Randomization tests for spillovers under interference: A graph-theoretic approach

Panagiotis (Panos) Toulis
panos.toulis@chicagobooth.edu

Econometrics and Statistics
University of Chicago, Booth School of Business

Interference

Interference exists when the outcomes of some unit depend on the treatment of others.

—(Hong and Raudenbush, 2006); (Hudgens and Halloran, 2008); (Aronow, 2012); (Bowers, 2013); (Toulis and Kao, 2013); (Ogburn and VanderWeele, 2014); (Eckles et. al., 2016); (Aronow and Samii, 2017); (Ogburn et. al., 2017); (Savje et al, 2017); (Athey et. al, 2018), (Basse and Feller, 2018); (Basse et. al., 2019); (Jagadeesan et. al., 2020) (Forastiere et. al., 2020);

Includes spillovers, peer effects, contagion, equilibrium effects, etc.

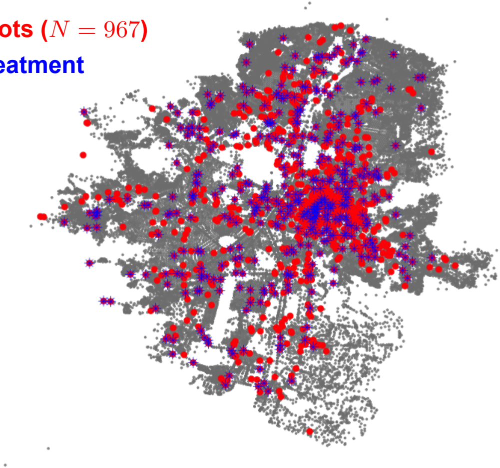
Pervasive in most social studies. Can be either a nuisance to be addressed by design, or the quantity of interest.

Motivation for this work: Crime spillovers across streets from policing experiment in Medellin, Colombia.

Medellin application

crime hotspots ($N = 967$)

observed treatment



unit = street; *treatment* = intensified policing; *outcome* = crime index.

We will test whether there are spillovers on control streets from nearby treated streets.

Current approaches

Several model-based approaches exist. Typically include regressions of unit outcomes on group/peer treatments and outcomes.
—(Durlauf and Young, 2001); (Brock and Durlauf, 2001); (Jackson, 2010); (Graham, 2008)

A model-based approach has identification and interpretation issues.
—(Deaton, 1990); (Manski, 1993); (Boozer and Cacciola, 2001); (Moffit, 2001); (Angrist, 2014)

Design-based approaches have emerged as a robust alternative. They mostly aim to generalize the classical Fisher randomization test.
—(Aronow, 2012); (Athey et. al., 2018); (Basse et. al., 2019)

Outline

1. Setup and notation
2. Classical Fisher randomization test (FRT)
3. FRTs under Interference
4. The null exposure graph
5. Application in Medellin
6. (if time) Backup slides: computation, test power

Setup and notation

There is a set $\mathbb{U} = \{1, \ldots, N\}$ of N units indexed by i.

<u>Denote:</u>

$Z = (Z_1, \ldots, Z_N) \in \{0,1\}^N =: \mathbb{Z}$	binary treatment
$Y(z) = (Y_1(z), \ldots, Y_N(z)) \in \mathbb{R}^N$	potential outcomes under $z \in \mathbb{Z}$
$Z^{\text{obs}} \in \mathbb{Z},\ Y^{\text{obs}} \in \mathbb{R}^N$	observed quantities
Z^*, Y^*	randomization draws
$P(Z) \in [0,1]$	design, assumed known

As usual, potential outcomes are assumed to be fixed, and randomness comes only from $P(Z)$.

Fisher randomization test (FRT, 1935)

We start with the simplest "global null" hypothesis of no effect:

$$H_0 : Y_i(z) = Y_i(z'), \text{ for all } z, z' \in \mathbb{Z}.$$

Choose test statistic $T = t(y, z)$ —(e.g., difference in means).

1. $T^{\text{obs}} = t(Y^{\text{obs}}, Z^{\text{obs}})$.
2. Sample $Z^* \sim P(Z^*)$, store $T_R = t(Y^{\text{obs}}, Z^*)$.
3. p-value $= \mathbb{E}\left[\mathbb{1}\left\{T_R \geq T^{\text{obs}}\right\}\right]$.

Proof of validity:

$$t(Y^{\text{obs}}, Z^*) \overset{H_0}{=} t(Y^*, Z^*) \overset{d}{=} t(Y^{\text{obs}}, Z^{\text{obs}})$$

—"$T_R \sim T^{\text{obs}}$ (under null)"

More about the FRT

Advantages of FRT include:

- Simple and exact. The test is valid in finite samples.
- Minimal assumptions. No model for Y.
- Robust. Same answer under some transformations of Ys.

Main critique of FRT:

- Can only test “strong” nulls.
- Only inference in-sample.

Another argument for FRTs (under no interference)

- Suppose completely randomized design (half-treated/half-control):

Unit (i)	Assignment (Z_i)	Outcome (Y_i)
1	1	8
2	0	$3+\epsilon$
3	0	$3-\epsilon$
4	1	8

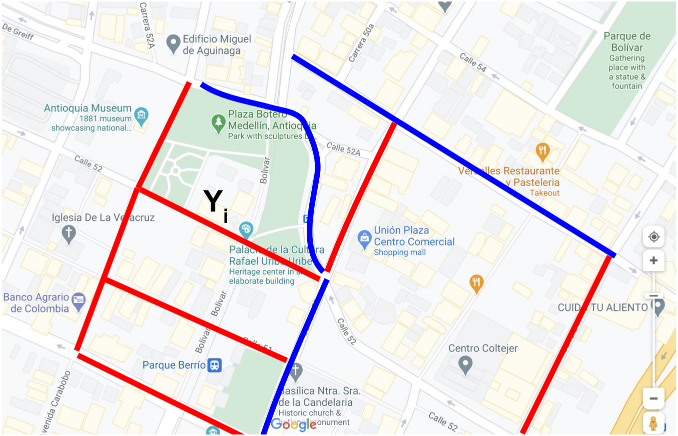
- **Model-based approach:** Regress $Y_i \sim Z_i$. The estimate of "causal effect" is +5, with standard error $O(\epsilon)$.

  $\hookrightarrow$ Standard error estimation is **conflated** with model fit.

- **Design-based approach:** No significance. We could have observed -5 with same probability, even when there is no effect whatsoever.

Crime spillovers in Medellin

The global null is not useful for **crime spillovers**.



e.g., what is the effect of a nearby **treated street** on a **control street**?

A crime spillover hypothesis

To write the null hypothesis compactly, use a *treatment exposures*:

$$f_i(z) = \begin{cases} \texttt{short}, & z_i = 0, \text{dist}_i < 125\text{m} \\ \texttt{control}, & z_i = 0, \text{dist}_i > 500\text{m} \\ \texttt{neither}, & \text{otherwise.} \end{cases}$$

where $\text{dist}_i = \min_{j \neq i : z_j = 1} d(j, i)$ = distance to closest treated street.

$$H_0 : \; Y_i(z) = Y_i(z') \text{ for every } i, z, z',$$

$$\text{such that } f_i(z), f_i(z') \in \{\texttt{short}, \texttt{control}\}.$$

A crime spillover hypothesis

To write the null hypothesis compactly, use a *treatment exposures*:

$$f_i(z) = \begin{cases} \texttt{short}, & z_i = 0, \text{dist}_i < 125\text{m} \\ \texttt{control}, & z_i = 0, \text{dist}_i > 500\text{m} \\ \texttt{neither}, & \text{otherwise.} \end{cases}$$

where $\text{dist}_i = \min_{j \neq i : z_j = 1} d(j, i)$ = distance to closest treated street.

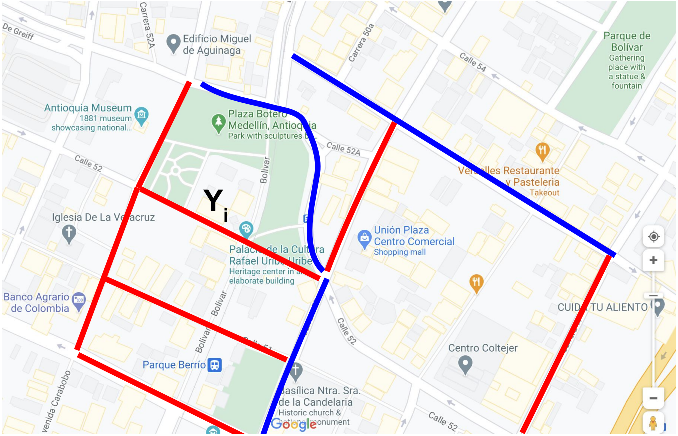
$$H_0 : \; Y_i(z) = Y_i(z') \text{ for every } i, z, z',$$

$$\text{such that } f_i(z), f_i(z') \in \{\texttt{short}, \texttt{control}\}.$$

$\star$ Can we use the standard FRT?

FRT problems under interference

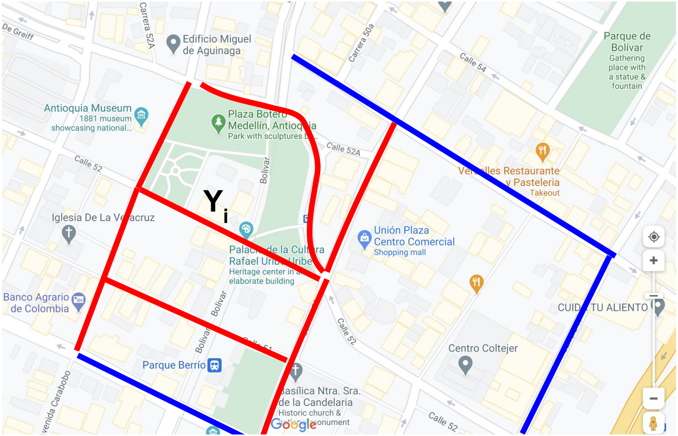
Consider unit i and the depicted Z^{obs} below:



Unit i is exposed to `short` range spillovers.
Observed outcome is $Y_i^{\text{obs}} = Y_i(\texttt{short})$— say $Y_i^{\text{obs}} = 2.5$ (“crime score”).

FRT problems under interference

Consider a counterfactual treatment assignment Z^*:

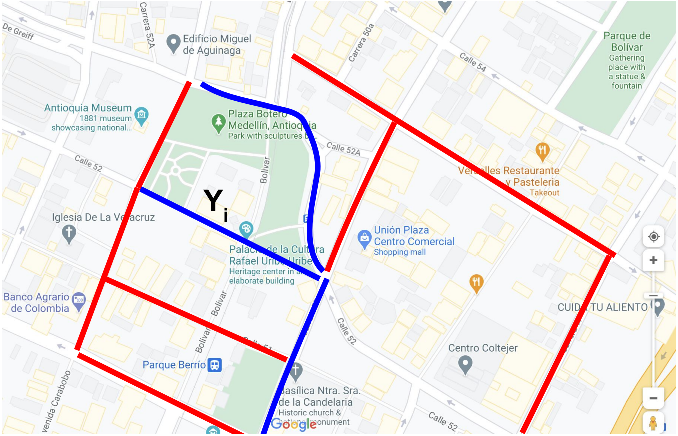


Unit i is exposed to $\texttt{control}$ under Z^*.

$\star$ Outcome $Y_i(\texttt{control})$ was not actually observed but <u>can be imputed</u> from Y_i^{obs} under the null—since $Y_i(\texttt{control}) = Y_i(\texttt{short})$ under H_0.

FRT problems under interference

Consider another Z^* in randomization:



Unit i is exposed to $\texttt{neither}$ under Z^*.
Outcome $Y_i(\texttt{neither})$ **cannot be imputed** from Y_i^{obs}.

Generalization: Conditional FRT

So, we have to **condition** on subsets of units and assignments. —Aronow (2012); Athey et. al. (2018); Basse et al (2019)

We denote this conditioning as $C = (U, \mathcal{Z})$, where $U \subset \mathbb{U}, \mathcal{Z} \subset \mathbb{Z}$.

It is generally probabilistic and can be described through a **conditioning mechanism**, $P(C|Z^{\text{obs}})$.

Generalization: Conditional FRT

So, we have to **condition** on subsets of units and assignments. —Aronow (2012); Athey et. al. (2018); Basse et al (2019)

We denote this conditioning as $C = (U, \mathcal{Z})$, where $U \subset \mathbb{U}, \mathcal{Z} \subset \mathbb{Z}$.

It is generally probabilistic and can be described through a **conditioning mechanism**, $P(C|Z^{\text{obs}})$.

We can execute the FRT conditionally on C as long as:

1. The potential outcomes **can be imputed** for any $i \in U$ and any $z \in \mathcal{Z}$ using H_0.
2. The resampling distribution is:

$$P(Z^*|C) \propto \underbrace{P(C|Z^*)}_{\text{conditioning mechanism}} \times \underbrace{P(Z^*)}_{\text{design}}$$

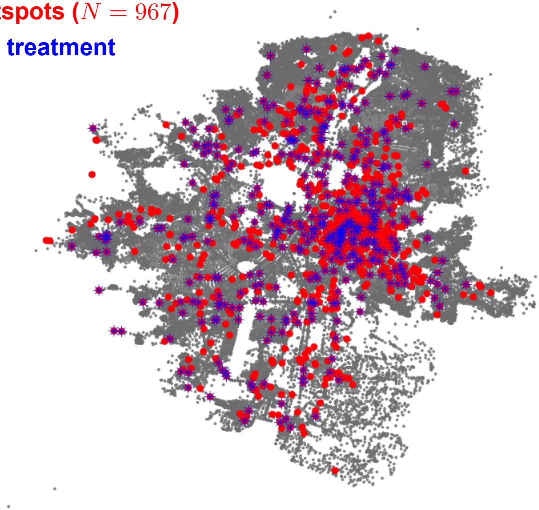
Requirement #2 is trivial (under our control).

The key challenge is #1. How to ensure this constraint?

Medellin application: What's a good conditioning mechanism?

crime hotspots ($N = 967$)

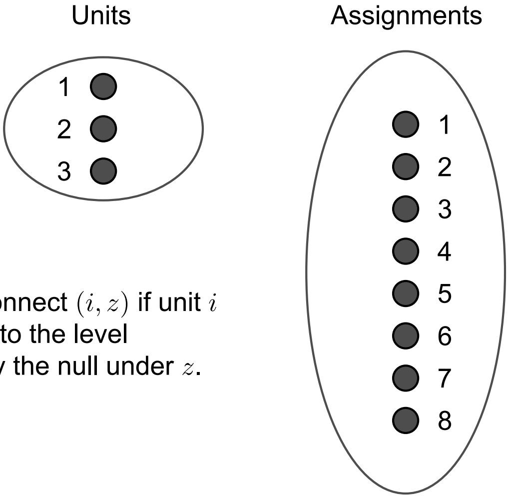
observed treatment



$\star$ What should $P(C|Z)$ be? —unclear, interference structure is complex.

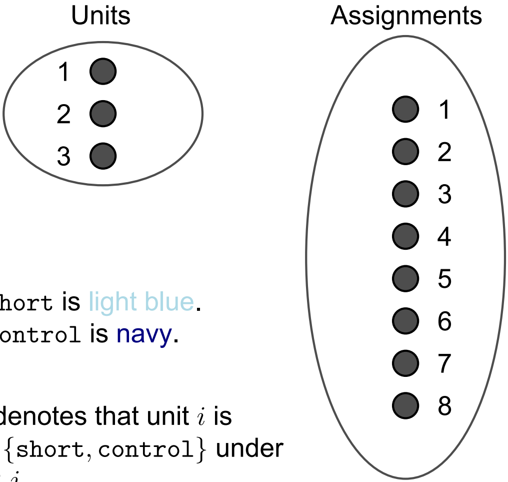
The null exposure graph

We set all the units one side and all the assignments on the other.



Then we connect (i, z) if unit i is exposed to the level specified by the null under z.

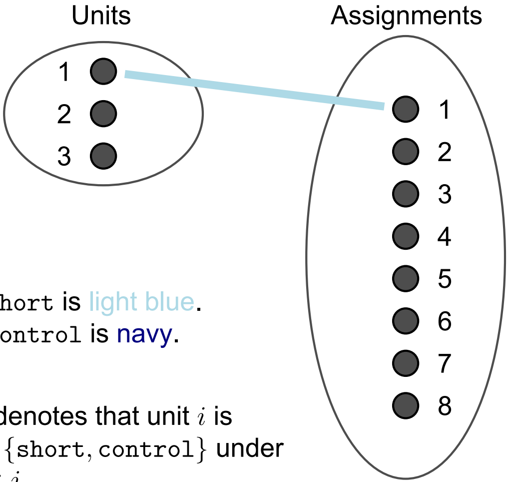
The null exposure graph



Exposure $\texttt{short}$ is light blue.
Exposure $\texttt{control}$ is navy.

edge (i, j) denotes that unit i is exposed to $\{\texttt{short}, \texttt{control}\}$ under assignment j.

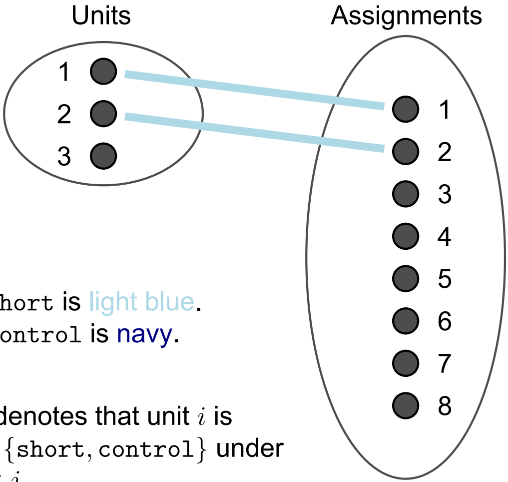
The null exposure graph



Exposure `short` is light blue.
Exposure `control` is navy.

edge (i, j) denotes that unit i is exposed to $\{\texttt{short}, \texttt{control}\}$ under assignment j.

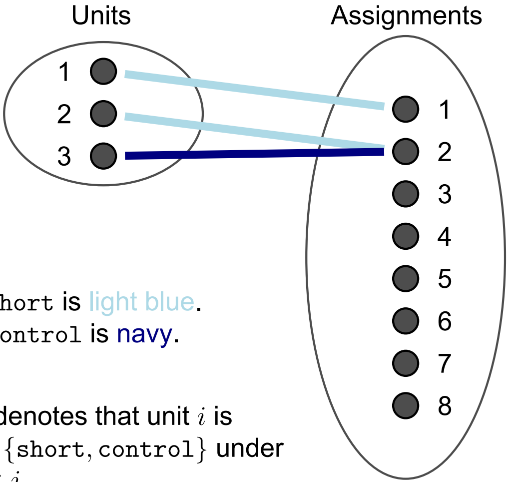
The null exposure graph



Exposure `short` is light blue.
Exposure `control` is navy.

edge (i, j) denotes that unit i is exposed to $\{\texttt{short}, \texttt{control}\}$ under assignment j.

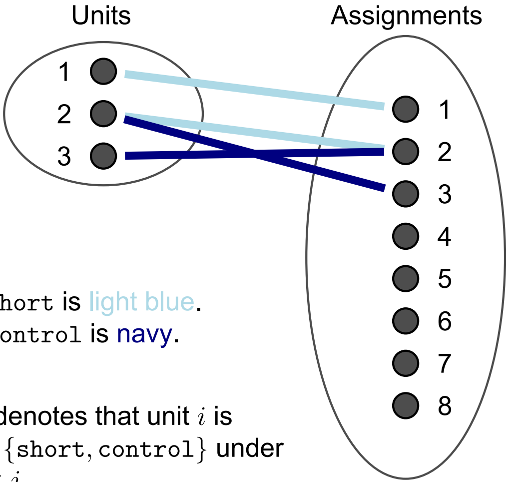
The null exposure graph



Exposure `short` is light blue.
Exposure `control` is navy.

edge (i, j) denotes that unit i is exposed to $\{\texttt{short}, \texttt{control}\}$ under assignment j.

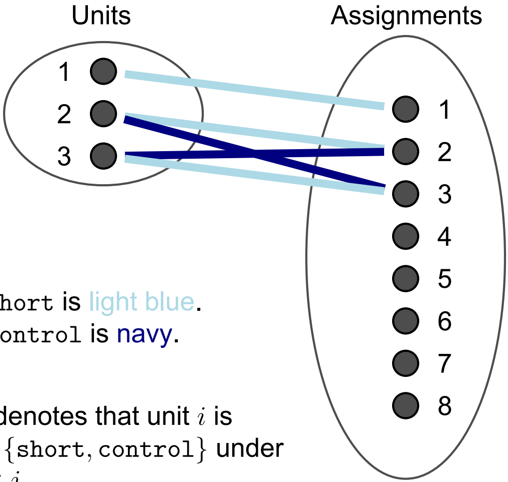
The null exposure graph



Exposure `short` is light blue.
Exposure `control` is navy.

edge (i, j) denotes that unit i is exposed to $\{\texttt{short}, \texttt{control}\}$ under assignment j.

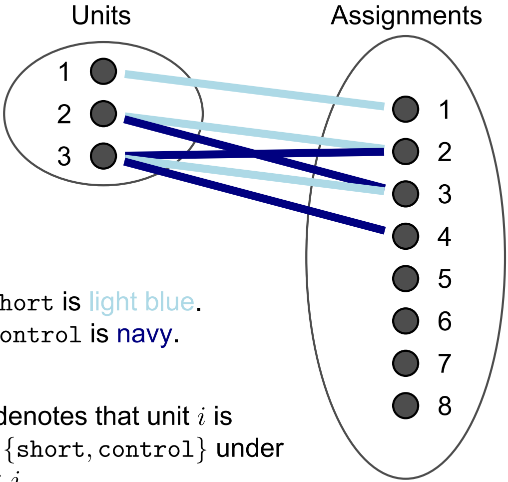
The null exposure graph



Exposure `short` is light blue.
Exposure `control` is navy.

edge (i, j) denotes that unit i is exposed to $\{\texttt{short}, \texttt{control}\}$ under assignment j.

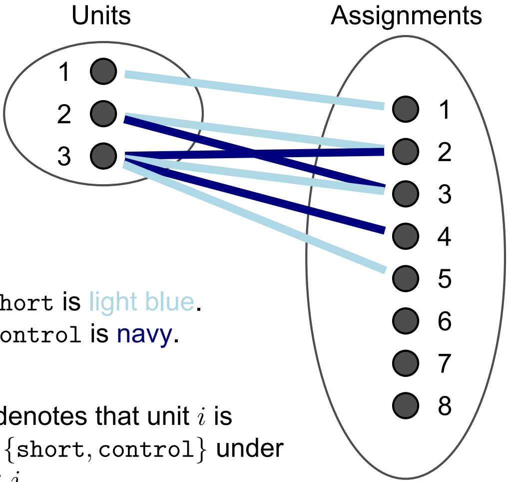
The null exposure graph



Exposure `short` is light blue.
Exposure `control` is navy.

edge (i, j) denotes that unit i is exposed to $\{\texttt{short}, \texttt{control}\}$ under assignment j.

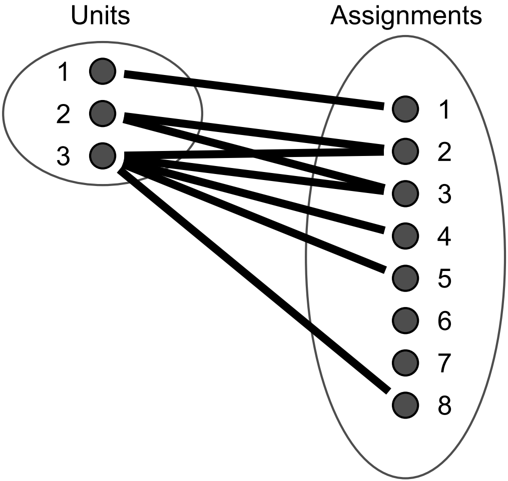
The null exposure graph

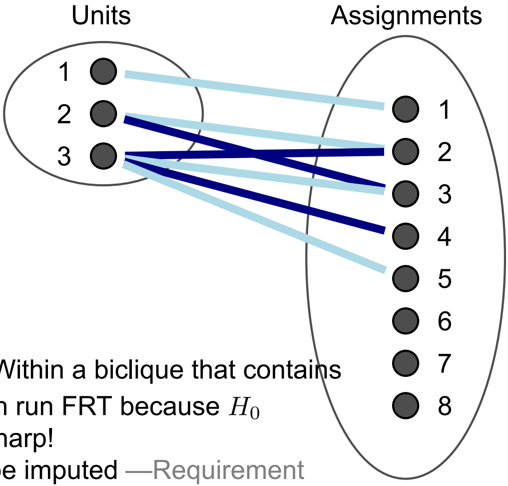


Exposure `short` is light blue.
Exposure `control` is navy.

edge (i, j) denotes that unit i is exposed to $\{\texttt{short}, \texttt{control}\}$ under assignment j.

The null exposure graph

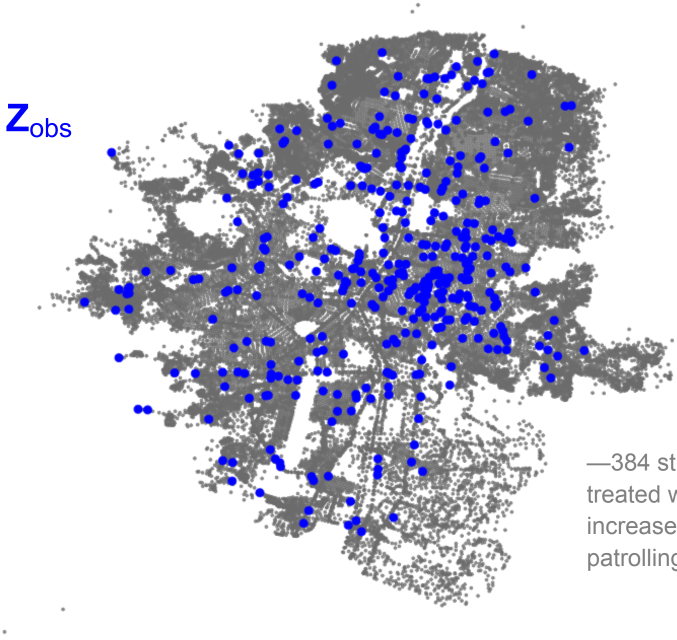




Key idea: Within a biclique that contains Z^{obs} we can run FRT because H_0 becomes sharp!
(all Y can be imputed —Requirement #2).

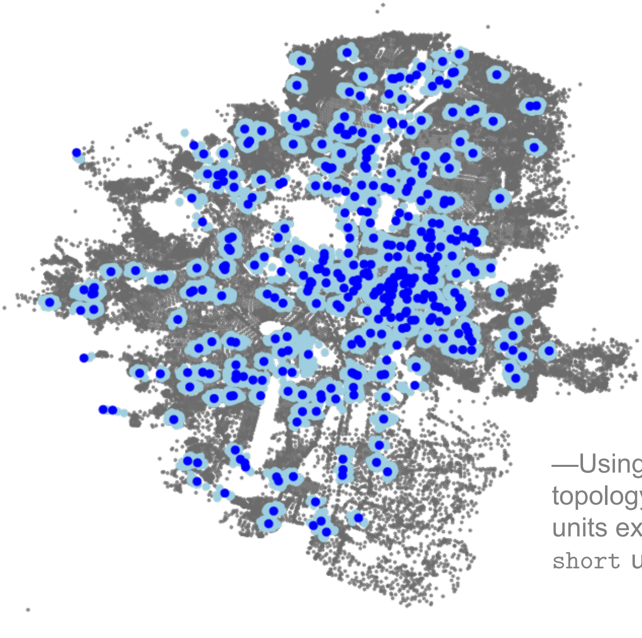
Returning to the map

The observed assignment



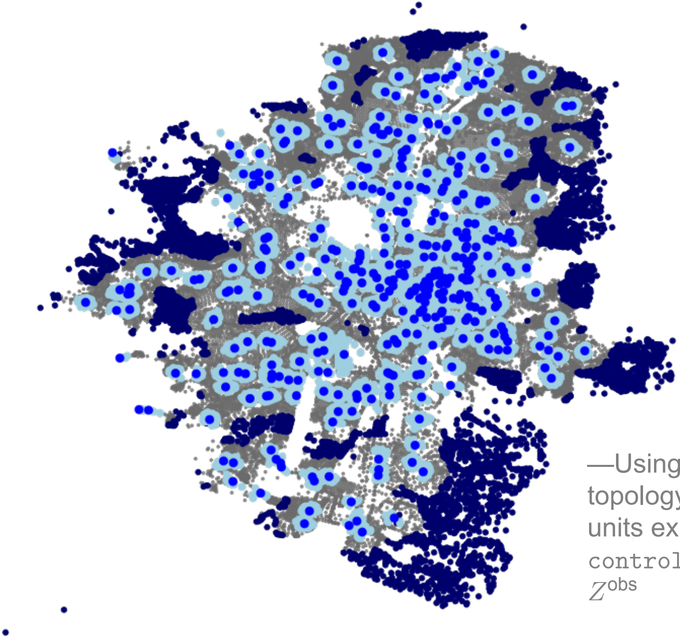
—384 streets are treated with increased police patrolling

Short-range spillover units (`short`)



—Using network topology, color units exposed to `short` under Z^{obs}

Pure control units (`control`)

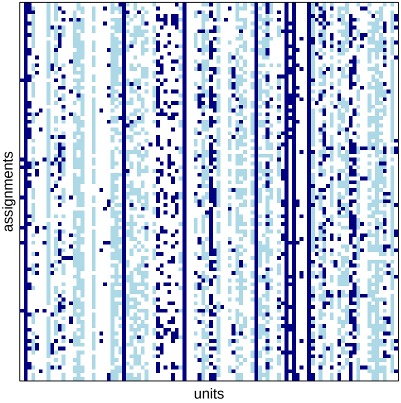


—Using network topology, color units exposed to `control` under Z^{obs}

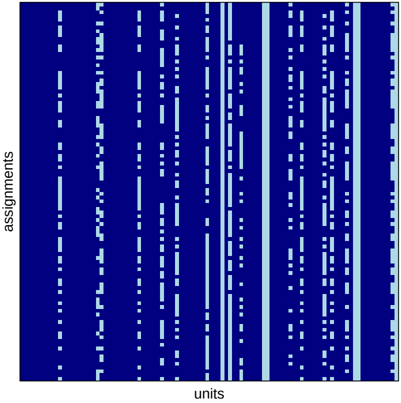
We can remake these pictures for every assignment Z drawn from design $P(Z)$ ...

—The output is our null exposure graph!

Null exposure graph and clique



null exposure graph



clique (zoomed-in)

Null-exposure graphs: summary

- A null-exposure graph, G, is thus uniquely defined given H_0 and treatment exposures.
- H_0 is sharp in a biclique of G. So, we can run a conditional FRT within a biclique. But which biclique to condition on?
- Our full procedure first produces a **biclique decomposition** of this graph. Then, conditions on the biclique that contains Z^{obs}.

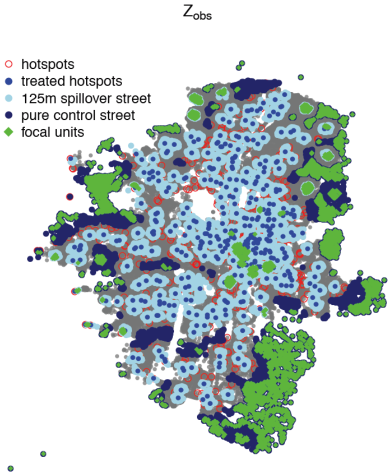
Spatial interference: Medellin data

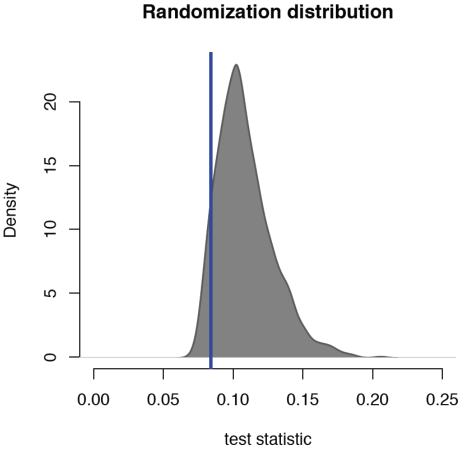
Statistics of the null-exposure graph:

- #units = 37,055.
- #assignments = 10,000 (design is uniform over this fixed set).
- #edges = 163,836,445.
- density (#edges / #total possible edges) = 44.2%

Statistics of the clique we condition on:

- #units in clique = 3,981.
- #assignments in clique $\approx$ 1,000.





Focal units (in green) are in downtown and outskirts.
Biclique test **automatically** discovered this pattern.

Varying radius of short-range effect

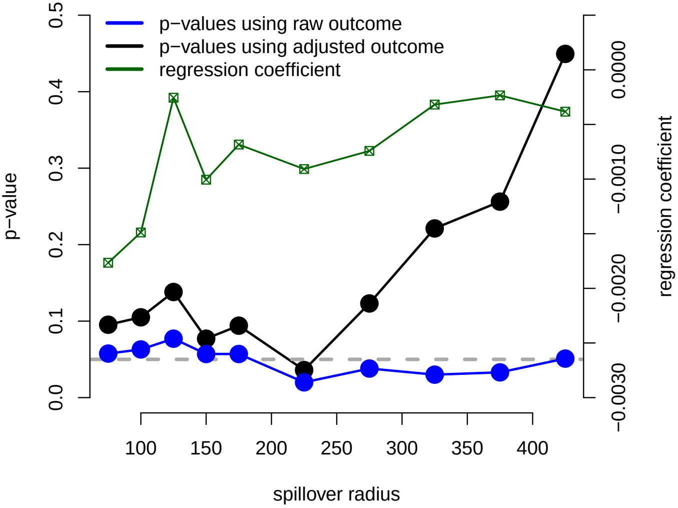


Figure: P-values for clique tests with varying spillover radius.

Experimental design

Suppose a design space $(p_0, p_1) \in [0,1]^2$ where p_0=prob. of treatment in city-center, and p_1 prob. of treatment in outskirts.

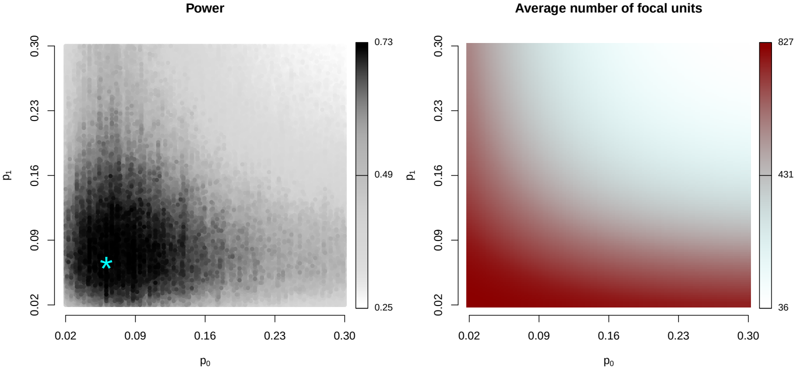


Figure: *Left:* The power of the test for different combinations of p_0, p_1 calculated via simulation. Darker colors denote larger power values, while lighter colors denote smaller power values. *Right:* Null-exposure graph density for different combinations of p_0, p_1.

Concluding thoughts

- Structure is placed on null hypotheses under interference through exposure functions.

- We represent the problem through the null exposure graph, and we condition on bicliques of this graph.

- Translates the testing problem into graphical operations on the null exposure graph.

- Can study power through properties of the null exposure graph (e.g., density).

Thank you!

(Puelz et. al., 2020) “A Graph-Theoretic Approach to Randomization Tests of Causal Effects Under General Interference”; *arXiv:1910.10862* (under revision).

Power

The size of the clique is crucial for the test power.

Theorem (high level)

For $C = (U, \mathcal{Z})$ *let* $|C| = (n, m)$ *imply that* $|U| = n$ *and* $|\mathcal{Z}| = m$. *Suppose:*

(A1) n *is scale parameter* ($1/\sqrt{n}$) *for null distribution of test statistic;*

(A2) *spillover effect* τ *is additive;*

(A3) *the* m *test statistic values are i.i.d. from the null;*

(A4) *the null distribution cdf can be* ϵ*-approximated by a sigmoid.*

Then,

$$\mathrm{E}\big(\text{reject} \mid H_1, |C| = (n, m)\big) \geq \frac{1}{1 + Ae^{-a\tau\sqrt{n}}} - O(m^{-r}) - \epsilon,$$

where $a, A > 0, r \in (1/2, 1]$.

Interpretation:

- Number of focal units controls "sensitivity" of the test.
- Number of focal assignments controls maximum power.

A naive test (which doesn't work)

Not all approaches lead to a valid test. For example:

1. Given Z^{obs} calculate maximum clique in null-exposure graph, G_f, that contains Z^{obs}, say,

$$C = \texttt{mc}(Z^{\text{obs}}; G_f); \quad (\texttt{mc} = \texttt{"max clique"}).$$

2. Condition the randomization test on C^*, i.e., resample assignments according to

$$r(Z^*) = \frac{\mathbb{1}\{Z^* \in C\} P(Z^*)}{P(C)}.$$

A naive test (which doesn't work)

Not all approaches lead to a valid test. For example:

1. Given Z^{obs} calculate maximum clique in null-exposure graph, G_f, that contains Z^{obs}, say,

$$C = \texttt{mc}(Z^{\text{obs}}; G_f); \quad (\texttt{mc} = \texttt{"max clique"}).$$

2. Condition the randomization test on C^*, i.e., resample assignments according to

$$r(Z^*) = \frac{\mathbb{1}\{Z^* \in C\} P(Z^*)}{P(C)}.$$

Proof of invalidity:

The correct conditional distribution is:

$$P(Z^*|C) = \frac{P(C|Z^*)P(Z^*)}{P(C)} = \frac{\mathbb{1}\{\texttt{mc}(Z^*; G_f) = C\} P(Z^*)}{P(C)} \neq r(Z^*).$$

Main method: Clique-based randomization test

1. **Decompose:** Compute *biclique decomposition* $\mathcal{C}$ of G_f.
2. **Condition:** Pick out clique containing Z^{obs}, call it C.
3. **Summarize:** Compute $T^{\text{obs}} = t(Y^{\text{obs}}, Z^{\text{obs}}; C)$, then

$$\text{p-value} = \mathbb{E}\left[\mathbb{1}\left\{t(Y^{\text{obs}}, Z^*; C) \geq T^{\text{obs}}\right\} \mid C\right]$$

Here, we resample with respect to

$$r(Z^*) \propto \underbrace{\mathbb{1}\{Z^* \in C\}}_{\text{cond. mechanism}} \cdot \underbrace{P(Z^*)}_{\text{design}}$$

Main method: Clique-based randomization test

1. **Decompose:** Compute *biclique decomposition* $\mathcal{C}$ of G_f.
2. **Condition:** Pick out clique containing Z^{obs}, call it C.
3. **Summarize:** Compute $T^{\text{obs}} = t(Y^{\text{obs}}, Z^{\text{obs}}; C)$, then

$$\text{p-value} = \mathbb{E}\left[\mathbb{1}\left\{t(Y^{\text{obs}}, Z^*; C) \geq T^{\text{obs}}\right\} \mid C\right]$$

Here, we resample with respect to

$$r(Z^*) \propto \underbrace{\mathbb{1}\{Z^* \in C\}}_{\text{cond. mechanism}} \cdot \underbrace{P(Z^*)}_{\text{design}}$$

Proof of validity:

The correct conditional distribution is:

$$P(Z^*|C) = \frac{P(C|Z^*)P(Z^*)}{P(C)} = \frac{\mathbb{1}\{C \in \mathcal{C}\}\,\mathbb{1}\{Z^* \in C\}\,P(Z^*)}{P(C)} = r(Z^*).$$

—first eq. from Bayes; second from definition of conditioning mechanism.

- Finding cliques is NP-hard—Peeters, 2003; Zhang et al, 2014).

- We use the "Binary Inclusion-Maximal Biclustering Algorithm", which uses a "divide and conquer" method to find cliques (`Bimax`, Prelic et. al, 2006).
  —works fine for hundred nodes/thousands edges.

- Our method is **constructive**, still can be optimized.
  —i.e., different biclique decompositions will have different power properties, but all are **valid**.