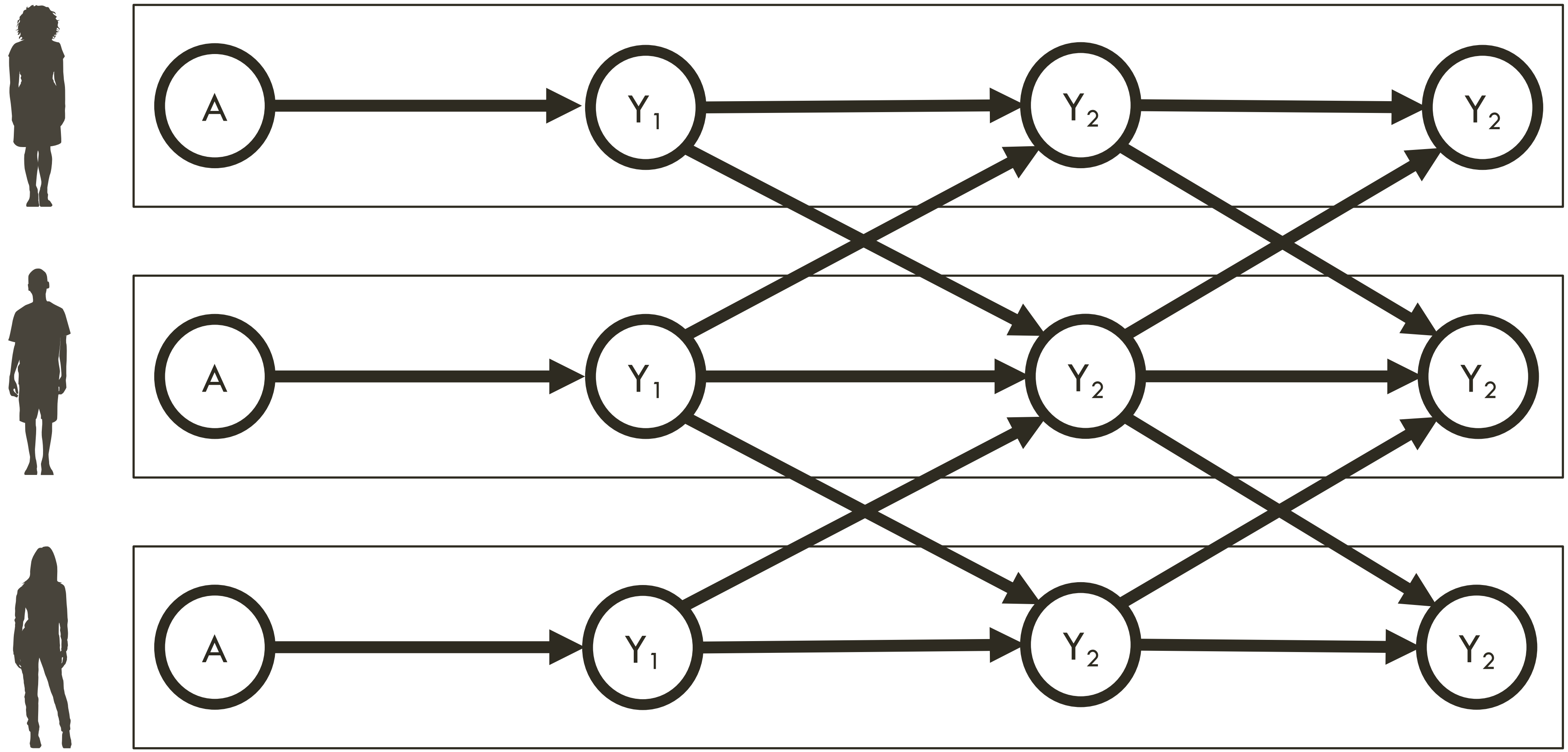
Chain and Segregated Graphs

ACYCLICITY

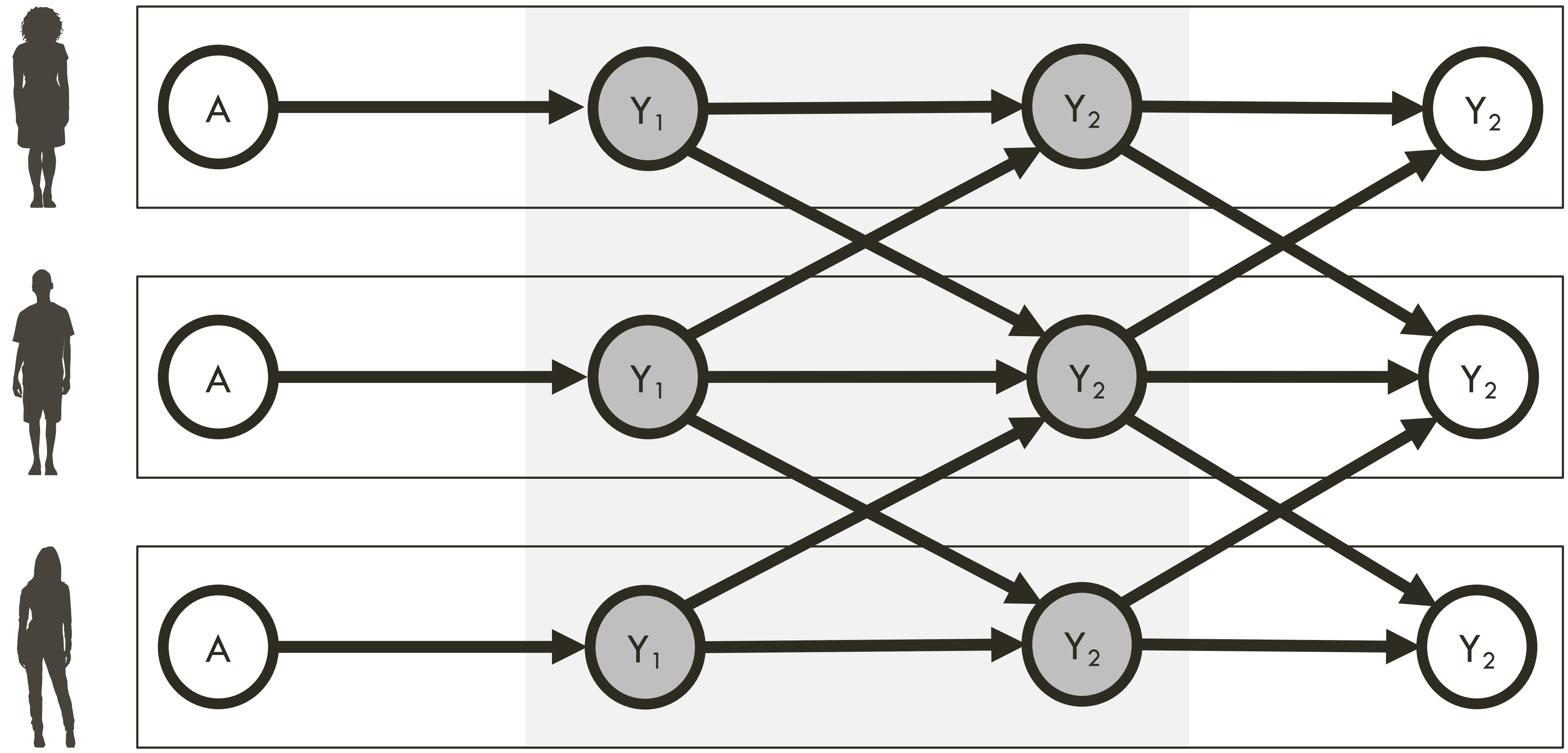


FEEDBACK

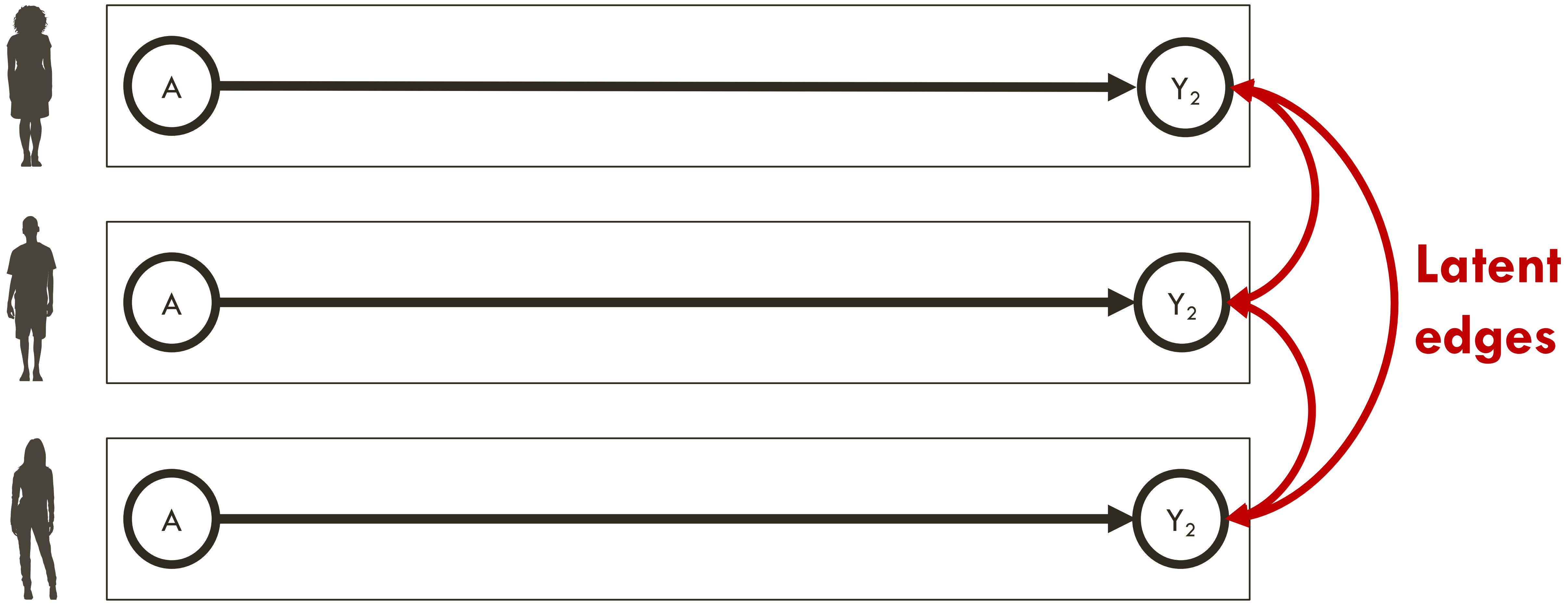




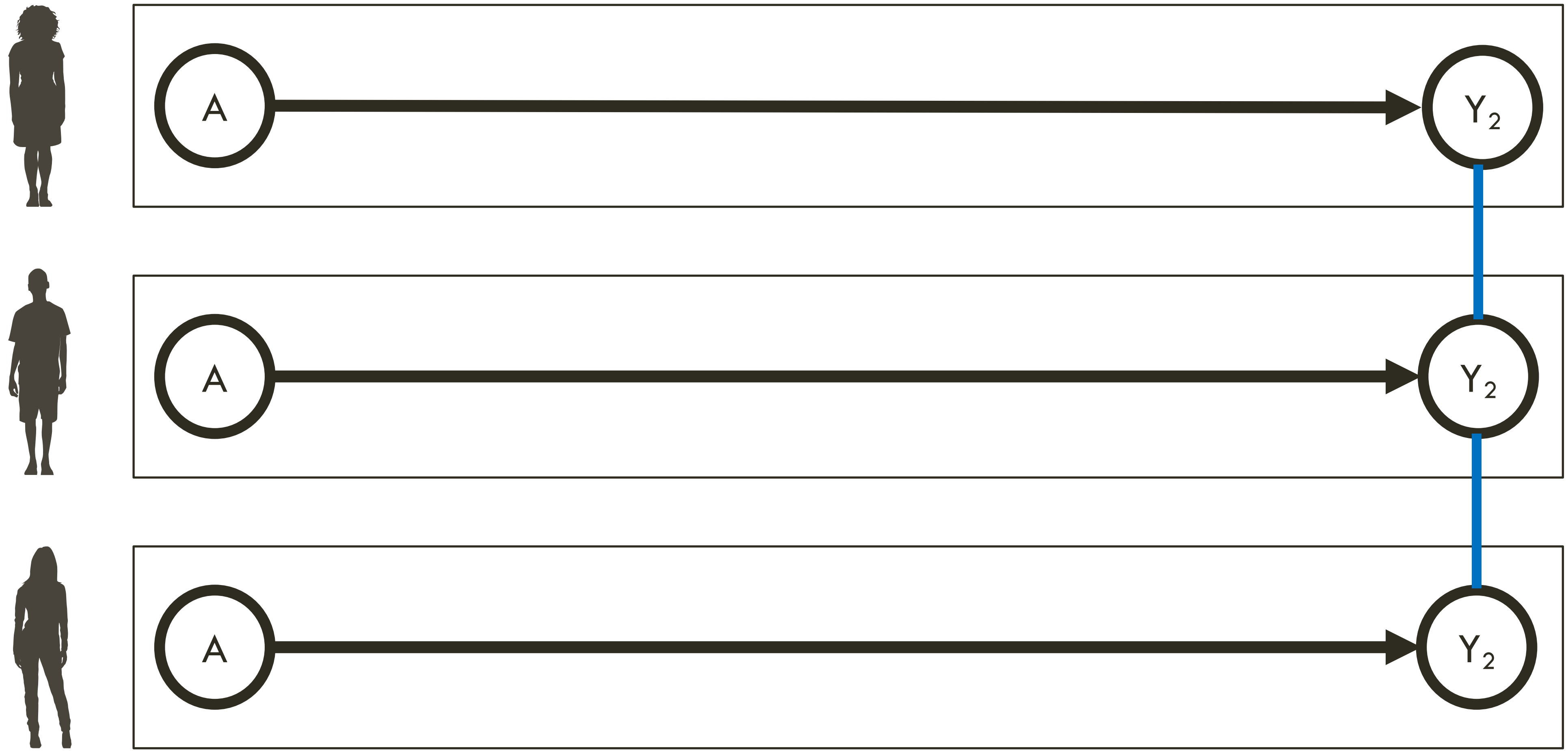
Ogburn, Shpitser and Lee. *Causal inference, social networks and chain graphs*. JRSSB 2020.



Ogburn, Shpitser and Lee. *Causal inference, social networks and chain graphs*. JRSSB 2020.



Ogburn, Shpitser and Lee. *Causal inference, social networks and chain graphs*. JRSSB 2020.

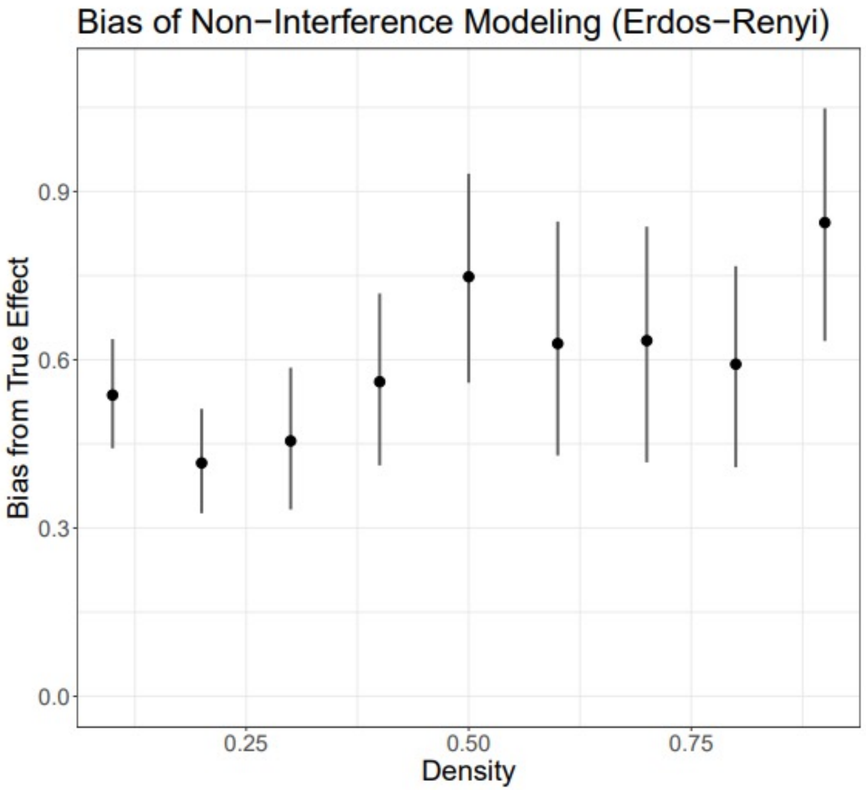


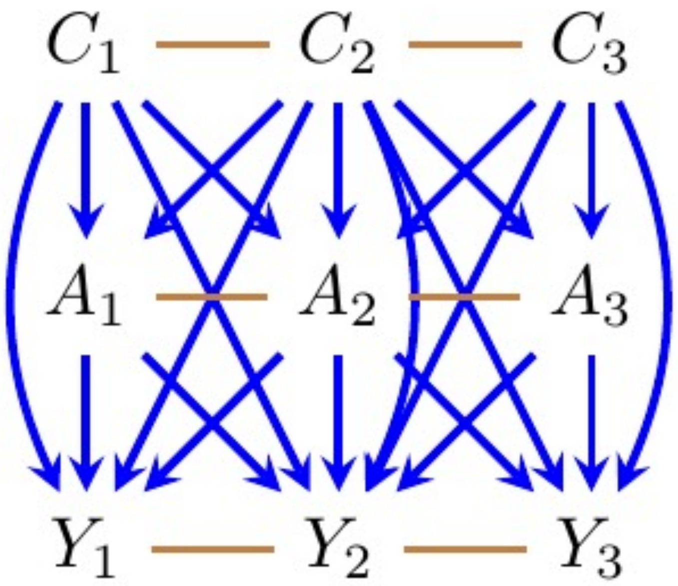
non causal
undirected
edges

CHAIN GRAPHS

Ogburn, Shpitser and Lee. *Causal inference, social networks and chain graphs*. JRSSB 2020.
Lauritzen & Richardson. *Chain Graph Models and Their Causal Interpretation*. JRSSB. 2002.

WHY DEPENDENCE-AWARE MODELING?[1]





Lee & Ogburn. *Network Dependence Can Lead to Spurious Associations and Invalid Inference*. Journal of American Statistical Association. 2020.

Sherman, Arbour, and Shpitser. *General Identification of Dynamic Treatment Regimes Under Interference*. AISTATS. 2020.

CHAIN GRAPHS

Undirected edges represent stable equilibrium between 2+ edges

'DAG of blocks' with 2-level factorization

$$V \leftarrow f_V(\mathcal{B}(V), \mathrm{pa}_{\mathcal{G}}(\mathcal{B}(V)), \epsilon_V)$$

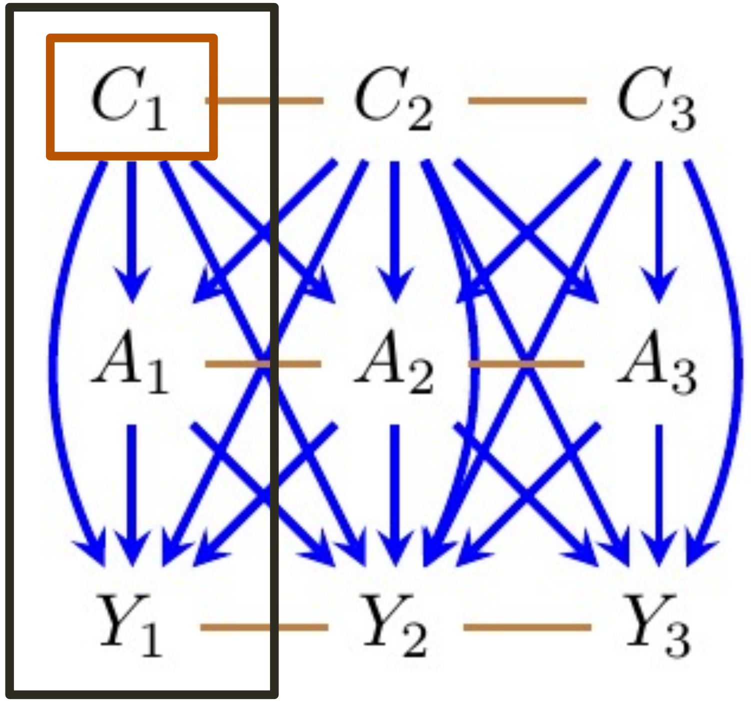
$$p(\mathbf{V}) = \prod_{\mathbf{B} \in \mathcal{B}(\mathcal{G})} p(\mathbf{B} | \mathrm{pa}_{\mathcal{G}}(\mathbf{B})) = \prod_{\mathbf{B} \in \mathcal{B}(\mathcal{G})} \frac{1}{Z(pa_{\mathcal{G}}(\mathbf{B}))} \prod_{\mathbf{C} \in \mathcal{C}^{\star}} \phi_{\mathbf{C}}(\mathbf{C}),$$

Lauritzen & Richardson. *Chain Graph Models and Their Causal Interpretation.* JRSSB. 2002.

DATA GENERATING PROCESS

Procedure 1 CG Data Generating Process

procedure CG-DGP($\mathcal{G}, \{f_B : B \in \mathbf{V}\}$)
  for each block $\mathbf{B}_i \in \mathcal{B}(\mathcal{G})$ **do**
    repeat
      for each variable $B_j \in \mathbf{B}_i$ **do**
        $B_j \leftarrow f_{B_j}(\mathbf{B}_i \setminus B_j, \mathrm{pa}_{\mathcal{G}}(\mathbf{B}_i), \epsilon_{B_j})$
    until equilibrium
  return V

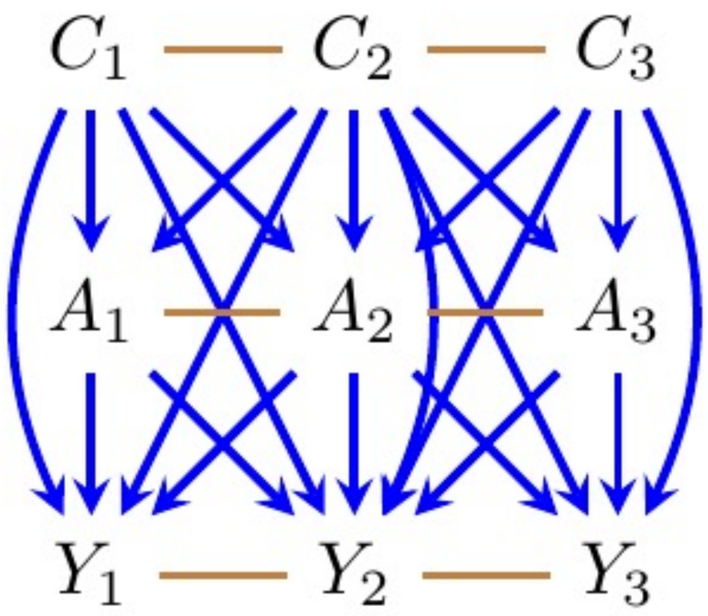


Lauritzen & Richardson. *Chain Graph Models and Their Causal Interpretation.* JRSSB. 2002.

IDENTIFICATION

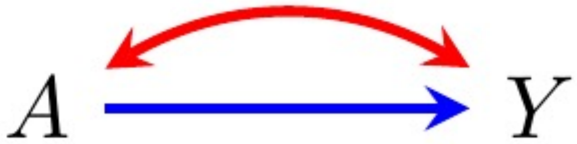
$$p(\mathbf{V}_C(\mathbf{a})) = \prod_{\mathbf{B} \in \mathcal{B}(\mathcal{G})} p(\mathbf{B} \setminus \mathbf{A} | \operatorname{pa}_{\mathcal{G}}(\mathbf{B}), \mathbf{B} \cap \mathbf{A})|_{\mathbf{A}=\mathbf{a}}$$

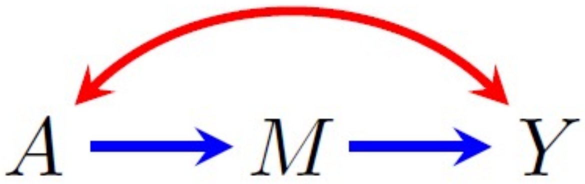
$$p(\mathbf{V}_D(\mathbf{a})) = \prod_{V \in \mathbf{V}_D \setminus \mathbf{A}} p(V | \operatorname{pa}_{\mathcal{G}}(V))|_{\mathbf{A}=\mathbf{a}}$$



Lauritzen & Richardson. *Chain Graph Models and Their Causal Interpretation.* JRSSB. 2002.

HANDLING LATENT VARIABLES





Acyclic Directed Mixed Graphs (ADMGs) – latent projection DAGs

- A↔️ B means A and B share a common cause

Markov Kernels

- ADMGs factorize as product of densities that relate *district* variables[1]

$$p(V) = \prod_{D \in \mathcal{D}(\mathcal{G})} q_D(D \mid \mathrm{pa}_{\mathcal{G}}(D)),$$

Richardson. *Markov Properties for Acyclic Directed Mixed Graphs*. Scandinavian Journal of Statistics. 2002.

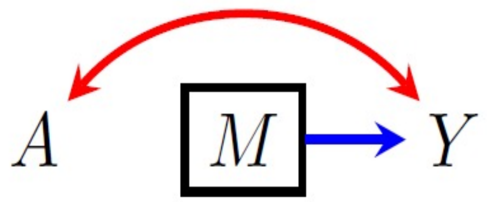
1: COMPARE TO DAGS, WHICH FACTORIZE ACCORDING TO UNIVARIATE CONDITIONALS

THE ID ALGORITHM

Fixing

- Truncated factorization provided notion of ‘fixing’ a variable in a DAG
- Corresponding notion in ADMGs – yields conditional ADMG (CADMG)
  - Reframe Pearl’s ‘graph surgery’ via fixing operator

$$\phi_V(\mathcal{G}) \qquad\qquad \phi_V(q; \mathcal{G})$$



Richardson, Robins and Shpitser. Nested *Markov Properties for Acyclic Directed Mixed Graphs*. *UAI*. 2012.

HANDLING LATENTS IN CHAIN GRAPHS

Segregation Property

- Do not permit ↔ and – edge at the same node
  - No known likelihood to support violations

Block-safeness

- Enforces segregation property in underlying chain graph
- Block-safe CGs can undergo latent projection operation to yield *segregated graph*

Shpitser. *Segregated Graphs and Marginals of Chain Graph Models*. NeurIPS. 2015.

THE ID ALGORITHM

For any disjoint subsets $\mathbf{Y}$, $\mathbf{A}$ of $\mathbf{V}$ in a latent projection $\mathcal{G}(\mathbf{V})$ representing a causal DAG $\mathcal{G}(\mathbf{V} \cup \mathbf{H})$, define $\mathbf{Y}^* \equiv \mathrm{an}_{\mathcal{G}(\mathbf{V})_{\mathbf{V}\setminus\mathbf{A}}}(\mathbf{Y})$. Then $p(\mathbf{Y}|\mathrm{do}(\mathbf{a}))$ is identified in $\mathcal{G}$ if *and only if* every set $\mathbf{D} \in \mathcal{D}(\mathcal{G}(\mathbf{V})_{\mathbf{Y}^*})$ is reachable (in fact, intrinsic). Moreover, if identification holds, we have [16]:

$$p(\mathbf{Y}|\mathrm{do}(\mathbf{a})) = \sum_{\mathbf{Y}^*\setminus\mathbf{Y}} \prod_{\mathbf{D}\in\mathcal{D}(\mathcal{G}(\mathbf{V})_{\mathbf{Y}^*})} \phi_{\mathbf{V}\setminus\mathbf{D}}(p(\mathbf{V}); \mathcal{G}(\mathbf{V}))|_{\mathbf{A}=\mathbf{a}}. \qquad (2)$$

Richardson, Robins and Shpitser. Nested *Markov Properties for Acyclic Directed Mixed Graphs. UAI*. 2012.

THE ID ALGORITHM

Y's ancestors are the only thing that is relevant for identifying effects on Y

For any disjoint subsets $\mathbf{Y}, \mathbf{A}$ of $\mathbf{V}$ in a latent projection $\mathcal{G}(\mathbf{V})$ representing a causal DAG $\mathcal{G}(\mathbf{V} \cup \mathbf{H})$, define $\mathbf{Y}^* \equiv \text{an}_{\mathcal{G}(\mathbf{V})_{\mathbf{V} \setminus \mathbf{A}}}(\mathbf{Y})$. Then $p(\mathbf{Y}|\text{do}(\mathbf{a}))$ is identified in $\mathcal{G}$ if *and only if* every set $\mathbf{D} \in \mathcal{D}(\mathcal{G}(\mathbf{V})_{\mathbf{Y}^*})$ is reachable (in fact, intrinsic). Moreover, if identification holds, we have [16]:

$$p(\mathbf{Y}|\text{do}(\mathbf{a})) = \sum_{\mathbf{Y}^* \setminus \mathbf{Y}} \prod_{\mathbf{D} \in \mathcal{D}(\mathcal{G}(\mathbf{V})_{\mathbf{Y}^*})} \phi_{\mathbf{V} \setminus \mathbf{D}}(p(\mathbf{V}); \mathcal{G}(\mathbf{V}))|_{\mathbf{A}=\mathbf{a}}. \quad (2)$$

Richardson, Robins and Shpitser. Nested *Markov Properties for Acyclic Directed Mixed Graphs. UAI.* 2012.

THE ID ALGORITHM

For any disjoint subsets $\mathbf{Y}, \mathbf{A}$ of $\mathbf{V}$ in a latent projection $\mathcal{G}(\mathbf{V})$ representing a causal DAG $\mathcal{G}(\mathbf{V} \cup \mathbf{H})$, define $\mathbf{Y}^* \equiv \mathrm{an}_{\mathcal{G}(\mathbf{V})_{\mathbf{V}\setminus\mathbf{A}}}(\mathbf{Y})$. Then $p(\mathbf{Y}|\mathrm{do}(\mathbf{a}))$ is identified in $\mathcal{G}$ if *and only if* every set $\mathbf{D} \in \mathcal{D}(\mathcal{G}(\mathbf{V})_{\mathbf{Y}^*})$ is reachable (in fact, intrinsic). Moreover, if identification holds, we have [16]:

$$p(\mathbf{Y}|\mathrm{do}(\mathbf{a})) = \sum_{\mathbf{Y}^*\setminus\mathbf{Y}} \prod_{\mathbf{D}\in\mathcal{D}(\mathcal{G}(\mathbf{V})_{\mathbf{Y}^*})} \phi_{\mathbf{V}\setminus\mathbf{D}}(p(\mathbf{V});\mathcal{G}(\mathbf{V}))|_{\mathbf{A}=\mathbf{a}}. \quad (2)$$

Algorithm is complete

Intuition: need to 'identify' (reach) each district

Richardson, Robins and Shpitser. Nested *Markov Properties for Acyclic Directed Mixed Graphs. UAI.* 2012.

THE ID ALGORITHM

For any disjoint subsets $\mathbf{Y}, \mathbf{A}$ of $\mathbf{V}$ in a latent projection $\mathcal{G}(\mathbf{V})$ representing a causal DAG $\mathcal{G}(\mathbf{V} \cup \mathbf{H})$, define $\mathbf{Y}^* \equiv \mathrm{an}_{\mathcal{G}(\mathbf{V})_{\mathbf{V} \setminus \mathbf{A}}}(\mathbf{Y})$. Then $p(\mathbf{Y}|\mathrm{do}(\mathbf{a}))$ is identified in $\mathcal{G}$ if *and only if* every set $\mathbf{D} \in \mathcal{D}(\mathcal{G}(\mathbf{V})_{\mathbf{Y}^*})$ is reachable (in fact, intrinsic). Moreover, if identification holds, we have [16]:

$$p(\mathbf{Y}|\mathrm{do}(\mathbf{a})) = \sum_{\mathbf{Y}^* \setminus \mathbf{Y}} \prod_{\mathbf{D} \in \mathcal{D}(\mathcal{G}(\mathbf{V})_{\mathbf{Y}^*})} \phi_{\mathbf{V} \setminus \mathbf{D}}(p(\mathbf{V}); \mathcal{G}(\mathbf{V}))|_{\mathbf{A}=\mathbf{a}}. \quad (2)$$

Marginalize

$$p(\mathbf{V}_D(\mathbf{a})) = \prod_{V \in \mathbf{V}_D \setminus \mathbf{A}} p(V|\,\mathrm{pa}_{\mathcal{G}}(V))|_{\mathbf{A}=\mathbf{a}}$$

Richardson, Robins and Shpitser. Nested *Markov Properties for Acyclic Directed Mixed Graphs*. *UAI*. 2012.

THE ID ALGORITHM

For any disjoint subsets $\mathbf{Y}, \mathbf{A}$ of $\mathbf{V}$ in a latent projection $\mathcal{G}(\mathbf{V})$ representing a causal DAG $\mathcal{G}(\mathbf{V} \cup \mathbf{H})$, define $\mathbf{Y}^* \equiv \mathrm{an}_{\mathcal{G}(\mathbf{V})_{\mathbf{V} \setminus \mathbf{A}}}(\mathbf{Y})$. Then $p(\mathbf{Y}|\mathrm{do}(\mathbf{a}))$ is identified in $\mathcal{G}$ if *and only if* every set $\mathbf{D} \in \mathcal{D}(\mathcal{G}(\mathbf{V})_{\mathbf{Y}^*})$ is reachable (in fact, intrinsic). Moreover, if identification holds, we have [16]:

$$p(\mathbf{Y}|\mathrm{do}(\mathbf{a})) = \sum_{\mathbf{Y}^* \setminus \mathbf{Y}} \boxed{\prod_{\mathbf{D} \in \mathcal{D}(\mathcal{G}(\mathbf{V})_{\mathbf{Y}^*})}} \boxed{\phi_{\mathbf{V} \setminus \mathbf{D}}(p(\mathbf{V}); \mathcal{G}(\mathbf{V}))|_{\mathbf{A}=\mathbf{a}}}. \quad (2)$$

$$p(\mathbf{V}_D(\mathbf{a})) = \boxed{\prod_{V \in \mathbf{V}_D \setminus \mathbf{A}}} \boxed{p(V|\,\mathrm{pa}_{\mathcal{G}}(V))|_{\mathbf{A}=\mathbf{a}}}$$

Richardson, Robins and Shpitser. Nested *Markov Properties for Acyclic Directed Mixed Graphs. UAI.* 2012.

HANDLING LATENTS IN CHAIN GRAPHS

Segregation Property

- Do not permit ↔ and – edge at the same node
  - No known likelihood to support violations

Block-safeness

  - Enforces segregation property in underlying chain graph
  - Block-safe CGs can undergo latent projection operation to yield *segregated graph*

Shpitser. *Segregated Graphs and Marginals of Chain Graph Models*. NeurIPS. 2015.

HANDLING LATENTS IN CHAIN GRAPHS

Factorization-Blocks and districts

Conditional Chain Graph

$$q(\mathbf{B}^\star |\, \mathrm{pa}^s_{\mathcal{G}}(\mathbf{B}^\star)) \;=\; \prod_{\mathbf{B}\in\mathcal{B}^{nt}(\mathcal{G})} p(\mathbf{B}|\, \mathrm{pa}_{\mathcal{G}}(\mathbf{B}))$$

CADMG

$$q(\mathbf{D}^\star |\, \mathrm{pa}^s_{\mathcal{G}}(\mathbf{D}^\star)) \;=\; \frac{p(\mathbf{V})}{q(\mathbf{B}^\star |\, \mathrm{pa}^s_{\mathcal{G}}(\mathbf{B}^\star))}$$

Shpitser. *Segregated Graphs and Marginals of Chain Graph Models*. NeurIPS. 2015.

THE SEGREGATED GRAPH ID ALGORITHM

Theorem 2 *Assume $\mathcal{G}(\mathbf{V} \cup \mathbf{H})$ is a causal CG, where $\mathbf{H}$ is block-safe. Fix disjoint subsets $\mathbf{Y}, \mathbf{A}$ of $\mathbf{V}$. Let $\mathbf{Y}^* = \text{ant}_{\mathcal{G}(\mathbf{V})_{\mathbf{V} \setminus \mathbf{A}}} \mathbf{Y}$. Then $p(\mathbf{Y}|do(\mathbf{a}))$ is identified from $p(\mathbf{V})$ if and only if every element in $\mathcal{D}(\widetilde{\mathcal{G}}^d)$ is reachable in $\mathcal{G}^d$, where $\widetilde{\mathcal{G}}^d$ is the induced CADMG of $\mathcal{G}(\mathbf{V})_{\mathbf{Y}^*}$.*

Moreover, if $p(\mathbf{Y}|do(\mathbf{a}))$ is identified, it is equal to

$$\sum_{\mathbf{Y}^* \setminus \mathbf{Y}} \left[\prod_{\mathbf{D} \in \mathcal{D}(\widetilde{\mathcal{G}}^d)} \phi_{\mathbf{D}^* \setminus \mathbf{D}}(q(\mathbf{D}^* | \text{pa}_{\mathcal{G}(\mathbf{V})}(\mathbf{D}^*)); \mathcal{G}^d) \right] \left[\prod_{\mathbf{B} \in \mathcal{B}(\widetilde{\mathcal{G}}^b)} p(\mathbf{B} \setminus \mathbf{A} | \text{pa}_{\mathcal{G}(\mathbf{V})_{\mathbf{Y}^*}}(\mathbf{B}), \mathbf{B} \cap \mathbf{A}) \right] \Bigg|_{\mathbf{A}=\mathbf{a}}$$

where $q(\mathbf{D}^* | \text{pa}_{\mathcal{G}(\mathbf{V})}(\mathbf{D}^*)) = p(\mathbf{V}) / (\prod_{\mathbf{B} \in \mathcal{B}^{nt}(\mathcal{G}(\mathbf{V}))} p(\mathbf{B} | \text{pa}_{\mathcal{G}(\mathbf{V})}(\mathbf{B}))$, and $\widetilde{\mathcal{G}}^b$ is the induced CCG of $\mathcal{G}(\mathbf{V})_{\mathbf{Y}^*}$.

$$p(\mathbf{Y}|\text{do}(\mathbf{a})) = \sum_{\mathbf{Y}^* \setminus \mathbf{Y}} \prod_{\mathbf{D} \in \mathcal{D}(\mathcal{G}(\mathbf{V})_{\mathbf{Y}^*})} \phi_{\mathbf{V} \setminus \mathbf{D}}(p(\mathbf{V}); \mathcal{G}(\mathbf{V}))|_{\mathbf{A}=\mathbf{a}}.$$

$$p(\mathbf{V}_C(\mathbf{a})) = \prod_{\mathbf{B} \in \mathcal{B}(\mathcal{G})} p(\mathbf{B} \setminus \mathbf{A} | \text{pa}_{\mathcal{G}}(\mathbf{B}), \mathbf{B} \cap \mathbf{A})|_{\mathbf{A}=\mathbf{a}}$$

Shpitser. *Segregated Graphs and Marginals of Chain Graph Models.* NeurIPS. 2015.
Sherman & Shpitser. *Identification of Causal Effects from Dependent Data.* NeurIPS. 2018.

GENERALIZING NODE INTERVENTIONS

Policy analysis

- Want to evaluate treatments ‘tailored’ to the subject
- Ultimately want to perform policy optimization[1]
- Intervene with a function $f_A(\mathbf{W}_A)$ where $\mathbf{W}_A \subseteq \acute{\mathbf{V}}_{\prec A}$
- Pearlian ‘graph’ surgery
  - Remove edges *into* A, add edges from W to A

Sherman, Arbour, and Shpitser. *General Identification of Dynamic Treatment Regimes Under Interference*. AISTATS. 2020.

THE POLICY ID ALGORITHM

$p(\mathbf{Y}(\mathbf{f_A}))$ is identified in $\mathcal{G}$ if and only if $p(\mathbf{Y}^\star(\mathbf{a}))$ is identified in $\mathcal{G}$. If it is identified, then

$$p(\mathbf{Y}(\mathbf{f_A})) = \sum_{(\mathbf{Y}^\star \cup \mathbf{A}) \setminus \mathbf{Y}} \prod_{\mathbf{D} \in \mathcal{D}(\mathcal{G}_{\mathbf{Y}^\star})} \phi_{\mathbf{V} \setminus \mathbf{D}}(p(\mathbf{V}); \mathcal{G})|_{\tilde{\mathbf{a}}_{\mathrm{pa}^s_{\mathcal{G}}(\mathbf{D}) \cap \mathbf{A}}}$$

where $\tilde{\mathbf{a}}_{\mathrm{pa}^s_{\mathcal{G}}(\mathbf{D}) \cap \mathbf{A}} = \{A = f_A(\mathbf{W}_A) | A \in \mathrm{pa}_{\mathcal{G}}(\mathbf{D}) \cap \mathbf{A}\}$ if $\mathrm{pa}_{\mathcal{G}}(\mathbf{D}) \cap \mathbf{A} \neq \emptyset$ and $\tilde{\mathbf{a}} = \emptyset$ otherwise.

$$p(\mathbf{Y} | \mathrm{do}(\mathbf{a})) = \sum_{\mathbf{Y}^* \setminus \mathbf{Y}} \prod_{\mathbf{D} \in \mathcal{D}(\mathcal{G}(\mathbf{V})_{\mathbf{Y}^*})} \phi_{\mathbf{V} \setminus \mathbf{D}}(p(\mathbf{V}); \mathcal{G}(\mathbf{V}))|_{\mathbf{A}=\mathbf{a}}.$$

Sherman, Arbour, and Shpitser. *General Identification of Dynamic Treatment Regimes Under Interference*. AISTATS. 2020.

SEGREGATION-PRESERVING POLICIES

Need to assume policies maintain segregation property

- Cannot induce a (partially-directed) cycle
- Allow for a variety of intervention types
  - Add/remove directed edge
  - Modify existing (directed or undirected) edge

Procedure 2 Obtaining $\mathcal{G}_{\mathbf{f_A}}$ from $\mathcal{G}$

```
procedure InterveneGraph(𝒢, f_A(Z_A))
    Initialize 𝒢_{f_A} ← 𝒢
    for each A ∈ A do
        Replace all V − A with A → V in 𝒢_{f_A}
        Remove all · → A, · ↔ A from 𝒢_{f_A}
        Add edges Z_A → A in 𝒢_{f_A}
    for each V_i, V_j ∈ V do
        if V_i → V_j and V_j → V_i in 𝒢_{f_A} then
            Remove V_i → V_j and V_j → V_i from 𝒢_{f_A}
            Add V_i − V_j in 𝒢_{f_A}
    return 𝒢_{f_A}
```

Sherman, Arbour, and Shpitser. *General Identification of Dynamic Treatment Regimes Under Interference*. AISTATS. 2020.

POLICY ID FOR SEGREGATED GRAPHS

Theorem 1 *Let $\mathcal{G}(\mathbf{V} \cup \mathbf{H})$ be a causal LV-CG with $\mathbf{H}$ block-safe, and a topological order $\prec$. Fix disjoint $\mathbf{Y}, \mathbf{A} \subseteq \mathbf{V}$. Let $\mathbf{f_A}(\mathbf{Z_A})$ be a segregation preserving policy set. Let $\mathbf{Y}^\star \equiv \text{ant}_{\mathcal{G}_{\mathbf{f_A}}}(\mathbf{Y}) \setminus \mathbf{A}$. Let $\mathcal{G}^d, G^d$ be the induced CADMGs on $\mathcal{G}_{\mathbf{f_A}}$ and $\mathcal{G}_{\mathbf{Y}^\star}$, and $\tilde{G}^b$ the induced CCG on $\mathcal{G}_{\mathbf{Y}^\star}$. Let $q(\mathbf{D}^\star | \text{pa}^s_{\mathcal{G}_{\mathbf{f_A}}}(\mathbf{D}^\star)) = \prod_{\mathbf{D} \in \mathcal{G}_{\mathbf{f_A}}} q(\mathbf{D} | \text{pa}^s_{\mathcal{G}_{\mathbf{f_A}}}(\mathbf{D}))$, where $q(\mathbf{D} | \text{pa}^s_{\mathcal{G}_{\mathbf{f_A}}}(\mathbf{D})) = \prod_{D \in \mathbf{D}} p(D | \mathbf{V}_{\prec D})$ if $\mathbf{D} \cap \mathbf{A} = \emptyset$ and $q = f_A(\mathbf{Z}_A)$ if $\mathbf{D} \cap \mathbf{A} \neq \emptyset$. $p(\mathbf{Y}(\mathbf{f_A}(\mathbf{Z_A})))$ is identified in $\mathcal{G}$ if and only if $p(\mathbf{Y}^\star(\mathbf{a}))$ is identified in $\mathcal{G}$ for the unrestricted class of policies. If identified, $p(\mathbf{Y}(\mathbf{f_A}(\mathbf{Z_A}))) =$*

$$\sum_{\{\mathbf{Y}^\star \cup \mathbf{A}\} \setminus \mathbf{Y}} \left[\prod_{\mathbf{B} \in \mathcal{B}(\tilde{\mathcal{G}}^b)} p^\star(\mathbf{B} | \text{pa}_{\mathcal{G}_{\mathbf{f_A}}}(\mathbf{B})) \right] \times \left[\prod_{\mathbf{D} \in \mathcal{D}(\tilde{\mathcal{G}}^d)} \phi_{\mathbf{D}^\star \setminus \mathbf{D}}(q(\mathbf{D}^\star | \text{pa}^s_{\mathcal{G}_{\mathbf{f_A}}}(\mathbf{D}^\star)); \mathcal{G}^d) \right] \Bigg|_{\mathbf{A} = \tilde{\mathbf{a}}} \quad (11)$$

where (a) $\tilde{\mathbf{a}} = \{A = f_A(\mathbf{Z}_A) : A \in \text{pa}_{\mathcal{G}_{\mathbf{f_A}}}(\mathbf{D}) \cap \mathbf{A}\}$ if $\text{pa}_{\mathcal{G}_{\mathbf{f_A}}}(\mathbf{D}) \cap \mathbf{A} \neq \emptyset$ and $\tilde{\mathbf{a}}_\mathbf{D} = \emptyset$ otherwise, and (b) $p^\star$ is obtained by running Procedure 1 over functions $g_{B_i}(B_{-i}, \text{pa}_{\mathcal{G}_{\mathbf{f_A}}}(B_i), \epsilon_{B_i})$ where $g_{B_i} \in \mathbf{f_A}$ if $B_i \in \mathbf{A}$ and g_{B_i} is given by the observed distribution if $B_i \notin \mathbf{A}$[2].

Sherman, Arbour, and Shpitser. *General Identification of Dynamic Treatment Regimes Under Interference*. AISTATS. 2020.

EASY

Modeling feedback

Modeling latent variables

Identification

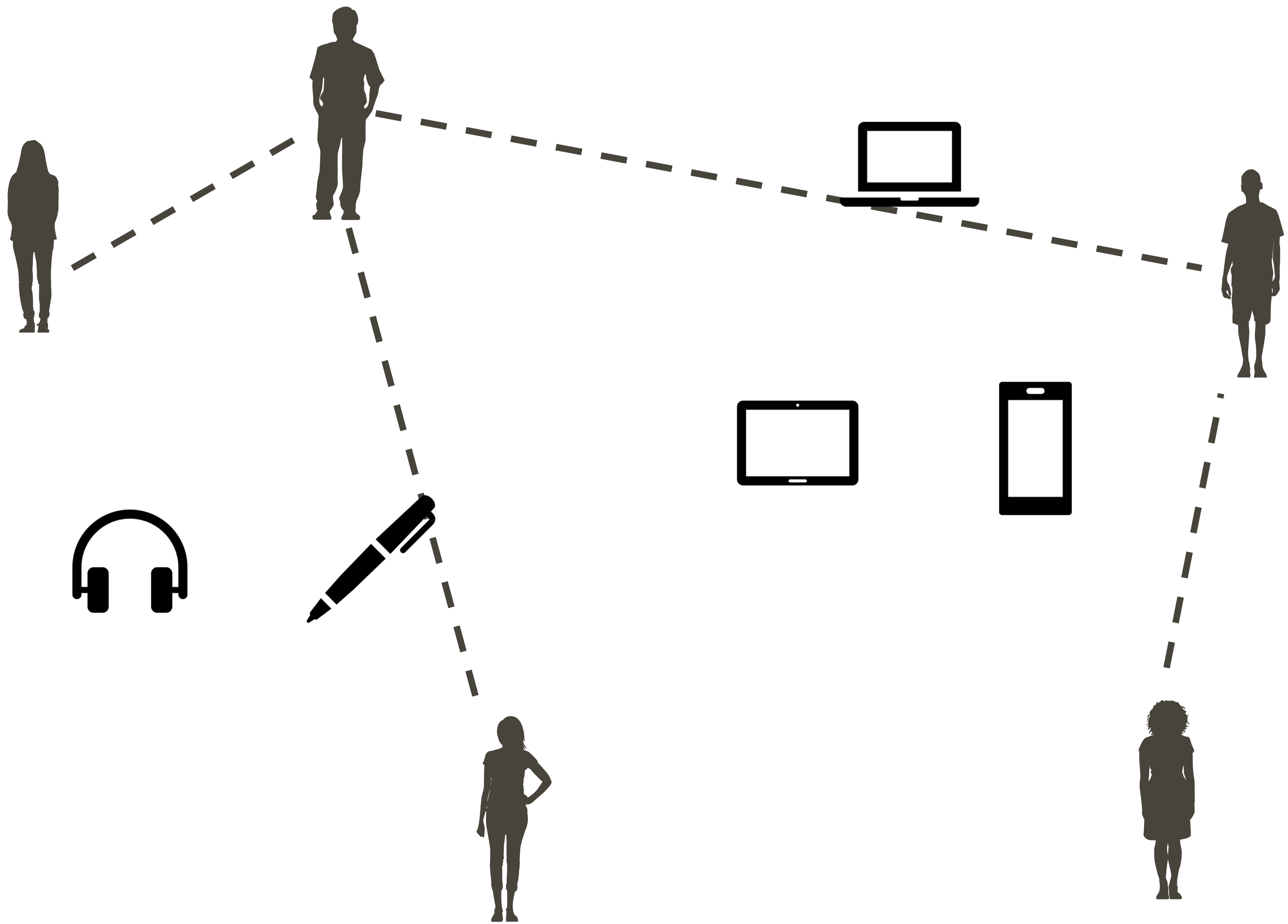
HARD

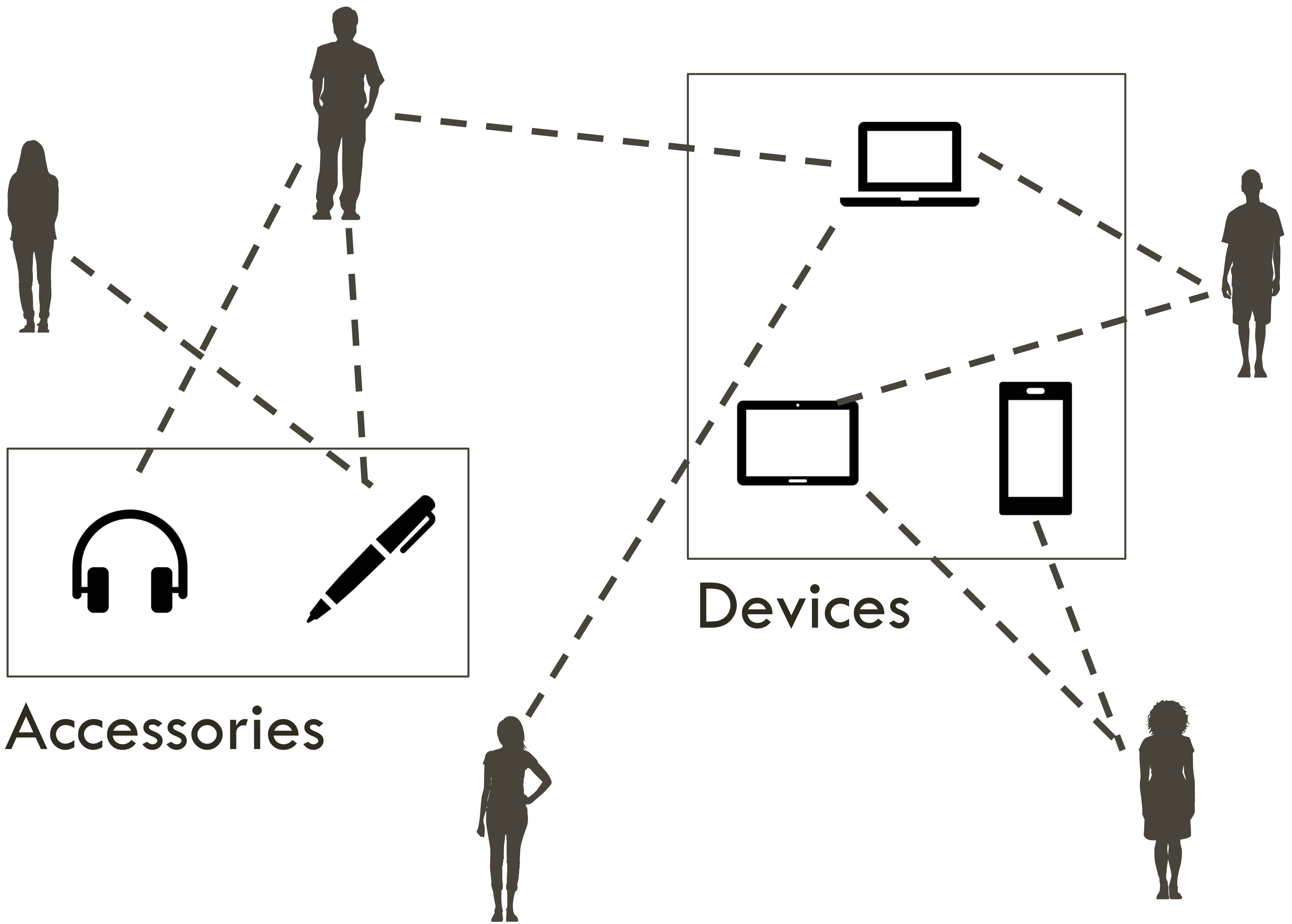
Expensive–Gibbs sampling is required for inference

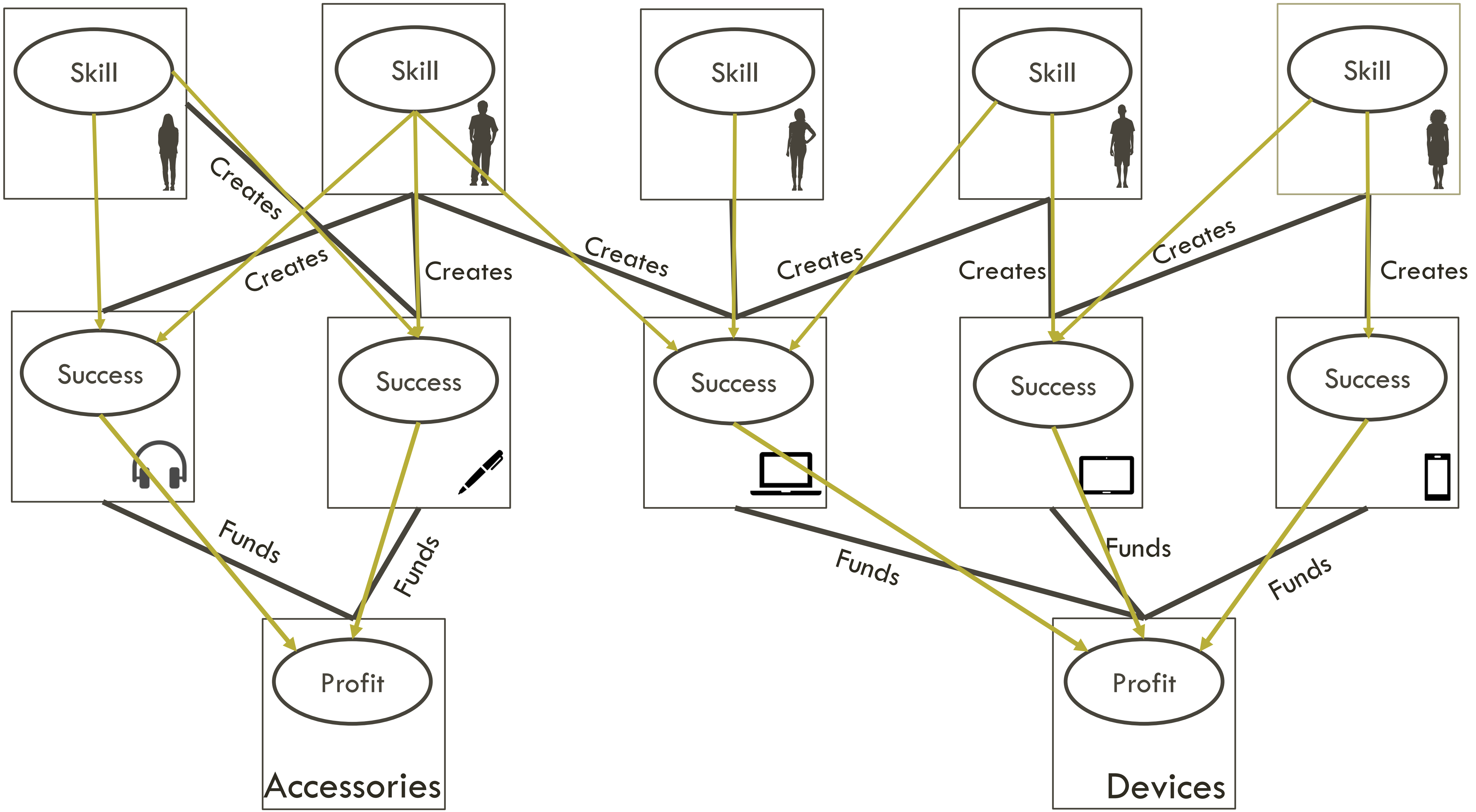
Difficult to represent interventions on distributions

COUNTERFACTUALS & CAUSAL EFFECTS IN OBSERVATIONAL NETWORK DATA

Multi-relational data and abstract ground graphs




Accessories
Devices

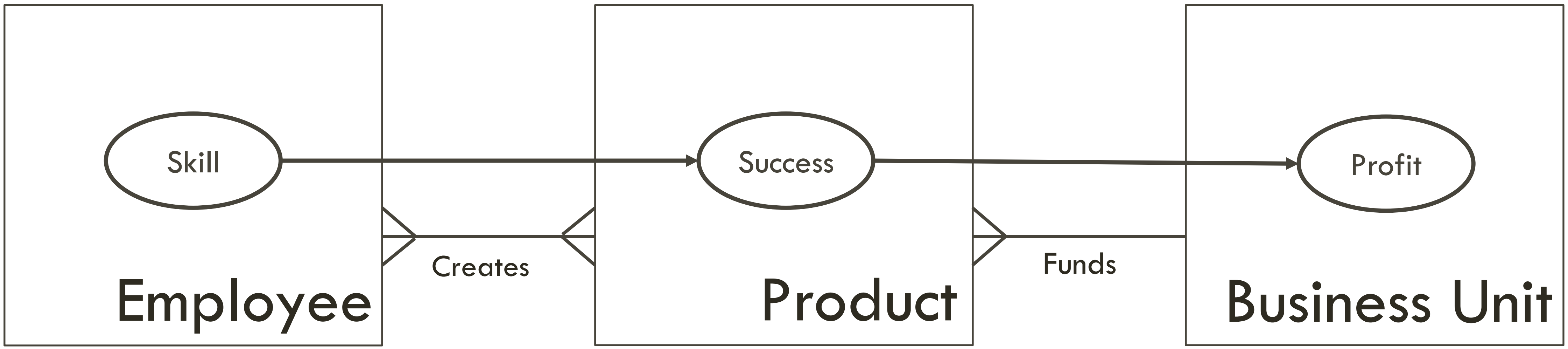

Skill
Skill
Skill
Skill
Skill
Creates
Creates
Creates
Creates
Creates
Creates
Creates
Creates
Success
Success
Success
Success
Success
Funds
Funds
Funds
Funds
Funds
Profit
Profit
Accessories
Devices

TEMPLATES

Assume shared marginal and conditional distributions

Allows a general model which represents relationships and dependencies more abstractly

Getoor, Friedman, Koller & Pfeffer. *Learning Probabilistic Relational Models. IJCAI*. 1999.



Getoor, Friedman, Koller & Pfeffer. *Learning Probabilistic Relational Models. IJCAI*. 1999.

OVERVIEW OF TEMPLATE MODELS

Schema $\xrightarrow{\text{Instantiate}}$ Skeleton

Model $\longrightarrow$ Ground Graph

Getoor, Friedman, Koller & Pfeffer. *Learning Probabilistic Relational Models. IJCAI.* 1999.

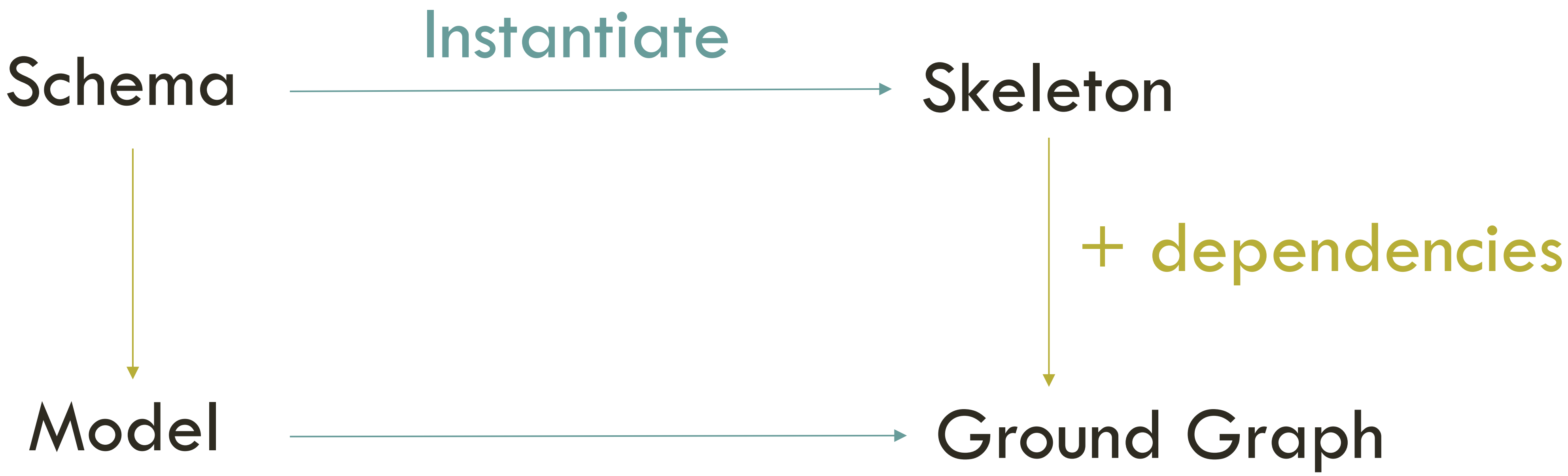
OVERVIEW OF TEMPLATE MODELS

$\text{Schema} \longrightarrow \text{Model}$

$\text{Skeleton} \xrightarrow{+\ \text{dependencies}} \text{Ground Graph}$

Getoor, Friedman, Koller & Pfeffer. *Learning Probabilistic Relational Models. IJCAI.* 1999.

OVERVIEW OF TEMPLATE MODELS

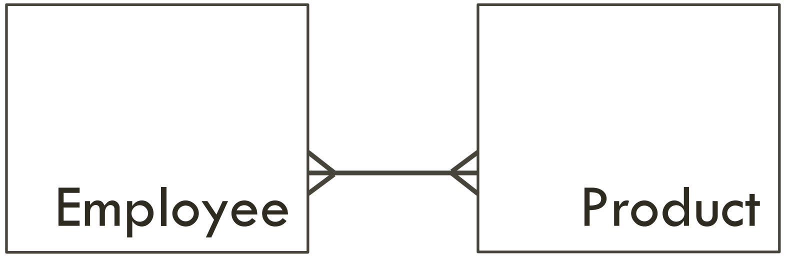


Getoor, Friedman, Koller & Pfeffer. *Learning Probabilistic Relational Models. IJCAI.* 1999.

RELATIONAL PATHS

Products an Employee works on

[Employee, Product]



Business units an Employee works in

[Employee, Product, Business Unit]

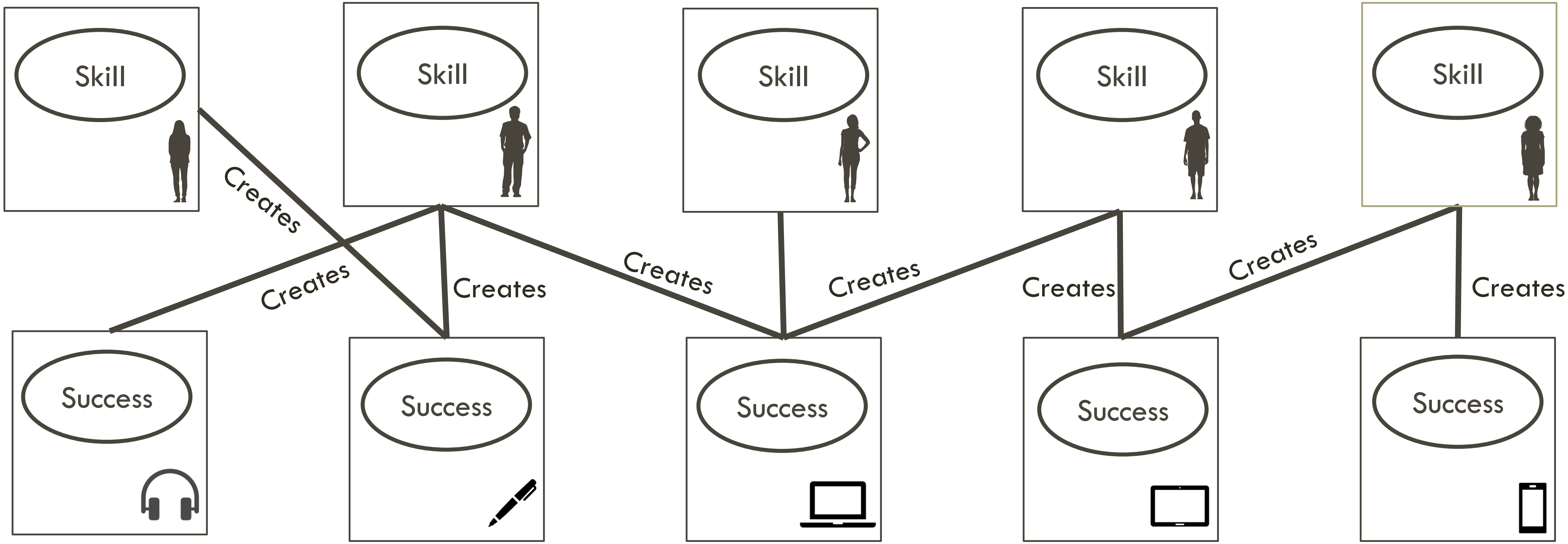
An employee's coworkers

[Employee, Product, Employee]

Heckerman, Meek, and Killer. Probablistic Models for Relational Data. MSR Tech Report. 2004.

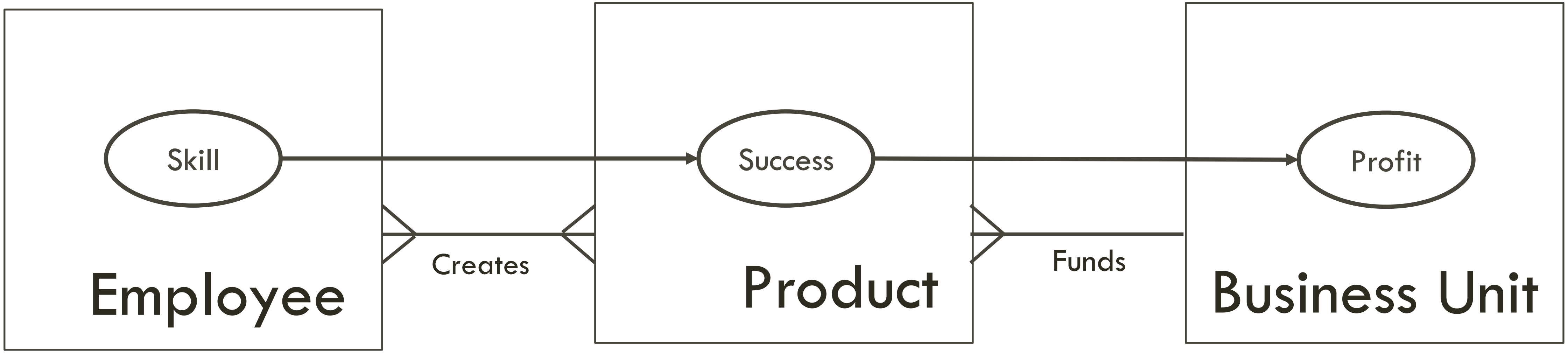
An employee's coworkers

[Employee, Product, Employee]



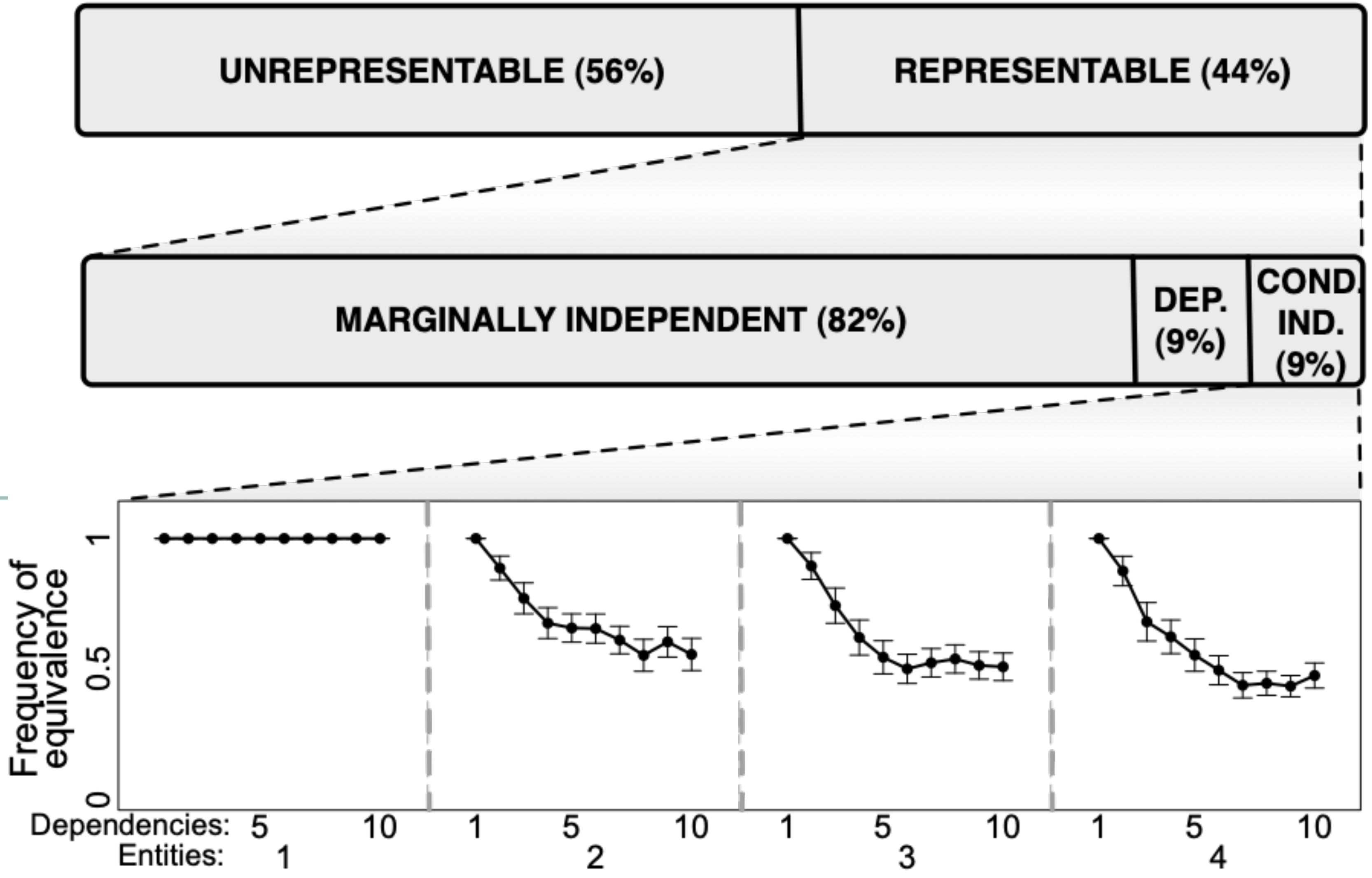
employee coworkers

D-SEPARATION ON TEMPLATES

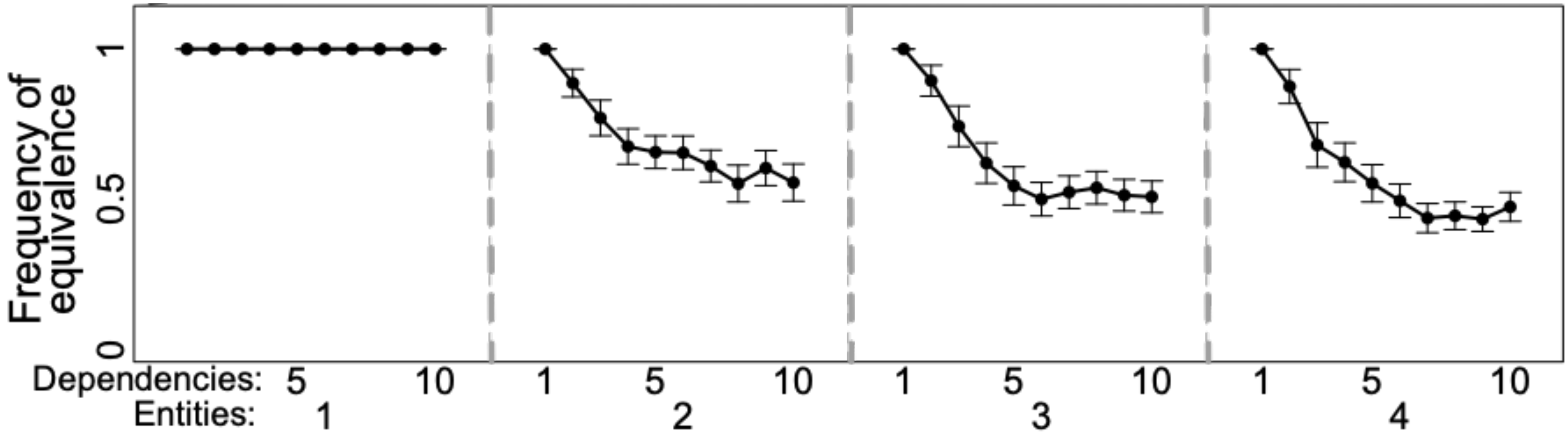


Maier, Marazopoulou, and Jensen. *Reasoning about Independence in Probabilistic Models of Relational Data. Arxiv*. 2013.
Lee and Hanovar. *A Characterization of Markov Equivalence Classes of Relational Causal Models under Path Semantics*. UAI. 2016

D-separation on templates often fails

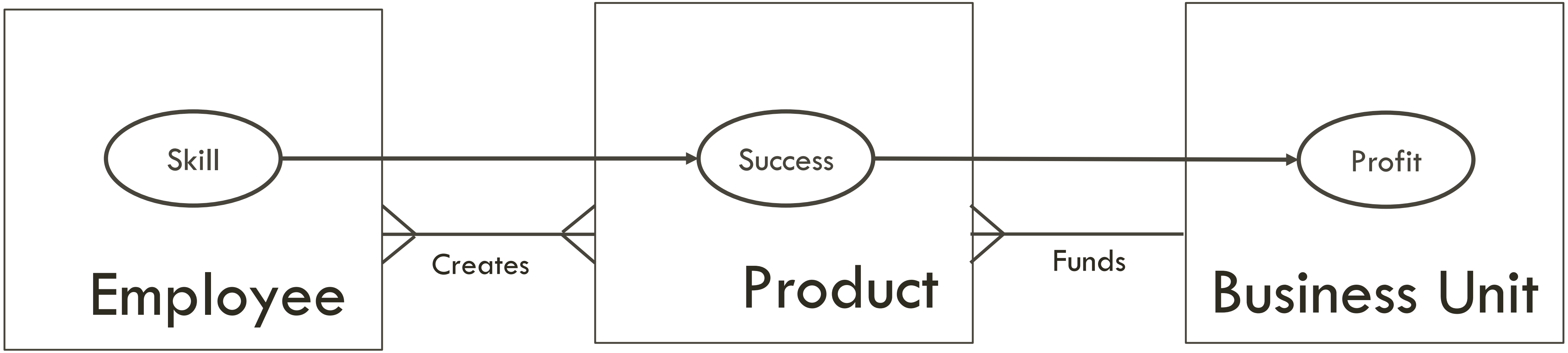


Maier, Marazopoulou, and Jensen. *Reasoning about Independence in Probabilistic Models of Relational Data. Arxiv*. 2013.
Lee and Hanovar. *A Characterization of Markov Equivalence Classes of Relational Causal Models under Path Semantics*. UAI. 2016



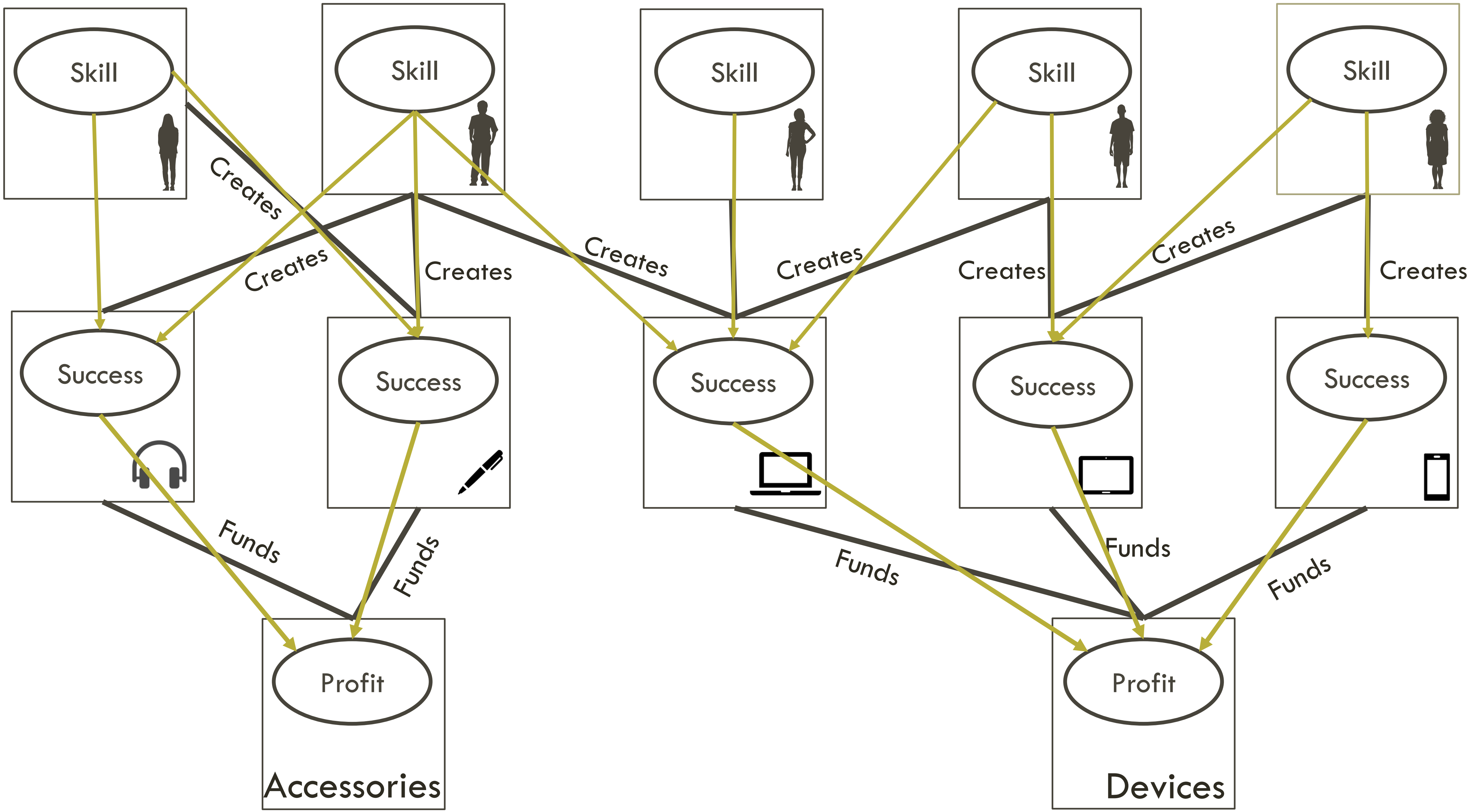
Maier, Marazopoulou, and Jensen. *Reasoning about Independence in Probabilistic Models of Relational Data. Arxiv*. 2013.
Lee and Hanovar. *A Characterization of Markov Equivalence Classes of Relational Causal Models under Path Semantics*. UAI. 2016

D-SEPARATION ON TEMPLATES



[Employee, Product, Employee].Skill $\coprod$ [Employee].Skill

Maier, Marazopoulou, and Jensen. *Reasoning about Independence in Probabilistic Models of Relational Data.* *Arxiv.* 2013.
Lee and Hanovar. *A Characterization of Markov Equivalence Classes of Relational Causal Models under Path Semantics.* UAI. 2016

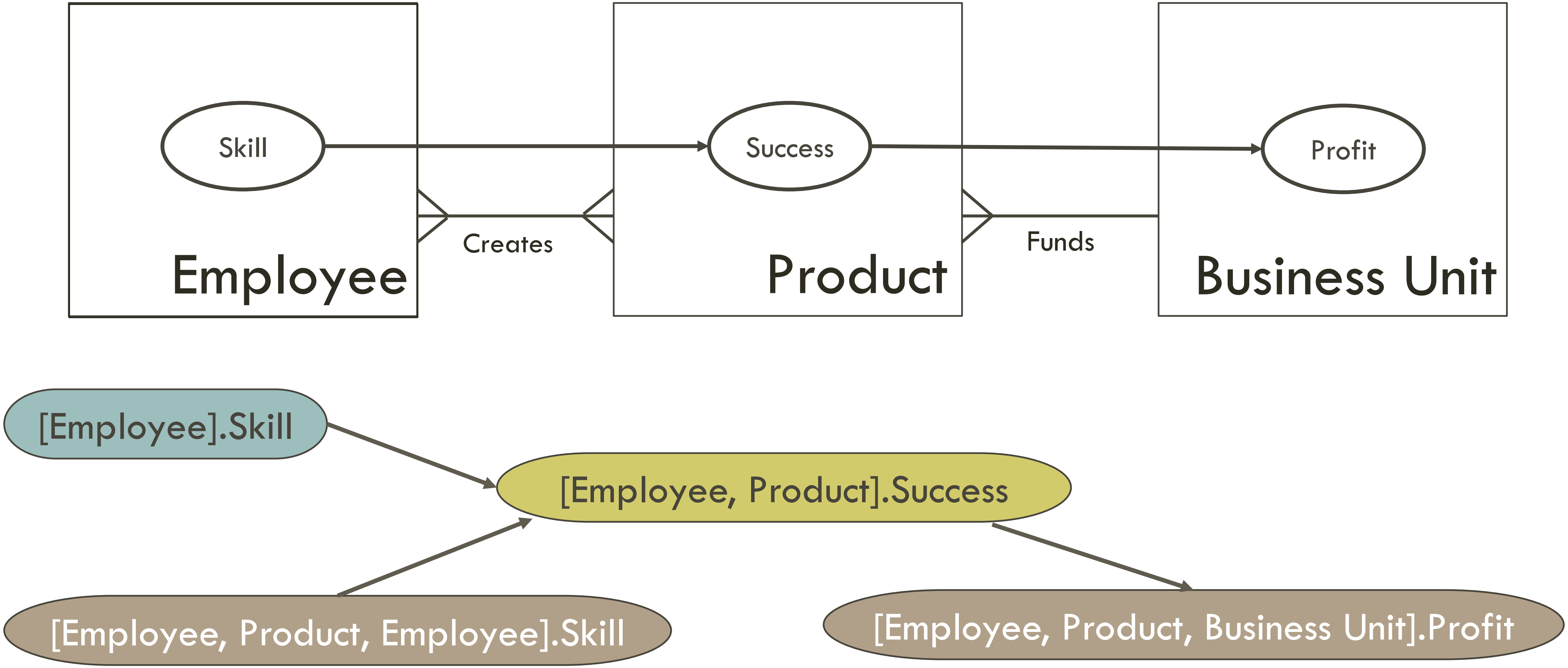

Skill
Skill
Skill
Skill
Skill
Creates
Creates
Creates
Creates
Creates
Creates
Creates
Creates
Success
Success
Success
Success
Success
Funds
Funds
Funds
Funds
Funds
Profit
Profit
Accessories
Devices

HOW DO WE FIND AN INTERMEDIATE REPRESENTATION THAT ALLOWS FOR D-SEPARATION?

ABSTRACT GROUND GRAPHS

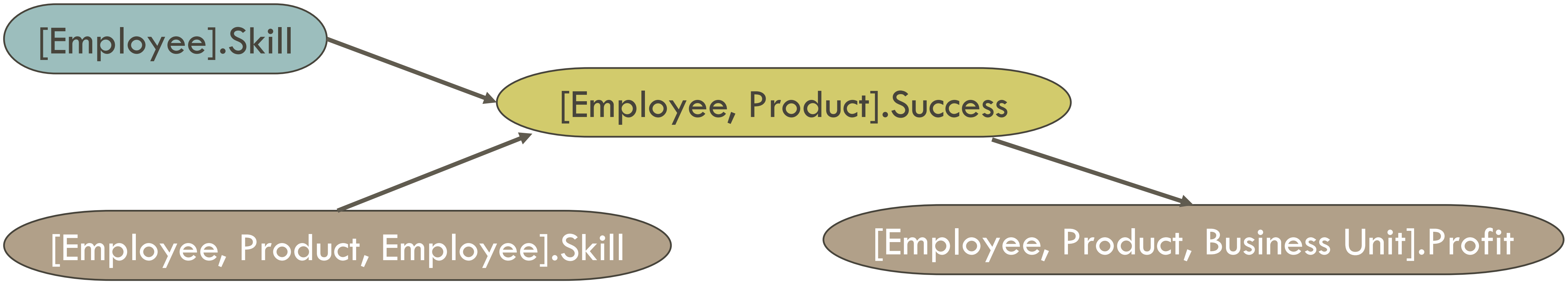
Lifted representation
with d-separation semantics

EMPLOYEE PERSPECTIVE | Hop threshold = 2



Maier, Marazopoulou, and Jensen. *Reasoning about Independence in Probabilistic Models of Relational Data. Arxiv.* 2013.
Lee and Hanovar. *A Characterization of Markov Equivalence Classes of Relational Causal Models under Path Semantics.* UAI. 2016

AGGS INHERIT THE PROPERTIES OF BAYES NETS

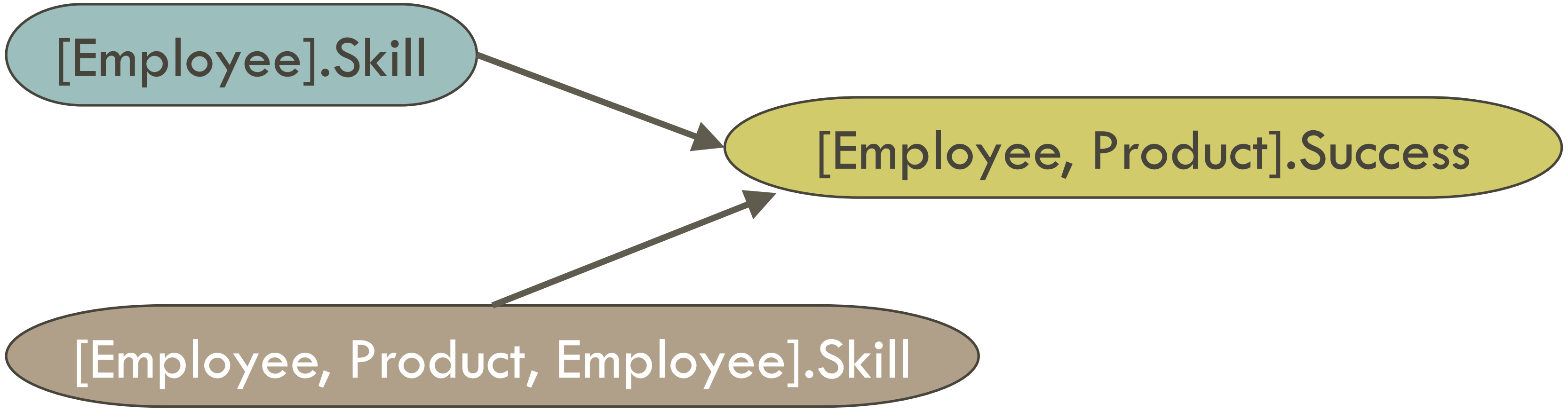


d-separation and identification theory from Bayesian networks can be applied directly.

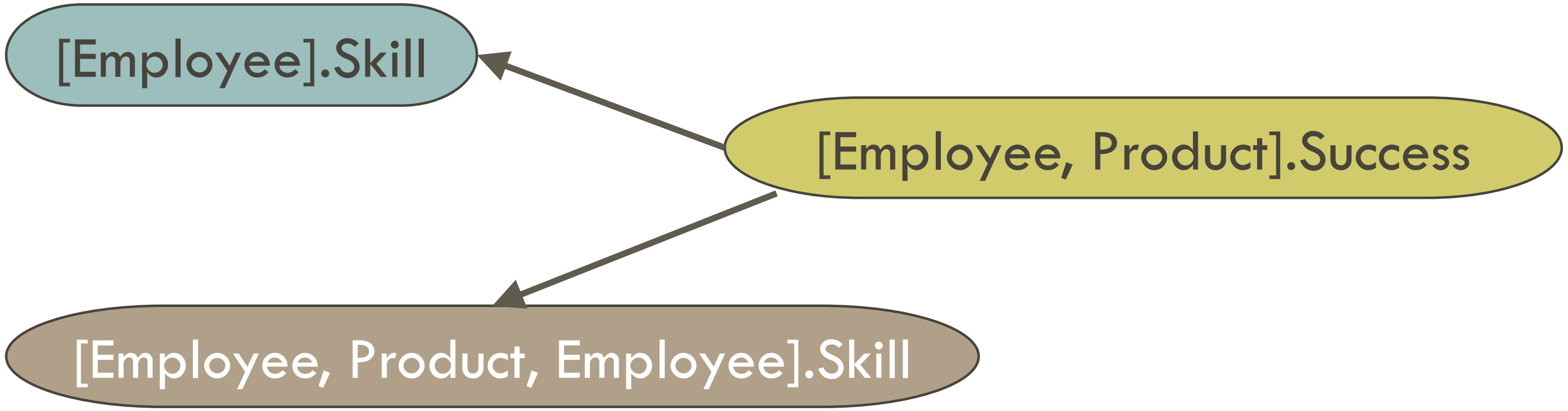
Maier, Marazopoulou, and Jensen. *Reasoning about Independence in Probabilistic Models of Relational Data. Arxiv*. 2013.
Arbour, Garant, and Jensen. *Inferring Network Effects from Observational Data*. KDD. 2016.

RELATIONAL BIVARIATE ORIENTATION

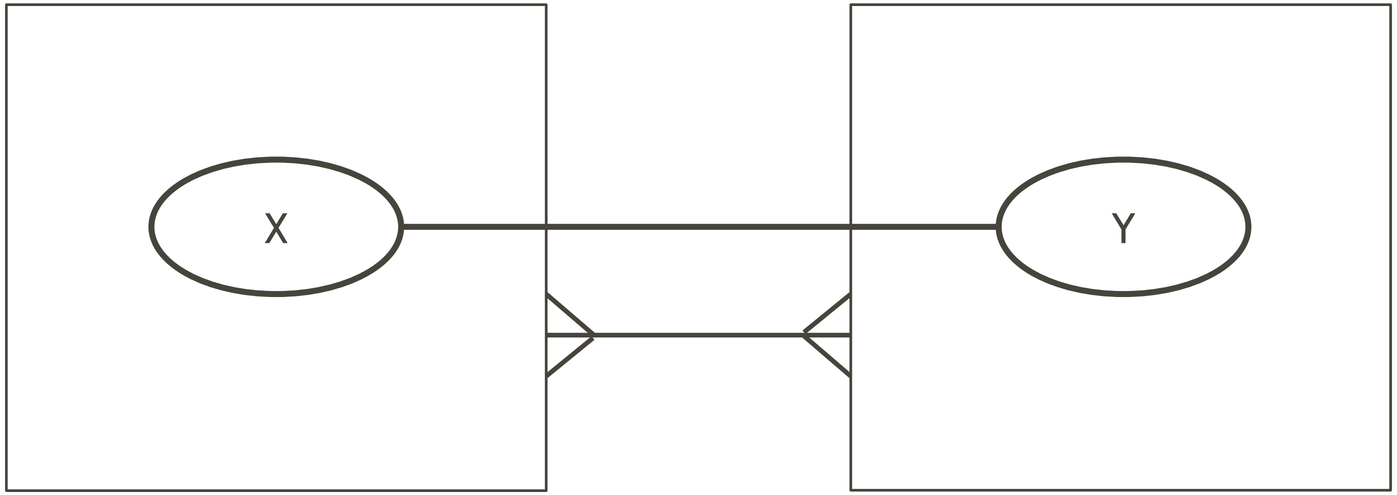
Unshielded collider



Non-collider



Maier, Marazopoulou, Arbour, and Jensen. *A Sound and Complete Algorithm for Learning Causal Models from Relational Data*. UAI. 2013.

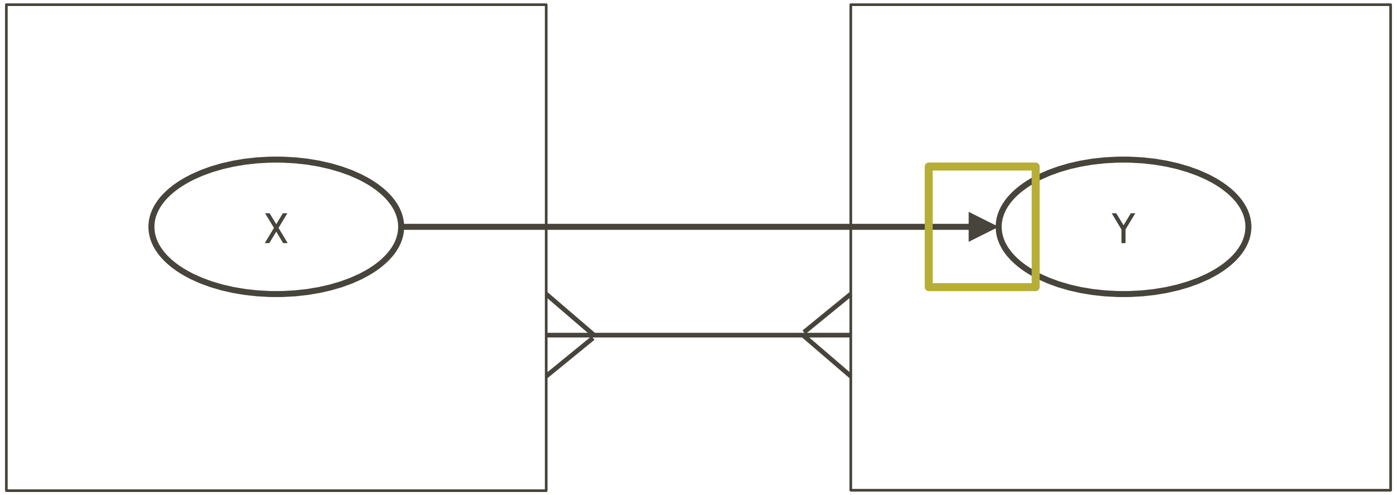


Compare:

$$cov([A].X, [A, B].Y)$$
$$cov([B].Y, [B, A].X)$$

INFERRING DIRECTION OF RELATIONAL DEPENDENCIES DIRECTLY

Arbour, Marazopoulou, and Jensen. *Inferring Causal Direction from Relational Data*. UAI. 2016.

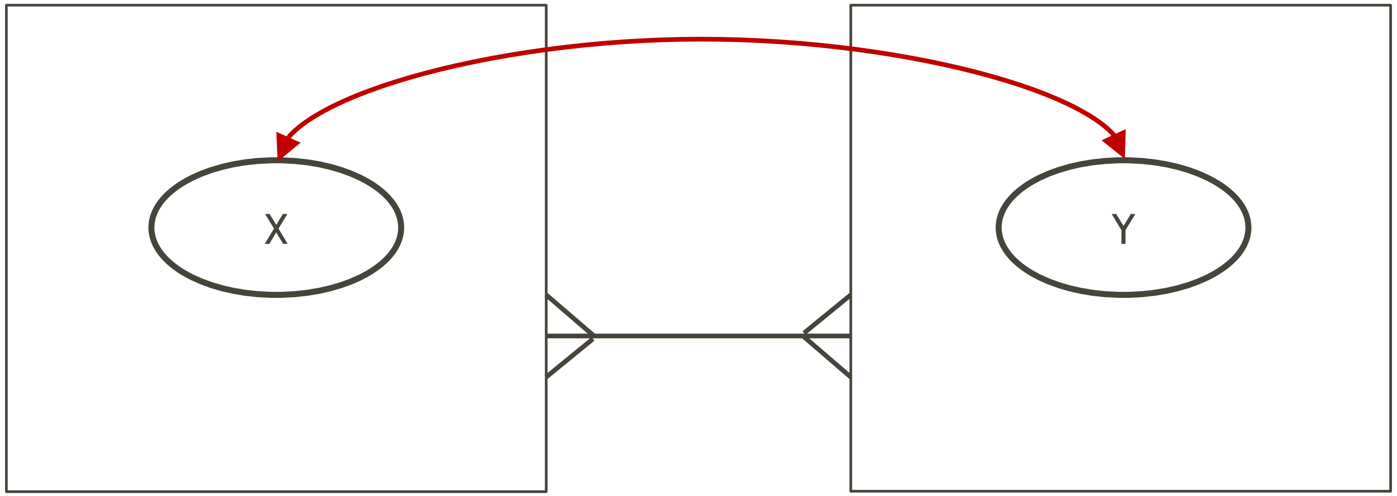


Larger covariance is true direction

Compare: $cov([A].X,[A,B].Y)$
$cov([B].Y,[B,A].X)$

INFERRING DIRECTION OF RELATIONAL DEPENDENCIES DIRECTLY

Arbour, Marazopoulou, and Jensen. *Inferring Causal Direction from Relational Data*. UAI. 2016.



Sufficient for detecting a latent confounder

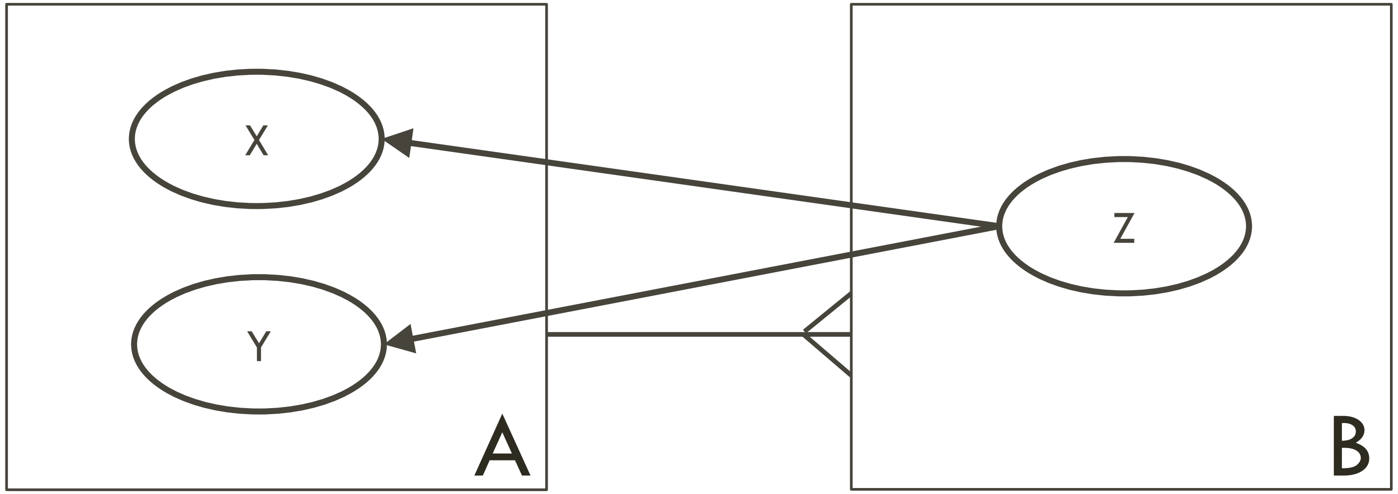
Compare: $cov([B,A].X,[A,B].Y)$ $cov([B].Y,[B,A].X)$ and $cov([B,A].X,[A,B].Y)$ $cov([A].X,[A,B].Y)$

INFERRING DIRECTION OF RELATIONAL DEPENDENCIES DIRECTLY

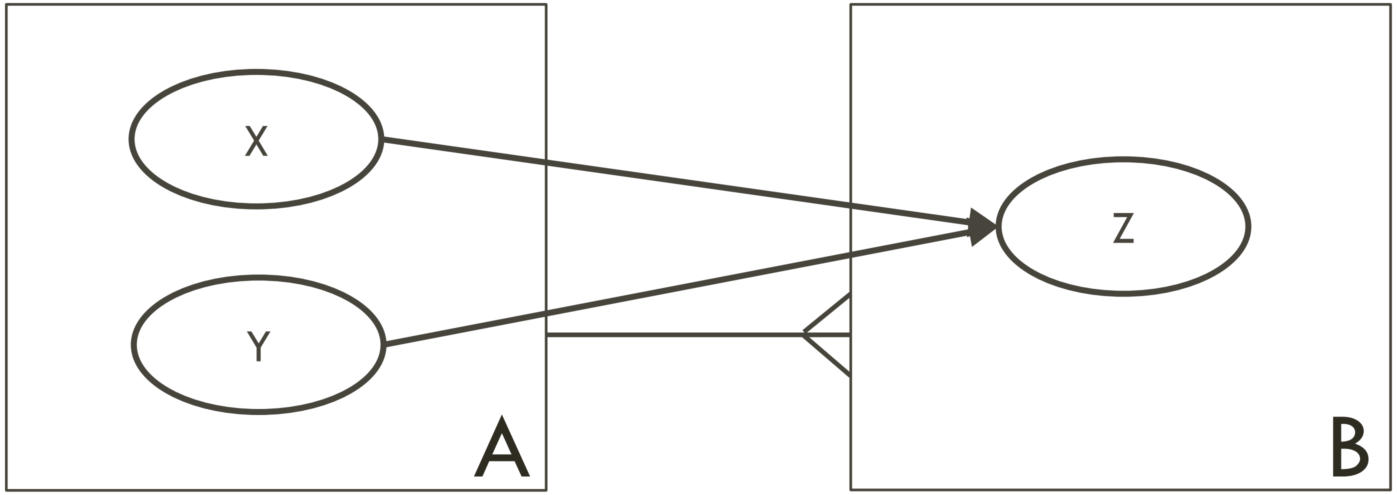
Arbour, Marazopoulou, and Jensen. *Inferring Causal Direction from Relational Data*. UAI. 2016.

OBJECT CONDITIONING

$[A].X \perp\!\!\!\perp [A].Y \mid [B].ID$

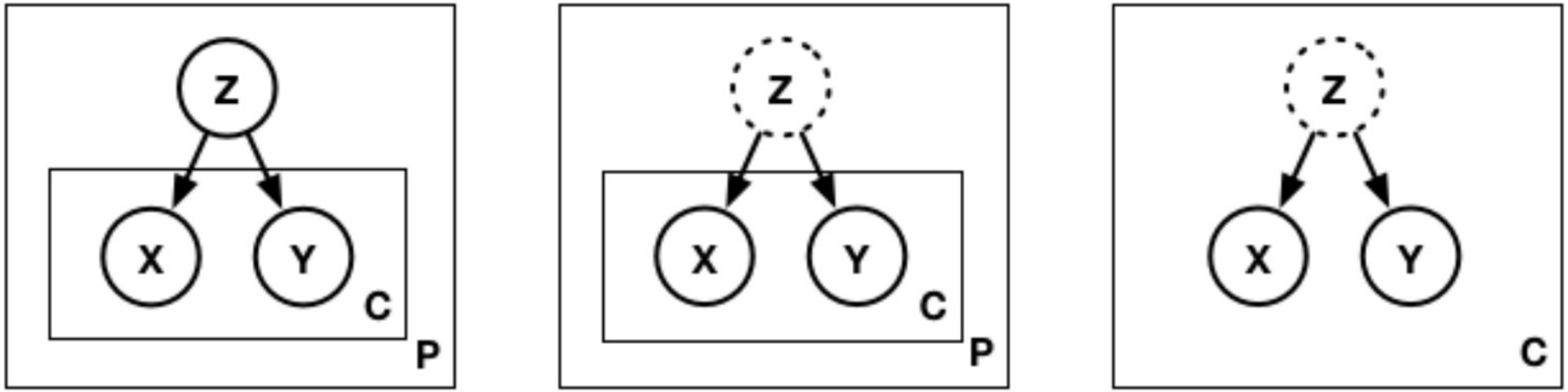


$[A].X \perp\!\!\!\perp [A].Y \mid [B].ID$



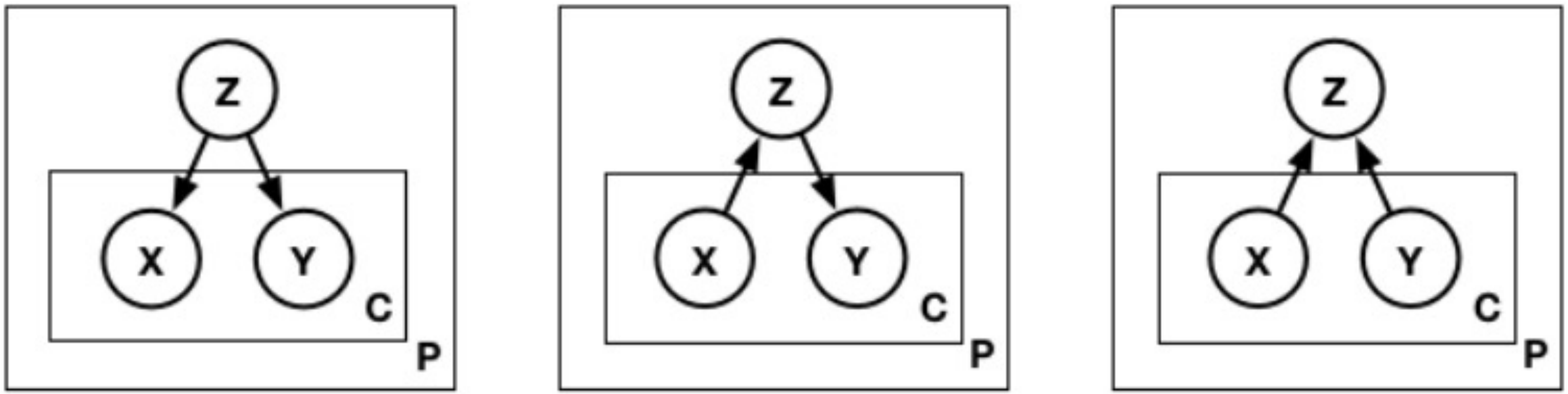
Jensen, Burroni, and Rattigan. *Object Conditioning for Causal Inference*. UAI. 2020.

OBJECT CONDITIONING



Jensen, Burroni, and Rattigan. *Object Conditioning for Causal Inference*. UAI. 2020.

OBJECT CONDITIONING



Jensen, Burroni, and Rattigan. *Object Conditioning for Causal Inference*. UAI. 2020.

EASY

Modeling multiple entity and relationships

ID for acyclic ground graphs

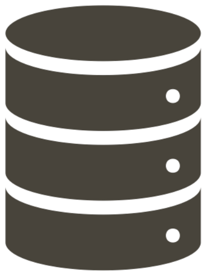
HARD

Specifying the right relational path semantic

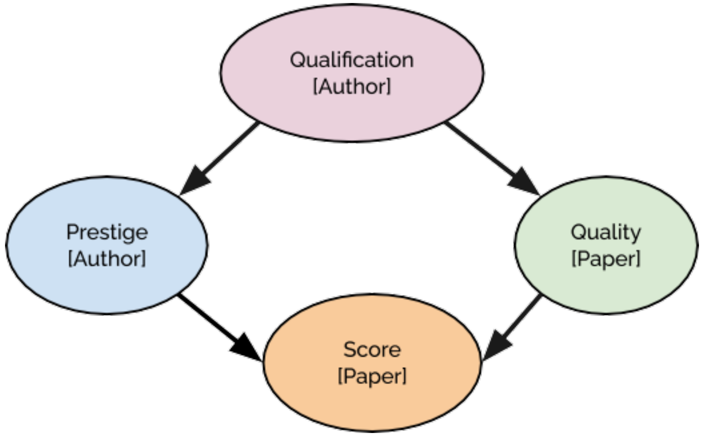
Feedback

Network uncertainty and topological features

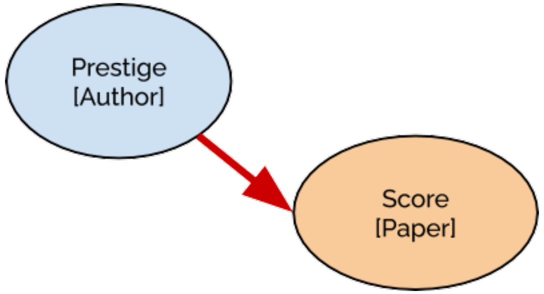
INFERENCE WITHIN THE CARL FRAMEWORK



Relational DB

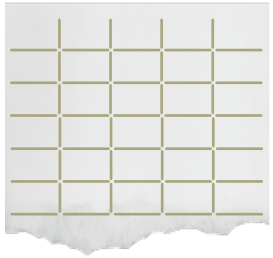


Background Knowledge



Causal Query

Single flat data-table



Causal Effects(s) Estimates

Salimi, Parikh, Kayali, Getoor, Roy, and Suciu. *Causal Relational Learning*. SIGMOD. 2020.

Content provided by Sudeepa Roy

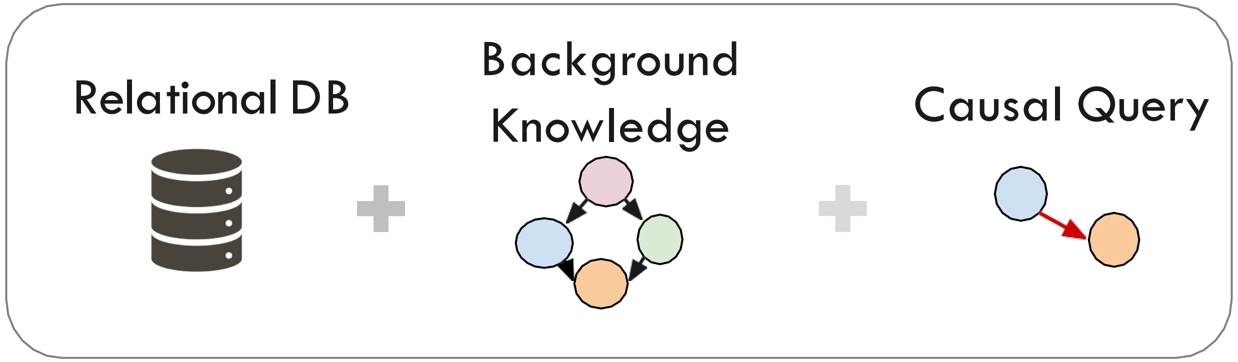
Skeleton Traversal

Grounding

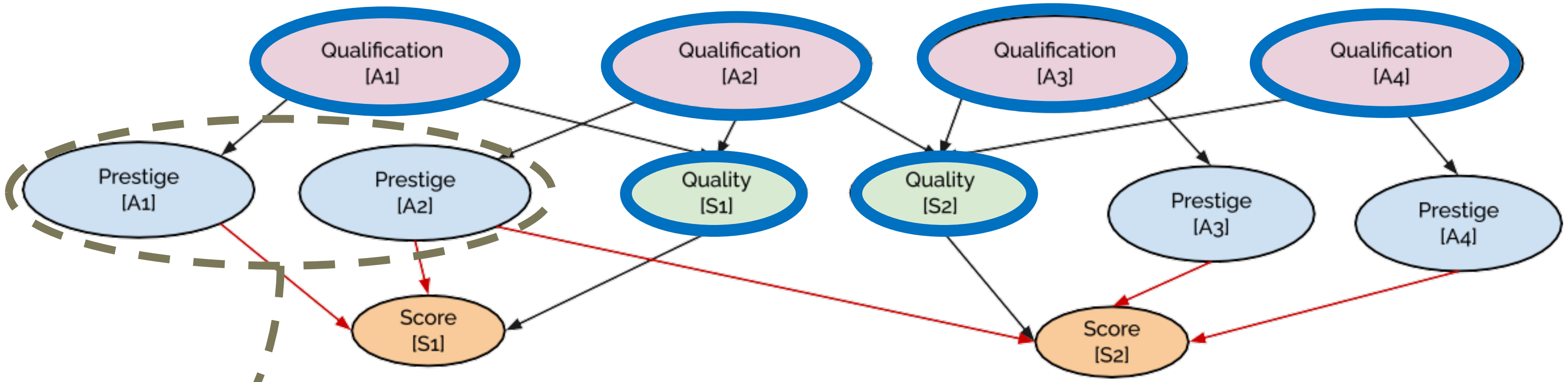
Confounder Identification

Summary Functions

Causal Inference



Grounded Causal DAG



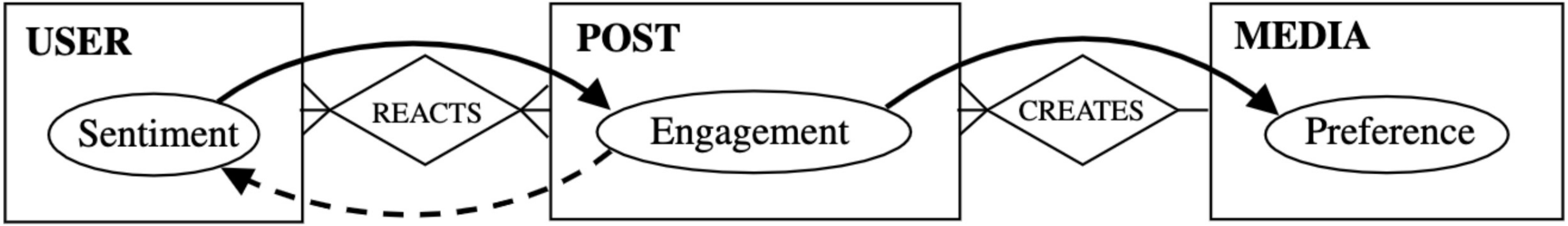
Flat Table

Summary-Prestige	Summary-Qualification	Score	Quantity
$g_{prestige}(\mathrm{Prestige}[A_1,A_2])$	$g_{qual}(\mathrm{Qualification}[A_1,A_2])$	Score[S1]	Quantity[S1]
$g_{prestige}(\mathrm{Prestige}[A_1,A_2,A_3])$	$g_{qual}(\mathrm{Qualification}[A_1,A_2,A_3])$	Score[S2]	Quantity[S2]

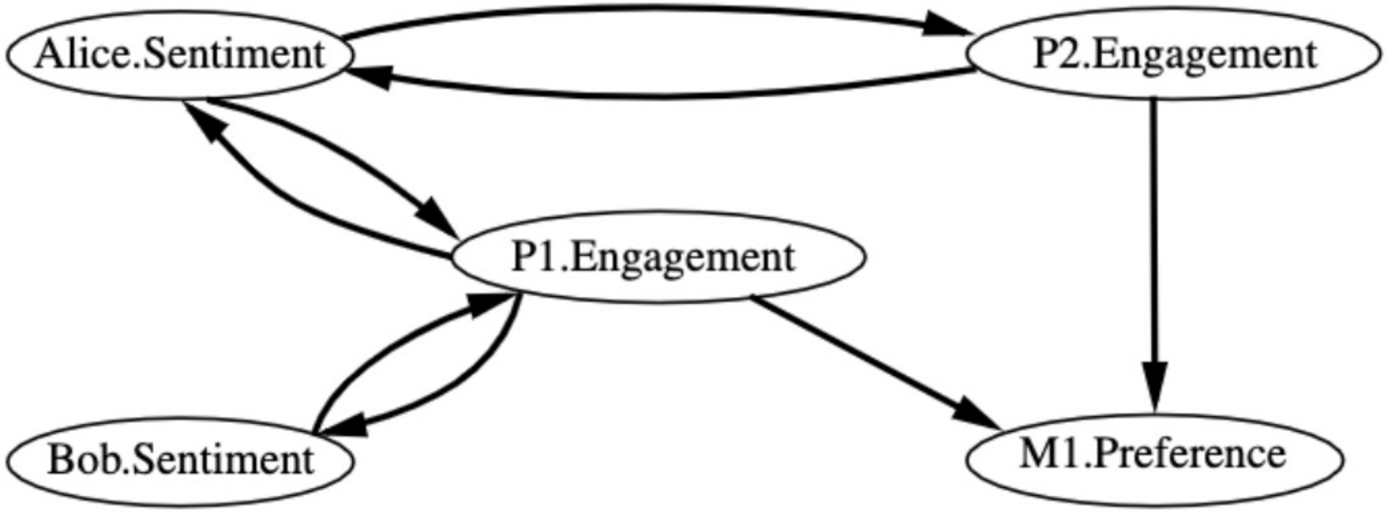
Salimi, Parikh, Kayali, Getoor, Roy, and Suciu. *Causal Relational Learning*. SIGMOD. 2020.

Content provided by Sudeepa Roy

REPRESENTING CYCLES IN TEMPLATED CAUSAL MODELS



Ahsan, Arbour, and Zheleva. Relational Causal Models with Cycles: Representation and Reasoning. CLEAR. 2022 (to appear).



Ahsan, Arbour, and Zheleva. Relational Causal Models with Cycles: Representation and Reasoning. CLEAR. 2022 (to appear).

Definition 2 (σ-separation) *(Forré and Mooij, 2017)*
A walk $\langle v_0 ... v_n \rangle$ in DCG $G = \langle \mathcal{V}, \mathcal{E} \rangle$ is σ-blocked by $C \subseteq V$ if:

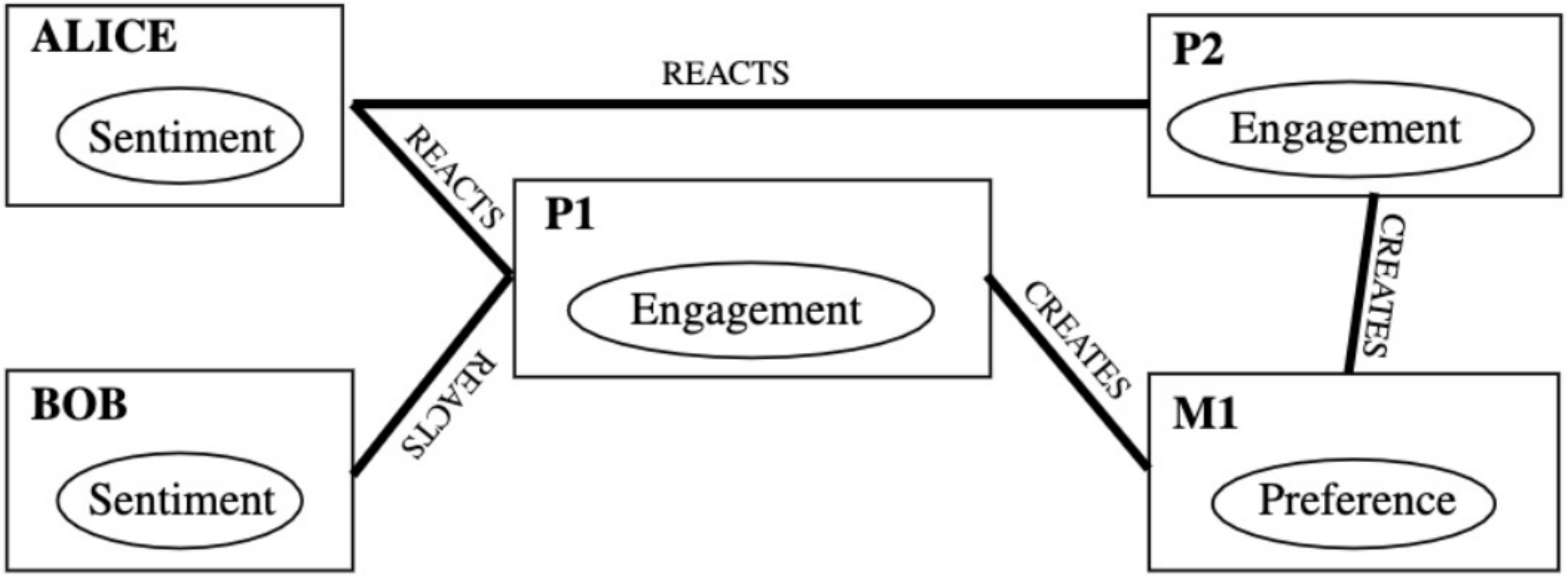
1. *its first node $v_0 \in C$ or its last node $v_n \in C$, or*
2. *it contains a collider $v_k \notin AN_{\mathcal{G}}(C)$, or*
3. *it contains a non-collider $v_k \in C$ that points to a node on the walk in another strongly connected component (i.e., $v_{k-1} \rightarrow v_k \rightarrow v_{k+1}$ with $v_{k+1} \notin SC_{\mathcal{G}}(v_k)$, $v_{k-1} \leftarrow v_k \leftarrow v_{k+1}$ with $v_{k-1} \notin SC_{\mathcal{G}}(v_k)$ or $v_{k-1} \leftarrow v_k \rightarrow v_{k+1}$ with $v_{k-1} \notin SC_{\mathcal{G}}(v_k)$ or $v_{k+1} \notin SC_{\mathcal{G}}(v_k)$).*

If all paths in $\mathcal{G}$ between any node in set $A \subseteq \mathcal{V}$ and any node in set $B \subseteq \mathcal{V}$ are σ-blocked by a set $C \subseteq \mathcal{V}$, we say that A is σ-separated from B by C, and we write $A \underset{\mathcal{G}}{\overset{\sigma}{\perp\!\!\!\perp}} B | C$.

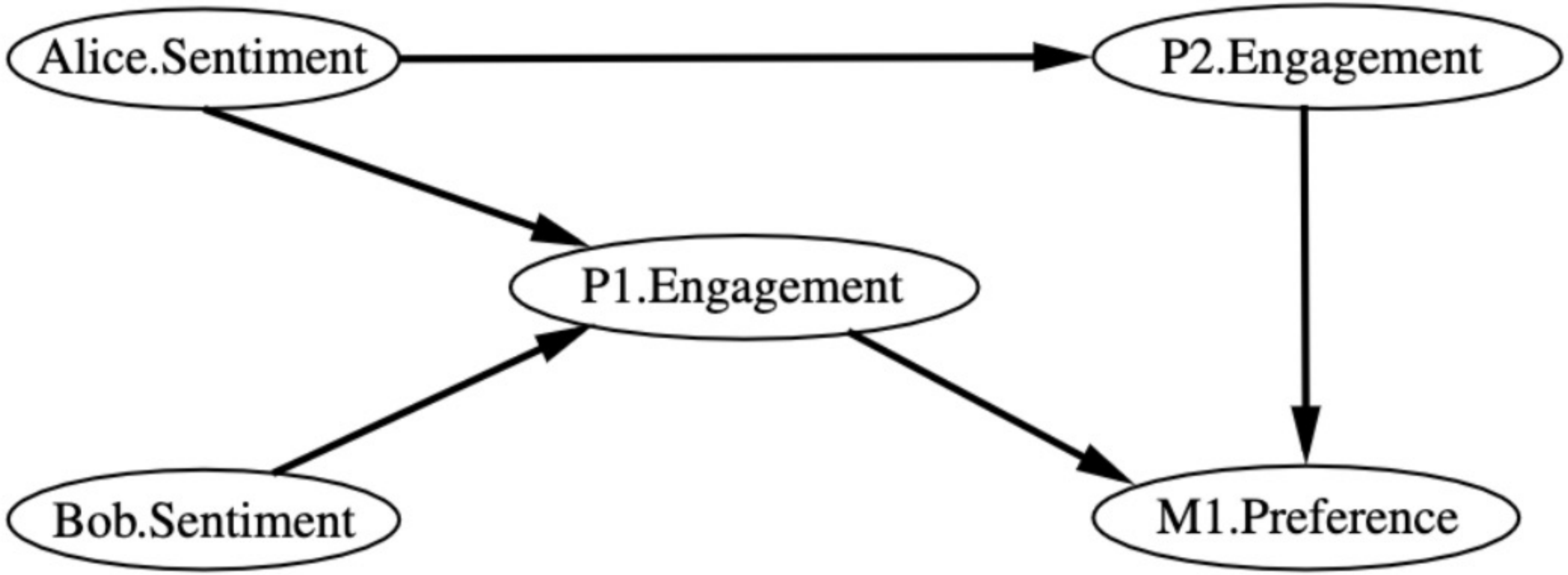
Ahsan, Arbour, and Zheleva. Relational Causal Models with Cycles: Representation and Reasoning. CLEAR. 2022 (to appear).

Definition 6 (Relational σ-separation) *Let $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$ be three distinct sets of relational variables with the same perspective $B \in \boldsymbol{\mathcal{E}} \cup \boldsymbol{\mathcal{R}}$ defined over relational schema $\mathcal{S}$. Then, for relational model structure $\mathcal{M}$, $\boldsymbol{X}$ and $\boldsymbol{Y}$ are σ-separated by $\boldsymbol{Z}$ if and only if, for all skeletons $s \in \sum_{\mathcal{S}}$, $\boldsymbol{X}|_b$ and $\boldsymbol{Y}|_b$ are σ-separated by $\boldsymbol{Z}|_b$ in ground graph $GG_{\mathcal{M}_s}$ for all instances $b \in s(B)$ where $s(B)$ refers to the instances of B in skeleton s.*

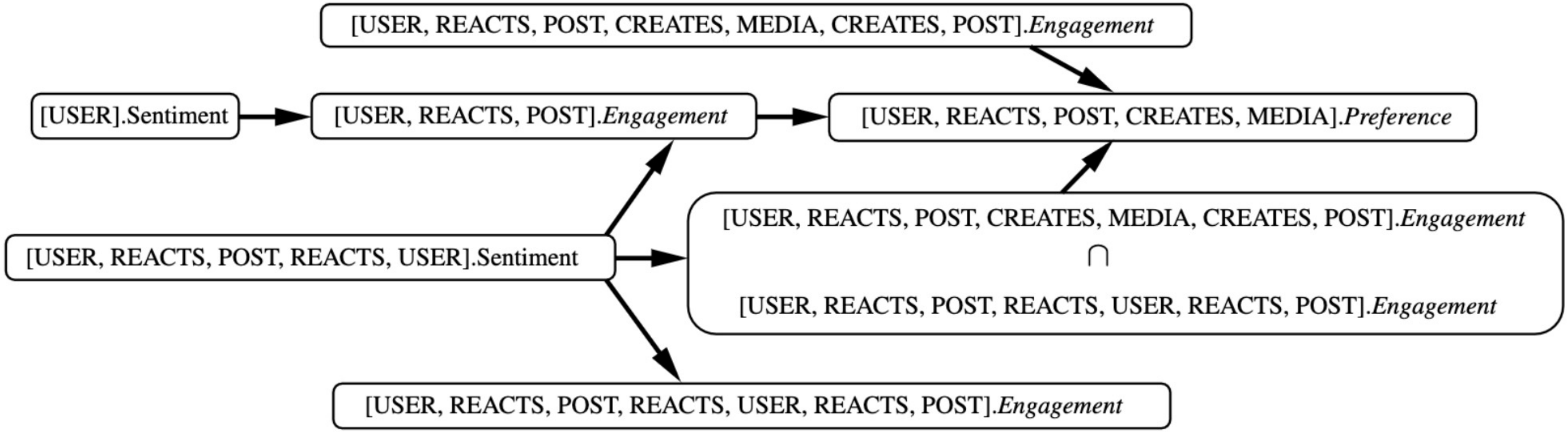
Ahsan, Arbour, and Zheleva. Relational Causal Models with Cycles: Representation and Reasoning. CLEAR. 2022 (to appear).



Ahsan, Arbour, and Zheleva. Relational Causal Models with Cycles: Representation and Reasoning. CLEAR. 2022 (to appear).



Ahsan, Arbour, and Zheleva. Relational Causal Models with Cycles: Representation and Reasoning. CLEAR. 2022 (to appear).

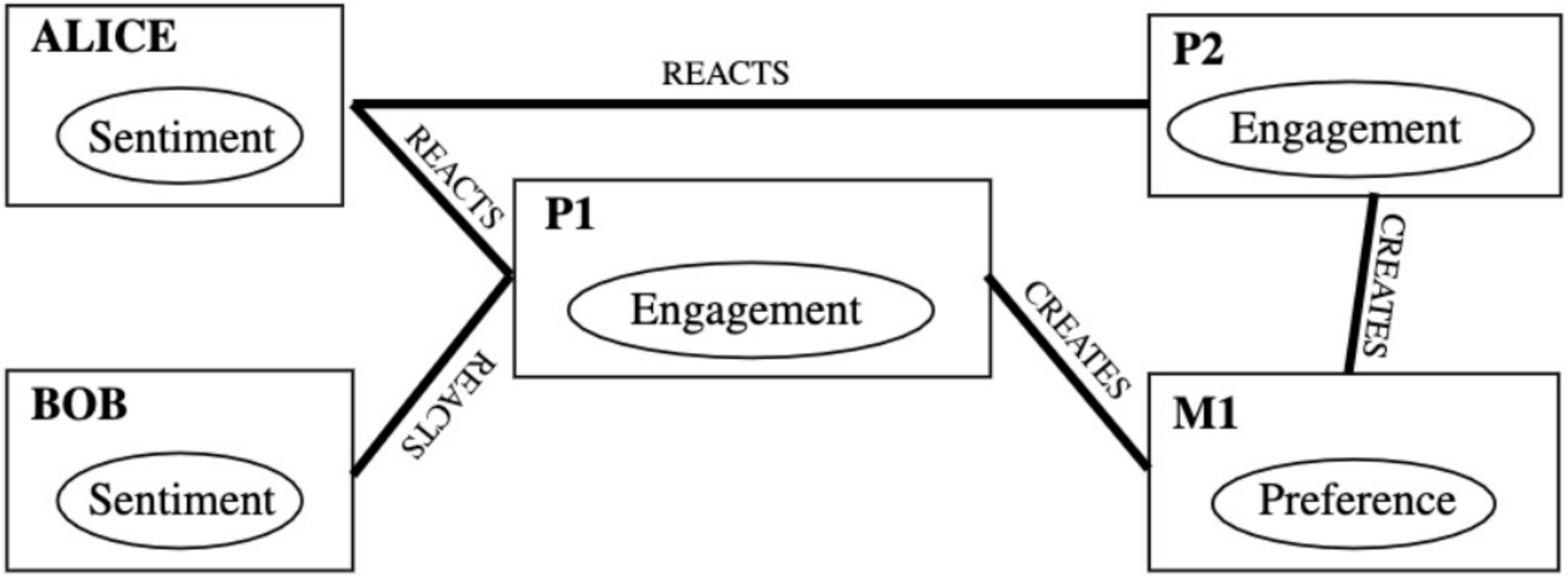


Ahsan, Arbour, and Zheleva. Relational Causal Models with Cycles: Representation and Reasoning. CLEAR. 2022 (to appear).

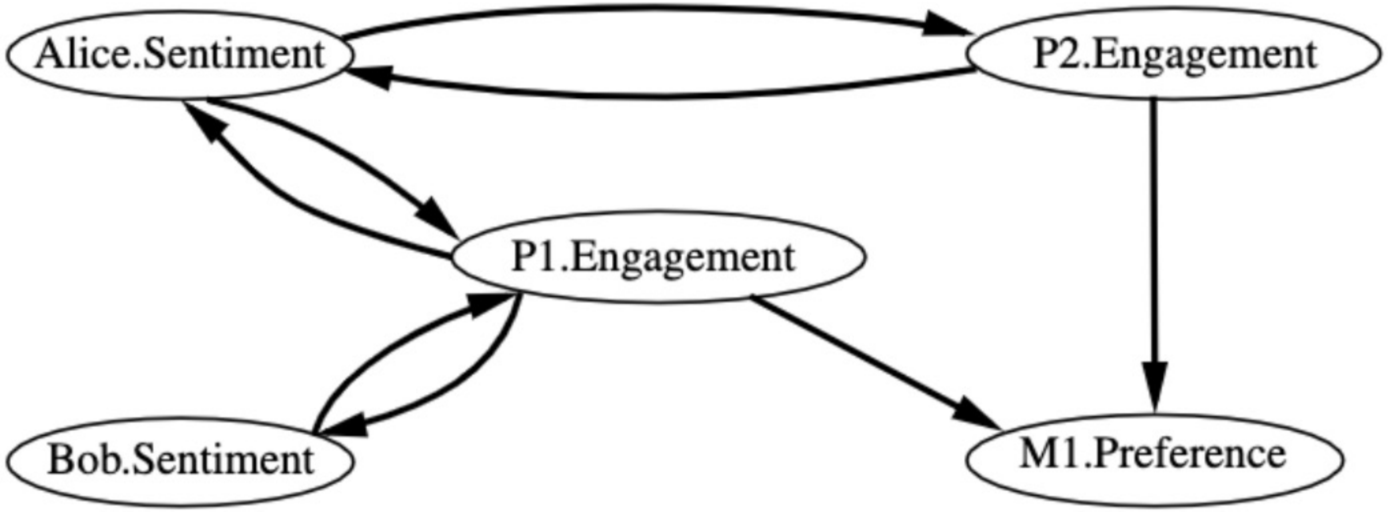
Definition 7 (σ-Abstract Ground Graph) *An abstract ground graph σ-AGG$_{\mathcal{M}} = (V, E)$ for relational model structure $\mathcal{M} = (\mathcal{S}, \mathcal{D})$, perspective $B \in \boldsymbol{\mathcal{E}} \cup \boldsymbol{\mathcal{R}}$, and hop threshold $h \in \mathbb{N}^0$ is a directed graph that abstracts the dependencies $\mathcal{D}$ for all ground graphs $GG_{\mathcal{M}_s}$, where $s \in \sum_{\mathcal{S}}$. The σ-AGG$_{\mathcal{M}_s}$ is a directed cyclic graph with the following nodes and edges:*

1. *$V = RV \cup IV$, where*
   - *(a) RV is the set of relational variables with a path of length at most h + 1.*
   - *(b) IV are intersection variables between pairs of relational variables that could intersect*
2. *$E = RVE \cup IVE$, where*
   - *(a) $RVE \subset RV \times RV$ are the relational variable edges*
   - *(b) $IVE \subset (IV \times RV) \cup (RV \times IV)$ are the intersection variable edges. This is the set of edges that intersection variables "inherit" from the relational variables that they were created from*

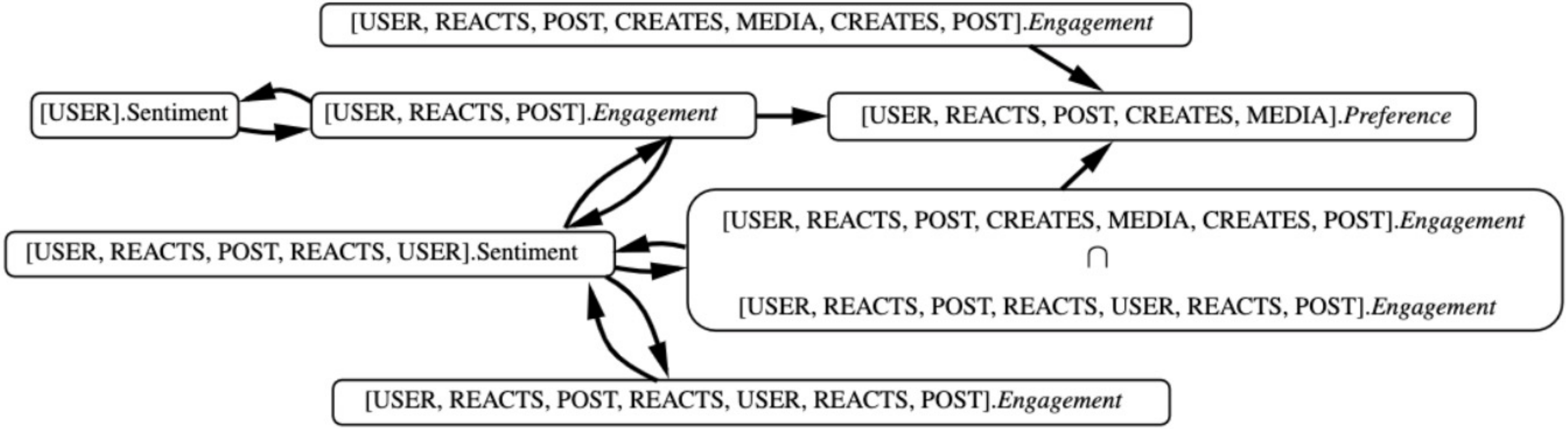
Ahsan, Arbour, and Zheleva. Relational Causal Models with Cycles: Representation and Reasoning. CLEAR. 2022 (to appear).



Ahsan, Arbour, and Zheleva. Relational Causal Models with Cycles: Representation and Reasoning. CLEAR. 2022 (to appear).



Ahsan, Arbour, and Zheleva. Relational Causal Models with Cycles: Representation and Reasoning. CLEAR. 2022 (to appear).



Ahsan, Arbour, and Zheleva. Relational Causal Models with Cycles: Representation and Reasoning. CLEAR. 2022 (to appear).

Definition 6 (Relational σ-separation) *Let* $\boldsymbol{X}$, $\boldsymbol{Y}$, *and* $\boldsymbol{Z}$ *be three distinct sets of relational variables with the same perspective* $B \in \boldsymbol{\mathcal{E}} \cup \boldsymbol{\mathcal{R}}$ *defined over relational schema* $\mathcal{S}$. *Then, for relational model structure* $\mathcal{M}$, $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *are* σ*-separated by* $\boldsymbol{Z}$ *if and only if, for all skeletons* $s \in \sum_{\mathcal{S}}$, $\boldsymbol{X}|_b$ *and* $\boldsymbol{Y}|_b$ *are* σ*-separated by* $\boldsymbol{Z}|_b$ *in ground graph* $GG_{\mathcal{M}_s}$ *for all instances* $b \in s(B)$ *where* $s(B)$ *refers to the instances of* B *in skeleton* s.

Ahsan, Arbour, and Zheleva. Relational Causal Models with Cycles: Representation and Reasoning. CLEAR. 2022 (to appear).

Definition 8 (Relational σ-separation Markov Condition) *Let X, Y, Z be relational variables for perspective $B \in \boldsymbol{\mathcal{E}} \cup \boldsymbol{\mathcal{R}}$ defined over relational schema $\mathcal{S}$. For any solution $(\boldsymbol{\mathcal{X}}, \boldsymbol{\epsilon})$ of a relational model $\mathcal{M}$ which follows a simple SCM,*

$$X \overset{\sigma}{\underset{\mathcal{M}}{\perp\!\!\!\perp}} Y | Z \implies \boldsymbol{\mathcal{X}}_X \underset{\mathbb{P}_{\mathcal{M}}(\boldsymbol{\mathcal{X}})}{\perp\!\!\!\perp} \boldsymbol{\mathcal{X}}_Y | \boldsymbol{\mathcal{X}}_Z, \quad \textit{if and only if}$$

$$x \overset{\sigma}{\underset{GG_{\mathcal{M}}}{\perp\!\!\!\perp}} y | z \implies \boldsymbol{\mathcal{X}}'_x \underset{\mathbb{P}_{GG_{\mathcal{M}}}(\boldsymbol{\mathcal{X}}')}{\perp\!\!\!\perp} \boldsymbol{\mathcal{X}}'_y | \boldsymbol{\mathcal{X}}'_z, \;\; \textit{for} \; \forall x \in X|_b, \; \forall y \in Y|_b, \; \forall z \in Z|_b$$

in ground graph $GG_{\mathcal{M}_s}$ for all skeletons $s \in \sum_{\mathcal{S}}$ and for all $b \in s(B)$ where $(\boldsymbol{\mathcal{X}}', \boldsymbol{\epsilon}')$ refers to the solution of the SCM corresponding to the ground graphs.

Ahsan, Arbour, and Zheleva. Relational Causal Models with Cycles: Representation and Reasoning. CLEAR. 2022 (to appear).

EASY

Modeling multiple entity and relationships

Representation of causal cyclic relationships

HARD

Specifying the right relational path semantic

Learning and inference

Network uncertainty and topological features

COUNTERFACTUALS & CAUSAL EFFECTS IN OBSERVATIONAL NETWORK DATA

Discovery

DISCOVERING RELATIONAL STRUCTURE OF CHAIN GRAPHS

Assume: **Causal** structure is known a priori

Learn: The **relational** structure

Bhattacharya, Malinksy, and Shpitser. *Causal inference under interference and network uncertainty*. UAI, 2019.

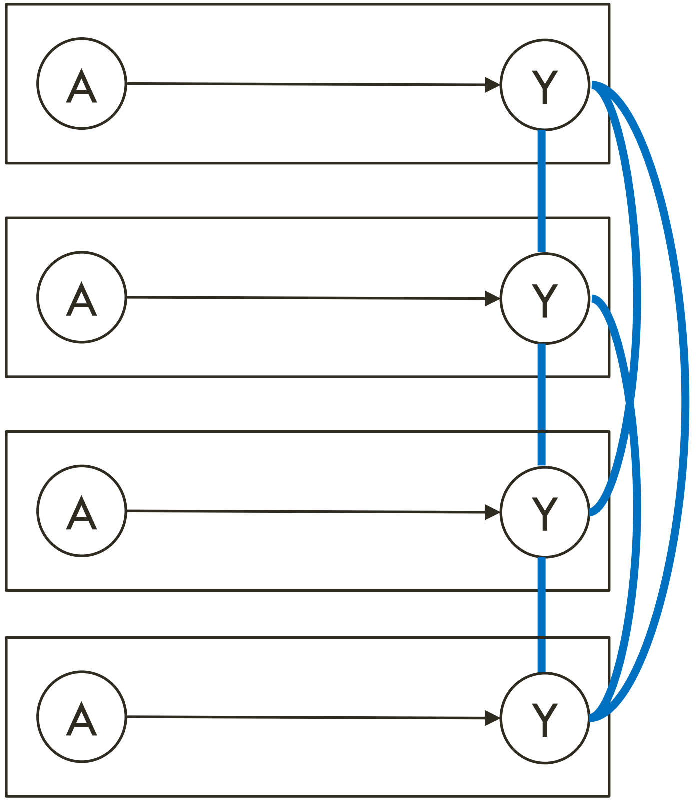
DISCOVERING RELATIONAL STRUCTURE

Assume: **Causal** graph is known

Learn: Greedily search for the relational structure that maximizes the pseudo-likelihood

$$PL(\boldsymbol{D}; G) \equiv \prod_{i=1}^{n} \prod_{j=1}^{d} p(x_{j,i} \mid x_{-j,i}; G)$$

Bhattacharya, Malinksy, and Shpitser. *Causal inference under interference and network uncertainty*. UAI, 2019.

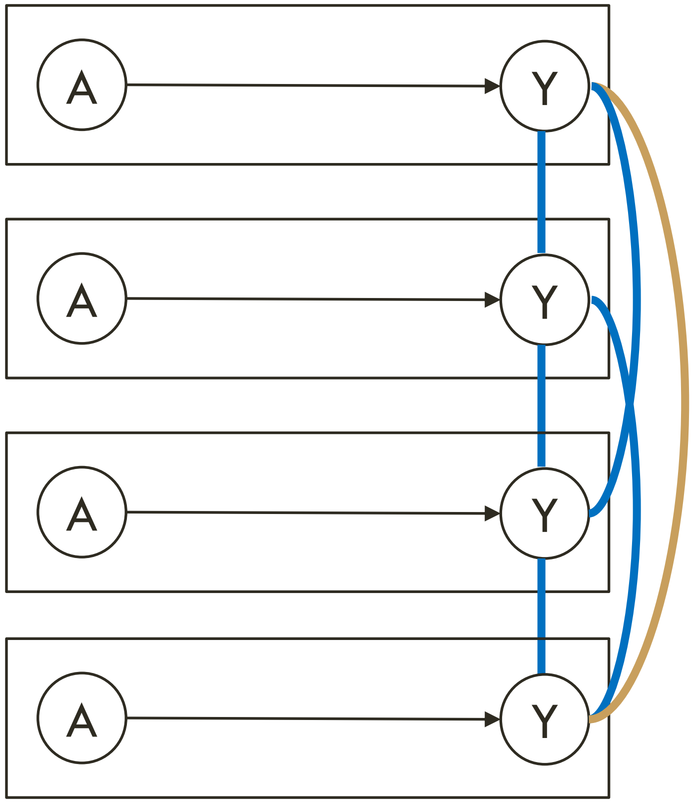


Algorithm 1 GREEDY NETWORK SEARCH($\mathcal{G}^{\text{init}}, \mathbf{D}$)

```
```

$\mathcal{G}^* \leftarrow \mathcal{G}^{\text{init}}$
score change $\leftarrow$ True
while score change **do**
  score change $\leftarrow$ False
  $\mathcal{E}^*_{\mathcal{N}} \leftarrow$ network ties in $\mathcal{G}^*$
  $\boxed{E_{max} \leftarrow \text{argmax}_{E \in \mathcal{E}^*_{\mathcal{N}}} \text{PBIC}(\mathbf{D}; \mathcal{G}^* \setminus E)}$
  if PBIC($\mathbf{D}; \mathcal{G}^* \setminus E_{max}$) $>$ PBIC($\mathbf{D}; \mathcal{G}^*$) **then**
    $\mathcal{G}^* \leftarrow \mathcal{G}^* \setminus E_{max}$ $\triangleright$ deleting edge E_{max}
    score change $\leftarrow$ True
return $\mathcal{E}^*_{\mathcal{N}}$

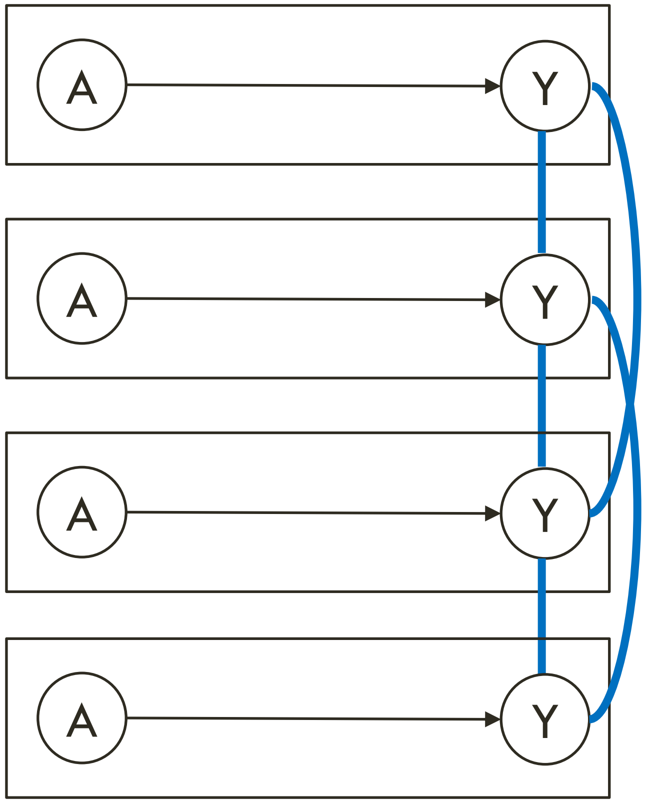
Bhattacharya, Malinksy, and Shpitser. *Causal inference under interference and network uncertainty*. UAI, 2019.



Algorithm 1 $\textsc{Greedy Network Search}(\mathcal{G}^{\text{init}}, \mathbf{D})$

$\mathcal{G}^* \leftarrow \mathcal{G}^{\text{init}}$
score change $\leftarrow$ True
while score change **do**
    score change $\leftarrow$ False
    $\mathcal{E}^*_{\mathcal{N}} \leftarrow$ network ties in $\mathcal{G}^*$
    $E_{max} \leftarrow \text{argmax}_{E \in \mathcal{E}^*_{\mathcal{N}}} \text{PBIC}(\mathbf{D}; \mathcal{G}^* \setminus E)$
    if $\text{PBIC}(\mathbf{D}; \mathcal{G}^* \setminus E_{max}) > \text{PBIC}(\mathbf{D}; \mathcal{G}^*)$ **then**
        $\mathcal{G}^* \leftarrow \mathcal{G}^* \setminus E_{max}$ $\triangleright$ deleting edge E_{max}
        score change $\leftarrow$ True
return $\mathcal{E}^*_{\mathcal{N}}$

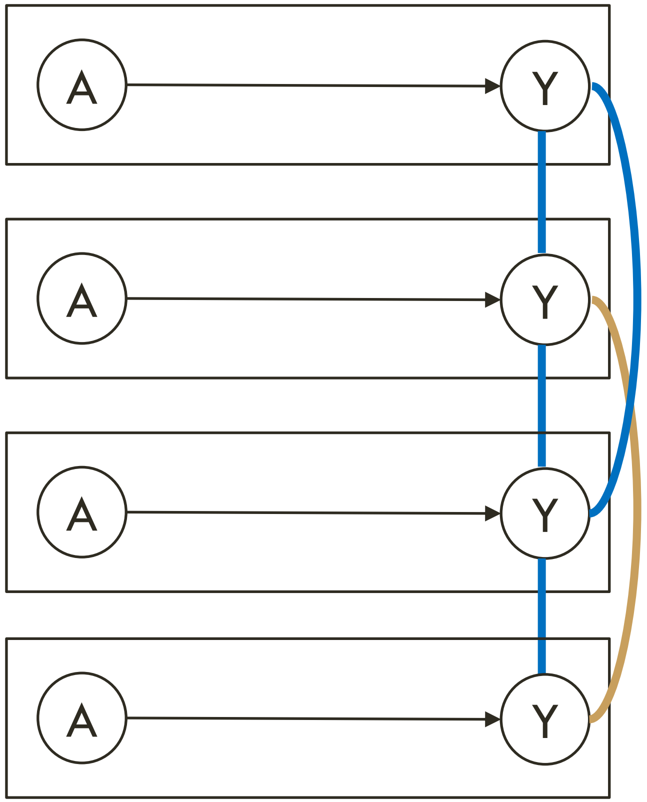
Bhattacharya, Malinksy, and Shpitser. *Causal inference under interference and network uncertainty*. UAI, 2019.



Algorithm 1 GREEDY NETWORK SEARCH($\mathcal{G}^{\text{init}}$, $\mathbf{D}$)

$\mathcal{G}^* \leftarrow \mathcal{G}^{\text{init}}$
score change $\leftarrow$ True
while score change **do**
    score change $\leftarrow$ False
    $\mathcal{E}^*_{\mathcal{N}} \leftarrow$ network ties in $\mathcal{G}^*$
    $\boxed{E_{max} \leftarrow \text{argmax}_{E \in \mathcal{E}^*_{\mathcal{N}}} \text{PBIC}(\mathbf{D}; \mathcal{G}^* \setminus E)}$
    if $\text{PBIC}(\mathbf{D}; \mathcal{G}^* \setminus E_{max}) > \text{PBIC}(\mathbf{D}; \mathcal{G}^*)$ **then**
        $\mathcal{G}^* \leftarrow \mathcal{G}^* \setminus E_{max}$ $\triangleright$ deleting edge E_{max}
        score change $\leftarrow$ True
return $\mathcal{E}^*_{\mathcal{N}}$

Bhattacharya, Malinksy, and Shpitser. *Causal inference under interference and network uncertainty*. UAI, 2019.

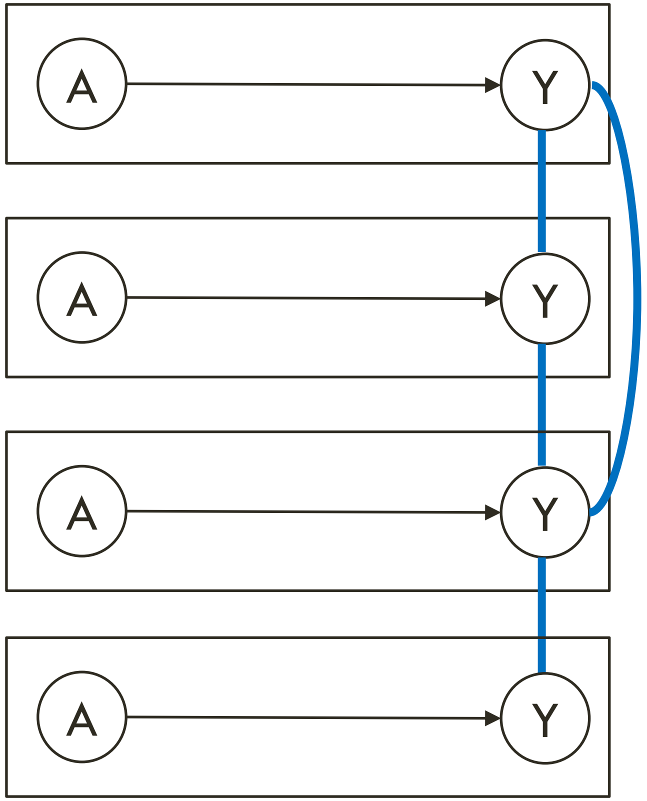


Algorithm 1 GREEDY NETWORK SEARCH($\mathcal{G}^{\text{init}}$, $\mathbf{D}$)

```
G* ← G^init
score change ← True
while score change do
    score change ← False
    E*_N ← network ties in G*
    E_max ← argmax_{E ∈ E*_N} PBIC(D; G* \ E)
    if PBIC(D; G* \ E_max) > PBIC(D; G*) then
        G* ← G* \ E_max          ▷ deleting edge E_max
        score change ← True
return E*_N
```

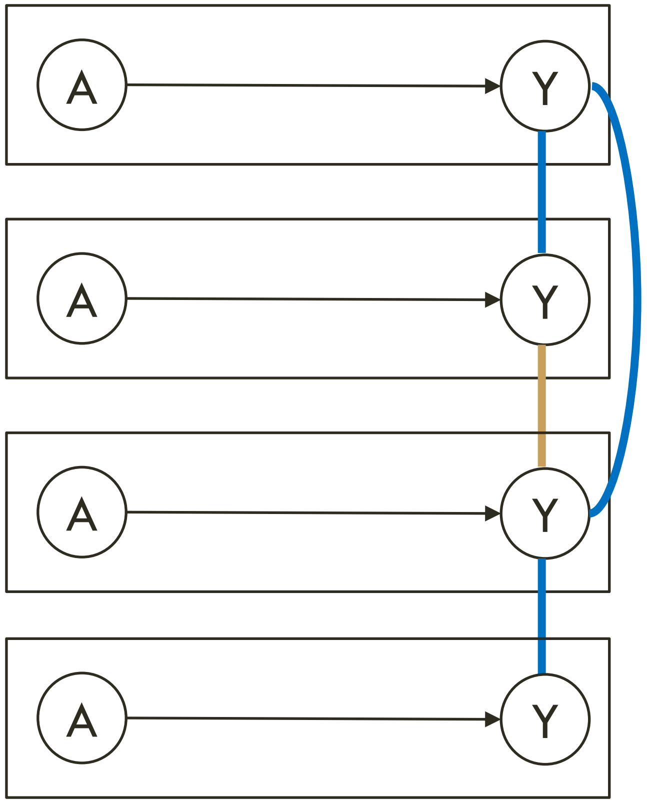
$\mathcal{G}^* \leftarrow \mathcal{G}^{\text{init}}$
score change $\leftarrow$ True
while score change **do**
  score change $\leftarrow$ False
  $\mathcal{E}^*_{\mathcal{N}} \leftarrow$ network ties in $\mathcal{G}^*$
  $E_{max} \leftarrow \text{argmax}_{E \in \mathcal{E}^*_{\mathcal{N}}} \text{PBIC}(\mathbf{D}; \mathcal{G}^* \setminus E)$
  if $\text{PBIC}(\mathbf{D}; \mathcal{G}^* \setminus E_{max}) > \text{PBIC}(\mathbf{D}; \mathcal{G}^*)$ **then**
    $\mathcal{G}^* \leftarrow \mathcal{G}^* \setminus E_{max}$ $\triangleright$ deleting edge E_{max}
    score change $\leftarrow$ True
return $\mathcal{E}^*_{\mathcal{N}}$

Bhattacharya, Malinksy, and Shpitser. *Causal inference under interference and network uncertainty*. UAI, 2019.



Algorithm 1 GREEDY NETWORK SEARCH($\mathcal{G}^{\text{init}}, \mathbf{D}$)

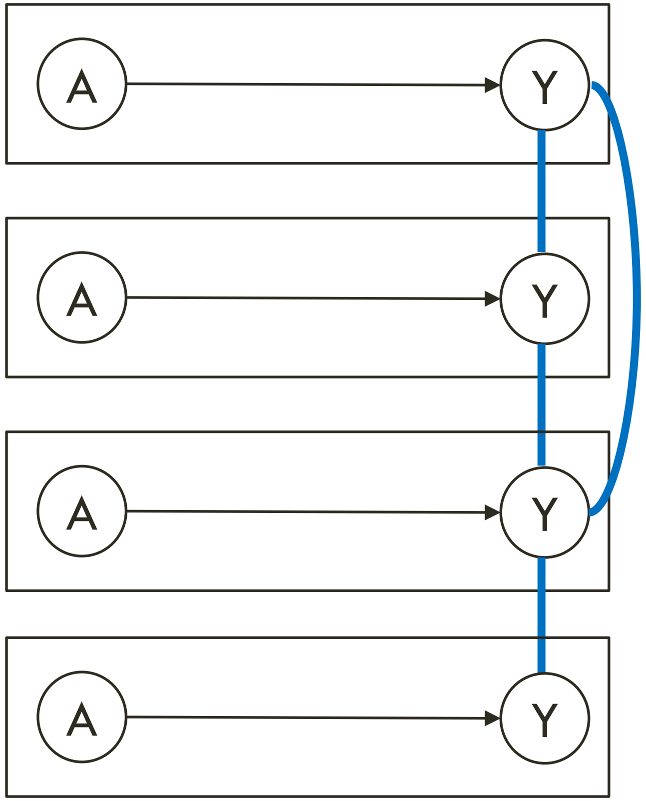
```
𝒢* ← 𝒢^init
score change ← True
while score change do
    score change ← False
    ℰ*_𝒩 ← network ties in 𝒢*
    [ E_max ← argmax_{E∈ℰ*_𝒩} PBIC(D; 𝒢* \ E) ]
    if PBIC(D; 𝒢* \ E_max) > PBIC(D; 𝒢*) then
        𝒢* ← 𝒢* \ E_max            ▷ deleting edge E_max
        score change ← True
return ℰ*_𝒩
```



Algorithm 1 GREEDY NETWORK SEARCH($\mathcal{G}^{\text{init}}$, $\mathbf{D}$)

$\mathcal{G}^* \leftarrow \mathcal{G}^{\text{init}}$
score change $\leftarrow$ True
while score change **do**
    score change $\leftarrow$ False
    $\mathcal{E}^*_{\mathcal{N}} \leftarrow$ network ties in $\mathcal{G}^*$
    $E_{max} \leftarrow \text{argmax}_{E \in \mathcal{E}^*_{\mathcal{N}}} \text{PBIC}(\mathbf{D}; \mathcal{G}^* \setminus E)$
    if $\text{PBIC}(\mathbf{D}; \mathcal{G}^* \setminus E_{max}) > \text{PBIC}(\mathbf{D}; \mathcal{G}^*)$ **then**
        $\mathcal{G}^* \leftarrow \mathcal{G}^* \setminus E_{max}$ $\triangleright$ deleting edge E_{max}
        score change $\leftarrow$ True
return $\mathcal{E}^*_{\mathcal{N}}$

Bhattacharya, Malinksy, and Shpitser. *Causal inference under interference and network uncertainty*. UAI, 2019.



Algorithm 1 GREEDY NETWORK SEARCH($\mathcal{G}^{\text{init}}$, $\mathbf{D}$)

$\mathcal{G}^* \leftarrow \mathcal{G}^{\text{init}}$
score change $\leftarrow$ True
while score change **do**
  score change $\leftarrow$ False
  $\mathcal{E}^*_{\mathcal{N}} \leftarrow$ network ties in $\mathcal{G}^*$
  $E_{max} \leftarrow \text{argmax}_{E \in \mathcal{E}^*_{\mathcal{N}}} \text{PBIC}(\mathbf{D}; \mathcal{G}^* \setminus E)$
  if $\text{PBIC}(\mathbf{D}; \mathcal{G}^* \setminus E_{max}) > \text{PBIC}(\mathbf{D}; \mathcal{G}^*)$ **then**
    $\mathcal{G}^* \leftarrow \mathcal{G}^* \setminus E_{max}$ ▷ deleting edge E_{max}
    score change $\leftarrow$ True
return $\mathcal{E}^*_{\mathcal{N}}$

Bhattacharya, Malinksy, and Shpitser. *Causal inference under interference and network uncertainty*. UAI, 2019.

DISCOVERING RELATIONAL STRUCTURE

1

Can additionally search over heterogenous relationship types

2

Consistent assuming true distribution is in the curved exponential family

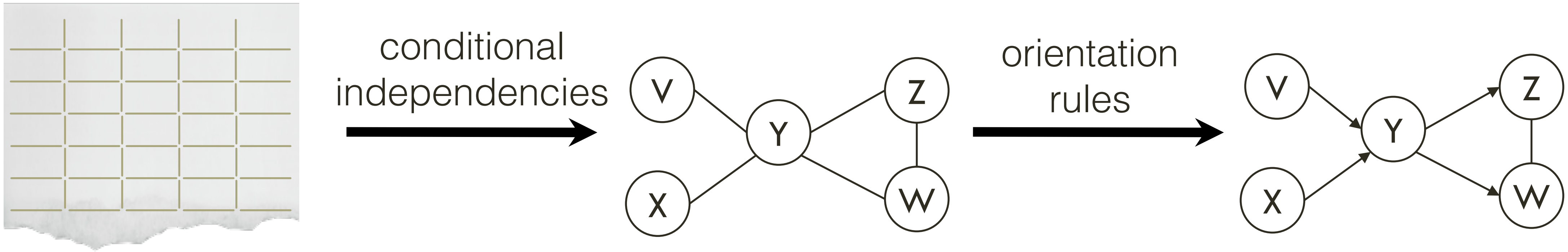
Bhattacharya, Malinksy, and Shpitser. *Causal inference under interference and network uncertainty*. UAI, 2019.

DISCOVERING THE CAUSAL STRUCTURE OF MULTI-RELATIONAL DATA

Assume: **Relational** structure is known a priori

Learn: The **causal** structure

PC ALGORITHM



DATA

SKELETON

MARKOV EQUIVALENCE CLASS

Spirtes, Glymour, Scheines. *Causation, Prediction, and Search.* MIT Press, 1993.

ORIENTATION RULES

Collider Detection (CD)

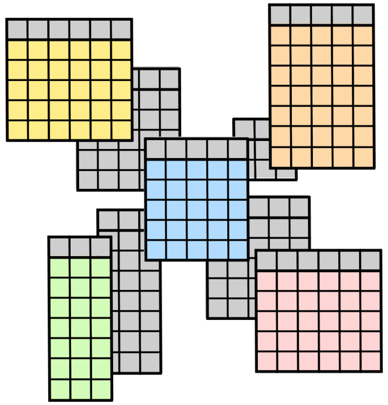
Cycle Avoidance (CA)

Known Non-Colliders (KNC)

Meek Rule 3 (MR3)

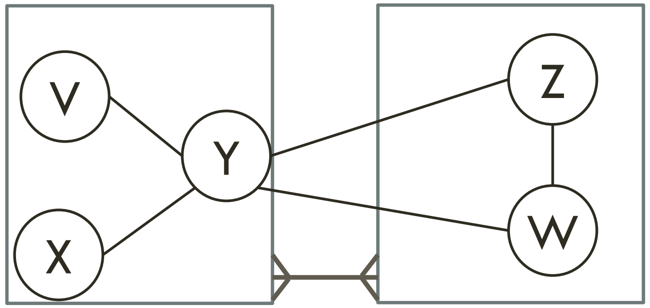
Spirtes, Glymour, Scheines. *Causation, Prediction, and Search.* MIT Press, 1993.

RELATIONAL CAUSAL DISCOVERY (RCD)



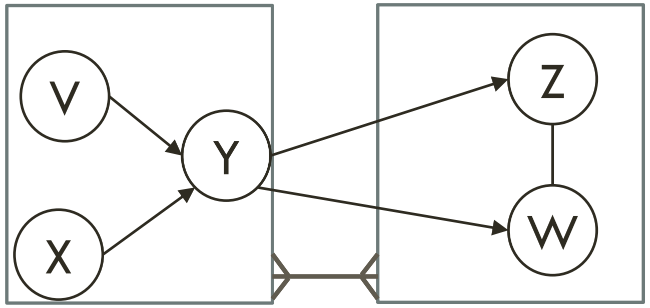
RELATIONAL DATA

conditional independencies



SKELETON

orientation rules

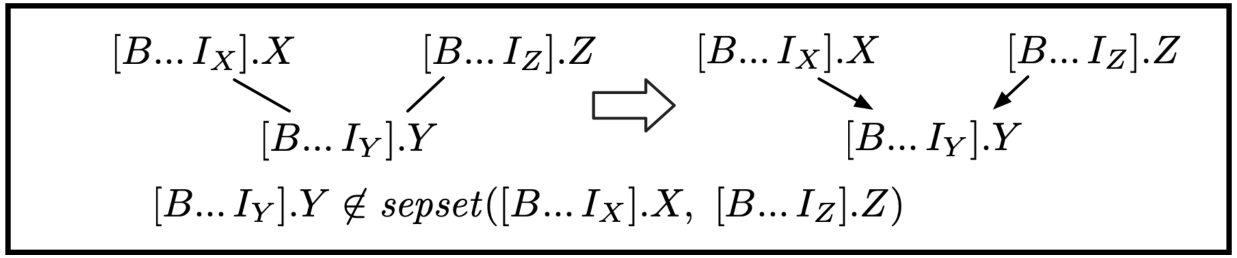


MARKOV EQUIVALENCE CLASS

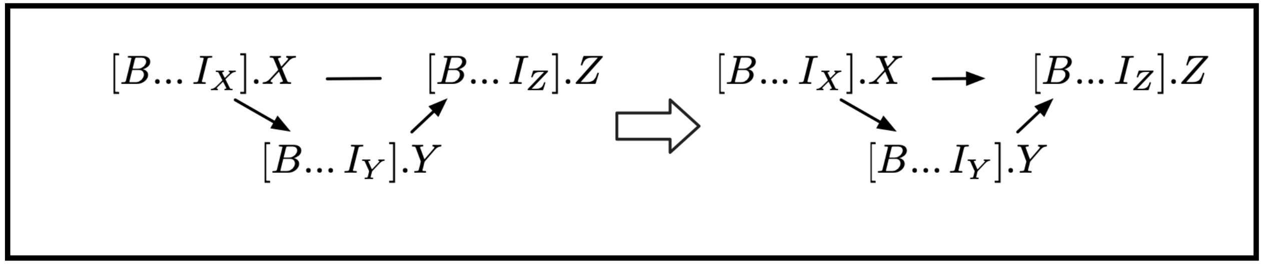
Maier, Marazopoulou, Arbour, and Jensen. *A Sound and Complete Algorithm for Learning Causal Models from Relational Data*. UAI. 2013.
Lee and Hanovar. *On Learning Causal Models from Relational Data*. AAAI. 2016.

RELATIONAL CAUSAL DISCOVERY (RCD)

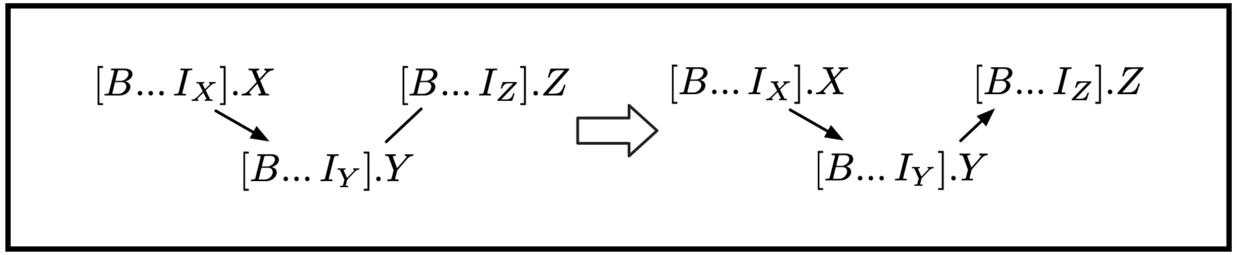
Collider Detection (CD)



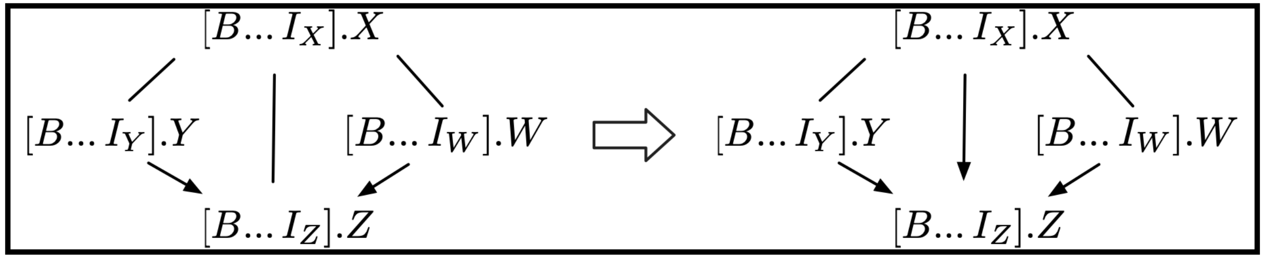
Cycle Avoidance (CA)



Known Non-Colliders (KNC)



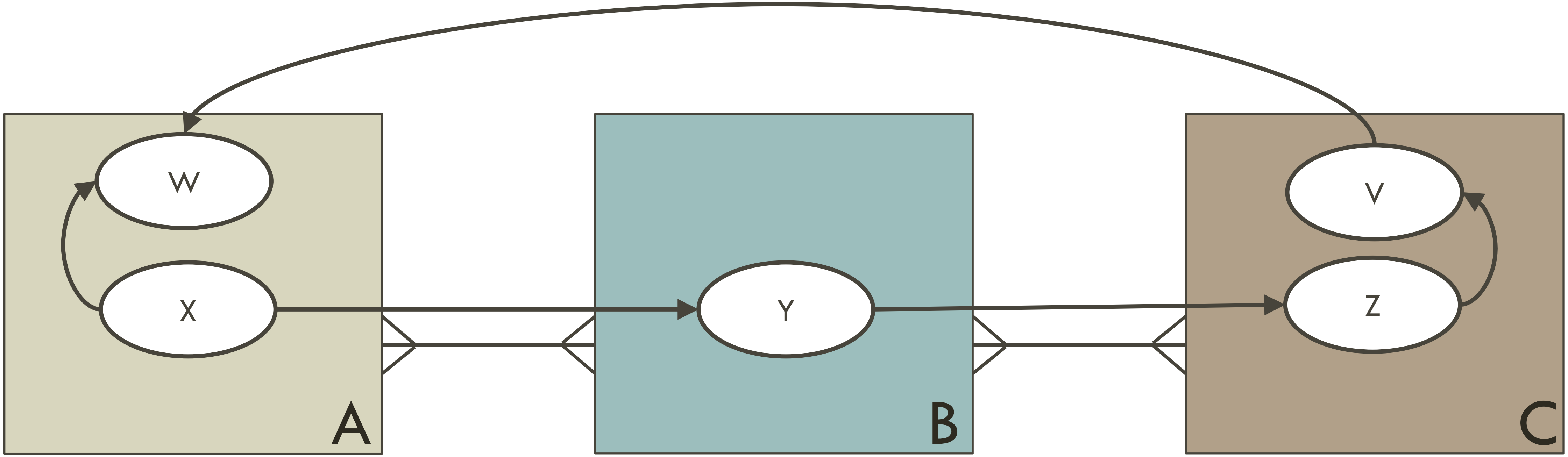
Meek Rule 3 (MR3)



Orientations are propagated across perspectives

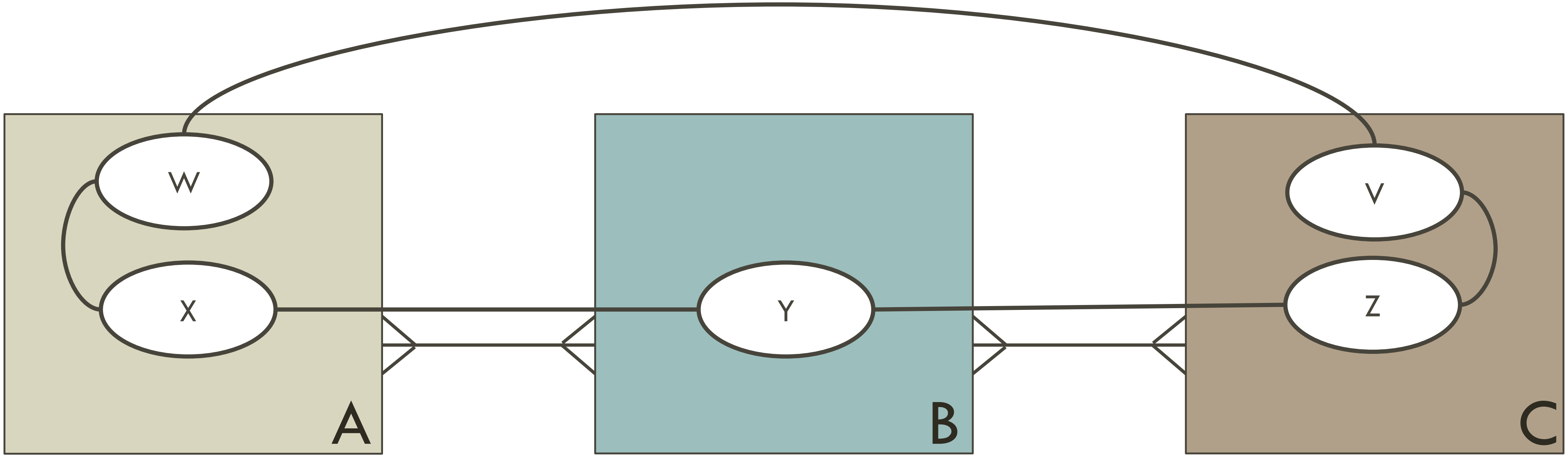
Maier, Marazopoulou, Arbour, and Jensen. *A Sound and Complete Algorithm for Learning Causal Models from Relational Data*. UAI. 2013.

TRACING THE EXECUTION OF RCD



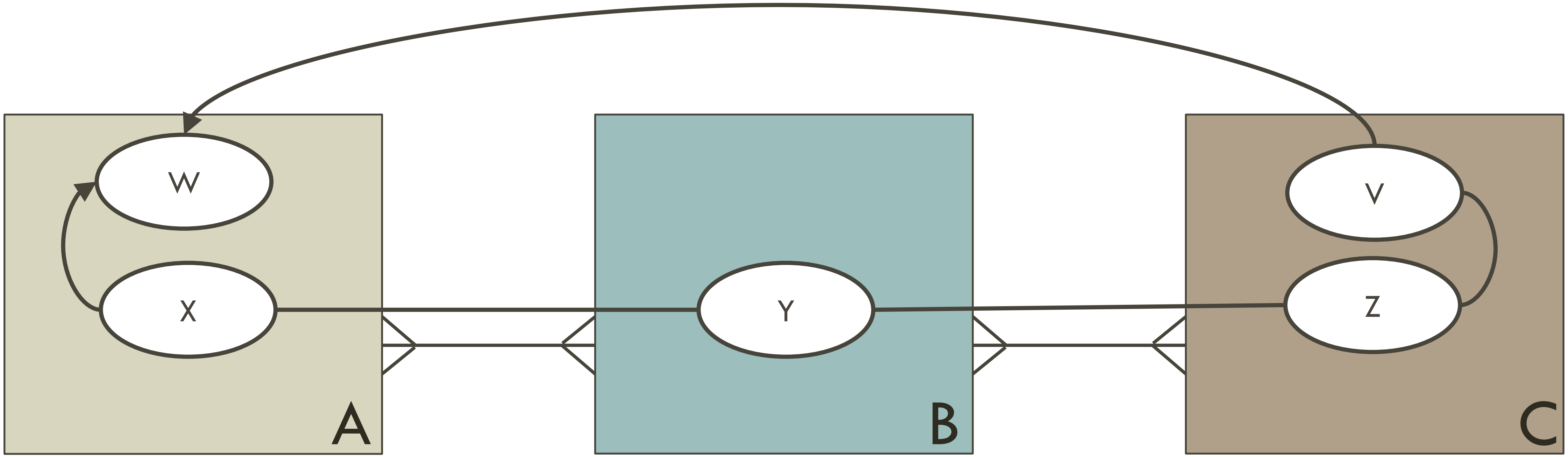
Maier, Marazopoulou, Arbour, and Jensen. *A Sound and Complete Algorithm for Learning Causal Models from Relational Data*. UAI. 2013.

IDENTIFY UNDIRECTED EDGES



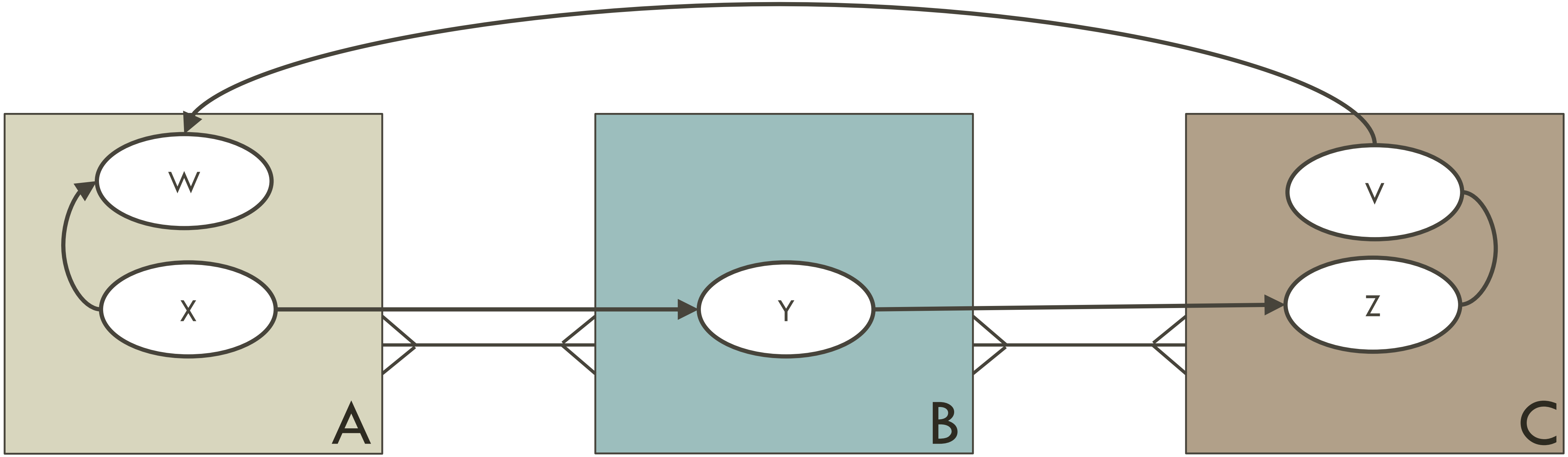
Maier, Marazopoulou, Arbour, and Jensen. *A Sound and Complete Algorithm for Learning Causal Models from Relational Data*. UAI. 2013.

APPLY COLLIDER DETECTION



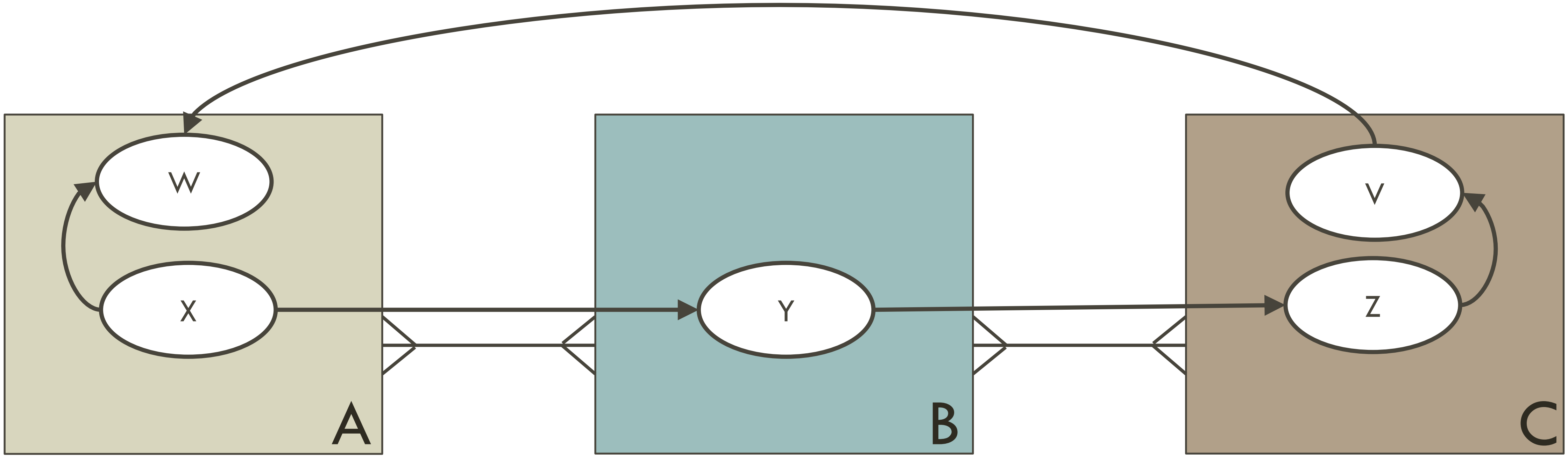
Maier, Marazopoulou, Arbour, and Jensen. *A Sound and Complete Algorithm for Learning Causal Models from Relational Data*. UAI. 2013.

ORIENT RELATIONAL DEPENDENCIES



Maier, Marazopoulou, Arbour, and Jensen. *A Sound and Complete Algorithm for Learning Causal Models from Relational Data*. UAI. 2013.

APPLY KNOWN NON-COLLIDERS



Maier, Marazopoulou, Arbour, and Jensen. *A Sound and Complete Algorithm for Learning Causal Models from Relational Data*. UAI. 2013.

SUMMARY

Relational domains hold considerable promise and unique challenges to causal inference

There is a growing literature with many open research problem in:

- Experimental design
- Graphical representations
- Observational causal inference
- Discovery

THANK YOU!

David Arbour, Adobe Research

@darbour26

Elena Zheleva, University of Illinois at Chicago

@elenadata

Website: https://netcause.github.io

- All materials, slides & references
- Our contact information

REFERENCES

Introduction

Christakis & Fowler. The Spread of Obesity in a Large Social Network Over 32 Years. New England Journal of Medicine. 2007.
Lee & Ogburn. Network Dependence Can Lead to Spurious Associations and Invalid Inference. Journal of American Statistical Association. 2020.
Bail, Argyle, Brown, Bumpus, Chen, Hunzaker, Lee, Mann, Merhout, Volfovsky. Exposure to opposing views on social media can increase political polarization. PNAS 2018.
T. Sun, S. Viswanathan, E. Zheleva. Creating Social Contagion through Firm-mediated Message Design: Evidence from A Randomized Field Experiment. Management Science 2021.
J. Pearl. The seven tools of causal inference, with reflections on machine learning. Communications of the ACM 2019.
Pearl. Causality: Models, Reasoning and Inference. 2009.

Causal Effects in Networks

Halloran, Struchiner. Causal inference in infectious diseases. Epidemiology 1995.
Hudgens, Halloran. Toward causal inference with interference. JASA 2008.
Ugander, Karrer, Backstrom, Kleinberg. Graph cluster randomization: Network exposure to multiple universes. KDD 2013.

Interventions and Network Experiments

H. Leavitt. Some effects of certain communication patterns on group performance. The Journal of Abnormal and Social Psychology, 46(1): 38. 1951.
S. Aral. Networked Experiments. The Oxford Handbook of the Economics of Networks. 2016.
Manski. Economic analysis of social interactions. Journal of Economic Perspectives. 2000
P. Toulis, E. Kao. Estimation of Causal Peer Influence Effects. ICML 2013.
D. Eckles, R. Kizilcec, E. Bakshy. Estimating peer effects in networks with peer encouragement designs. PNAS 2016
Z. Fatemi, E. Zheleva. Minimizing interference and selection bias in network experiment design. ICWSM 2020.
Saveski, Pouget-Abadie, Saint-Jacques, Duan, Ghosh, Xu, Airoldi. Detecting network effects: Randomizing over randomized experiments. KDD 2017.
Stuart. Matching methods for causal inference: a review and look forward. Stat. Science 2010.
R. Johari, H. Li, I. Liskovic, G. Weintraub. Experimental design in two-sided platforms: An analysis of bias. Arxiv 2020.
P. Bajari, B. Burdick, G. Imbens, J. McQueen, T. Richardson, I. Rosen. Multiple randomization designs for interference. ASSA Annual Meeting 2020.

Counterfactuals: Blocks

Ogburn, VanderWeele. Causal Diagrams for Interference. Statistical Science 2014.
Shalizi & Thomas. Homophily and Contagion Are Generically Confounded in Observational Social Network Studies. Sociological Methods & Research 2011.

REFERENCES

Representation: Chain and Segregated Graphs

Ogburn, et al. Causal Inference, Social Networks and Chain Graphs. JRSSB 2020.
Sherman, Arbour, and Shpitser. General Identification of Dynamic Treatment Regimes Under Interference. AISTATS. 2020.
Lauritzen & Richardson. Chain Graph Models and Their Causal Interpretation. JRSSB. 2002.
Richardson. Markov Properties for Acyclic Directed Mixed Graphs. Scandinavian Journal of Statistics. 2002.
Richardson, Robins and Shpitser. Nested Markov Properties for Acyclic Directed Mixed Graphs. UAI. 2012.
Tian and Pearl. A General Identification Condition for Causal Effects. AAAI. 2002.
Shpitser and Pearl. Identification of Joint Interventional Distributions in Recursive Semi-Markovian Causal Models. AAAI. 2006.
Huang and Voltorta. Pearl's Calculs of Intervention is Complete. UAI. 2006.
Shpitser. Segregated Graphs and Marginals of Chain Graph Models. NeurIPS. 2015.
Sherman & Shpitser. Identification of Causal Effects from Dependent Data. NeurIPS. 2018.

Representation: Multi-Relational Data

Getoor, Friedman, Koller & Pfeffer. Learning Probabilistic Relational Models. IJCAI. 1999.
Heckerman, Meek, and Killer. Probablistic Models for Relational Data. MSR Tech Report. 2004.
Maier, Marazopoulou, and Jensen. Reasoning about Independence in Probabilistic Models of Relational Data. Arxiv. 2013.
Lee and Hanovar. A Characterization of Markov Equivalence Classes of Relational Causal Models under Path Semantics. UAI. 2016
Arbour, Garant, and Jensen. Inferring Network Effects from Observational Data. KDD. 2016.
Maier, Marazopoulou, Arbour, and Jensen. A Sound and Complete Algorithm for Learning Causal Models from Relational Data. UAI. 2013.
Arbour, Marazopoulou, and Jensen. Inferring Causal Direction from Relational Data. UAI. 2016.
Jensen, Burroni, and Rattigan. Object Conditioning for Causal Inference. UAI. 2020.
Salimi, Parikh, Kayali, Getoor, Roy, and Suciu. Causal Relational Learning. SIGMOD. 2020.
Ahsan, Arbour, and Zheleva. Relational Causal Models with Cycles: Representation and Reasoning. CLEAR. 2022 (to appear).

Discovery

Bhattacharya, Malinksy, and Shpitser. Causal inference under interference and network uncertainty. UAI, 2019.
Spirtes, Glymour, Scheines. Causation, Prediction, and Search. MIT Press, 1993.
Lee and Hanovar. On Learning Causal Models from Relational Data. AAAI. 2016.