Randomization Tests of Causal Effects Under General Interference

Panos Toulis
University of Chicago, Booth School

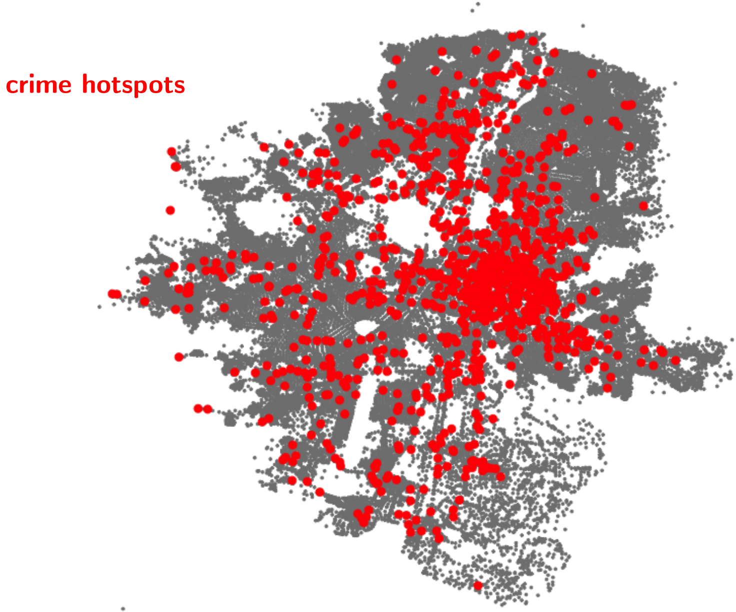
joint with: G Basse, A Feller, and D Puelz

Network Econometrics Conference, Northwestern University, 2019

Medellín, Colombia

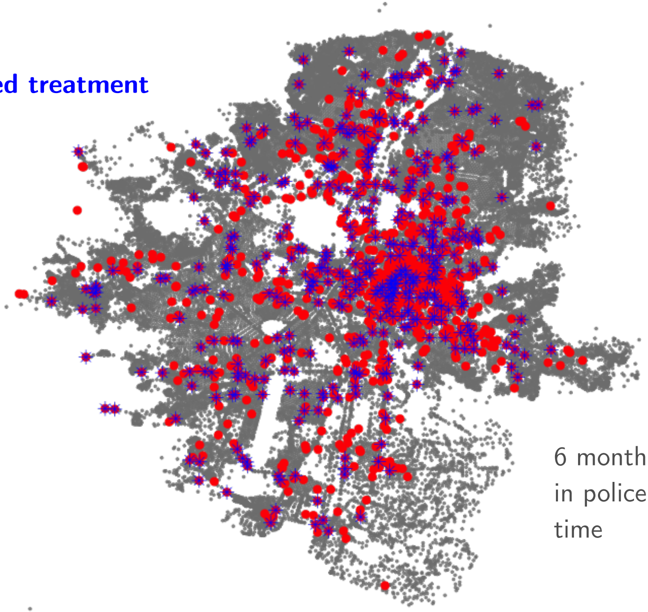
Medellín, Colombia

Medellín, Colombia



Medellín, Colombia

observed treatment

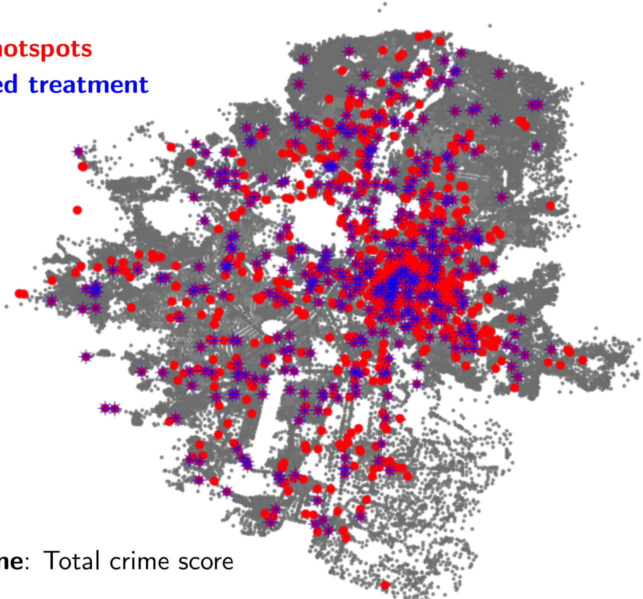


6 month increase in police patrolling time

Medellín, Colombia

crime hotspots

observed treatment

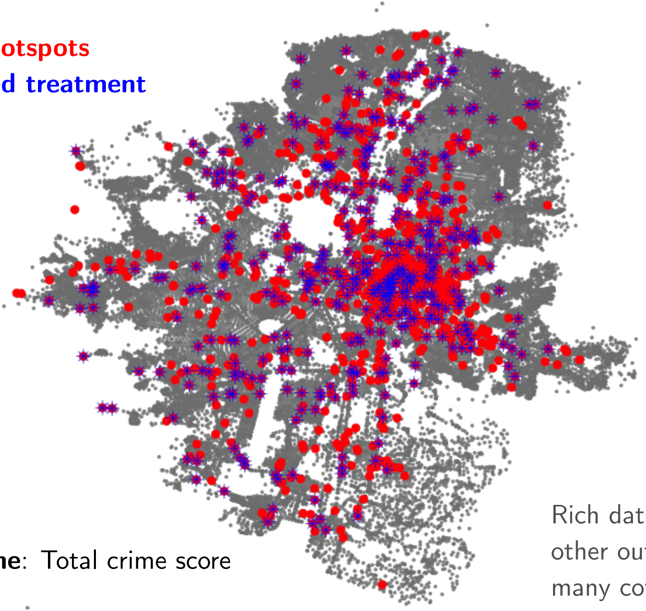


Outcome: Total crime score

Medellín, Colombia

crime hotspots

observed treatment



Outcome: Total crime score

Rich data set with other outcomes, many covariates

Question

How does the intervention affect crime?

$\rightarrow$ direct effect?

$\rightarrow$ spillovers to adjacent streets?

Question

How does the intervention affect crime?
$\rightarrow$ direct effect?
$\rightarrow$ spillovers to adjacent streets?

We will address these through hypothesis testing.

We would like to be (outcome) model-free, so we will use the randomization method of inference.

Notation and a classical test

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

Define observed data:
$Z = (Z_1, \ldots, Z_N)$ as binary treatment assignment;
$Y = (Y_1, \ldots, Y_N)$ as vector of observed outcomes.

The potential outcome of unit i under assignment z: $Y_i(z)$
i.e., total crime score

Assume no interference: $Y_i(z)$ depends only on z_i.
$\Rightarrow$ Only two potential outcomes, $Y_i(0)$, $Y_i(1)$, for every i.

Does treatment have an effect?

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

Fisher randomization test (1935)

$\mathbf{H_0}: \quad Y_i(0) = Y_i(1)$ 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 $Z' \sim \text{pr}(Z')$, 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)

$\mathbf{H_0}: \quad Y_i(0) = Y_i(1)$ 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 $Z' \sim \text{pr}(Z')$, store $T_r = T(Y', Z') \overset{H_0}{=} T(Y, Z')$.
3. p-value $= \mathbb{E}[\mathbb{1}\{T_r \geq T_{\text{obs}}\}]$.

Proof of validity:

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

"$T_{\text{obs}} \sim T_r$ (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).

No interference assumption is too strong ...

Assume: $Y_i(z)$ depends only on z_i (no interference)
$\rightarrow$ not very realistic for our application.

In reality, $Y_i(z)$ is **exposed** to (depends on) multiple parts of z.

No interference assumption is too strong ...

Assume: $Y_i(z)$ depends only on z_i (no interference)
$\rightarrow$ not very realistic for our application.

In reality, $Y_i(z)$ is **exposed** to (depends on) multiple parts of z.

One way to express more potential outcomes is through 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:

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

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

We can now ask questions in terms of **exposures**!

Question: Is there a short-range spillover effect?

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

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

$$f_i(Z) := \begin{cases} \texttt{short} & Z_i = 0, \mathrm{dist}_i < 125\mathrm{m} \\ \texttt{control} & Z_i = 0, \mathrm{dist}_i > 500\mathrm{m} \\ \texttt{neither} & \text{else} \end{cases}$$

$\mathrm{dist}_i :=$ distance to closest treated street.

Can we use the classical Fisher test again? Not quite ...

Recall, observed $T \sim$ randomized T for things to work:

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

The null only assumes 2 of the 3 exposures have equal outcomes

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

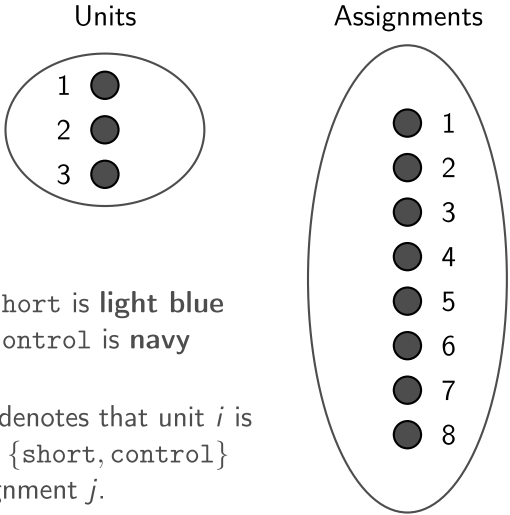
In this case, the null is **not sharp.** We cannot impute potential outcomes $Y^{'}$ freely under any $Z^{'}$.

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

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

This is a binary relationship!

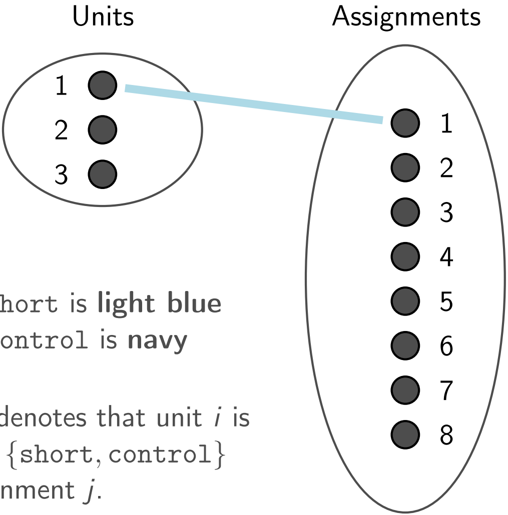
The null exposure graph



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.

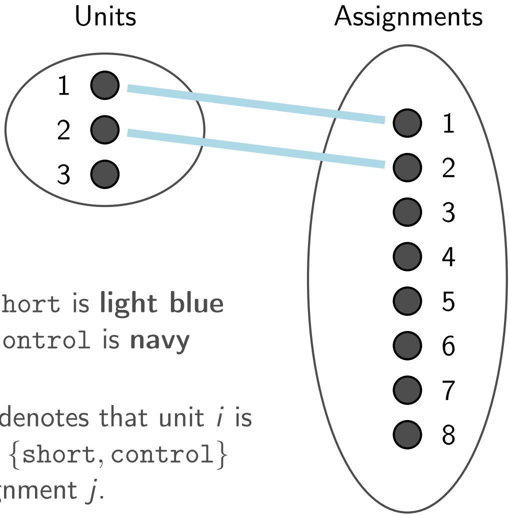
The null exposure graph



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.

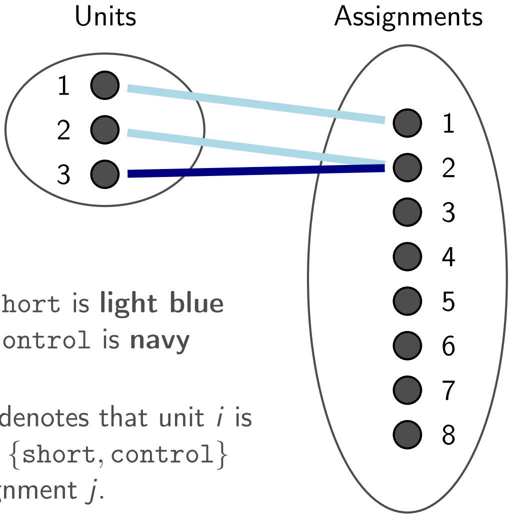
The null exposure graph



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.

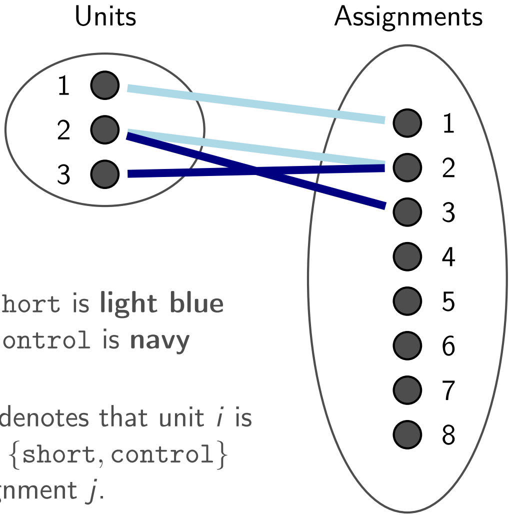
The null exposure graph



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.

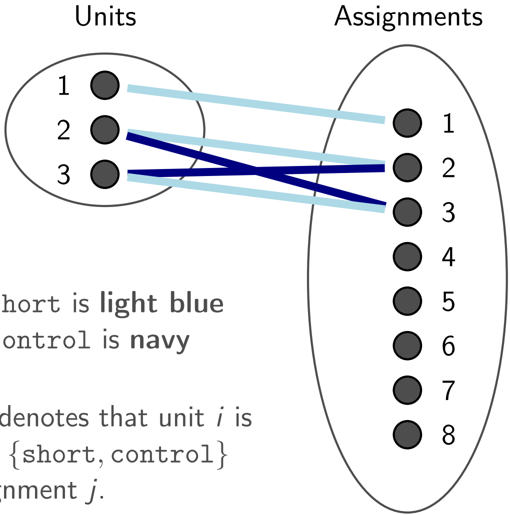
The null exposure graph



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.

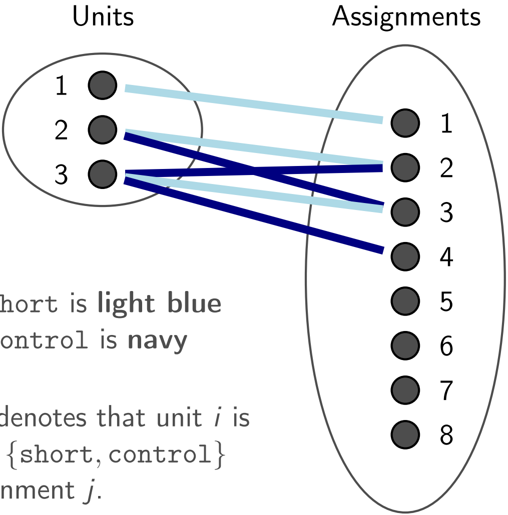
The null exposure graph



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.

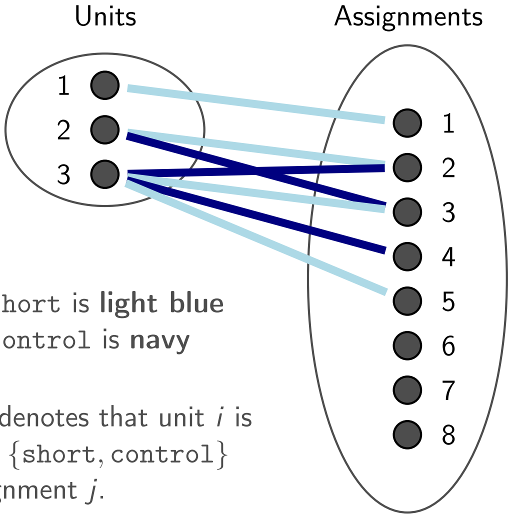
The null exposure graph



Exposure $\texttt{short}$ 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.

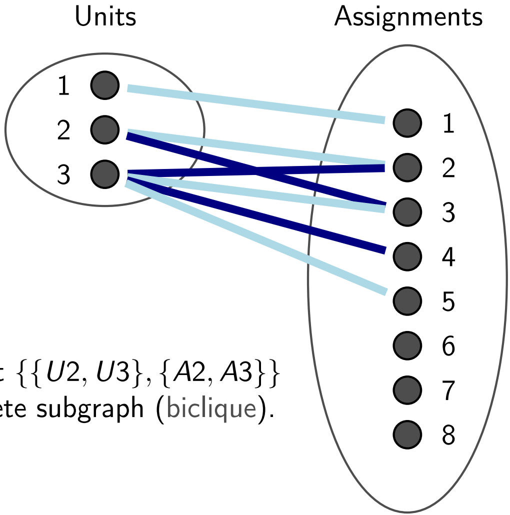
The null exposure graph



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

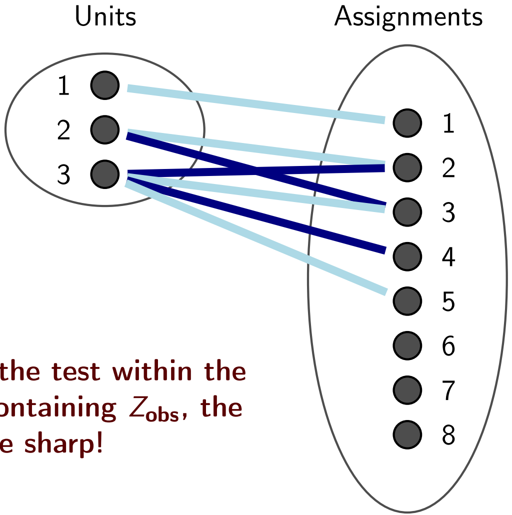
edge (i, j) denotes that unit i is exposed to $\{\texttt{short}, \texttt{control}\}$ under assignment j.

Our main contribution: The null exposure graph



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

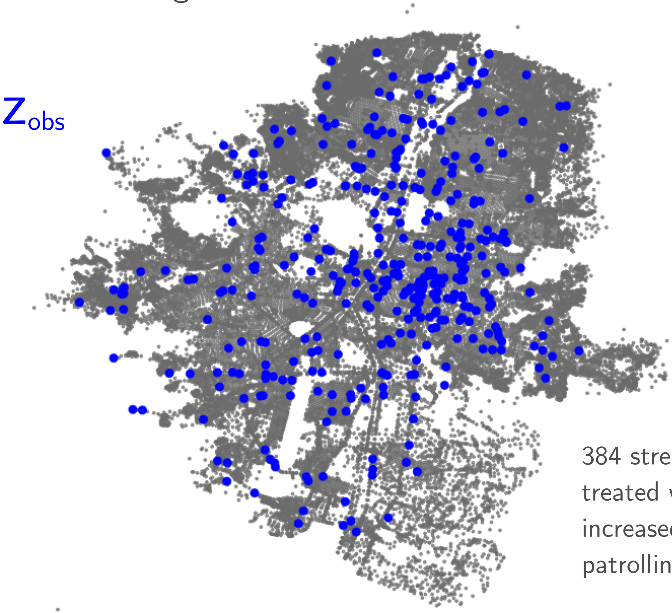
Our main contribution: The null exposure graph



If we run the test within the biclique containing $Z_{\mathbf{obs}}$, the null will be sharp!

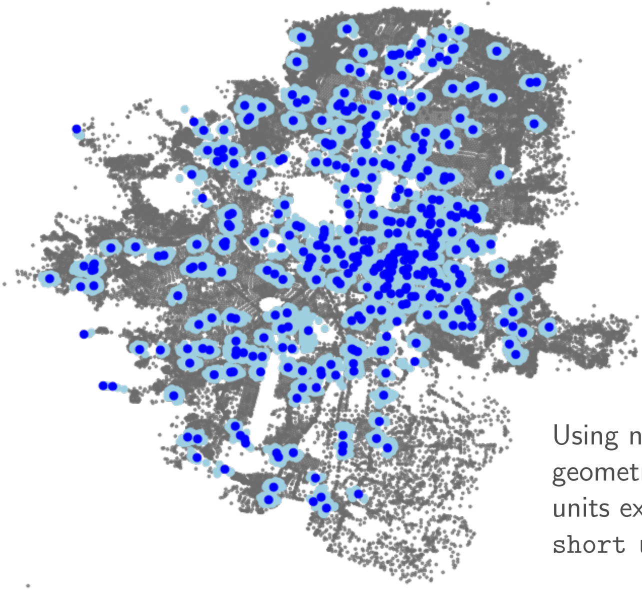
Returning to the map

The observed assignment



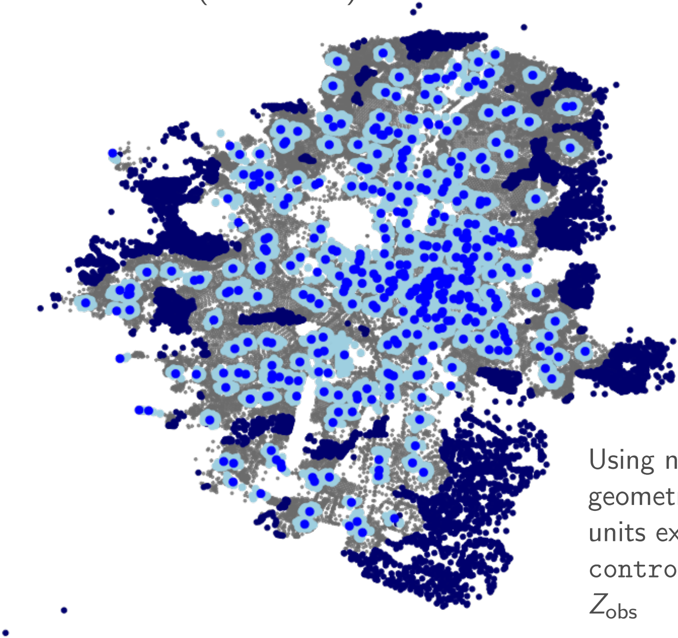
384 streets are treated with increased police patrolling

Short-range spillover units (`short`)



Using network geometry, color units exposed to `short` under Z_{obs}

Pure control units (`control`)



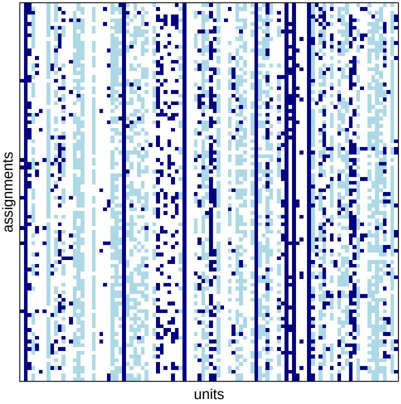
Using network geometry, color units exposed to `control` under Z_{obs}

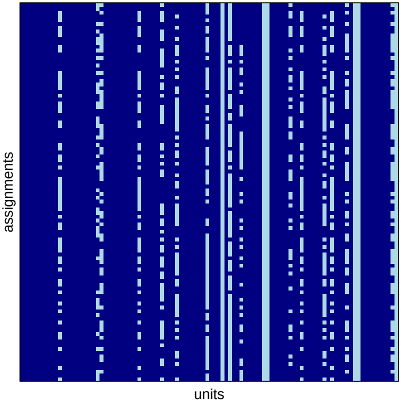
We can remake these pictures for every assignment Z drawn from $\mathrm{pr}(Z)$ ...

We can remake these pictures for every assignment Z drawn from pr(Z) ...

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

Null exposure graph and biclique





Clique-based randomization test

$\rightarrow$ A null exposure graph uniquely defined given $\mathbf{H_0}$.

$\rightarrow$ A test statistic $T = T(y, z)$.

1. **Decompose:** Compute biclique decomposition of null exposure graph. Pick out biclique with Z_{obs}, call it C.

2. **Condition:** Compute test statistic values with units and assignments only in C.

3. **Summarize:** p-value $= \mathbb{E}_{Z_C}\left[\mathbb{1}\{T_C \geq T_{\text{obs}}\}\right]$.

   Here, $P(Z_C) \propto pr(Z_C)\mathbb{1}\{Z_C \in C\}$

Conditioning in this way gives a valid method!

Clique test statistics: $\quad T_C = T(Y_C, Z_C)$

*T is defined only in C by **condition** step in method

For every Z, Z', we need to show $T(Y', Z') \stackrel{d}{=} T(Y, Z) \mid C$

Proof:

$$T(Y', Z') \stackrel{*}{=} T(Y'_C, Z'_C) \stackrel{H_0}{=} T(Y_C, Z'_C) \stackrel{d}{=} T(Y_C, Z_C) \stackrel{*}{=} T(Y, Z)$$

Related work

We can use our framework to describe related work:

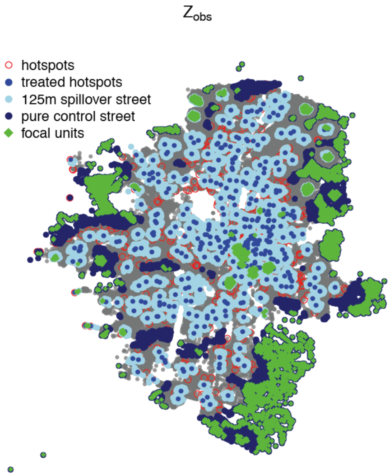
- Aronow (2012) and Athey et al (2018) effectively propose to randomly sample focal units on one side, and then find the *maximal induced* clique to condition on.

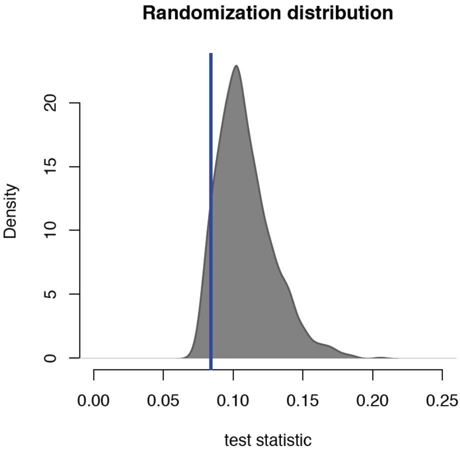
  - General procedure but the random selection does not exploit the problem structure $\Rightarrow$ Loss of power.

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

  - Case-by-case analysis. Cannot generalize.

A test of the null on 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.

- 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

Why is this a valid method?

Clique test statistics: $T_C = T(Y_C, Z_C)$

*T is defined only in C by **condition** step in method

For every Z, Z', we need to show $T(Y', Z') \overset{d}{=} T(Y, Z) \mid C$

Proof:

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

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, ...)

Considerations / alternative approaches

- Finding bicliques is hard, actually, **NP-hard**[1]

- The method is **constructive**, still needs to be optimized

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

- Other conditional testing methods:

  Aronow 2012, Athey et al. 2018. (Roughly) equivalent to randomly sampling units one one side, then computing the clique that contains those units and obs Z.
  $\Rightarrow$ loses power.

  Basse et al. 2019. Biclique sampling can depend on obs Z.
  $\Rightarrow$ easier when interference has structure.

[1] We use Binary Inclusion-Maximal Biclustering Algorithm, which uses a divide and conquer method to find bicliques.

What about simulated data?

We consider a **partial interference** setting.

Suppose we have N observations living in K blocks. The blocks could be classrooms or households.

Experiment: Randomly treat K/2 blocks. Within treated households, randomly treat 1 observation.

What about simulated data?

We consider a **partial interference** setting.

Suppose we have N observations living in K blocks. The blocks could be classrooms or households.

Experiment: Randomly treat K/2 blocks. Within treated households, randomly treat 1 observation.

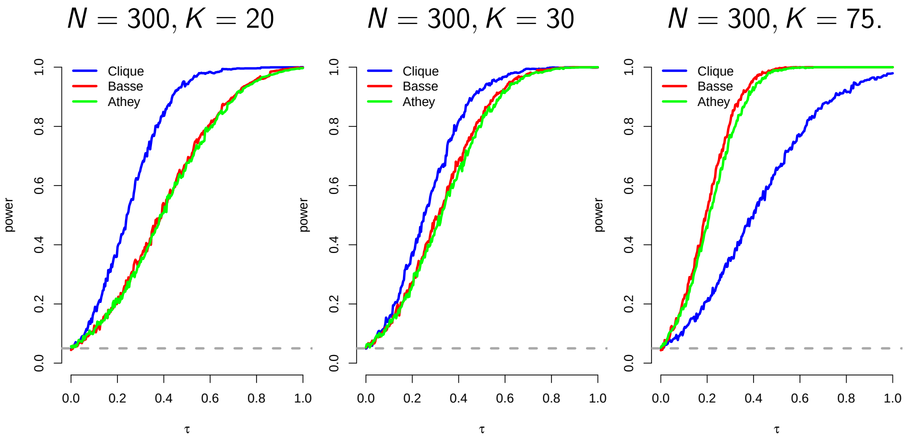
Do outcomes in control households differ from outcomes of control observations in treated households?

The null and competing methods

$$\mathbf{H}_0 : \qquad Y_i(\texttt{control}) = Y_i(\texttt{exposed}),\ \forall i$$

1. **Athey** et. al. JASA (2018): sample one focal per household. Run permutation test.
2. **Basse** et. al. Biometrika (2019): for treated households – sample one untreated focal, for untreated, sample one focal. Run permutation test.
3. **Clique** – proposed method.

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



The clique method improves upon existing methods as the block size increases!