Causal Inference With Contagion and Latent Homophily Under Full Interference

Yufeng Wu, Rohit Bhattacharya
{sw20,rb17}@williams.edu

ACIC 2024

Williams College

Social Interactions Create Dependence in Data [1]

[1] Shalizi and Thomas 2011, Ogburn and VanderWeele 2014, Lauritzen and Richardson 2002, Shpitser 2015

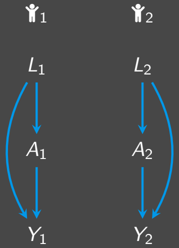
Social Interactions Create Dependence in Data [1]

L = confounders; A = therapy sessions; Y = job satisfaction.

[1] Shalizi and Thomas 2011, Ogburn and VanderWeele 2014, Lauritzen and Richardson 2002, Shpitser 2015

Social Interactions Create Dependence in Data [1]

L = confounders; A = therapy sessions; Y = job satisfaction.

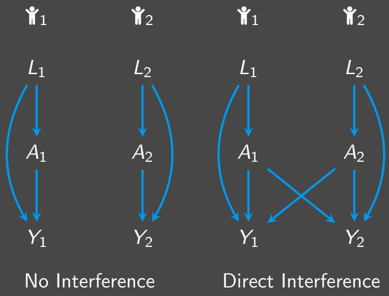


No Interference

[1] Shalizi and Thomas 2011, Ogburn and VanderWeele 2014, Lauritzen and Richardson 2002, Shpitser 2015

Social Interactions Create Dependence in Data [1]

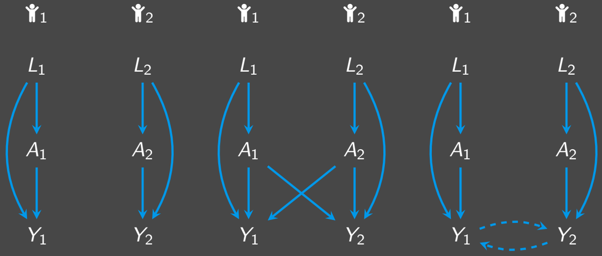
L = confounders; A = therapy sessions; Y = job satisfaction.



[1] Shalizi and Thomas 2011, Ogburn and VanderWeele 2014, Lauritzen and Richardson 2002, Shpitser 2015

Social Interactions Create Dependence in Data [1]

L = confounders; A = therapy sessions; Y = job satisfaction.

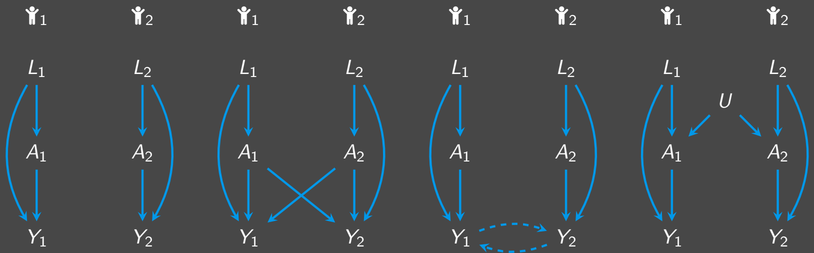


No Interference Direct Interference Contagion

[1]Shalizi and Thomas 2011, Ogburn and VanderWeele 2014, Lauritzen and Richardson 2002, Shpitser 2015

Social Interactions Create Dependence in Data [1]

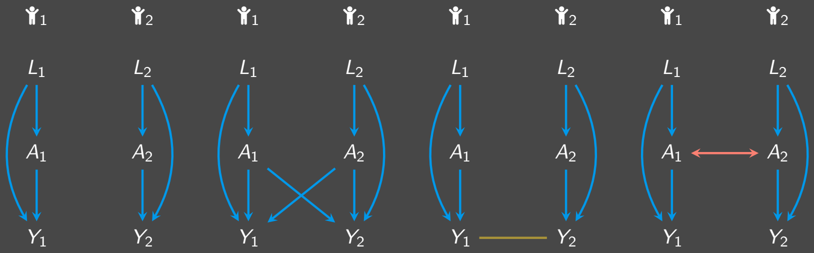
L = confounders; A = therapy sessions; Y = job satisfaction.



No Interference Direct Interference Contagion Latent Homophily

[1] Shalizi and Thomas 2011, Ogburn and VanderWeele 2014, Lauritzen and Richardson 2002, Shpitser 2015

L = confounders; A = therapy sessions; Y = job satisfaction.

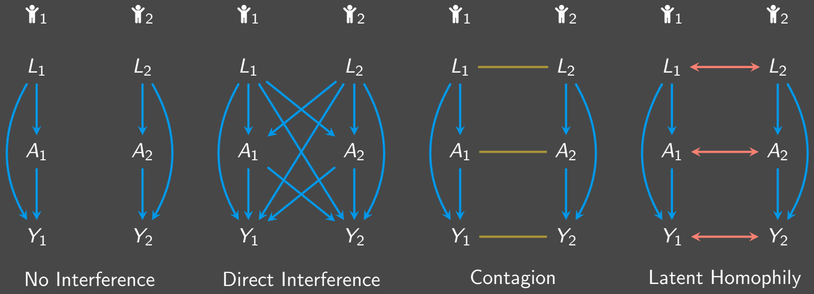


No Interference   Direct Interference   Contagion   Latent Homophily

[2] Shalizi and Thomas 2011, Ogburn and VanderWeele 2014, Lauritzen and Richardson 2002, Shpitser 2015

Social Interactions Create Dependence in Data [3]

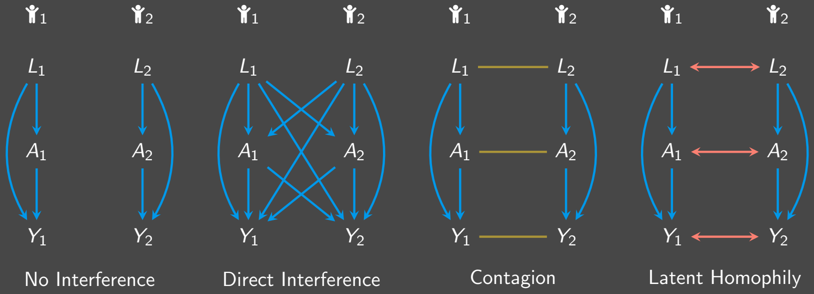
L = confounders; A = therapy sessions; Y = job satisfaction.



[3] Shalizi and Thomas 2011, Ogburn and VanderWeele 2014, Lauritzen and Richardson 2002, Shpitser 2015

Social Interactions Create Dependence in Data [3]

L = confounders; A = therapy sessions; Y = job satisfaction.



"**Interference**"

[3] Shalizi and Thomas 2011, Ogburn and VanderWeele 2014, Lauritzen and Richardson 2002, Shpitser 2015

How to Deal With Interference? (1)

[4]Bhattacharya, Malinsky, and Shpitser 2020, Kang and Imbens 2016, Tchetgen and VanderWeele 2012, Hudgens and Halloran 2008

How to Deal With Interference? (1)

Partial Interference: assume i.i.d. "chunks" of data. [4]

[4] Bhattacharya, Malinsky, and Shpitser 2020, Kang and Imbens 2016, Tchetgen and VanderWeele 2012, Hudgens and Halloran 2008

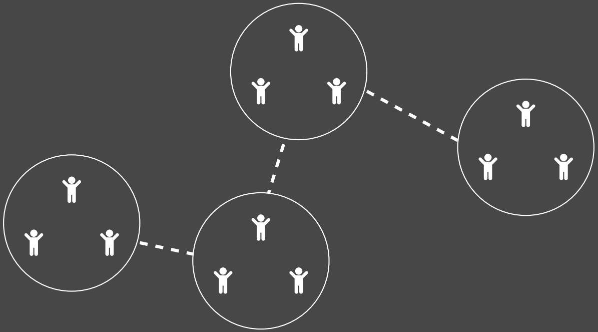
How to Deal With Interference? (1)

Partial Interference: assume i.i.d. "chunks" of data. [4]

[4] Bhattacharya, Malinsky, and Shpitser 2020, Kang and Imbens 2016, Tchetgen and VanderWeele 2012, Hudgens and Halloran 2008

How to Deal With Interference? (1)

Partial interference does not hold in general!



How to Deal With Interference? (2)

This work focuses on **Full Interference** [5]:

[5] Ogburn and VanderWeele 2014, Tchetgen Tchetgen, Fulcher, and Shpitser 2021

How to Deal With Interference? (2)

This work focuses on **Full Interference** [5]:

- Everyone may interfere with anyone else in the network.

[5] Ogburn and VanderWeele 2014, Tchetgen Tchetgen, Fulcher, and Shpitser 2021

How to Deal With Interference? (2)

This work focuses on **Full Interference** [5]:

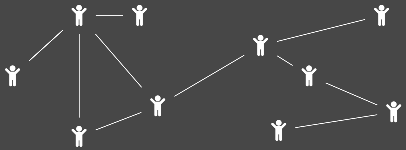
- Everyone may interfere with anyone else in the network.
- No restrictions on the structure of the network.

[5] Ogburn and VanderWeele 2014, Tchetgen Tchetgen, Fulcher, and Shpitser 2021

How to Deal With Interference? (2)

This work focuses on **Full Interference** [5]:

- Everyone may interfere with anyone else in the network.
- No restrictions on the structure of the network.

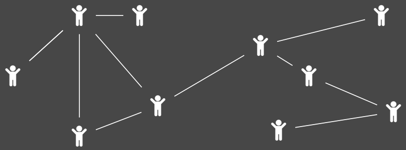


[5] Ogburn and VanderWeele 2014, Tchetgen Tchetgen, Fulcher, and Shpitser 2021

How to Deal With Interference? (2)

This work focuses on **Full Interference** [5]:

- Everyone may interfere with anyone else in the network.
- No restrictions on the structure of the network.



Usually assume **parameter-sharing**: $p(L_i, L_{\mathrm{nb}(i)}; \theta)$ is shared by everyone in the network.

[5] Ogburn and VanderWeele 2014, Tchetgen Tchetgen, Fulcher, and Shpitser 2021

Previous Work (1)

Auto-g computation[6]:

- can estimate causal effects under full interference.

[6] Tchetgen Tchetgen, Fulcher, and Shpitser 2021

Previous Work (1)

Auto-g computation[6]:

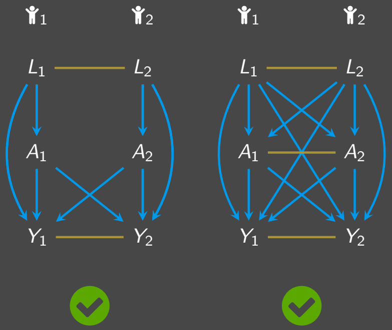
- can estimate causal effects under full interference.
- allows for direct interference ($\rightarrow$) and contagion ($-$), but assumes no latent homophily ($\leftrightarrow$).

[6] Tchetgen Tchetgen, Fulcher, and Shpitser 2021

Previous Work (1)

Auto-g computation[6]:

- can estimate causal effects under full interference.
- allows for direct interference (→) and contagion (–), but assumes no latent homophily (↔).

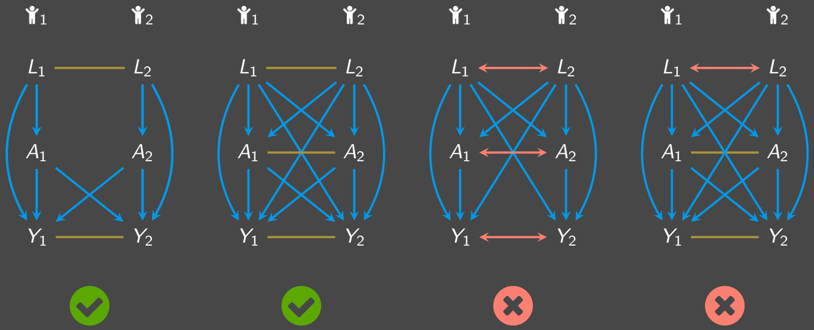


[6] Tchetgen Tchetgen, Fulcher, and Shpitser 2021

Previous Work (1)

Auto-g computation[6]:

- can estimate causal effects under full interference.
- allows for direct interference (→) and contagion (–), but assumes no latent homophily (↔).



[6] Tchetgen Tchetgen, Fulcher, and Shpitser 2021

Previous Work (2)

Causal Inference for Social Network Data [7].

- allows for direct interference and latent homophily between individuals.

[7] Ogburn et al. 2024

Open Problem

1. A method that simultaneously accounts for all three mechanisms: direct interference ($\rightarrow$), contagion ($-$), and latent homophily ($\leftrightarrow$).

Open Problem

1. A method that simultaneously accounts for all three mechanisms: direct interference ($\rightarrow$), contagion ($-$), and latent homophily ($\leftrightarrow$).

2. The assumptions of auto-g relies on prior knowledge & belief.

   We want a test to distinguish between interference due to contagion ($-$) and latent homophily ($\leftrightarrow$).

Open Problem

1. A method that simultaneously accounts for all three mechanisms: direct interference ($\rightarrow$), contagion ($-$), and latent homophily ($\leftrightarrow$).

2. The assumptions of auto-g relies on prior knowledge & belief.

   We want a test to distinguish between interference due to contagion ($-$) and latent homophily ($\leftrightarrow$).

Intuition

Claim: contagion vs. latent homophily is distinguishable using an independence test.

Claim: contagion vs. latent homophily is distinguishable using an independence test.

Undirected Edge:

$$Y_1 \text{ --- } Y_2 \text{ --- } Y_3$$

$$Y_1 \not\perp\!\!\!\perp Y_3 \text{ and } Y_1 \perp\!\!\!\perp Y_3 \mid Y_2$$

Intuition

Claim: contagion vs. latent homophily is distinguishable using an independence test.

Undirected Edge:

$$Y_1 - Y_2 - Y_3$$

$$Y_1 \not\perp\!\!\!\perp Y_3 \text{ and } Y_1 \perp\!\!\!\perp Y_3 \mid Y_2$$

Bidirected Edge:

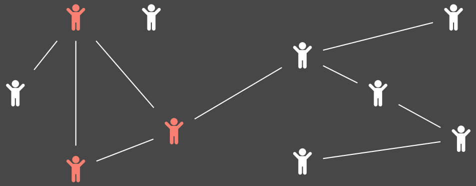
$$Y_1 \longleftrightarrow Y_2 \longleftrightarrow Y_3$$

$$Y_1 \perp\!\!\!\perp Y_3 \text{ and } Y_1 \not\perp\!\!\!\perp Y_3 \mid Y_2$$

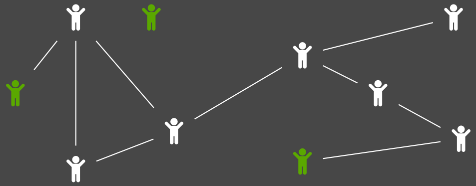
How To Get i.i.d. Samples for Independence Tests

How To Get i.i.d. Samples for Independence Tests

Intuition: further away in network $\approx$ less dependent.

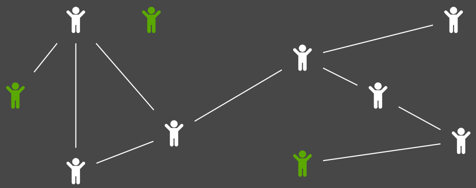


Intuition: further away in network $\approx$ less dependent.



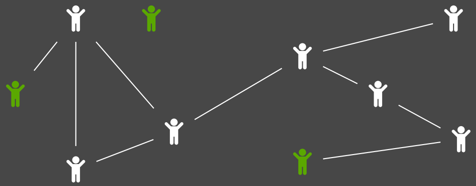
Our Proposed Test

Step 1: find a 5-hop independent set $\mathcal{I}$ from the network.



Our Proposed Test

Step 2: for each person in $\mathcal{I}$, collect information on their neighbors and their 2nd-order neighbors (i.e., neighbors' neighbors).



Our Proposed Test

Step 3: Is $\text{person}_i \perp\!\!\!\perp \text{2nd-order nb}(i) \mid \text{nb}(i)$?

Our Proposed Test

Step 3: Is $\text{person}_i \perp\!\!\!\perp \text{2nd-order nb}(i) \mid \text{nb}(i)$?

Likelihood ratio test:

- Model 1: $\text{person}_i \sim \text{nb}(i)$
- Model 2: $\text{person}_i \sim \text{nb}(i) + \text{2nd-order nb}(i)$

Our Proposed Test

Step 3: Is $\text{person}_i \perp\!\!\!\perp \text{2nd-order nb}(i) \mid \text{nb}(i)$?

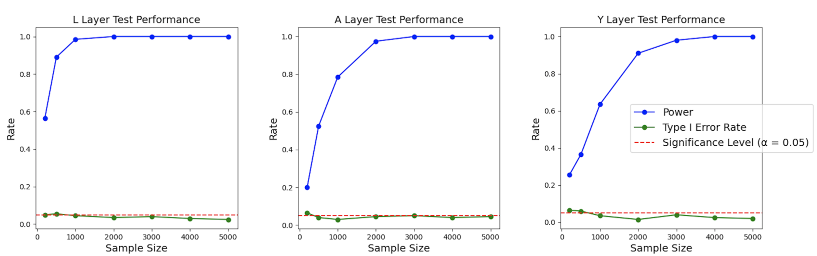
Likelihood ratio test:

- Model 1: $\text{person}_i \sim \text{nb}(i)$
- Model 2: $\text{person}_i \sim \text{nb}(i) + \text{2nd-order nb}(i)$

If $\perp\!\!\!\perp$, conclude contagion ($-$).

If $\not\perp\!\!\!\perp$, conclude latent homophily ($\leftrightarrow$).

Evaluating Our Test



Recap

1. A causal effect estimation method that allows for three types of dependence mechanisms: direct interference ($\rightarrow$), contagion ($-$), and latent homophily ($\leftrightarrow$).
2. The assumptions of auto-g relies on prior knowledge & belief.

   We want a test to distinguish between interference due to contagion ($-$) and latent homophily ($\leftrightarrow$).

Intuition

Why do we even need a new method when latent homophily ($\leftrightarrow$) is present?

[8] Lauritzen and Richardson 2002

Intuition

Why do we even need a new method when latent homophily ($\leftrightarrow$) is present?

Undirected Edge:

$$L_1 \;\text{———}\; L_2 \;\text{———}\; L_3$$

Gibbs factors [8]: $p(L_1 \mid L_2)$, $p(L_2 \mid L_1, L_3)$, and $p(L_3 \mid L_2)$

[8] Lauritzen and Richardson 2002

Intuition

Why do we even need a new method when latent homophily ($\leftrightarrow$) is present?

Undirected Edge:

$$L_1 \;—\; L_2 \;—\; L_3$$

Gibbs factors [8]: $p(L_1 \mid L_2)$, $p(L_2 \mid L_1, L_3)$, and $p(L_3 \mid L_2)$

Bidirected Edge:

$$L_1 \leftrightarrow L_2 \leftrightarrow L_3$$

$p(L_1, L_2, L_3)$

[8] Lauritzen and Richardson 2002

Intuition

Why do we even need a new method when latent homophily ($\leftrightarrow$) is present?

Undirected Edge:

$$L_1 \;\text{———}\; L_2 \;\text{———}\; L_3$$

Gibbs factors [8]: $p(L_1 \mid L_2)$, $p(L_2 \mid L_1, L_3)$, and $p(L_3 \mid L_2)$

Bidirected Edge:

$$L_1 \xleftrightarrow{\text{Cov(1,2)}} L_2 \xleftrightarrow{\text{Cov(2,3)}} L_3$$

$$p(L_1, L_2, L_3) \sim MVN(\mu, \Sigma)$$

[8] Lauritzen and Richardson 2002

Target for Causal Effect Estimation

Unit potential outcome expectation[9] for every $i \in V$:

$$\mathbb{E}[Y_i(a)] = \sum_{l} \mathbb{E}[Y_i \mid A = a, L] \times p(L)$$

[9]Tchetgen Tchetgen, Fulcher, and Shpitser 2021

Target for Causal Effect Estimation

Unit potential outcome expectation[9] for every $i \in V$:

$$\mathbb{E}[Y_i(a)] = \sum_l \mathbb{E}[Y_i \mid A = a, L] \times p(L)$$

Other targets such as average direct effect & average overall effect can be derived from $\mathbb{E}[Y_i(a)]$.

[9]Tchetgen Tchetgen, Fulcher, and Shpitser 2021

Target for Causal Effect Estimation

Unit potential outcome expectation[9] for every $i \in V$:

$$\mathbb{E}[Y_i(a)] = \sum_{l} \mathbb{E}[Y_i \mid A = a, L] \times p(L)$$

Other targets such as average direct effect & average overall effect can be derived from $\mathbb{E}[Y_i(a)]$.

[9] Tchetgen Tchetgen, Fulcher, and Shpitser 2021

Task 1: Estimate $p(L)$

Cannot estimate using empirical distribution under interference.

[10] Drton, Eichler, and Richardson 2009

Task 1: Estimate $p(L)$

Cannot estimate using empirical distribution under interference.

Contagion (–): use auto-g computation.

[10] Drton, Eichler, and Richardson 2009

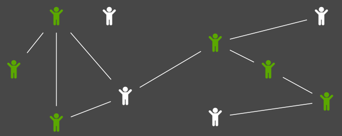
Task 1: Estimate $p(L)$

Cannot estimate using empirical distribution under interference.

Contagion (–): use auto-g computation.

Latent Homophily ($\leftrightarrow$):

1. Pre-processing: select triplets from the network s.t. none of these triplets are connected.



[10] Drton, Eichler, and Richardson 2009

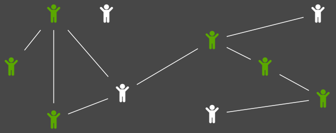
Task 1: Estimate p(L)

Cannot estimate using empirical distribution under interference.

Contagion ($-$): use auto-g computation.

Latent Homophily ($\leftrightarrow$):

1. Pre-processing: select triplets from the network s.t. none of these triplets are connected.



2. Assume $p(L) \sim MVN$ & Parameter-sharing.

[10] Drton, Eichler, and Richardson 2009

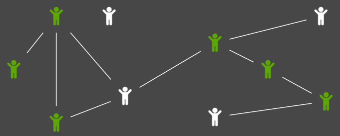
Task 1: Estimate p(L)

Cannot estimate using empirical distribution under interference.

Contagion (−): use auto-g computation.

Latent Homophily (↔):

1. Pre-processing: select triplets from the network s.t. none of these triplets are connected.



2. Assume $p(L) \sim MVN$ & Parameter-sharing.
3. Apply Residual Iterative Conditional Fitting (RICF) [10]: $\widehat{\mu}$, $\widehat{\Sigma}$.

[10] Drton, Eichler, and Richardson 2009

Task 2: Estimate $\mathbb{E}[Y_i \mid A, L]$

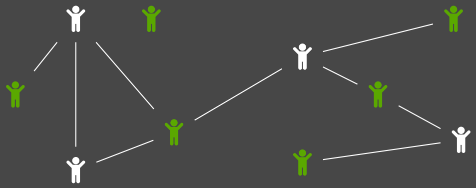
Task 2: Estimate $\mathbb{E}[Y_i \mid A, L]$

Contagion (–): use auto-g computation.

Task 2: Estimate $\mathbb{E}[Y_i \mid A, L]$

Contagion (–): use auto-g computation.

Latent Homophily ($\leftrightarrow$):



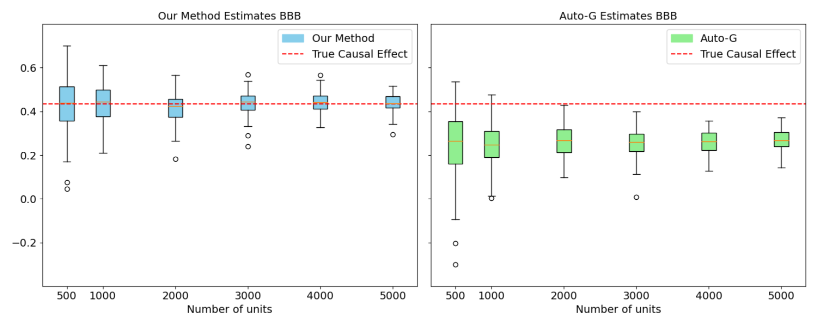
Directly estimate $\mathbb{E}[Y_i \mid A, L]$ using data from people in **an independent set** of the network.

Simulation Study 1

Latent homophily ($\leftrightarrow$) in all three (L, A, and Y) layers.

Simulation Study 1

Latent homophily ($\leftrightarrow$) in all three (L, A, and Y) layers.



Simulation Study 2

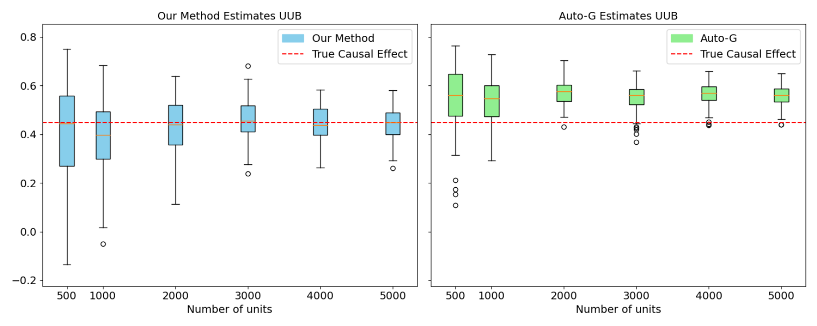
Contagion ($-$) in the L and A layers.

Latent homophily ($\leftrightarrow$) in the Y layers.

Simulation Study 2

Contagion (—) in the L and A layers.

Latent homophily ($\leftrightarrow$) in the Y layers.



References I

Bhattacharya, Rohit, Daniel Malinsky, and Ilya Shpitser (2020). "Causal inference under interference and network uncertainty". In: *Uncertainty in Artificial Intelligence*. PMLR, pp. 1028–1038.

Drton, Mathias, Michael Eichler, and Thomas S Richardson (2009). "Computing Maximum Likelihood Estimates in Recursive Linear Models with Correlated Errors.". In: *Journal of Machine Learning Research* 10.10.

Hudgens, Michael G and M Elizabeth Halloran (2008). "Toward causal inference with interference". In: *Journal of the American Statistical Association* 103.482, pp. 832–842.

Kang, Hyunseung and Guido Imbens (2016). "Peer encouragement designs in causal inference with partial interference and identification of local average network effects". In: *arXiv preprint arXiv:1609.04464*.

Lauritzen, Steffen L and Thomas S Richardson (2002). "Chain graph models and their causal interpretations". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 64.3, pp. 321–348.

Ogburn, Elizabeth L and Tyler J VanderWeele (2014). "Causal diagrams for interference". In.

References II

- Ogburn, Elizabeth L et al. (2024). "Causal inference for social network data". In: *Journal of the American Statistical Association* 119.545, pp. 597–611.
- Shalizi, Cosma Rohilla and Andrew C Thomas (2011). "Homophily and contagion are generically confounded in observational social network studies". In: *Sociological methods & research* 40.2, pp. 211–239.
- Shpitser, Ilya (2015). "Segregated graphs and marginals of chain graph models". In: *Advances in neural information processing systems* 28.
- Tchetgen, Eric J Tchetgen and Tyler J VanderWeele (2012). "On causal inference in the presence of interference". In: *Statistical methods in medical research* 21.1, pp. 55–75.
- Tchetgen Tchetgen, Eric J, Isabel R Fulcher, and Ilya Shpitser (2021). "Auto-g-computation of causal effects on a network". In: *Journal of the American Statistical Association* 116.534, pp. 833–844.

Thanks!

Questions?