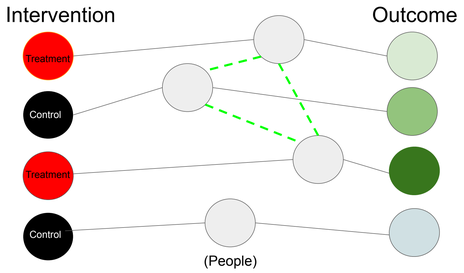
A General Method for Detecting & Characterizing Interference in Field Experiments

Connor T. Jerzak, Harvard University, Department of Government

Question: How can we understand causal effects when people could influence each other in the experiment, but we don't know exactly how?



Goal: (a) Discover the unobserved network & (b) understand how the treatment propagates through it

Statistical importance

- Network effects can ↑ standard errors [10]
- Unobserved network → bias [3]

Substantive importance

- Network effects can be > than main effects [6]
- Relevant for treatment targeting [11]

Existing approaches

– Test for spillovers in *a priori* clusters of units (e.g. two-stage randomization designs, [9, 2])

– Parametric models [8]

This approach

– Uses observed data to learn network dynamics

– One stage non-parametric estimator

Notation & Key Details

- n experimental subjects
- k pre-treatment covariates: $\mathbf{X} = (\mathbf{X}_1, ..., \mathbf{X}_n)'$
- Binary treatment vector, $\mathbf{W}$ (randomized)
- Y_i^{Obs} is the observed outcome
- **Treatment-fixed Network**, $\mathbf{A}$:

$$\mathbf{A}_{ij} = \begin{cases} 1 \text{ if } Y_i(w_1, w_2, ..., w_j, ..., w_n) \\ \quad \neq Y_i(w_1, w_2, ..., w'_j, ..., w_n) \\ \quad \text{for all } w_1, ..., w_n \\ 0 \text{ if } Y_i(w_1, w_2, ..., w_j, ..., w_n) \\ \quad = Y_i(w_1, w_2, ..., w'_j, ..., w_n) \\ \quad \text{for no } w_1, ..., w_n \end{cases}$$

- *Intuition*. Units (may) exist on network, which is independent of treatment assignment and specific to *outcome & intervention*

Step 1. Inferring Network

Can we use information from the realized treatment & outcome to learn about the network connection probabilities?

Yes, if we postulate an influence probability: $\mathbf{P}_{ij} = \Pr(\mathbf{A}_{ij} = 1|\mathbf{X})$. Estimation of $\hat{\mathbf{P}}$: latent variable model with parameters, $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = \text{argmax}_{\boldsymbol{\beta}} f(Y_i|\mathbf{W}, \mathbf{X}, \boldsymbol{\beta})$$
$$= \text{argmax}_{\boldsymbol{\beta}} \sum_{\mathbf{a} \in \mathcal{A}} f(Y_i|\mathbf{A}_i = \mathbf{a}, \mathbf{W}, \mathbf{X}, \boldsymbol{\beta}) \Pr(\mathbf{A}_i = \mathbf{a}|\mathbf{X}, \boldsymbol{\beta}).$$

$\hat{\mathbf{P}}_i = \Pr(\mathbf{A}_i|\mathbf{X}, \hat{\boldsymbol{\beta}}) \to \mathbf{P}_i$ if regularity conditions met [12]:

- $f(Y_i|\mathbf{A}_i, \cdot)$ smooth enough
- Network density decreases with n

Model example: "Homophily": $\mathbf{P}_{ij} \propto g(-d(\mathbf{x}_i, \mathbf{x}_j))$, where g is a monotonic function and d is a distance function between its two inputs.

- "Close units likely to influence each other"
- Generalizations available (d_i varies with $\mathbf{x}_i$)

Step 2. Defining a Feasible Network Effect

Can we use the network probabilities to define a feasible network effect on the outcome?

- Estimand using $\mathbf{A} \to$ not useful target ($\mathbf{A}$ is unobserved)
- Estimand using connection probabilities **is** useful target

Stochastic Intervention Estimand [4, 5, 7]:

$$\tau(\boldsymbol{\theta}) = \sum_{\mathbf{w} \in \mathcal{W}} \mathbb{E}[Y_i(\mathbf{w})] \Pr_{\boldsymbol{\theta}, \mathbf{P}_i}(\mathbf{W} = \mathbf{w}).$$

- **Intuition**: "What would my weighted expected outcome be if people were probabilistically assigned to receive treatment in proportion to their connection probability to me?" ($\boldsymbol{\theta}$ = parameters that define this policy)

Step 3. Estimating the Effect

What estimator can we use to target the network effect?

Non-parametric estimator for $\tau(\boldsymbol{\theta})$:

$$\hat{\tau}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} Y_i^{\text{Obs}} \cdot \frac{\Pr_{\boldsymbol{\theta}, \hat{\mathbf{P}}_i}(\mathbf{W} = \mathbf{w}^{\text{Obs}})}{\Pr_{\text{Gen}}(\mathbf{W} = \mathbf{w}^{\text{Obs}})},$$

where $\Pr_{\text{Gen}}(\mathbf{W} = \mathbf{w})$, is *known by design*.

- **Unbiased**. Estimator unbiased if $\hat{\mathbf{P}}_i = \mathbf{P}_i$ (network probabilities observed) and the following holds:

$$\frac{\Pr_{\boldsymbol{\theta}, \hat{\mathbf{P}}_i}(\mathbf{W} = \mathbf{w})}{\Pr_{\text{Gen}}(\mathbf{W} = \mathbf{w})} < \infty \quad \text{for all } \mathbf{w} \text{ almost surely}$$

- **Consistent**. Estimator consistent if $\hat{\mathbf{P}}_i \to \mathbf{P}_i$ and previous conditions hold.

Step 4. Assessing Statistical Significance Using Randomization Inference

How can we assess the significance of the network on the outcome?

- Sample $\mathbf{w}_{-i}^{(m)}$ from its distribution and re-calculate test statistic (i.e. $\hat{\tau}(\boldsymbol{\theta})$ or measure of fit)
  – Formally, $H_0 : Y_i \perp \mathbf{w}_{-i}|\mathbf{X}$ [3]
  ∗ p-value: $\frac{1}{M+1}\left(1 + \sum_{m=1}^{M} \mathbb{I}_{\{T(\mathbf{Y}, \mathbf{w}^{(m)}\mathbf{X}) \geq T(\mathbf{Y}, \mathbf{w}^{\text{Obs}}, \mathbf{X})\}}\right)$
  ∗ Under H_0, p-value, q, satisfies $\Pr(q \leq \alpha) \leq \alpha \; \forall \; \alpha$
  ∗ Computational burden: for each $\mathbf{w}^{(m)}$, re-train model
  – Differs from sharp null approach, hard to apply [1]

Validation via Simulation

$$\text{Outcome model} : Y_i = 0.1 \cdot w_i + \boldsymbol{\gamma}' \mathbf{x}_i + (\boldsymbol{\phi}' \mathbf{x}_i) \cdot w_i + m \cdot \log(\#\text{treated connections} + 1) + \epsilon_i$$

- *Interference structure*: homophily
- Vary $|m| \in \{0, 0.1, 0.5, 1\}$ (interference magnitude)
- Vary $k \in \{2, 10, 20\}$ (number of covariates)

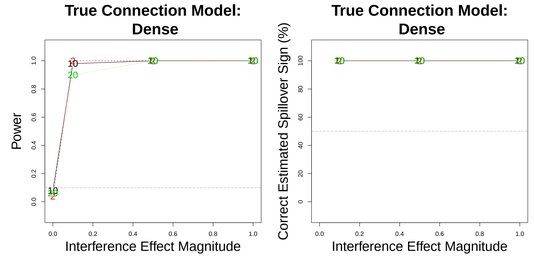


Figure 1: Statistical power increases with the interference magnitude & decreases with the number of covariates.

Real Data Validation: Can We Recover Known Networks? (Paluck et al., 2016)

- *Treatment*: Student assignment to anti-bullying program in 26 New Jersey schools (schools randomized to program; students randomized in schools)
- *Outcome*: # of student deviance events (logged)
- *True network observed*: Students listed up to 10 friends

Table: School results. The outcome of interest is the logarithm of the number of defiance/insubordination events plus 1. The p-values for the spillover effects are calculated via permutation inerference using the observed friend network.

	Estimate	p-value
Inferred Network Effect	-0.02	0.001
Main Effect, Controlling for Network Effect	-0.036	0.43
Main Effect, No Controls	0.0039	0.49
Spillover Effect for Treated Units: Having >2 Treated Friends	-0.013	0.29
Spillover Effect for Control Units: Having >2 Treated Friends	-0.024	0.049

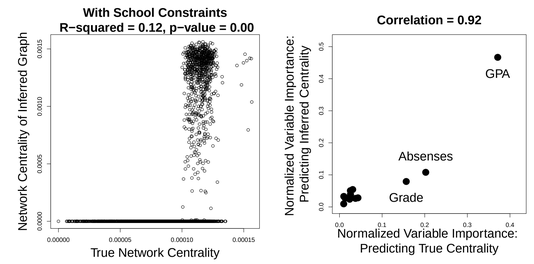


Figure 2: *Left panel*. Network centrality using observed or inferred network are strongly correlated (*centrality:* leading principle component of adjacency matrix). *Right panel*. Similar variable importance ($\rho > 0.90$) using covariates to predict network centrality of students in observed vs. true graphs.

How Often Do We Detect Interference? A Reanalysis of 12 Experiments

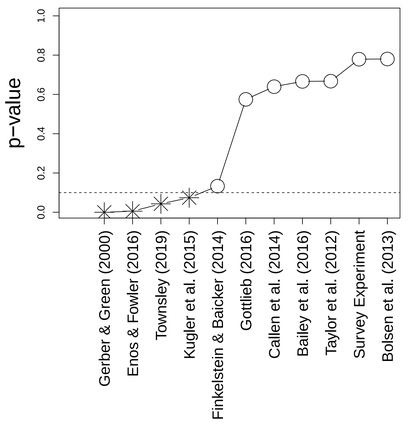


Figure 3: Using the proposed algorithm, we obtain evidence for treatment interference in 4 of 12 social science experiments at the 0.10 level.

Case Study on Turnout (Townsley, 2019)

- *Treatment*: Canvasing/campaign mail from the Liberal Party (UK) in a Suffolk community. *Outcome*: Turnout
- A positive main effect; our analysis suggests more treated inferred connections → lower turnout

Table: Analysis of data from Townsley (2019).

	Estimate	Std. Error
Treatment Effect	0.083	0.068
Inferred Spillover Effect	-0.043	0.019

References

[1] P. M. Aronow. A general method for detecting interference between units in randomized experiments. *Sociological Methods & Research*, 41(1):3–16, 2012.

[2] S. Baird, J. A. Bohren, C. McIntosh, and B. Özler. Optimal design of experiments in the presence of interference. *Review of Economics and Statistics*, 100(5):844–860, 2018.

[3] E. CandÀ, Y. Fan, L. Janson, J. Lv, et al. Panning for gold: Model-x knockoffs for high dimensional controlled variable selection. *Journal of the Royal Statistical Society Series B*, 80(3):551–577, 2018.

[4] D. Eaton and K. Murphy. Exact bayesian structure learning from uncertain interventions. In *Artificial intelligence and statistics*, pages 107–114, 2007.

[5] I. D. Muñoz and M. van der Laan. Population intervention causal effects based on stochastic interventions. *Biometrics*, 68:541–549, June 2012.

[6] E. L. Paluck, H. Shepherd, and P. M. Aronow. Changing climates of conflict: A social network experiment in 56 schools. *Proceedings of the National Academy of Sciences*, 113(3):566–571, 2016.

[7] G. Papadogeorgou, K. Imai, J. Lyall, and F. Li. Causal inference with spatio-temporal data: Estimating the effects of airstrikes on insurgent violence in iraq. *arXiv preprint arXiv:2003.13555*, 2020.

[8] P. R. Rosenbaum. Interference between units in randomized experiments. *Journal of the American Statistical Association*, 102(477):191–200, 2007.

[9] M. Saveski, J. Pouget-Abadie, G. Saint-Jacques, W. Duan, S. Ghosh, Y. Xu, and E. M. Airoldi. Detecting network effects: Randomizing over randomized experiments. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, pages 1027–1035, New York, NY, USA, 2017. ACM.

[10] F. Sävje, P. M. Aronow, and M. G. Hudgens. Average treatment effects in the presence of unknown interference. *arXiv preprint arXiv:1711.06399*, 2017.

[11] Y. Yang, R. Luo, M. Li, M. Zhou, W. Zhang, and J. Wang. Mean field multi-agent reinforcement learning. *arXiv preprint arXiv:1802.05438*, 2018.

[12] J. Zhang. Consistency of mle, lse and m-estimation under mild conditions. *Statistical Papers*, 61(1):189–199, 2020.