Randomization Tests for Bipartite Experiments

Jizhou Liu — PKU

Sizhu Lu — UC Berkeley

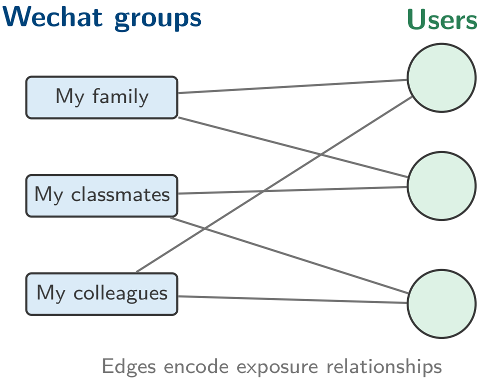
Peng Ding — UC Berkeley

Motivation: experiments are often bipartite

- Treatment is assigned to one population.
- Outcomes are measured on another population.
- The link between them is a known graph.

Examples

Intervention units: Wechat Group
Outcome units: Users

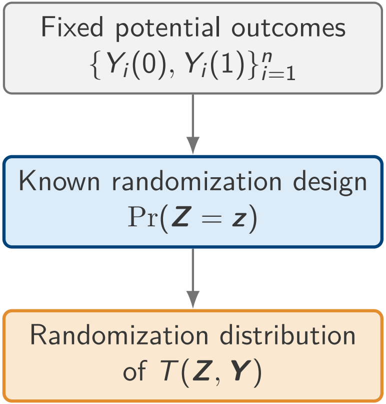


Why randomization inference?

Design-based perspective

Potential outcomes fixed
$\boldsymbol{Z} \sim \Pr(\boldsymbol{Z})$

- The assignment mechanism is known.
- No regression model for outcomes is required.
- The reference distribution comes from the experiment itself.



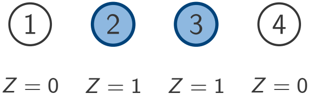
Classical potential outcomes without interference

- Unit i has two potential outcomes: $Y_i(1)$ and $Y_i(0)$.
- Observed outcome:

$$Y_i^{\mathrm{obs}} = Z_i Y_i(1) + (1 - Z_i) Y_i(0).$$

- Fundamental problem: one of the two is missing.

i	Z_i	Y_i^{obs}	Missing
1	0	$Y_1(0)$	$Y_1(1)$
2	1	$Y_2(1)$	$Y_2(0)$
3	1	$Y_3(1)$	$Y_3(0)$
4	0	$Y_4(0)$	$Y_4(1)$



The Fisher sharp null

No individual effect

$$H_0 : \quad Y_i(1) = Y_i(0) \qquad \forall i.$$

Under H_0, all missing potential outcomes can be imputed from the observed outcomes.

i	Z_i	Y_i^{obs}	$Y_i(0)$	$Y_i(1)$
1	0	15	15	15
2	1	20	20	20
3	1	21	21	21
4	0	14	14	14

Sharp null $\Rightarrow$ full schedule known
$\Rightarrow$ counterfactual assignments are usable

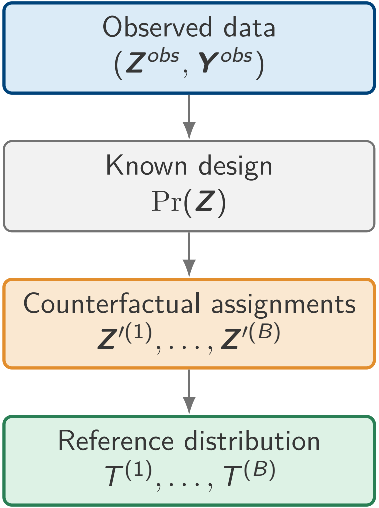
Fisher randomization test: algorithm

1. Choose a statistic, e.g.

$$T(\boldsymbol{Z}, \boldsymbol{Y}) = \bar{Y}_1 - \bar{Y}_0.$$

2. Compute $T^{obs} = T(\boldsymbol{Z}^{obs}, \boldsymbol{Y}^{obs})$.
3. Draw $\boldsymbol{Z}' \sim \Pr(\boldsymbol{Z})$ from the original design.
4. Recompute $T(\boldsymbol{Z}', \boldsymbol{Y}^{obs})$ and compare to T^{obs}.

$$\text{p-value} = \frac{1}{B} \sum_{b=1}^{B} \mathbf{1}\{T^{(b)} \geq T^{obs}\}.$$



Why the Fisher test is exact

$T(\boldsymbol{Z}', \boldsymbol{Y}')$ counterfactual statistic $\xrightarrow{H_0}$ $T(\boldsymbol{Z}', \boldsymbol{Y}^{obs})$ imputed by sharp null $\xrightarrow{\text{same design}}$ $T(\boldsymbol{Z}^{obs}, \boldsymbol{Y}^{obs})$ observed statistic

Key identity

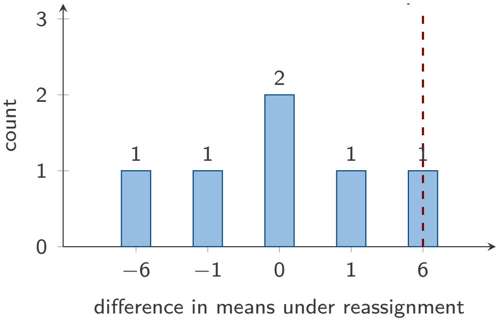
$$T(\boldsymbol{Z}', \boldsymbol{Y}^{obs}) \overset{d}{=} T(\boldsymbol{Z}^{obs}, \boldsymbol{Y}^{obs}) \qquad \text{under } H_0.$$

- Finite-sample validity: no asymptotic normal approximation.
- Model-free: no parametric assumptions on the outcome distribution.

A small randomization distribution

- Four units; two are treated under complete randomization.
- Under the sharp null, outcomes are fixed across reassignments.
- Thus, we can enumerate all six possible assignments.

i	Y_i^{obs}	Z_i^{obs}	label
1	15	0	C
2	20	1	T
3	21	1	T
4	14	0	C



A large observed statistic is assessed relative to the distribution induced by the randomization design.

What breaks under interference?

Without interference:

$$Y_i(Z_i) \in \{Y_i(0), Y_i(1)\}.$$

With interference:

$$Y_i(\boldsymbol{Z}), \qquad \boldsymbol{Z} \in \{0,1\}^m.$$

Potential outcome schedule

0	mixed	mixed	**1**
known	?	?	known

classical FRT needs all entries to be imputable

Total-effect null only links the two extremes

A meaningful null is often not sharp

$$H_0^{total}: \quad Y_i(\mathbf{1}) - Y_i(\mathbf{0}) = 0.$$

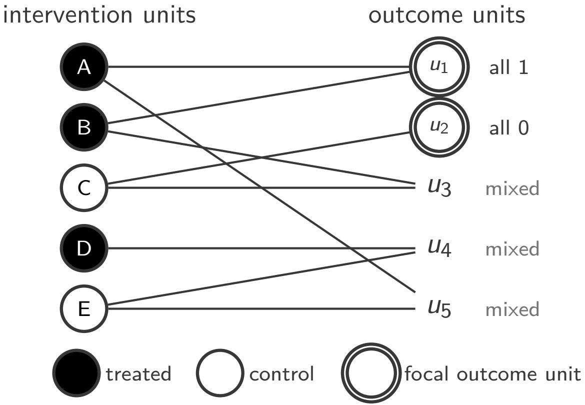
It says nothing about mixed assignments such as $(1,0,1,0)$.

Related literature: conditional randomization tests

- **Testing under interference**
  - Interference detection and focal-unit tests: Aronow (2012)
  - Exact tests for network interference: Athey et al. (2018)
- **Conditional randomization and conditioning mechanisms**
  - General conditioning framework: Basse et al. (2019)
  - Graph-based conditioning and clique ideas: Puelz et al. (2021)
  - Recent extensions: Basse et al. (2024); Liu et al. (2025)

Step 1: choose focal units from the realized bipartite graph

- Left side: intervention units with the *realized* assignment.
- Filled node = assigned 1; empty node = assigned 0.
- We choose focal outcome units whose observed exposure is already:
  - **all 1**, or
  - **all 0**.



In this example,

$$\mathcal{U} = \{u_1, u_2\},$$

because u_1 is exposed to all 1 and u_2 is exposed to all 0.

Step 2: which assignments keep the null sharp?

- From the realized graph, suppose we selected

$$\mathcal{U} = \{u_1, u_2\}.$$

- To keep the null sharp on these focal units:
  u_1 and u_2 must remain exposed to **all 1** or **all 0**.
- $\mathcal{Z}$ contains a set of these "good" assignment vectors.

assignment vector

$\mathbf{z} = (z_A, z_B, z_C, z_D, z_E)$

$\mathbf{z}^{\text{obs}} = (1, 1, 0, 1, 0)$
observed assignment $\checkmark$

$(0, 0, 1, 0, 1)$
valid: preserves focal exposures $\checkmark$

$(1, 0, 0, 1, 0)$
invalid: u_1 no longer exposed to all 1 $\times$

Conditional randomization test

Conditioning event

Let Ω be the original assignment space. A conditioning event is

$$\mathcal{C} = (\mathcal{U}, \mathcal{Z}),$$

where

$$\mathcal{U} \subseteq \{1, \ldots, n\} \quad \text{and} \quad \mathcal{Z} \subseteq \Omega.$$

$\mathcal{U}$ is the focal set of outcome units; $\mathcal{Z}$ is the restricted set of assignments used in the test.

Sampling from the conditioning event

A conditioning mechanism specifies

$$\Pr(\mathcal{C} \mid \mathbf{Z} = \mathbf{z}).$$

The conditional assignment law is

$$\Pr(\mathbf{Z} = \mathbf{z} \mid \mathcal{C}) \propto \Pr(\mathcal{C} \mid \mathbf{Z} = \mathbf{z}) \Pr(\mathbf{Z} = \mathbf{z}), \qquad \mathbf{z} \in \Omega.$$

Bipartite framework

Treatment and outcome populations are distinct but linked

Related literature: bipartite experiments

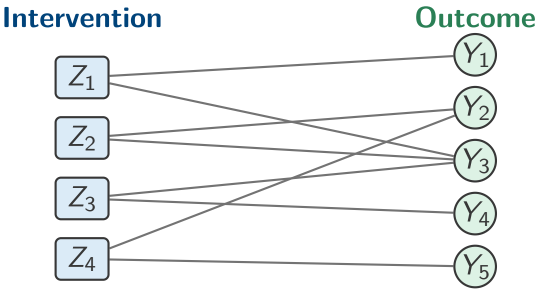
- **Bipartite causal inference and exposure mapping**
  - Doudchenko et al. (2020); Zigler and Papadogeorgou (2021); Papadogeorgou et al. (2025)
- **Design-based analysis of bipartite experiments**
  - Lu et al. (2025); Shi et al. (2025)
- **Design and optimization with bipartite interference**
  - Harshaw et al. (2023); Zigler et al. (2025)

This paper: finite-sample valid randomization tests for total and incremental effects in bipartite experiments.

Bipartite experiment in one picture

- m intervention units receive binary treatments.
- n outcome units have observed outcomes.
- $G_{ki} = 1$ indicates intervention k may affect outcome i.

$$\boldsymbol{Z} = (Z_1, \ldots, Z_m), \qquad Y_i(\boldsymbol{z}).$$



$\mathcal{G}_{+i} = \{k : G_{ki} = 1\}$ is outcome i's intervention neighborhood

Bipartite interference assumption

Local dependence through the bipartite graph

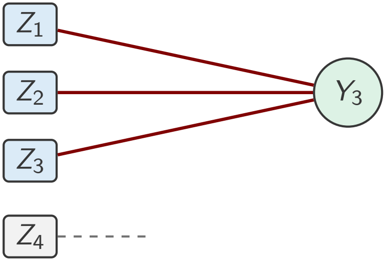
$$\boldsymbol{z}_{\mathcal{G}_{+i}} = \boldsymbol{z}'_{\mathcal{G}_{+i}} \quad \Rightarrow \quad Y_i(\boldsymbol{z}) = Y_i(\boldsymbol{z}').$$

Then we write

$$Y_i(\boldsymbol{z}_{\mathcal{G}_{+i}})$$

instead of a potential outcome indexed by all m interventions.

Only neighbors matter



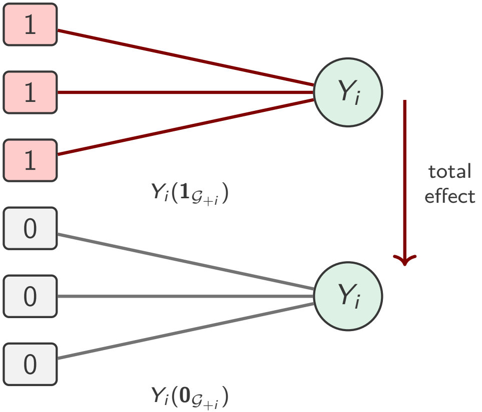
Z_4 not connected: no effect on Y_3

Two empirically relevant null hypotheses

Total effect

$$H_0^{\text{total}}: \quad Y_i(\mathbf{1}_{\mathcal{G}_{+i}}) = Y_i(\mathbf{0}_{\mathcal{G}_{+i}}) \quad \forall i.$$

Compare unit i when all connected interventions are treated versus when none are treated.



Exposure mapping assumption

Exchangeability among connected interventions

For outcome unit i, two local assignments with the same number of treated connected intervention units imply the same potential outcome:

$$\mathbf{1}^\top \mathbf{z}_{\mathcal{G}_{+i}} = \mathbf{1}^\top \mathbf{z}'_{\mathcal{G}_{+i}} \implies Y_i(\mathbf{z}_{\mathcal{G}_{+i}}) = Y_i(\mathbf{z}'_{\mathcal{G}_{+i}}).$$

This is an exposure mapping: the identities of treated neighbors do not matter once the count is fixed.

Exposure count

Define

$$D_i(\mathbf{z}) = \sum_{k \in \mathcal{G}_{+i}} z_k,$$

the number of treated intervention units connected to outcome unit i.

Then we can write

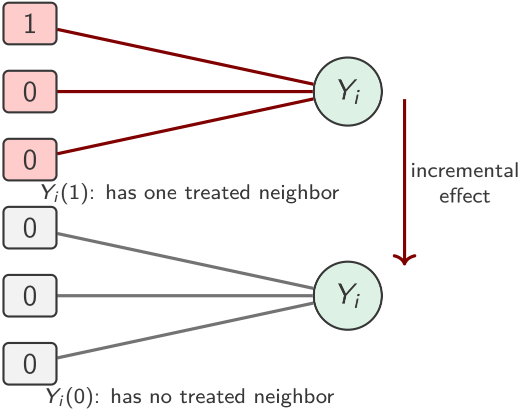
$$Y_i(\mathbf{z}_{\mathcal{G}_{+i}}) = Y_i(D_i(\mathbf{z}))\,.$$

Two empirically relevant null hypotheses

Incremental effect

$$H_0^{\text{inc}}: \quad Y_i(d+1) = Y_i(d) \quad \forall i.$$

One additional treated connected intervention does not change the outcome.



Total-effect test

Make the all-ones vs all-zeros null sharp by using pure blocks

Why the total-effect null is not sharp on the full design

The null $(H_0^{\text{total}} : Y_i(\mathbf{1}_{G_{+i}}) = Y_i(\mathbf{0}_{G_{+i}}), \quad \forall i)$ only equates two extreme exposure states: all connected interventions treated versus all connected interventions in control.

It does not restrict potential outcomes under mixed local assignments:

$$Y_i(1,0,0,\ldots), \quad Y_i(1,1,0,\ldots), \quad Y_i(0,1,0,\ldots), \quad \ldots$$

Strategy

Construct a conditioning event

$$\mathcal{C} = (\mathcal{U}, \mathcal{Z})$$

such that, for every focal outcome unit $i \in \mathcal{U}$ and every reassignment $\boldsymbol{z} \in \mathcal{Z}$,

$$\boldsymbol{z}_{\mathcal{G}_{+i}} \in \{\mathbf{1}_{G_{+i}}, \mathbf{0}_{G_{+i}}\}.$$

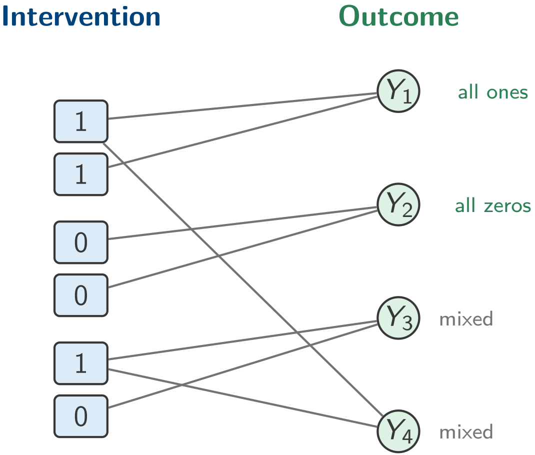
Then the total-effect null becomes sharp on the focal units.

Total-effect test: start from the observed graph

- Fix the observed assignment $\boldsymbol{Z}^{obs}$.
- We want to test the all-ones vs all-zeros null:

$$H_0: \; Y_i(\mathbf{1}_{G_{+i}}) = Y_i(\mathbf{0}_{G_{+i}}) \quad \forall i.$$

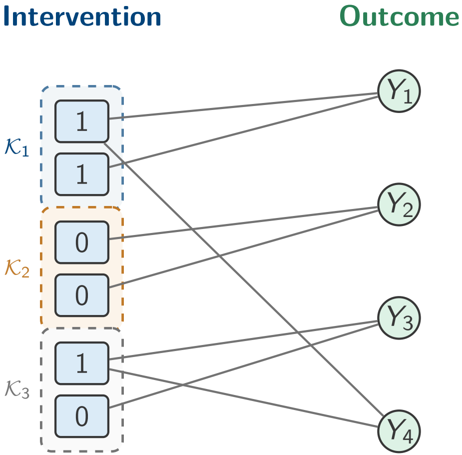
- Problem: under $\boldsymbol{Z}^{obs}$, not every outcome is observed under an all-ones or all-zeros exposure.



Total-effect test: partition intervention units

- Partition intervention units into blocks
$$\mathcal{K}_1, \ldots, \mathcal{K}_K.$$
- The partition is chosen using the graph only. Cannot use $\mathbf{Z}^{obs}$.
- Here we use three blocks:

$$\mathcal{K}_1 = \{1, 2\}, \mathcal{K}_2 = \{3, 4\}, \mathcal{K}_3 = \{5, 6\}.$$



Total-effect test: internal outcomes

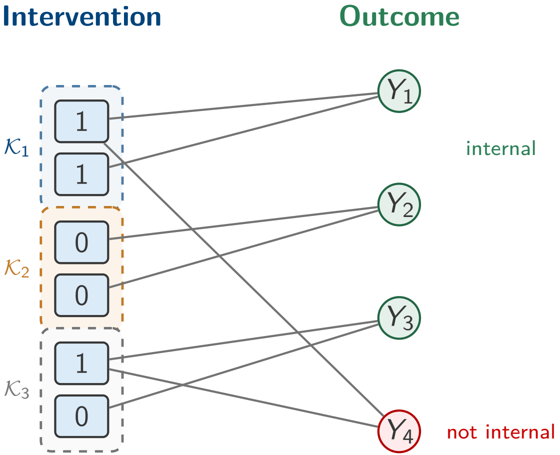
- Internal outcomes of block j:

$$\mathcal{V}_j = \{i : \mathcal{G}_{+i} \subseteq \mathcal{K}_j\}.$$

- Here:

$$\mathcal{V}_1 = \{Y_1\}, \mathcal{V}_2 = \{Y_2\}, \mathcal{V}_3 = \{Y_3\}.$$

- Y_4 is **not internal** because its neighbors span two blocks.



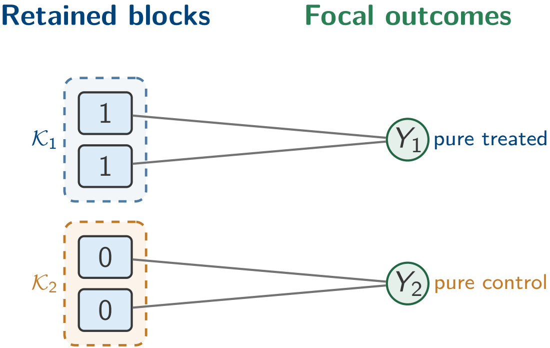
Total-effect test: keep only pure blocks

- Under $\boldsymbol{Z}^{obs}$:

  $$\mathcal{K}_1 = (1, 1), \mathcal{K}_2 = (0, 0), \mathcal{K}_3 = (1, 0).$$

- Keep only blocks that are **pure treated** or **pure control**.
- Drop mixed blocks and non-internal outcomes.
- Focal set here:

  $$\mathcal{U} = \{Y_1, Y_2\}.$$



Total-effect test: permute block labels

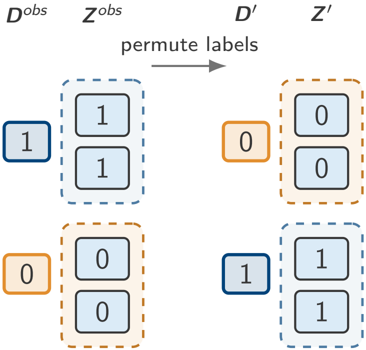
- On the retained blocks, form the block-level label vector

$$\boldsymbol{D}^{obs} = (1, 0).$$

- Permute the pure-block labels:

$$\boldsymbol{D}' \in \Pi(\boldsymbol{D}^{obs}).$$

- For focal outcomes, the exposure under permutation remains either all ones or all zeros.

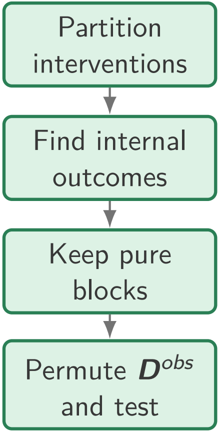


same number of treated blocks
valid conditional randomization

Total-effect test: algorithm summary

Procedure

1. Choose a graph-only partition $\{\mathcal{K}_1, \ldots, \mathcal{K}_K\}$.
2. Compute internal sets
$$\mathcal{V}_j = \{i : \mathcal{G}_{+i} \subseteq \mathcal{K}_j\}.$$
3. Keep pure blocks with $|\mathcal{V}_j| > 0$, and form $\boldsymbol{D}^{obs}$.
4. Permute $\boldsymbol{D}^{obs}$, recompute the test statistic, and compare with the observed value.



Why the total-effect test is exact

Sharpness

Focal units see only **1** or **0**

Invariance

The same $\mathcal{C}$ remain selected under resampling

Uniformity

Equal block sizes + fixed treated-block count

Finite-sample conclusion

Under complete randomization or i.i.d. Bernoulli assignment,

$$\Pr\{\text{pval} \leq \alpha\} \leq \alpha.$$

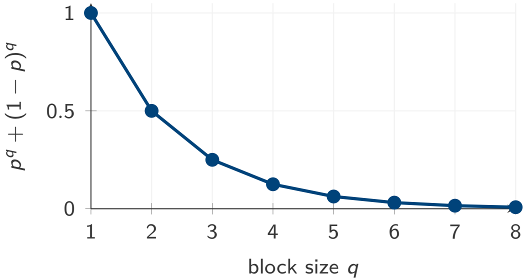
The validity holds for any test statistic imputable on $(\mathcal{U}, \mathcal{Z})$.

Power trade-off: blocks versus focal units

For Bernoulli(p), an equal-size block of size q is pure with probability

$$\Pr(\text{pure}) = p^q + (1-p)^q.$$

- Larger blocks capture more internal outcome units.
- But larger blocks are less likely to be pure.



Shown for $p = 0.5$. Sparse graphs often favor small-to-moderate blocks.

Incremental-effect test

Use switchers to move focal outcomes between d and $d + 1$

Exposure mapping: formal definition

Assumption from the paper

For every outcome unit i and any two local assignments,

$$\mathbf{1}_{G_{+i}}^{\top} \boldsymbol{z}_{\mathcal{G}_{+i}} = \mathbf{1}_{G_{+i}}^{\top} \boldsymbol{z}'_{\mathcal{G}_{+i}} \quad \Longrightarrow \quad Y_i(\boldsymbol{z}_{\mathcal{G}_{+i}}) = Y_i(\boldsymbol{z}'_{\mathcal{G}_{+i}}).$$

The identities of the treated neighbors are exchangeable once the treated-neighbor count is fixed.

Exposure count

$$E_i(\boldsymbol{z}) = \sum_{k \in \mathcal{G}_{+i}} z_k.$$

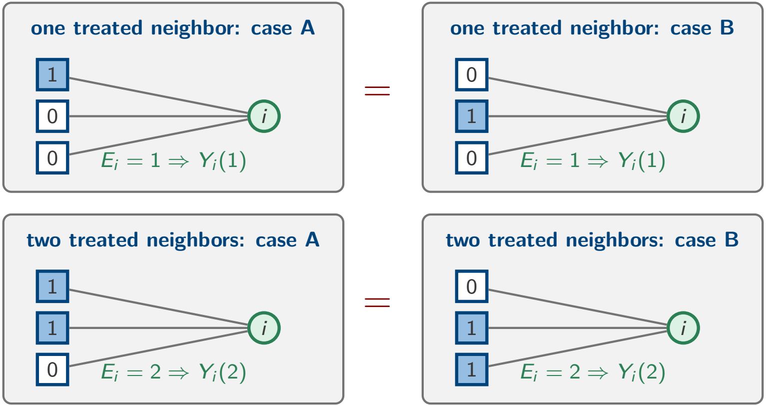
Then write

$$Y_i(\boldsymbol{z}_{\mathcal{G}_{+i}}) = Y_i(E_i(\boldsymbol{z})).$$

Incremental null

$$H_0^{\mathrm{inc}}(d) : \quad Y_i(d+1) = Y_i(d) \quad \forall i.$$

Exposure mapping: exchangeability visualization



Why the incremental null is not sharp on the full design

The null ($H_0^{\text{inc}}(d): Y_i(d+1) = Y_i(d), \quad \forall i$) only equates two adjacent exposure levels. It does not restrict potential outcomes at other exposure levels:

$$\ldots, Y_i(d-2), Y_i(d-1), \ldots, Y_i(d+2),\ Y_i(d+3), \ldots$$

or at exposure levels below d.

Strategy

Construct a conditioning event

$$\mathcal{C} = (\mathcal{U}, \mathcal{Z})$$

such that, for every focal outcome unit $i \in \mathcal{U}$ and every reassignment $\boldsymbol{z} \in \mathcal{Z}$,

$$E_i(\boldsymbol{z}) \in \{d, d+1\}.$$

Then the incremental null becomes sharp on the focal units.

Step 1: identify the target exposure classes

For illustration set $d = 0$.

$$E_i^{\text{obs}} = \sum_{k \in \mathcal{G}_{+i}} Z_k^{\text{obs}}.$$

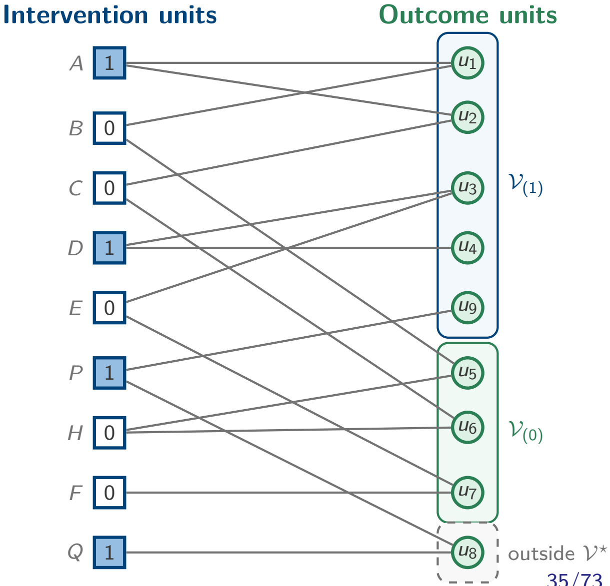
Let

$$\mathcal{V}_{(0)} = \{i : E_i^{\text{obs}} = 0\},$$
$$\mathcal{V}_{(1)} = \{i : E_i^{\text{obs}} = 1\}.$$

The initial target set is

$$\mathcal{V}^\star = \mathcal{V}_{(0)} \cup \mathcal{V}_{(1)}.$$

Outcome units outside $\mathcal{V}^\star$ are not allowed to enter the test after resampling.



Step 2: internal intervention units

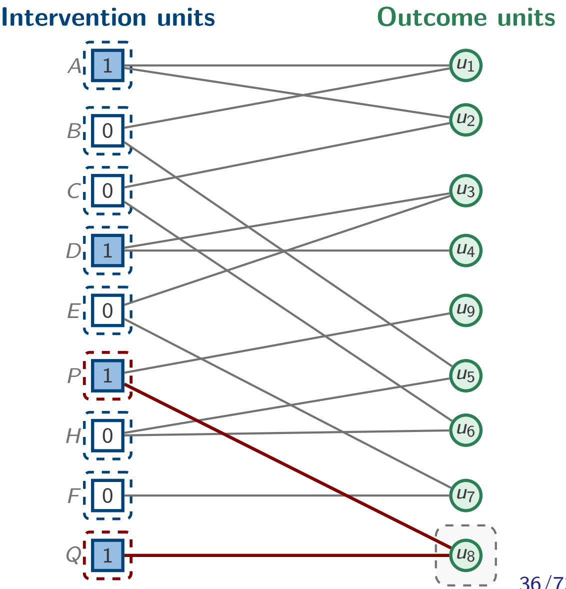
An intervention unit is internal if all outcome units it can affect are already in the target exposure classes:

$$\mathcal{I}^{\mathrm{int}} = \{k : \mathcal{G}_{k+} \subseteq \mathcal{V}^{\star}\}.$$

Why this matters: flipping an internal unit cannot create a new focal outcome outside $\mathcal{V}^{\star}$.

In the toy graph

A, B, C, D, E, H, F are internal.
P, Q are not internal because they affect u_8, which has $E = 2$.



Step 3: choose a disjoint internal set $\mathcal{W}$

From the internal units, choose a deterministic disjoint set

$$\mathcal{W} = \{w_1, \ldots, w_R\} \subseteq \mathcal{I}^{\text{int}},$$

with

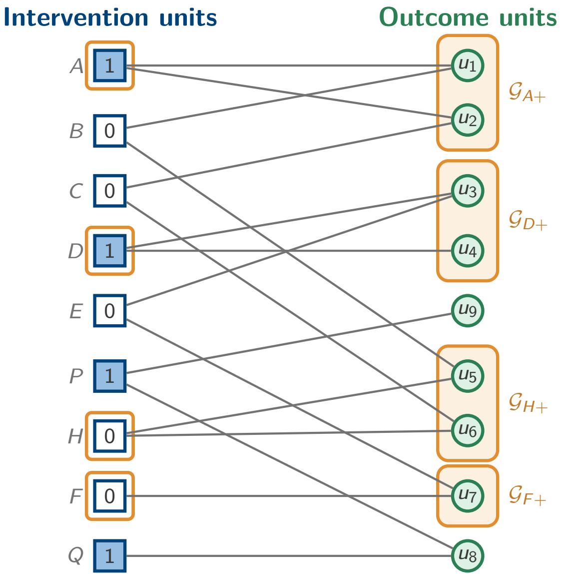
$$\mathcal{G}_{w_r+} \cap \mathcal{G}_{w_{r'}+} = \varnothing \qquad (r \neq r').$$

Disjointness makes each focal outcome block controlled by at most one switcher.

Toy graph

A possible choice is

$$\mathcal{W} = \{A, D, H, F\}.$$



Step 4: which internal units are eligible switchers?

For a unit $k \in \mathcal{W}$, check whether every affected outcome has exactly d treated neighbors *other than* k:

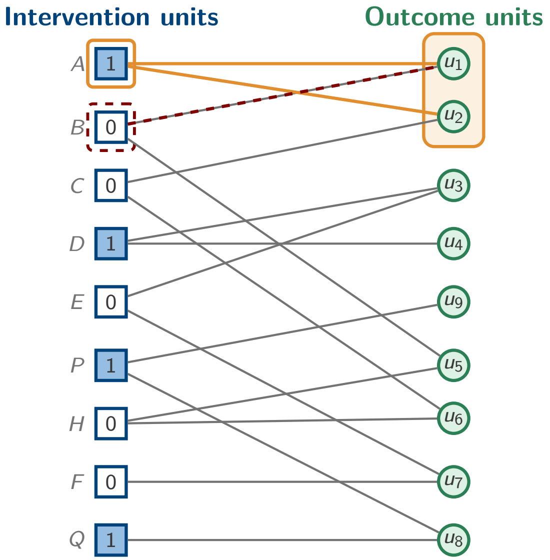
$$\sum_{\ell \in \mathcal{G}_{+i} \setminus \{k\}} Z_\ell^{\mathrm{obs}} = d \quad \text{for all } i \in \mathcal{G}_{k+}.$$

Then toggling k gives

$$E_i = d + Z_k \in \{d, d+1\}.$$

Example with $d = 0$

A is eligible: the other neighbors of u_1, u_2 are $B = 0$, $C = 0$.
B is not eligible: for u_1, the other neighbor is $A = 1$.



Step 5: switchers and focal outcome blocks

After the eligibility check,

$$\mathcal{S}^L = \{s_1, \ldots, s_L\} = \{A, D, H, F\}.$$

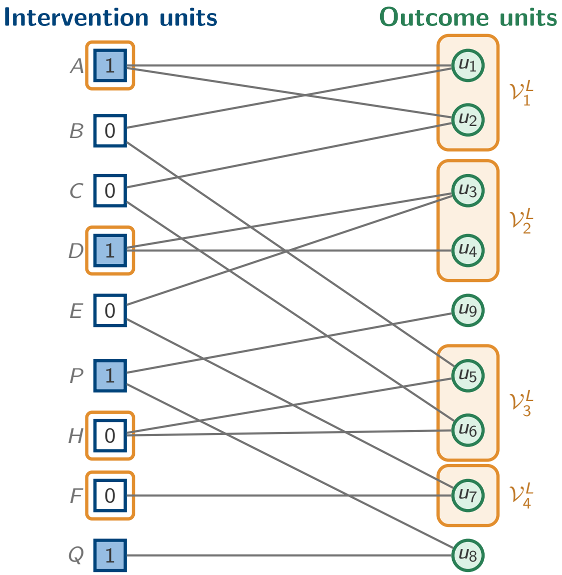
For each switcher,

$$\mathcal{V}_j^L = \mathcal{G}_{s_j+}, \qquad \mathcal{U} = \bigcup_{j=1}^{L} \mathcal{V}_j^L.$$

The observed switcher labels are

$$\boldsymbol{D}^{\mathrm{obs}} = (Z_A, Z_D, Z_H, Z_F) = (1, 1, 0, 0).$$

Each block is either under exposure d or exposure $d+1$ depending only on its switcher.



Step 6: condition and permute switcher labels

Fix all non-switcher intervention assignments:

$$\boldsymbol{z}_{\mathcal{J}} = \boldsymbol{Z}_{\mathcal{J}}^{\mathrm{obs}}, \qquad \mathcal{J} = \{1, \dots, m\} \setminus \mathcal{S}^L.$$

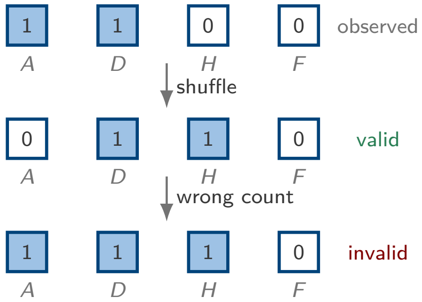
Condition on the number of treated switchers:

$$\sum_{j=1}^{L} z_{s_j} = L_1, \qquad L_1 = \sum_{j=1}^{L} D_j^{\mathrm{obs}}.$$

Thus

$$\mathcal{Z} = \left\{ \boldsymbol{z} : \boldsymbol{z}_{\mathcal{J}} = \boldsymbol{Z}_{\mathcal{J}}^{\mathrm{obs}}, \sum_{j=1}^{L} z_{s_j} = L_1 \right\}.$$

Switcher labels $\boldsymbol{D} = (Z_A, Z_D, Z_H, Z_F)$



Valid resamples preserve exactly two treated switchers

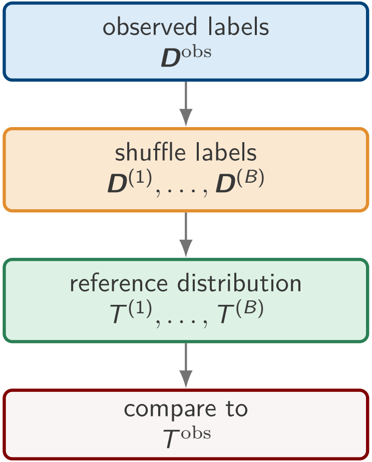
Step 7: compute the permutation statistic

For any permuted switcher-label vector $\boldsymbol{D}$, a difference-in-means statistic is

$$T(\boldsymbol{D} \mid \boldsymbol{Y}^{\mathrm{obs}}, \mathcal{V}^L) = \frac{1}{n_{d+1}} \sum_{j=1}^{L} D_j \sum_{i \in \mathcal{V}_j^L} Y_i^{\mathrm{obs}} - \frac{1}{n_d} \sum_{j=1}^{L} (1 - D_j) \sum_{i \in \mathcal{V}_j^L} Y_i^{\mathrm{obs}},$$

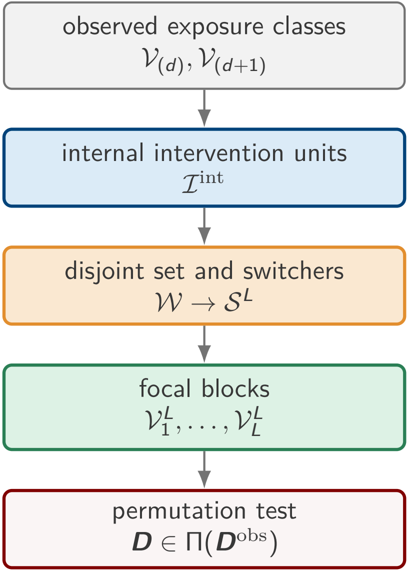
where

$$n_{d+1} = \sum_{j=1}^{L} D_j |\mathcal{V}_j^L|, \qquad n_d = \sum_{j=1}^{L} (1 - D_j) |\mathcal{V}_j^L|.$$



Incremental-effect test: complete procedure

1. Compute observed exposures E_i^{obs}.
2. Form $\mathcal{V}^\star = \mathcal{V}_{(d)} \cup \mathcal{V}_{(d+1)}$.
3. Find internal interventions $\mathcal{I}^{\mathrm{int}}$.
4. Choose a deterministic disjoint set $\mathcal{W}$.
5. Keep eligible switchers $\mathcal{S}^L$.
6. Define focal blocks $\mathcal{V}_j^L = \mathcal{G}_{s_j+}$.
7. Permute $\boldsymbol{D}^{\mathrm{obs}} = (Z_{s_1}^{\mathrm{obs}}, \ldots, Z_{s_L}^{\mathrm{obs}})$ holding $\sum_j D_j$ fixed.



Why the incremental-effect test is exact

Sharpness

Focal outcomes always have $E_i = d$ or $d + 1$

Invariance

Internal and disjoint switchers keep $\mathcal{C} = (\mathcal{U}, \mathcal{Z})$ fixed

Uniformity

Given $\boldsymbol{z}_{\mathcal{J}}$ and L_1, switcher labels are exchangeable

Finite-sample conclusion

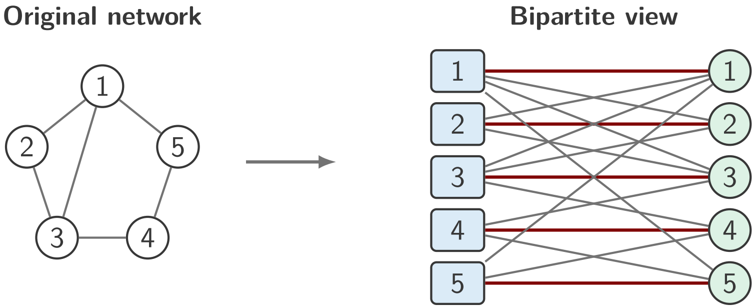
Under complete randomization or i.i.d. Bernoulli assignment with common probability,

$$\Pr\{\text{pval} \leq \alpha\} \leq \alpha.$$

The argument applies to any test statistic imputable on the selected focal blocks.

Network experiments

A network experiment can be embedded as a bipartite experiment



Original network
Bipartite view
1
2
3
4
5
1
2
3
4
5
1
2
3
4
5

Step 1: start from a network experiment

Suppose units form a network with adjacency matrix A_{ik}.

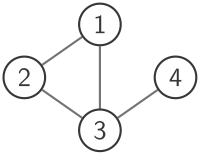
Unit i may be affected by

- its own treatment Z_i,
- the treatments of its neighbors.

Idea

A network experiment can be rewritten as a special bipartite experiment.

Original network



Edges indicate interference links

Step 2: embed the network as a bipartite experiment

Create two copies of the same units:

$$\mathcal{I} = \mathcal{O} = \{1, \ldots, n\}.$$

Interpret

- the left copy as intervention units,
- the right copy as outcome units.

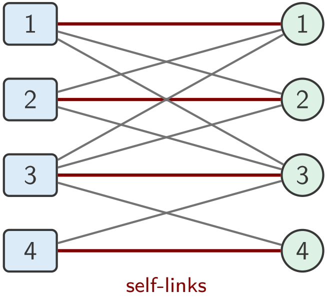
Define the bipartite graph by

$$G_{ki} = \begin{cases} A_{ik}, & i \neq k, \\ 1, & i = k. \end{cases}$$

Thus outcome unit i is connected to

- itself,
- all of its network neighbors.

Bipartite representation



gray edges reproduce the original network links

Consequence: total effect tests follow directly

After the embedding, a network experiment is just a special bipartite experiment.

Under local interference, we test the null of global treatment effect:

$$H_0^{\text{network,total}}: \quad Y_i(\mathbf{1}_{\mathcal{G}_{+i}}) = Y_i(\mathbf{0}_{\mathcal{G}_{+i}}) \qquad \forall i,$$

where

$$\mathcal{G}_{+i} = \{i\} \cup \{k : A_{ik} = 1\}.$$

Key message

For total effects, nothing new is needed: after embedding the network as a bipartite experiment, the same conditioning and permutation procedure applies directly.

Network incremental effects: one extra complication

In a network experiment, write

$$Y_i(z, d),$$

where

$$\begin{aligned} z &= \text{unit } i\text{'s own treatment status,} \\ d &= \text{number of treated neighbours of unit } i. \end{aligned}$$

The incremental-effect null is

$$H_0^{\text{network,inc}}: \quad Y_i(z, d+1) = Y_i(z, d) \qquad \forall i.$$

New issue in networks

A switcher is also an outcome unit.

If the switcher is focal, permuting its treatment changes its own z.

Fix: switchers are not focal outcomes

Use the same switcher idea as in the bipartite incremental-effect test.

Extra rule for networks

A switcher can affect focal units, but the switcher itself is not used as a focal outcome.

For every focal unit:

$$\begin{aligned} \text{own treatment } z &\quad \text{stays fixed,} \\ \text{treated neighbours } d &\quad \text{move between } d \text{ and } d+1. \end{aligned}$$

Then

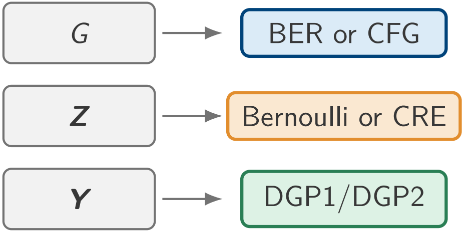
The same conditioning and permutation test applies.

Simulation evidence

Size is controlled; power depends on certified focal units

Simulation design

- $m = 400$ interventions, $n = 1600$ outcomes.
- Sparse bipartite graphs:
    - BER: homogeneous sparse graph.
    - CFG: degree-heterogeneous graph.
- Assignment: Bernoulli(0.5) or complete randomization.
- 2000 Monte Carlo replications, 2000 permutations.



Simulation: total-effect test

Table 1: size under the null ($c = 0$)

Comparison of the conditional Fisher randomization test and the Hájek t-test for the total-effect null.

Graph	Design	FRT Reject	Hájek Reject	Median L	Median $\lvert\mathcal{U}\rvert/n$
BER	Bernoulli	0.0445	0.0245	100.0	0.4050
BER	CRE	0.0520	0.0215	100.0	0.4056
CFG	Bernoulli	0.0475	0.0205	99.0	0.3262
CFG	CRE	0.0500	0.0125	99.0	0.3250

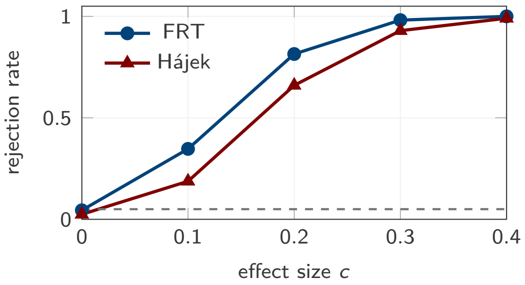
Entries are empirical rejection rates at $\alpha = 0.05$ under the total-effect null. The FRT stays close to nominal size, while the Hájek benchmark is conservative.

Comparison: conditional FRT vs Hájek/Wald benchmark

The Hájek benchmark (Lu et al. (2025)) uses all observed all-one/all-zero exposure units, but relies on a conservative variance bound.

In these designs

The conditional FRT is closer to nominal size and more powerful.



BER/Bernoulli shown.

Simulation: incremental-effect test

Graph	Design	Reject					Median L	Median $\lvert\mathcal{U}\rvert/n$
		$c = 0$	0.1	0.2	0.3	0.4		
BER	Bernoulli	0.0495	0.1825	0.4235	0.7200	0.8940	67	0.150
BER	CRE	0.0475	0.1760	0.4330	0.7025	0.8995	67	0.149
CFG	Bernoulli	0.0436	0.0881	0.1461	0.2075	0.3292	15	0.031
CFG	CRE	0.0495	0.0866	0.1316	0.1975	0.3050	15	0.031

At $c = 0$, rejection rates remain close to the nominal 5% level. Smaller certified sets under CFG are associated with substantially lower power.

Application I: China's high-speed rail network

HSR lines are intervention units; cities are outcome units

HSR application: data and empirical question

Empirical setting

- Intervention units: HSR lines.
- Outcome units: cities.
- Edge: an HSR line passes through a city.
- Treatment: line completed by 2016.
- Outcome: city employment growth from 2007 to 2016.

Substantive question

Does HSR line completion affect city employment growth through changes in market access?

Quantity	Value
HSR lines	150
Cities with outcome data	237
Average city degree	2.1
Maximum city degree	9
Average line degree	3.8
Maximum line degree	18

The graph is sparse: most cities are connected to only a few lines. This is favorable for finding focal exposure sets.

HSR application: observed exposure groups

Total-effect test

Compare cities with all connected HSR lines completed versus cities with no connected HSR line completed:

$$Y_i(\mathbf{1}_{G_{+i}}) \quad \text{vs.} \quad Y_i(\mathbf{0}_{G_{+i}}).$$

Incremental-effect test

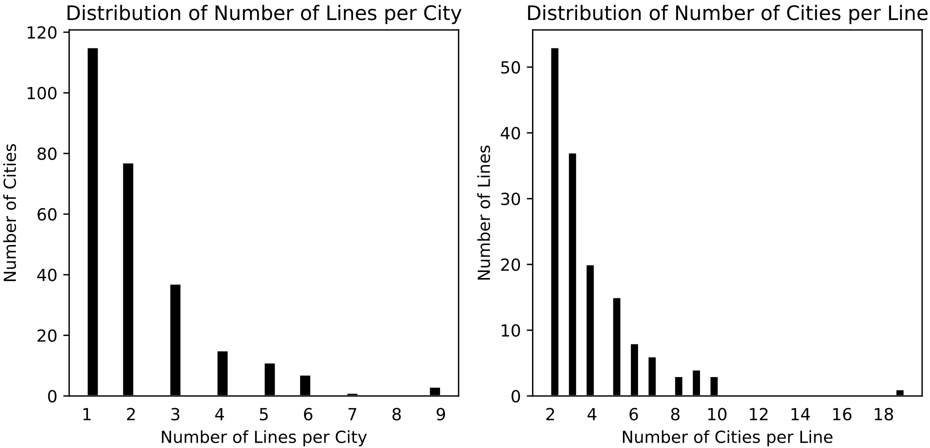
Compare cities with exactly one completed connected line versus no completed connected line:

$$Y_i(1) \quad \text{vs.} \quad Y_i(0).$$

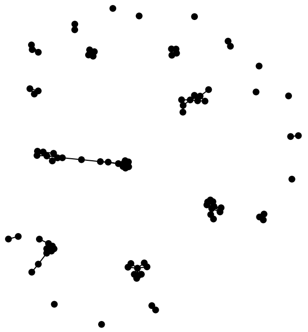
Observed exposure group	Cities
All connected lines completed	97
No connected lines completed	55
All-one or all-zero exposure	152
Exactly one completed connected line	104
No completed connected line	55

The total-effect test starts from 152 all-one/all-zero cities. The incremental test uses the single-one versus all-zero comparison.

HSR application: degree distributions from the paper



HSR application: focal graph components for total effects



What this figure shows

The appendix figure visualizes the intervention-disjoint graph among cities observed under the all-one exposure state.

- All-one focal graph: 24 clusters.
- All-zero focal graph: 26 clusters.
- These clusters create a rich block-permutation space.
- Boundary screening ensures the conditioning event remains invariant.

HSR application: how the tests are implemented

Total effect

$$H_0: \quad Y_i(\mathbf{1}_{G_{+i}}) = Y_i(\mathbf{0}_{G_{+i}}) \quad \forall i.$$

- Start from cities observed under all-one or all-zero exposure.
- Decompose the induced bipartite graph into intervention-disjoint components.
- Permute component labels after boundary screening.

Incremental effect

$$H_0: \quad Y_i(1) = Y_i(0) \quad \forall i.$$

- Start from cities with zero or one completed connected line.
- Find switchers that move cities only between exposure 0 and exposure 1.
- Permute switcher labels while fixing the treated-switcher count.

HSR application: testing results

Effect	Statistic	p-value	Estimate
All ones vs all zeros	Diff-in-Means	0.1254	0.05
	IPW	0.1649	0.10
Single one vs all zeros	Diff-in-Means	0.4186	0.01
	IPW	0.2478	0.02

Main finding

No p-value is below 5%.

- No strong evidence of a total effect of HSR completion on employment growth.
- No strong evidence of a single-line incremental effect.

HSR application: synthetic-effect threshold exercise

To assess how large an effect would be needed for rejection, the paper adds a constant shift to treated-exposure outcomes:

$$\widetilde{Y}_i^{\mathrm{obs}} = Y_i^{\mathrm{obs}} + \mathbf{1}\{\text{treated exposure}\}\tau.$$

For the all-ones versus all-zeros test, treated exposure means

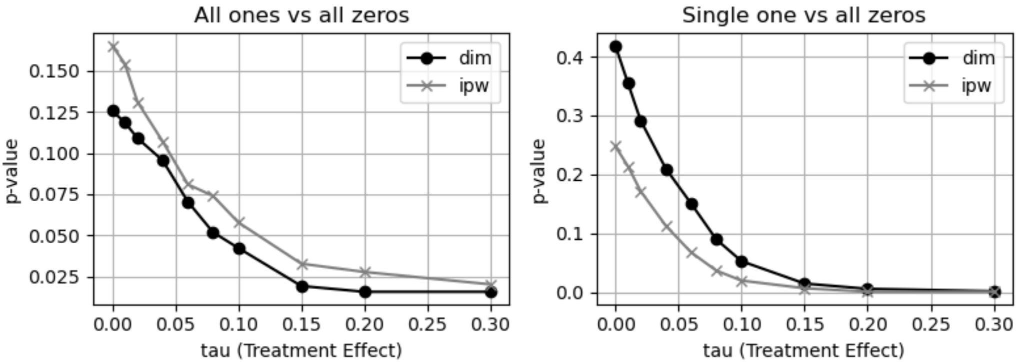
$$\boldsymbol{Z}_{\mathcal{G}_{+i}} = \mathbf{1}_{G_{+i}}.$$

Interpretation

This is a thresholding exercise under the observed design and observed outcome realization.

- It is not a formal power calculation.
- The same focal units and permutation design are reused.
- The question is: how large must a synthetic shift be before the test rejects?
- In this application, rejection occurs around $\tau \approx 0.10$ for the total-effect exercise.

HSR application: synthetic-effect p-value curve



Paper figure: p-values under synthetic constant shifts for the all-ones versus all-zeros HSR test.

Application II: social network experiment in schools

Students are both intervention units and outcome units

School network application: data description

Empirical setting

- Anti-conflict intervention in schools.
- Original experiment conducted across 56 schools.
- Analysis focuses on the five largest treated-school networks.
- Sample: 320 treatment-eligible students.
- Network ties come from student friendship nominations.

Outcome

Self-reported endorsement of anti-conflict norms, measured through reported wristband wearing.

Why the bipartite view helps

Each student appears twice:

- as an intervention unit whose treatment may affect peers;
- as an outcome unit whose norms may respond to peer treatment.

School network application: hypotheses

Total effect

$$H_0 : \quad Y_i(\mathbf{1}_{G_{+i}}) = Y_i(\mathbf{0}_{G_{+i}}) \quad \forall i.$$

This compares a policy where the student and relevant peers are treated versus a policy where none are treated.

Incremental peer effect

For ego treatment z and number of treated friends d,

$$H_0 : \quad Y_i(z, d+1) = Y_i(z, d) \quad \forall i.$$

The application focuses on the single treated peer versus no treated peer comparison.

For incremental effects, switchers are excluded from the focal outcome set so that every focal student's own treatment status remains fixed while peer exposure changes.

School network application: testing results

Effect	Statistic	p-value	Estimate
All ones vs all zeros	Diff-in-Means	0.0360	0.26
	IPW	0.0210	0.24
Single one vs all zeros	Diff-in-Means	0.1126	0.11
	IPW	0.1122	0.12

Main finding

The total-effect test rejects at 5%; the single-peer incremental test does not.

- Evidence supports a broad treated-neighborhood effect.
- The marginal effect of one treated peer is weaker.
- Diff-in-Means and IPW give similar conclusions.

Empirical lessons from the two applications

HSR network

- Sparse bipartite graph creates many all-one/all-zero focal units.
- No rejection for total or incremental effects.
- Synthetic shift exercise suggests rejection around $\tau \approx 0.10$ for the total-effect test.

School network

- Network experiment is embedded as a bipartite experiment.
- Total-effect null is rejected.
- Single-peer incremental null is not rejected at 5%.

Main empirical message

Conditional randomization tests can be applied directly to real bipartite and network data once the graph supplies enough certified focal units and admissible assignment permutations.

Takeaways

What the method delivers

- Exact finite-sample inference for non-classical interference hypotheses.
- Graph-aware conditioning events that restore sharpness on focal outcome units.
- Flexible test statistics, including Diff-in-Means and IPW-style statistics.

What determines power

- Total-effect tests benefit from many pure all-one/all-zero components.
- Incremental-effect tests need enough eligible switchers.
- Dense or highly overlapping graphs can reduce the certified focal set.

Thank you

Questions and comments welcome

References I

Aronow, P. M. (2012). A general method for detecting interference between units in randomized experiments. *Sociological Methods & Research*, **41** 3–16.

Athey, S., Eckles, D. and Imbens, G. W. (2018). Exact p-values for network interference. *Journal of the American Statistical Association*, **113** 230–240.

Basse, G., Ding, P., Feller, A. and Toulis, P. (2024). Randomization tests for peer effects in group formation experiments. *Econometrica*, **92** 567–590.

Basse, G. W., Feller, A. and Toulis, P. (2019). Randomization tests of causal effects under interference. *Biometrika*, **106** 487–494.

Doudchenko, N., Zhang, M., Drynkin, E., Airoldi, E., Mirrokni, V. and Pouget-Abadie, J. (2020). Causal inference with bipartite designs. 2010.02108, URL https://arxiv.org/abs/2010.02108.

Harshaw, C., Sävje, F., Eisenstat, D., Mirrokni, V. and Pouget-Abadie, J. (2023). Design and analysis of bipartite experiments under a linear exposure-response model. *Electronic Journal of Statistics*, **17** 464 – 518.

Liu, J., Shaikh, A. M. and Toulis, P. (2025). Randomization inference in two-sided market experiments. 2504.06215, URL https://arxiv.org/abs/2504.06215.

Lu, S., Shi, L., Fang, Y., Zhang, W. and Ding, P. (2025). Design-based causal inference in bipartite experiments. 2501.09844, URL https://arxiv.org/abs/2501.09844.

Papadogeorgou, G., Song, Z., Imbens, G. and Mealli, F. (2025). Causal inference when intervention units and outcome units differ. 2507.20231, URL https://arxiv.org/abs/2507.20231.

References II

Puelz, D., Basse, G., Feller, A. and Toulis, P. (2021). A Graph-Theoretic Approach to Randomization Tests of Causal Effects under General Interference. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **84** 174–204.

Shi, L., Bakhitov, E., Hung, K., Karrer, B., Walker, C., Bhole, M. and Schrijvers, O. (2025). Scalable analysis of bipartite experiments. `2402.11070`, URL `https://arxiv.org/abs/2402.11070`.

Zigler, C., Liu, V., Mealli, F. and Forastiere, L. (2025). Bipartite interference and air pollution transport: estimating health effects of power plant interventions. *Biostatistics*, **26** kxae051.

Zigler, C. and Papadogeorgou, G. (2021). Bipartite causal inference with interference. *Statistical Science*, **36** 109–123.