Detecting Network Effects
Randomizing Over Randomized Experiments

Martin Saveski
MIT

Jean Pouget-Abadie
Harvard

Guillaume Saint-Jacques
MIT

Weitao Duan
LinkedIn

Souvik Ghosh
LinkedIn

Ya Xu
LinkedIn

Edo Airoldi
Harvard
Treatment

\[ Z_i = 1 \]

New Feed

Ranking Algorithm

→

User
Treatment

\[ Z_i = 1 \]

New Feed Ranking Algorithm

Control

\[ Z_j = 0 \]

Old Feed Ranking Algorithm
Treatment

\[ Z_i = 1 \]

New Feed
Ranking Algorithm

Control

\[ Z_j = 0 \]

Old Feed
Ranking Algorithm
Treatment
\[ Z_i = 1 \]

New Feed Ranking Algorithm

Control
\[ Z_j = 0 \]

Old Feed Ranking Algorithm

Engagement
\[ Y_i \]
Treatment

\[ Z_i = 1 \]

New Feed Ranking Algorithm

\[ Y_i \]

Engagement

Control

\[ Z_j = 0 \]

Old Feed Ranking Algorithm
Treatment

\[ Z_i = 1 \]

New Feed Ranking Algorithm

Engagement

\[ Y_i \]

Control

\[ Z_j = 0 \]

Old Feed Ranking Algorithm

Engagement

\[ Y_j \]
Completely-randomized Experiment
Completely-randomized Experiment

Treatment (B)
Completely-randomized Experiment
Completely-randomized Experiment

\[ \mu = \frac{\sum Y(\bigcirc)}{\big| \bigcirc \big|} - \frac{\sum Y(\bigcirc)}{\big| \bigcirc \big|} \]

- Control (A)
- Treatment (B)
\[ \mu = \frac{\sum Y(\bigcirc)}{\bigcirc} - \frac{\sum Y(\bigcirc)}{\bigcirc} \]

**SUTVA**: Stable Unit Treatment Value Assumption

Every user’s behavior is affected only by their treatment and NOT by the treatment of any other user

**Completely-randomized Experiment**
Cluster-based Randomized Experiment
Cluster-based Randomized Experiment
Cluster-based Randomized Experiment
Cluster-based Randomized Experiment

Control (A)

Treatment (B)
Completely-randomized Experiment

Cluster-based Randomized Experiment

More Spillovers
Lower Variance

Less Spillovers
Higher Variance
Design for Detecting Network Effects
Completely Randomized Experiment
Cluster-based Randomized Experiment

Completely Randomized Experiment

Cluster-based Randomized Experiment
Cluster-based Randomized Experiment

Completely Randomized Experiment

μ_{completely-randomized} \ ? \ μ_{cluster-based}
Hypothesis Test
Hypothesis Test

$H_0$: SUTVA Holds
Hypothesis Test

$H_0$: SUTVA Holds

$$E_{W,Z} [\hat{\mu}_{cbr} - \hat{\mu}_{cr}] = 0$$
Hypothesis Test

\( H_0: \text{SUTVA Holds} \)

\[
E_{W,Z}[\hat{\mu}_{cbr} - \hat{\mu}_{cr}] = 0
\]

\[
\text{var}_{W,Z}[\hat{\mu}_{cr} - \hat{\mu}_{cbr}] \leq E_{W,Z}[\hat{\sigma}^2]
\]
Hypothesis Test

\( \mathbf{H}_0: \text{SUTVA Holds} \)

\[
E_{\mathbf{w}, \mathbf{z}} [\hat{\mu}_{cbr} - \hat{\mu}_{cr}] = 0
\]

\[
\text{var}_{\mathbf{w}, \mathbf{z}} [\hat{\mu}_{cr} - \hat{\mu}_{cbr}] \leq E_{\mathbf{w}, \mathbf{z}} [\hat{\sigma}^2]
\]

Reject the null when:
Hypothesis Test

$H_0$: SUTVA Holds

$$E_{\mathbf{w}, \mathbf{z}} [\hat{\mu}_{cbr} - \hat{\mu}_{cr}] = 0$$

$$\text{var}_{\mathbf{w}, \mathbf{z}} [\hat{\mu}_{cr} - \hat{\mu}_{cbr}] \leq E_{\mathbf{w}, \mathbf{z}}[\hat{\sigma}^2]$$

Reject the null when:

$$\frac{|\hat{\mu}_{cr} - \hat{\mu}_{cbr}|}{\sqrt{\hat{\sigma}^2}} \geq \frac{1}{\sqrt{\alpha}}$$
Hypothesis Test

\( H_0: \text{SUTVA Holds} \)

\[
E_{\mathbf{w}, \mathbf{z}} [\hat{\mu}_{cbr} - \hat{\mu}_{cr}] = 0
\]

\[
\text{var}_{\mathbf{w}, \mathbf{z}} [\hat{\mu}_{cr} - \hat{\mu}_{cbr}] \leq E_{\mathbf{w}, \mathbf{z}} [\hat{\sigma}^2]
\]

Reject the null when:

\[
\left| \frac{\hat{\mu}_{cr} - \hat{\mu}_{cbr}}{\sqrt{\hat{\sigma}^2}} \right| \geq \frac{1}{\sqrt{\alpha}}
\]

Type I error is no greater than \( \alpha \)
Nuts and Bolts of Running
Cluster-based Randomized Experiments
Why Balanced Clustering?
Why Balanced Clustering?

• Theoretical Motivation
  – Constants VS random variables
Why Balanced Clustering?

• Theoretical Motivation
  – Constants VS random variables

• Practical Motivations
Why Balanced Clustering?

- Theoretical Motivation
  - Constants VS random variables

- Practical Motivations
  - Variance reduction
Why Balanced Clustering?

• Theoretical Motivation
  – Constants VS random variables

• Practical Motivations
  – Variance reduction
  – Balance on pre-treatment covariates
    (homophily => large homogenous clusters)
Algorithms for Balanced Clustering
Most clustering methods find skewed distributions of cluster sizes (Leskovec, 2009; Fortunato, 2010)
Algorithms for Balanced Clustering

Most clustering methods find skewed distributions of cluster sizes
(Leskovec, 2009; Fortunato, 2010)

=> Algorithms that enforce equal cluster sizes
Algorithms for Balanced Clustering

Most clustering methods find skewed distributions of cluster sizes
(Leskovec, 2009; Fortunato, 2010)

=>$\text{Algorithms that enforce equal cluster sizes}$

Restreaming Linear Deterministic Greedy
(Nishimura & Ugander, 2013)
Algorithms for Balanced Clustering

Most clustering methods find skewed distributions of cluster sizes
(Leskovec, 2009; Fortunato, 2010)

=> Algorithms that enforce equal cluster sizes

Restreaming Linear Deterministic Greedy
(Nishimura & Ugander, 2013)

– Streaming
– Parallelizable
– Stable
Clustering the LinkedIn Graph

- Graph: >100M nodes, >10B edges
- 350 Hadoop nodes
- 1% leniency
Clustering the LinkedIn Graph

- Graph: >100M nodes, >10B edges
- 350 Hadoop nodes
- 1% leniency
Clustering the LinkedIn Graph

- Graph: >100M nodes, >10B edges
- 350 Hadoop nodes
- 1% leniency

% edges within clusters

<table>
<thead>
<tr>
<th>k</th>
<th>% edges within clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>35.6%</td>
</tr>
<tr>
<td>3000</td>
<td></td>
</tr>
<tr>
<td>5000</td>
<td></td>
</tr>
<tr>
<td>7000</td>
<td></td>
</tr>
<tr>
<td>10000</td>
<td></td>
</tr>
</tbody>
</table>
Clustering the LinkedIn Graph

- Graph: >100M nodes, >10B edges
- 350 Hadoop nodes
- 1% leniency
Choosing the Number of Clusters
Choosing the Number of Clusters

small $k$  
large $k$
Choosing the Number of Clusters

small $k$  
large clusters  

large $k$  
small clusters
Choosing the Number of Clusters

small $k$ vs. large $k$

- large clusters
  - large network effect
  - large variance
- small clusters
  - small network effect
  - small variance
Choosing the Number of Clusters

Understanding the Type II error
Choosing the Number of Clusters
Understanding the Type II error

Assuming an interference model
Choosing the Number of Clusters
Understanding the Type II error

Assuming an interference model

\[ Y_i = \beta_0 + \beta_1 Z_i + \beta_2 \rho_i + \epsilon_i \]

\( \rho_i \): fraction of treated friends
Choosing the Number of Clusters

Understanding the Type II error

Assuming an interference model

\[ Y_i = \beta_0 + \beta_1 Z_i + \beta_2 \rho_i + \epsilon_i \]

\( \rho_i \): fraction of treated friends

\[ E [\hat{\mu}_{cbr} - \hat{\mu}_{cr}] \approx \rho \cdot \beta_2 \]

\( \rho \): average fraction of a unit's neighbors contained in the cluster
Choosing the Number of Clusters
Understanding the Type II error

Assuming an interference model

\[ Y_i = \beta_0 + \beta_1 Z_i + \beta_2 \rho_i + \epsilon_i \]

\( \rho_i \): fraction of treated friends

\[ E [\hat{\mu}_{cbr} - \hat{\mu}_{cr}] \approx \rho \cdot \beta_2 \]

\( \rho \): average fraction of a unit's neighbors contained in the cluster

Choose number of clusters \( M \) and clustering \( C \) such that

\[ \max_{M,C} \frac{\rho}{\sqrt{\hat{\sigma}^2_C}} \]
Experiments on LinkedIn
Cluster-based Randomized Experiment

\( \mu_{\text{completely-randomized}} \stackrel{?}{=} \mu_{\text{cluster-based}} \)
Experiment 1
Experiment 1

- Population: 20% of all LinkedIn users [Bernoulli: 10%, Cluster-based: 10%]
Experiment 1

- **Population**: 20% of all LinkedIn users [Bernoulli: 10%, Cluster-based: 10%]
- **Time period**: 2 weeks
Experiment 1

- Population: 20% of all LinkedIn users [Bernoulli: 10%, Cluster-based: 10%]
- Time period: 2 weeks
- Number of clusters: $k = 3,000$
Experiment 1

- Population: 20% of all LinkedIn users [Bernoulli: 10%, Cluster-based: 10%]
- Time period: 2 weeks
- Number of clusters: \( k = 3,000 \)
- Outcome: feed engagement
Experiment 1

- **Population**: 20% of all LinkedIn users [Bernoulli: 10%, Cluster-based: 10%]
- **Time period**: 2 weeks
- **Number of clusters**: $k = 3,000$
- **Outcome**: feed engagement

<table>
<thead>
<tr>
<th>Treatment effect</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernoulli Randomization (BR)</td>
<td>0.0559</td>
</tr>
<tr>
<td>Cluster-based Randomization (CBR)</td>
<td>0.0771</td>
</tr>
<tr>
<td>Delta (CBR – BR)</td>
<td>-0.0211</td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td>0.4246</td>
</tr>
</tbody>
</table>
Experiment 1

- **Population:** 20% of all LinkedIn users [Bernoulli: 10%, Cluster-based: 10%]
- **Time period:** 2 weeks
- **Number of clusters:** $k = 3,000$
- **Outcome:** feed engagement

<table>
<thead>
<tr>
<th></th>
<th>Treatment effect</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernoulli Randomization (BR)</td>
<td>0.0559</td>
<td>0.0050</td>
</tr>
<tr>
<td>Cluster-based Randomization (CBR)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delta (CBR – BR)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Experiment 1**

- **Population:** 20% of all LinkedIn users [Bernoulli: 10%, Cluster-based: 10%]
- **Time period:** 2 weeks
- **Number of clusters:** $k = 3,000$
- **Outcome:** feed engagement

<table>
<thead>
<tr>
<th></th>
<th>Treatment effect</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernoulli Randomization (BR)</td>
<td>0.0559</td>
<td>0.0050</td>
</tr>
<tr>
<td>Cluster-based Randomization (CBR)</td>
<td>0.0771</td>
<td>0.0260</td>
</tr>
<tr>
<td>Delta (CBR – BR)</td>
<td>-0.0211</td>
<td>0.0265</td>
</tr>
</tbody>
</table>

**p-value:** 0.4246
Experiment 1

- Population: 20% of all LinkedIn users [Bernoulli: 10%, Cluster-based: 10%]
- Time period: 2 weeks
- Number of clusters: $k = 3,000$
- Outcome: feed engagement

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Treatment effect</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernoulli Randomization (BR)</td>
<td>0.0559</td>
<td>0.0050</td>
</tr>
<tr>
<td>Cluster-based Randomization (CBR)</td>
<td>0.0771</td>
<td>0.0260</td>
</tr>
<tr>
<td>Delta (CBR – BR)</td>
<td>-0.0211</td>
<td>0.0265</td>
</tr>
</tbody>
</table>
Experiment 1

- Population: 20% of all LinkedIn users [Bernoulli: 10%, Cluster-based: 10%]
- Time period: 2 weeks
- Number of clusters: $k = 3,000$
- Outcome: feed engagement

<table>
<thead>
<tr>
<th></th>
<th>Treatment effect</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernoulli Randomization (BR)</td>
<td>0.0559</td>
<td>0.0050</td>
</tr>
<tr>
<td>Cluster-based Randomization (CBR)</td>
<td>0.0771</td>
<td>0.0260</td>
</tr>
<tr>
<td>Delta (CBR – BR)</td>
<td>-0.0211</td>
<td>0.0265</td>
</tr>
</tbody>
</table>

$p$-value: 0.4246
Experiment 2
Experiment 2

- Population: 36% of all LinkedIn users [Bernoulli: 20%, Cluster-based: 16%]
Experiment 2

- Population: 36% of all LinkedIn users [Bernoulli: 20%, Cluster-based: 16%]
- Time period: 4 weeks
Experiment 2

- Population: 36% of all LinkedIn users [Bernoulli: 20%, Cluster-based: 16%]
- Time period: 4 weeks
- Number of clusters: $k = 10,000$
Experiment 2

- Population: 36% of all LinkedIn users [Bernoulli: 20%, Cluster-based: 16%]
- Time period: 4 weeks
- Number of clusters: $k = 10,000$
- Outcome: feed engagement
**Experiment 2**

- **Population:** 36% of all LinkedIn users [Bernoulli: 20%, Cluster-based: 16%]
- **Time period:** 4 weeks
- **Number of clusters:** $k = 10,000$
- **Outcome:** feed engagement

<table>
<thead>
<tr>
<th>Treatment effect</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernoulli Randomization (BR)</td>
<td></td>
</tr>
<tr>
<td>Cluster-based Randomization (CBR)</td>
<td></td>
</tr>
<tr>
<td>Delta (CBR – BR)</td>
<td></td>
</tr>
</tbody>
</table>

- **p-value:** 0.0483
Experiment 2

- Population: 36% of all LinkedIn users [Bernoulli: 20%, Cluster-based: 16%]
- Time period: 4 weeks
- Number of clusters: $k = 10,000$
- Outcome: feed engagement

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Treatment effect</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernoulli Randomization (BR)</td>
<td>0.2108</td>
<td>0.2911</td>
</tr>
<tr>
<td>Cluster-based Randomization (CBR)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delta (CBR – BR)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Experiment 2

- Population: 36% of all LinkedIn users [Bernoulli: 20%, Cluster-based: 16%]
- Time period: 4 weeks
- Number of clusters: $k = 10,000$
- Outcome: feed engagement

<table>
<thead>
<tr>
<th></th>
<th>Treatment effect</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernoulli Randomization (BR)</td>
<td>0.2108</td>
<td>0.2911</td>
</tr>
<tr>
<td>Cluster-based Randomization (CBR)</td>
<td>0.5390</td>
<td>0.5613</td>
</tr>
<tr>
<td>Delta (CBR – BR)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Experiment 2

- Population: 36% of all LinkedIn users [Bernoulli: 20%, Cluster-based: 16%]
- Time period: 4 weeks
- Number of clusters: $k = 10,000$
- Outcome: feed engagement

<table>
<thead>
<tr>
<th>Treatment Method</th>
<th>Treatment effect</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernoulli Randomization (BR)</td>
<td>0.2108</td>
<td>0.2911</td>
</tr>
<tr>
<td>Cluster-based Randomization (CBR)</td>
<td>0.5390</td>
<td>0.5613</td>
</tr>
<tr>
<td>Delta (CBR – BR)</td>
<td>-0.3281</td>
<td>0.5712</td>
</tr>
</tbody>
</table>
Experiment 2

- Population: 36% of all LinkedIn users [Bernoulli: 20%, Cluster-based: 16%]
- Time period: 4 weeks
- Number of clusters: \( k = 10,000 \)
- Outcome: feed engagement

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Treatment effect</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernoulli Randomization (BR)</td>
<td>0.2108</td>
<td>0.2911</td>
</tr>
<tr>
<td>Cluster-based Randomization (CBR)</td>
<td>0.5390</td>
<td>0.5613</td>
</tr>
<tr>
<td>Delta (CBR – BR)</td>
<td>-0.3281</td>
<td>0.5712</td>
</tr>
</tbody>
</table>

p-value: 0.0483
Test SUTVA null
Test SUTVA null

reject
Use cluster-based experiment to estimate treatment effects
Test SUTVA null

reject

Use cluster-based experiment to estimate treatment effects

(higher variance)
Test SUTVA null

- reject
- fail to reject

Use cluster-based experiment to estimate treatment effects

(higher variance)
Test SUTVA null

- Reject: Use cluster-based experiment to estimate treatment effects (higher variance)
- Fail to reject: Use Bernoulli experiment to estimate treatment effects
Test SUTVA null

reject

Use cluster-based experiment to estimate treatment effects
(higher variance)

fail to reject

Use Bernoulli experiment to estimate treatment effects
(lower variance)
Papers available online

KDD’17
Arxiv
Detecting Network Effects
Randomizing Over Randomized Experiments

Martin Saveski
(@msaveski)
MIT