Randomization Tests of Causal Effects Under General Interference

Panos Toulis

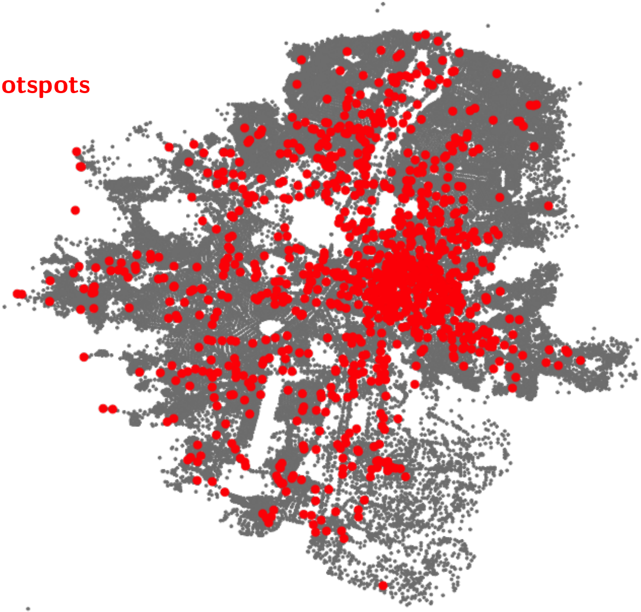
University of Chicago,
Booth School of Business

April 24, 2019

Medellín, Colombia

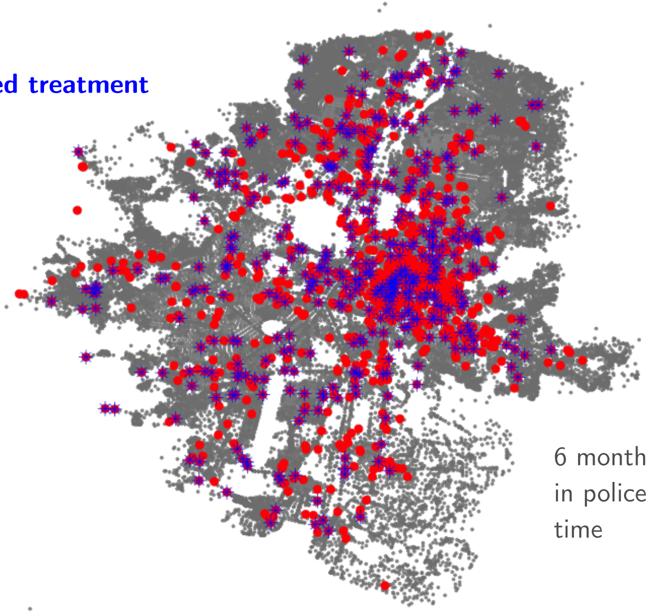
Medellín, Colombia

Medellín, Colombia



Medellín, Colombia

observed treatment



6 month increase in police patrolling time

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)

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, aggregate crime score)
- Survey data on hotspot streets
- Characteristics of hotspots (distance from school, bus stop, rec center, church, neighborhood, ...)

Questions we aim to answer

How does the intervention affect crime?

* direct effect?
* spillovers to adjacent streets?

Questions we aim to answer

How does the intervention affect crime?

- ∗ direct effect?
- ∗ spillovers to adjacent streets?

We will answer these through hypothesis testing.

We use the randomization mode of inference. It is robust and model-free.

The classical randomization test

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

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?

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

Fisher randomization test (1935)

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

The procedure:

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

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

Fisher randomization test (1935)

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

The procedure:

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

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

Proof of validity:

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

"$T^{\text{obs}} \overset{d}{=} T^{\text{rand}}$ (under null)"

Advantages of Fisherian randomization

- Exact. The test is valid in finite samples.

- Minimal assumptions. No model for Y. Regression analysis of peer effects can be tricky (Angrist, 2014).

- Robust. Test gives the same answer with different Y-scales (the same cannot be said for regression).

Possible limitation:

The test requires imputation from Y to any Y'; i.e., H_0 has to be a sharp null.

No interference assumption is too strong...

No interference is not realistic in our application.

We expect $Y_i(z)$ to depend on multiple components of z_i.

We cannot write "$Y_i(0) = Y_i(1)$". There are more potential outcomes.

One way to express more potential outcomes is through the concept of treatment exposure.

Treatment exposures

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

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

$\mathcal{E}$ is the set of possible exposures (one neighboring street treated, no neighboring streets treated, etc.)

Both $\mathcal{E}$ and f_i need to be defined by the analyst. Any choice will likely be contentious.

We can now ask questions in terms of exposures:

> *Is there a difference in outcome between short-range and pure control streets?*

Question: Is there a short-range spillover effect?

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

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

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

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

We cannot use the classical Fisher test

Recall: we need $T^{\text{obs}} \overset{d}{=} T^{\text{rand}}$ for things to work.

$$T^{\text{rand}} = T(Y', Z') \overset{H_0}{\not=} T(Y, Z') \overset{d}{=} T(Y, Z) = T^{\text{obs}}.$$

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

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

Here, the null is **not sharp.** We cannot impute potential outcomes Y' freely under any Z'.

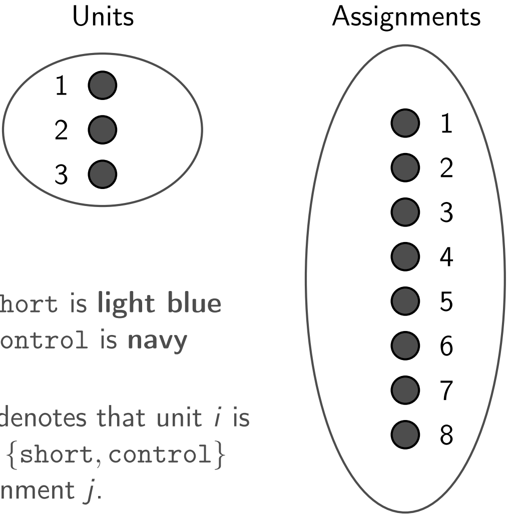
Fisherian randomization works only with sharp/global nulls.

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!
How can we visualize?

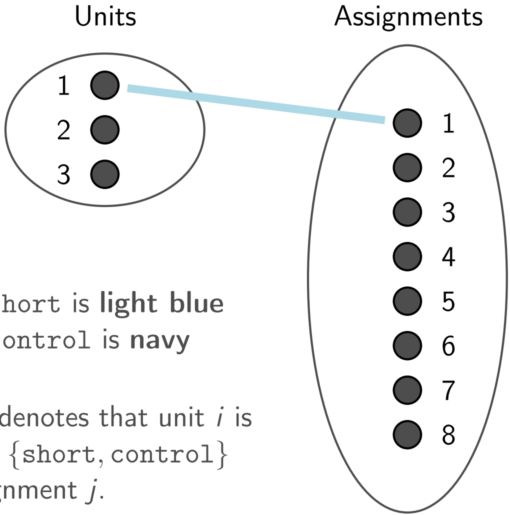
Our main contribution: The null exposure graph



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.

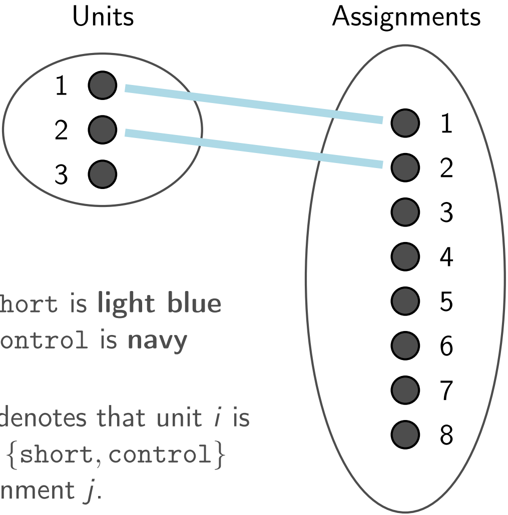
Our main contribution: The null exposure graph



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.

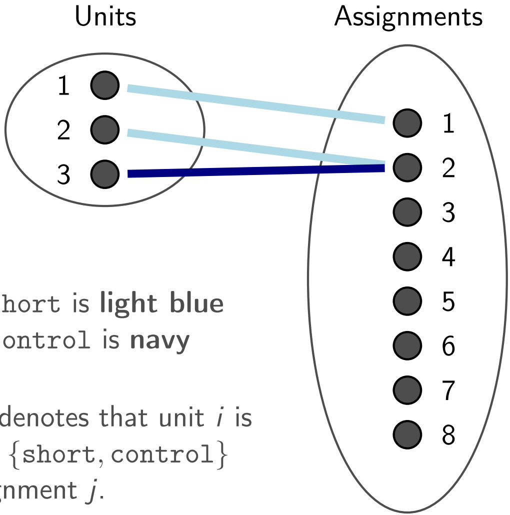
Our main contribution: The null exposure graph



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.

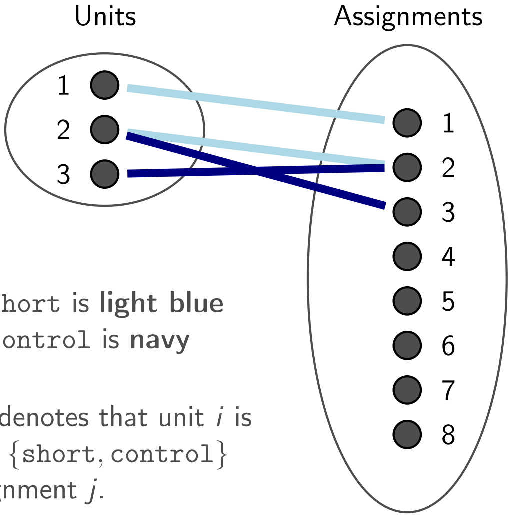
Our main contribution: The null exposure graph



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.

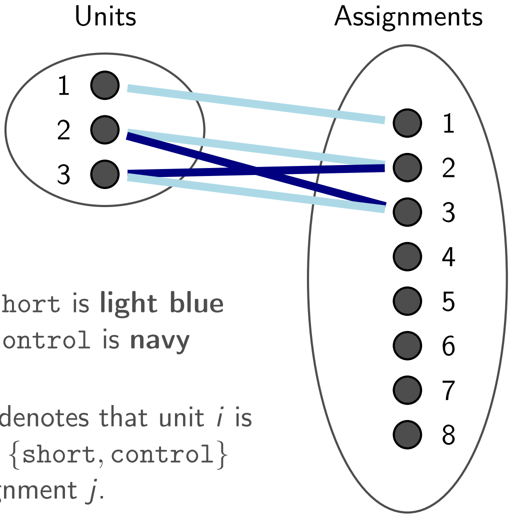
Our main contribution: The null exposure graph



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.

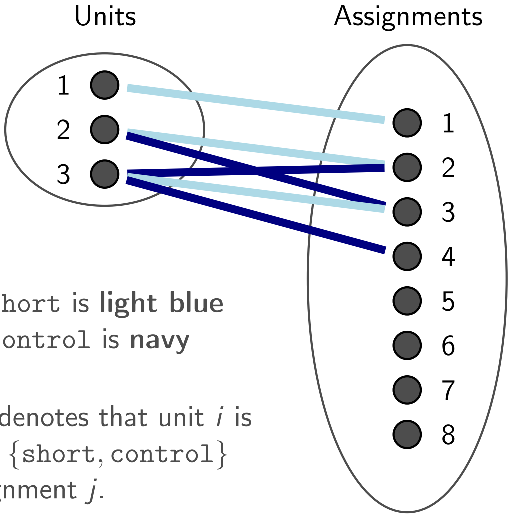
Our main contribution: The null exposure graph



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.

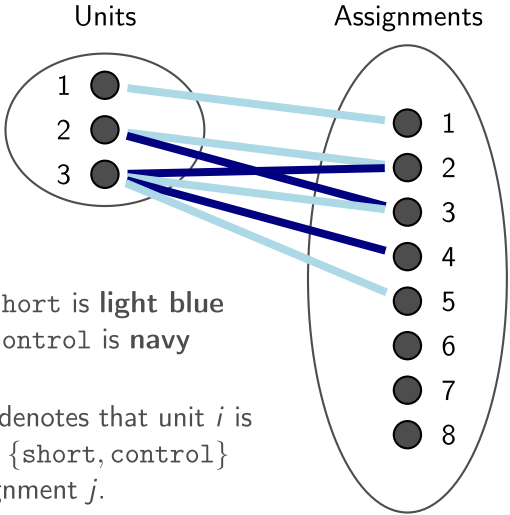
Our main contribution: The null exposure graph



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.

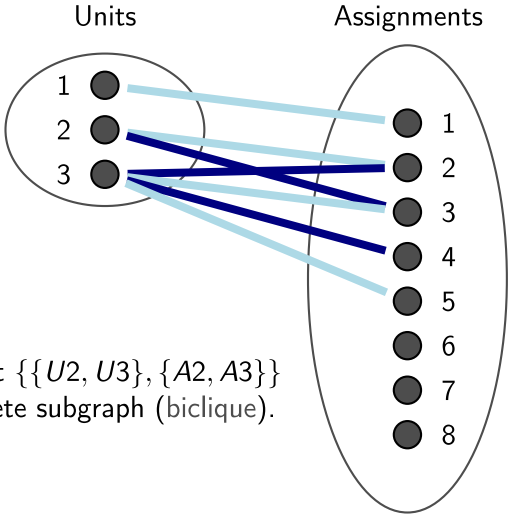
Our main contribution: The null exposure graph



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.

Introducing the null exposure graph



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

Why are these bicliques useful?

Within a biclique, every unit is exposed to either `short` or `control` under any assignment.

i.e.: If obs Z is in a biclique, we can impute potential outcomes, H_0 **is sharp** within the biclique.

Let's outline the method ...

Biclique method

Input:

$*$A null exposure graph uniquely defined given $\mathsf{H_0}$.

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

1. **Decompose:** Compute biclique decomposition of null exposure graph. Pick out biclique with obs Z, say 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, $P(Z_C) \propto pr(Z_C)\mathbb{1}\{Z_C \in C\}$

Why is this a valid method?

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

$\star T$ is defined only in C by step 2.

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

Proof:

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

Considerations

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

- Our method could be optimized;

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

- Other conditional testing methods:

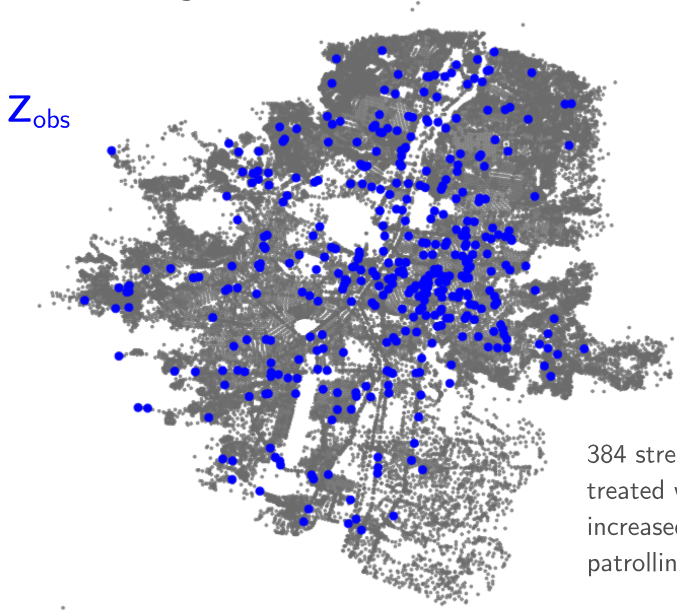
  Aronow 2012, Athey et al. 2018. (Roughly) equivalent to randomly sampling units one one side, then computing the clique that contains those units and obs Z.
  $\Rightarrow$ loses power.

  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.

Returning to the map

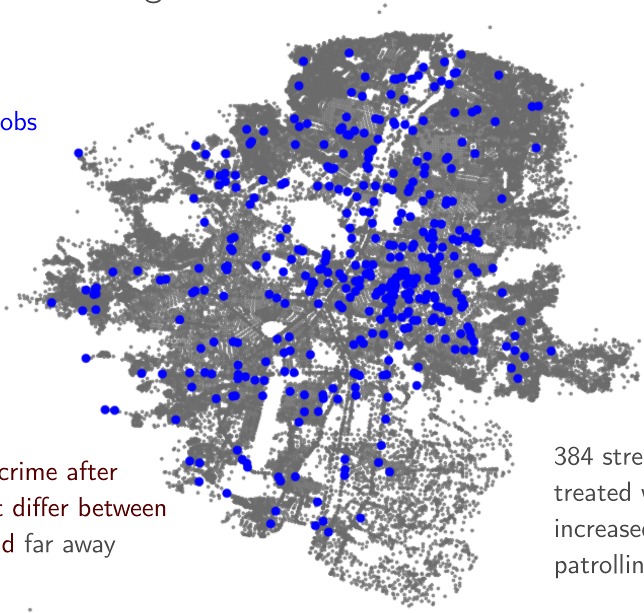
The observed assignment



384 streets are treated with increased police patrolling

The observed assignment

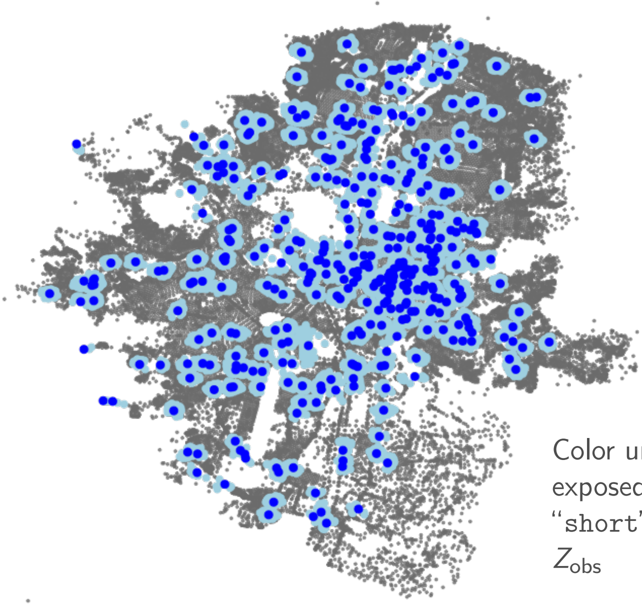
Z_{obs}



Q: Does crime after treatment differ between nearby and far away streets?

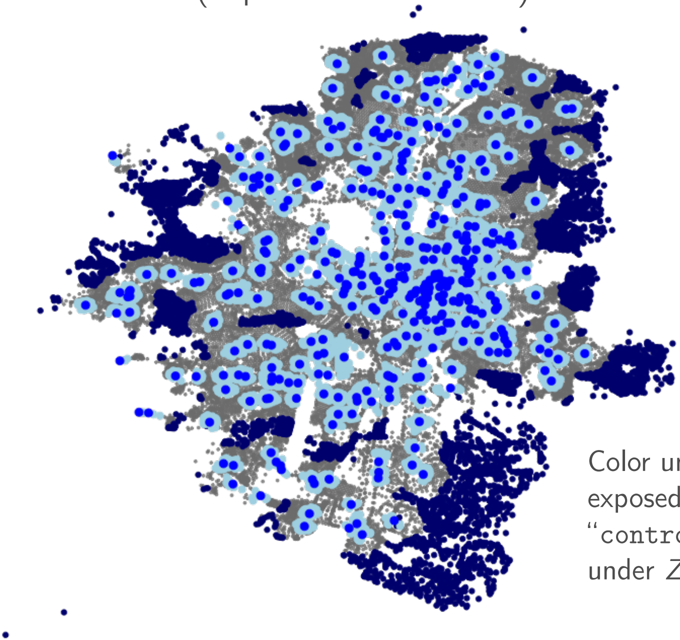
384 streets are treated with increased police patrolling

Short-range spillover units (exposure "short")



Color units exposed to "short" under Z_{obs}

Pure control units (exposure "`control`")



Color units exposed to "`control`" under Z_{obs}

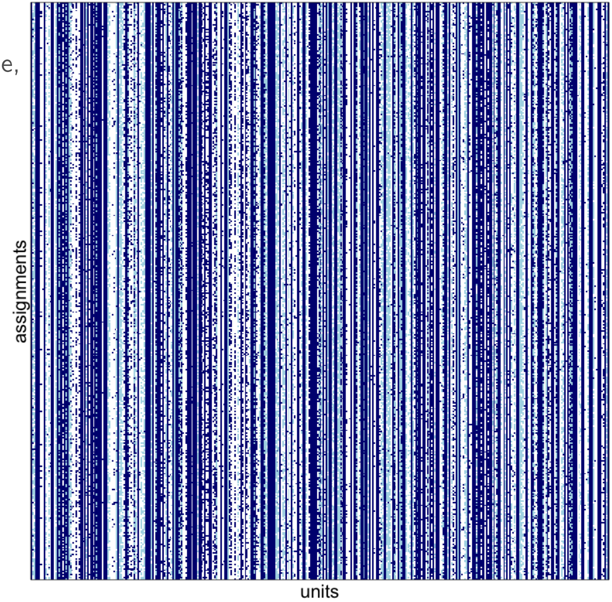
We can remake these pictures for every assignment Z drawn from $\text{pr}(Z)$ ...

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

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

Null exposure graph

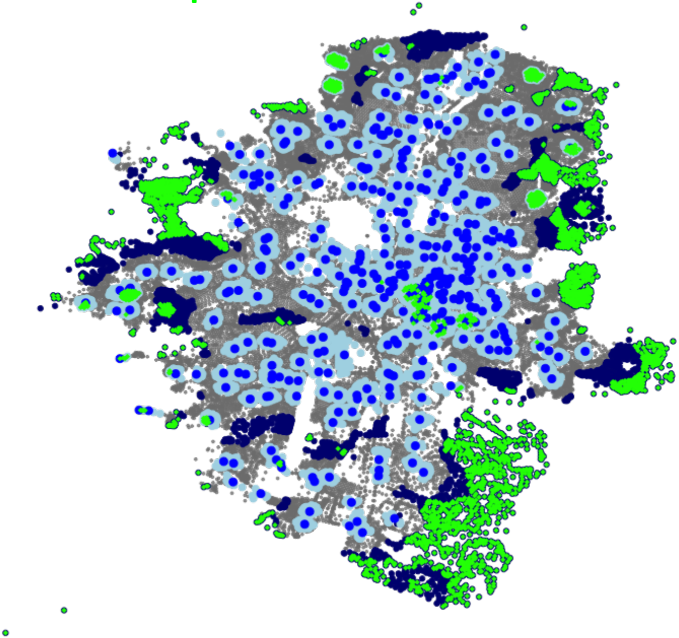
navy, light blue, and white



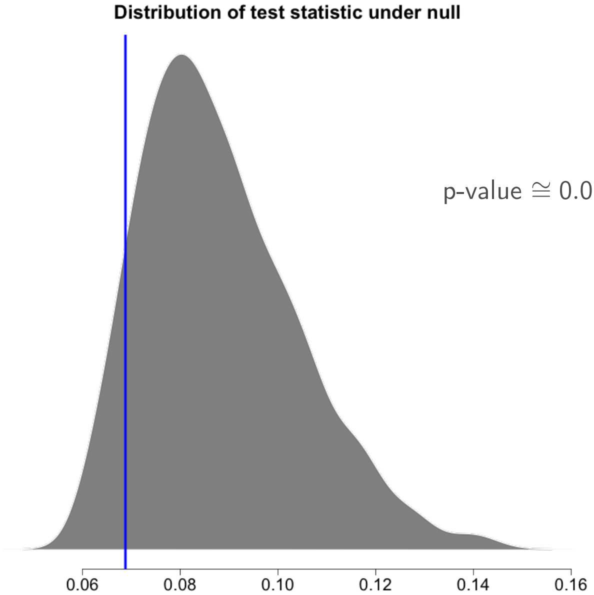
Biclique containing the observed assignment

only navy and
light blue!

Where are the **clique units?**



A test of the null



p-value $\cong 0.07$

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!

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)