Randomization Tests of Causal Effects Under General Interference

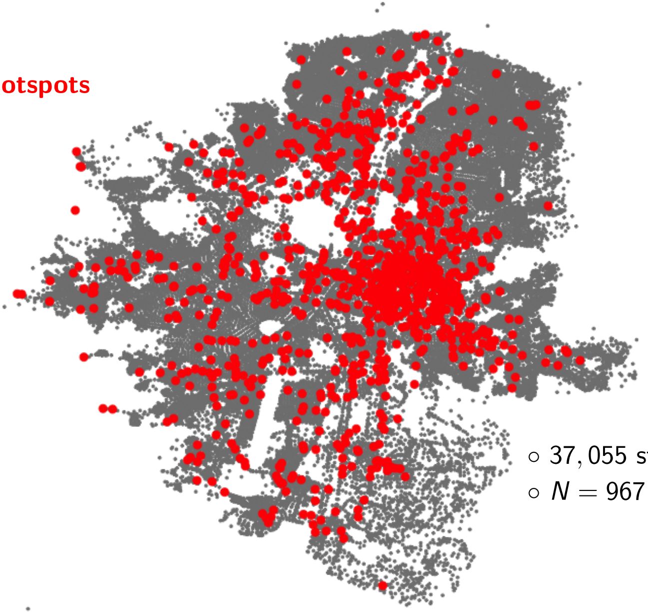
Panos Toulis
University of Chicago, Booth School

joint with: G Basse, A Feller, and D Puelz

Medellín, Colombia

Medellín, Colombia

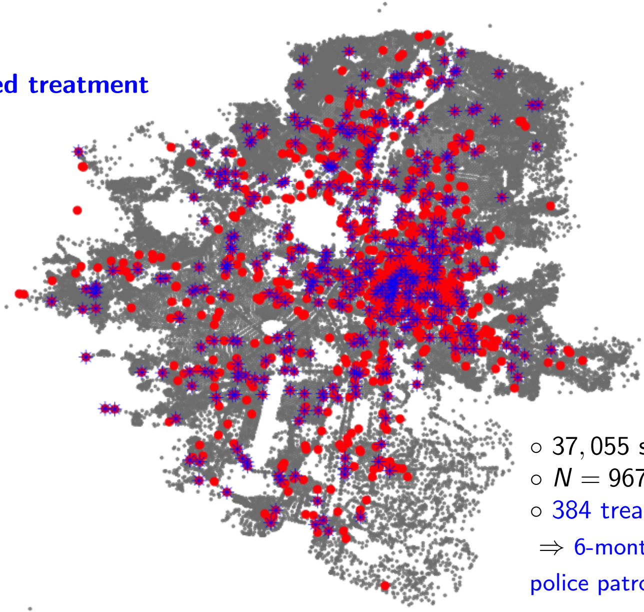
Medellín, Colombia



- 37,055 street segments
- $N = 967$ "hotspots".

Medellín, Colombia

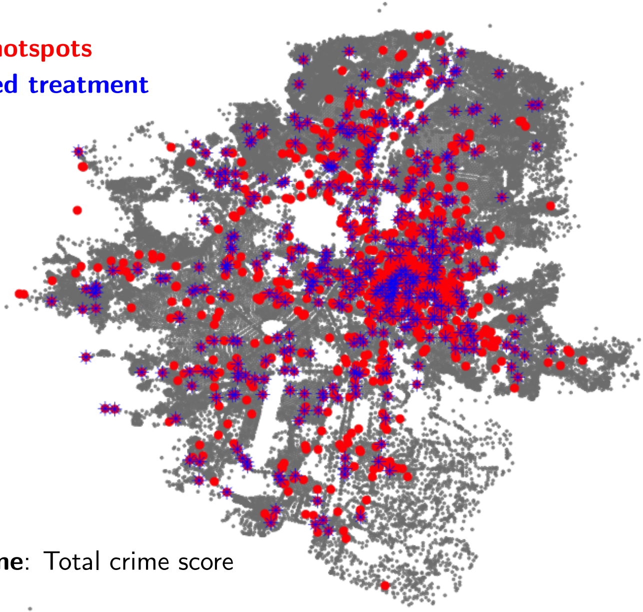
observed treatment



- 37,055 street segments
- $N = 967$ "hotspots".
- 384 treated.

$\Rightarrow$ 6-month increase in police patrolling time

Medellín, Colombia

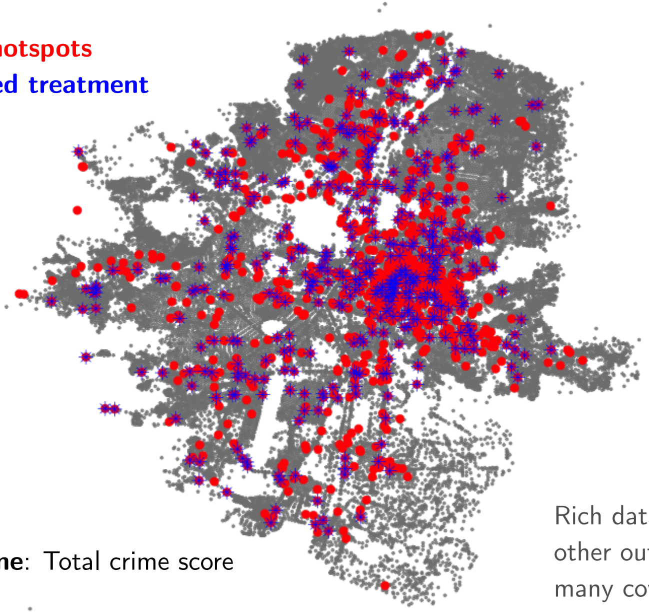


Outcome: Total crime score

Medellín, Colombia

crime hotspots

observed treatment



Outcome: Total crime score

Rich data set with other outcomes, many covariates

Questions

How does the intervention affect crime?

$\rightarrow$ direct effect?

$\rightarrow$ spillovers to adjacent streets?

Questions

How does the intervention affect crime?

$\rightarrow$ direct effect?

$\rightarrow$ spillovers to adjacent streets?

We will focus on spillovers through hypothesis testing.

We prefer a model-free approach, so we will use the randomization method of inference.

Notation and data

N units (streets) indexed by $i = 1, 2, \ldots, N$.

Define observed data:

$Z = (Z_1, \ldots, Z_N)$ as binary treatment; $P(Z) =$ design;

$Y = (Y_1, \ldots, Y_N)$ as vector of observed outcomes.

$\hookrightarrow$ *will also use Z^{obs}, Y^{obs} for emphasis.*

Hotspots received increased policing, while non-hotspots lost about $\sim$5min of patrol time.

Outcome is a weighted average of crime indicators:

$\hookrightarrow$*0.550 for homicides, 0.112 for assaults, 0.221 for car and motorbike theft, and 0.116 for personal robbery (Collazos et al., 2019).*

Classical approach: no interference

Assume no interference: Outcome of unit depends only on its own treatment assignment.

$\hookrightarrow$ *Only two potential outcomes, $Y_i(0)$, $Y_i(1)$, for every i.*

Unrealistic in this application. But helps build the intuition.

Classical approach: no interference

Assume no interference: Outcome of unit depends only on its own treatment assignment.

↪ *Only two potential outcomes,* $Y_i(0), Y_i(1)$, *for every* i.

Unrealistic in this application. But helps build the intuition.

Classical question of randomization inference: Does treatment have an effect *at all*?

$$\mathbf{H_0}: \quad Y_i(0) = Y_i(1) \text{ for every } i.$$

Key implication of H_0 is that Y is fixed across all possible randomizations.

Fisher randomization test (1935)

$\mathbf{H_0}: \quad Y_i(0) = Y_i(1), \text{ for every } i.$

The procedure:

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

1. $T^{\text{obs}} = t(Y, Z)$.
2. Sample $Z' \sim P(Z')$, store $T_r = t(Y', Z') \overset{H_0}{=} t(Y, Z')$.
3. p-value $= \mathbb{E}\left[\mathbb{1}\{T_r \geq T^{\text{obs}}\}\right]$.

Fisher randomization test (1935)

$\mathbf{H_0}: \quad Y_i(0) = Y_i(1), \text{ for every } i.$

The procedure:

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

1. $T^{\text{obs}} = t(Y, Z)$.
2. Sample $Z' \sim P(Z')$, store $T_r = t(Y', Z') \overset{H_0}{=} t(Y, Z')$.
3. p-value $= \mathbb{E}\left[\mathbb{1}\{T_r \geq T^{\text{obs}}\}\right]$.

Proof of validity:

$$t(Y', Z') \overset{H_0}{=} t(Y, Z') \overset{d}{=} t(Y, Z)$$

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

Advantages of Fisherian randomization

- Exact. The test is valid in finite samples.
- Minimal assumptions. No model for Y.
- Robust. Test gives the same (or very similar) answers with different Y-scales (the same cannot be said for regression).

Goal is to use Fisherian randomization under interference.

“No interference” assumption is too strong ...

Assumption of no interference is not realistic in our application.
Spillovers are actually a quantity of interest.

First problem is notational:
In general, $Y_i(z)$ is the potential outcome of i under pop assignment z.

$\hookrightarrow$ *More outcomes than just* $Y_i(0), Y_i(1)$.

"No interference" assumption is too strong ...

Assumption of no interference is not realistic in our application.
Spillovers are actually a quantity of interest.

First problem is notational:
In general, $Y_i(z)$ is the potential outcome of i under pop assignment z.

$\hookrightarrow$ *More outcomes than just* $Y_i(0), Y_i(1)$.

In principle there could be 2^N potential outcomes. Impractical.

To make progress we can use the concept of **exposure functions**.

Treatment exposures

For any given Z, unit i is exposed to "something more" than Z_i.
We assume unit i's exposure is defined by a function:

$$f_i : \{0,1\}^N \rightarrow \mathcal{E}.$$

$\hookrightarrow$*e.g.,* $f_i(z) = (z_i, \sum_{j \neq i} g_{ij} z_j)$, *where* g_{ij} *indicates whether* i *and* j *can influence each other (classmates or neighbors).*

$\mathcal{E}$ is the set of possible exposures (short-range spillover, medium-range spillover, pure control, etc.)

Definition of $\mathcal{E}, \{f_i\}$ depends on the substantive scientific question.

What's in a treatment exposure?

Intuitively, we think that $f_i(z)$ defines "equivalence classes" across the population assignments; e.g.,

$$Y_i(z) = Y_i(z'), \text{ if } f_i(z) = f_i(z').$$

This assumption is not necessary for our method but helps with interpretation of our results.

Alternative: when f_i are unspecified we may consider marginal estimands, e.g., $E\{Y_i(z_i = 1, z_{-i})\} - E\{Y_i(z_i = 0, z_{-i})\}$;

$\hookrightarrow$*see Aronow and Samii (2019); Savje (2019). Interpretation is hard though.*

Question: Is there a short-range spillover effect?

We can use exposures $\{f_i\}$ to study spillovers/interference.

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

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

Here, we defined:

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

dist_i = distance to closest treated street.

Use classical Fisherian randomization? Not quite ...

Recall, $T_r \sim T^{\text{obs}}$ under H_0 for things to work. However,

$$T_r = t(Y^{'}, Z^{'}) \overset{\cancel{H_0}}{\neq} t(Y, Z^{'}) \overset{d}{=} t(Y, Z) = T^{\text{obs}}.$$

The null is relevant for only 2 out of the 3 possible exposures:

$\mathbf{H_0}$: $\quad Y_i(\texttt{short}) = Y_i(\texttt{control}) \overset{?}{=} Y_i(\texttt{neither})$, for every i.

In this case, the null is **not sharp**. We cannot impute missing potential outcomes $Y^{'}$ freely under any $Z^{'}$.

Testing $Y_i(\texttt{short}) = Y_i(\texttt{control}), \forall i$.

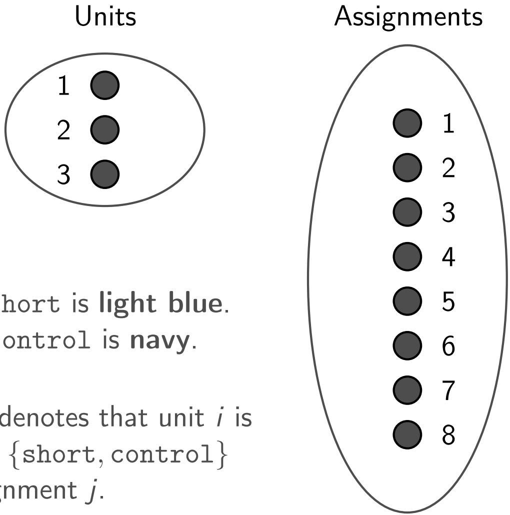
Idea: If we focus on units only exposed to `short` or `control` then we can impute their missing outcomes in the randomization test.

$\hookrightarrow$*i.e., conditional randomization test.*

Given a null hypothesis and assignment from $P(Z)$, we know which units are exposed to `short` or `control` using $f_i(\cdot)$.

This is a binary relationship!

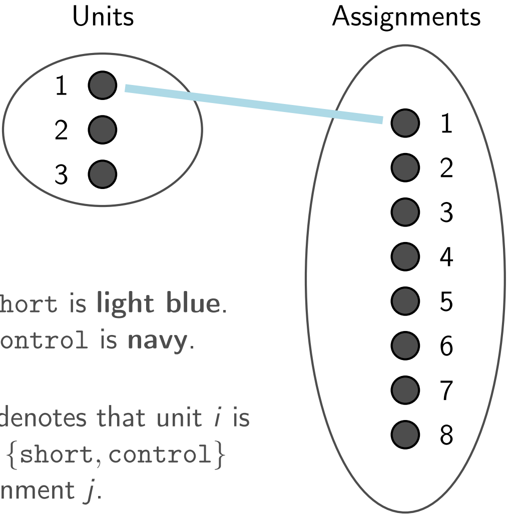
The null exposure graph



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.

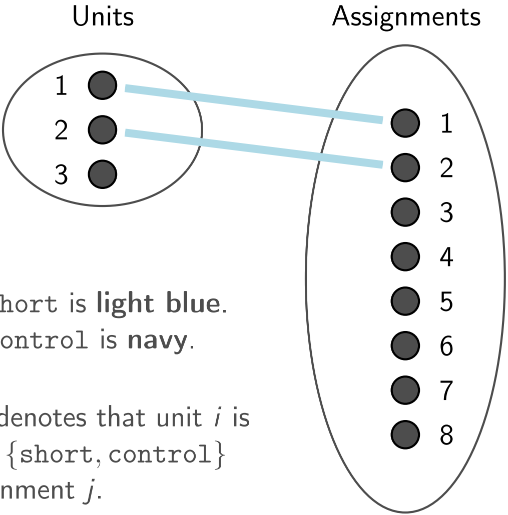
The null exposure graph



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.

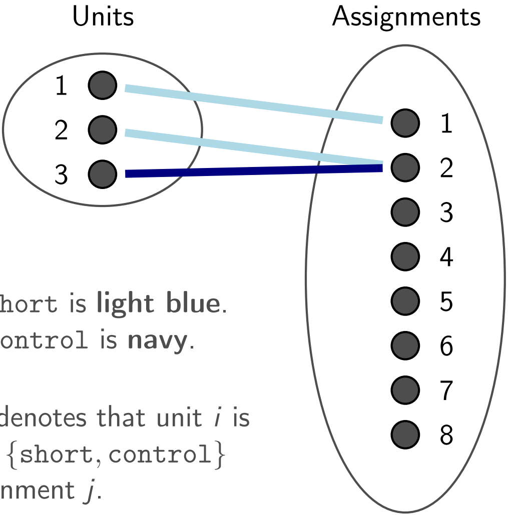
The null exposure graph



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.

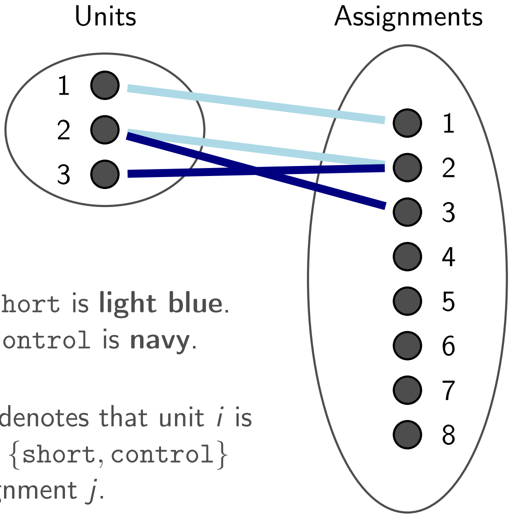
The null exposure graph



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.

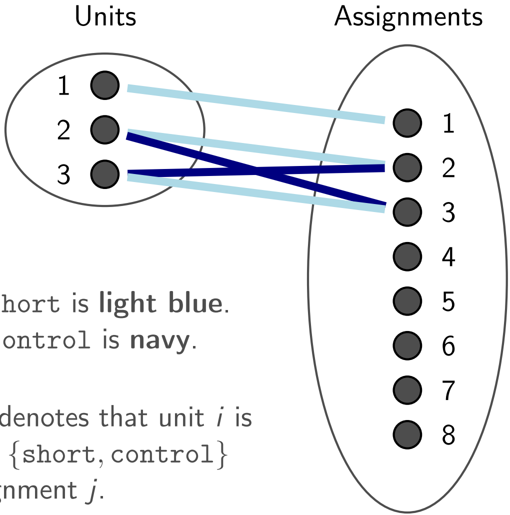
The null exposure graph



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.

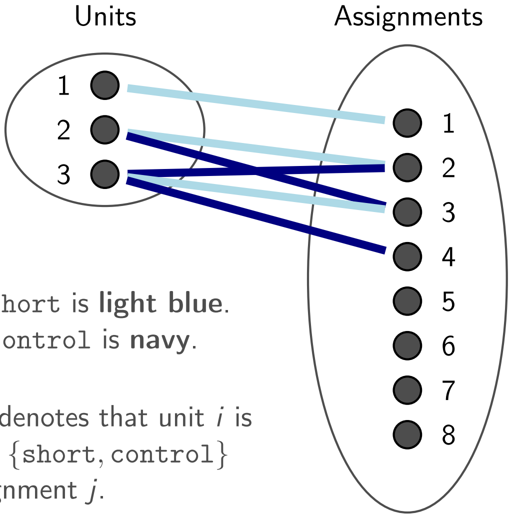
The null exposure graph



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.

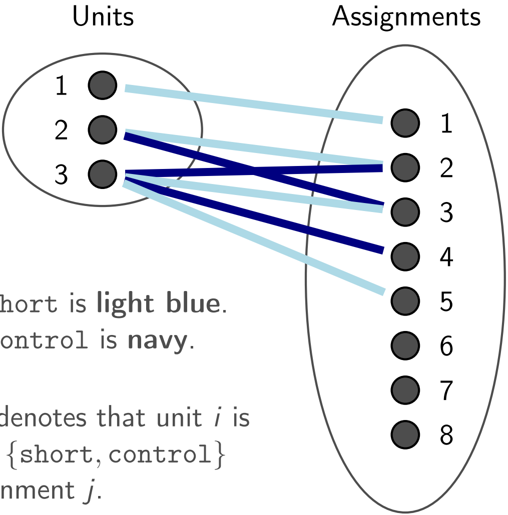
The null exposure graph



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.

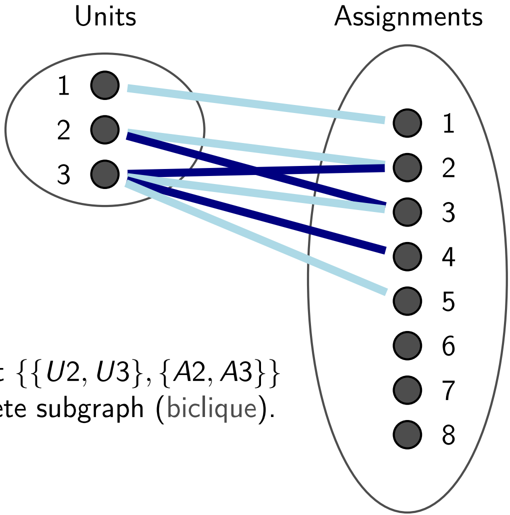
The null exposure graph



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

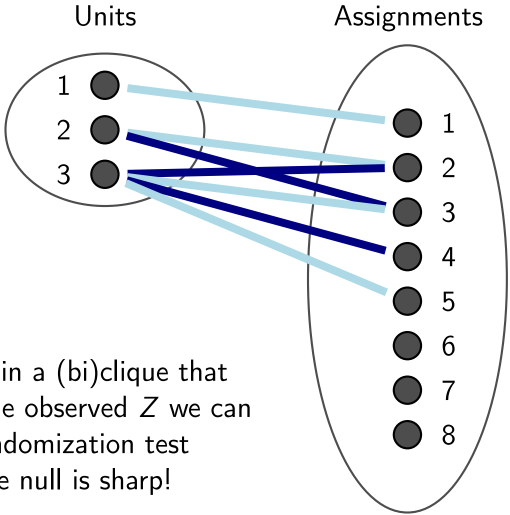
edge (i, j) denotes that unit i is exposed to $\{\mathtt{short}, \mathtt{control}\}$ under assignment j.

The null exposure graph



Notice that $\{\{U2, U3\}, \{A2, A3\}\}$ is a complete subgraph (biclique).

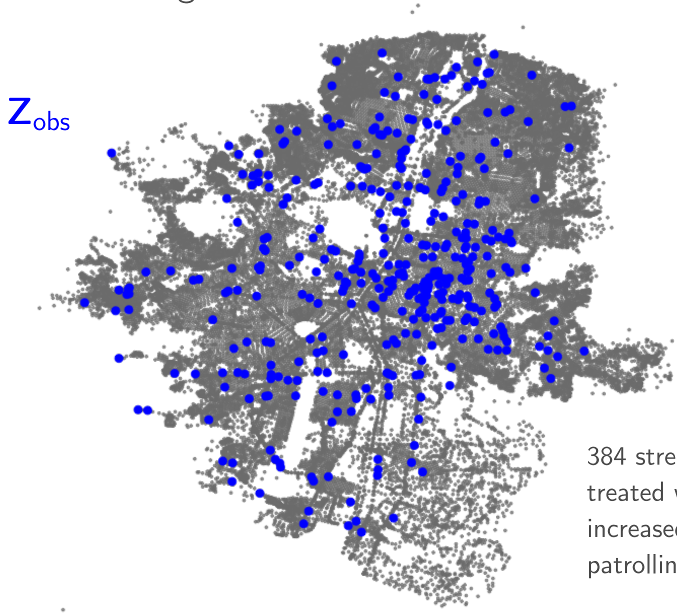
The null exposure graph



Idea: Within a (bi)clique that contains the observed Z we can run the randomization test because the null is sharp!

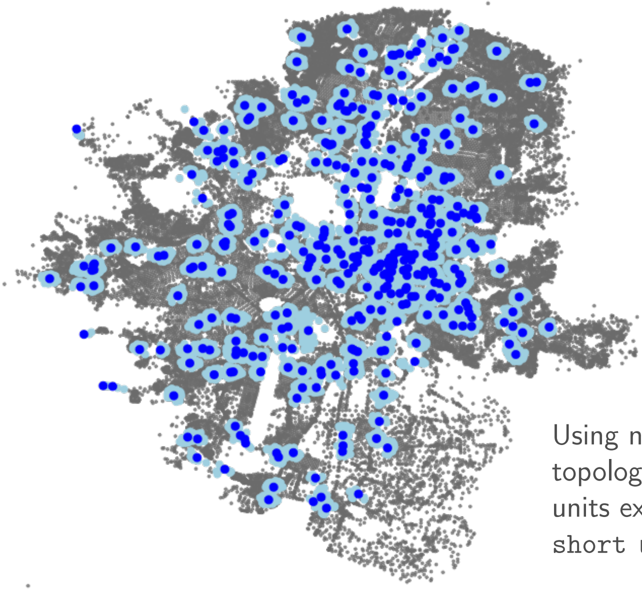
Returning to the map

The observed assignment



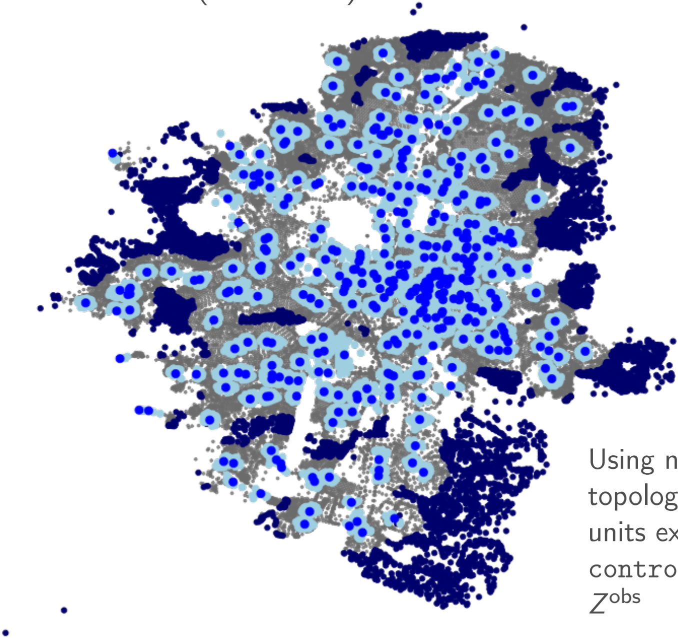
384 streets are treated with increased police patrolling

Short-range spillover units (`short`)



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

Pure control units (`control`)



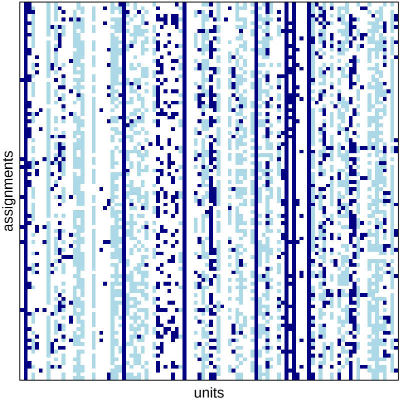
Using network topology, color units exposed to `control` under Z^{obs}

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

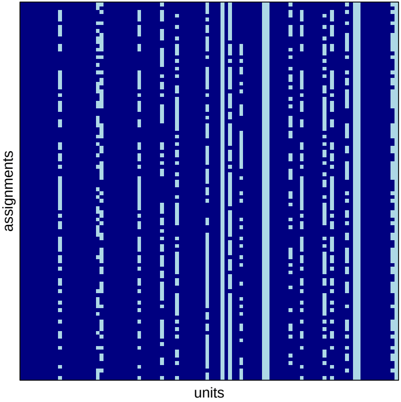
We can remake these pictures for every assignment Z drawn from design $P(Z)$ ...

$\rightarrow$ The output is our null exposure graph!

Null exposure graph and clique



null exposure graph



clique (zoomed-in)

Clique-based randomization test

- A null exposure graph uniquely defined given $\mathrm{H}_{\mathbf{0}}, \{f_i\}$.
- A "clique test stat." $t(y, z; C)$ where C is a clique in G_f such that

$$t(y, z; C) = t(y', z'; C), \text{ if } y_C = y'_C \text{ and } z_C = z'_C.$$

$\hookrightarrow$ *y_C, z_C are sub-vectors of y, z only with units/assignments in C.*

Clique-based randomization test

- A null exposure graph uniquely defined given $\mathrm{H}_{\mathbf{0}}, \{f_i\}$.
- A "clique test stat." $t(y, z; C)$ where C is a clique in G_f such that

$$t(y, z; C) = t(y', z'; C), \text{ if } y_C = y'_C \text{ and } z_C = z'_C.$$

$\hookrightarrow$ *y_C, z_C are sub-vectors of y, z only with units/assignments in C.*

1. **Decompose:** Compute biclique decomposition* of G_f. Pick out clique containing observed Z, call it C.
2. **Condition:** Compute $T^{\text{obs}} = t(Y, Z; C)$ given C.
3. **Summarize:** p-value $= \mathbb{E}\left[\mathbb{1}\{t(Y, Z'; C) \geq T^{\text{obs}}\} \mid C\right]$.

   $\hookrightarrow$ *Here, we sample with respect to*

$$P(Z'|C) \propto \underbrace{P(C|Z')}_{\textit{clique gen.}} \cdot \underbrace{P(Z')}_{\textit{design}}$$

Why this works

$*$ Test statistic T is defined only on clique C by Step 2.

Proof of validity:

$$t(Y', Z'; C) \overset{*}{=} t(Y'_C, Z'_C; C) \overset{H_0}{=} t(Y_C, Z'_C; C) \overset{d}{=} t(Y_C, Z_C) \overset{*}{=} t(Y, Z)$$

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

$\circ$ Imputation is possible because $Y'_C = Y_C$ within clique C.

Biclique decomposition

- 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 needs to be optimized.

  ↪*i.e., different biclique decompositions will have different power properties, but all are **valid**.*

Related work

We can also use our framework to describe related work:

- Aronow (2012) and Athey et al (2018) effectively propose to randomly sample *focal units* on one side, and then find the *maximum induced* clique to condition on.

  $\hookrightarrow$ *General procedure but the random selection of focals does not exploit the problem structure — Loss of power.*

- Basse et al (2019) develop a clique decomposition that provably leads to permutation test in clustered interference.

  $\hookrightarrow$ *Case-by-case analysis — Cannot generalize.*

Simulated study: clustered interference

To illustrate, we consider a **clustered interference** setting.

Suppose we have N units spread equally in K clusters.The clusters could be classrooms or households.

Experiment: Randomly treat K/2 clusters. Within each treated cluster, randomly treat 1 unit.

$\hookrightarrow$ *Motivated by student absenteeism study (Basse et al, 2019).*

Simulated study: clustered interference

To illustrate, we consider a **clustered interference** setting.

Suppose we have N units spread equally in K clusters.The clusters could be classrooms or households.

Experiment: Randomly treat K/2 clusters. Within each treated cluster, randomly treat 1 unit.

$\hookrightarrow$ *Motivated by student absenteeism study (Basse et al, 2019).*

◦ Do outcomes of a control unit in control cluster differ from outcomes of a control unit in a treated cluster?

The null and competing methods

$$H_0: \quad Y_i(\texttt{control}) = Y_i(\texttt{exposed}),\ \forall i,$$

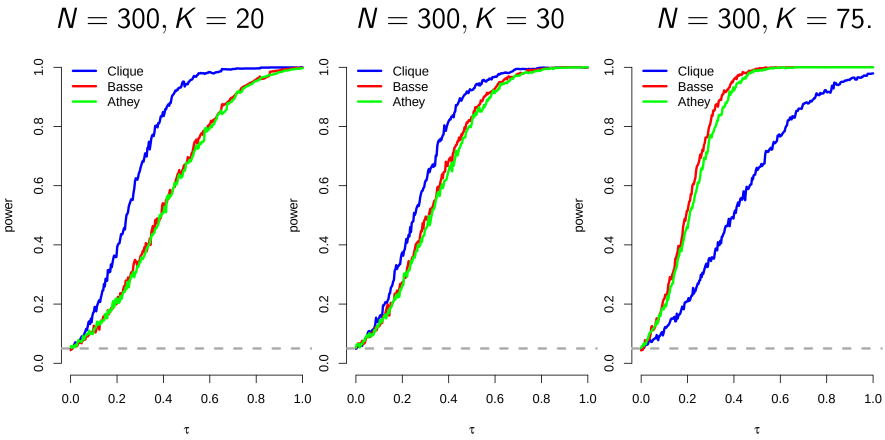
where:
$f_i(Z) = \texttt{control}$, if $Z_i = 0$ and $\sum_{j\in[i]} Z_j = 0$;
$f_i(Z) = \texttt{exposed}$, if $Z_i = 0$ and $\sum_{j\in[i]} Z_j = 1$, and $[i]$ denotes i's cluster.

1. Athey et. al. (2018): sample one focal per household. Run randomization test*.

2. Basse et. al. (2019): For treated households, sample one untreated focal unit (uniformly). For untreated households, sample one focal. Run permutation test on the focals.

3. Clique test – proposed method.

Power comparison: $Y_i(\texttt{exposed}) = Y_i(\texttt{control}) + \tau$

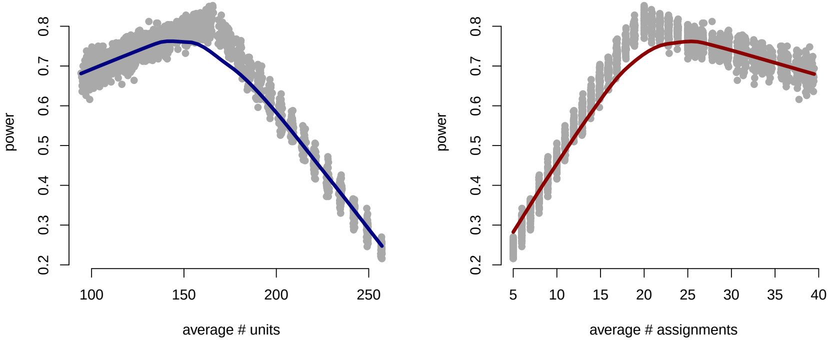


The clique test improves upon existing methods as the cluster size increases (smaller K)!

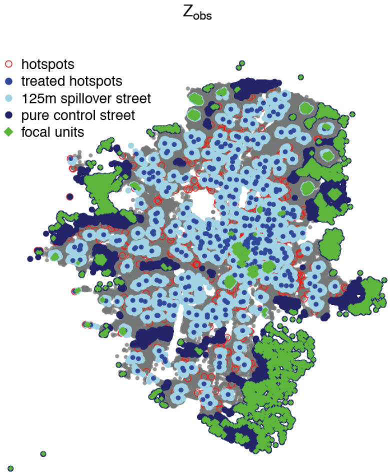
$\hookrightarrow$ *It achieves more flexible conditioning (i.e., many units/cluster).*

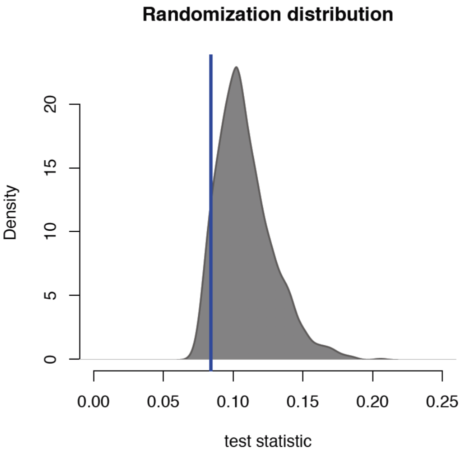
Power characteristics

Trade-off between #units,#assignments in the cliques.



Spatial interference: Medellin data





Concluding thoughts

- New method is presented for testing causal effects under general interference using null exposure graphs and bicliques.
- Structure is placed on null hypothesis through exposure functions.
- Translates the testing problem into graphical operations on the null exposure graph.
- Future work: understand power properties; optimized biclique decomposition; more hypotheses.

Thank You!

Working paper: “A Graph-Theoretic Approach to Randomization Tests of Causal Effects Under General Interference"

Athey, Eckles, Imbens, "Exact p-Values for Network Interference" (JASA, 2018)

Basse, Feller, Toulis, "Randomization tests of causal effects under interference" (Biometrika, 2019)

Aronow, "A general method for detecting interference between units in randomized experiments." (Sociol. Methods Res., 2012)

Extra slides

Experiment and data

Units and treatment assignment

- 37,055 total streets (units)
- 967 streets are identified as crime "hotspots"
- 384 are treated with increased police presence

Access to randomizations based on the design, pr(Z)

Outcomes and covariates

- Crime counts on all streets (murders, car and motorbike thefts, personal robberies, assaults)
- Survey data on hotspot streets
- Characteristics of hotspots (distance from school, bus stop, rec center, church, neighborhood, ...)

Definition. Let $\mathbb{U}, \mathbb{Z}$ denote the units and assignments, respectively. Let $\mathrm{a}, \mathrm{b} \in \mathcal{E}$ be any two exposures and consider the hypothesis:

$H_0^{\mathrm{a,b}} : Y_i(z) = Y_i(z'),$ for all i, z, z' such that $f_i(z), f_i(z') \in \{\mathrm{a}, \mathrm{b}\}$.

Define the vertex set as $V = \mathbb{U} \cup \mathbb{Z}$, and the edge set as

$$E = \{(i, z) \in \mathbb{U} \times \mathbb{Z} : f_i(z) \in \{\mathrm{a}, \mathrm{b}\}\}. \tag{1}$$

Then, $G_f = (V, E)$ is the null-exposure graph of $H_0^{\mathrm{a,b}}$ wrt f.

- For given $H_0^{\mathrm{a,b}}$ and $\{f_i\}$ the null exposure graph G_f is unique.
- Imputation is possible within the clique that contains obs. Z:

Definition. Let $\mathbb{U}, \mathbb{Z}$ denote the units and assignments, respectively. Let $\mathrm{a}, \mathrm{b} \in \mathcal{E}$ be any two exposures and consider the hypothesis:

$H_0^{\mathrm{a,b}}$: $Y_i(z) = Y_i(z')$, for all i, z, z' such that $f_i(z), f_i(z') \in \{\mathrm{a}, \mathrm{b}\}$.

Define the vertex set as $V = \mathbb{U} \cup \mathbb{Z}$, and the edge set as

$$E = \{(i, z) \in \mathbb{U} \times \mathbb{Z} : f_i(z) \in \{\mathrm{a}, \mathrm{b}\}\}. \quad (1)$$

Then, $G_f = (V, E)$ is the null-exposure graph of $H_0^{\mathrm{a,b}}$ wrt f.

- For given $H_0^{\mathrm{a,b}}$ and $\{f_i\}$ the null exposure graph G_f is unique.
- Imputation is possible within the clique that contains obs. Z:

Proposition. Consider a null-exposure graph, G_f, with some clique $C = (U, \mathcal{Z})$. If $Z^{\mathrm{obs}} \in \mathcal{Z}$, then $Y_i(z) = Y_i(Z^{\mathrm{obs}})$ under $H_0^{\mathrm{a,b}}$, for all $i \in U$ and all $z \in \mathcal{Z}$.