Regression under Interference in Connected Populations

Michael Schweinberger

with: Cornelius Fritz, Subhankar Bhadra, Jonathan R. Stewart, and David R. Hunter

The Pennsylvania State University

Outline

Interference in Connected Populations

Scalable Models

Interpretable Models

Scalable Methods

Theoretical Guarantees

Hate Speech on X

Structure

Interference in Connected Populations

Scalable Models

Interpretable Models

Scalable Methods

Theoretical Guarantees

Hate Speech on X

Networks Affect Outcomes

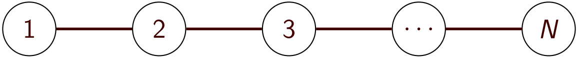
Networks Affect Outcomes

Networks affect individual and collective outcomes:

- **Global economy:** Markets (exchange networks) affect the economic welfare of billions of people around the world.
- **War and peace:** Relations between state and non-state actors affect the welfare of nations.
- **Epidemics:** Networks of contacts affect the spread of infectious diseases and public health.
- **Human brain:** Connections between neurons affect the ability of human beings to learn, create, and invent.
- **Causal inference under interference:** Connections allow the treatment assignments of units to affect the outcomes of other units directly (spillover) and indirectly (contagion).

Interference

Graph representing data structure: Population of N units connected by personal or professional relationships:



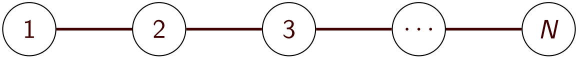
Graph representing model structure: Predictors $X_1, \ldots, X_N$ and outcomes $Y_1, \ldots, Y_N$:

- Main effect
- Spillover
- Contagion

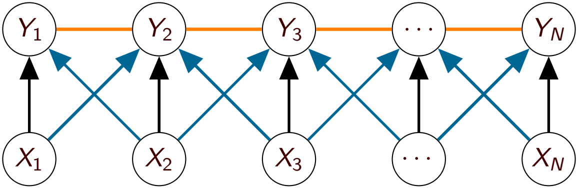
Dependent data without replications: e.g., X_1 can affect Y_N.

Interference

Graph representing data structure: Population of N units connected by personal or professional relationships:



Graph representing model structure: Predictors $X_1, \ldots, X_N$ and outcomes $Y_1, \ldots, Y_N$:

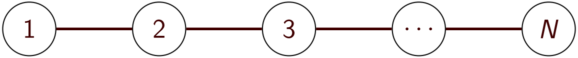


- Main effect
- Spillover
- Contagion

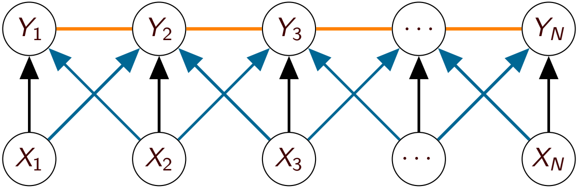
Dependent data without replications: e.g., X_1 can affect Y_N.

Interference

Graph representing data structure: Population of N units connected by personal or professional relationships:



Graph representing model structure: Predictors $X_1, \ldots, X_N$ and outcomes $Y_1, \ldots, Y_N$:



- Main effect
- Spillover
- Contagion

Dependent data without replications: e.g., X_1 can affect Y_N.

Literature

- **Conditional models of $\boldsymbol{Y} \mid (\boldsymbol{X}, \boldsymbol{Z})$:**

  Network-aware regression: e.g., Robins et al. (2001), Li et al. (2019), Zhu et al. (2020), Koskinen & Daraganova (2022), Le & Li (2022), Lei et al. (2024), Lunde et al. (2025).

  Causal inference under network interference: e.g., Ugander et al. (2013), Aronow & Samii (2017), Tchetgen Tchetgen et al. (2021), Li & Wager (2022), Ogburn et al. (2024).

- **Conditional models of $\boldsymbol{Z} \mid \boldsymbol{X}$:** e.g., Binkiewicz et al. (2017), Chandra et al. (2021), Huang et al. (2024), Huang et al. (2024), Stein et al. (2025).

- **Models of $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$:** Snijders et al. (2007), Fellows & Handcock (2012), Fosdick & Hoff (2015), but limited in scope and scalability, and without theoretical guarantees.

Models of $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ offer more insight and generalizability. By contrast, combining models of $\boldsymbol{Y} \mid (\boldsymbol{X}, \boldsymbol{Z})$ and $(\boldsymbol{X}, \boldsymbol{Z})$ assumes that $\boldsymbol{Z}$ was generated first and $\boldsymbol{Z}$ is not affected by $\boldsymbol{Y}$.

Regression Under Interference

Introduce comprehensive regression framework for studying relationships among attributes $(\boldsymbol{X}, \boldsymbol{Y})$ and connections $\boldsymbol{Z}$:

- **Scalable models,** capturing heterogeneity along with dependence among $(\boldsymbol{X}, \boldsymbol{Y})$ and $\boldsymbol{Z}$ in large populations.
- **Interpretable models,** based on extensions of Generalized Linear Models (GLMs) to dependent $(\boldsymbol{X}, \boldsymbol{Y})$ and $\boldsymbol{Z}$.
- **Scalable methods,** based on convex optimization of pseudolikelihoods, implemented by a divide and conquer approach using minorization-maximization (MM) methods.
- **Provable theoretical guarantees,** based on a single observation of dependent random variables $(\boldsymbol{X}, \boldsymbol{Y})$ and $\boldsymbol{Z}$.
- **Software:** R package `iglm` available on CRAN, which extends R functions `lm()` and `glm()` to regression under interference.

Regression Under Interference

Introduce comprehensive regression framework for studying relationships among attributes $(\boldsymbol{X}, \boldsymbol{Y})$ and connections $\boldsymbol{Z}$:

- **Scalable models,** capturing heterogeneity along with dependence among $(\boldsymbol{X}, \boldsymbol{Y})$ and $\boldsymbol{Z}$ in large populations.
- **Interpretable models,** based on extensions of Generalized Linear Models (GLMs) to dependent $(\boldsymbol{X}, \boldsymbol{Y})$ and $\boldsymbol{Z}$.
- **Scalable methods,** based on convex optimization of pseudolikelihoods, implemented by a divide and conquer approach using minorization-maximization (MM) methods.
- **Provable theoretical guarantees,** based on a single observation of dependent random variables $(\boldsymbol{X}, \boldsymbol{Y})$ and $\boldsymbol{Z}$.
- **Software:** R package `iglm` available on CRAN, which extends R functions `lm()` and `glm()` to regression under interference.

Regression Under Interference

Introduce comprehensive regression framework for studying relationships among attributes $(\boldsymbol{X}, \boldsymbol{Y})$ and connections $\boldsymbol{Z}$:

- **Scalable models,** capturing heterogeneity along with dependence among $(\boldsymbol{X}, \boldsymbol{Y})$ and $\boldsymbol{Z}$ in large populations.
- **Interpretable models,** based on extensions of Generalized Linear Models (GLMs) to dependent $(\boldsymbol{X}, \boldsymbol{Y})$ and $\boldsymbol{Z}$.
- **Scalable methods,** based on convex optimization of pseudolikelihoods, implemented by a divide and conquer approach using minorization-maximization (MM) methods.
- **Provable theoretical guarantees,** based on a single observation of dependent random variables $(\boldsymbol{X}, \boldsymbol{Y})$ and $\boldsymbol{Z}$.
- **Software:** `R` package `iglm` available on CRAN, which extends `R` functions `lm()` and `glm()` to regression under interference.

Regression Under Interference

Introduce comprehensive regression framework for studying relationships among attributes $(\boldsymbol{X}, \boldsymbol{Y})$ and connections $\boldsymbol{Z}$:

- **Scalable models,** capturing heterogeneity along with dependence among $(\boldsymbol{X}, \boldsymbol{Y})$ and $\boldsymbol{Z}$ in large populations.
- **Interpretable models,** based on extensions of Generalized Linear Models (GLMs) to dependent $(\boldsymbol{X}, \boldsymbol{Y})$ and $\boldsymbol{Z}$.
- **Scalable methods,** based on convex optimization of pseudolikelihoods, implemented by a divide and conquer approach using minorization-maximization (MM) methods.
- **Provable theoretical guarantees,** based on a single observation of dependent random variables $(\boldsymbol{X}, \boldsymbol{Y})$ and $\boldsymbol{Z}$.
- **Software:** `R` package `iglm` available on CRAN, which extends `R` functions `lm()` and `glm()` to regression under interference.

Regression Under Interference

Introduce comprehensive regression framework for studying relationships among attributes $(\boldsymbol{X}, \boldsymbol{Y})$ and connections $\boldsymbol{Z}$:

- **Scalable models,** capturing heterogeneity along with dependence among $(\boldsymbol{X}, \boldsymbol{Y})$ and $\boldsymbol{Z}$ in large populations.
- **Interpretable models,** based on extensions of Generalized Linear Models (GLMs) to dependent $(\boldsymbol{X}, \boldsymbol{Y})$ and $\boldsymbol{Z}$.
- **Scalable methods,** based on convex optimization of pseudolikelihoods, implemented by a divide and conquer approach using minorization-maximization (MM) methods.
- **Provable theoretical guarantees,** based on a single observation of dependent random variables $(\boldsymbol{X}, \boldsymbol{Y})$ and $\boldsymbol{Z}$.
- **Software:** `R` package `iglm` available on CRAN, which extends `R` functions `lm()` and `glm()` to regression under interference.

Regression Under Interference

Introduce comprehensive regression framework for studying relationships among attributes $(\boldsymbol{X}, \boldsymbol{Y})$ and connections $\boldsymbol{Z}$:

- **Scalable models,** capturing heterogeneity along with dependence among $(\boldsymbol{X}, \boldsymbol{Y})$ and $\boldsymbol{Z}$ in large populations.
- **Interpretable models,** based on extensions of Generalized Linear Models (GLMs) to dependent $(\boldsymbol{X}, \boldsymbol{Y})$ and $\boldsymbol{Z}$.
- **Scalable methods,** based on convex optimization of pseudolikelihoods, implemented by a divide and conquer approach using minorization-maximization (MM) methods.
- **Provable theoretical guarantees,** based on a single observation of dependent random variables $(\boldsymbol{X}, \boldsymbol{Y})$ and $\boldsymbol{Z}$.
- **Software:** `R` package `iglm` available on CRAN, which extends `R` functions `lm()` and `glm()` to regression under interference.

Data

Connected population of $N \geq 2$ units:

- one or more binary, count-, or real-valued predictors X_i: e.g., covariates or treatment assignments;
- binary, count-, or real-valued responses or outcomes Y_i;
- binary, count-, or real-valued connections $Z_{i,j}$, directed or not.

Structure

Interference in Connected Populations

Scalable Models

Interpretable Models

Scalable Methods

Theoretical Guarantees

Hate Speech on X

Scalable Models

How to extend GLMs for independent attributes (X_i, Y_i) to dependent attributes (X_i, Y_i) and connections $Z_{i,j}$?

$$
\begin{aligned}
f_{\theta}(\boldsymbol{y}, \boldsymbol{z} \mid \boldsymbol{x}) \;\propto\; & \prod_{i=1}^{N} a_{\mathcal{Y}}(y_i)\, e^{\boldsymbol{\theta}_g^{\top} g_i(\boldsymbol{x}_i,\, y_i)} \quad \text{GLM: } (X_i, Y_i) \\
\times\; & \prod_{i<j}^{N} a_{\mathcal{Z}}(z_{i,j})\, e^{\boldsymbol{\theta}_h^{\top} h_{i,j}(\boldsymbol{x}, y_i, y_j, \boldsymbol{z})} \quad \text{Interference}
\end{aligned}
$$

- Models with complex dependence can be constructed using two simple building blocks: g_i and $h_{i,j}$.
- The functions $h_{i,j}$ can induce dependence among responses Y_i and connections $Z_{i,j}$.
- To build scalable models that can accommodate small and large populations, each unit i has a (known) neighborhood $\mathcal{N}_i$, which can affect i's attributes (X_i, Y_i) and connections $Z_{i,j}$.

Scalable Models

How to extend GLMs for independent attributes (X_i, Y_i) to dependent attributes (X_i, Y_i) and connections $Z_{i,j}$?

$$f_{\boldsymbol{\theta}}(\boldsymbol{y}, \boldsymbol{z} \mid \boldsymbol{x}) \;\propto\; \prod_{i=1}^{N} a_{\mathcal{Y}}(y_i)\, e^{\boldsymbol{\theta}_g^\top g_i(\boldsymbol{x}_i,\, y_i)} \quad \text{GLM: } (X_i, Y_i)$$

$$\times \prod_{i<j}^{N} a_{\mathcal{Z}}(z_{i,j})\, e^{\boldsymbol{\theta}_h^\top h_{i,j}(\boldsymbol{x}, y_i, y_j, \boldsymbol{z})} \quad \text{Interference}$$

- Models with complex dependence can be constructed using two simple building blocks: g_i and $h_{i,j}$.
- The functions $h_{i,j}$ can induce dependence among responses Y_i and connections $Z_{i,j}$.
- To build scalable models that can accommodate small and large populations, each unit i has a (known) neighborhood $\mathcal{N}_i$, which can affect i's attributes (X_i, Y_i) and connections $Z_{i,j}$.

Scalable Models

How to extend GLMs for independent attributes (X_i, Y_i) to dependent attributes (X_i, Y_i) and connections $Z_{i,j}$?

$$f_{\boldsymbol{\theta}}(\boldsymbol{y}, \boldsymbol{z} \mid \boldsymbol{x}) \;\propto\; \prod_{i=1}^{N} a_{\mathcal{Y}}(y_i)\, e^{\boldsymbol{\theta}_g^\top g_i(\boldsymbol{x}_i, y_i)} \quad \text{GLM: } (X_i, Y_i)$$

$$\times \prod_{i<j}^{N} a_{\mathcal{Z}}(z_{i,j})\, e^{\boldsymbol{\theta}_h^\top h_{i,j}(\boldsymbol{x}, y_i, y_j, \boldsymbol{z})} \quad \text{Interference}$$

- Models with complex dependence can be constructed using two simple building blocks: g_i and $h_{i,j}$.
- The functions $h_{i,j}$ can induce dependence among responses Y_i and connections $Z_{i,j}$.
- To build scalable models that can accommodate small and large populations, each unit i has a (known) neighborhood $\mathcal{N}_i$, which can affect i's attributes (X_i, Y_i) and connections $Z_{i,j}$.

Scalable Models

How to extend GLMs for independent attributes (X_i, Y_i) to dependent attributes (X_i, Y_i) and connections $Z_{i,j}$?

$$f_{\boldsymbol{\theta}}(\boldsymbol{y}, \boldsymbol{z} \mid \boldsymbol{x}) \;\propto\; \prod_{i=1}^{N} a_{\mathcal{Y}}(y_i)\, e^{\boldsymbol{\theta}_g^\top g_i(\boldsymbol{x}_i, y_i)} \quad \text{GLM: } (X_i, Y_i)$$

$$\times \; \prod_{i<j}^{N} a_{\mathcal{Z}}(z_{i,j})\, e^{\boldsymbol{\theta}_h^\top h_{i,j}(\boldsymbol{x}, y_i, y_j, \boldsymbol{z})} \quad \text{Interference}$$

- Models with complex dependence can be constructed using two simple building blocks: g_i and $h_{i,j}$.
- The functions $h_{i,j}$ can induce dependence among responses Y_i and connections $Z_{i,j}$.
- To build scalable models that can accommodate small and large populations, each unit i has a (known) neighborhood $\mathcal{N}_i$, which can affect i's attributes (X_i, Y_i) and connections $Z_{i,j}$.

Scalable Models

How to extend GLMs for independent attributes (X_i, Y_i) to dependent attributes (X_i, Y_i) and connections $Z_{i,j}$?

$$
\begin{aligned}
f_{\boldsymbol{\theta}}(\boldsymbol{y}, \boldsymbol{z} \mid \boldsymbol{x}) \;\propto\; & \prod_{i=1}^{N} a_{\mathcal{Y}}(y_i)\, e^{\boldsymbol{\theta}_g^\top g_i(\boldsymbol{x}_i,\, y_i)} \quad \text{GLM: } (X_i, Y_i) \\
\times\; & \prod_{i<j}^{N} a_{\mathcal{Z}}(z_{i,j})\, e^{\boldsymbol{\theta}_h^\top h_{i,j}(\boldsymbol{x}, y_i, y_j, \boldsymbol{z})} \quad \text{Interference}
\end{aligned}
$$

- Models with complex dependence can be constructed using two simple building blocks: g_i and $h_{i,j}$.
- The functions $h_{i,j}$ can induce dependence among responses Y_i and connections $Z_{i,j}$.
- To build scalable models that can accommodate small and large populations, each unit i has a (known) neighborhood $\mathcal{N}_i$, which can affect i's attributes (X_i, Y_i) and connections $Z_{i,j}$.

Structure

Interference in Connected Populations

Scalable Models

Interpretable Models

Scalable Methods

Theoretical Guarantees

Hate Speech on X

Conditionals of Y_i can be represented by GLMs

Assume that

$$g_i(\boldsymbol{x}_i, y_i) \;\coloneqq\; g_{i,0}(\boldsymbol{x}_i) + g_{i,1}(\boldsymbol{x}_i)\, y_i$$

$$h_{i,j}(\boldsymbol{x}, y_i, y_j, \boldsymbol{z}) \;\coloneqq\; h_{i,j,0}(\boldsymbol{x}, y_j, \boldsymbol{z}) + h_{i,j,1}(\boldsymbol{x}, y_j, \boldsymbol{z})\, y_i$$

Then the conditional distribution of response Y_i can be represented by a GLM with linear predictor

$$\eta_i \;\coloneqq\; \boldsymbol{\theta}^\top \left(g_{i,1}(\boldsymbol{x}_i), \sum_{j:\, \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset} h_{i,j,1}(\boldsymbol{x}, y_j, \boldsymbol{z}) \right)$$

and conditional mean

$$\mu_i(\eta_i) \;=\; \nabla_{\eta_i}\, \psi \log \int_{\mathcal{Y}_i} a_{\mathcal{Y}}(y) \exp\left(\frac{\eta_i\, y}{\psi}\right) \mathrm{d}\,\nu_{\mathcal{Y}}(y)$$

Example of g_i and $h_{i,j}$

Models with complex dependence can be constructed using two simple building blocks:

$$g_i := \begin{pmatrix} y_i \\ x_i\, y_i \end{pmatrix} \quad \begin{pmatrix} \textit{intercept } \alpha_{\mathcal{Y}} \\ \textit{effect } \beta_{\mathcal{X},\mathcal{Y}} \textit{ of } x_i \textit{ on } Y_i \end{pmatrix}$$

$$h_{i,j} := \begin{pmatrix} \boldsymbol{e}_{i,j}\, z_{i,j} \\ -(1 - c_{i,j})\, z_{i,j} \log N \\ d_{i,j}(\boldsymbol{z})\, z_{i,j} \\ c_{i,j}\, (x_i\, y_j + x_j\, y_i)\, z_{i,j} \\ c_{i,j}\, y_i\, y_j\, z_{i,j} \end{pmatrix} \quad \begin{pmatrix} \textit{degree heterogeneity } \alpha_{\mathcal{Z},i},\, \alpha_{\mathcal{Z},j} \\ \textit{sparsity } \lambda \\ \textit{local transitivity } \gamma_{\mathcal{Z},\mathcal{Z}} \\ \textit{spillover } \gamma_{\mathcal{X},\mathcal{Y},\mathcal{Z}} \\ \textit{contagion } \gamma_{\mathcal{Y},\mathcal{Y},\mathcal{Z}} \end{pmatrix}$$

where the dimension of $\boldsymbol{\theta} := (\boldsymbol{\theta}_g,\, \boldsymbol{\theta}_h) \in \mathbb{R}^{N+6}$ increases with N.

Example: $Y_i \in \mathbb{R}$

Let

$$a_{\mathcal{Y}}(y_i) \quad := \quad \frac{1}{\sqrt{2\,\pi\,\psi}}\, \exp\left(-\frac{y_i^2}{2\,\psi}\right)\, \mathbb{I}(y_i \in \mathbb{R})$$

1. *Conditional distribution of response Y_i: $N(\mu_i(\eta_i),\, \psi)$*
2. *Conditional mean of response Y_i:*

$$\mu_i(\eta_i) \quad = \quad \alpha_{\mathcal{Y}} \quad + \quad \beta_{\mathcal{X},\mathcal{Y}}\, x_i \quad + \quad \gamma_{\mathcal{X},\mathcal{Y},\mathcal{Z}} \sum_{j:\, \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset} x_j\, z_{i,j} \quad + \quad \gamma_{\mathcal{Y},\mathcal{Y},\mathcal{Z}} \sum_{j:\, \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset} y_j\, z_{i,j}$$

capturing **spillover** and **contagion.**

Example: $Y_i \in \{0, 1, \ldots\}$

Let

$$a_{\mathcal{Y}}(y_i) \;\; := \;\; \frac{1}{y_i!}\, \mathbb{I}(y_i \in \{0, 1, \ldots\})$$

1. *Conditional distribution of response Y_i:* Poisson$(\mu_i(\eta_i))$
2. *Conditional mean of response Y_i:*

$$\mu_i(\eta_i) \;\; = \;\; \nabla_{\eta_i}\, b_i(\eta_i) \;\; = \;\; \nabla_{\eta_i}\, \exp(\eta_i) \;\; = \;\; \exp(\eta_i)$$

where

$$\begin{aligned} \eta_i \;\; = \;\; \alpha_{\mathcal{Y}} \;\; + \;\; \beta_{\mathcal{X},\mathcal{Y}}\, x_i \;\; &+ \;\; \gamma_{\mathcal{X},\mathcal{Y},\mathcal{Z}} \sum_{j:\, \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset} x_j\, z_{i,j} \\ &+ \;\; \gamma_{\mathcal{Y},\mathcal{Y},\mathcal{Z}} \sum_{j:\, \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset} y_j\, z_{i,j} \end{aligned}$$

capturing **spillover** and **contagion.**

Example: $Y_i \in \{0, 1\}$

Let

$$a_{\mathcal{Y}}(y_i) \quad := \quad \mathbb{I}(y_i \in \{0, 1\})$$

1. *Conditional distribution of response Y_i:* Bernoulli$(\mu_i(\eta_i))$
2. *Conditional mean of response Y_i:*

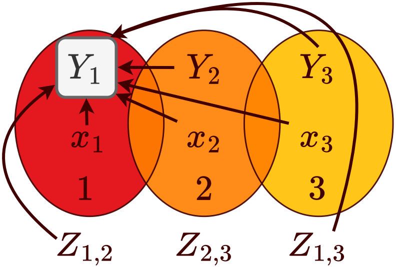
$$\mu_i(\eta_i) \;=\; \nabla_{\eta_i}\, b_i(\eta_i) \;=\; \nabla_{\eta_i}\, \log(1 + \exp(\eta_i)) \;=\; \text{logit}^{-1}(\eta_i)$$

where

$$\begin{aligned} \eta_i \quad = \quad \alpha_{\mathcal{Y}} \quad + \quad \beta_{\mathcal{X},\mathcal{Y}}\, x_i \quad &+ \quad \gamma_{\mathcal{X},\mathcal{Y},\mathcal{Z}} \sum_{j:\, \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset} x_j\, z_{i,j} \\ &+ \quad \gamma_{\mathcal{Y},\mathcal{Y},\mathcal{Z}} \sum_{j:\, \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset} y_j\, z_{i,j} \end{aligned}$$

capturing **spillover** and **contagion.**

Example: What affects response Y_1?



A population with $N = 3$ members 1, 2, and 3 with neighborhoods $\mathcal{N}_1 := \{2\}$, $\mathcal{N}_2 := \{1, 3\}$, and $\mathcal{N}_3 := \{2\}$.

The conditional distribution of reponse Y_1 is affected by

(a) predictor x_1,

(b) predictor x_2 and response Y_2 of neighbor $2 \in \mathcal{N}_1$,

(c) predictor x_3 and response Y_3 of non-neighbor $3 \notin \mathcal{N}_1$, because 1 and 3 have a common neighbor in $\mathcal{N}_1 \cap \mathcal{N}_3 = \{2\}$,

(d) connections $Z_{1,2}$ and $Z_{1,3}$.

Structure

Interference in Connected Populations

Scalable Models

Interpretable Models

Scalable Methods

Theoretical Guarantees

Hate Speech on X

Scalable Methods

Log-pseudolikelihood:

$$\ell(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z}) \;\; := \;\; \log \left[\prod_{i=1}^{N} f_{\boldsymbol{\theta}}(y_i \mid \text{others})\right] \left[\prod_{i<j}^{N} f_{\boldsymbol{\theta}}(z_{i,j} \mid \text{others})\right]$$

based on full conditionals of Y_i and $Z_{i,j}$, all of which are tractable.

Set of near-maximum pseudolikelihood estimators:

$$\widehat{\boldsymbol{\Theta}}(c_N) \; := \; \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : \; \|\nabla_{\boldsymbol{\theta}} \, \ell(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z})\|_\infty \; \leq \; c_N\}$$

where $c_N \geq 0$ can be interpreted as a convergence criterion.

Divide and Conquer

Partition the parameter vector $\boldsymbol{\theta} \coloneqq (\boldsymbol{\theta}_1,\, \boldsymbol{\theta}_2) \in \mathbb{R}^{N+6}$ into

- $\boldsymbol{\theta}_1 \coloneqq (\alpha_{\mathcal{Z},1},\, \ldots,\, \alpha_{\mathcal{Z},N}) \in \mathbb{R}^N$
- $\boldsymbol{\theta}_2 \coloneqq (\alpha_{\mathcal{Y}},\, \beta_{\mathcal{X},\mathcal{Y}},\, \gamma_{\mathcal{Z},\mathcal{Z}},\, \gamma_{\mathcal{X},\mathcal{Y},\mathcal{Z}},\, \gamma_{\mathcal{Y},\mathcal{Y},\mathcal{Z}},\, \lambda) \in \mathbb{R}^6$

and partition the negative Hessian of ℓ in accordance:

$$-\nabla^2_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z}) \quad \coloneqq \quad \begin{pmatrix} \boldsymbol{A}(\boldsymbol{\theta}) & \boldsymbol{B}(\boldsymbol{\theta}) \\ \boldsymbol{B}(\boldsymbol{\theta})^\top & \boldsymbol{C}(\boldsymbol{\theta}) \end{pmatrix}$$

The block $\boldsymbol{A}(\boldsymbol{\theta}) \in \mathbb{R}^{N \times N}$ contains N^2 elements.

Divide and Conquer with MM

Iterate:

Step 1: Given $\boldsymbol{\theta}_2^{(t)}$, find $\boldsymbol{\theta}_1^{(t+1)}$ such that

$$\ell(\boldsymbol{\theta}_1^{(t+1)}, \boldsymbol{\theta}_2^{(t)}; \boldsymbol{y}, \boldsymbol{z}) \;\; \geq \;\; \ell(\boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_2^{(t)}; \boldsymbol{y}, \boldsymbol{z})$$

Step 2: Given $\boldsymbol{\theta}_1^{(t+1)}$, find $\boldsymbol{\theta}_2^{(t+1)}$ such that

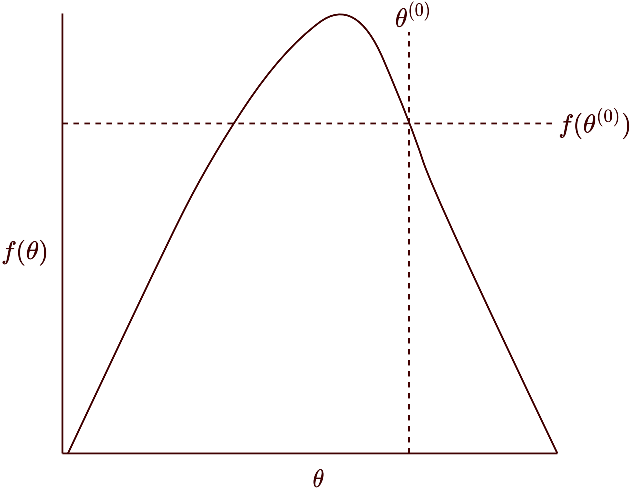
$$\ell(\boldsymbol{\theta}_1^{(t+1)}, \boldsymbol{\theta}_2^{(t+1)}; \boldsymbol{y}, \boldsymbol{z}) \;\; \geq \;\; \ell(\boldsymbol{\theta}_1^{(t+1)}, \boldsymbol{\theta}_2^{(t)}; \boldsymbol{y}, \boldsymbol{z})$$

Step 1 is implemented by a MM algorithm, replacing the inverse Hessian of ℓ (cost: $O(N^3)$) by a constant matrix (cost: $O(N)$).

Step 2 is implemented by a NR algorithm.

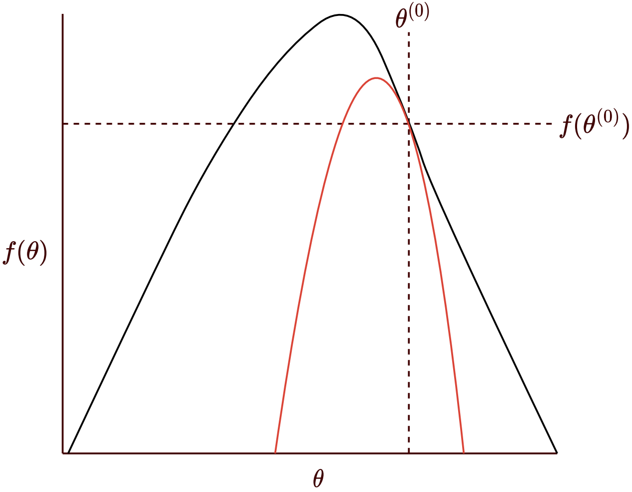
MM: Minorize-Maximize

Consider a generic objective function $f : \mathbb{R} \mapsto \mathbb{R}$:



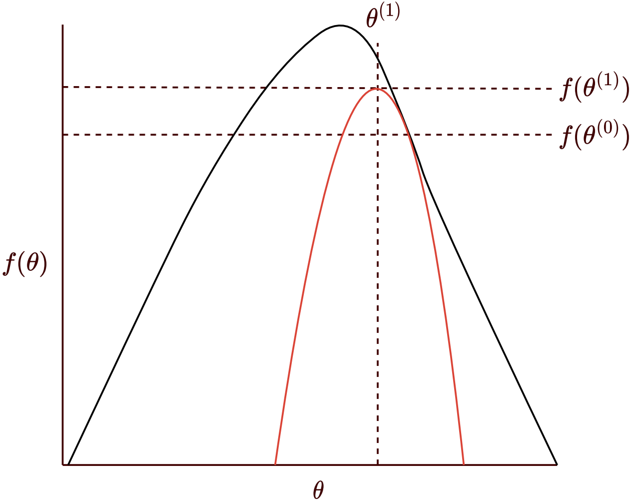
MM: Minorize-Maximize

Consider a generic objective function $f : \mathbb{R} \mapsto \mathbb{R}$:



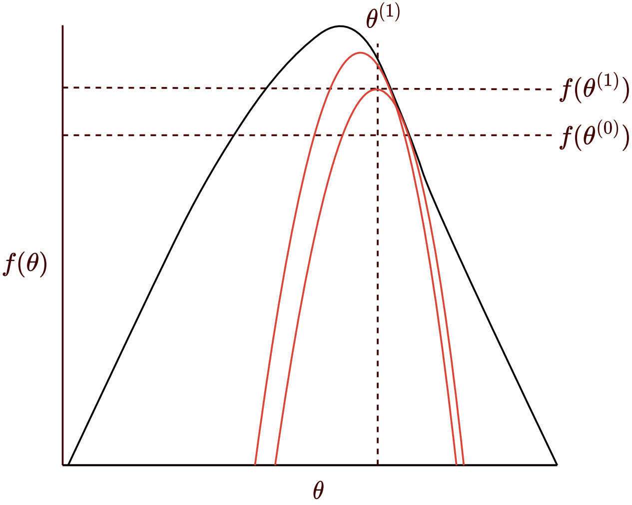
MM: Minorize-Maximize

Consider a generic objective function $f : \mathbb{R} \mapsto \mathbb{R}$:



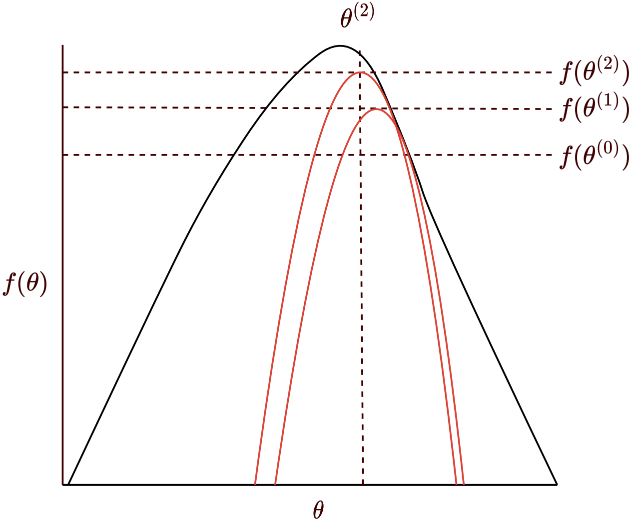
MM: Minorize-Maximize

Consider a generic objective function $f : \mathbb{R} \mapsto \mathbb{R}$:



MM: Minorize-Maximize

Consider a generic objective function $f : \mathbb{R} \mapsto \mathbb{R}$:



Minorizer of ℓ when $Z_{i,j} \in \{0, 1\}$

Let

$$\boldsymbol{A}^\star \quad := \quad \frac{1}{4}\left[(N-2) + \mathbf{1}\mathbf{1}^\top\right]$$

which is constant and can be inverted with $O(N)$ operations. Then

$$\begin{aligned} m(\boldsymbol{\theta}_1;\, \boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_2^{(t)}) \quad &:= \; \ell(\boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_2^{(t)}) \\ &+ \; \left(\nabla_{\boldsymbol{\theta}_1} \ell(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{(t)})\Big|_{\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{(t)}}\right)^\top (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_1^{(t)}) \\ &+ \; \frac{1}{2}\,(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_1^{(t)})^\top\,(-\boldsymbol{A}^\star)\,(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_1^{(t)}) \end{aligned}$$

is a minorizer of $\ell(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{(t)})$ at $\boldsymbol{\theta}_1^{(t)}$ for fixed $\boldsymbol{\theta}_2^{(t)}$:

$$\begin{aligned} m(\boldsymbol{\theta}_1;\, \boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_2^{(t)}) \quad &\leq \quad \ell(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{(t)}) \;\text{ for all }\; \boldsymbol{\theta}_1 \in \mathbb{R}^N \\ m(\boldsymbol{\theta}_1^{(t)};\, \boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_2^{(t)}) \quad &= \quad \ell(\boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_2^{(t)}) \end{aligned}$$

Structure

Interference in Connected Populations

Scalable Models

Interpretable Models

Scalable Methods

Theoretical Guarantees

Hate Speech on X

Theoretical Guarantees

Theorem 1 of Fritz et al. (2026)

Consider a single observation of $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$*, where* $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ *is a finite, countably infinite, or uncountable set. Then, under weak conditions,*

$$\widehat{\boldsymbol{\Theta}}(c_N) \quad \subseteq \quad \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \rho_N)$$

with high probability, provided N *is large enough.*

The error $\rho_N \propto c_N \, \Lambda_N(\boldsymbol{\theta}^\star)$ of the estimator depends on

- the strength of concentration of the gradient of ℓ via c_N;
- the inverse negative Hessian of ℓ via $\Lambda_N(\boldsymbol{\theta}^\star)$.

Example

Corollary 1 of Fritz et al. (2026): $Y_i, Z_{i,j} \in \{0, 1\}$

Under certain conditions on predictors, parameters, and dependence, for all large enough N*,*

$$L\sqrt{N \log N} \quad \leq \quad c_N \quad \leq \quad U\sqrt{N \log N}$$

and

$$\widehat{\boldsymbol{\Theta}}(c_N) \quad \subseteq \quad \mathcal{B}_\infty\left(\boldsymbol{\theta}^\star,\; C\sqrt{\frac{\log N}{N}}\right)$$

with probability at least $1 - 6/N^2$*.*

Convergence rate is comparable to the special case of the β-model, which is the best convergence rate one can hope for.

Peeking Under the Hood

Lemma 1 of Stewart & S (2026)

Let $\boldsymbol{\mu} : \mathbb{R}^p \mapsto \mathbb{R}^p$ be a homeomorphism and $\|\cdot\|$ be a vector norm with induced matrix norm $\|\!|\cdot|\!\|$. Consider any $\boldsymbol{\theta}^\star \in \mathbb{R}^p$ and $\epsilon \in (0, \infty)$, and define

$$\delta(\epsilon) \;\; := \inf_{\boldsymbol{\theta} \in \operatorname{bd} \mathcal{B}(\boldsymbol{\theta}^\star, \epsilon)} \|\boldsymbol{\mu}(\boldsymbol{\theta}) - \boldsymbol{\mu}(\boldsymbol{\theta}^\star)\|$$

If $\boldsymbol{\mu}(\boldsymbol{\theta})$ is continuously differentiable for all $\boldsymbol{\theta} \in \mathcal{B}(\boldsymbol{\theta}^\star, \epsilon)$ and $\mathcal{I}(\boldsymbol{\theta}) := \nabla_{\boldsymbol{\theta}}\, \boldsymbol{\mu}(\boldsymbol{\theta})$ is invertible for all $\boldsymbol{\theta} \in \mathcal{B}(\boldsymbol{\theta}^\star, \epsilon)$, then

$$\frac{\epsilon}{\sup_{\boldsymbol{\theta} \in \mathcal{B}(\boldsymbol{\theta}^\star, \epsilon)} \|\!|\mathcal{I}(\boldsymbol{\theta})^{-1}|\!\|} \;\; \leq \;\; \delta(\epsilon)$$

Lemma 1 helps "transport" concentration-of-measure between homeomorphic spaces, facilitating rates of convergence.

Peeking Under the Hood

Lemma 1 of Stewart & S (2026)

$$\frac{\epsilon}{\sup_{\boldsymbol{\theta} \in \mathcal{B}(\boldsymbol{\theta}^\star, \epsilon)} \|\mathcal{I}(\boldsymbol{\theta})^{-1}\|} \leq \delta(\epsilon)$$

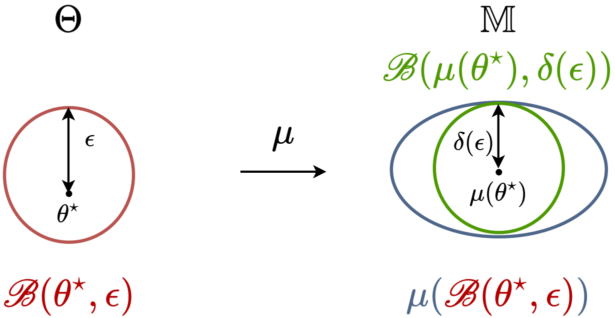
Example: Exponential family $f_{\boldsymbol{\theta}}(\boldsymbol{t}) \propto e^{\boldsymbol{\theta}^\top \boldsymbol{t}}$ (e.g., GLMs, Gaussians, Ising models, Markov random fields, undirected graphical models, Boltzmann machines in AI $\rightarrow$ Nobel prize 2024):

- $\boldsymbol{\theta} \in \mathbb{R}^p$ is the canonical parameter.
- $\boldsymbol{\mu}(\boldsymbol{\theta}) := \mathbb{E}_{\boldsymbol{\theta}} \boldsymbol{T} \in \mathbb{R}^p$ is the mean-value parameter.
- $\widehat{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}^\star$.
- $\boldsymbol{\mu}(\widehat{\boldsymbol{\theta}}) = \boldsymbol{T}$ is the MLE of $\boldsymbol{\mu}(\boldsymbol{\theta}^\star) = \mathbb{E}_{\boldsymbol{\theta}^\star} \boldsymbol{T}$.
- $\mathcal{I}(\boldsymbol{\theta})$ is the Fisher information matrix.

Peeking Under the Hood

Lemma 1 of Stewart & S (2026)

$$\frac{\epsilon}{\sup_{\boldsymbol{\theta} \in \mathcal{B}(\boldsymbol{\theta}^\star, \epsilon)} \| \mathcal{I}(\boldsymbol{\theta})^{-1} \|} \leq \delta(\epsilon)$$



We are interested in $\boldsymbol{\Theta}$, but $\mathbb{M}$ is more convenient.

Peeking Under the Hood

Lemma 1 of Stewart & S (2026)

$$\frac{\epsilon}{\sup_{\boldsymbol{\theta} \in \mathcal{B}(\boldsymbol{\theta}^\star, \epsilon)} \|\mathcal{I}(\boldsymbol{\theta})^{-1}\|} \leq \delta(\epsilon)$$

The map $\boldsymbol{\mu}$ is a homeomorphism. So, for all $\epsilon > 0$,

$$\begin{aligned}
\mathbb{P}(\widehat{\boldsymbol{\theta}} \in \mathcal{B}(\boldsymbol{\theta}^\star, \epsilon)) &= \mathbb{P}(\boldsymbol{\mu}(\widehat{\boldsymbol{\theta}}) \in \boldsymbol{\mu}(\mathcal{B}(\boldsymbol{\theta}^\star, \epsilon))) \\
&\geq \mathbb{P}(\boldsymbol{\mu}(\widehat{\boldsymbol{\theta}}) \in \mathcal{B}(\boldsymbol{\mu}(\boldsymbol{\theta}^\star), \delta(\epsilon))) \\
&\geq 1 - a(\delta(\epsilon)) \\
&\geq 1 - a\left(\frac{\epsilon}{\sup_{\boldsymbol{\theta} \in \mathcal{B}(\boldsymbol{\theta}^\star, \epsilon)} \|\mathcal{I}(\boldsymbol{\theta})^{-1}\|}\right)
\end{aligned}$$

Peeking Under the Hood

Lemma 1 of Stewart & S (2026)

$$\frac{\epsilon}{\sup_{\boldsymbol{\theta} \in \mathcal{B}(\boldsymbol{\theta}^\star, \epsilon)} \|\mathcal{I}(\boldsymbol{\theta})^{-1}\|} \leq \delta(\epsilon)$$

The map $\boldsymbol{\mu}$ is a homeomorphism. So, for all $\epsilon > 0$,

$$\begin{aligned}
\mathbb{P}(\widehat{\boldsymbol{\theta}} \in \mathcal{B}(\boldsymbol{\theta}^\star, \epsilon)) &= \mathbb{P}(\boldsymbol{\mu}(\widehat{\boldsymbol{\theta}}) \in \boldsymbol{\mu}(\mathcal{B}(\boldsymbol{\theta}^\star, \epsilon))) \\
&\geq \mathbb{P}(\boldsymbol{\mu}(\widehat{\boldsymbol{\theta}}) \in \mathcal{B}(\boldsymbol{\mu}(\boldsymbol{\theta}^\star), \delta(\epsilon))) \\
&\geq 1 - a(\delta(\epsilon)) \\
&\geq 1 - a\left(\frac{\epsilon}{\sup_{\boldsymbol{\theta} \in \mathcal{B}(\boldsymbol{\theta}^\star, \epsilon)} \|\mathcal{I}(\boldsymbol{\theta})^{-1}\|}\right)
\end{aligned}$$

where $a(.)$ is a non-increasing concentration function.

Peeking Under the Hood

Lemma 1 of Stewart & S (2026)

$$\frac{\epsilon}{\sup_{\boldsymbol{\theta} \in \mathcal{B}(\boldsymbol{\theta}^\star, \epsilon)} \|\mathcal{I}(\boldsymbol{\theta})^{-1}\|} \leq \delta(\epsilon)$$

The map $\boldsymbol{\mu}$ is a homeomorphism. So, for all $\epsilon > 0$,

$$\begin{aligned}
\mathbb{P}(\widehat{\boldsymbol{\theta}} \in \mathcal{B}(\boldsymbol{\theta}^\star, \epsilon)) &= \mathbb{P}(\boldsymbol{\mu}(\widehat{\boldsymbol{\theta}}) \in \boldsymbol{\mu}(\mathcal{B}(\boldsymbol{\theta}^\star, \epsilon))) \\
&\geq \mathbb{P}(\boldsymbol{\mu}(\widehat{\boldsymbol{\theta}}) \in \mathcal{B}(\boldsymbol{\mu}(\boldsymbol{\theta}^\star), \delta(\epsilon))) \\
&\geq 1 - a(\delta(\epsilon)) \\
&\geq 1 - a\left(\frac{\epsilon}{\sup_{\boldsymbol{\theta} \in \mathcal{B}(\boldsymbol{\theta}^\star, \epsilon)} \|\mathcal{I}(\boldsymbol{\theta})^{-1}\|}\right)
\end{aligned}$$

where $a(.)$ is a non-increasing concentration function.

Peeking Under the Hood

Lemma 1 of Stewart & S (2026)

$$\frac{\epsilon}{\sup_{\boldsymbol{\theta} \in \mathcal{B}(\boldsymbol{\theta}^\star, \epsilon)} \|\mathcal{I}(\boldsymbol{\theta})^{-1}\|} \leq \delta(\epsilon)$$

The map $\boldsymbol{\mu}$ is a homeomorphism. So, for all $\epsilon > 0$,

$$\begin{aligned}
\mathbb{P}(\widehat{\boldsymbol{\theta}} \in \mathcal{B}(\boldsymbol{\theta}^\star, \epsilon)) &= \mathbb{P}(\boldsymbol{\mu}(\widehat{\boldsymbol{\theta}}) \in \boldsymbol{\mu}(\mathcal{B}(\boldsymbol{\theta}^\star, \epsilon))) \\
&\geq \mathbb{P}(\boldsymbol{\mu}(\widehat{\boldsymbol{\theta}}) \in \mathcal{B}(\boldsymbol{\mu}(\boldsymbol{\theta}^\star), \delta(\epsilon))) \\
&\geq 1 - a(\delta(\epsilon)) \\
&\geq 1 - a\left(\frac{\epsilon}{\sup_{\boldsymbol{\theta} \in \mathcal{B}(\boldsymbol{\theta}^\star, \epsilon)} \|\mathcal{I}(\boldsymbol{\theta})^{-1}\|}\right)
\end{aligned}$$

where $a(.)$ is a non-increasing concentration function.

Peeking Under the Hood

Lemma 1 of Stewart & S (2026)

$$\frac{\epsilon}{\sup_{\boldsymbol{\theta} \in \mathcal{B}(\boldsymbol{\theta}^\star, \epsilon)} \|\mathcal{I}(\boldsymbol{\theta})^{-1}\|} \leq \delta(\epsilon)$$

The map $\boldsymbol{\mu}$ is a homeomorphism. So, for all $\epsilon > 0$,

$$\begin{aligned}
\mathbb{P}(\widehat{\boldsymbol{\theta}} \in \mathcal{B}(\boldsymbol{\theta}^\star, \epsilon)) &= \mathbb{P}(\boldsymbol{\mu}(\widehat{\boldsymbol{\theta}}) \in \boldsymbol{\mu}(\mathcal{B}(\boldsymbol{\theta}^\star, \epsilon))) \\
&\geq \mathbb{P}(\boldsymbol{\mu}(\widehat{\boldsymbol{\theta}}) \in \mathcal{B}(\boldsymbol{\mu}(\boldsymbol{\theta}^\star), \delta(\epsilon))) \\
&\geq 1 - a(\delta(\epsilon)) \\
&\geq 1 - a\left(\frac{\epsilon}{\sup_{\boldsymbol{\theta} \in \mathcal{B}(\boldsymbol{\theta}^\star, \epsilon)} \|\mathcal{I}(\boldsymbol{\theta})^{-1}\|}\right)
\end{aligned}$$

where $a(.)$ is a non-increasing concentration function.

Peeking Under the Hood

In general, the rate of convergence depends on:

- the information $\sup_{\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)} |||\mathcal{I}(\boldsymbol{\theta})^{-1}|||$;
- the strength of concentration:
  - the sample space and the tails of the distribution of $\boldsymbol{T}$;
  - the smoothness and dependence of $\boldsymbol{T}$.

Remarks:

- Can be extended from MLE to *M*-estimators (e.g., MPLE).
- Facilitates rates of convergence for *M*-estimators when $|\text{parameters}| = o(|\text{data}|)$.
- The data can be dependent: e.g., GLMs for dependent attributes and connections (special case: ERGMs), using additional structure to control dependence.

Structure

Interference in Connected Populations

Scalable Models

Interpretable Models

Scalable Methods

Theoretical Guarantees

Hate Speech on X

Hate Speech on X

Data shared by B. Desmarais: 109,974 posts by $N = 2{,}191$ U.S. state legislators between June 6, 2020 and January 6, 2021:

- $x_{i,1} = 1$ if legislator i is Republican and is 0 otherwise,
- $Y_i = 1$ if legislator i made at least one post with hate speech and is 0 otherwise based on Large Language Models (LLMs),
- $Z_{i,j} = 1$ if legislator i mentioned or retweeted legislator j in a tweet and is 0 otherwise,
- $\mathcal{N}_i$ is the set of legislators followed by legislator i, i.e., the set of legislators who can influence i.

Other predictors: gender ($x_{i,2}$), race ($x_{i,3}$), and state ($x_{i,4}$).

Hate Speech on X

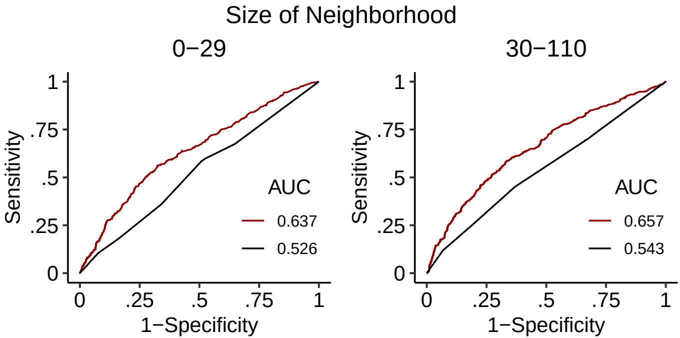
Hate Speech on X

Weight	Estimate	S.E.	Weight	Estimate	S.E.
$\alpha_{\mathcal{Y}}$	$-.893$	.143	$\gamma_{\mathcal{Z},\mathcal{Z},1}$	.604	.037
$\beta_{\mathcal{X},\mathcal{Y},1}$	$-.257$	.159	$\gamma_{\mathcal{Z},\mathcal{Z},2}$	2.57	.031
$\beta_{\mathcal{X},\mathcal{Y},2}$	$-.034$	.117	$\gamma_{\mathcal{X},\mathcal{Z},1}$	.035	.005
$\beta_{\mathcal{X},\mathcal{Y},3}$	.069	.095	$\gamma_{\mathcal{X},\mathcal{Z},2}$	.236	.015
$\gamma_{\mathcal{Y},\mathcal{Z}}$	$-.007$	.055	$\gamma_{\mathcal{X},\mathcal{Z},3}$	.756	.028
$\gamma_{\mathcal{X},\mathcal{Y},\mathcal{Z}}$	.038	.014	$\gamma_{\mathcal{X},\mathcal{Z},4}$	4.729	.041
			λ	.184	.005

The positive sign of $\widehat{\gamma}_{\mathcal{X},\mathcal{Y},\mathcal{Z}} = .038$ *suggests that the more Republicans interact with legislator i, the higher is the conditional probability of the event that legislator i uses offensive text in a post, holding everything else constant.*

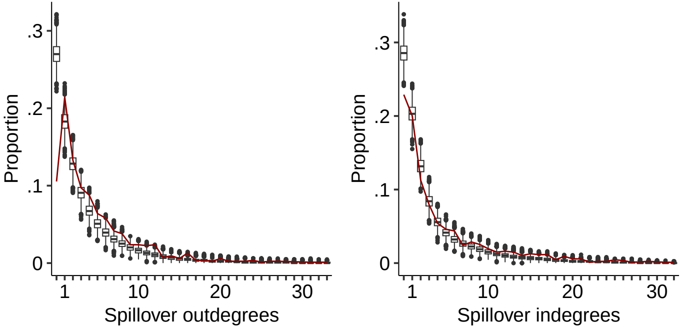
Compare model-based predictions:

- $Y_i \mid \boldsymbol{X}_i = \boldsymbol{x}_i$: logistic regression
- $Y_i \mid (\boldsymbol{X}, \boldsymbol{Y}_{-i}, \boldsymbol{Z}) = (\boldsymbol{x}, \boldsymbol{y}_{-i}, \boldsymbol{z})$



Hate Speech on X

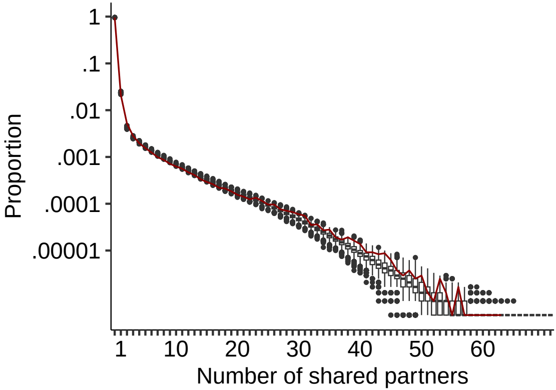
Model-based predictions of spillover in- and outdegrees of U.S. legislators in the subnetwork with *i* being Republican, *j* using offensive language, and the neighborhoods of *i* and *j* overlapping:



By construction, the connections in the subnetwork act as potential channels of spillover.

Hate Speech on X

Model-based predictions of the shared partners distribution of the network of repost and mention interactions of U.S. legislators:



Capturing closure in the network of U.S. legislators.

Publications

Fritz, S, Bhadra & Hunter (2026). *A regression framework for studying relationships among attributes under network interference.* Journal of the American Statistical Association, Theory & Methods. To appear.

Stewart & S (2026). *Pseudo-likelihood-based M-estimation of random graphs with dependent edges and parameter vectors of increasing dimension.* The Annals of Statistics, 54, 74–92.

Fritz & S (2026). `R` *package* `iglm`*: Regression under interference in connected populations.* `arXiv:2604.22791`

Support: NSF award DMS-2515763 (with Jonathan R. Stewart) and ARO award W911NF-21-1-0237 (with Johan Koskinen and Pavel N. Krivitsky).

Thank you!

Quantifying Uncertainty, Without Asymptotic Normality

The mean-value theorem (Ortega & Rheinboldt, pp. 68–69) implies that there exist real numbers $t_1, \ldots, t_p \in (0,\, 1)$ such that

$$\nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta};\, \boldsymbol{y},\, \boldsymbol{z})\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} - \nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta};\, \boldsymbol{y},\, \boldsymbol{z})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} = \boldsymbol{H}(\widehat{\boldsymbol{\theta}},\, \boldsymbol{\theta}^\star;\, \boldsymbol{y},\, \boldsymbol{z})\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star)$$

where

$$\boldsymbol{H}(\widehat{\boldsymbol{\theta}},\, \boldsymbol{\theta}^\star;\, \boldsymbol{y},\, \boldsymbol{z}) \;\; := \;\; \begin{pmatrix} g_1'(\boldsymbol{\theta}^\star + t_1\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star);\, \boldsymbol{y},\, \boldsymbol{z}) \\ \ldots \\ g_p'(\boldsymbol{\theta}^\star + t_p\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star);\, \boldsymbol{y},\, \boldsymbol{z}) \end{pmatrix}$$

As a result, the exact covariance matrix of $\widehat{\boldsymbol{\theta}}$ is

$$\mathbb{V}_{\boldsymbol{\theta}^\star}(\widehat{\boldsymbol{\theta}}) \;\; = \;\; \mathbb{V}_{\boldsymbol{\theta}^\star}\left\{-\boldsymbol{H}(\widehat{\boldsymbol{\theta}},\, \boldsymbol{\theta}^\star;\, \boldsymbol{Y}, \boldsymbol{Z})^{-1}\, \nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta};\, \boldsymbol{Y},\, \boldsymbol{Z})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star}\right\}$$

where $\boldsymbol{\theta}^\star$ can be approximated by $\widehat{\boldsymbol{\theta}}$ provided N is large enough.

Quantifying Uncertainty, Without Asymptotic Normality

Asymptotic normality in the presence of complex dependence is an open problem:

- Ising models and Markov random fields of $\boldsymbol{Y}$ and $\boldsymbol{Y} \mid \boldsymbol{X}$ with $p = 1$ and $p = 2$:
  - no asymptotic normality results (e.g., Ghosal and Mukherjee, 2020, Bhattacharya and Sen, 2024).
  - asymptotic normality in special cases, but the asymptotic variance cannot be characterized as a function of the model parameters (e.g., Martin-Löf, 1973; Bhattacharya et al., 2025).
- Coding estimators for models of $\boldsymbol{Y} \mid (\boldsymbol{X}, \boldsymbol{Z})$ (Tchetgen Tchetgen et al., 2021): strong independence assumptions and p is fixed.
- MLE (Stewart, 2025) and Stein estimators (Fischer et al., 2025) for models of $\boldsymbol{Z}$: local dependence assumptions.