

Detecting Network Effects

Randomizing Over Randomized Experiments

Martin Saveski

(@msaveski)

MIT

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

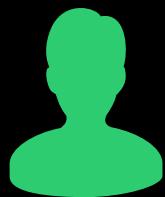


Edo Airoldi
Harvard

Treatment

$$Z_i = 1$$

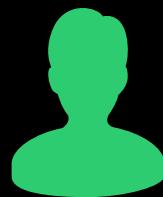
New Feed
Ranking Algorithm



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

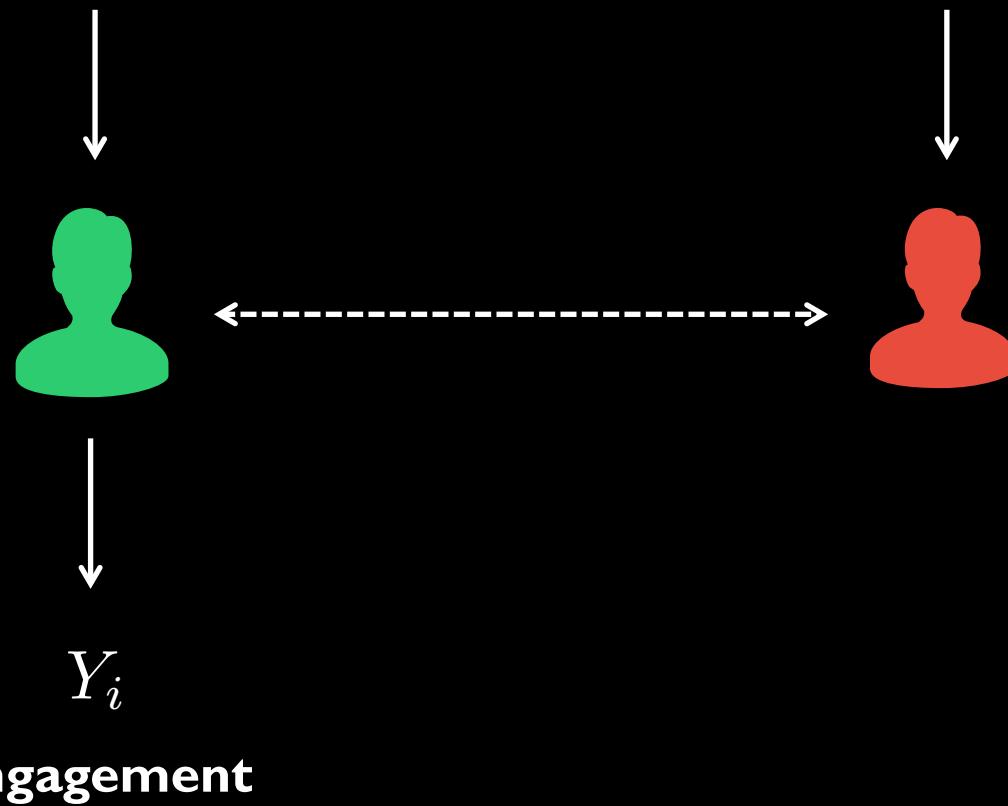
$$Z_i = 1$$

New Feed
Ranking Algorithm

Control

$$Z_j = 0$$

Old Feed
Ranking Algorithm



Treatment

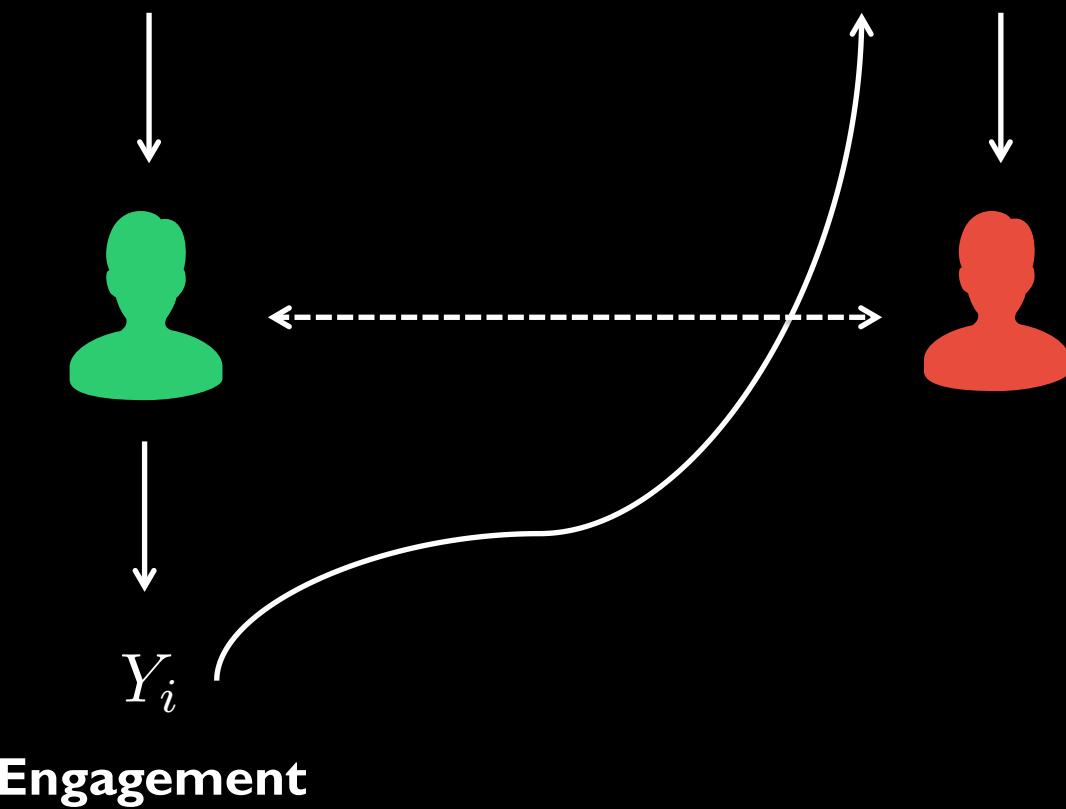
$$Z_i = 1$$

New Feed
Ranking Algorithm

Control

$$Z_j = 0$$

Old Feed
Ranking Algorithm



Treatment

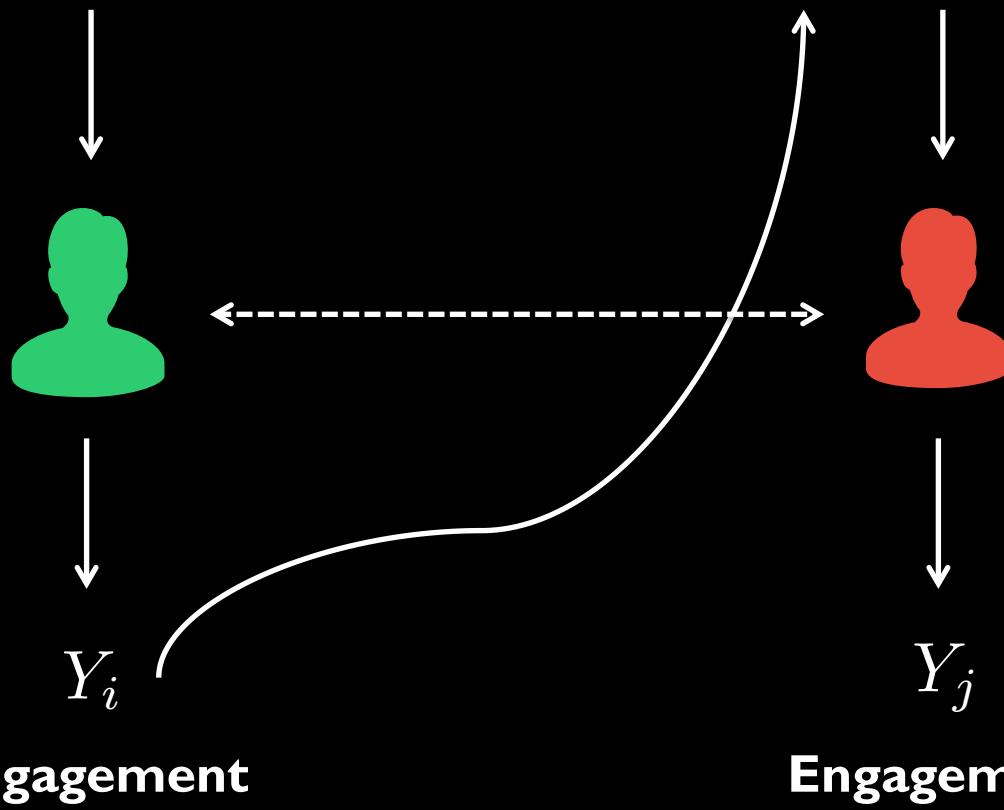
$$Z_i = 1$$

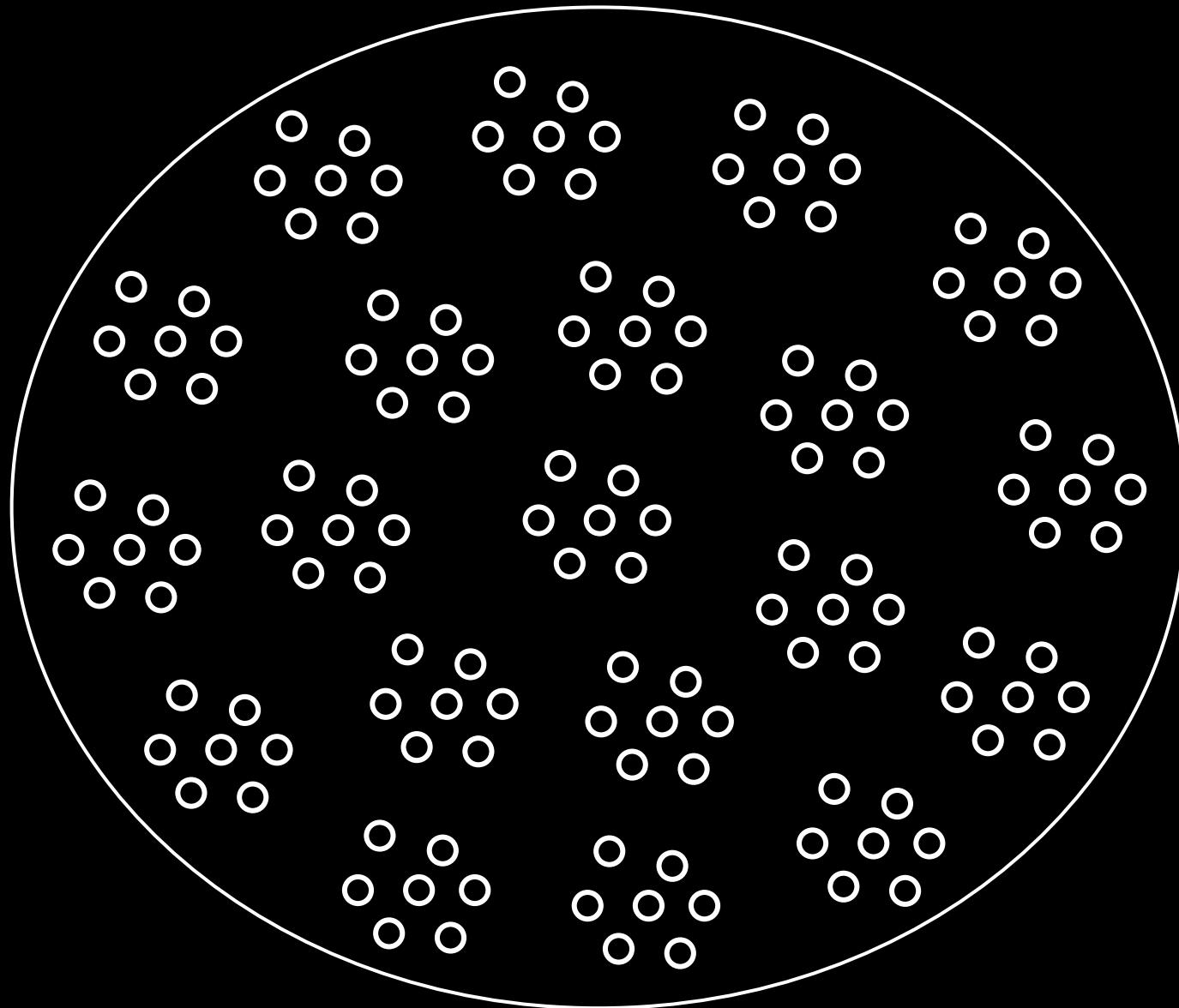
New Feed
Ranking Algorithm

Control

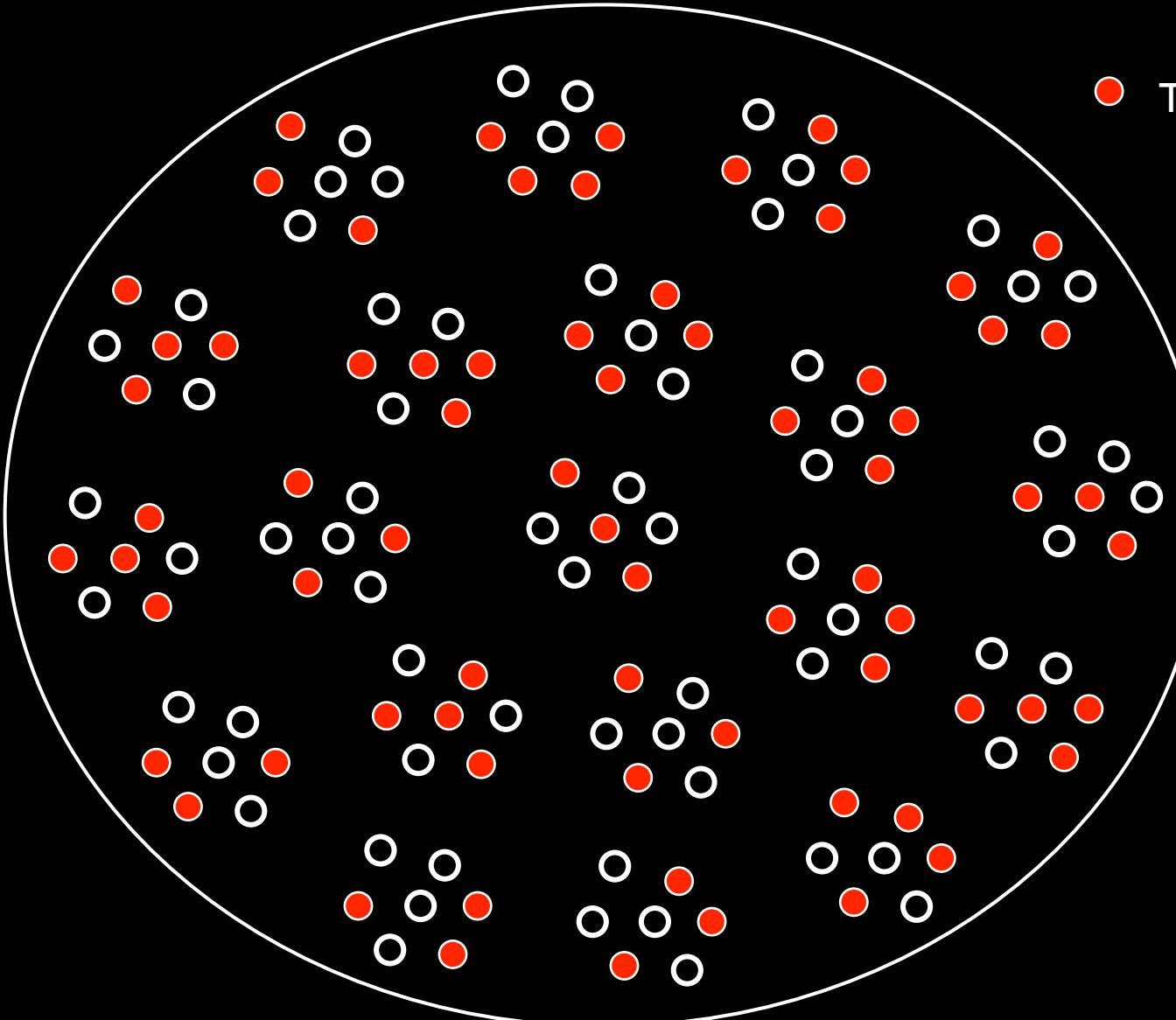
$$Z_j = 0$$

Old Feed
Ranking Algorithm



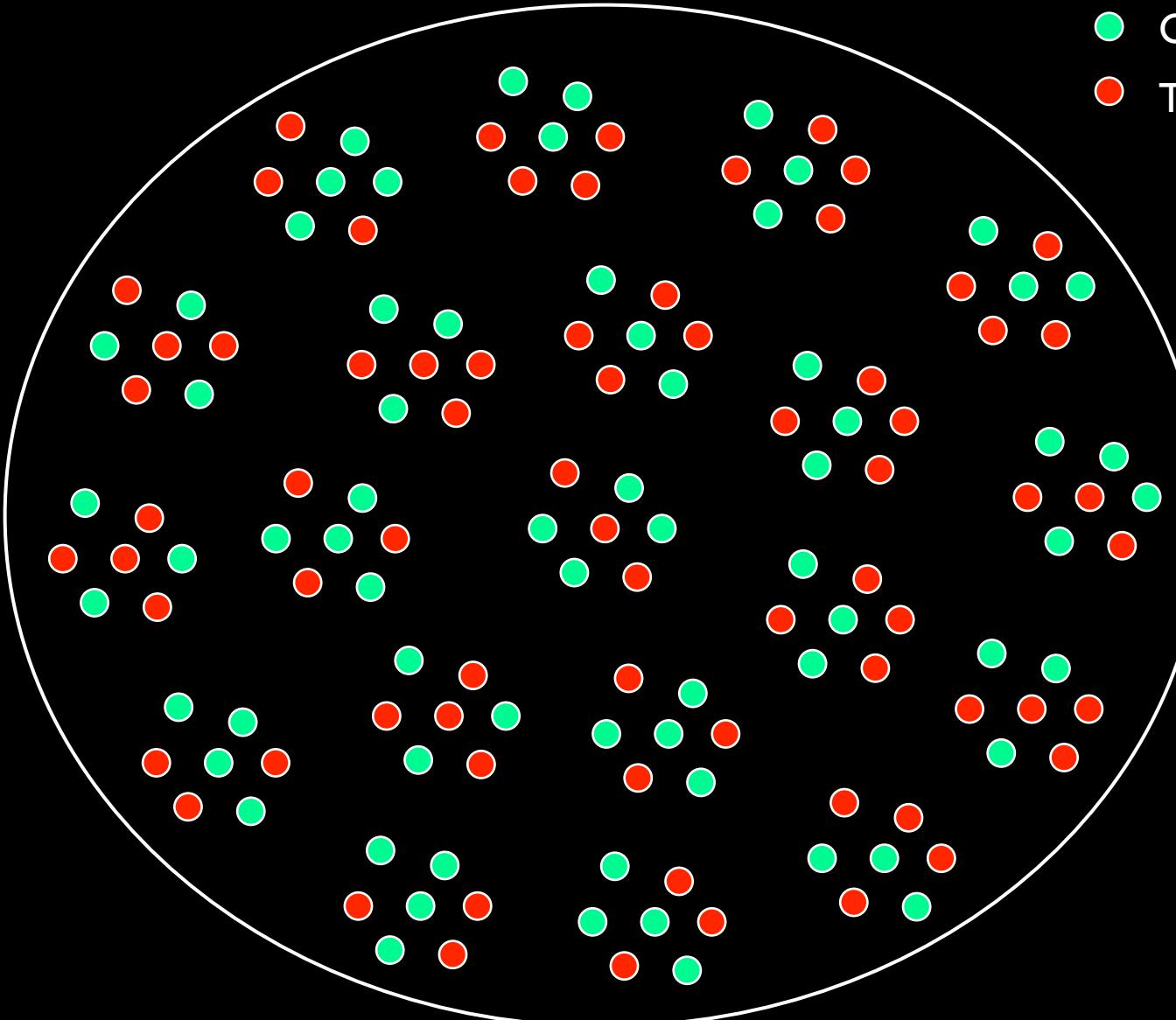


Completely-randomized Experiment



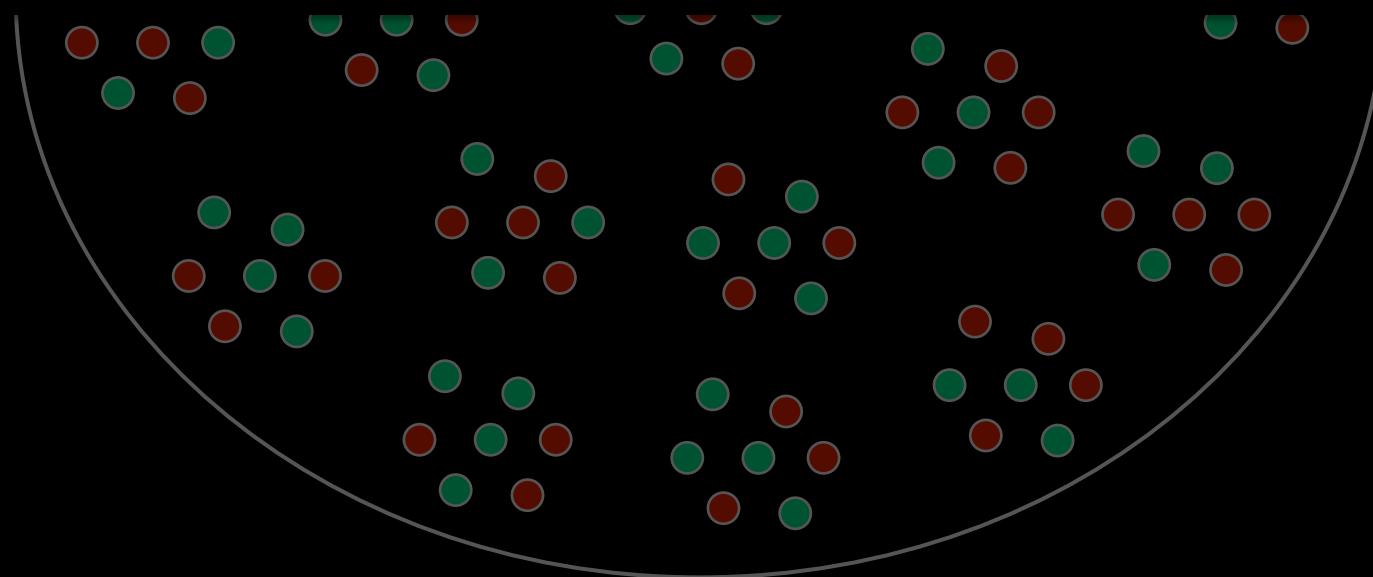
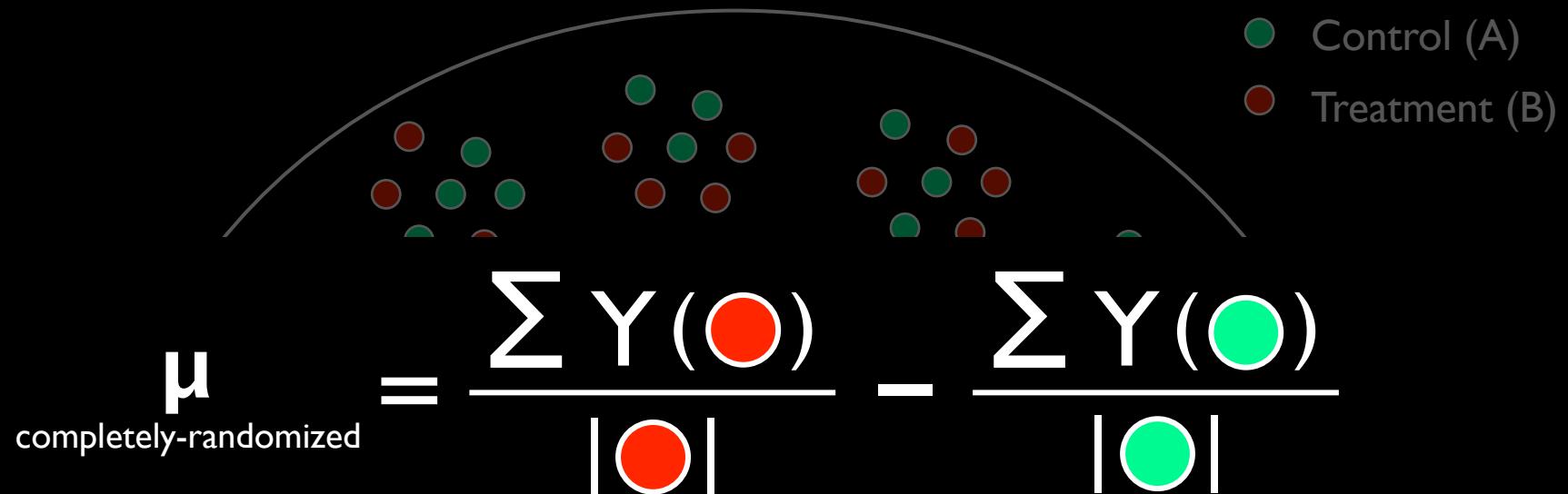
Treatment (B)

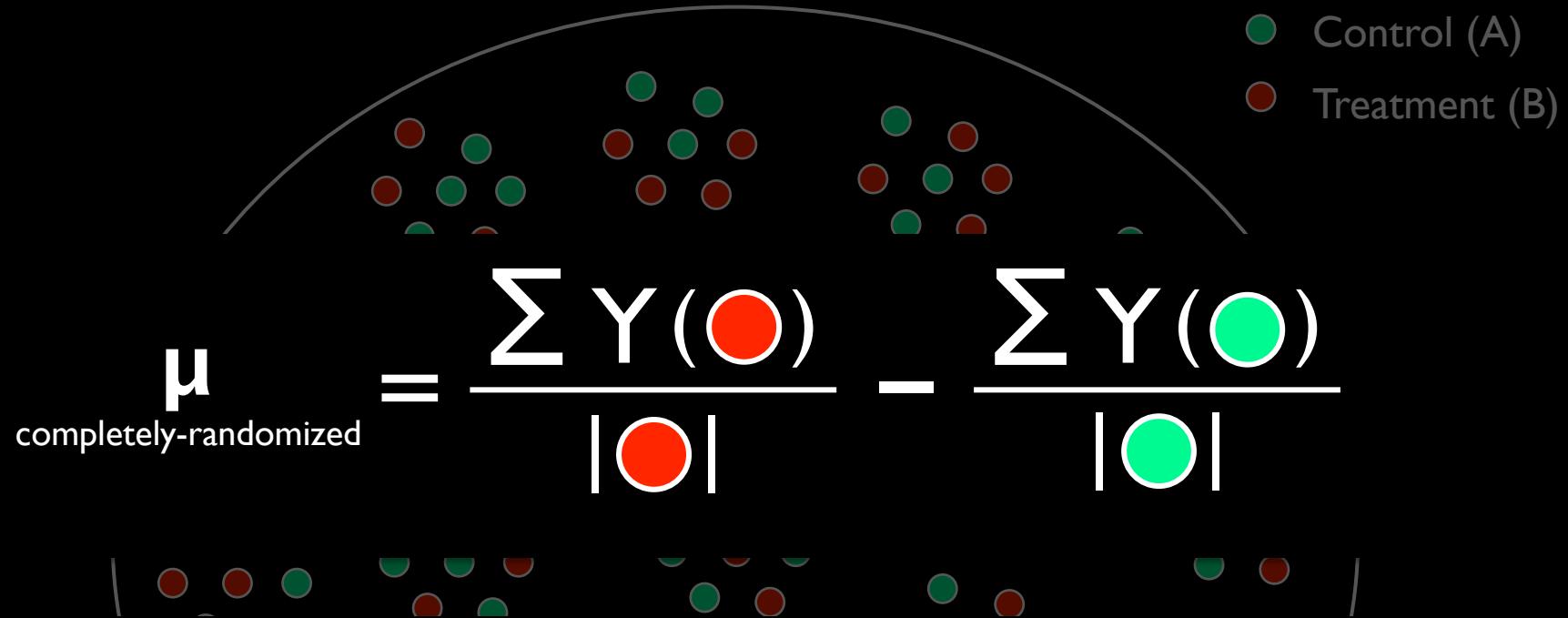
Completely-randomized Experiment



● Control (A)
● Treatment (B)

Completely-randomized Experiment



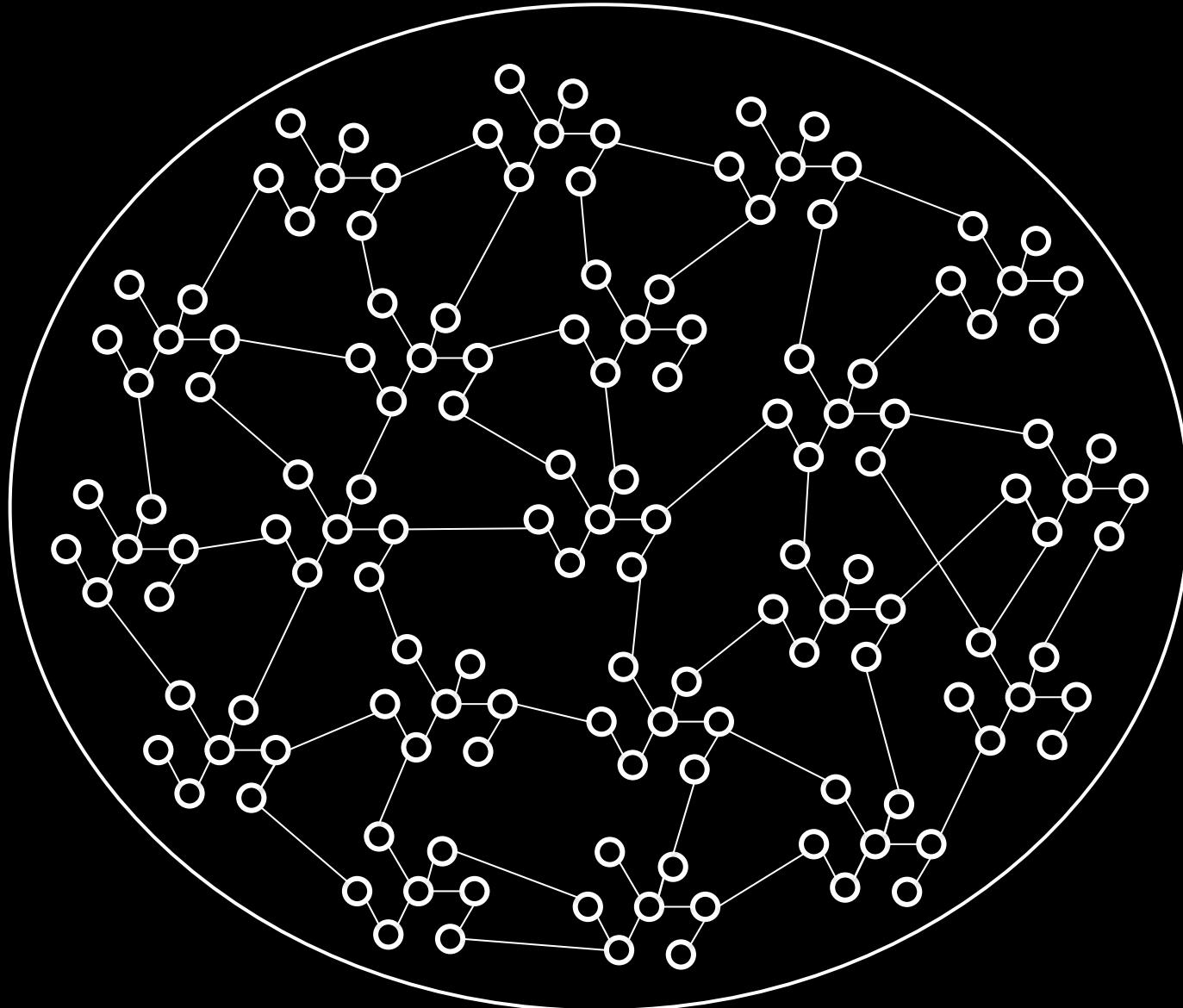


SUTVA: Stable Unit Treatment Value Assumption

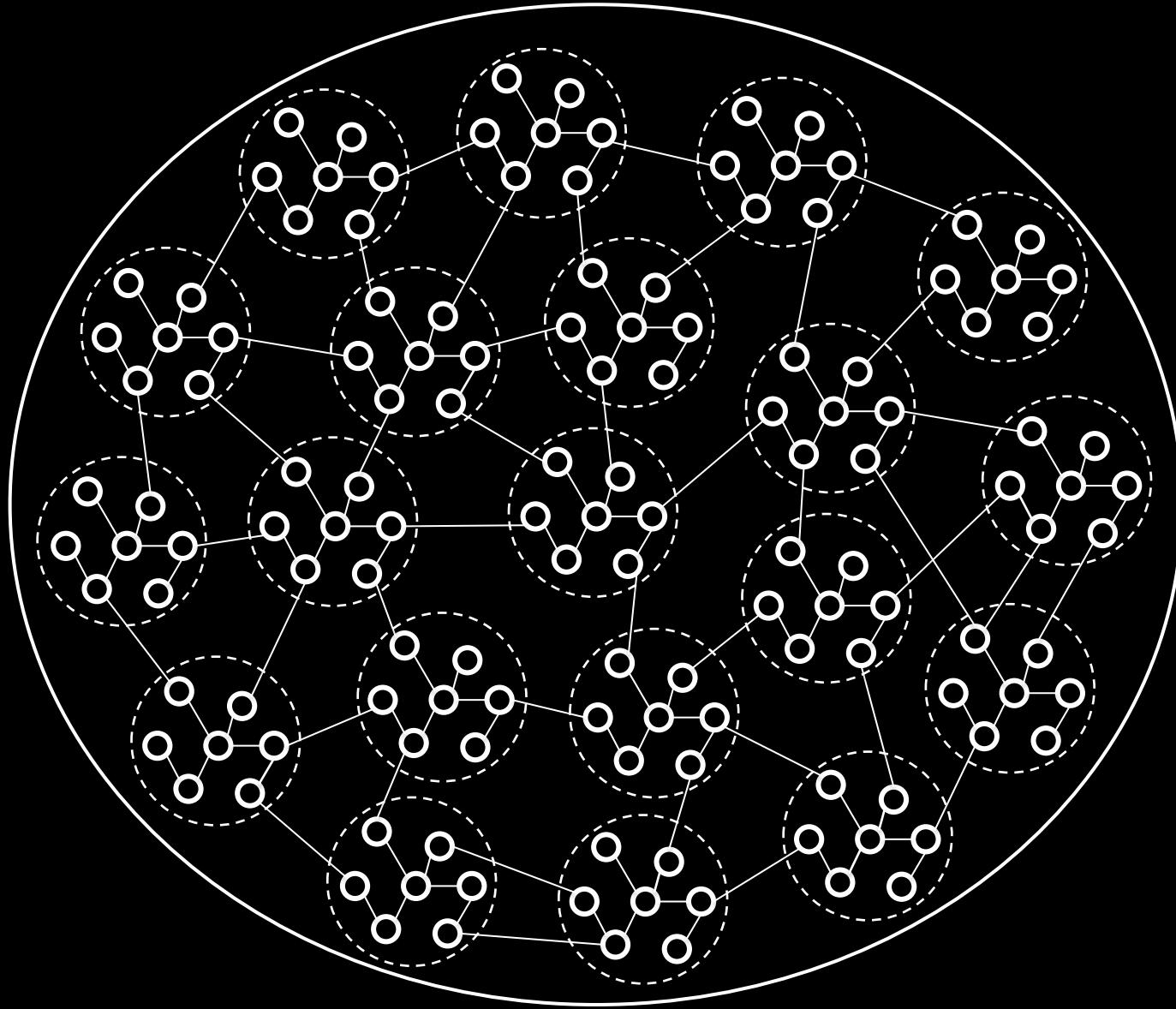
Every user's behavior is affected only by their treatment
and NOT by the treatment of any other user

Completely-randomized Experiment

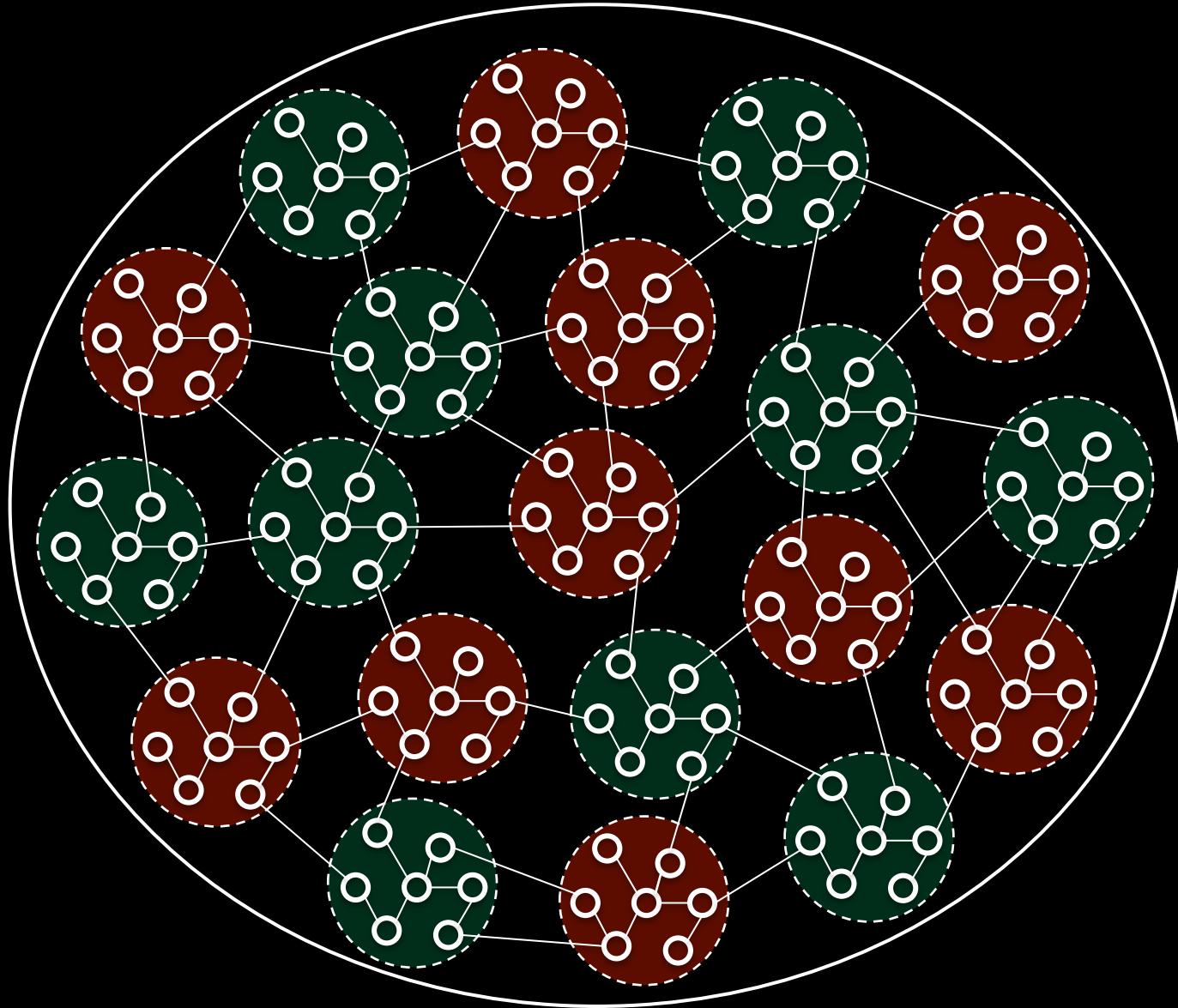
Cluster-based Randomized Experiment



Cluster-based Randomized Experiment

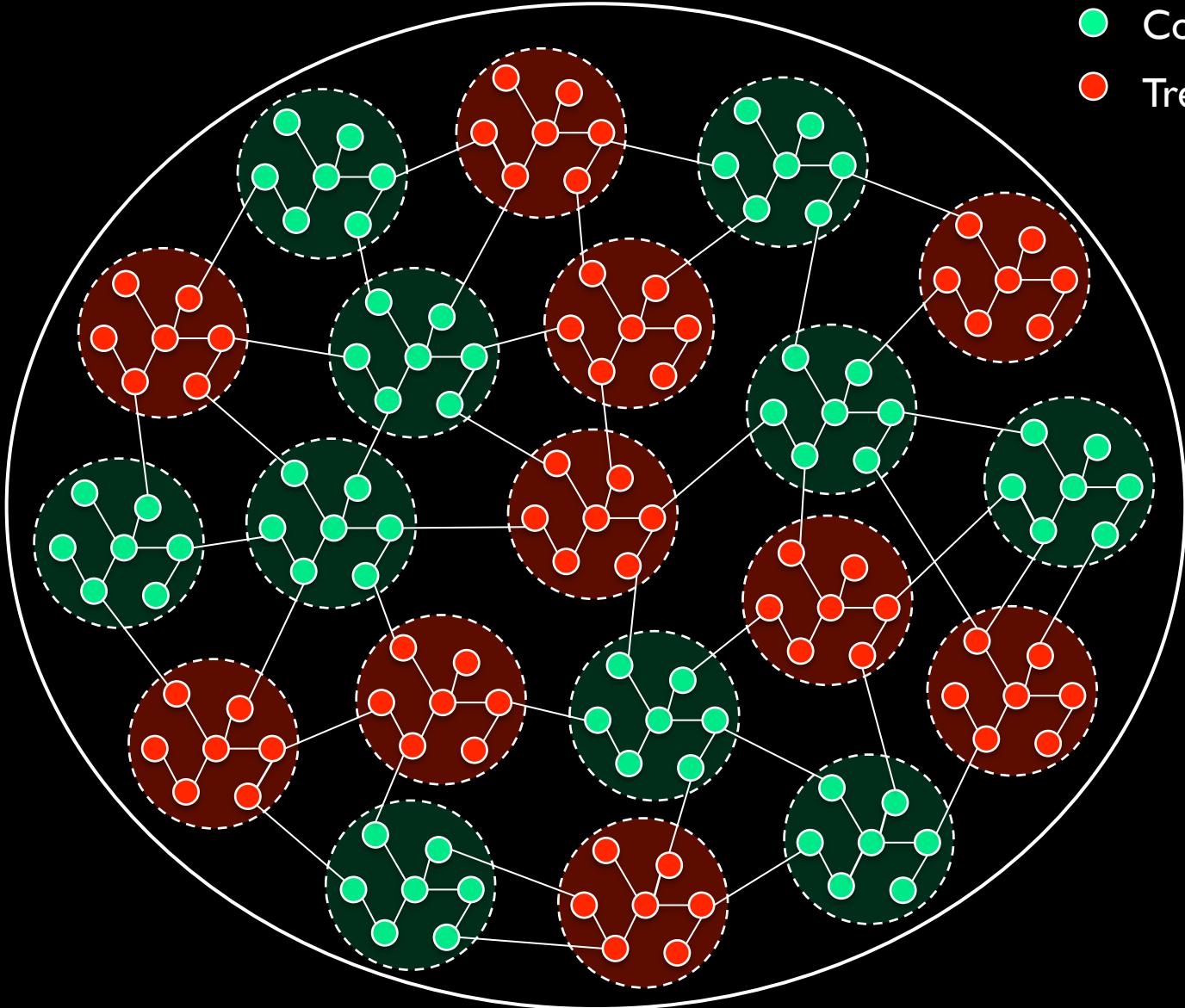


Cluster-based Randomized Experiment



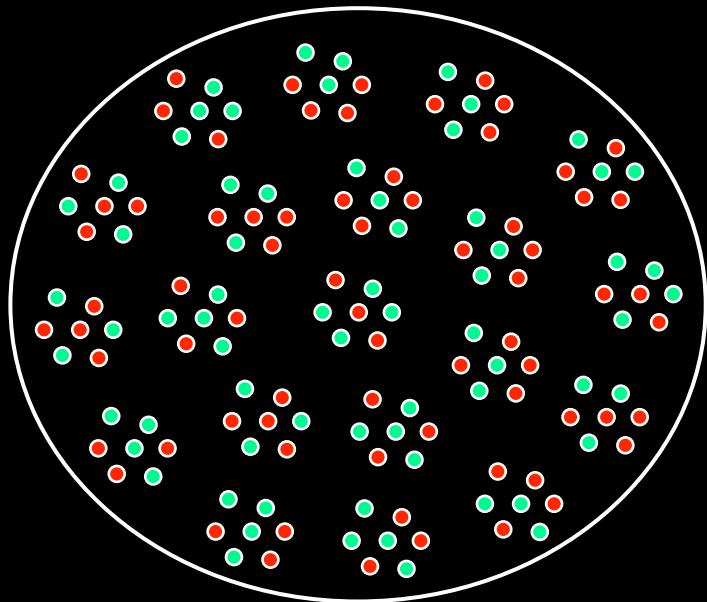
Cluster-based Randomized Experiment

- Control (A)
- Treatment (B)

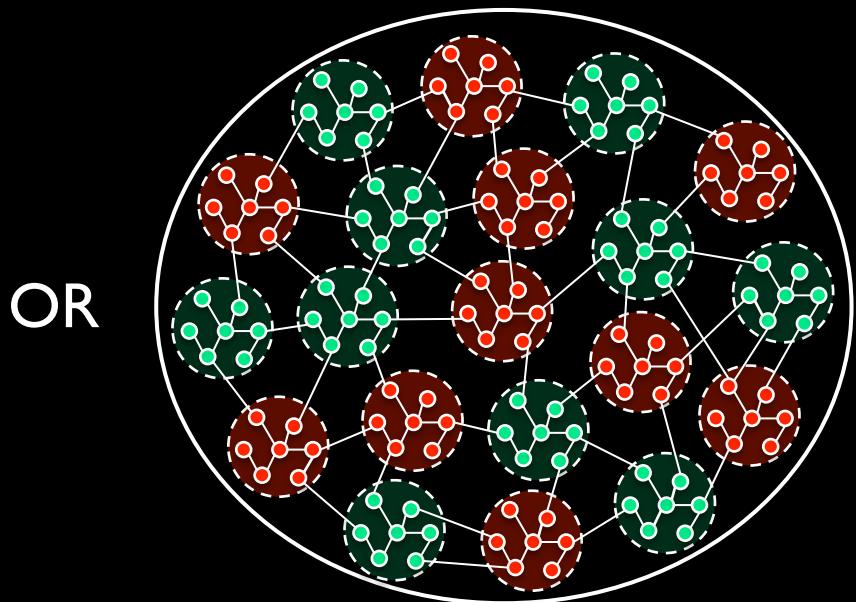


Cluster-based Randomized Experiment

Completely-randomized Experiment

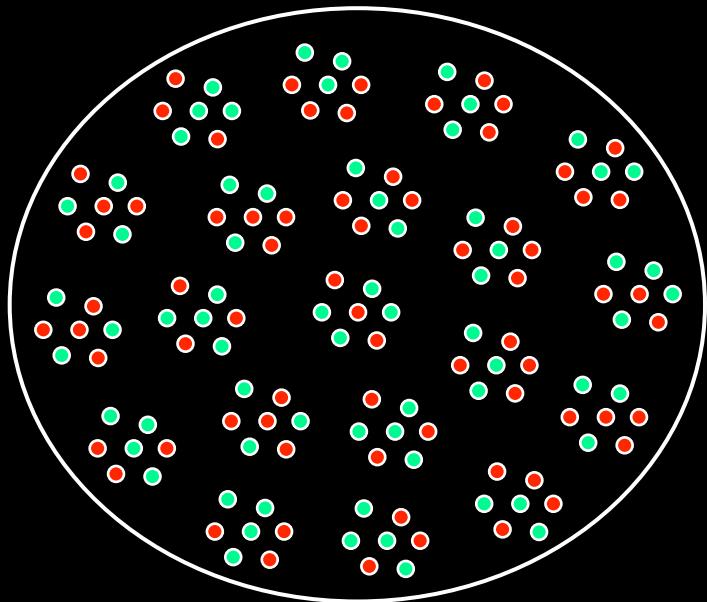


Cluster-based Randomized Experiment

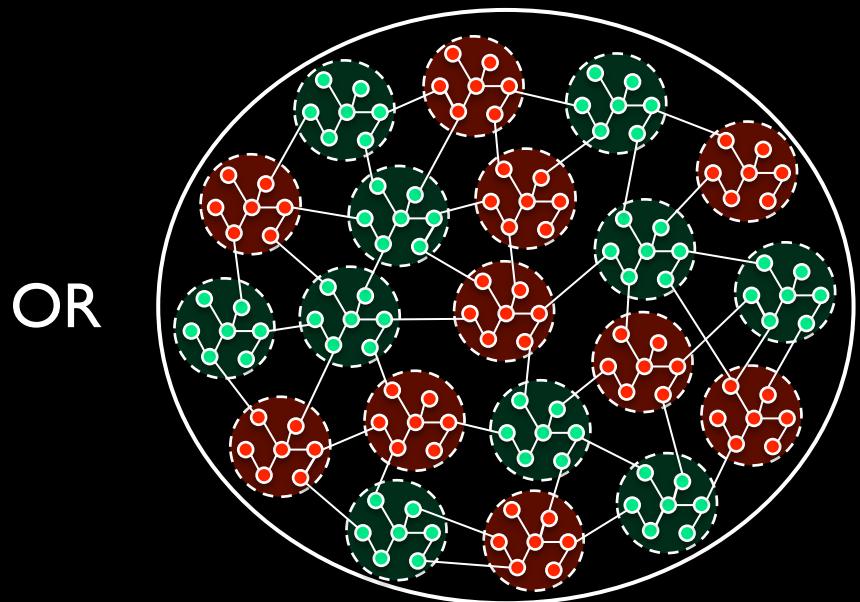


OR

Completely-randomized Experiment



Cluster-based Randomized Experiment



OR

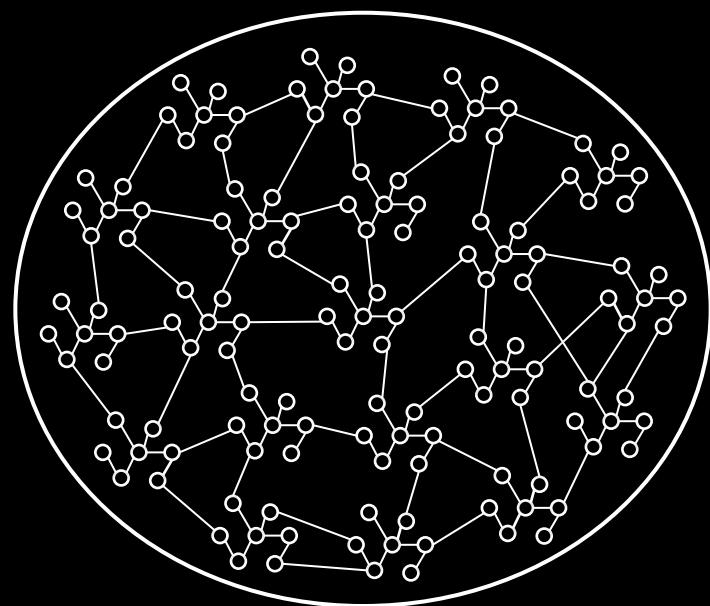
More Spillovers

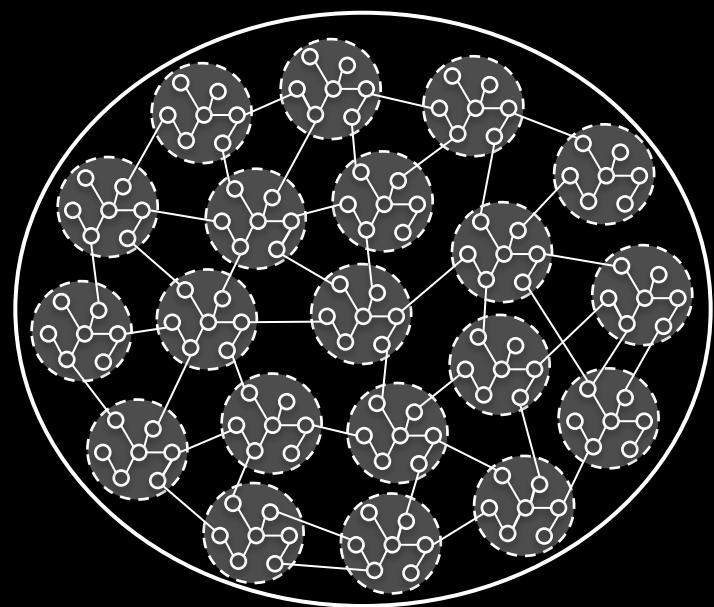
Lower Variance

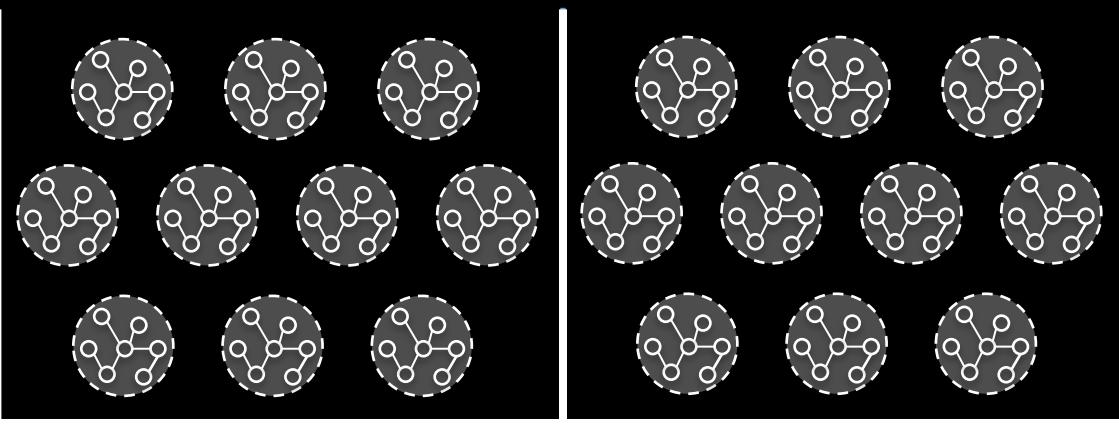
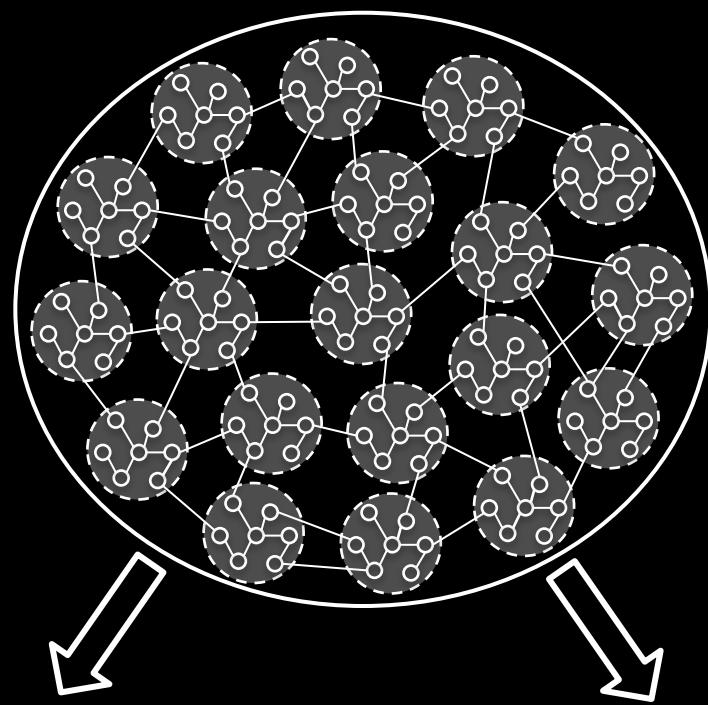
Less Spillovers

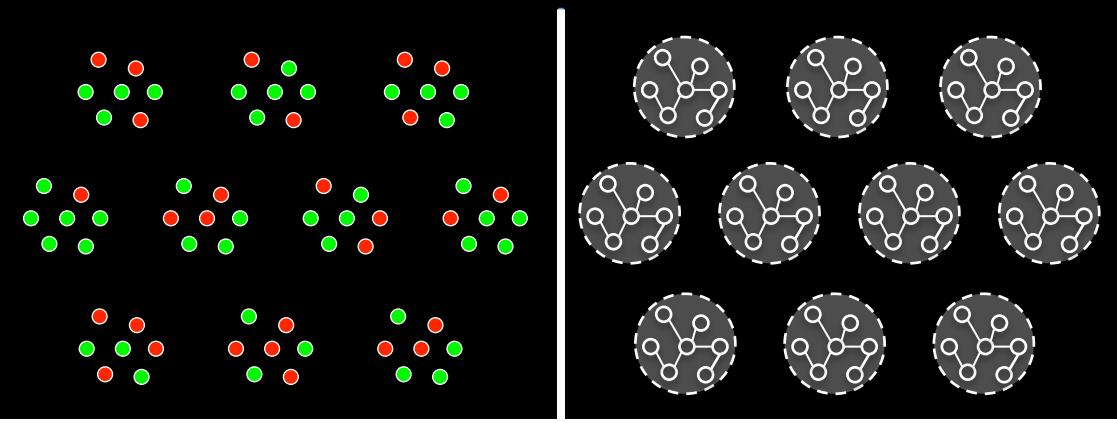
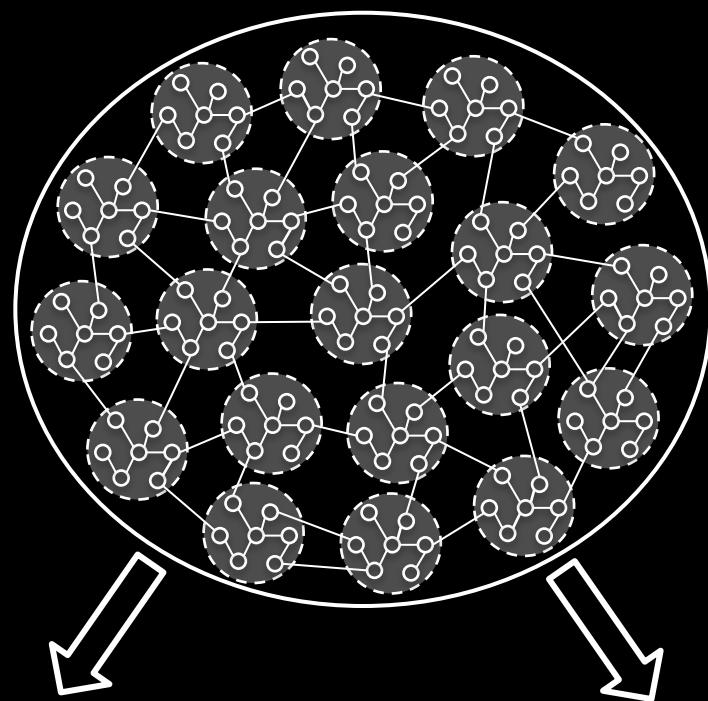
Higher Variance

Design for Detecting Network Effects

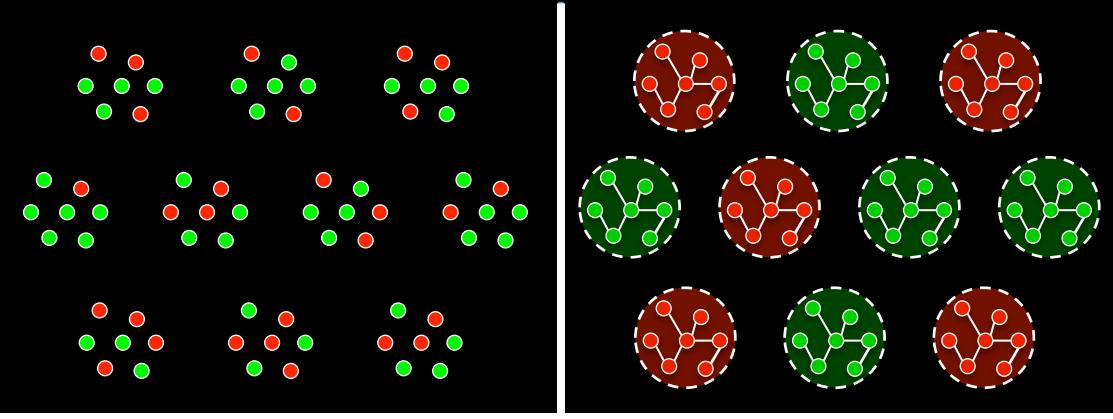
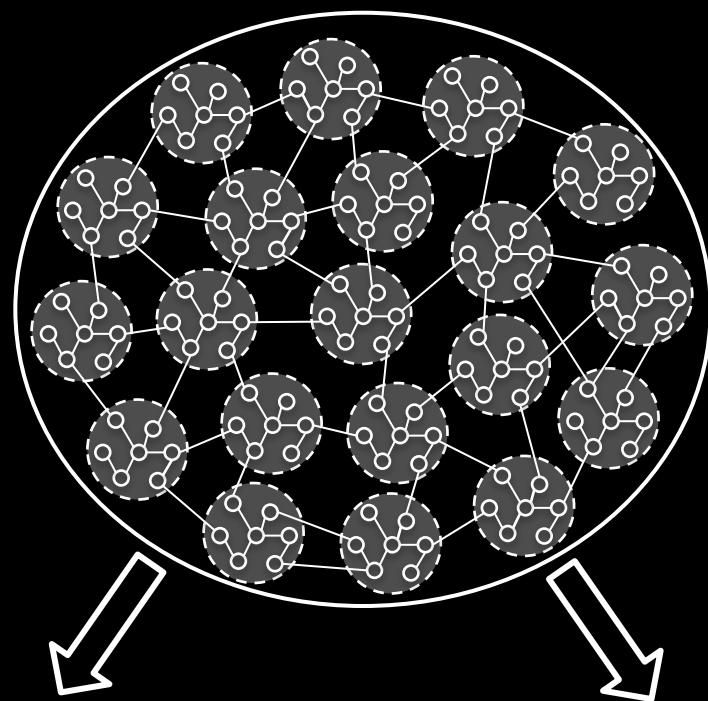






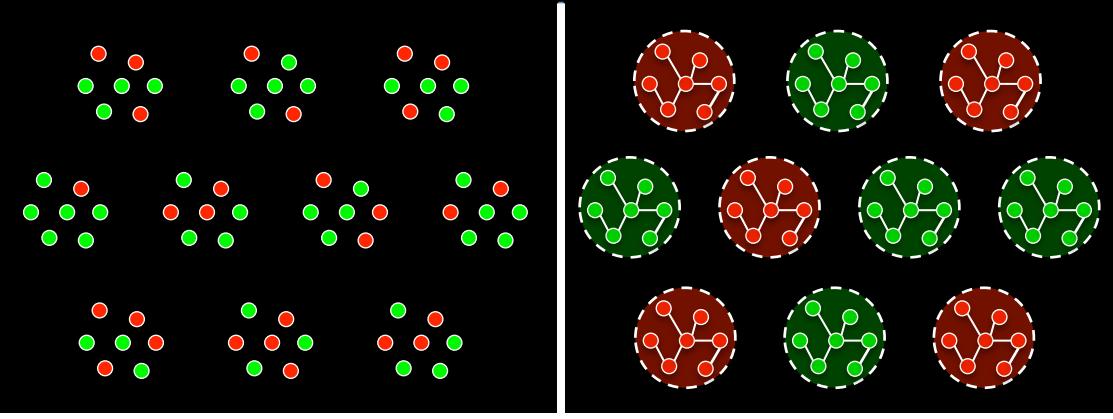
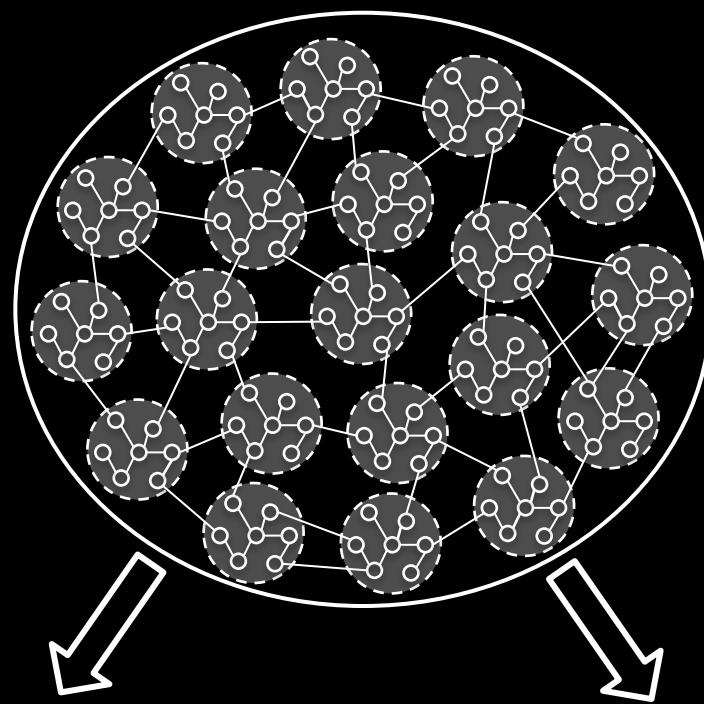


Completely Randomized
Experiment



Completely Randomized
Experiment

Cluster-based Randomized
Experiment



Completely Randomized
Experiment

Cluster-based Randomized
Experiment

$$\mu_{\text{completely-randomized}} \stackrel{?}{=} \mu_{\text{cluster-based}}$$

Hypothesis Test

Hypothesis Test

H_0 : SUTVA Holds

Hypothesis Test

H_0 : SUTVA Holds

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

Hypothesis Test

H_0 : SUTVA Holds

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

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

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:

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:

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

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:

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

Type I error is no greater than α

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

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)

=> 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)

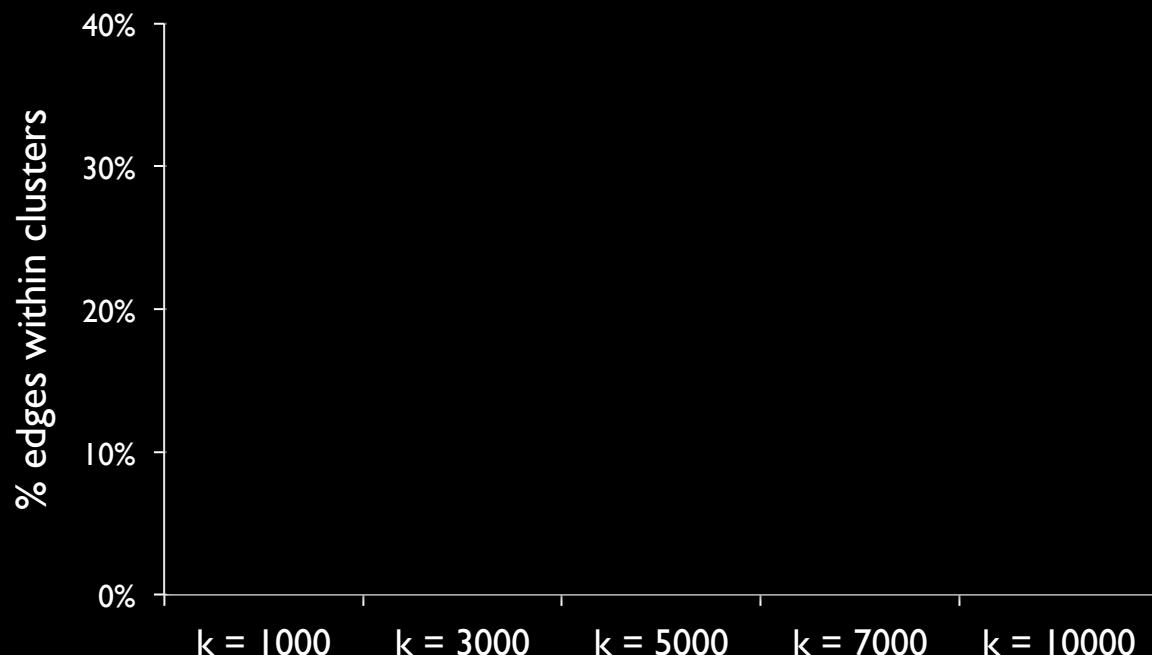
- Streaming
- Parallelizable
- Stable

Clustering the LinkedIn Graph

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

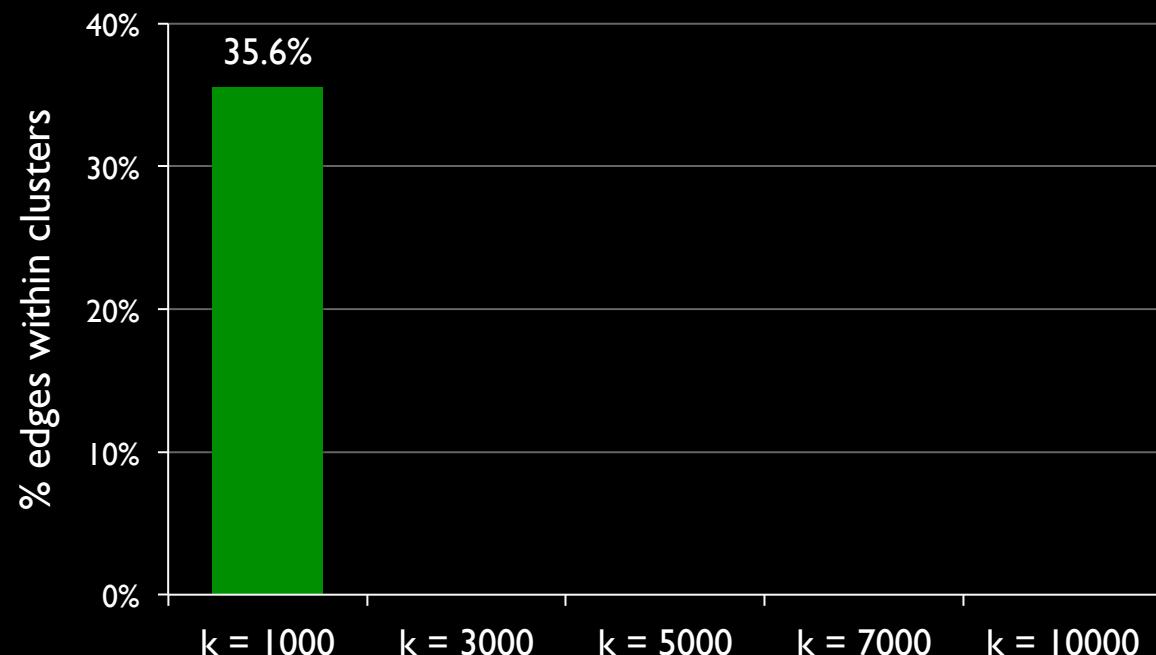
Clustering the LinkedIn Graph

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



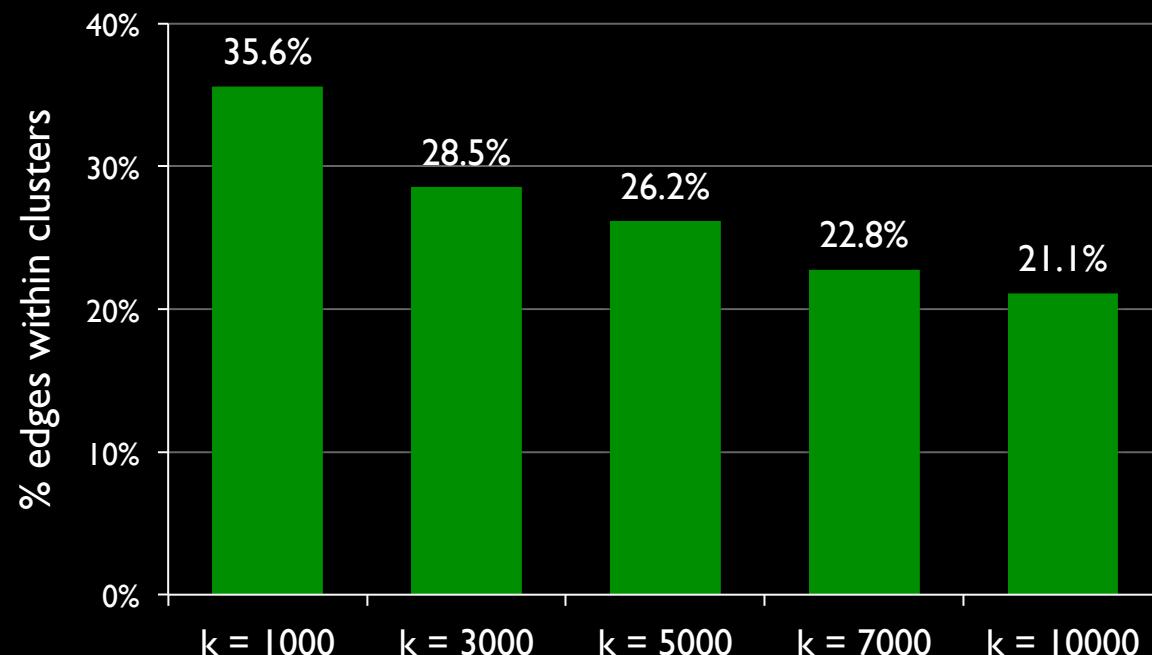
Clustering the LinkedIn Graph

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



Clustering the LinkedIn Graph

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



Choosing the Number of Clusters

Choosing the Number of Clusters

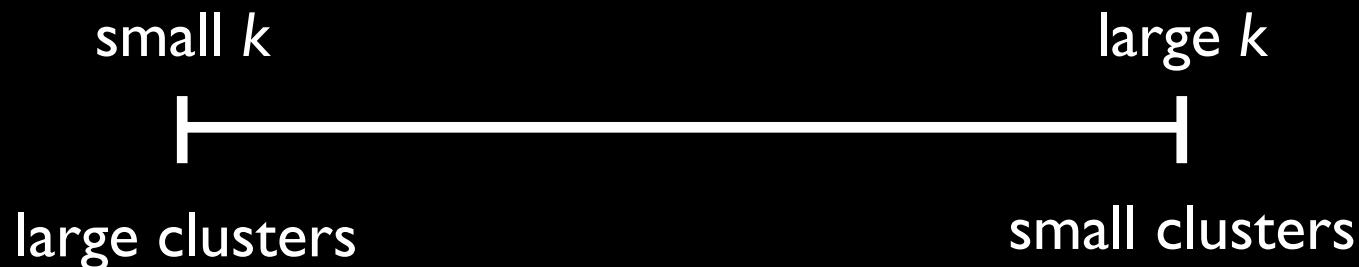
small k



large k



Choosing the Number of Clusters



Choosing the Number of Clusters



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

$$Y_i = \beta_0 + \beta_1 Z_i + \beta_2 \rho_i + \epsilon_i$$

ρ_i : fraction of treated friends

Choosing the Number of Clusters

Understanding the Type II error

Assuming an interference model

$$Y_i = \beta_0 + \beta_1 Z_i + \beta_2 \rho_i + \epsilon_i$$

ρ_i : fraction of treated friends

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

ρ : 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

$$Y_i = \beta_0 + \beta_1 Z_i + \beta_2 \rho_i + \epsilon_i$$

ρ_i : fraction of treated friends

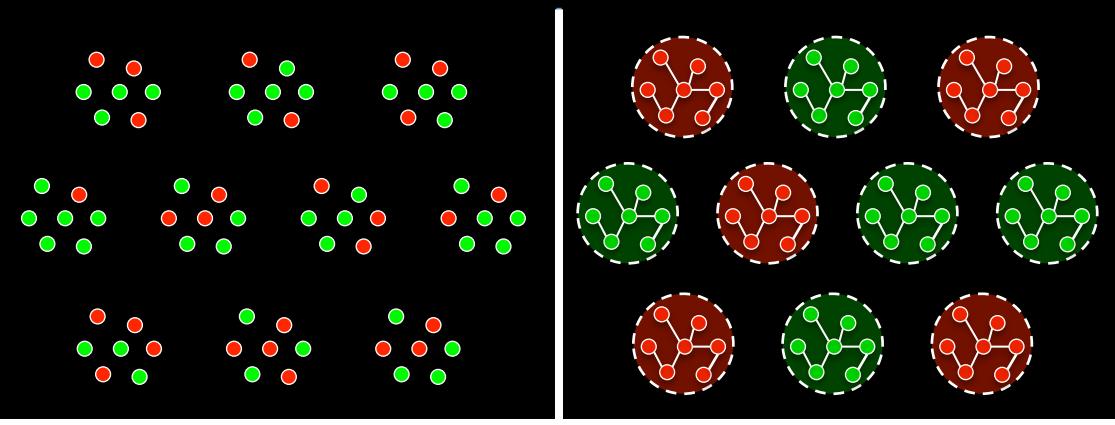
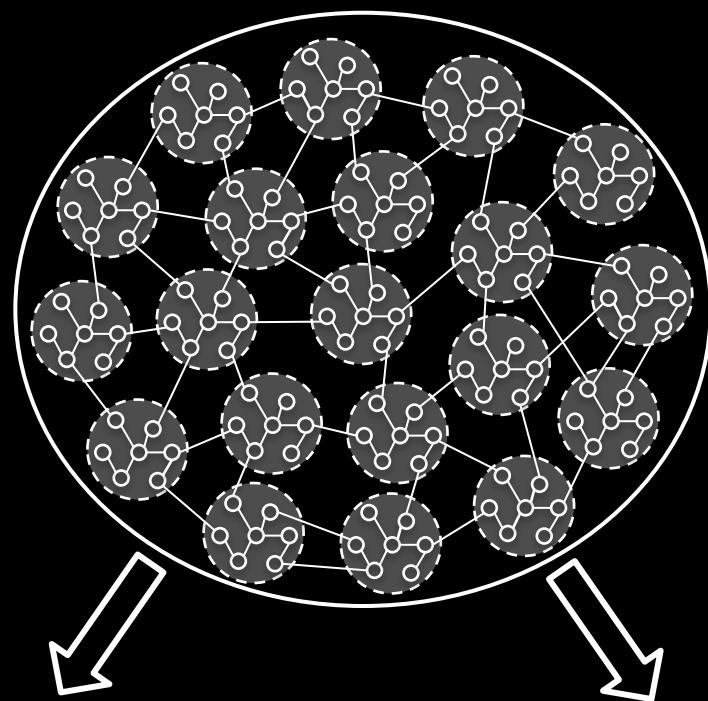
$$E [\hat{\mu}_{cbr} - \hat{\mu}_{cr}] \approx \rho \cdot \beta_2$$

ρ : average fraction of a unit's neighbors contained in the cluster

Choose number of clusters M and clustering C such that

$$\max_{M,C} \frac{\rho}{\sqrt{\hat{\sigma}_C^2}}$$

Experiments on LinkedIn



Bernoulli
Randomized
Experiment

$(\mu_{\text{bernoulli}})$

Completely Randomized
Experiment

Cluster-based Randomized
Experiment

$$\mu_{\text{completely-randomized}} \stackrel{?}{=} \mu_{\text{cluster-based}}$$

Experiment I

Experiment I

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

Experiment I

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

Experiment I

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

Experiment I

- 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 I

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

	Treatment effect	Standard Deviation
Bernoulli Randomization (BR)		
Cluster-based Randomization (CBR)		
Delta (CBR – BR)		

Experiment I

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

	Treatment effect	Standard Deviation
Bernoulli Randomization (BR)	0.0559	0.0050
Cluster-based Randomization (CBR)		
Delta (CBR – BR)		

Experiment I

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

	Treatment effect	Standard Deviation
Bernoulli Randomization (BR)	0.0559	0.0050
Cluster-based Randomization (CBR)	0.0771	0.0260
Delta (CBR – BR)		

Experiment I

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

	Treatment effect	Standard Deviation
Bernoulli Randomization (BR)	0.0559	0.0050
Cluster-based Randomization (CBR)	0.0771	0.0260
Delta (CBR – BR)	-0.0211	0.0265

Experiment I

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

	Treatment effect	Standard Deviation
Bernoulli Randomization (BR)	0.0559	0.0050
Cluster-based Randomization (CBR)	0.0771	0.0260
Delta (CBR – BR)	-0.0211	0.0265

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

	Treatment effect	Standard Deviation
Bernoulli Randomization (BR)		
Cluster-based Randomization (CBR)		
Delta (CBR – BR)		

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

	Treatment effect	Standard Deviation
Bernoulli Randomization (BR)	0.2108	0.2911
Cluster-based Randomization (CBR)		
Delta (CBR – BR)		

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

	Treatment effect	Standard Deviation
Bernoulli Randomization (BR)	0.2108	0.2911
Cluster-based Randomization (CBR)	0.5390	0.5613
Delta (CBR – BR)		

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

	Treatment effect	Standard Deviation
Bernoulli Randomization (BR)	0.2108	0.2911
Cluster-based Randomization (CBR)	0.5390	0.5613
Delta (CBR – BR)	-0.3281	0.5712

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

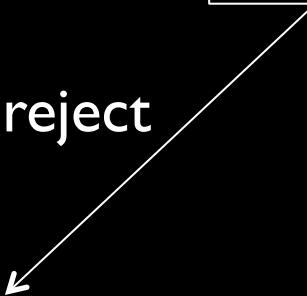
	Treatment effect	Standard Deviation
Bernoulli Randomization (BR)	0.2108	0.2911
Cluster-based Randomization (CBR)	0.5390	0.5613
Delta (CBR – BR)	-0.3281	0.5712

p-value: 0.0483

Test SUTVA null

Test SUTVA null

reject



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

fail to reject

Use cluster-based
experiment to estimate
treatment effects

(higher variance)

Use Bernoulli
experiment to estimate
treatment effects

Test SUTVA null

reject

fail to reject

Use cluster-based
experiment to estimate
treatment effects

(higher variance)

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