# Correlated Cluster-Based Randomized Experiments: Robust Variance Minimization

Chen Chen

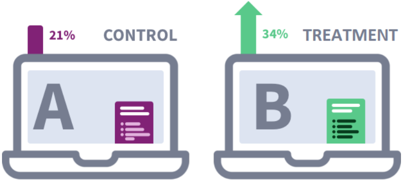
上海纽约大学
NYU SHANGHAI

ACM EC 2023
July 9, 2023

*Joint work with:*
Ozan Candogan and Rad Niazadeh (Chicago Booth)

# Randomized experiments

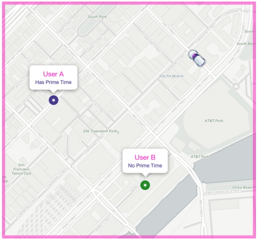
- Experimental design: the science of designing randomized tests (e.g., A/B testing), a.k.a. experiments, to measure the effectiveness of an intervention.




- **High-level goal:** estimate the total market effect, i.e., the difference in total potential outcomes of the users if the intervention is introduced to the entire market.

# Network effects and interference

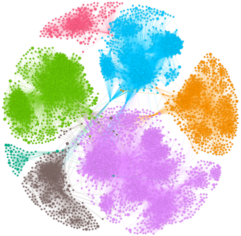
Social network
Eckles et al. (2016)



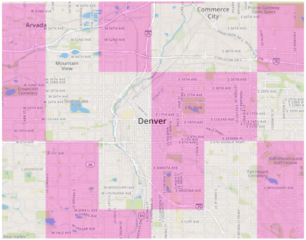

Ride-sharing
Chamandy (2016)

- Experiments in online platforms/networks often suffer from **interference:** one user's assignment to the treatment or control affects another user's outcome.

# Cluster-based experimental design



Social network
Eckles et al. (2016)

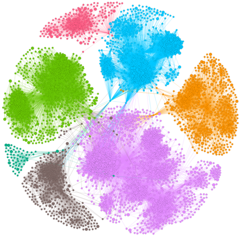



Ride-sharing
Chamandy (2016)

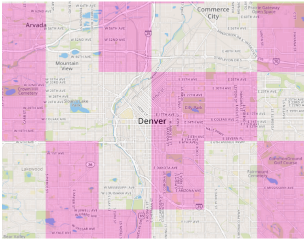
- Experiments in online platforms/networks often suffer from **interference:** one user's assignment to the treatment or control affects another user's outcome.
- **Common practice:** (i) group users who likely have substantial impact on each others' outcomes into clusters, (ii) assign all users in a cluster to the same variant.

$\implies$ often leads to a small **bias** (if any)

# Cluster-based experimental design



Social network
Eckles et al. (2016)




Ride-sharing
Chamandy (2016)

- Experiments in online platforms/networks often suffer from **interference:** one user's assignment to the treatment or control affects another user's outcome.
- **Common practice:** (i) group users who likely have substantial impact on each others' outcomes into clusters, (ii) assign all users in a cluster to the same variant.

  $\Longrightarrow$ often leads to a small **bias** (if any)
- **Research question:** obtain the "best" (correlated) randomized assignment to minimize **variance**.

# Problem formulation

A platform conducts binary experiment over $n$ disjoint (and heterogeneous) clusters:

- Each cluster $i$ receives a **treatment** $=$"**1**" or **control** $=$"**0**" variant
  - $Z_i = 1$ or $0$: indicator of assignment

# Problem formulation

A platform conducts binary experiment over $n$ disjoint (and heterogeneous) clusters:

- Each cluster $i$ receives a **treatment** $=$"**1**" or **control** $=$"**0**" variant
  - $Z_i = 1$ or $0$: indicator of assignment
- A cluster $i$ has a treatment (control) outcome $y_{i1}$ ($y_{i0}$) if it receives the treatment (control) assignment

# Problem formulation

A platform conducts binary experiment over $n$ disjoint (and heterogeneous) clusters:

- Each cluster $i$ receives a **treatment** $=$"**1**" or **control** $=$"**0**" variant
  - $Z_i = 1$ or $0$: indicator of assignment
- A cluster $i$ has a treatment (control) outcome $y_{i1}$ ($y_{i0}$) if it receives the treatment (control) assignment
- **Objective:** estimate the total market effect $\tau \triangleq \sum_{i\in[n]} y_{i1} - \sum_{i\in[n]} y_{i0}$

# Problem formulation

A platform conducts binary experiment over $n$ disjoint (and heterogeneous) clusters:

- Each cluster $i$ receives a **treatment** $=$"**1**" or **control** $=$"**0**" variant
  - $Z_i = 1$ or $0$: indicator of assignment
- A cluster $i$ has a treatment (control) outcome $y_{i1}$ ($y_{i0}$) if it receives the treatment (control) assignment
- **Objective:** estimate the total market effect $\tau \triangleq \sum_{i\in[n]} y_{i1} - \sum_{i\in[n]} y_{i0}$
- Use the unbiased Horvitz-Thompson estimator:

$$\hat{\tau} \triangleq \sum_{i\in[n]} y_{i1} \frac{Z_i}{\mathbb{P}\big[Z_i = 1\big]} - \sum_{i\in[n]} y_{i0} \frac{1 - Z_i}{\mathbb{P}\big[Z_i = 0\big]}$$

# Problem formulation

A platform conducts binary experiment over $n$ disjoint (and heterogeneous) clusters:

- Each cluster $i$ receives a **treatment** $=$ "**1**" or **control** $=$ "**0**" variant
  - $Z_i = 1$ or $0$: indicator of assignment
- A cluster $i$ has a treatment (control) outcome $y_{i1}$ ($y_{i0}$) if it receives the treatment (control) assignment
- **Objective:** estimate the total market effect $\tau \triangleq \sum_{i\in[n]} y_{i1} - \sum_{i\in[n]} y_{i0}$
- Use the unbiased Horvitz-Thompson estimator:

$$\hat{\tau} \triangleq \sum_{i\in[n]} y_{i1} \frac{Z_i}{\mathbb{P}\left[Z_i = 1\right]} - \sum_{i\in[n]} y_{i0} \frac{1 - Z_i}{\mathbb{P}\left[Z_i = 0\right]}$$

- **Goal:** design joint distribution (or correlation) of $(Z_i)_{i\in[n]}$ to minimize variance of the estimation

# Problem formulation

**Problem:** minimize variance $\mathbb{V}\mathrm{ar}\big[\hat{\tau}\big]$ against the worst-case potential outcomes

$$V^{\mathrm{OPT}} = \min_{P \in \mathcal{P}_q} \max_{\substack{y_{i0} \in [0, w_{i0}], \\ y_{i1} \in [0, w_{i1}], \, \forall \, i \in [n]}} \mathbb{V}\mathrm{ar}\big[\hat{\tau}\big].$$

$\mathcal{P}_q \triangleq$ set of joint assignment distributions with marginal prob $q \in (0, 1)$

# Problem formulation

**Problem:** minimize variance $\mathbb{V}\mathrm{ar}\left[\hat{\tau}\right]$ against the worst-case potential outcomes

$$V^{\mathrm{OPT}} = \min_{P \in \mathcal{P}_q} \max_{\substack{y_{i0} \in [0, w_{i0}], \\ y_{i1} \in [0, w_{i1}], \forall i \in [n]}} \mathbb{V}\mathrm{ar}\left[\hat{\tau}\right].$$

$\mathcal{P}_q \triangleq$ set of joint assignment distributions with marginal prob $q \in (0, 1)$

- Potential outcomes: non-negative and bounded
  - Upper bounds $w_{i1}$, $w_{i0}$ can vary cross clusters and treatment/control variants

# Problem formulation

**Problem:** minimize variance $\mathbb{V}\mathrm{ar}\big[\hat{\tau}\big]$ against the worst-case potential outcomes

$$V^{\mathrm{OPT}} = \min_{P \in \mathcal{P}_q} \max_{\substack{y_{i0} \in [0, w_{i0}], \\ y_{i1} \in [0, w_{i1}], \, \forall\, i \in [n]}} \mathbb{V}\mathrm{ar}\big[\hat{\tau}\big].$$

$\mathcal{P}_q \triangleq$ set of joint assignment distributions with marginal prob $q \in (0,1)$

- Potential outcomes: non-negative and bounded
  - Upper bounds $w_{i1}$, $w_{i0}$ can vary cross clusters and treatment/control variants
- Joint assignment dists: $\forall\, i \in [n]$: $\mathbb{P}[Z_i = 1] = q \in (0,1)$

# Optimal correlation design among clusters

**Lemma** *With any cluster-based randomized experiment, the worst-case potential outcome is such that for any cluster* $i \in [n]$*, either* $y_{i1} = y_{i0} = 0$*, or* $y_{i1} = w_{i1}$ *and* $y_{i0} = w_{i0}$*. The variance of the HT estimator is*

$$\mathbb{V}\mathrm{ar}\left[\hat{\tau}\right] = y^{\mathrm{T}} \Sigma y,$$

*where* $y_i = \sqrt{q(1-q)} \cdot \left( \frac{y_{i1}}{q} + \frac{y_{i0}}{1-q} \right)$*, and* $\Sigma$ *is the correlation matrix of the assignments* $(Z_i)_{i \in [n]}$*.*

# Optimal correlation design among clusters

**Lemma** *With any cluster-based randomized experiment, the worst-case potential outcome is such that for any cluster $i \in [n]$, either $y_{i1} = y_{i0} = 0$, or $y_{i1} = w_{i1}$ and $y_{i0} = w_{i0}$. The variance of the HT estimator is*

$$\mathbb{V}\mathrm{ar}\left[\hat{\tau}\right] = y^{\mathrm{T}} \Sigma y,$$

*where $y_i = \sqrt{q(1-q)} \cdot \left(\frac{y_{i1}}{q} + \frac{y_{i0}}{1-q}\right)$, and $\Sigma$ is the correlation matrix of the assignments $(Z_i)_{i \in [n]}$.*

The **optimization problem** becomes

correlation matrix under joint distribution $P$

$$V^{\mathrm{OPT}} = \min_{P \in \mathcal{P}_q} \max_{y_i \in [0, w_i], \forall i \in [n]} y^{\mathrm{T}} \Sigma(P) y,$$

with $w_i \triangleq \sqrt{q(1-q)} \cdot \left(\frac{w_{i1}}{q} + \frac{w_{i0}}{1-q}\right)$.

# Optimal correlation design among clusters

The **optimization problem:**

$$V^{\mathrm{OPT}} = \min_{P \in \mathcal{P}_q} \max_{y \in \times_{i \in [n]} \{0, w_i\}} y^{\mathrm{T}} \Sigma(P) y$$

- **Difficult to solve:** number of decision variables/comstraints exponential in number of clusters
- **Difficult to interpret and implement:** correlation structure can be complicated

# Optimal correlation design among clusters

The **optimization problem:**

$$V^{\mathrm{OPT}} = \min_{P \in \mathcal{P}_q} \max_{y \in \times_{i \in [n]} \{0, w_i\}} y^{\mathrm{T}} \Sigma(P) y$$

- **Difficult to solve:** number of decision variables/comstraints exponential in number of clusters
- **Difficult to interpret and implement:** correlation structure can be complicated

For the Facebook example (cluster sizes $w_i$: $1190, 747, 741, 537, 315, 203, 59$), correlation matrix of the optimal experiment with $q = \frac{1}{2}$ is:

decreasing sizes →

$$\Sigma^* = \begin{bmatrix} 1 & -0.314 & -0.311 & -0.226 & -0.132 & -0.085 & -0.087 \\ -0.314 & 1 & -0.268 & 0 & 0 & 0 & -0.242 \\ -0.311 & -0.268 & 1 & -0.402 & -0.019 & 0 & 0.100 \\ -0.226 & 0 & -0.402 & 1 & 0 & 0 & 0 \\ -0.132 & 0 & -0.019 & 0 & 1 & 0 & 0.237 \\ -0.085 & 0 & 0 & 0 & 0 & 1 & 0 \\ -0.087 & -0.242 & 0.100 & 0 & 0.237 & 0 & 1 \end{bmatrix}$$

↓ decreasing sizes

- Randomize over $36$ possible assignment vectors with different probabilities
- It even deliberately introduces *positive correlation* between pairs

# Warm-up example

When clusters have **equal** "sizes" $w_i$ (WLOG, $w_i = 1$ for all $i \in [n]$):

# Warm-up example

When clusters have **equal** "sizes" $w_i$ (WLOG, $w_i = 1$ for all $i \in [n]$):

**Lemma** *The optimal experiment randomly assigns a fraction* $q$ *of the clusters to treatment. The correlation between assignments of any two clusters is* $\sigma \approx -\frac{1}{n-1}$*. The worst-case variance is* $V^{\mathrm{OPT}} \approx \frac{n}{4}$*.*

# Warm-up example

When clusters have **equal** "sizes" $w_i$ (WLOG, $w_i = 1$ for all $i \in [n]$):

**Lemma** *The optimal experiment randomly assigns a fraction $q$ of the clusters to treatment. The correlation between assignments of any two clusters is $\sigma \approx -\frac{1}{n-1}$. The worst-case variance is $V^{\mathrm{OPT}} \approx \frac{n}{4}$.*

$$\begin{bmatrix} 1 & & & \\ & 1 & & \frac{-1}{n-1} \\ & & 1 & \\ \frac{-1}{n-1} & & & \ddots \\ & & & & 1 \end{bmatrix}$$

correlation matrix of
the optimal experiment

# Warm-up example

When clusters have **equal** "sizes" $w_i$ (WLOG, $w_i = 1$ for all $i \in [n]$):

**Lemma** *The optimal experiment randomly assigns a fraction $q$ of the clusters to treatment. The correlation between assignments of any two clusters is $\sigma \approx -\frac{1}{n-1}$. The worst-case variance is $V^{\mathrm{OPT}} \approx \frac{n}{4}$.*

$$\begin{bmatrix} 1 & & & \frac{-1}{n-1} \\ & 1 & & \\ & & 1 & \\ \frac{-1}{n-1} & & & \ddots \\ & & & & 1 \end{bmatrix} \quad vs. \quad \begin{bmatrix} 1 & & & \varnothing \\ & 1 & & \\ & & 1 & \\ \varnothing & & & \ddots \\ & & & & 1 \end{bmatrix}$$

correlation matrix of the optimal experiment

correlation matrix of independent assignments

**Price of independence:** independently assigning each cluster to treatment has worst-case variance $n \Longrightarrow$ multiplicative gap $= 4$.

# Independent block randomization (IBR)

We consider a family of **independent block randomization** experiments:

1. Sort clusters in (decreasing) sizes $w_i$;
2. Partition clusters into blocks so that each block contains clusters of similar sizes;
3. Randomly treat a fraction $q$ of the clusters in each block, and do so independently across blocks.

$w_1 \quad w_2 \quad w_3 \quad w_4 \qquad w_5 \quad w_6 \quad w_7 \quad w_8 \qquad \cdots \qquad \cdots \qquad \cdots \; w_{n-1} \qquad w_n$

# Independent block randomization (IBR)

We consider a family of **independent block randomization** experiments:

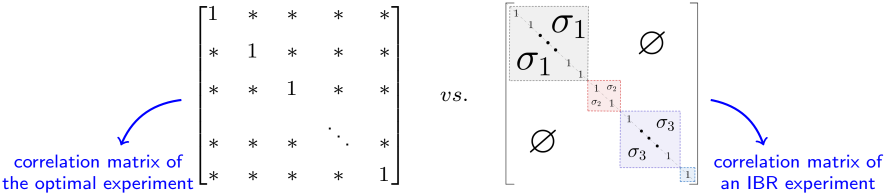
1. Sort clusters in (decreasing) sizes $w_i$;
2. Partition clusters into blocks so that each block contains clusters of similar sizes;
3. Randomly treat a fraction $q$ of the clusters in each block, and do so independently across blocks.

$$\underbrace{w_1 \quad w_2 \quad w_3 \quad w_4}_{\text{block } 1} \quad \underbrace{w_5 \quad w_6 \quad w_7 \quad w_8}_{\text{block } 2} \quad \cdots \quad \underbrace{\cdots \quad \cdots}_{\text{block } k} \quad \cdots \quad \underbrace{w_{n-1} \quad w_n}_{\text{block } K}$$

# Independent block randomization (IBR)

We consider a family of **independent block randomization** experiments:

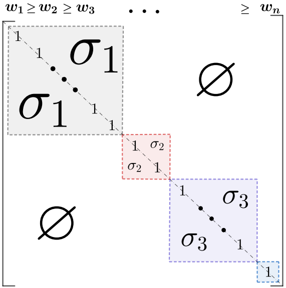
1. Sort clusters in (decreasing) sizes $w_i$;
2. Partition clusters into blocks so that each block contains clusters of similar sizes;
3. Randomly treat a fraction $q$ of the clusters in each block, and do so independently across blocks.




- Simple characterization of the worst-case outcome for a block
- The worst-case variance is additive over blocks

$\Longrightarrow$ Simple dynamic program to compute the optimal partition!

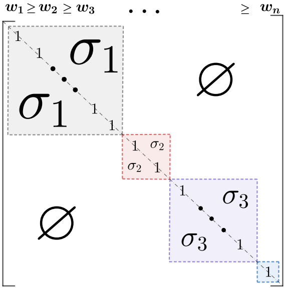
# Performance Trade-off of IBR vs. Optimal




Correlation matrix of an IBR experiment

- **Gain:** a larger negative correlation within a block
- **Losses** come from two sources:
  - independent assignments across blocks
  - ignoring cluster size difference within a block

# Performance Trade-off of IBR vs. Optimal




Correlation matrix of an IBR experiment

- **Gain:** a larger negative correlation within a block
- **Losses** come from two sources:
  - independent assignments across blocks
  - ignoring cluster size difference within a block

Both losses can be small with a careful design of cluster partitions

# Performance analysis

**Theorem** *Let $V^{\mathrm{DP}}$ denote the worst-case variance of the optimal IBR experiment.*

1. *Approximation ratio: for any problem instance,*

$$\frac{V^{\mathrm{DP}}}{V^{\mathrm{OPT}}} \le C(q).$$

*($C(\frac{1}{2}) = \frac{7}{3} \approx 2.33$, $C(\frac{1}{3}) = 2$, $C(\frac{1}{4}) = \frac{7}{3} \approx 2.33$ ...)*

2. *Asymptotic optimality:*

$$\frac{V^{\mathrm{DP}} - V^{\mathrm{OPT}}}{V^{\mathrm{OPT}}} = O\left(\sqrt{\frac{w_1^2}{\sum_{i\in[n]} w_i^2}}\right)$$

# Performance analysis

**Theorem** *Let* $V^{\text{DP}}$ *denote the worst-case variance of the optimal IBR experiment.*

1. *Approximation ratio: for any problem instance,*

$$\frac{V^{\text{DP}}}{V^{\text{OPT}}} \le C(q).$$

*(*$C(\frac{1}{2}) = \frac{7}{3} \approx 2.33$*,* $C(\frac{1}{3}) = 2$*,* $C(\frac{1}{4}) = \frac{7}{3} \approx 2.33$ *...)*

2. *Asymptotic optimality: as* $n \to \infty$ *and if* $w_1^2 = o\Big(\sum_{i\in[n]} w_i^2\Big)$*,*

the largest cluster size

$V^{\text{OPT}} = \Theta\Big(\sum_{i\in[n]} w_i^2\Big)$

$$\frac{V^{\text{DP}} - V^{\text{OPT}}}{V^{\text{OPT}}} = O\left(\sqrt{\frac{w_1^2}{\sum_{i\in[n]} w_i^2}}\right) \to 0.$$

# Performance analysis

**Theorem** *Let $V^{\mathrm{DP}}$ denote the worst-case variance of the optimal IBR experiment.*

1. *Approximation ratio: for any problem instance,*

$$\frac{V^{\mathrm{DP}}}{V^{\mathrm{OPT}}} \leq C(q).$$

$(C(\frac{1}{2}) = \frac{7}{3} \approx 2.33,\ C(\frac{1}{3}) = 2,\ C(\frac{1}{4}) = \frac{7}{3} \approx 2.33\ ...)$

2. *Asymptotic optimality: as $n \to \infty$ and if $w_1^2 = o\Big(\sum_{i\in[n]} w_i^2\Big)$,*

the largest cluster size ($w_1^2$)

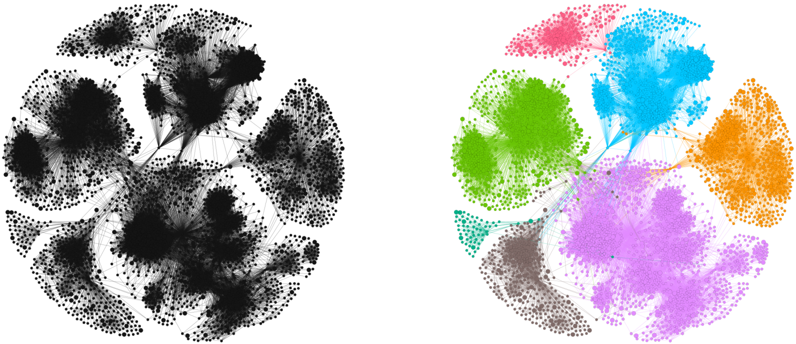
$\Big(\sum_{i\in[n]} w_i^2\Big)$ → $V^{\mathrm{OPT}} = \Theta\Big(\sum_{i\in[n]} w_i^2\Big)$

$$\frac{V^{\mathrm{DP}} - V^{\mathrm{OPT}}}{V^{\mathrm{OPT}}} = O\left(\sqrt{\frac{w_1^2}{\sum_{i\in[n]} w_i^2}}\right) \to 0.$$

- Performance is substantially better in numerical study.
- Stronger results for more specific settings.

# Back to Facebook example

- Partition the network using the classic Louvain algorithm, and merge two clusters if one contaminates more than $10\%$ of the second one.
- $n = 7$ clusters of different sizes; "contaminated" users $\sim 6\%$.
- Consider marginal assignment probability $q = \frac{1}{2}$.



Source: Stanford SNAP Datasets

# Facebook example

- **Reminder**: The optimal cluster-based assignment is difficult to solve and has a complicated correlation structure:
  - Randomizes over $36$ possible assignment vectors with different probabilities.
  - Even deliberately introduces positive correlation between small clusters.

$$\Sigma_c^* = \begin{bmatrix} 1 & -0.314 & -0.311 & -0.226 & -0.132 & -0.085 & -0.087 \\ -0.314 & 1 & -0.268 & 0 & 0 & 0 & -0.242 \\ -0.311 & -0.268 & 1 & -0.402 & -0.019 & 0 & 0.100 \\ -0.226 & 0 & -0.402 & 1 & 0 & 0 & 0 \\ -0.132 & 0 & -0.019 & 0 & 1 & 0 & 0.237 \\ -0.085 & 0 & 0 & 0 & 0 & 1 & 0 \\ -0.087 & -0.242 & 0.100 & 0 & 0.237 & 0 & 1 \end{bmatrix}$$

# Facebook example

- **Reminder**: The optimal cluster-based assignment is difficult to solve and has a complicated correlation structure:
  - Randomizes over $36$ possible assignment vectors with different probabilities.
  - Even deliberately introduces positive correlation between small clusters.

- The optimal IBR experiment has a simple structure:

$$\Sigma^{\mathrm{DP}} = \begin{bmatrix} 1 & -1/3 & -1/3 & -1/3 & 0 & 0 & 0 \\ -1/3 & 1 & -1/3 & -1/3 & 0 & 0 & 0 \\ -1/3 & -1/3 & 1 & -1/3 & 0 & 0 & 0 \\ -1/3 & -1/3 & -1/3 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1/3 & -1/3 \\ 0 & 0 & 0 & 0 & -1/3 & 1 & -1/3 \\ 0 & 0 & 0 & 0 & -1/3 & -1/3 & 1 \end{bmatrix}$$

# Facebook example

- **Reminder**: The optimal cluster-based assignment is difficult to solve and has a complicated correlation structure:
  - Randomizes over $36$ possible assignment vectors with different probabilities.
  - Even deliberately introduces positive correlation between small clusters.

- The optimal IBR experiment has a simple structure:

$$\Sigma^{\mathrm{DP}} = \begin{bmatrix} 1 & -1/3 & -1/3 & -1/3 & 0 & 0 & 0 \\ -1/3 & 1 & -1/3 & -1/3 & 0 & 0 & 0 \\ -1/3 & -1/3 & 1 & -1/3 & 0 & 0 & 0 \\ -1/3 & -1/3 & -1/3 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1/3 & -1/3 \\ 0 & 0 & 0 & 0 & -1/3 & 1 & -1/3 \\ 0 & 0 & 0 & 0 & -1/3 & -1/3 & 1 \end{bmatrix}$$

- The optimal IBR experiment: $(V^{\mathrm{DP}} - V^{\mathrm{OPT}})/V^{\mathrm{OPT}} = 7\%$.
- Randomly treating half of the clusters: $(V^{\mathrm{half}} - V^{\mathrm{OPT}})/V^{\mathrm{OPT}} = 31.3\%$.
- Pair matching experiment: $(V^{\mathrm{pair}} - V^{\mathrm{OPT}})/V^{\mathrm{OPT}} = 46.0\%$.
- Independent cluster-based assignment: $(V^{\mathrm{ind}} - V^{\mathrm{OPT}})/V^{\mathrm{OPT}} = 108.5\%$.

# Takeaways and future directions

- We study robust experimental design for cluster-based randomization.
- We develop simple IBR experiments that (i) attain good approximation ratio and (ii) are asymptotically optimal with many clusters under mild conditions.
- **Operational takeaway:** collecting similar clusters together and randomly treating a fraction $q$ in each block is near-optimal.
- Future work:
  - Optimal bias-variance trade-off
  - Careful empirical study based on real data

**Reference:** O. Candogan, C. Chen, and R. Niazadeh, "Correlated Cluster-Based Randomized Experiments: Robust Variance Minimization." *Management Science* (forthcoming).

Working paper available at `https://papers.ssrn.com/abstract=3852100`

# Appendix

# Related Literature—CS/ECON/OR/STAT

**Experimental design in networks:**

- General networks: [Eckles et al., 2016],[Aronow et al., 2017],[Ugander and Yin, 2020]
- Bipartite networks: [Zigler and Papadogeorgou, 2018], [Pouget-Abadie et al., 2019], [Doudchenko et al., 2020],[Harshaw et al., 2021]

**Experimental design in online platforms:**

- Analysis of bias: [Johari et al., 2020]
- Analysis of marketplace equilibrium: [Wager and Xu, 2021]
- Switchback experiments: [Bojinov et al., 2020],[Glynn et al., 2020]
- Relevant empirical work: [Ostrovsky and Schwarz, 2011],[Blake and Coey, 2014],[Zhang et al., 2020],[Holtz et al., 2020],[Holtz and Aral, 2020]

**Models of potential outcomes:**

- Covariate model: [Bertsimas et al., 2015],[Bertsimas et al. 2019],[Kallus, 2018],[Bhat et al., 2020],[Harshaw et al., 2019], etc.
- Worst-case potential outcomes: [Bojinov et al., 2020]
  *(we follow the same framework)*

**And many more!**