Bayesian inference of causal effects with proxy measurements of the interference network

Bar Weinstein
Tel Aviv University

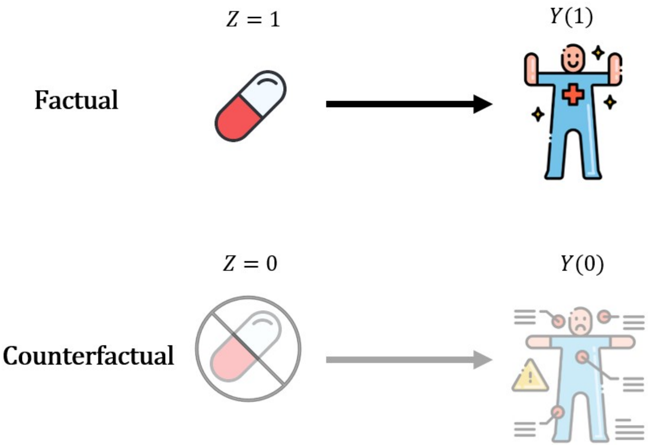
ISDSA Conference, May 2024

Joint work with Daniel Nevo

Outline

- Causal inference 101.
- Networked interference.
- Available proxies of the interference network.
- Inferential framework.

Causal Inference 101



Observation:

$$Y^{obs} = Y(1)$$

Causal effect:

$$Y(1) - Y(0)$$

Associations:

$$E[Y^{obs}|Z = 1] - E[Y^{obs}|Z = 0]$$

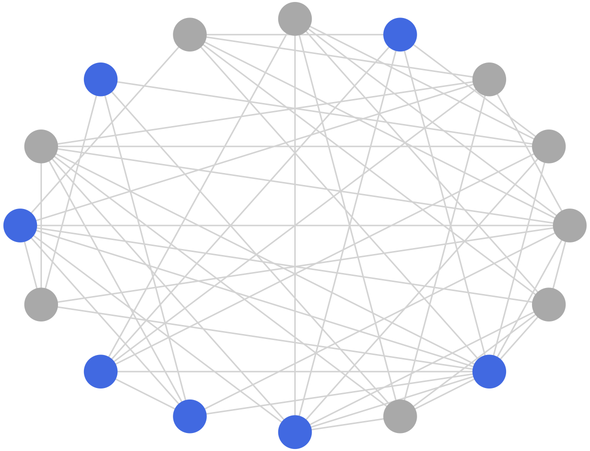
Interference

- Interference occurs when treatments of one unit affect the outcomes of other units.
- Units interact, resulting in the spread of treatment effects.
- Examples include
  1. Protection against infectious diseases (Halloran et al. 1995; Hayek et al. 2022)
  2. Spread of addictive drugs (Buchanan et al. 2023)
  3. A/B testing in online platforms (Aral 2016)
  4. Social influence between friends (Bearman et al. 2004; Christakis et al. 2007)
  5. Diffusion of financial information (Banerjee et al. 2013)

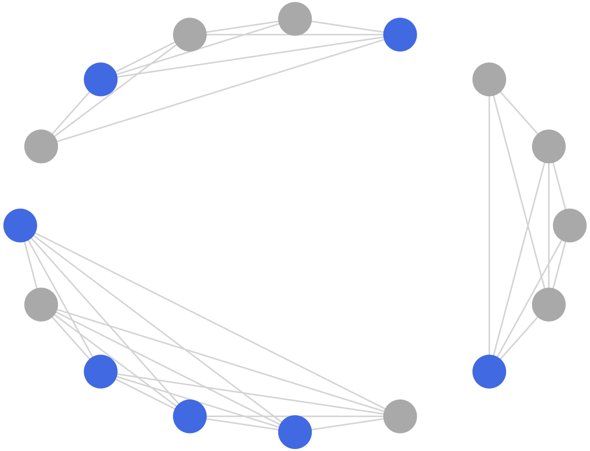
Interference

- Interference occurs when treatments of one unit affect the outcomes of other units.
- Units interact, resulting in the spread of treatment effects.
- Examples include
  1. Protection against infectious diseases (Halloran et al. 1995; Hayek et al. 2022)
  2. Spread of addictive drugs (Buchanan et al. 2023)
  3. A/B testing in online platforms (Aral 2016)
  4. Social influence between friends (Bearman et al. 2004; Christakis et al. 2007)
  5. Diffusion of financial information (Banerjee et al. 2013)
- Encompass various terms such as spillovers, peer effects, contamination, etc.
- Interference structure between units can be represented within a network.

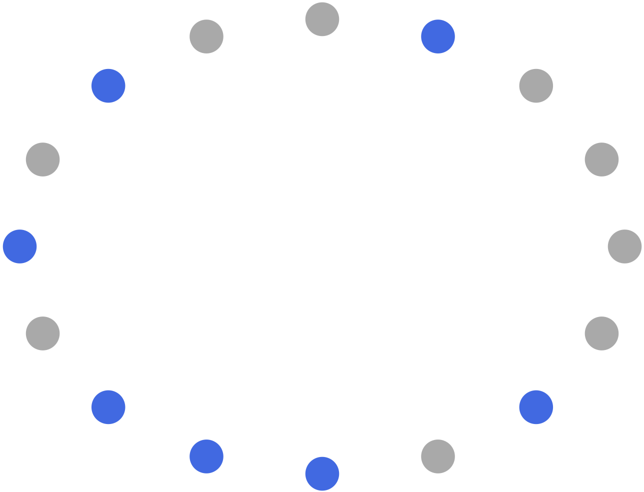
Networked interference



Networked interference



Networked interference



Formal settings

- Finite population of units $i \in \{1, \ldots, N\}$.
- $\boldsymbol{Z}$: population-level treatment vector with treatments' space $\mathcal{Z}$.
  - For example, for binary treatments $\mathcal{Z} \subseteq \{0, 1\}^N$.
- $Y_i(\boldsymbol{z})$: *potential* or *counterfactual* outcomes for each $\boldsymbol{z} \in \mathcal{Z}$.
  - Outcomes under hypothetical intervention that sets $\boldsymbol{Z} = \boldsymbol{z}$.

Formal settings

- Finite population of units $i \in \{1, \dots, N\}$.
- $\boldsymbol{Z}$: population-level treatment vector with treatments' space $\mathcal{Z}$.
  - For example, for binary treatments $\mathcal{Z} \subseteq \{0, 1\}^N$.
- $Y_i(\boldsymbol{z})$: *potential* or *counterfactual* outcomes for each $\boldsymbol{z} \in \mathcal{Z}$.
  - Outcomes under hypothetical intervention that sets $\boldsymbol{Z} = \boldsymbol{z}$.
- $\boldsymbol{X} \subseteq \mathbb{R}^{N \times q}$: covariates matrix.
- $\boldsymbol{A}^*$: adjacency matrix of the true interference network.
  - For simplicity, undirected and unweighted.

Causal estimands

- Comparisons of population-level interventions:
  1. **Static.** $\mu(\boldsymbol{z}) = N^{-1} \sum_{i=1}^{N} \mathbb{E}[Y_i(\boldsymbol{z}) | \boldsymbol{A}^*]$
  2. **Dynamic.** $\mu(h) = N^{-1} \sum_{i=1}^{N} \mathbb{E}[Y_i(h) | \boldsymbol{A}^*]$, for function $h \equiv h(\boldsymbol{X}, \boldsymbol{A}^*)$.
  3. **Stochastic.** $\mu(\alpha_0) = N^{-1} \sum_{i=1}^{N} \sum_{\boldsymbol{z} \in \mathcal{Z}} \pi_{\alpha_0}(\boldsymbol{z}) \mathbb{E}[Y_i(\boldsymbol{z}) | \boldsymbol{A}^*]$, for stochastic policy $\pi_{\alpha_0}(\boldsymbol{z})$.

Causal estimands

- Comparisons of population-level interventions:
  1. **Static.** $\mu(\boldsymbol{z}) = N^{-1} \sum_{i=1}^{N} \mathbb{E}[Y_i(\boldsymbol{z}) | \boldsymbol{A}^*]$
  2. **Dynamic.** $\mu(h) = N^{-1} \sum_{i=1}^{N} \mathbb{E}[Y_i(h) | \boldsymbol{A}^*]$, for function $h \equiv h(\boldsymbol{X}, \boldsymbol{A}^*)$.
  3. **Stochastic.** $\mu(\alpha_0) = N^{-1} \sum_{i=1}^{N} \sum_{\boldsymbol{z} \in \mathcal{Z}} \pi_{\alpha_0}(\boldsymbol{z}) \mathbb{E}[Y_i(\boldsymbol{z}) | \boldsymbol{A}^*]$, for stochastic policy $\pi_{\alpha_0}(\boldsymbol{z})$.

Assumptions and estimands are expressed via $\boldsymbol{A}^*$. Can we accurately measure it?

Measuring $\boldsymbol{A}^*$

- Accurately measuring networks of social interactions is a formidable task.
  - Social networks from self-reported elicitation of social relations.
  - Measured relations might not reflect those relevant to the studied mechanism.

Measuring $\boldsymbol{A}^*$

- Accurately measuring networks of social interactions is a formidable task.
  - Social networks from self-reported elicitation of social relations.
  - Measured relations might not reflect those relevant to the studied mechanism.
- However, it is commonly assumed that the measured network accurately represents the interference structure (Forastiere et al. 2020; Ogburn et al. 2022; Tchetgen Tchetgen et al. 2020).
- Misspecified network result in biased estimation (Weinstein et al. 2023).

Measuring $\boldsymbol{A}^*$

- Accurately measuring networks of social interactions is a formidable task.
  - Social networks from self-reported elicitation of social relations.
  - Measured relations might not reflect those relevant to the studied mechanism.
- However, it is commonly assumed that the measured network accurately represents the interference structure (Forastiere et al. 2020; Ogburn et al. 2022; Tchetgen Tchetgen et al. 2020).
- Misspecified network result in biased estimation (Weinstein et al. 2023).
- Assume that researchers observe proxy measurements $\boldsymbol{A} = (\boldsymbol{A}_1, \ldots, \boldsymbol{A}_B)$ of $\boldsymbol{A}^*$.
  - Single or multiple imprecisely measured networks.
  - Multiple networks from varying sources.
  - Multi-layer networks.

Measuring $\boldsymbol{A}^*$

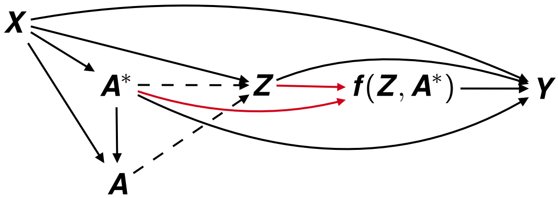
- Accurately measuring networks of social interactions is a formidable task.
  - Social networks from self-reported elicitation of social relations.
  - Measured relations might not reflect those relevant to the studied mechanism.
- However, it is commonly assumed that the measured network accurately represents the interference structure (Forastiere et al. 2020; Ogburn et al. 2022; Tchetgen Tchetgen et al. 2020).
- Misspecified network result in biased estimation (Weinstein et al. 2023).
- Assume that researchers observe proxy measurements $\boldsymbol{A} = (\boldsymbol{A}_1, \dots, \boldsymbol{A}_B)$ of $\boldsymbol{A}^*$.
  - Single or multiple imprecisely measured networks.
  - Multiple networks from varying sources.
  - Multi-layer networks.
- **Our goal: estimate causal effects using the proxy measurements $\boldsymbol{A}$.**

Structural model

- $\boldsymbol{Z}_{-i}$ affects Y_i only through values of fixed, possibly unknown, summarizing function of neighbors' treatments (Aronow et al. 2017).
- **Assumption 1.** *For any two treatments* $\boldsymbol{z}, \boldsymbol{z}' \in \mathcal{Z}$, *if* $z_i = z_i'$ *and* $f(\boldsymbol{z}_{-i}, \boldsymbol{A}_i^*) = f(\boldsymbol{z}'_{-i}, \boldsymbol{A}_i^*)$, *then* $Y_i(\boldsymbol{z}) = Y_i(\boldsymbol{z}')$.

Structural model

- $\boldsymbol{Z}_{-i}$ affects Y_i only through values of fixed, possibly unknown, summarizing function of neighbors' treatments (Aronow et al. 2017).
- **Assumption 1.** *For any two treatments $\boldsymbol{z}, \boldsymbol{z}' \in \mathcal{Z}$, if $z_i = z_i'$ and $f(\boldsymbol{z}_{-i}, \boldsymbol{A}_i^*) = f(\boldsymbol{z}'_{-i}, \boldsymbol{A}_i^*)$, then $Y_i(\boldsymbol{z}) = Y_i(\boldsymbol{z}')$.*
- Data are generated by sequentially evaluating a set of structural equations.
- Represented by a population-level directed acyclic graph (DAG):



Structural model

- $\boldsymbol{A}^*$ and $\boldsymbol{X}$ block all back-door paths between $\boldsymbol{Z}$ and $\boldsymbol{Y}$, i.e., $Y_i(\boldsymbol{z}) \perp\!\!\!\perp \boldsymbol{Z} | \boldsymbol{A}^*, \boldsymbol{X}$.
  - Accounting for latent network confounding $\boldsymbol{A}^* \leftarrow \boldsymbol{U} \rightarrow \boldsymbol{Y}$ is possible.

Structural model

- $\boldsymbol{A}^*$ and $\boldsymbol{X}$ block all back-door paths between $\boldsymbol{Z}$ and $\boldsymbol{Y}$, i.e., $Y_i(\boldsymbol{z}) \perp\!\!\!\perp \boldsymbol{Z} | \boldsymbol{A}^*, \boldsymbol{X}$.
  - Accounting for latent network confounding $\boldsymbol{A}^* \leftarrow \boldsymbol{U} \rightarrow \boldsymbol{Y}$ is possible.
- Require the specification of the following probabilistic models:
  1. *True network.* $p(\boldsymbol{A}^* | \boldsymbol{X})$.
     - Random graphs, stochastic block model, ERGM, latent space models, etc.

Structural model

- $\boldsymbol{A}^*$ and $\boldsymbol{X}$ block all back-door paths between $\boldsymbol{Z}$ and $\boldsymbol{Y}$, i.e., $Y_i(\boldsymbol{z}) \perp\!\!\!\perp \boldsymbol{Z} | \boldsymbol{A}^*, \boldsymbol{X}$.
  - Accounting for latent network confounding $\boldsymbol{A}^* \leftarrow \boldsymbol{U} \rightarrow \boldsymbol{Y}$ is possible.
- Require the specification of the following probabilistic models:
  1. *True network.* $p(\boldsymbol{A}^* | \boldsymbol{X})$.
     - Random graphs, stochastic block model, ERGM, latent space models, etc.
  2. *Observed proxies.* $p(\boldsymbol{A} | \boldsymbol{A}^*, \boldsymbol{X})$.
     - Measurement error models (differential or not) (e.g., Young et al. 2021).
     - Multi-layer network models with $\boldsymbol{A}^*$ being the “consensus" network (e.g., Sosa et al. 2022).

Structural model

- $\boldsymbol{A}^*$ and $\boldsymbol{X}$ block all back-door paths between $\boldsymbol{Z}$ and $\boldsymbol{Y}$, i.e., $Y_i(\boldsymbol{z}) \perp\!\!\!\perp \boldsymbol{Z} | \boldsymbol{A}^*, \boldsymbol{X}$.
  - Accounting for latent network confounding $\boldsymbol{A}^* \leftarrow \boldsymbol{U} \rightarrow \boldsymbol{Y}$ is possible.
- Require the specification of the following probabilistic models:
  1. *True network.* $p(\boldsymbol{A}^* | \boldsymbol{X})$.
     - Random graphs, stochastic block model, ERGM, latent space models, etc.
  2. *Observed proxies.* $p(\boldsymbol{A} | \boldsymbol{A}^*, \boldsymbol{X})$.
     - Measurement error models (differential or not) (e.g., Young et al. 2021).
     - Multi-layer network models with $\boldsymbol{A}^*$ being the "consensus" network (e.g., Sosa et al. 2022).
  3. *Outcome model.* $p(\boldsymbol{Y} | \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X})$.

Structural model

- $\boldsymbol{A}^*$ and $\boldsymbol{X}$ block all back-door paths between $\boldsymbol{Z}$ and $\boldsymbol{Y}$, i.e., $Y_i(\boldsymbol{z}) \perp\!\!\!\perp \boldsymbol{Z}|\boldsymbol{A}^*, \boldsymbol{X}$.
    - Accounting for latent network confounding $\boldsymbol{A}^* \leftarrow \boldsymbol{U} \rightarrow \boldsymbol{Y}$ is possible.
- Require the specification of the following probabilistic models:
    1. *True network.* $p(\boldsymbol{A}^*|\boldsymbol{X})$.
        - Random graphs, stochastic block model, ERGM, latent space models, etc.
    2. *Observed proxies.* $p(\boldsymbol{A}|\boldsymbol{A}^*, \boldsymbol{X})$.
        - Measurement error models (differential or not) (e.g., Young et al. 2021).
        - Multi-layer network models with $\boldsymbol{A}^*$ being the “consensus" network (e.g., Sosa et al. 2022).
    3. *Outcome model.* $p(\boldsymbol{Y}|\boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X})$.
- Few limitations (inc. dependence between units). The primary constraint is the computational burden.

Nonparameteric identification

- Observed outcomes Y_i are related to the potential outcomes $Y_i(\boldsymbol{z})$ as follows.
- **Assumption 2** (Consistency). *If* $\boldsymbol{Z} = \boldsymbol{z}$, *then* $Y_i = Y_i(\boldsymbol{z})$, $\forall \boldsymbol{z} \in \mathcal{Z}$

Nonparameteric identification

- Observed outcomes Y_i are related to the potential outcomes $Y_i(\boldsymbol{z})$ as follows.
- **Assumption 2** (Consistency). *If* $\boldsymbol{Z} = \boldsymbol{z}$, *then* $Y_i = Y_i(\boldsymbol{z})$, $\forall \boldsymbol{z} \in \mathcal{Z}$
- Causal estimands are identified via: $\mathbb{E}[Y_i(\boldsymbol{z}) | \boldsymbol{A}^*] = \mathbb{E}_{\boldsymbol{X}} \mathbb{E}[Y_i | \boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{A}^*, \boldsymbol{X}]$.

Nonparameteric identification

- Observed outcomes Y_i are related to the potential outcomes $Y_i(\boldsymbol{z})$ as follows.
- **Assumption 2** (Consistency). *If* $\boldsymbol{Z} = \boldsymbol{z}$, *then* $Y_i = Y_i(\boldsymbol{z})$, $\forall \boldsymbol{z} \in \mathcal{Z}$
- Causal estimands are identified via: $\mathbb{E}[Y_i(\boldsymbol{z})|\boldsymbol{A}^*] = \mathbb{E}_{\boldsymbol{X}}\mathbb{E}[Y_i|\boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{A}^*, \boldsymbol{X}]$.

- Further assuming that Y_i depends on $\boldsymbol{A}^*$ through summary statistics:
- **Assumption 3.** *For some fixed functions* ϕ_1, ϕ_2, ϕ_3,

$$p(Y_i|\boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}) = p(Y_i|Z_i, f(\phi_1(\boldsymbol{Z}_{-i}, \boldsymbol{A}_i^*)), \boldsymbol{X}_i, \phi_2(\boldsymbol{X}_{-i}, \boldsymbol{A}_i^*), \phi_{3,i}(\boldsymbol{A}^*)),\ \forall i$$

- yields

$$\mathbb{E}[Y_i(\boldsymbol{z})|\boldsymbol{A}^*] = \mathbb{E}_{\boldsymbol{X}}\mathbb{E}[Y_i|z_i, f(\phi_1), \boldsymbol{X}_i, \phi_2, \phi_{3,i}]$$

Bayesian inference

- We propose a Bayesian framework for inference.
- Denote the observed data by $\boldsymbol{O} = (\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{A})$.

Bayesian inference

- We propose a Bayesian framework for inference.
- Denote the observed data by $\boldsymbol{O} = (\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{A})$.
- Let η, θ, γ parameterize the outcome, network, and proxies distributions, respectively. They can be in a finite or infinite space. Assume prior independence.

Bayesian inference

- We propose a Bayesian framework for inference.
- Denote the observed data by $\boldsymbol{O} = (\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{A})$.
- Let η, θ, γ parameterize the outcome, network, and proxies distributions, respectively. They can be in a finite or infinite space. Assume prior independence.
- Posterior can be written as

$$
\begin{aligned}
p(\eta, \theta, \gamma, \boldsymbol{A}^* | \boldsymbol{O}) \propto\ & p(\boldsymbol{Y} | \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}, \eta) p(\eta) \\
& \cdot p(\boldsymbol{A} | \boldsymbol{A}^*, \boldsymbol{X}, \gamma) p(\gamma) \\
& \cdot p(\boldsymbol{A}^* | \boldsymbol{X}, \theta) p(\theta),
\end{aligned}
$$

- Mixed space of continuous and discrete latent variables. Discrete space consists of $\mathcal{O}(2^{N^2})$ terms!

Sampling from the posterior

- Metroplis-Hasting (and Gibbs) have slow convergence and poor mixing.
- Modern methods, e.g., Hamiltonian Monte Carlo, require continuous posterior. Extension for mixed spaces doesn't scale well (e.g., Nishimura et al. 2020; Zhou 2020).

Sampling from the posterior

- Metroplis-Hasting (and Gibbs) have slow convergence and poor mixing.
- Modern methods, e.g., Hamiltonian Monte Carlo, require continuous posterior. Extension for mixed spaces doesn't scale well (e.g., Nishimura et al. 2020; Zhou 2020).
- We propose a Bayesian modularization approach (Bayarri et al. 2009; Jacob et al. 2017). Posterior can be written as:

$$p(\eta, \theta, \gamma, \boldsymbol{A}^* | \boldsymbol{O}) \propto \underbrace{p(\eta | \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X})}_{\text{Outcome module}} \underbrace{p(\theta, \gamma, \boldsymbol{A}^* | \boldsymbol{X}, \boldsymbol{A})}_{\text{Network module}} \underbrace{p(\boldsymbol{Y} | \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X})}_{\text{Feedback term}}$$

Sampling from the posterior

- Metroplis-Hasting (and Gibbs) have slow convergence and poor mixing.
- Modern methods, e.g., Hamiltonian Monte Carlo, require continuous posterior. Extension for mixed spaces doesn't scale well (e.g., Nishimura et al. 2020; Zhou 2020).
- We propose a Bayesian modularization approach (Bayarri et al. 2009; Jacob et al. 2017). Posterior can be written as:

$$p(\eta, \theta, \gamma, \boldsymbol{A}^* | \boldsymbol{O}) \propto \underbrace{p(\eta | \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X})}_{\text{Outcome module}} \underbrace{p(\theta, \gamma, \boldsymbol{A}^* | \boldsymbol{X}, \boldsymbol{A})}_{\text{Network module}} \underbrace{p(\boldsymbol{Y} | \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X})}_{\text{Feedback term}}$$

- Use the cut-posterior that removes the feedback between modules:

$$\tilde{p}(\eta, \theta, \gamma, \boldsymbol{A}^* | \boldsymbol{O}) \propto p(\eta | \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}) p(\theta, \gamma, \boldsymbol{A}^* | \boldsymbol{X}, \boldsymbol{A})$$

Sampling from the posterior

- Sampling from the posterior is performed sequentially by sampling from each module:
  1. **Network.** $p(\theta, \gamma, \boldsymbol{A}^* | \boldsymbol{X}, \boldsymbol{A}) = p(\boldsymbol{A}^* | \boldsymbol{X}, \boldsymbol{A}, \theta, \gamma) \sum_{\boldsymbol{A}^*} p(\theta, \gamma, \boldsymbol{A}^* | \boldsymbol{X}, \boldsymbol{A}) \Rightarrow$ can be simplified!
  2. **Outcome.** $p(\eta | \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}) \propto p(\eta) p(\boldsymbol{Y} | \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}, \eta) \Rightarrow$ Bayesian regression.

Sampling from the posterior

- Sampling from the posterior is performed sequentially by sampling from each module:
  1. **Network.** $p(\theta, \gamma, \boldsymbol{A}^*|\boldsymbol{X}, \boldsymbol{A}) = p(\boldsymbol{A}^*|\boldsymbol{X}, \boldsymbol{A}, \theta, \gamma) \sum_{\boldsymbol{A}^*} p(\theta, \gamma, \boldsymbol{A}^*|\boldsymbol{X}, \boldsymbol{A}) \Rightarrow$ can be simplified!
  2. **Outcome.** $p(\eta|\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}) \propto p(\eta) p(\boldsymbol{Y}|\boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}, \eta) \Rightarrow$ Bayesian regression.
- In practice, performed via either:
  A. Sample (θ_m, γ_m). Then, sample $\boldsymbol{A}^*_m$. For each $\boldsymbol{A}^*_m$, sample η from the outcome module. Networks can be sampled via either
     i. For each (θ_m, γ_m), sample one network (*"Three-stage"*).
     ii. Sample networks using the posterior mean $\mathbb{E}[\theta, \gamma|\cdot]$ (*"Two-stage"*).

Sampling from the posterior

- Sampling from the posterior is performed sequentially by sampling from each module:
  1. **Network.** $p(\theta, \gamma, \boldsymbol{A}^*|\boldsymbol{X}, \boldsymbol{A}) = p(\boldsymbol{A}^*|\boldsymbol{X}, \boldsymbol{A}, \theta, \gamma) \sum_{\boldsymbol{A}^*} p(\theta, \gamma, \boldsymbol{A}^*|\boldsymbol{X}, \boldsymbol{A}) \Rightarrow$ can be simplified!
  2. **Outcome.** $p(\eta|\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}) \propto p(\eta)p(\boldsymbol{Y}|\boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}, \eta) \Rightarrow$ Bayesian regression.
- In practice, performed via either:
  - A. Sample (θ_m, γ_m). Then, sample $\boldsymbol{A}^*_m$. For each $\boldsymbol{A}^*_m$, sample η from the outcome module. Networks can be sampled via either
    - i. For each (θ_m, γ_m), sample one network (*"Three-stage"*).
    - ii. Sample networks using the posterior mean $\mathbb{E}[\theta, \gamma|\cdot]$ (*"Two-stage"*).
  - B. Sample multiple $\boldsymbol{A}^*$. Estimate summary statistics $\hat{\phi} = M^{-1} \sum_m \phi(\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{A}^*_m)$. Sample η by replacing ϕ with the plug-in estimator $\hat{\phi}$ (*"Plug-in"*).

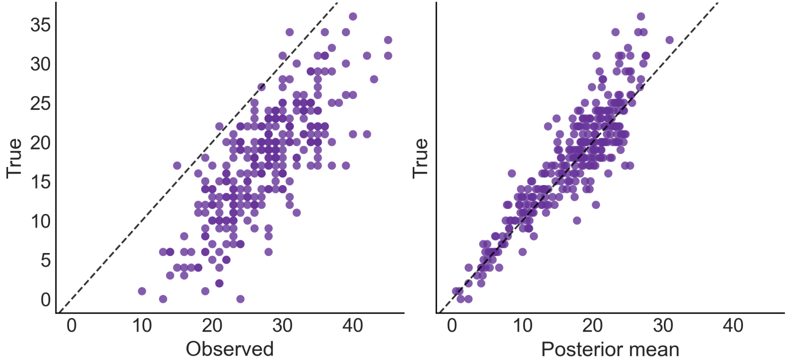
Numerical illustration

- Implemented in the probabilistic programming language *NumPyro*.
- Observed proxy network from a random noise measurement error model. $N = 300$.

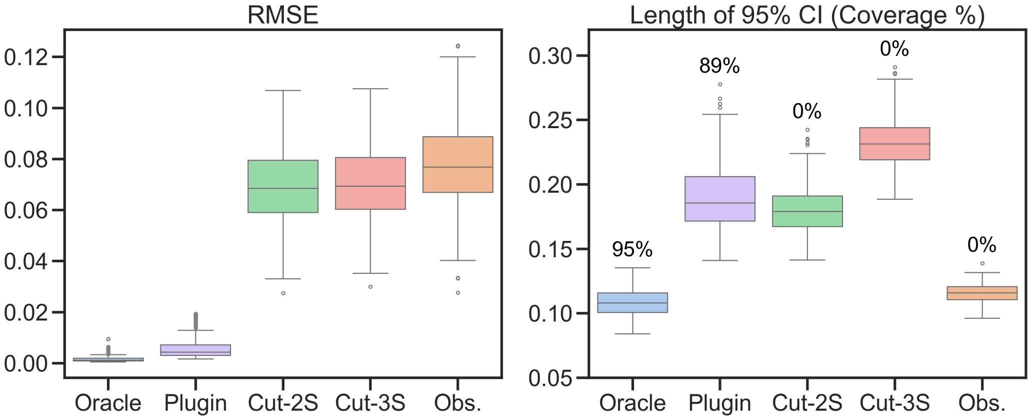
Numerical illustration

- Implemented in the probabilistic programming language *NumPyro*.
- Observed proxy network from a random noise measurement error model. $N = 300$.

Degree distribution



Numerical illustration



THANK YOU!

GitHub:// barwein

barwein@mail.tau.ac.il

Part I

Appendix

Postivity assumption

Assumption (positivity). $p(\boldsymbol{Z}_i = z, f(\boldsymbol{Z}_{-i}, \boldsymbol{A}_i^*) = c \mid \boldsymbol{X} = \boldsymbol{x}) > 0,\ \forall i, z \in \{0, 1\}, c \in \mathcal{C}, \boldsymbol{x}$.

Nonparameteric identification

$$
\begin{aligned}
\mathbb{E}[Y_i(\boldsymbol{z})|\boldsymbol{A}^*] &= \mathbb{E}_{\boldsymbol{X}}\mathbb{E}[Y_i(\boldsymbol{z})|\boldsymbol{A}^*, \boldsymbol{X}] \\
&= \mathbb{E}_{\boldsymbol{X}}\mathbb{E}[Y_i(\boldsymbol{z})|\boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{A}^*, \boldsymbol{X}] \\
&= \mathbb{E}_{\boldsymbol{X}}\mathbb{E}[Y_i|\boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{A}^*, \boldsymbol{X}] \\
&= \mathbb{E}_{\boldsymbol{X}}\mathbb{E}[Y_i|z_i, f(\phi_1(\boldsymbol{z}_{-i}, \boldsymbol{A}_i^*)), \boldsymbol{X}_i, \phi_2(\boldsymbol{X}_{-i}, \boldsymbol{A}_i^*), \phi_{3,i}(\boldsymbol{A}^*)]
\end{aligned}
$$

Nonparameteric identification – stochastic estimand

$$
\begin{aligned}
N^{-1}\sum_{i=1}^{N}\sum_{\boldsymbol{z}\in\mathcal{Z}}\pi_{\alpha_0}(\boldsymbol{z})\mathbb{E}[Y_i(\boldsymbol{z})|\boldsymbol{A}^*] &= N^{-1}\sum_{\boldsymbol{z}\in\mathcal{Z}}\pi_{\alpha_0}(\boldsymbol{z})\sum_{i=1}^{N}\mathbb{E}[Y_i(\boldsymbol{z})|\boldsymbol{A}^*] \\
&= N^{-1}\sum_{\boldsymbol{z}\in\mathcal{Z}}\pi_{\alpha_0}(\boldsymbol{z})\sum_{i=1}^{N}\mathbb{E}_{\boldsymbol{X}}\mathbb{E}[Y_i|\boldsymbol{Z}=\boldsymbol{z},\boldsymbol{A}^*,\boldsymbol{X}],
\end{aligned}
$$

DAG factorization

$$p(\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{A}, \boldsymbol{A}^*, \boldsymbol{X}) = p(\boldsymbol{Y}|\boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X})p(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{A}, \boldsymbol{A}^*)p(\boldsymbol{A}|\boldsymbol{A}^*, \boldsymbol{X})p(\boldsymbol{A}^*|\boldsymbol{X})p(\boldsymbol{X}),$$

Example – random noise

We observe single $\boldsymbol{A}$ generated by measuring $\boldsymbol{A}^*$ with random error independently across edges with false positive and negative rates γ_0, γ_1, respectively. Namely, $\Pr(A_{ij} = 1 - k | A^*_{ij} = k) = \gamma_k,\ k = 0, 1$, and we can write

$$p(\boldsymbol{A}|\boldsymbol{A}^*, \boldsymbol{X}, \gamma) = \prod_{i>j} \sum_{k=0,1} \mathbb{I}\{A^*_{ij} = k\} \gamma_k^{A_{ij}} (1 - \gamma_k)^{1-A_{ij}} = \prod_{i>j} \xi_{ij}(A^*_{ij}, \gamma),$$

where $\xi_{ij}(k, \gamma) = \gamma_k^{A_{ij}} (1 - \gamma_k)^{1-A_{ij}},\ k = 0, 1$. Extending to heterogeneous random noise that depends on covariates is possible as well.

Example – Repeated noisy measurements

Researchers observe $\boldsymbol{A} = (\boldsymbol{A}_1, \ldots, \boldsymbol{A}_B)$ networks each pertain to different noisy measurement of the true network (De Bacco et al. 2023; Redhead et al. 2023). A model can be assumed for each of the networks:

$$p(\boldsymbol{A}_b | \boldsymbol{A}^*, \boldsymbol{X}, \gamma_b) = \prod_{i>j} \xi_{b,ij}(A^*_{ij}, \gamma_b),$$

for some function $\xi_{b,ij}$. The joint distribution of $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_B$ can be assumed to be independent or dependent (e.g., via a hierarchical model).

Example – multi-layer networks

Researchers observe multi-layer collection of networks $\boldsymbol{A} = (\boldsymbol{A}_1, \dots, \boldsymbol{A}_B)$. Each $\boldsymbol{A}_b$ measures different social relationships. It is possible to model the observed multi-layer networks using LSM with shared latent positions:

$$p(\boldsymbol{A}_b | \boldsymbol{X}, \gamma_b, \boldsymbol{w}) = \prod_{i>j} \psi(\gamma_b - \|w_i - w_j\|)^{A_{b,ij}} (1 - \psi(\gamma_b - \|w_i - w_j\|))^{1 - A_{b,ij}},$$

for $\psi : \mathbb{R} \to [0, 1]$ and latent positions $w_i \in \mathbb{R}^d$. Covariates can also be included. Dependence between the networks can be modeled with multinomial (Salter-Townshend et al. 2017) or hierarchical (Sosa et al. 2022) models. The true network $\boldsymbol{A}^*$ can be viewed as a “consensus" or aggregate version of the observed networks. That can be achieved with a simple union (Salter-Townshend et al. 2017), or in a model-based approach (Sosa et al. 2022).

Feedback between modules

The posterior can be written as: $p(\eta, \theta, \gamma, \boldsymbol{A}^*|\boldsymbol{O}) = p(\eta|\boldsymbol{O}, \theta, \gamma, \boldsymbol{A}^*)p(\theta, \gamma, \boldsymbol{A}^*|\boldsymbol{O})$. Thus

$$\begin{aligned}
p(\theta, \gamma, \boldsymbol{A}^*|\boldsymbol{O}) &= \int_\eta p(\eta, \theta, \gamma, \boldsymbol{A}^*|\boldsymbol{O}) d\eta \\
&\propto p(\theta)p(\gamma)p(\boldsymbol{A}|\boldsymbol{A}^*, \boldsymbol{X}, \gamma)p(\boldsymbol{A}^*|\boldsymbol{X}, \theta)p(\boldsymbol{Y}|\boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}) \\
&\equiv p(\theta, \gamma, \boldsymbol{A}^*|\boldsymbol{X}, \boldsymbol{A})p(\boldsymbol{Y}|\boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X})
\end{aligned}$$

where $p(\boldsymbol{Y}|\boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}) = \int_\eta p(\boldsymbol{Y}|\boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}, \eta)p(\eta)d\eta$. In addition,

$$p(\eta|\boldsymbol{O}, \theta, \gamma, \boldsymbol{A}^*) = \frac{p(\eta, \theta, \gamma, \boldsymbol{A}^*|\boldsymbol{O})}{p(\theta, \gamma, \boldsymbol{A}^*|\boldsymbol{O})} \propto p(\boldsymbol{Y}|\boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}, \eta)p(\eta)$$

Sampling from network module

Assume that both the network generation model $p(\boldsymbol{A}^*|\boldsymbol{X},\theta)$ and the observed network model $p(\boldsymbol{A}|\boldsymbol{A}^*,\boldsymbol{X},\gamma)$ are dyad independent, i.e., can be written as

$$p(\boldsymbol{A}^*|\boldsymbol{X},\theta) = \prod_{i>j} \nu_{ij}(\boldsymbol{A}^*_{ij},\theta), \quad p(\boldsymbol{A}|\boldsymbol{A}^*,\boldsymbol{X},\gamma) = \prod_{i>j} \xi_{ij}(\boldsymbol{A}^*_{ij},\gamma),$$

for some functions ν_{ij}, ξ_{ij}. Thus,

$$\begin{aligned} p(\theta,\gamma|\boldsymbol{X},\boldsymbol{A}) &\propto p(\theta)p(\gamma)\sum_{\boldsymbol{A}^*}\prod_{i>j}\xi_{ij}(\boldsymbol{A}^*_{ij},\gamma)\nu_{ij}(\boldsymbol{A}^*_{ij},\theta) \\ &= p(\theta)p(\gamma)\prod_{i>j}\sum_{k=0,1}\xi_{ij}(k,\gamma)\nu_{ij}(k,\theta) \end{aligned}$$

Sampling networks

Generating samples of $\boldsymbol{A}^*$ is reduced to sampling edges independently from

$$p(\boldsymbol{A}^*|\boldsymbol{X},\boldsymbol{A},\theta,\gamma) = \frac{p(\theta,\gamma,\boldsymbol{A}^*|\boldsymbol{X},\boldsymbol{A})}{\sum_{\boldsymbol{A}^*} p(\theta,\gamma,\boldsymbol{A}^*|\boldsymbol{X},\boldsymbol{A})}$$
$$= \prod_{i>j} \frac{\xi_{ij}(\boldsymbol{A}^*_{ij},\gamma)\nu_{ij}(\boldsymbol{A}^*_{ij},\theta)}{\sum_{k=0,1}\xi_{ij}(k,\gamma)\nu_{ij}(k,\theta)},$$

Algorithm 1: Three-stage sampling

foreach $m = 1, \ldots, M$ **do**

  Sample $\theta_m, \gamma_m \sim p(\theta, \gamma | \boldsymbol{X}, \boldsymbol{A})$;

  Sample $\boldsymbol{A}^*_m \sim p(\boldsymbol{A}^* | \boldsymbol{X}, \boldsymbol{A}, \theta_m, \gamma_m)$;

  foreach $\ell = 1, \ldots, L$ **do**

    Sample $\eta_\ell \sim p(\eta | \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{A}^*_m, \boldsymbol{X})$;

  end

end

Two-stage sampling

Algorithm 2: Two-stage sampling

foreach $m = 1, \ldots, M$ **do**

  Sample $\theta_m, \gamma_m \sim p(\theta, \gamma | \boldsymbol{X}, \boldsymbol{A})$;

end

Compute posterior means $\bar{\theta} = M^{-1} \sum_m \theta_m$ and similarly for $\bar{\gamma}$;

foreach $m = 1, \ldots, M$ **do**

  Sample $\boldsymbol{A}_m^* \sim p(\boldsymbol{A}^* | \boldsymbol{X}, \boldsymbol{A}, \bar{\theta}, \bar{\gamma})$;

  foreach $\ell = 1, \ldots, L$ **do**

    Sample $\eta_\ell \sim p(\eta | \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{A}_m^*, \boldsymbol{X})$;

  end

end

Algorithm 3: Plug-in sampling

foreach $m = 1, \ldots, M$ **do**

  Sample $\theta_m, \gamma_m \sim p(\theta, \gamma | \boldsymbol{X}, \boldsymbol{A})$;

  Sample $\boldsymbol{A}_m^* \sim p(\boldsymbol{A}^* | \boldsymbol{X}, \boldsymbol{A}, \theta_m, \gamma_m)$;

end

Estimate the outcome models' summary statistics

$$\boldsymbol{A}^* \equiv \phi(\boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}) = M^{-1} \sum_m \phi(\boldsymbol{Z}, \boldsymbol{A}_m^*, \boldsymbol{X}) \; ;$$

foreach $\ell = 1, \ldots, L$ **do**

  Sample $\eta_\ell \sim p(\eta | \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X})$;

end

Simulations DGP

- Outcome model $Y_i = \eta_0 + \eta_1 Z_i + \eta_2 \sum_{j \neq i} Z_j A^*_{ij} + \eta_3 X_i + \varepsilon_i$
- Network generation $logit(\Pr(A^*_{ij} = 1)) = \theta_0 + \theta_1 |X_i - X_j|$.
- Observed network from a random noise measurement error model.
- $N = 300$ for 300 iterations.

References I

Aral, S. (2016). “Networked experiments”. In: *The Oxford handbook of the economics of networks*, pp. 376–411.

Aronow, P. M. and C. Samii (Dec. 2017). “Estimating average causal effects under general interference, with application to a social network experiment”. en. In: *The Annals of Applied Statistics* 11.4. ISSN: 1932-6157. DOI: 10.1214/16-AOAS1005. URL: https://projecteuclid.org/journals/annals-of-applied-statistics/volume-11/issue-4/Estimating-average-causal-effects-under-general-interference-with-application-to/10.1214/16-AOAS1005.full (visited on 08/01/2022).

Banerjee, A. et al. (July 2013). “The Diffusion of Microfinance”. In: *Science* 341.6144. DOI: 10.1126/science.1236498.

Bayarri, M. J., J. O. Berger, and F. Liu (Mar. 2009). “Modularization in Bayesian analysis, with emphasis on analysis of computer models”. In: *Bayesian Analysis* 4.1. Publisher: International Society for Bayesian Analysis, pp. 119–150. ISSN: 1936-0975, 1931-6690. DOI: 10.1214/09-BA404. URL: https://projecteuclid.org/journals/bayesian-analysis/volume-4/issue-1/Modularization-in-Bayesian-analysis-with-emphasis-on-analysis-of-computer/10.1214/09-BA404.full (visited on 03/18/2024).

Bearman, P. S. and J. Moody (2004). “Suicide and Friendships Among American Adolescents”. In: *American Journal of Public Health* 94.1. PMID: 14713704, pp. 89–95. DOI: 10.2105/AJPH.94.1.89. eprint: https://doi.org/10.2105/AJPH.94.1.89. URL: https://doi.org/10.2105/AJPH.94.1.89.

Buchanan, A. L. et al. (Feb. 2023). “Methods for Assessing Spillover in Network-Based Studies of HIV/AIDS Prevention among People Who Use Drugs”. In: *Pathogens* 12.2, p. 326. DOI: 10.3390/pathogens12020326.

Christakis, N. A. and J. H. Fowler (July 2007). “The Spread of Obesity in a Large Social Network over 32 Years”. In: *New England Journal of Medicine* 357.4, pp. 370–379. DOI: 10.1056/nejmsa066082.

References II

De Bacco, C. et al. (2023). “Latent network models to account for noisy, multiply reported social network data”. In: *Journal of the Royal Statistical Society Series A: Statistics in Society* 186.3, pp. 355–375.

Forastiere, L., E. M. Airoldi, and F. Mealli (June 2020). “Identification and Estimation of Treatment and Interference Effects in Observational Studies on Networks”. In: *Journal of the American Statistical Association* 116.534, pp. 901–918. DOI: 10.1080/01621459.2020.1768100.

Halloran, M. E. and C. J. Struchiner (Mar. 1995). “Causal Inference in Infectious Diseases”. In: *Epidemiology* 6.2, pp. 142–151. DOI: 10.1097/00001648-199503000-00010.

Hayek, S. et al. (2022). “Indirect protection of children from SARS-CoV-2 infection through parental vaccination”. In: *Science* 375.6585, pp. 1155–1159.

Jacob, P. E. et al. (Aug. 2017). “Better together? Statistical learning in models made of modules”. In: arXiv:1708.08719 [stat]. DOI: 10.48550/arXiv.1708.08719. URL: http://arxiv.org/abs/1708.08719 (visited on 03/18/2024).

Nishimura, A., D. B. Dunson, and J. Lu (June 2020). “Discontinuous Hamiltonian Monte Carlo for discrete parameters and discontinuous likelihoods”. In: *Biometrika* 107.2, pp. 365–380. ISSN: 0006-3444. DOI: 10.1093/biomet/asz083. URL: https://doi.org/10.1093/biomet/asz083 (visited on 03/14/2024).

Ogburn, E. L. et al. (2022). “Causal Inference for Social Network Data”. In: *Journal of the American Statistical Association*, pp. 1–46. DOI: 10.1080/01621459.2022.2131557. eprint: https://doi.org/10.1080/01621459.2022.2131557. URL: https://doi.org/10.1080/01621459.2022.2131557.

References III

Redhead, D., R. McElreath, and C. T. Ross (2023). “Reliable network inference from unreliable data: A tutorial on latent network modeling using STRAND.”. In: *Psychological methods*.

Salter-Townshend, M. and T. H. McCormick (2017). “Latent space models for multiview network data”. In: *The annals of applied statistics* 11.3, p. 1217.

Sosa, J. and B. Betancourt (May 2022). “A latent space model for multilayer network data”. In: *Computational Statistics & Data Analysis* 169, p. 107432. ISSN: 0167-9473. DOI: 10.1016/j.csda.2022.107432.

Tchetgen Tchetgen, E. J., I. R. Fulcher, and I. Shpitser (Oct. 2020). “Auto-G-Computation of Causal Effects on a Network”. In: *Journal of the American Statistical Association* 116.534, pp. 833–844. DOI: 10.1080/01621459.2020.1811098.

Weinstein, B. and D. Nevo (2023). “Causal inference with misspecified network interference structure”. In: *arXiv preprint arXiv:2302.11322*.

Young, J.-G., G. T. Cantwell, and M. E. J. Newman (Mar. 2021). “Bayesian inference of network structure from unreliable data”. en. In: *Journal of Complex Networks* 8.6. arXiv:2008.03334 [physics, stat], cnaa046. ISSN: 2051-1310, 2051-1329. DOI: 10.1093/comnet/cnaa046. URL: http://arxiv.org/abs/2008.03334 (visited on 08/01/2022).

Zhou, G. (2020). “Mixed Hamiltonian Monte Carlo for Mixed Discrete and Continuous Variables”. In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., pp. 17094–17104. URL: https://proceedings.neurips.cc/paper/2020/hash/c6a01432c8138d46ba39957a8250e027-Abstract.html (visited on 03/05/2024).