Higher-Order Causal Message Passing for Experimentation with Complex Interference

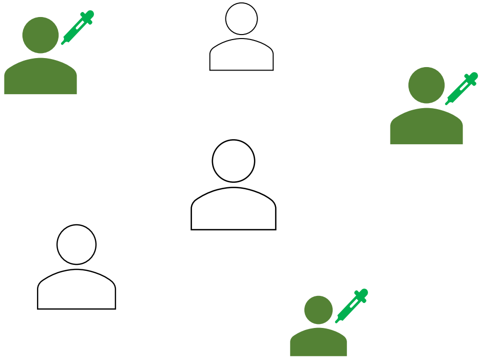
Mohsen Bayati[1], Yuwei Luo [1], Will Overman[1] Sadegh Shirani[1], and Ruoxuan Xiong[2]

1. Stanford Graduate School of Business, 2. Emory University

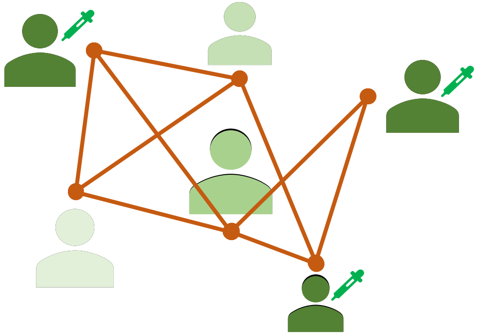
Motivating Example

- Estimating the causal effect of a proposed **treatment:**

Without Network Interference



With Network Interference



SUTVA (Stable Unit Treatment Values Assumption) fails due to **network interference**
(Cox 1958, Rubin 1978, Manski 1990, Imbens and Rubin 2015, Sussman and Airoldi 2017)

Causal Message-passing: Main Theory

- Observation: $\boldsymbol{W} = \begin{bmatrix} W_1^1 & \cdots & W_T^1 \\ \vdots & \ddots & \vdots \\ W_1^N & \cdots & W_T^N \end{bmatrix}$ and $\boldsymbol{Y}(\boldsymbol{W}) = \begin{bmatrix} Y_1^1(\boldsymbol{W}) & \cdots & Y_T^1(\boldsymbol{W}) \\ \vdots & \ddots & \vdots \\ Y_1^N(\boldsymbol{W}) & \cdots & Y_T^N(\boldsymbol{W}) \end{bmatrix}$
- Outcome model: $\vec{Y}_{t+1}(\boldsymbol{W}) = \boldsymbol{A}g\left(\vec{Y}_t(\boldsymbol{W}), \vec{W}_t, \boldsymbol{X}\right) + \text{noise}$
- Sample mean: $\nu_t^{\boldsymbol{W}} = \frac{1}{N}\sum_{i=1}^{N} Y_t^i(\boldsymbol{W})$ and sample variance: $\left(\rho_t^{\boldsymbol{W}}\right)^2 = \frac{1}{N}\sum_{i=1}^{N}\left(Y_t^i(\boldsymbol{W})\right)^2 - \left(\nu_t^{\boldsymbol{W}}\right)^2$

Causal-MP Main Theory (informal)
Under some regularity assumptions, **state evolution** equation holds.*

State Evolution equations

$$(\nu_{t+1}^W, \rho_{t+1}^W) = f_t(\nu_t^W, \rho_t^W, W)$$

*Shirani and Bayati. Causal message-passing for experiments with unknown and general network interference. *PNAS* 121.40 (2024).

Causal Message-passing: Estimation

- Observation: $\boldsymbol{W} = \begin{bmatrix} W_1^1 & \cdots & W_T^1 \\ \vdots & \ddots & \vdots \\ W_1^N & \cdots & W_T^N \end{bmatrix}$ and $\boldsymbol{Y}(\boldsymbol{W}) = \begin{bmatrix} Y_1^1(\boldsymbol{W}) & \cdots & Y_T^1(\boldsymbol{W}) \\ \vdots & \ddots & \vdots \\ Y_1^N(\boldsymbol{W}) & \cdots & Y_T^N(\boldsymbol{W}) \end{bmatrix}$
- Outcome specification: $\vec{Y}_{t+1}(\boldsymbol{W}) = \boldsymbol{A}g\left(\vec{Y}_t(\boldsymbol{W}), \vec{W}_t, \boldsymbol{X}\right) + \text{noise}$
- Sample mean: $\nu_t^{\boldsymbol{W}} = \frac{1}{N}\sum_{i=1}^N Y_t^i(\boldsymbol{W})$ and sample variance: $\left(\rho_t^{\boldsymbol{W}}\right)^2 = \frac{1}{N}\sum_{i=1}^N\left(Y_t^i(\boldsymbol{W})\right)^2 - \left(\nu_t^{\boldsymbol{W}}\right)^2$

The goal is to estimate f_t using the experimental data.

First-order CMP*

$$\nu_{t+1}^W = f_t(\nu_t^W, W)$$

Higher-order CMP#

$$(\nu_{t+1}^W, \rho_{t+1}^W) = f_t(\nu_t^W, \rho_t^W, W)$$

* Shirani and Bayati. Causal message-passing for experiments with unknown and general network interference. *PNAS* 121.40 (2024).

Bayati, Luo, Overman, Shirani, and Xiong. Higher-Order Causal Message Passing for Experimentation Under Unknown Interference. *NeurIPS* (2024)

Higher-Order Causal Message Passing Framework

- Goal: offer rich flexibility in estimating the unknown state evolution

$$(\nu_{t+1}^W, \rho_{t+1}^W) = f_t(\nu_t^W, \rho_t^W, W)$$

Feature vector: $\vec{x}_t = \vec{\phi}(\nu_t^W, (\rho_t^W)^2, W) = [\phi_1(\nu_t^W, (\rho_t^W)^2, W), \ldots, \phi_K(\nu_t^W, (\rho_t^W)^2, W)]$

- Proper features facilitate the extraction of informative patterns for learning the unknown state evolution
- Can be chosen based on heuristics, domain knowledge and prior information

Machine learning model: $(\hat{\nu}_{t+1}^W, (\hat{\rho}_{t+1}^W)^2) = f_\theta(\vec{x}_t)$

- Regression, MLP, tree-based models, etc.

Higher-Order Causal Message Passing Framework

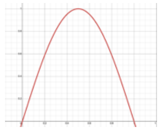
Examples of HO-CMP algorithms and their feature functions

Algorithms	Feature functions $\{\phi_k(\hat{\nu}_t(\boldsymbol{w}), \hat{\rho}_t(\boldsymbol{w})^2, \boldsymbol{w})\}_{k\in[K]}$	$f_{\boldsymbol{\theta}}(\cdot)$
FO-CMP	$\{\hat{\nu}_t(\boldsymbol{w}), \bar{w}_{t+1}, \hat{\nu}_t(\boldsymbol{w}) \cdot \bar{w}_t\}$	linear regression
HO-CMP	$\{\hat{\nu}_t(\boldsymbol{w}), \bar{w}_{t+1}, \hat{\nu}_t(\boldsymbol{w}) \cdot \bar{w}_t, \hat{\rho}_t(\boldsymbol{w})^2, \bar{w}_{t+1}^2\}$	linear regression

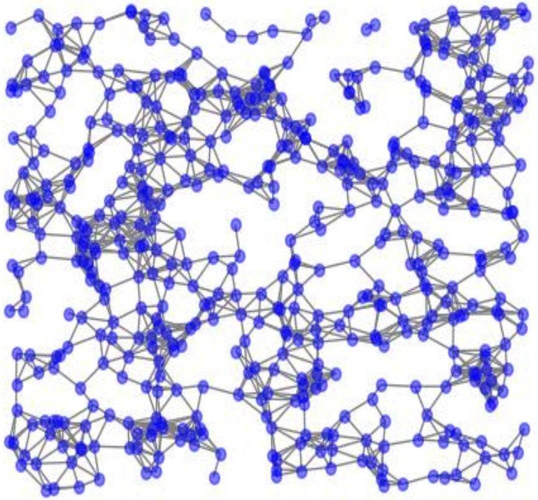
- FO-CMP: ν_{t+1}^{W} is a linear function of dynamics, treatments, and their interactions
- HO-CMP: introduces higher-order terms to model nonlinear effects

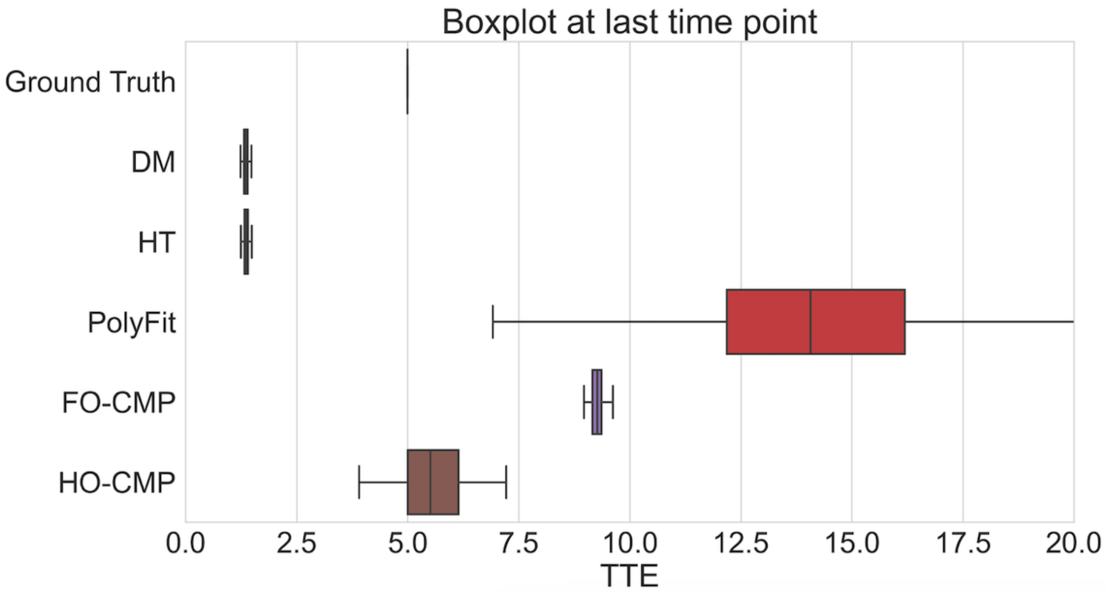
HO-CMP uses the observations of both sample mean and variance, hence modeling their potential interactions and improving the data efficiency

Estimation for Non-monotone Interference Effect

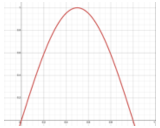


- Outcome(node i)= −1 + 0.8 Avg (outcomes of neighbors of i) + $1_{\text{i is treated}}$ + φ(fraction of treated neighbors of i)
- Random geometric graph; $\varphi(x) = \sin(\pi x)$; $\mathrm{T} = 40$

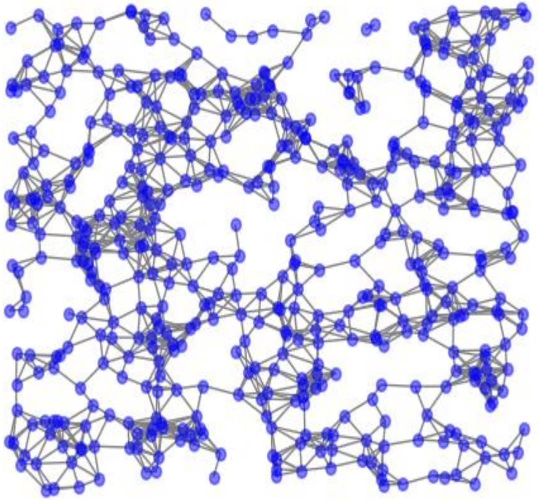


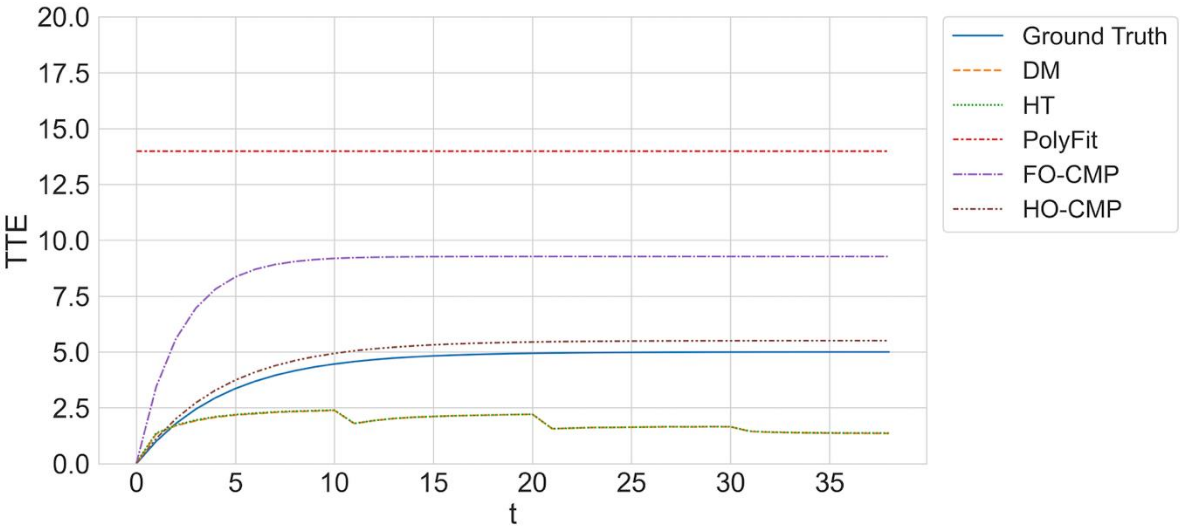


Estimation for Non-monotone Interference Effect

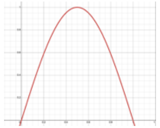


- Outcome(node i)= −1 + 0.8 Avg (outcomes of neighbors of i) + $1_{\text{i is treated}}$ + φ(fraction of treated neighbors of i)
- Random geometric graph; $\varphi(x) = \sin(\pi x)$; $\mathrm{T} = 40$

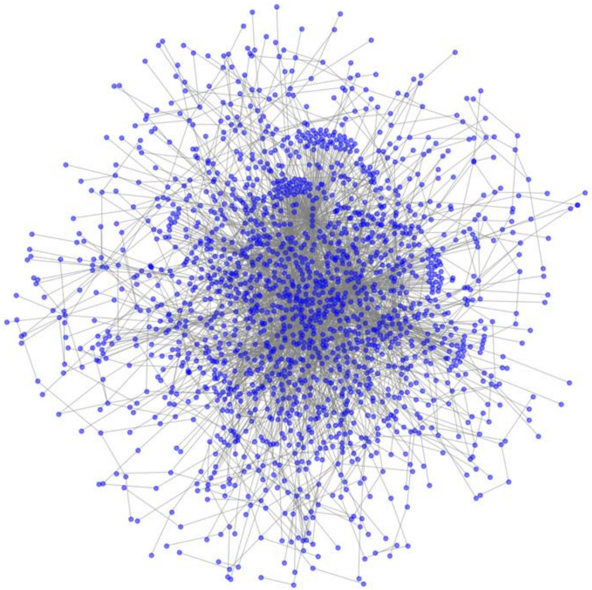


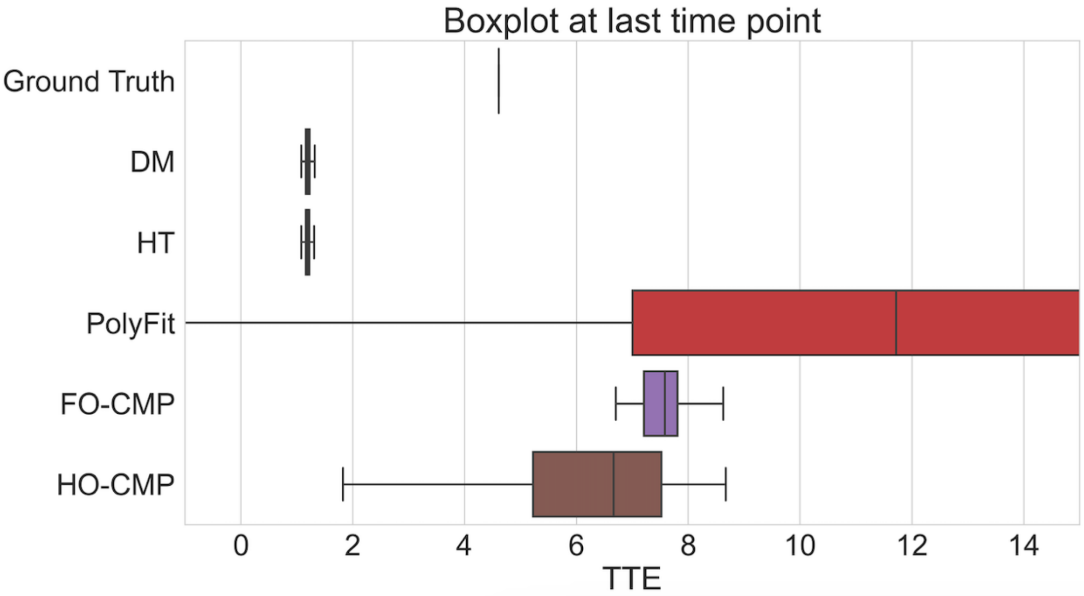


Estimation using Real Network Data (Non-Monotone Effect)



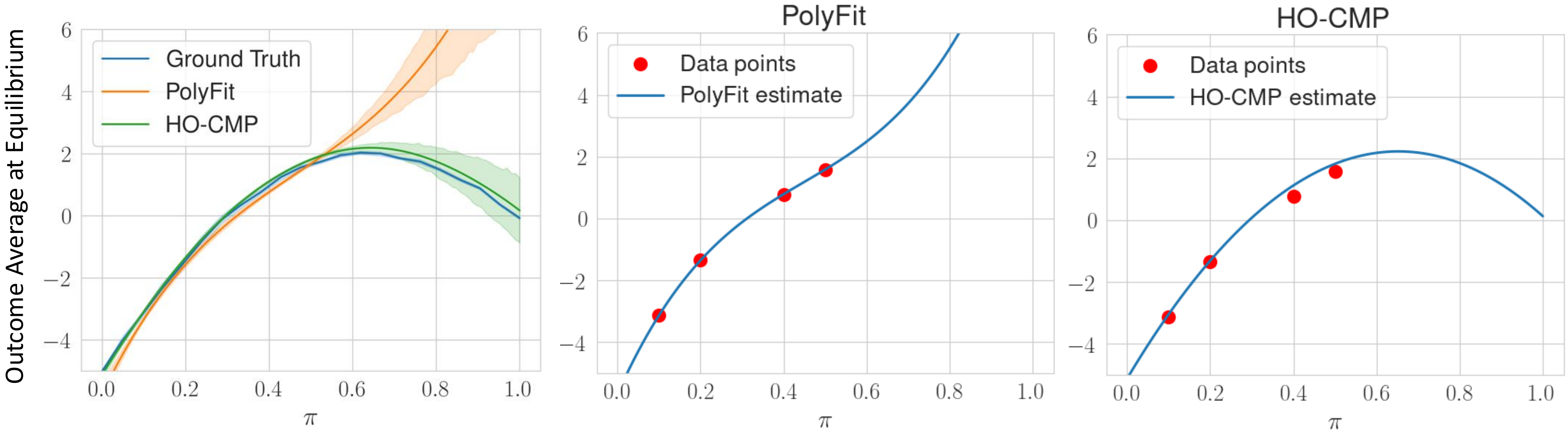
- Outcome(node i)= −1 + 0.8 Avg (outcomes of neighbors of i) + $1_{\text{i is treated}}$ + φ(fraction of treated neighbors of i)
- Social network of Twitch users; $\varphi(x) = \sin(\pi x)$; $\mathrm{T} = 40$





Improved Data Efficiency

Improved data efficiency enables HO-CMP to identify non-monotone effect with non-equilibrium data



Conclusion

HO-CMP:

- A method for estimating causal effects in experiments with unknown and general network interference.
- Efficient data usage using the whole dynamics rather than only the equilibrium
- Estimation robust to effect types (monotone vs. non-monotone) and graph structures (random vs. Twitch graph)

Thank you!