Randomization Inference for Spillover Effects

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

Econometrics and Statistics
University of Chicago, Booth School of Business

Bocconi University, Statistics Seminar, 2022

Introduction

The standard methods of causal inference assume no interference;
i.e., treatment on an individual unit cannot affect other units.

This assumes a simple, static world.

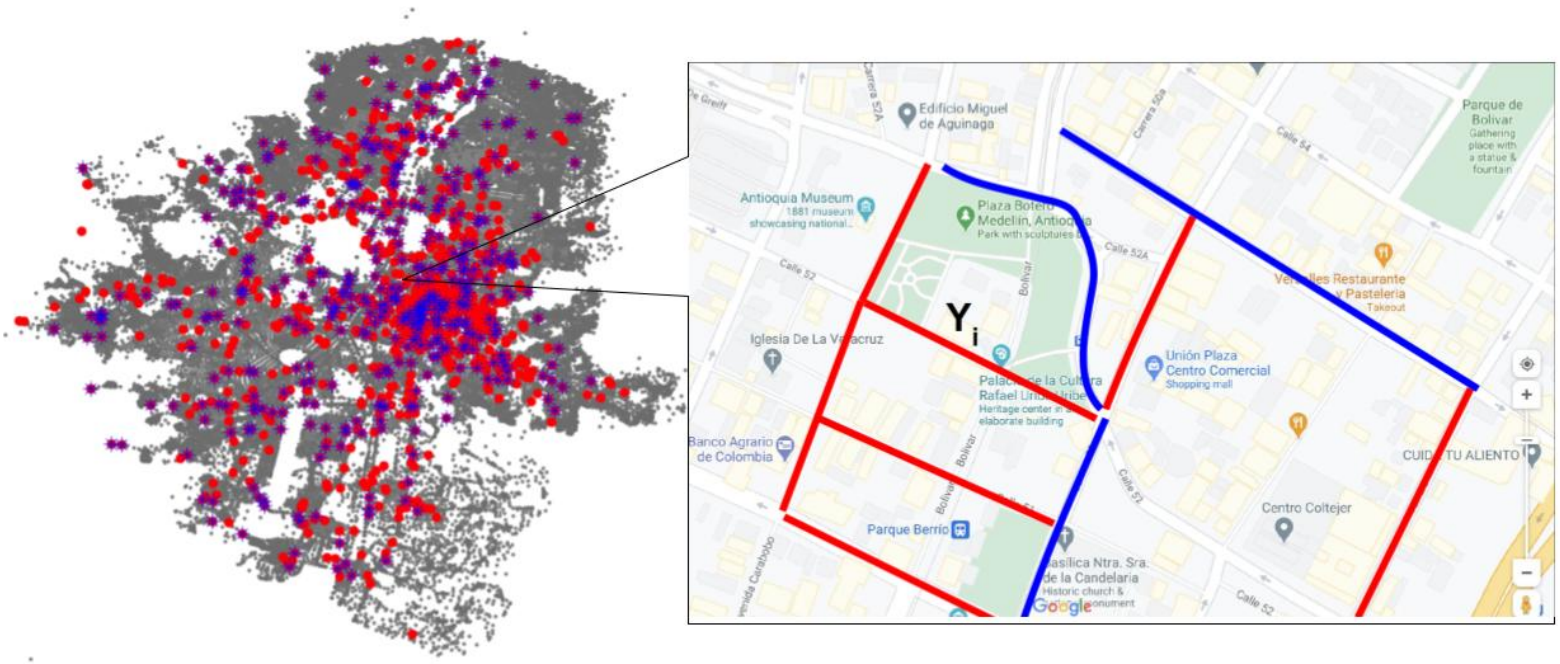
However, many interesting problems exist in settings where units interact in a complex way and influence each other.
—spillovers, peer effects, contagion, equilibrium effects, etc.

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

New methods and tools are needed. Many applications:
e.g., policy making, marketplace algorithms, climate science, healthcare, etc.

Motivation: Crime in Medellin, Colombia (Collazos et al, 2019)

Crime spillovers from nearby treated streets on control streets?



treatment = increased policing; control = baseline policing.

- What is a proper definition of a spillover effect?
- How to estimate it? Quantify uncertainty?

Current Approaches

- Model-based approach. Typically, regressions of the form:

$$\text{outcome}_i \sim \text{treatment}_i + \text{treatment\_group}_i.$$

  Validate with experiment, IV, "parallel trends", etc.

- Design-based approach. Exploit known variation in Z (e.g., from an experiment). The "potential outcomes" are fixed.

  e.g., Randomized studies. Remains the gold standard of causal inference.

- Causal graphs: Not today.

- DSGE-style / structural models: "Model-based approach on steroids".
  Still popular in macro policy making.

In this talk, I will highlight the contrast between model-based and design-based approaches.

But there exists a fruitful synergy: A model-based method could supply the estimator, and the design-based method could quantify uncertainty.

Pitfalls of model-based approach

A model-based approach requires correct specification, and is always open to potential biases.

A more pernicious problem is how this approach quantifies uncertainty.

Pitfalls of model-based approach

A model-based approach requires correct specification, and is always open to potential biases.

A more pernicious problem is how this approach quantifies uncertainty.

Example: Suppose a completely randomized design (50% treated/control):

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

Regress $Y_i \sim Z_i$. The estimate of "causal effect" is +5.

What is the standard error?

Pitfalls of model-based approach

A model-based approach requires correct specification, and is always open to potential biases.

A more pernicious problem is how this approach quantifies uncertainty.

Example: Suppose a completely randomized design (50% treated/control):

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

Regress $Y_i \sim Z_i$. The estimate of "causal effect" is +5.

What is the standard error? $O(\epsilon)$. Arbitrary level of certainty.

$\hookrightarrow$ But this is artificial. Standard error estimation is conflated with model fit.
(Here, the data fit a line very well).

Design-based approach

A design-based approach exploits the actual variation in the experiment.

The idea is to "predict" outcomes under counterfactual treatment assignments. Then compare with what was observed.

When possible, this is a more accurate way of quantifying uncertainty.

Design-based approach

A design-based approach exploits the actual variation in the experiment.

The idea is to "predict" outcomes under counterfactual treatment assignments. Then compare with what was observed.

When possible, this is a more accurate way of quantifying uncertainty.

To illustrate, suppose a counterfactual assignment, $Z^* = (0, 1, 1, 0)$.

— This assignment is equally probable to the observed one according to the design

What would be the outcomes Y^* under Z^* ?

Unit (i)	Treatment (Z_i^*)	Outcome (Y_i^*)
1	0	?
2	1	?
3	1	?
4	0	?

Design-based approach

If the treatment does not affect outcomes, then Y^* should be equal to the observed Y.

Design-based approach

If the treatment does not affect outcomes, then Y^* should be equal to the observed Y.

That is, the observed data would be as follows:

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

Under Z^*, we would have observed an effect of -5 instead of $+5$.

We can repeat this procedure for all 6 possible randomizations.

Observing an effect of +5, although extreme, has a $1/6 > 16\%$ chance of happening. No significance. (cf. linear model).

The goal will be to generalize this idea to settings with interference.

Outline

1. 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

Notation

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

$Z_i \in \{0, 1\}$	random treatment of unit i
Y_i	observed outcome of unit i
$y_i(0),\ y_i(1)$	potential outcomes of unit i

In causal inference, the potential outcomes are used to define the quantity of interest (e.g., $\bar{y}(1) - \bar{y}(0)$, the average treatment effect).

The potential outcomes are fixed, and the observed outcome is random only due to the random treatment assignment:

$$Y_i = Z_i y_i(1) + (1 - Z_i) y_i(0). \quad \text{[stability assumption, "SUTVA"]}$$

This assumption already precludes interference. We will revise later.

Notation

Population quantities:

$Z = (Z_1, \ldots, Z_N) \in \{0,1\}^N =: \mathbb{Z}$ population treatment

$Y = (Y_1, \ldots, Y_N) \in \mathbb{R}^N$ population outcomes

Z^*, Y^* randomization quantities (counterfactual)

In the potential outcomes framework, Y is determined by Z.
e.g., under SUTVA, $Y = Z \cdot y(1) + (1 - Z) \cdot y(0)$.

Similarly, Y^* is determined only by Z^*, under the same relationship.

Thus, whenever $Z^* \stackrel{d}{=} Z$, it holds that $(Z^*, Y^*) \stackrel{d}{=} (Z, Y)$.

The problem is that while Y is observed, Y^* is never observed.

In a design-based approach, we need to impute Y^* for various Z^*.

General Idea: Fisher's Randomization Test (1935)

$P(Z) : \mathbb{Z} \to [0, 1]$ is the experimental design. Assumed known.

As before, we want to test that there is no effect from the treatment:

$$H_0 : y_i(0) = y_i(1).$$

General Idea: Fisher's Randomization Test (1935)

$P(Z) : \mathbb{Z} \to [0, 1]$ is the experimental design. Assumed known.

As before, we want to test that there is no effect from the treatment:

$$H_0 : y_i(0) = y_i(1).$$

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

1. $T^{\text{obs}} = t(Y, Z)$.
2. Build randomization distr.: $F_R = \{\, t(Y, Z^*) : Z^* \sim P\}$
3. p-value $= 1 - F_R(T^{\text{obs}})$.

Proof of validity: A key implication of H_0 is that $Y^* = Y$ ("sharp null"). So,

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

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

An assessment of FRT

Major benefits:

- The test is exact in finite samples. No asymptotics.
- Not necessary to have correct Y-model specification.
- Robustness. Same answer under transformations of Y.
  (cf. regression/ML on $\log Y$ may yield completely different results than on Y)

Some disadvantages:

- Can only test "strong" hypotheses.
  (Currently, a lot of research activity in this area).
- Cannot generalize to population.
  (Personal opinion: this is a feature, not a bug.)

Randomization inference is widely considered the "gold standard" in experimental studies.
(Imbens and Rubin, 2015); (Rosenbaum, 2002); (Gerber and Green, 2012)

Interference

This machinery (both design-based and model-based) breaks down under interference.

e.g., spillovers, peer effects, contagion, dynamics.

Under interference, a unit is exposed to "something more" than Z_i. It is exposed to a "sum effect" from the entire population treatment, Z.

There are more unit outcomes than just $y_i(0), y_i(1)$.

In principle, there is a different potential outcome, $y_i(Z)$, for a different population treatment Z.

We need a new formalism to describe these settings.

Recent literature has converged to the concept of treatment exposure.

Treatment exposures

Under interference, every unit is exposed to "something more" than Z_i.

A popular convention is to call this a treatment exposure, $f_i(\mathbf{Z}) \in \mathbb{F}$.
(Hong and Raudenbush, 2006); (T. and Kao, 2013), (Aronow and Samii, 2017)

Here, $\mathbb{F}$ denotes the set of all possible treatment exposure levels.

Related: $f_i(Z)$ is the "effective treatment" (Manski, 2009), exclusion restriction in econometrics, etc.

Examples of treatment exposure:

- $f_i(\mathbf{z}) = z_i$. Standard setting. No interference.
- $f_i(\mathbf{z}) = z_i + \gamma \sum_{j \in \text{group}_i} z_j$. Clustered design.
- $f_i(\mathbf{z}) = z_i + \gamma \sum_{j \in \text{group}_i} z_j / |\text{group}_i|$. Saturation design.
- $f_i(\mathbf{z}) = (z_i, \mathbf{z}_{\text{group}_i})$. Multivalued exposure.

Group could be a $\{class, dorm, firm, household, city\}$ etc.

Could I just fit a regression?

Indeed, a popular approach is to fit:

$$Y_i = \alpha + \beta Z_i + \gamma \underbrace{f_i(\mathrm{Z})}_{\text{exposure}} + \delta' X_i + \epsilon_i.$$

The coefficient γ corresponds to the "spillover effect".

- As before, model specification is crucial.
- $f_i(\mathrm{Z})$ usually has a complex correlation structure with other covariates and Z, and possibly an underlying network.
- Asymptotics on $\hat{\gamma}$ may well be intractable.
- Cannot accurately quantify uncertainty, in general.
  (cf. simple linear example in the introduction)

Finally, it is not uncommon to use a model with Ys on the "left and right" of the regression. This is almost never a good idea. (Angrist, 2019)

Using exposures — Hypotheses for spillovers

In many settings, we want to test whether the exposures in a set $\mathbb{F}_0$ are equivalent (in terms of their outcomes).

In other words, all assignments producing an exposure in F_0 are equivalent in terms of their outcomes.

This may be expressed as:

$$H_0 : y_i(\mathrm{z}) = y_i(\mathrm{z}') \text{ for all } i, \mathrm{z}, \mathrm{z}' \text{ st } f_i(\mathrm{z}), f_i(\mathrm{z}') \in \mathbb{F}_0.$$

(Manski, 2009), (Aronow, 2012), (T. and Kao, 2013), (Bowers et al., 2013), (Athey et al., 2019), Basse et al, 2019), (Puelz et al, 2021).

When $\mathbb{F}_0 = \mathbb{F}$ then the problem reduces to the classical FRT.

If $\mathbb{F}_0 \subset \mathbb{F}$ we run into (many) problems. (null is not sharp, it's "weak")

A crime spillover hypothesis

In the Medellin example, the policy experts were interested in spillover effects on untreated units in a radius of 125m.

We can define the following exposure:

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

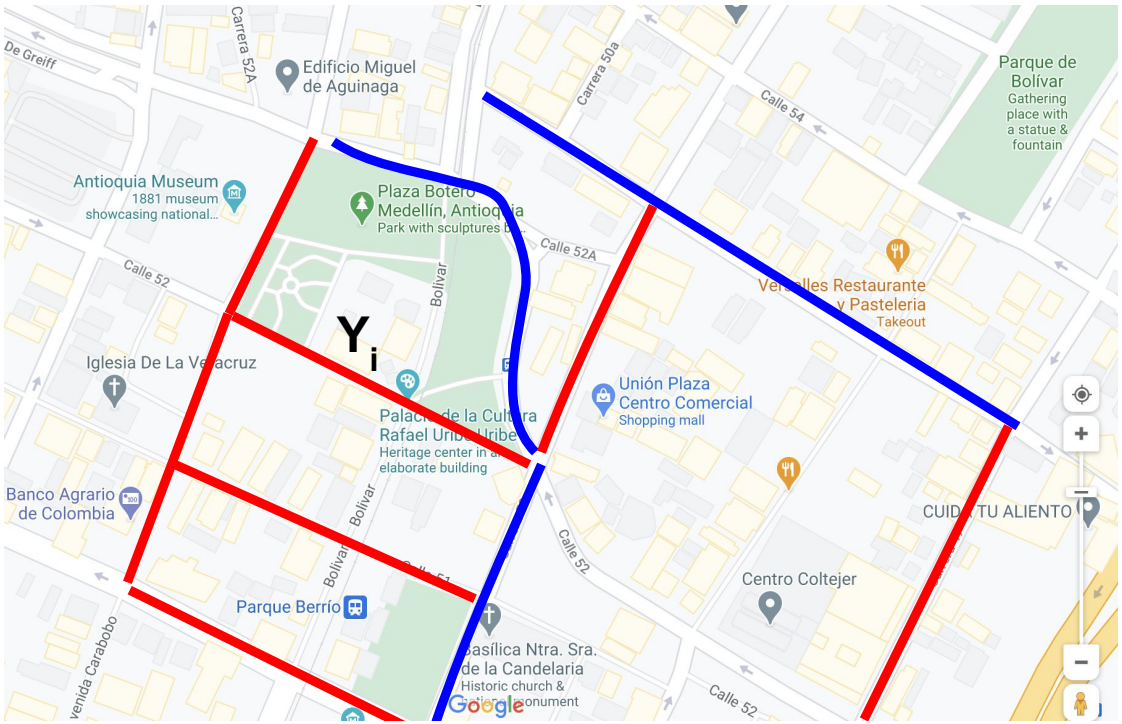
where $\text{dist}_i = \min_{j \neq i : z_j = 1} d(j, i)$ = distance to closest treated street.

This implies the null:

$$H_0 : \; Y_i(z) = Y_i(z') \text{ for every } i, z, z',$$
$$\text{s.t.} \;\; f_i(z), f_i(z') \in \{\texttt{short}, \texttt{control}\}.$$

Illustration from Medellin

Crime spillovers from nearby treated streets on control streets?



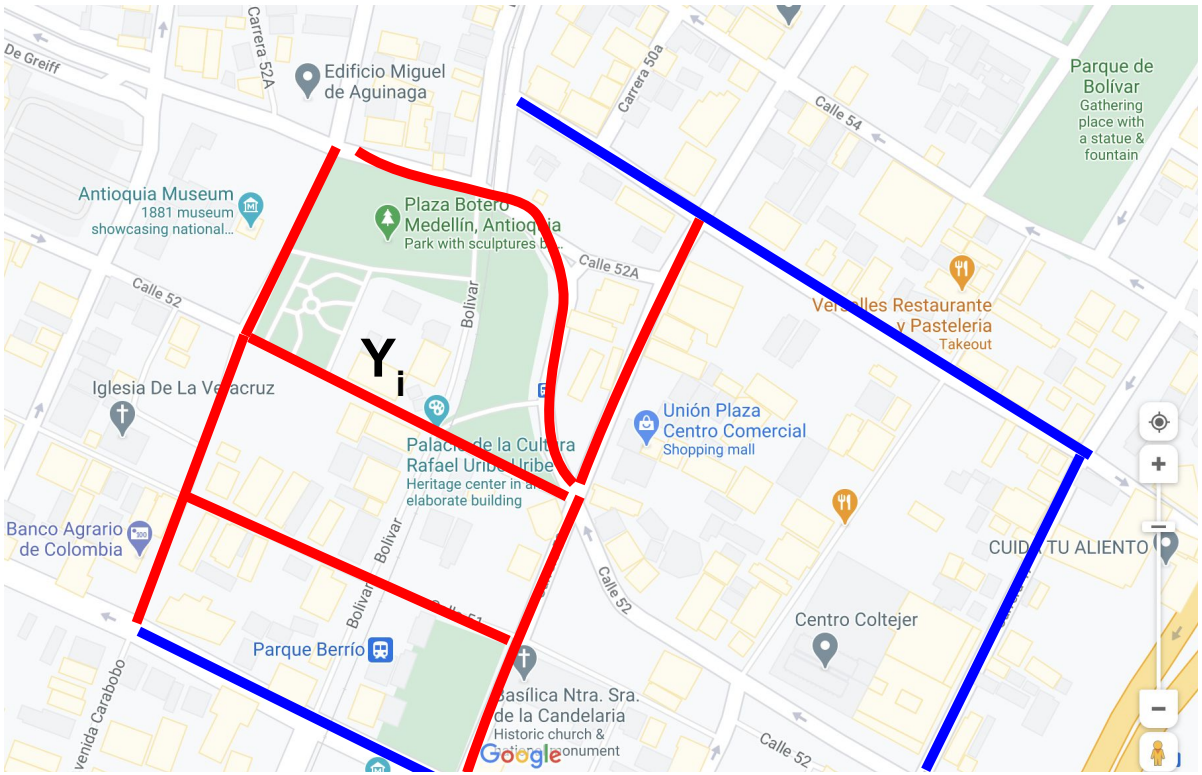
Here, $\mathbb{F}_0 = \{\texttt{short}, \texttt{control}\}$.

Treated streets (blue) are already excluded from the analysis.

Consider a control street i, and suppose $Y_i = 3.5$ ("agg crime score").
This means that $y_i(\texttt{short}) = 3.5$.

FRT problems under interference

Consider a counterfactual Z^* depicted below:

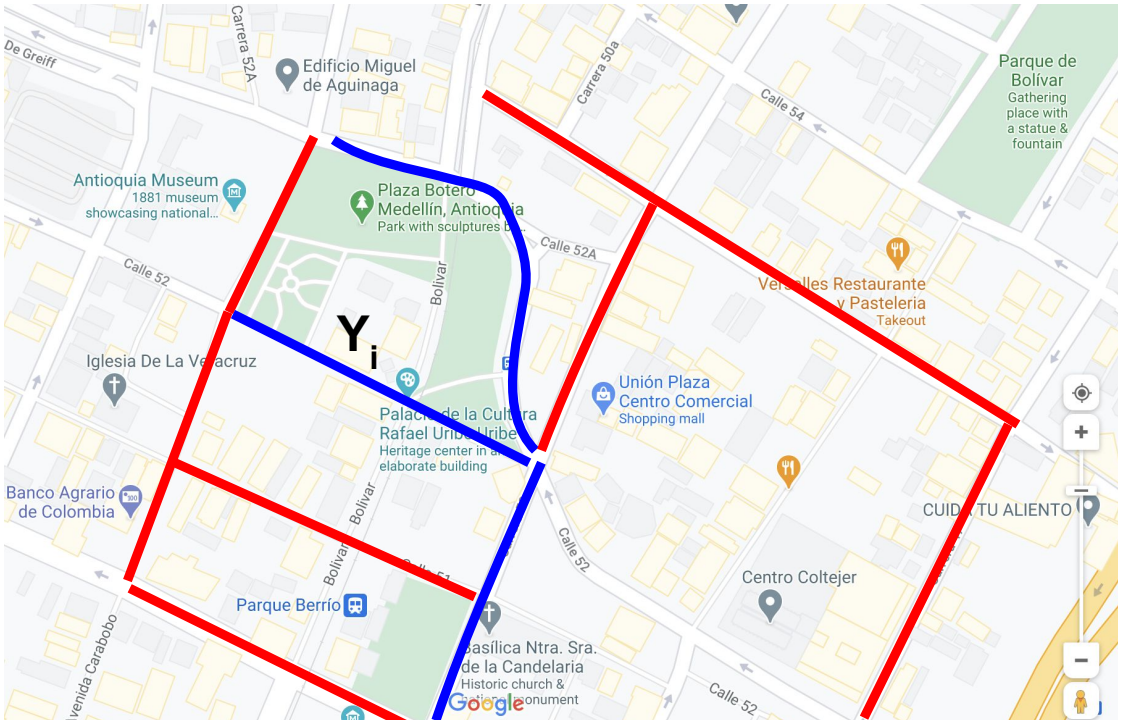


Unit i is exposed to `control` under Z^*, i.e., $Y_i^* = y_i(\texttt{control})$.

Outcome Y_i^* was not actually observed but can be imputed under the null. (since $y_i(\texttt{control}) = y_i(\texttt{short}) = 3.5$ under H_0.)

FRT problems under interference

Consider another counterfactual Z^* as depicted below:



Now i is treated, and its exposure is not in $\mathbb{F}_0$. Thus, $y_i(z^*)$ cannot be imputed under H_0.

So, either we need to drop unit i or drop assignment z^* from the analysis.

Main insight of recent literature: We have to condition on a subset of units/assignments where imputation is possible —"focal units" in (Athey et al, 2019).

Generalization: Conditional FRT

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

Earlier work (Aronow, 2012); (Athey et al, 2019) advocated for choosing U randomly from $\mathbb{U}$ to ensure validity. However, this may lose (a lot of) power, and is also unnecessary.

Generalization: Conditional FRT

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

Earlier work (Aronow, 2012); (Athey et al, 2019) advocated for choosing U randomly from $\mathbb{U}$ to ensure validity. However, this may lose (a lot of) power, and is also unnecessary.

(Basse, Feller, T., 2019) argued that conditioning can be probabilistic, and generally described through a conditioning mechanism, $P(C|Z)$.

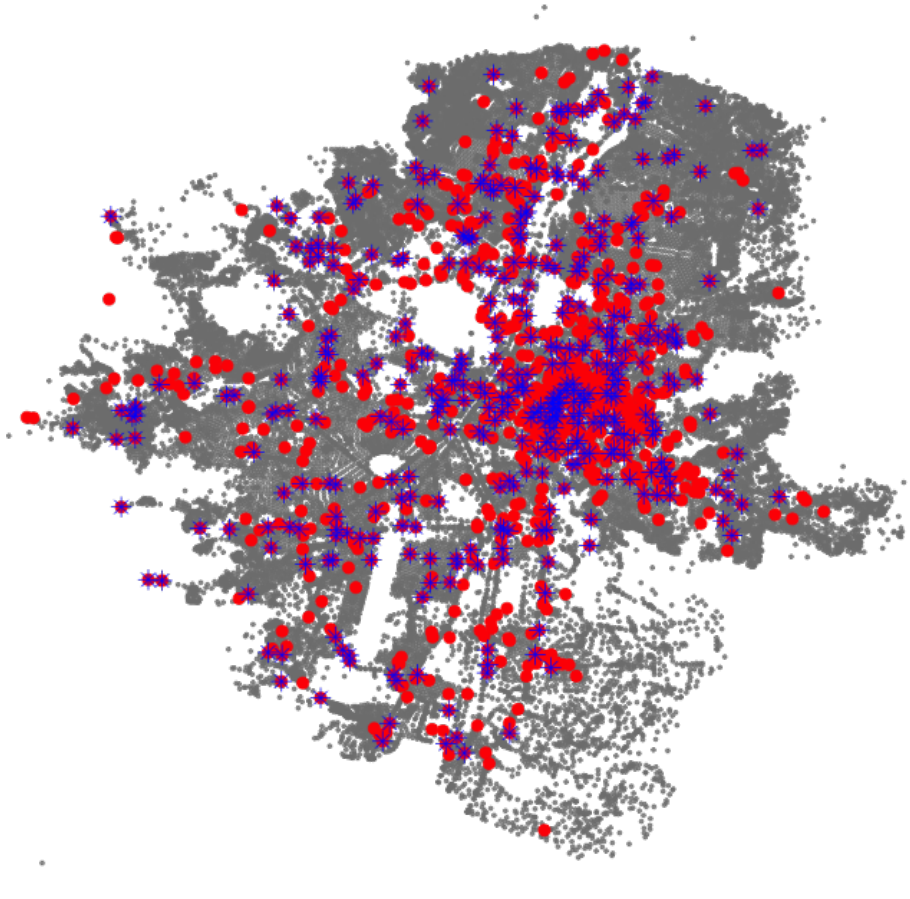
Theorem (Basse, Feller, T., 2019)

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.(null is sharp in C)
2. The resampling distribution is:

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

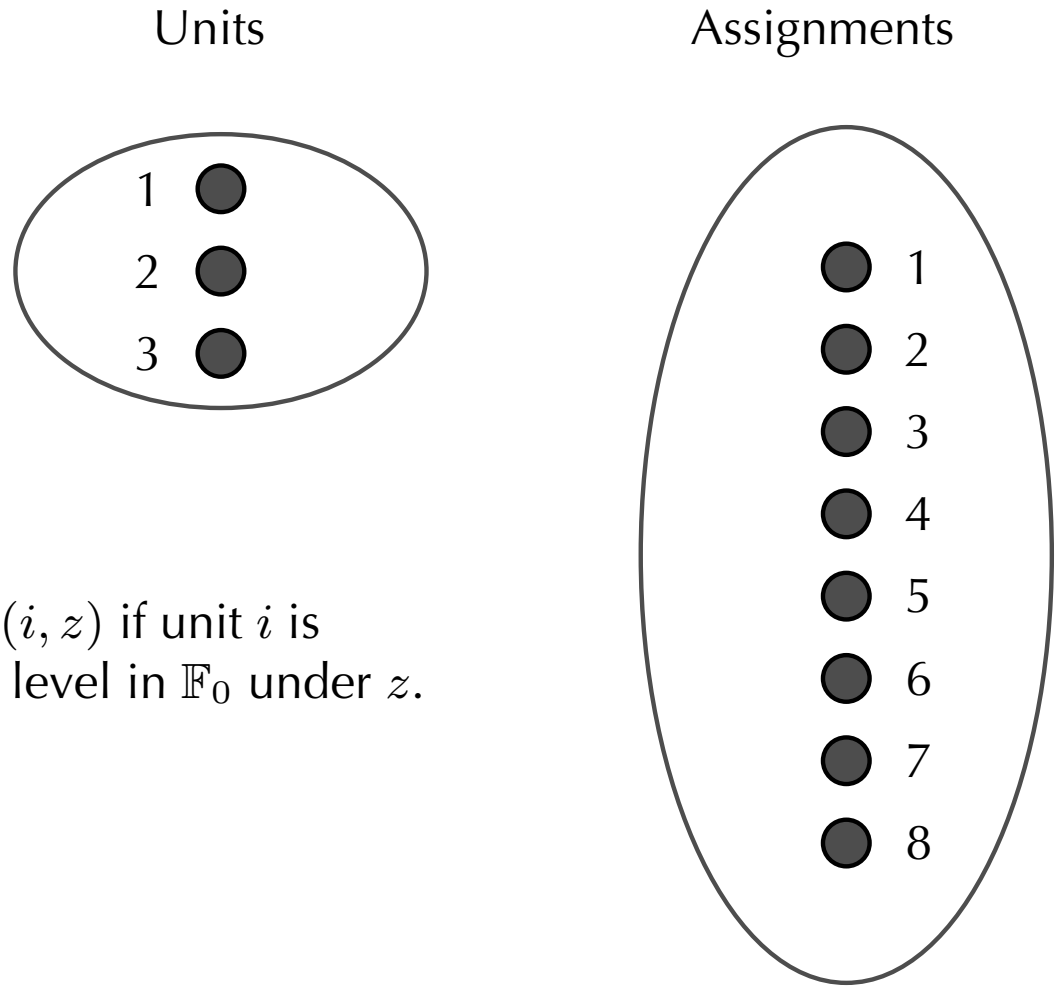
Medellin application: What's a good conditioning mechanism?



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

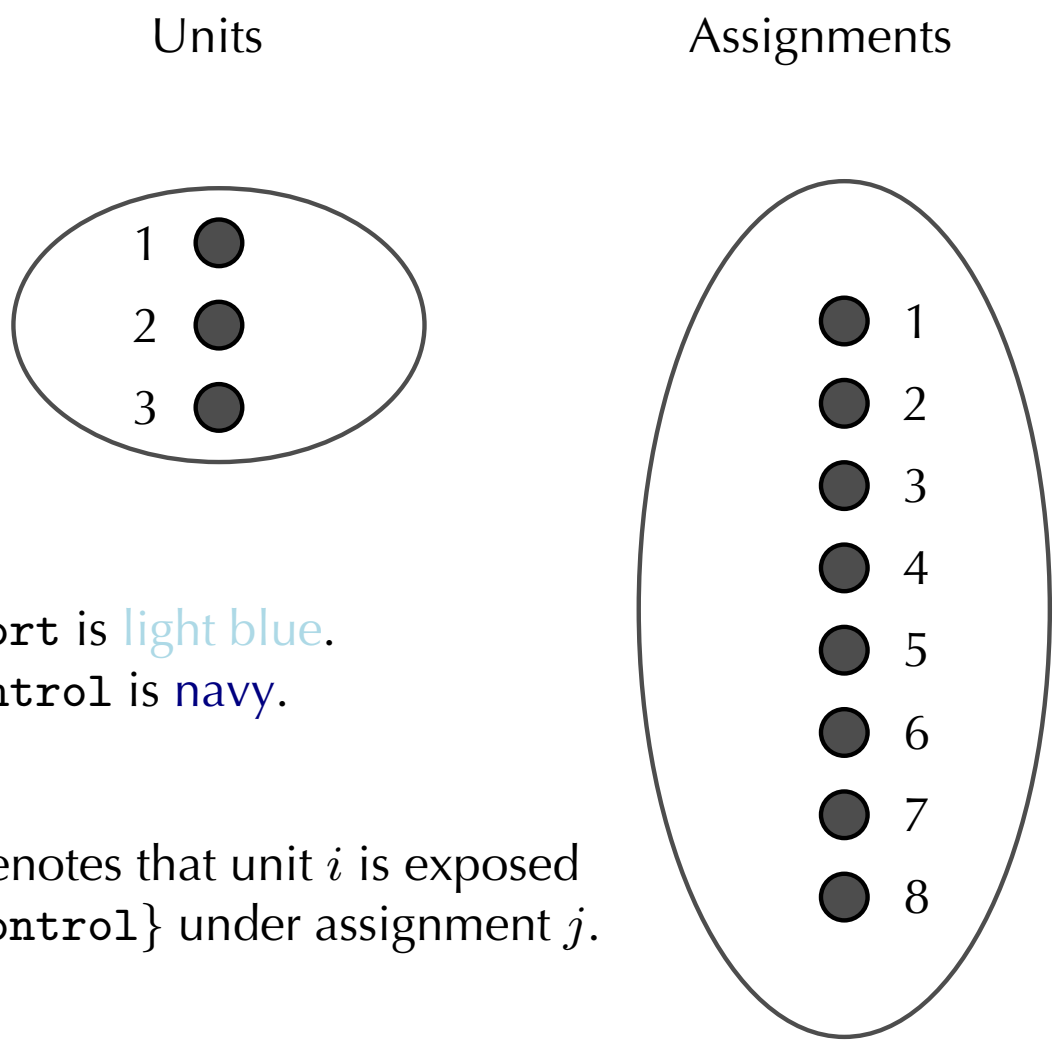
A key conceptual contribution is the null exposure graph (Puelz, Basse, Feller, T., 2021).

The null exposure graph



We connect (i, z) if unit i is exposed to a level in $\mathbb{F}_0$ under z.

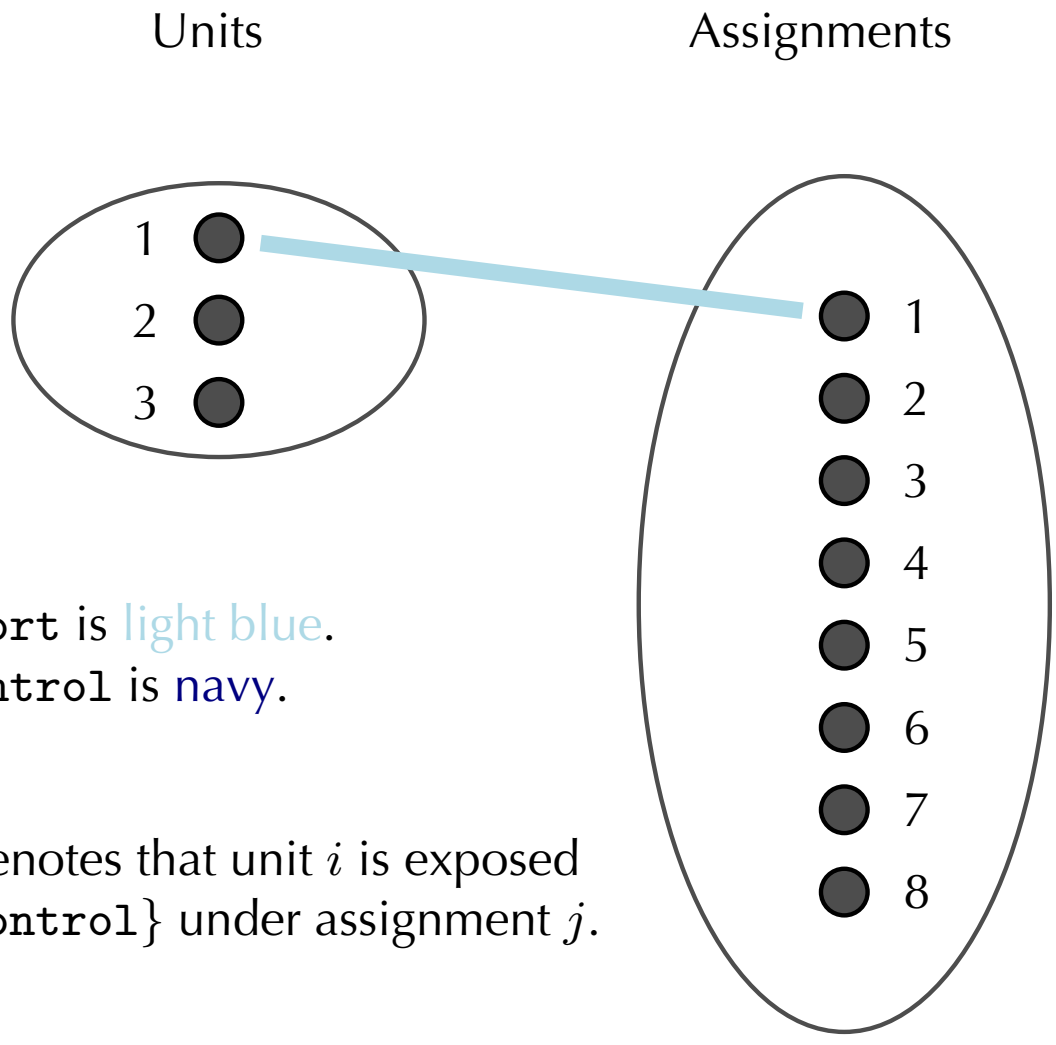
The null exposure graph



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

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

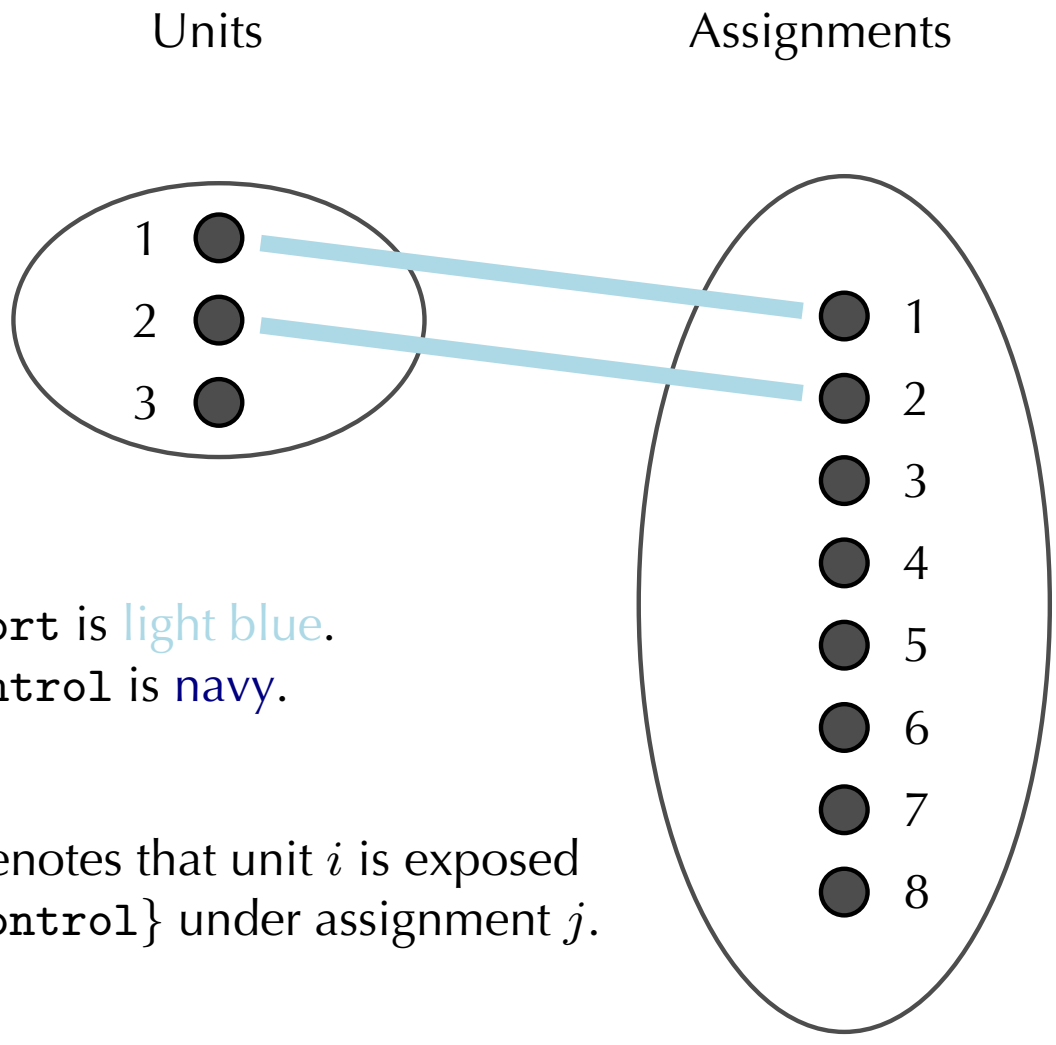
The null exposure graph



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

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

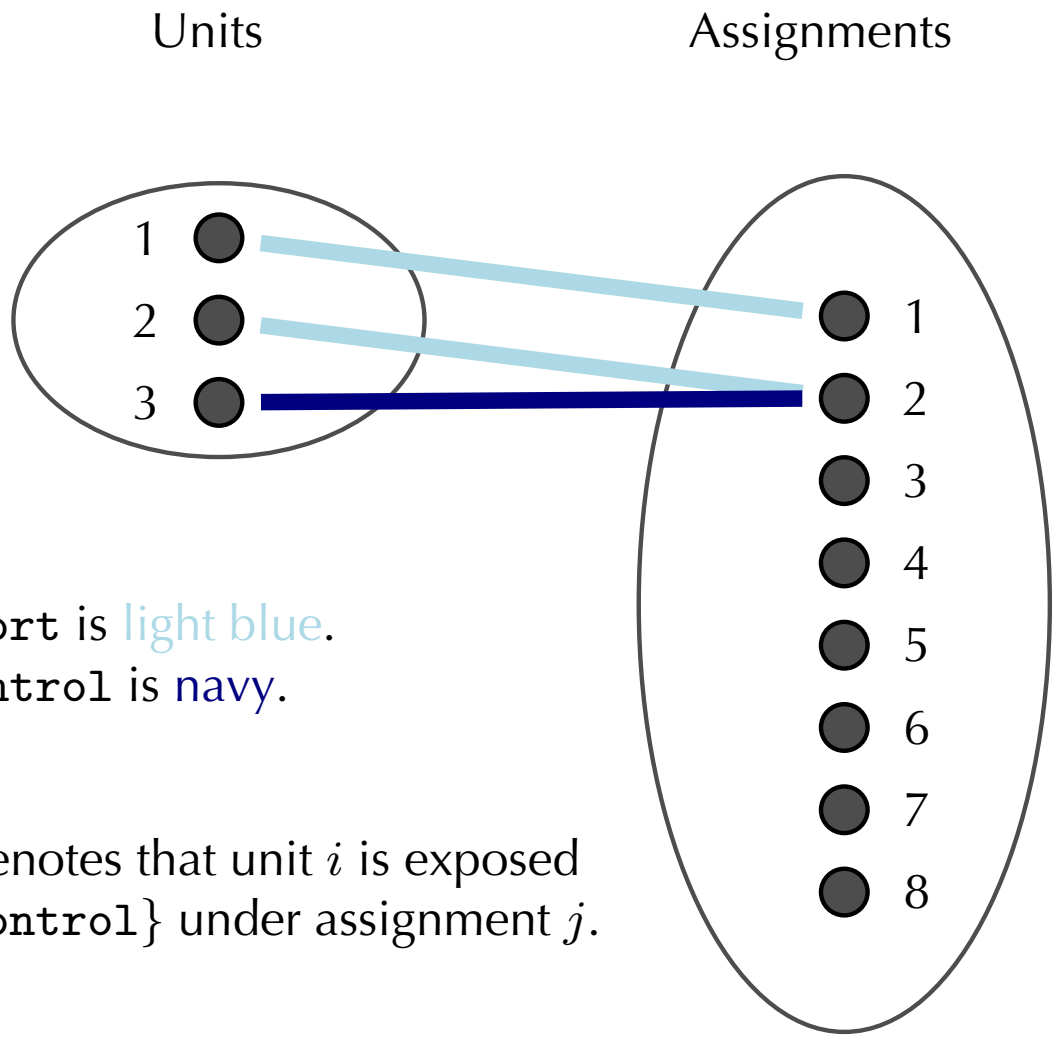
The null exposure graph



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

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

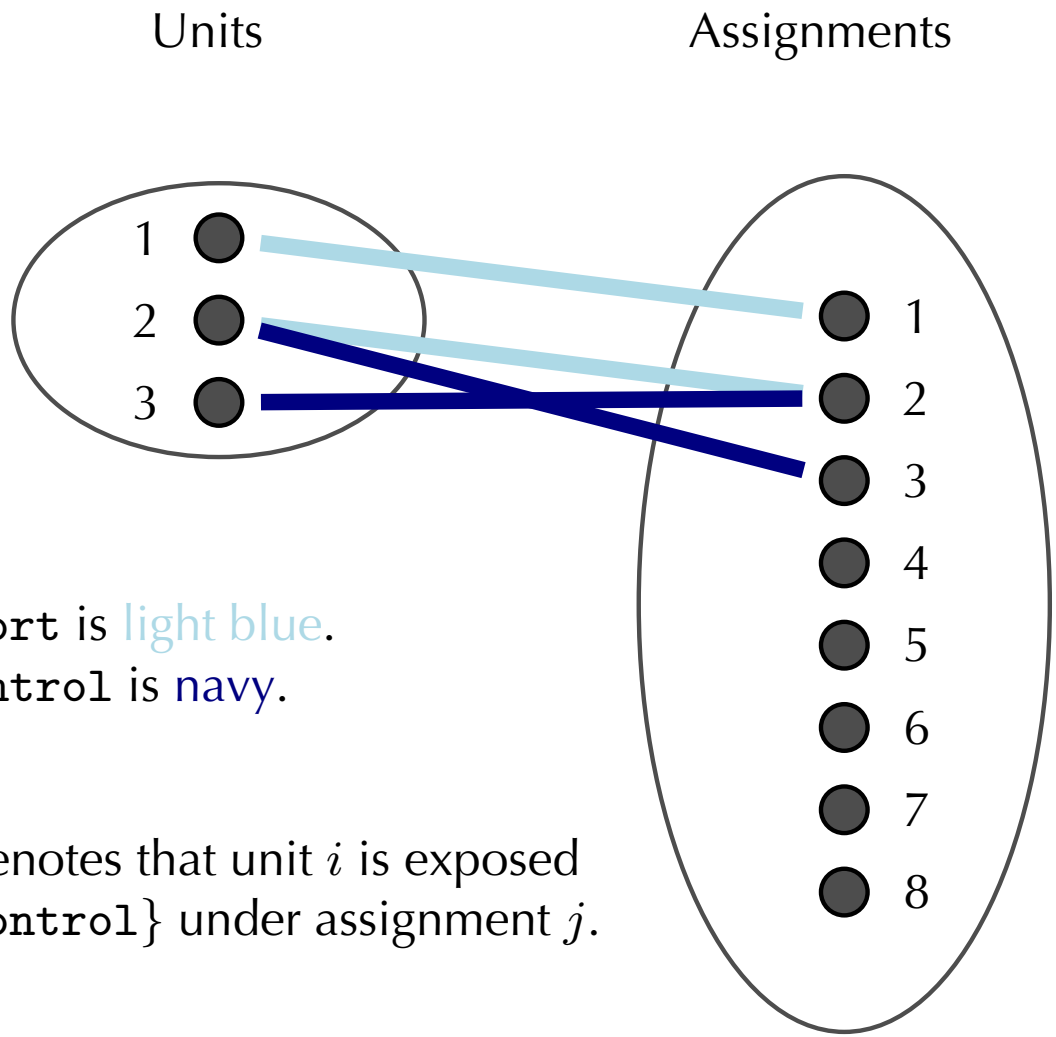
The null exposure graph



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

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

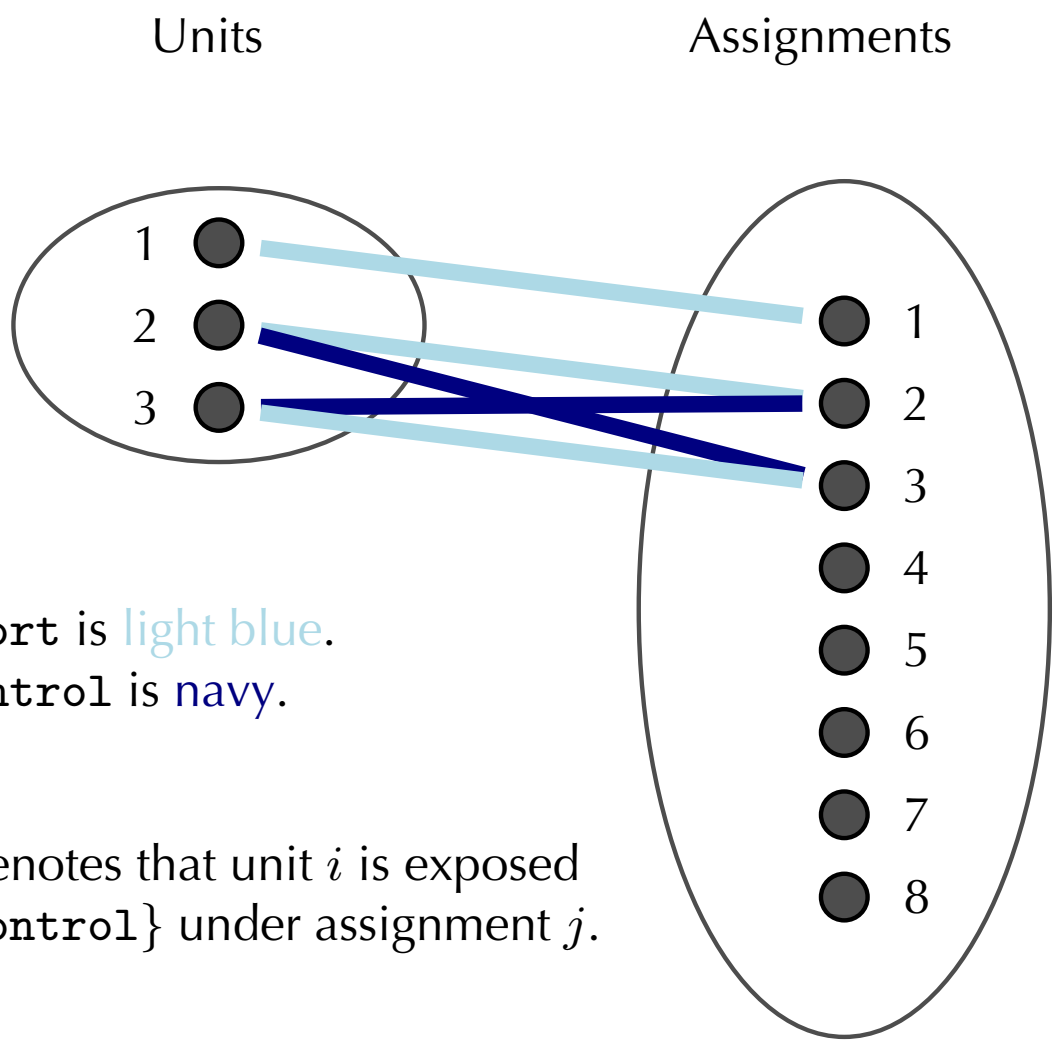
The null exposure graph



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

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

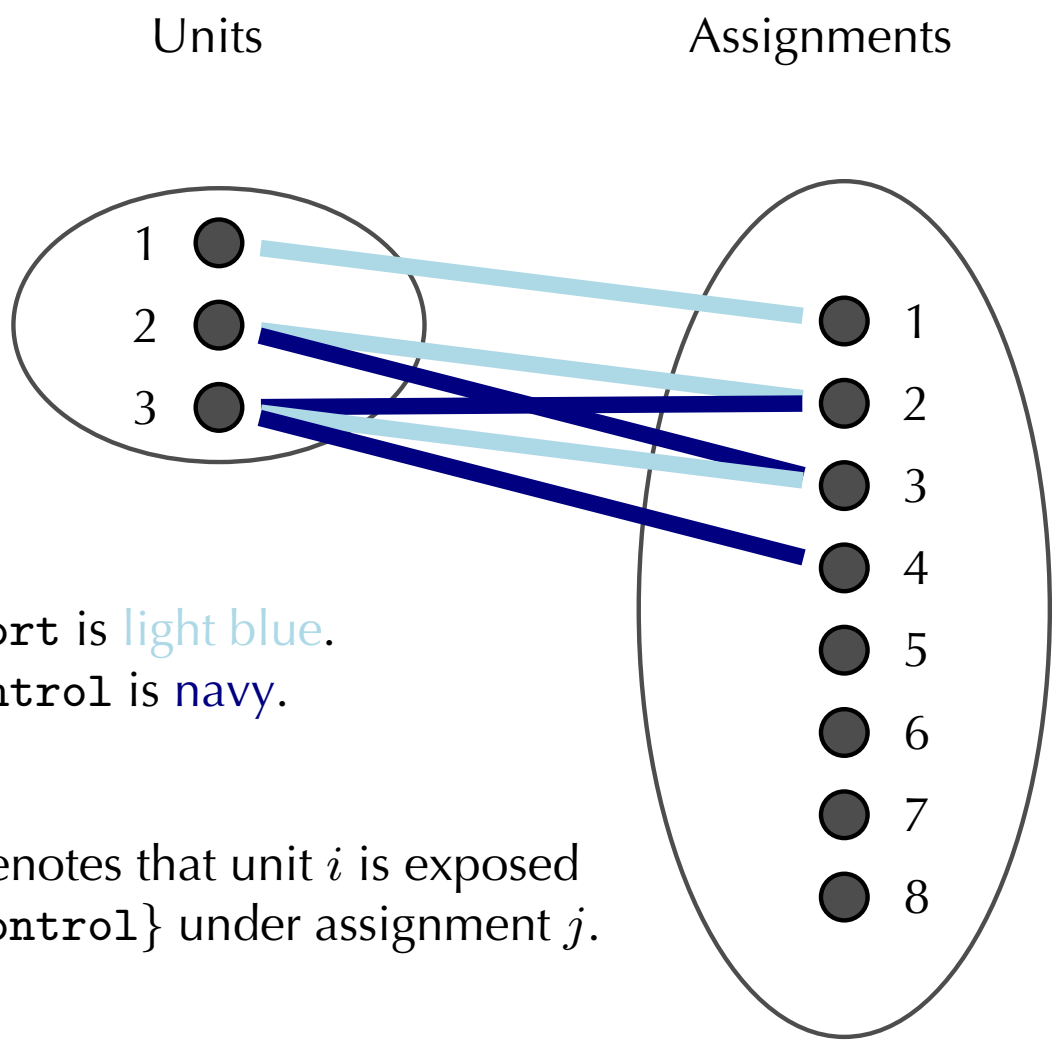
The null exposure graph



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

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

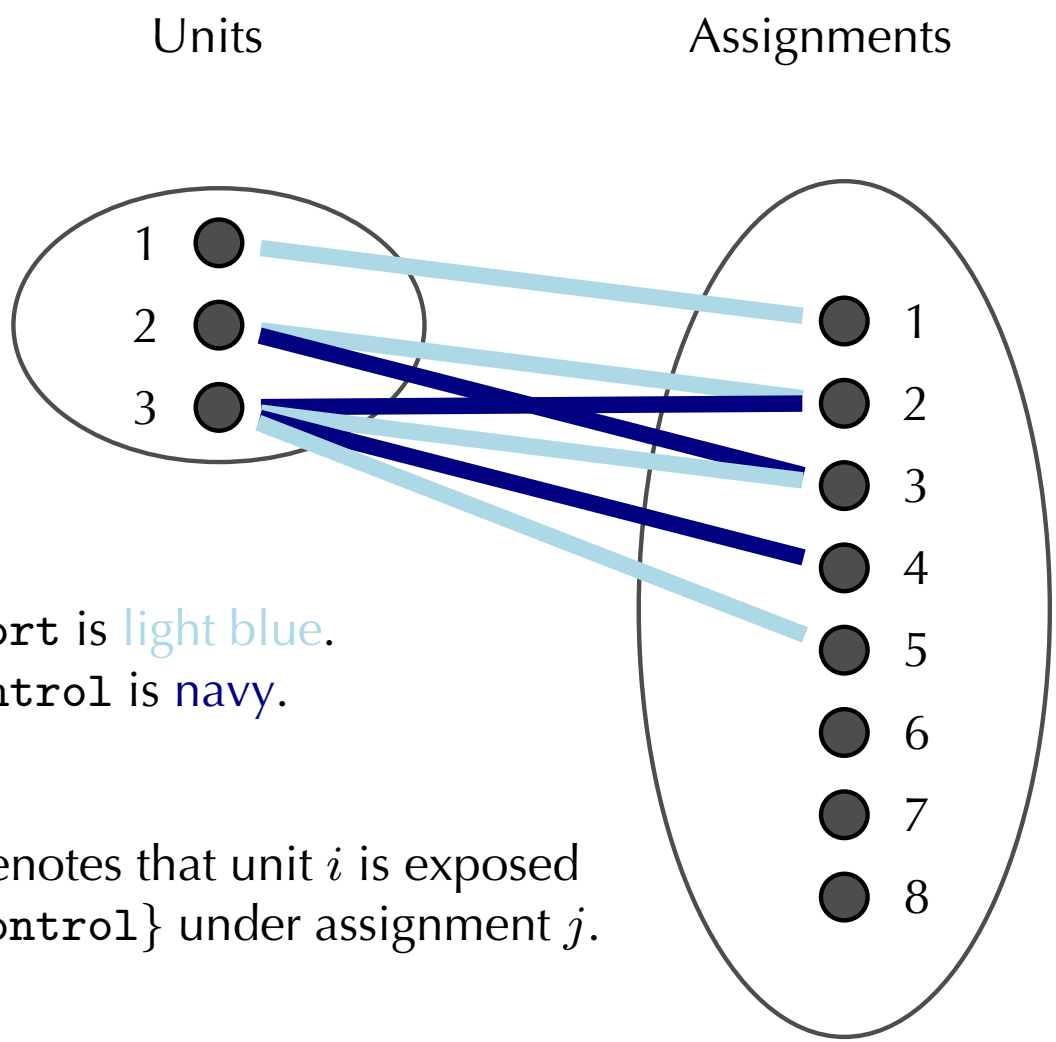
The null exposure graph



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

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

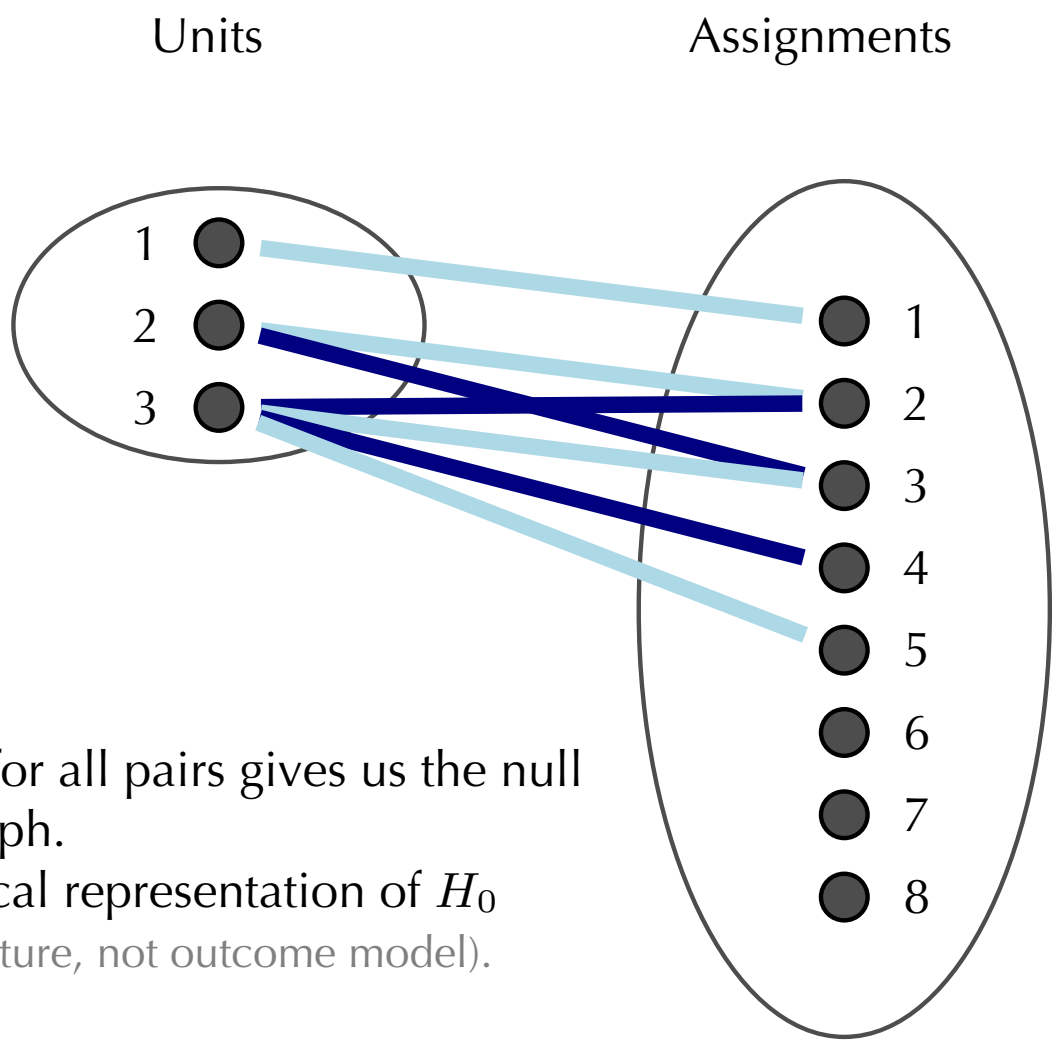
The null exposure graph



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

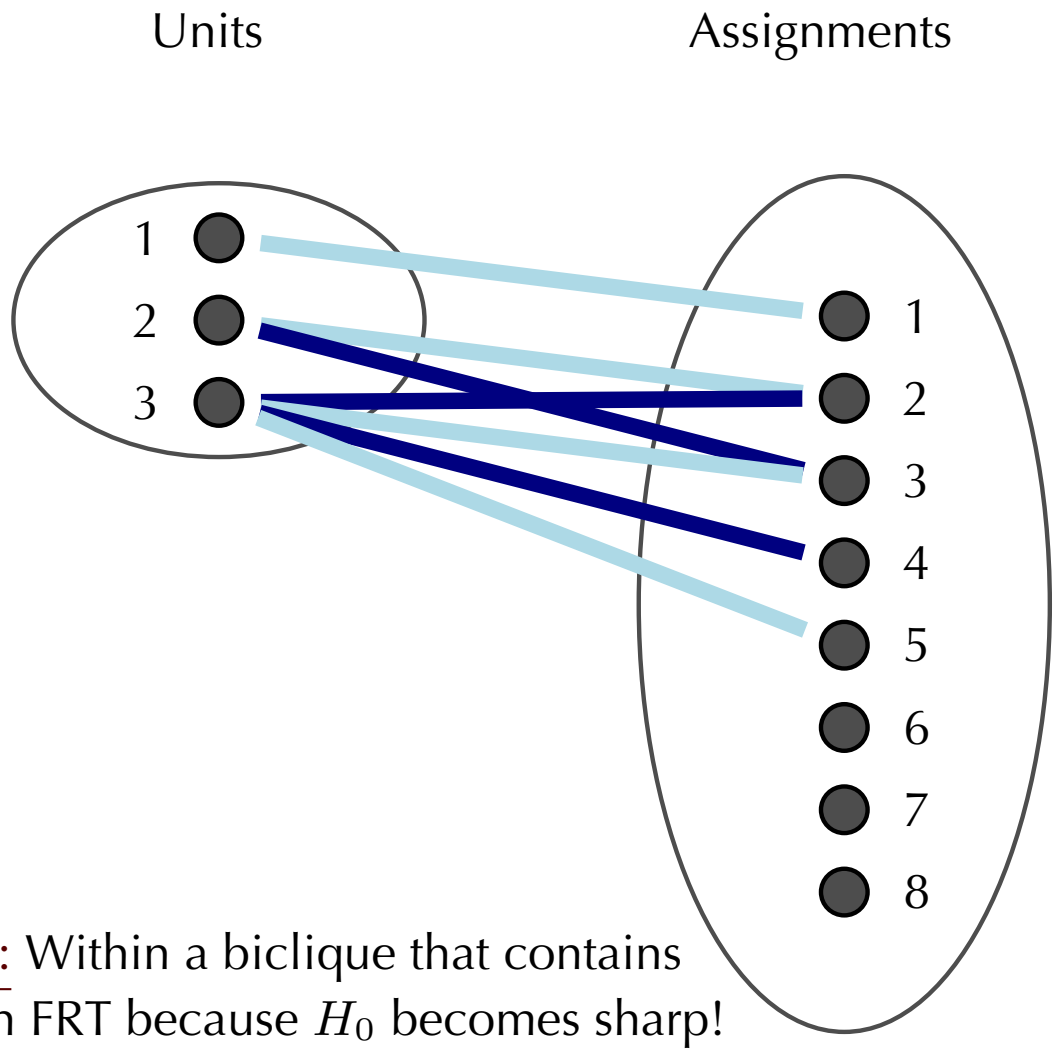
edge (i, j) denotes that unit i is exposed to $\{\texttt{short}, \texttt{control}\}$ under assignment j.

The null exposure graph



Completing for all pairs gives us the null exposure graph.
It is a graphical representation of H_0
(problem structure, not outcome model).

The null exposure graph



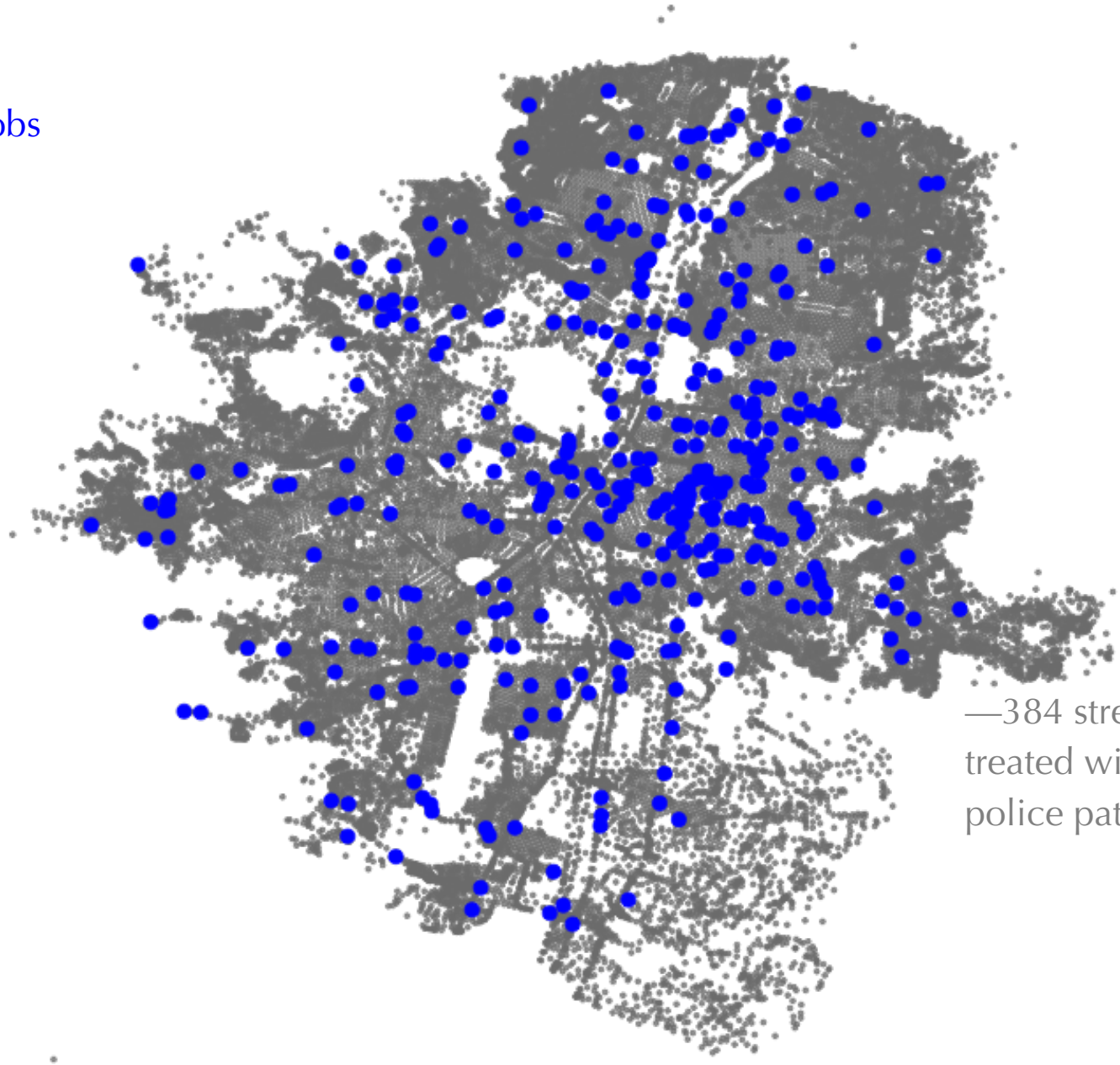
Key property: Within a biclique that contains Z we can run FRT because H_0 becomes sharp!

(all Y^* can be imputed)

Returning to the map

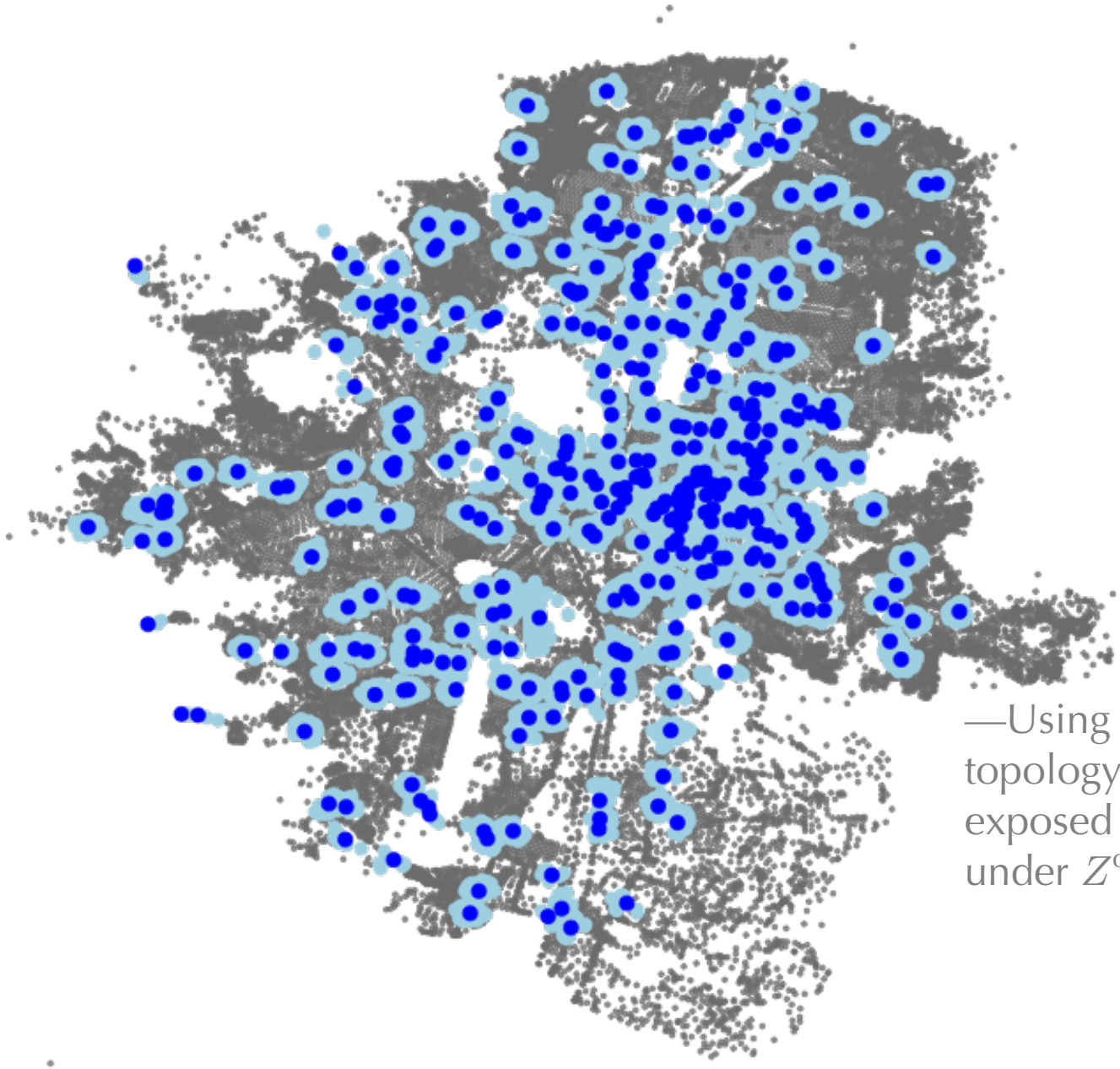
The observed assignment

Z_{obs}



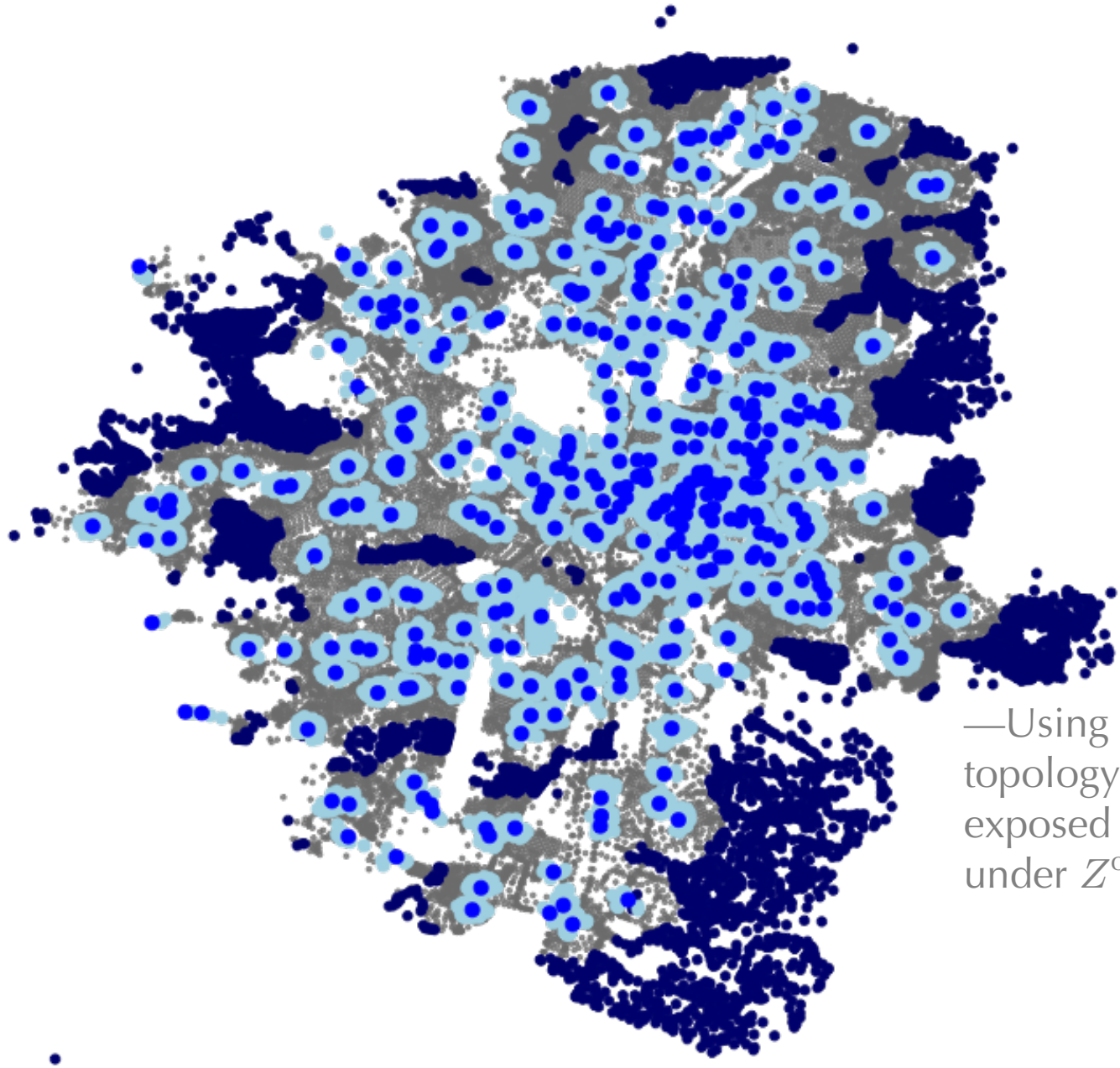
—384 streets are treated with increased police patrolling

Short-range spillover units (`short`)



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

Pure control units (`control`)

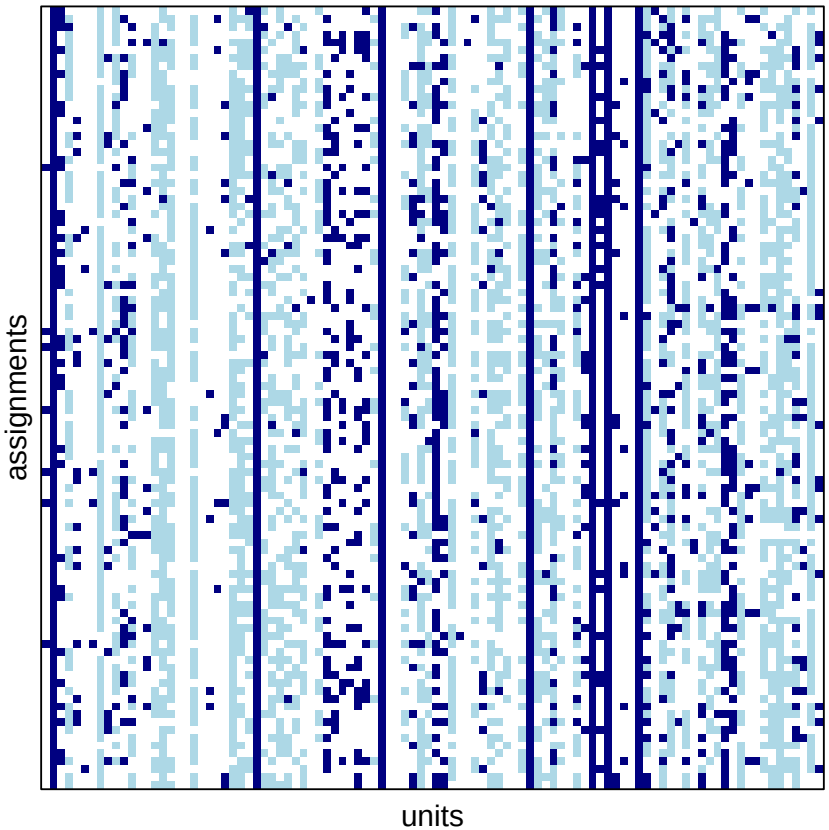


—Using network topology, color units exposed to `control` under Z^{obs}

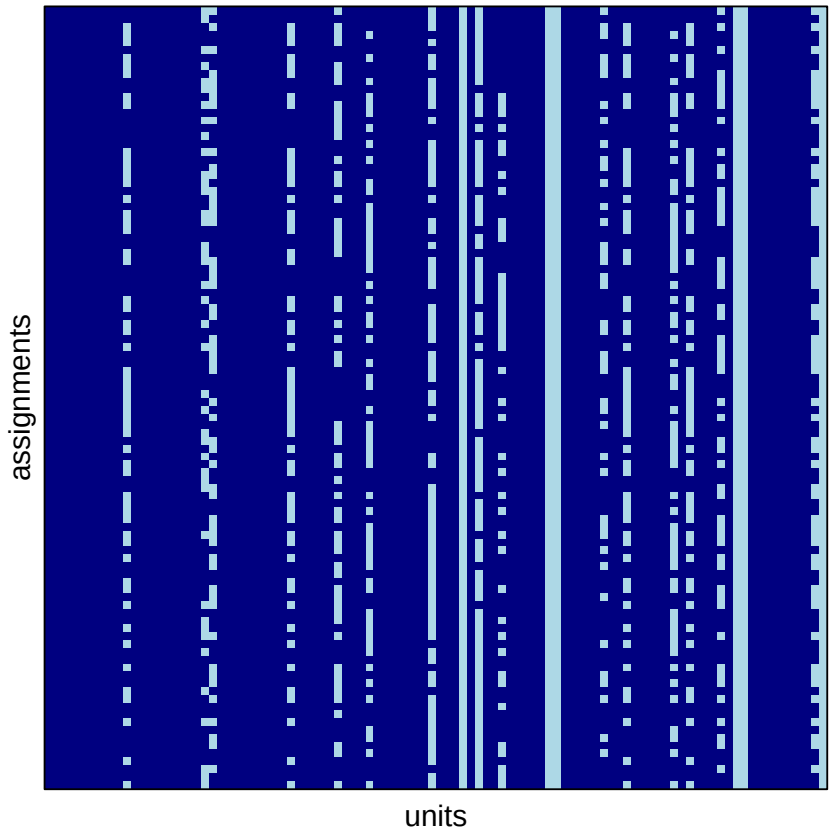
We can remake these pictures for every assignment Z drawn from design $P(Z)$ ...

—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 the observed Z.
  (Greedily conditioning on the largest biclique of NE is invalid!)
- The size of the conditioning biclique also affects the power of the randomization test. Generally speaking, bigger clique is better.

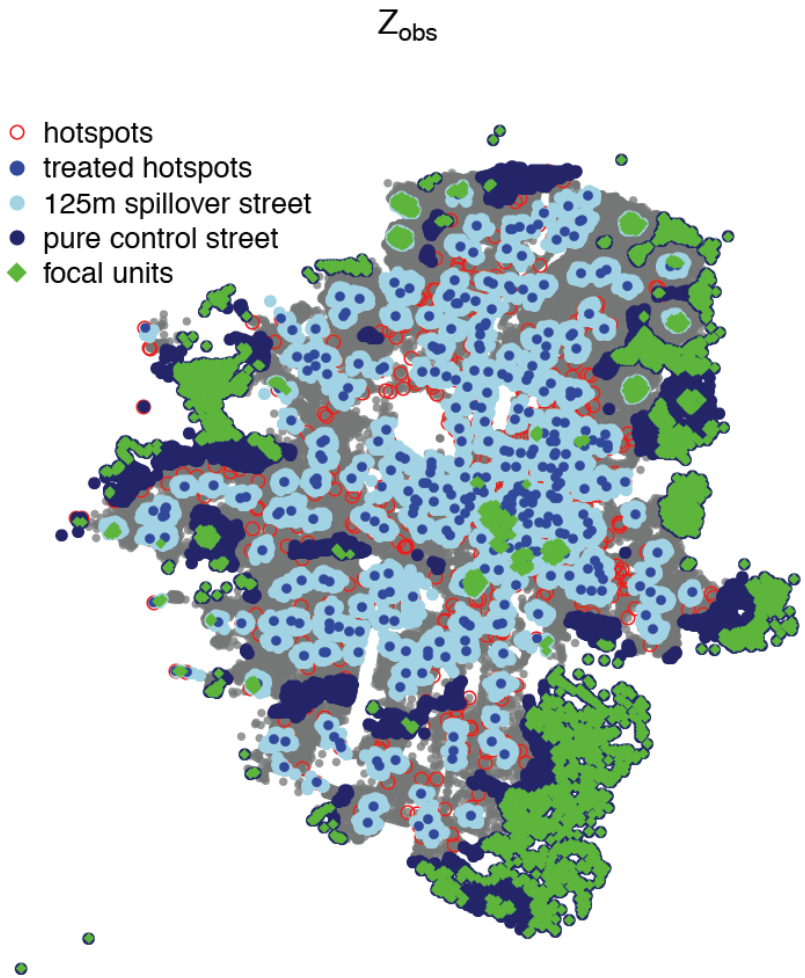
Spatial interference: Medellin data

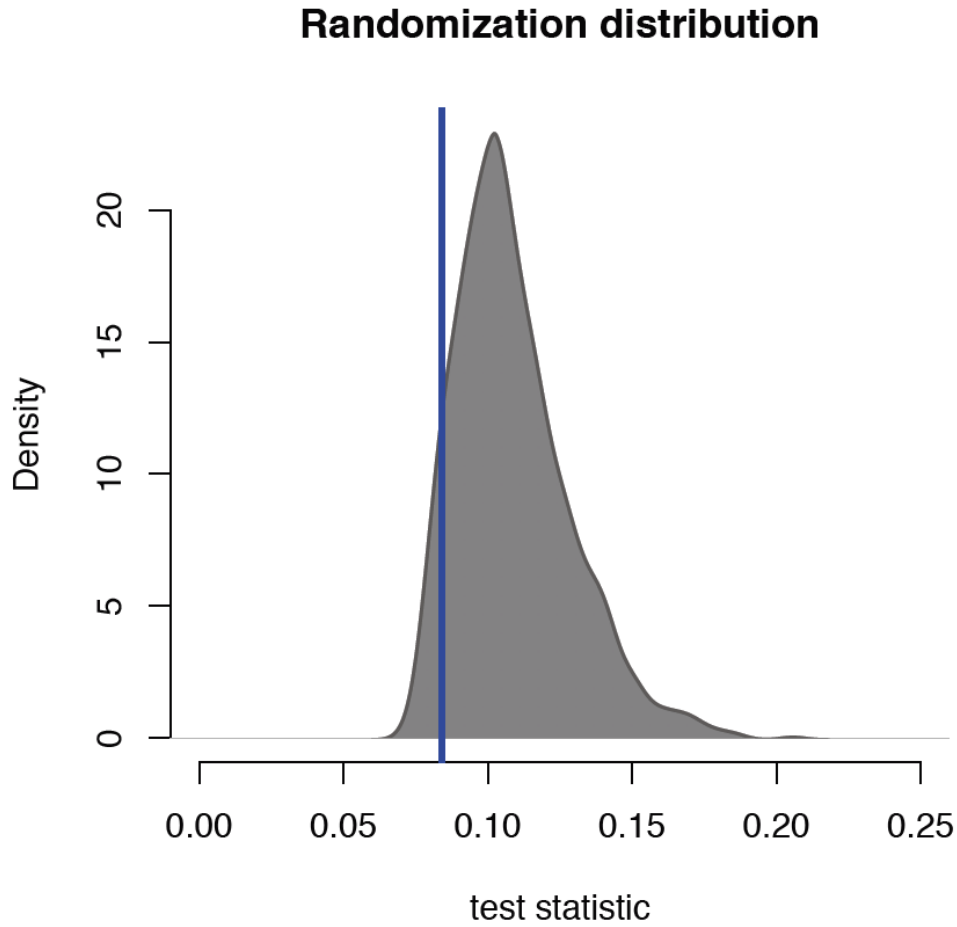
Statistics of the null-exposure graph:

- #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

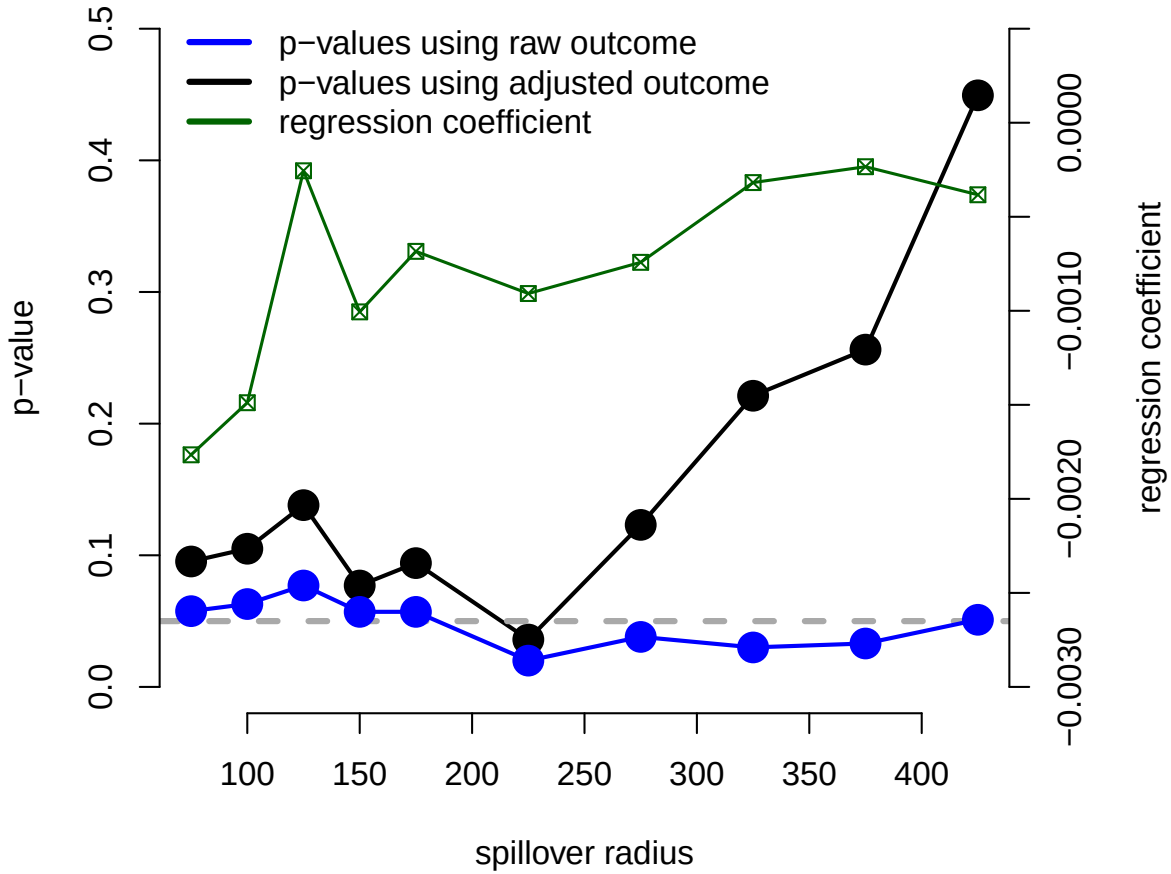


Figure: P-values for clique tests with varying spillover radius.

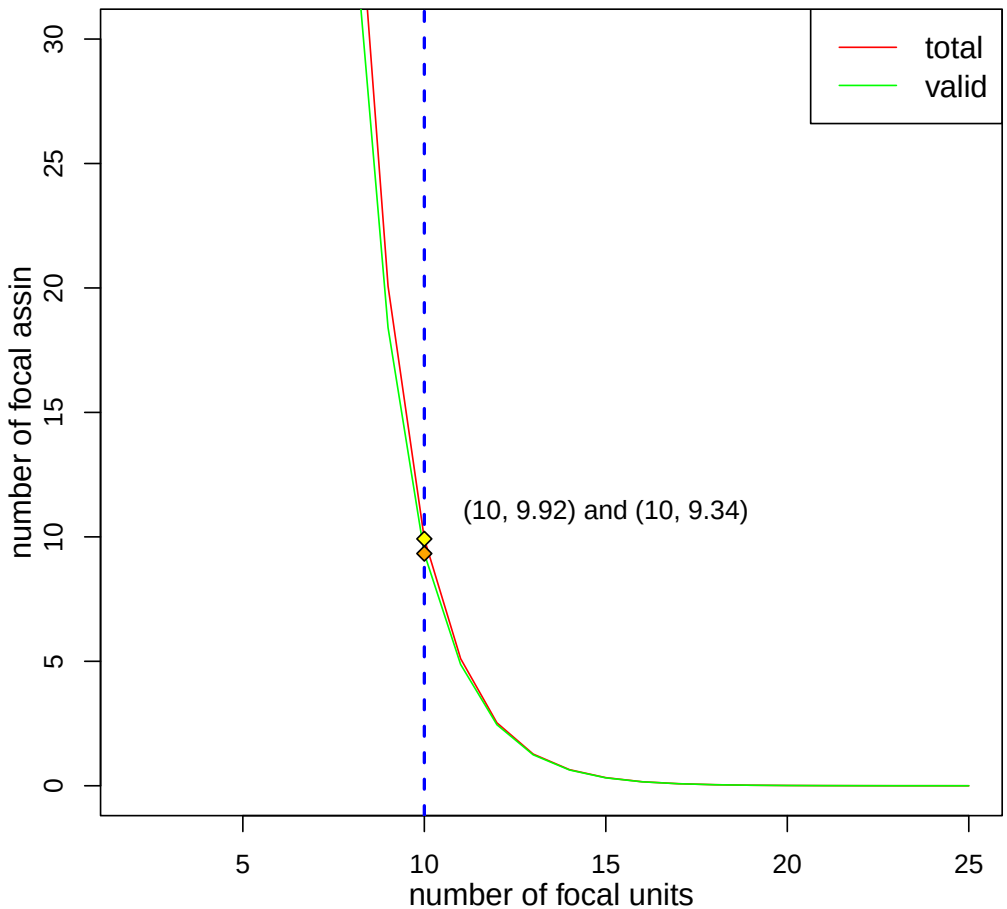
Spin-off 1: Diagnostic

This gives us an idea to "warn" the user when H_0 is hard to test.

Spin-off 1: Diagnostic

This gives us an idea to "warn" the user when H_0 is hard to test.

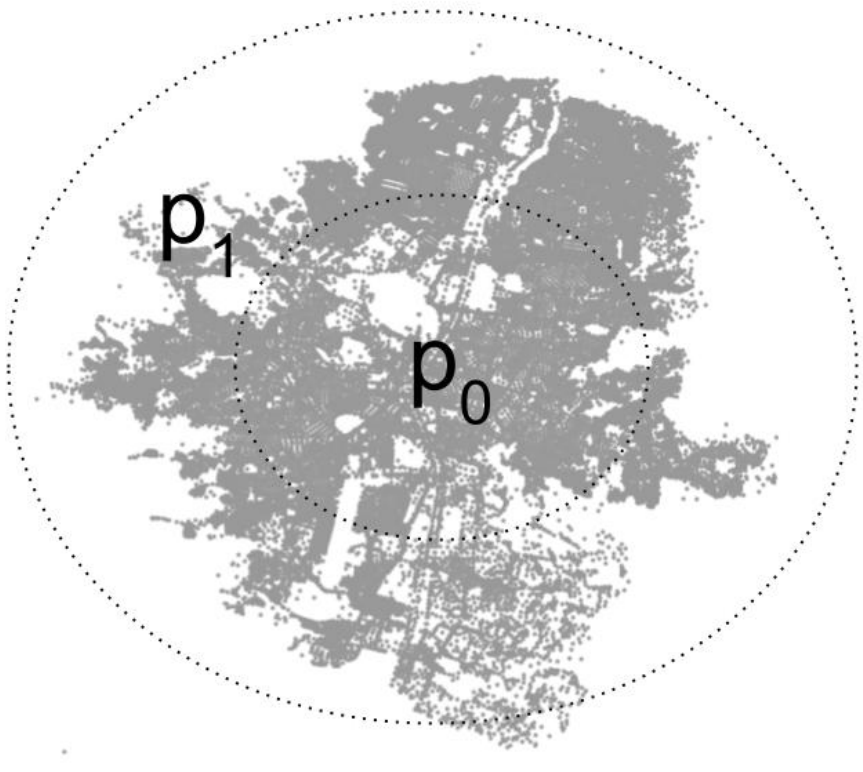
Example: "Effects of a large-scale social media advertising campaign on holiday travel and COVID-19 infections: a cluster randomized controlled trial" (Breza et al, 2021)



Spin-off 2: Improving the experimental design

We could use the NE graph to optimize the experimental design for a specific null hypothesis, H_0.

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

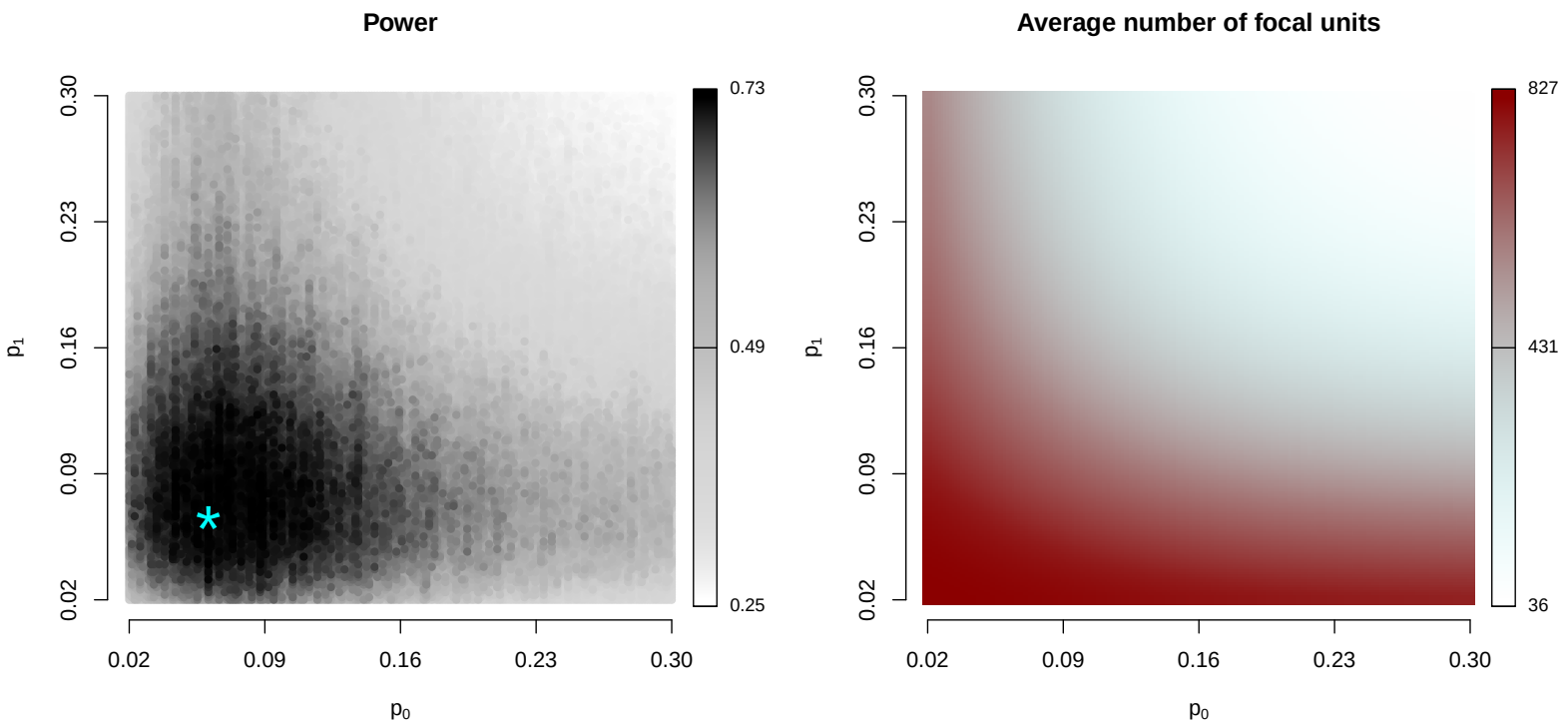


Spin-off 2: Improving the experimental design

Left: Power calculated under a simulated model for Y over the design space. (darker=higher power).

Right: Average clique sizes in NE graph over the design space.

This shows that we should favor designs that lead to larger bicliques.



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, biclique size).

A bigger agenda

The randomization method can be applied beyond experimental settings.

The fundamental idea is to rely on an invariance of the data: $Y \stackrel{d}{=} gY$.
e.g., g= permutations when data are exchangeable.

I have been exploring a new framework of inference based on such data invariances — invariant inference.

For instance, consider the linear model $y = X\beta + \varepsilon$. The goal is to perform inference on β under an assumption

$$\varepsilon \stackrel{d}{=} g\varepsilon,$$

without assuming i.i.d. data or assuming a distribution on ε.

This can be accomplished via a residual randomization method. This method unifies inference in a coherent way and under many complex error structures (one-way clustered errors, two-way clustered, etc.)

Thank you!

(Basse, Feller, T., 2019) “Randomization Tests for Causal Effects under General Interference”, Biometrika

(Puelz, Basse, Feller, T., 2021) “A Graph-Theoretic Approach to Randomization Tests of Causal Effects Under General Interference”, JRSS-B

(Shaikh and T., 2021) “Randomization Tests under Staggered Adoption of Treatment”, JASA

(Basse, Ding, Feller, T., 2022) “Randomization Tests for Peer Effects in Group Formation Experiments”, R&R

“Invariant Inference under Residual Randomization”, Under review

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,

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

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,

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

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:

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.

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 can be optimized.
  —i.e., different biclique decompositions will have different power properties, but all are valid.

Group formation experiments

Cai and Szeidl (2015) ran an experiment where CEOs of various companies were randomly grouped together. The CEOs from the same group met regularly for a year to exchange information and business ideas.

The goal was to understand whether participation in this group activity had an effect on the participating companies.

In this case, the treatment (Z_i) for unit i is the set of peers that unit i was grouped with.

The authors ran a panel regression to estimate the treatment effects.

Despite the design complexity, an exact randomization test is actually straightforward. For a null hypothesis that there is no effect on a specific type of company from a specific peer group type, the randomization test simply consists of randomizing (permuting) across companies with the same type.