Network Synthetic Interventions

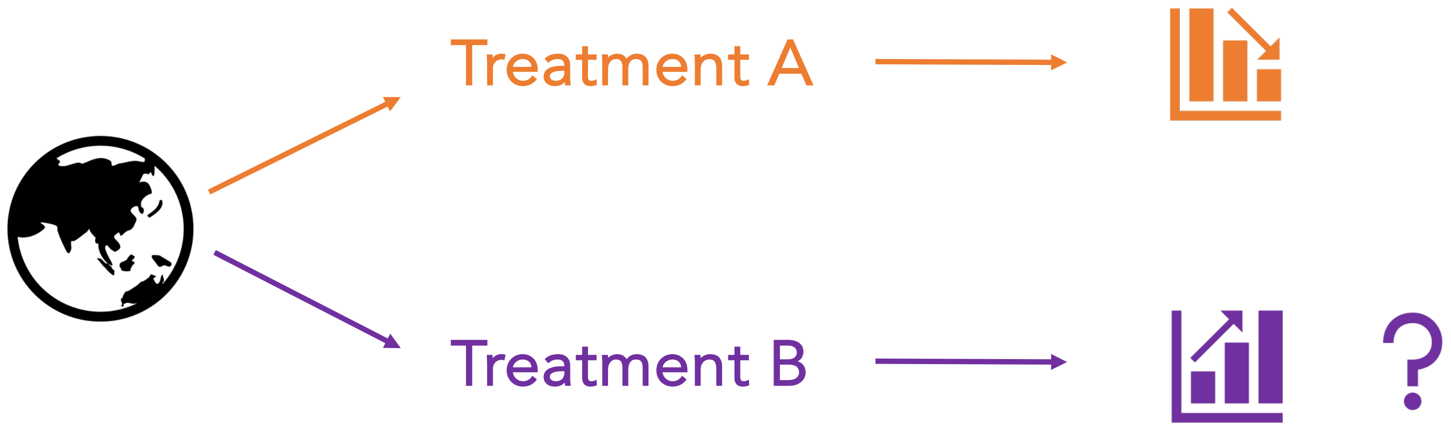
A Causal Framework for Panel Data Under Network Interference

Anish Agarwal, **Sarah Cen**, Devavrat Shah, and Christina Lee Yu

INFORMS

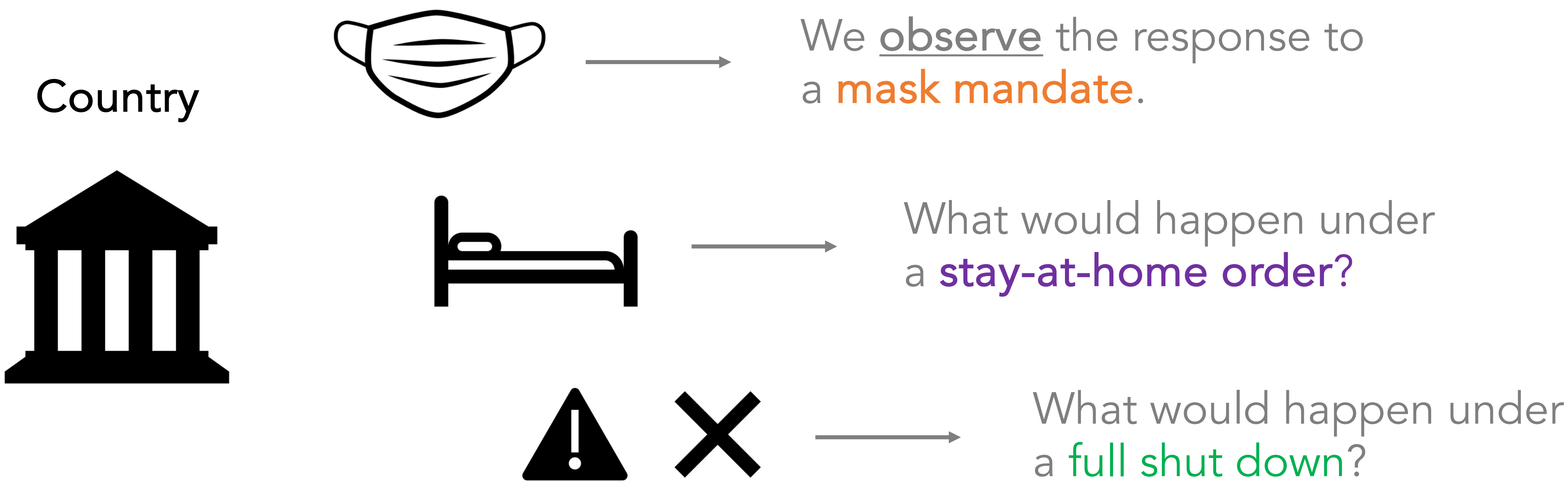
October 17, 2023

What would have happened if we had done A instead of B?



Example: Public Policy

What types of restrictions should a country adopt for COVID-19?



Potential Outcomes in Panel Data

How would unit n respond at time t to treatment a?

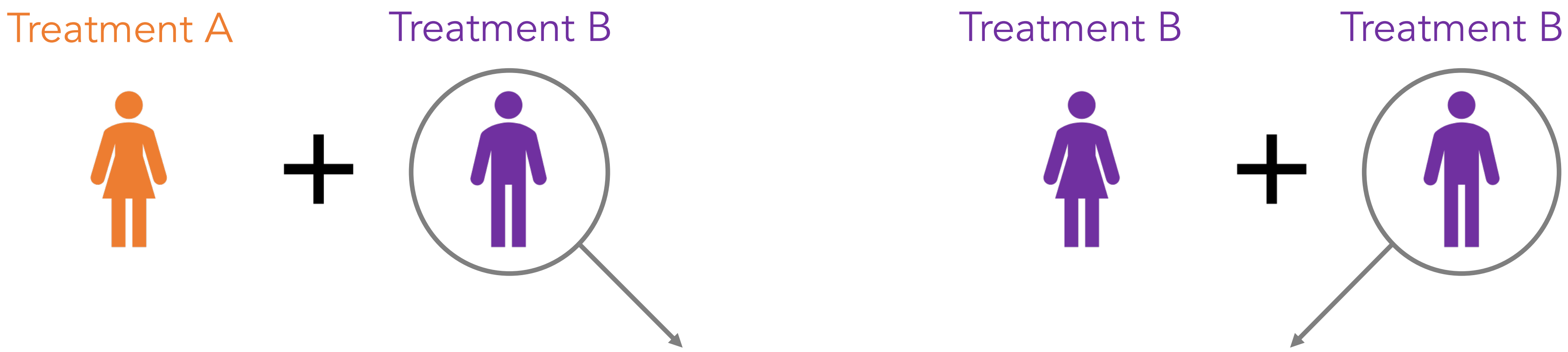
$$Y_{t,n}^{(a)}$$

	$t=1$	$t=2$	$t=3$	…	$t=T-2$	$t=T-1$	$t=T$
Unit 1	nothing	stay-at-home	stay-at-home	…	mask mandate	mask mandate	nothing
Unit 2							

Common Assumption

Stable Unit Treatment Value Assumption

SUTVA: Treatment of one unit does not affect outcome of another → **no spillover**.

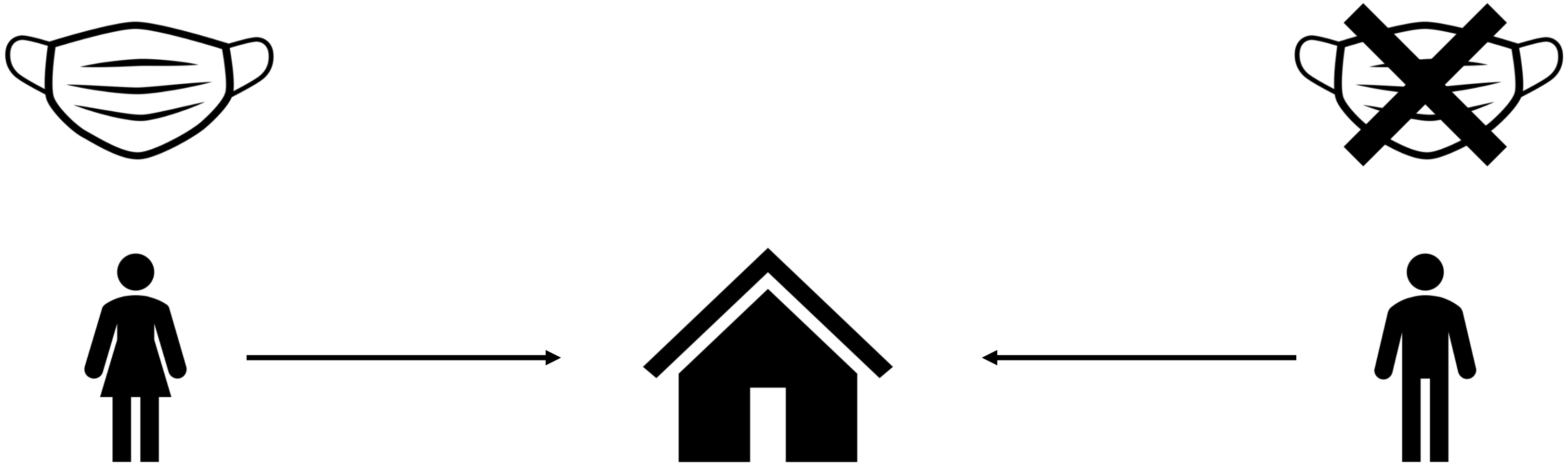


Are the outcomes for Bob the same in both cases?

Not if there is spillover!

Example: COVID-19 Policies

Whether a person is infected depends on who they live with.



Example: Amazon Product Ranking

Whether an item is purchased depends on what else is shown.

Search result #1 vs #2:

How do other products' discounts affect sale of ?

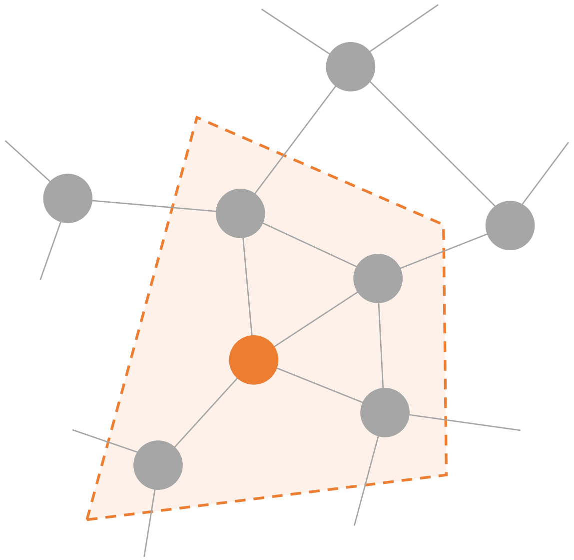
Potential Outcomes Under Spillover

How would unit n in graph G respond at time t to treatments $\boldsymbol{a}$?

$$Y_{t,n}^{(\boldsymbol{a})}$$

Model spillover as **network interference**:

$(j, i) \in \mathcal{E} \implies$ unit j's treatment affects unit i's outcome



Related Work

Average treatment effect estimation [Ugander et al. ‘13, Eckles et al. ‘17, Sussman & Airoldi ‘17, Basse & Airoldi ’18, Jagadeesan et al. ’20, Leung ’19, Chin ’19, Ma & Tresp ’21, Yu et al. ‘22, Cortez et al. ‘22]

Panel data [Bertrand et al. ‘04, Abadie ‘21, Arkhangelsky et al. ’19, Agarwal et al. ‘20]

Spillover effects or interference [Manski ’13, Aronow et al. ‘17, Sussman & Airoldi ‘17, Eckles et al. ‘17, Basse & Airoldi ‘18, Karwa and Airoldi ’18, Bhattacharya et al. ‘20, Li et al. ‘21, Bajari et al. ‘21, Yu et al. ‘22, Cortez et al. ‘22]

Network Synthetic Interventions (NSI)

How would unit n in graph G respond at time t to treatments $\boldsymbol{a}$?

$$Y_{t,n}^{(\boldsymbol{a})}$$

- Simple procedure that provides **point estimate and confidence interval**
- Formal results: guarantee **consistency & asymptotic normality** under LF model
- **Two validity tests**: check important assumption
- **Experiment design** guarantees $O(\mathrm{poly(d)}/\varepsilon^4)$ samples

Estimator: NSI

Network Synthetic Interventions

Setting: Panel Data

$$Y_{t,n}^{(\boldsymbol{a})}$$

	training period $\mathcal{T}_{tr}$					prediction period $\mathcal{T}_{pr}$				
Unit 1	1	1	…	…	…	1	1	…	…	…
Unit 2	0	1	…	…	…	0	0	…	…	…
Unit 3	0	0	…	…	…	0	0	…	…	…
Unit 4	1	1	…	…	…	0	0	…	…	…
Unit 5	1	1	…	…	…	0	0	…	…	…

Treatment matrix A

Setting: Panel Data

$$Y_{t,n}^{(\boldsymbol{a})}$$

	training period $\mathcal{T}_{tr}$					prediction period $\mathcal{T}_{pr}$				
Unit 1	1	1	…	…	…	1	1	…	…	…
Unit 2	0	1	…	…	…	0	0	…	…	…
Unit 3	0	0	…	…	…	0	0	…	…	…
Unit 4	1	1	…	…	…	0	0	…	…	…
Unit 5	1	1	…	…	…	0	0	…	…	…

Treatment matrix A

Setting: Panel Data

$$Y_{t,n}^{(\boldsymbol{a})}$$

	training period $\mathcal{T}_{tr}$					prediction period $\mathcal{T}_{pr}$				
Unit 1	1	1	…	…	…	1	1	…	…	…
Unit 2	0	1	…	…	…	0	0	…	…	…
Unit 3	0	0	…	…	…	0	0	…	…	…
Unit 4	1	1	…	…	…	0	0	…	…	…
Unit 5	1	1	…	…	…	0	0	…	…	…

Treatment matrix A

Setting: Panel Data

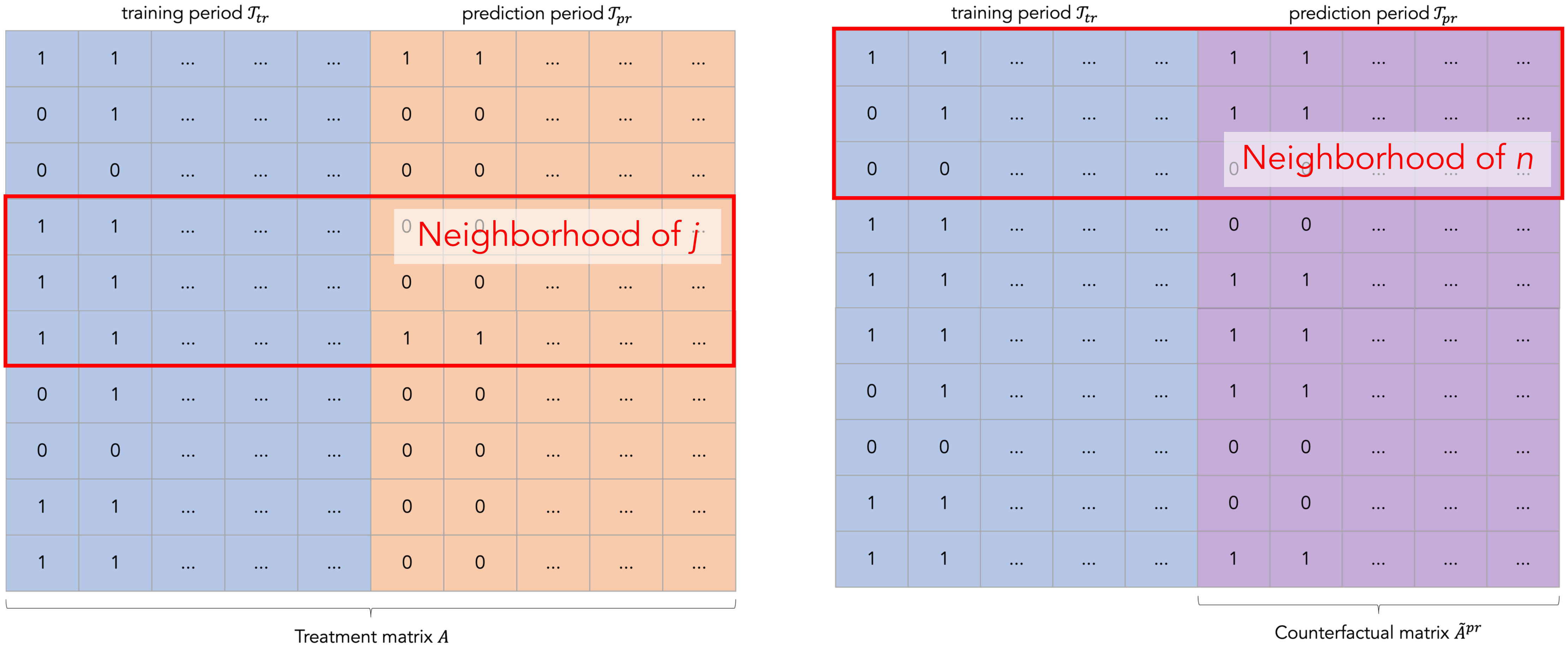
$Y_{t,n}^{(\boldsymbol{a})}$

training period $\mathcal{T}_{tr}$ | prediction period $\mathcal{T}_{pr}$

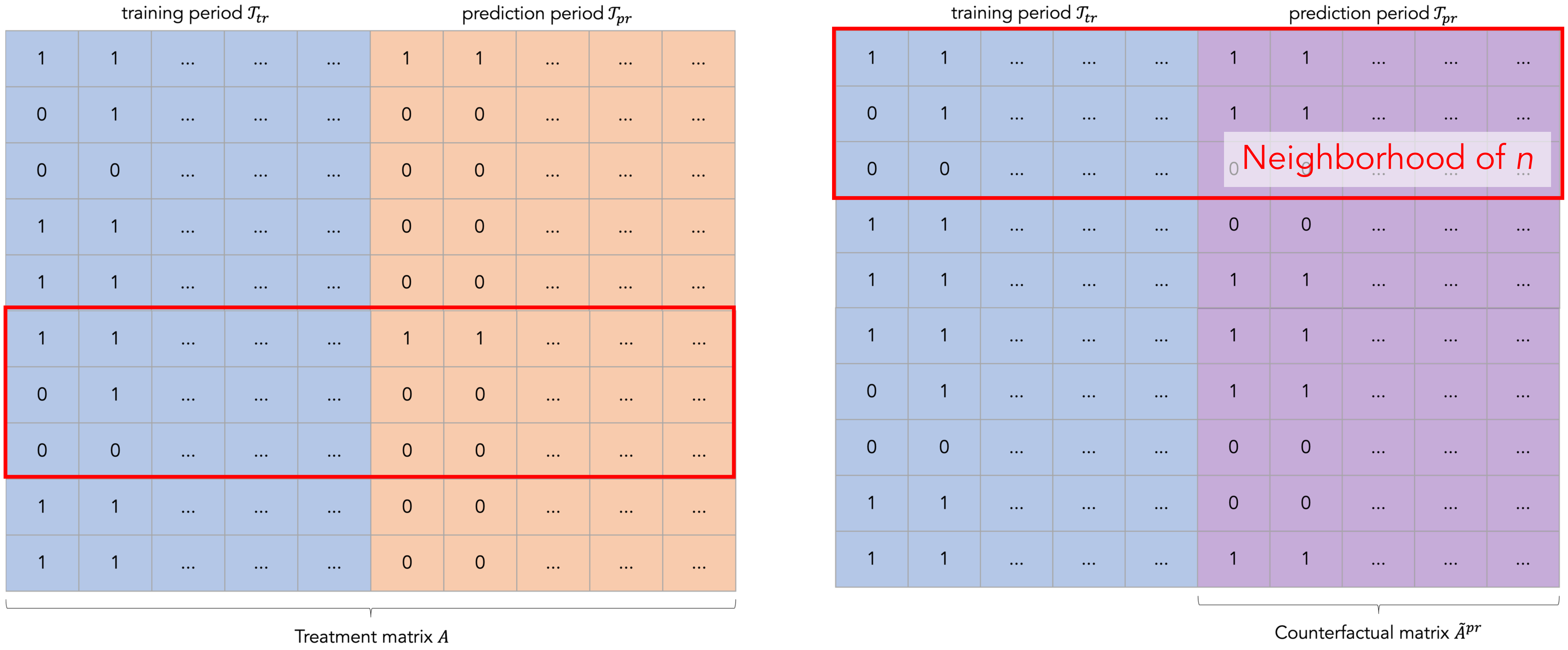
Unit 1	1	1	…	…	…	1	1	…	…	…
Unit 2	0	1	…	…	…	1	1	…	…	…
Unit 3	0	0	…	…	…	0	0	…	…	…
Unit 4	1	1	…	…	…	0	0	…	…	…
Unit 5	1	1	…	…	…	1	1	…	…	…

Counterfactual matrix $\tilde{A}^{pr}$

Step 1: Identify Donors for Unit n



Step 1: Identify Donors for Unit n



Step 1: Identify Donors for Unit n

training period $\mathcal{T}_{tr}$					prediction period $\mathcal{T}_{pr}$				
1	1	...	...	...	1	1	...	...	...
0	1	...	...	...	0	0	...	...	...
0	0	...	...	...	0	0	...	...	...
1	1	...	...	...	0	0	...	...	...
1	1	...	...	...	0	0	...	...	...
1	1	...	...	...	1	1	...	...	...
0	1	...	...	...	0	0	...	...	...
0	0	...	...	...	0	0	...	...	...
1	1	...	...	...	0	0	...	...	...
1	1	...	...	...	0	0	...	...	...

Treatment matrix A

training period $\mathcal{T}_{tr}$					prediction period $\mathcal{T}_{pr}$				
1	1	...	...	...	1	1	...	...	...
0	1	...	...	...	1	1	...	...	...
0	0	...	...	...	0	Neighborhood of n			
1	1	...	...	...	0	0	...	...	...
1	1	...	...	...	1	1	...	...	...
1	1	...	...	...	1	1	...	...	...
0	1	...	...	...	1	1	...	...	...
0	0	...	...	...	0	0	...	...	...
1	1	...	...	...	0	0	...	...	...
1	1	...	...	...	1	1	...	...	...

Counterfactual matrix $\tilde{A}^{pr}$

Step 1: Identify Donors for Unit n

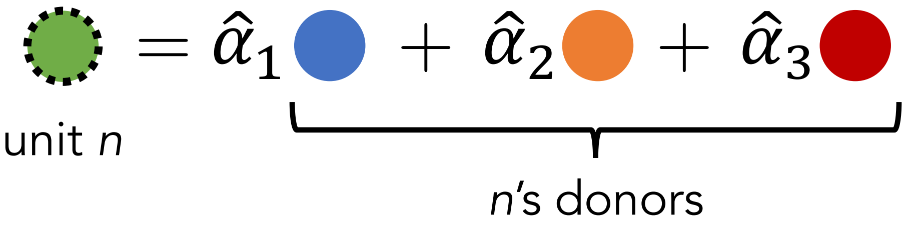
	training treatments					CF treatments				
unit n's neighborhood	1	1	...	...	...	1	...	...	...	...
	0	1	...	...	...	0	...	...	...	...
	0	0	...	...	...	0	...	...	...	...
donor's neighborhood	0	0	...	...	...	0	...	...	...	...
	0	1	...	...	...	0	...	...	...	...
	1	1	...	...	...	1	...	...	...	...

Step 2: Construct Estimate and CIs

Principal Component Regression

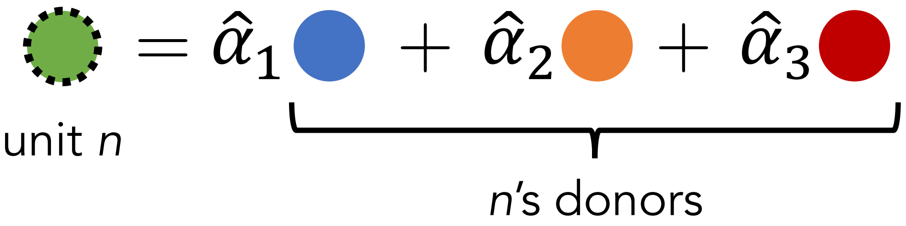
Run PCR on training observations to obtain $\{\hat{\alpha}_j\}_{j \in \text{Donors}}$

$$\widehat{\mathbb{E}}Y_{t,n}^{(a)} = \sum_{j \in \text{Donors}} \hat{\alpha}_j \cdot (\text{observation of unit j at time t})$$



(See paper for confidence interval)

Intuition



Main idea: Potential outcome of *n* can be written as a linear combination of the donors' potential outcomes.

Example:

sale of designer sunglasses = sale of sunscreen + sale of sunhats − sale of summer sports equipment

Formal Results

Model

How would unit n in graph G respond at time t to treatment $\boldsymbol{a}$?

$$Y_{t,n}^{(a)} = \langle u_{nn}, w_{t,a_n} \rangle + \sum_{j \in \mathcal{N}(n)} \langle u_{jn}, w_{t,a_j} \rangle + \text{noise}$$

n's neighbors

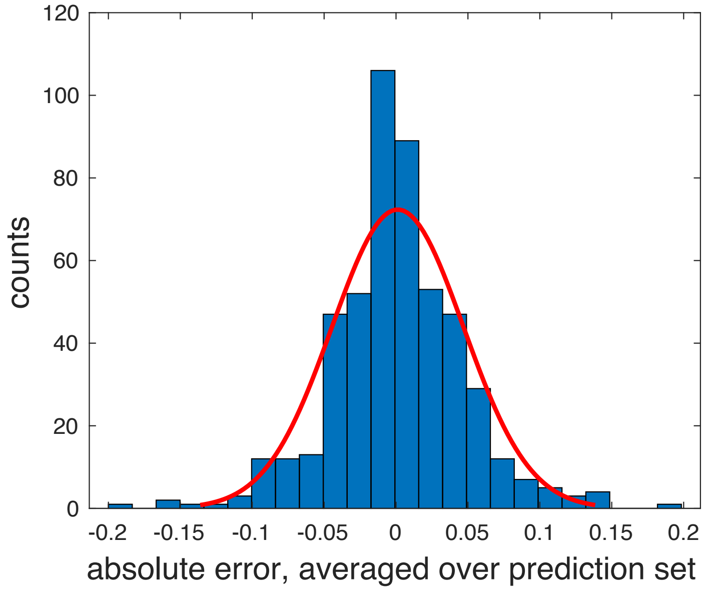
Spillover effect
of each neighbor

Potential outcome	=	Potential outcome if no network effects	+	Network effects

Theoretical Guarantees

Under **latent-factor model**,

1. NSI produces consistent point estimates.
2. NSI estimates are asymptotically normal.



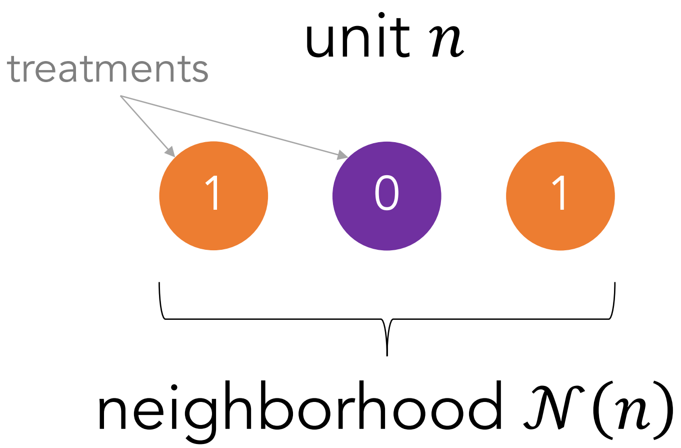
So, why is this problem hard?

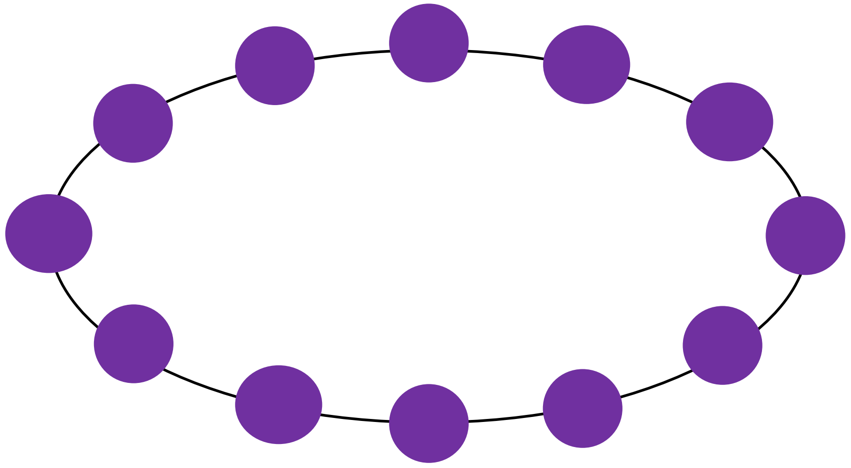
Key enabling assumption: **subspace inclusion**

Do this assumption typically hold?

Not always.

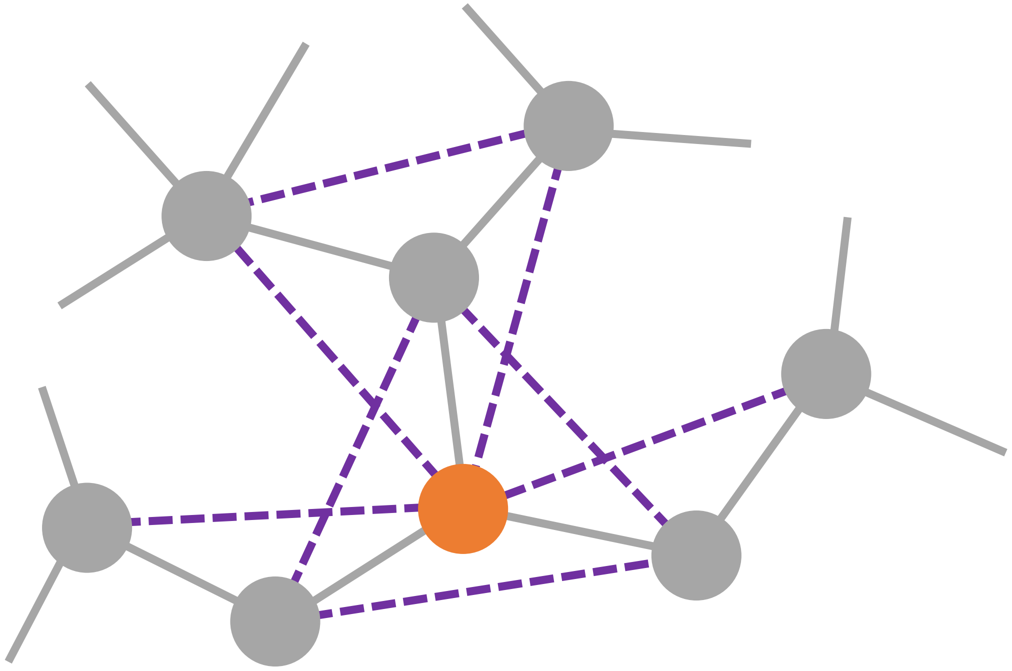
Subspace inclusion assumption: Training treatments are rich enough s.t. we can infer spillover effects from training data.



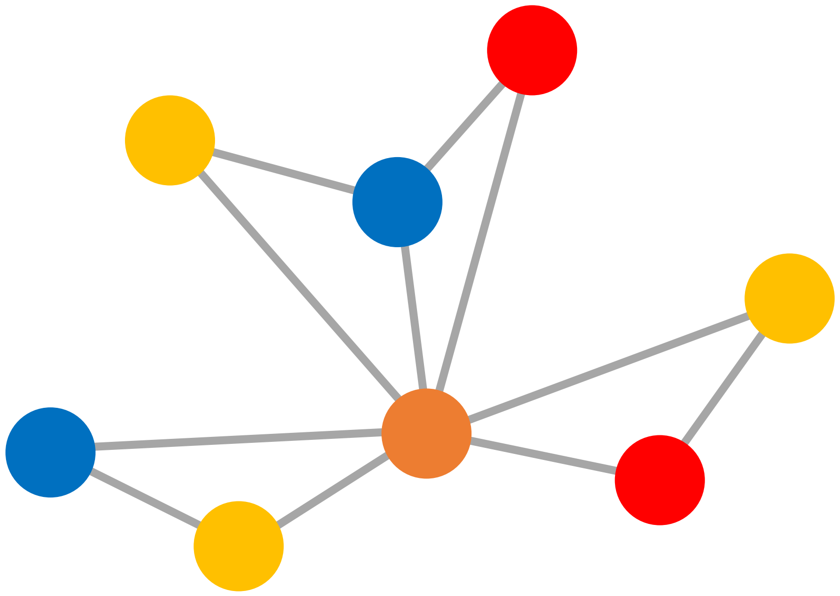


Experiment Design

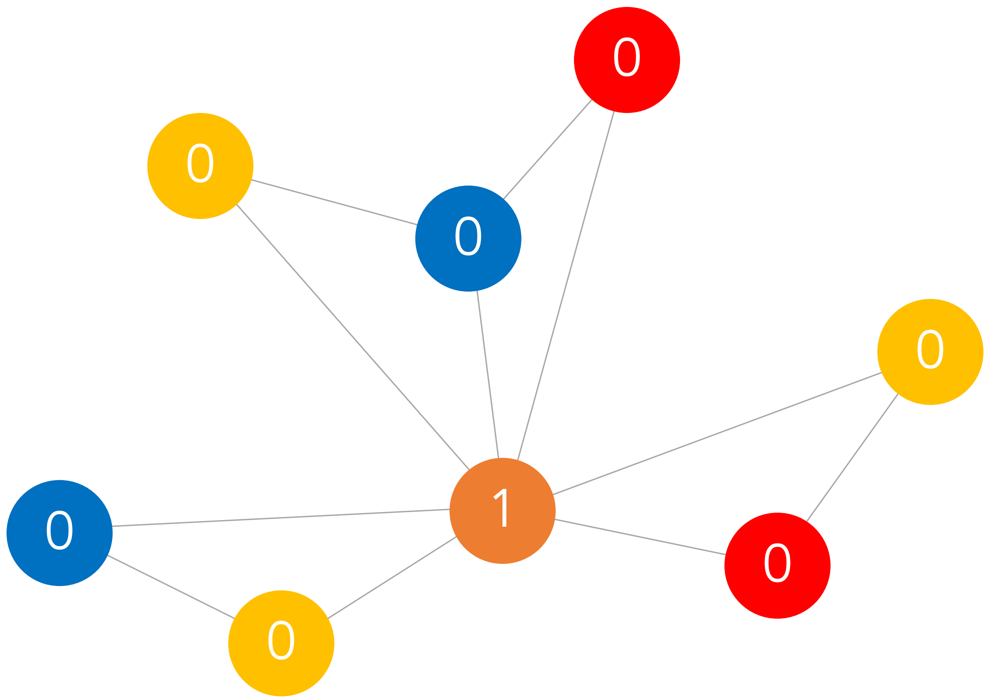
Step 1: Connect each unit to its two-hop neighbors to get $\mathcal{G}'$.



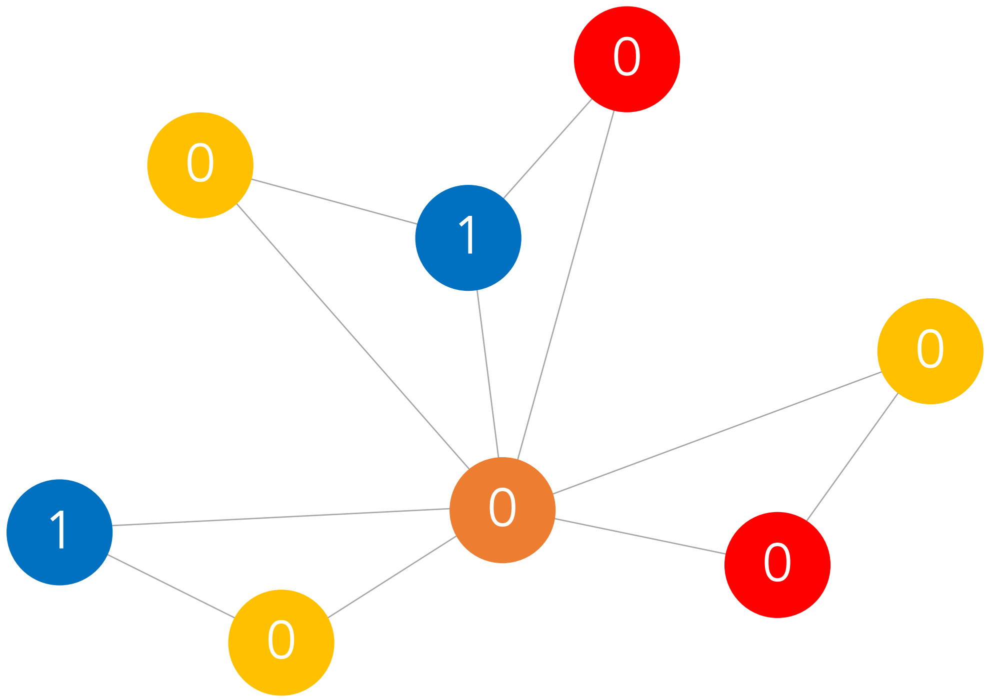
Step 2: Color $\mathcal{G}'$ such that no adjacent units in $\mathcal{G}'$ share a color.



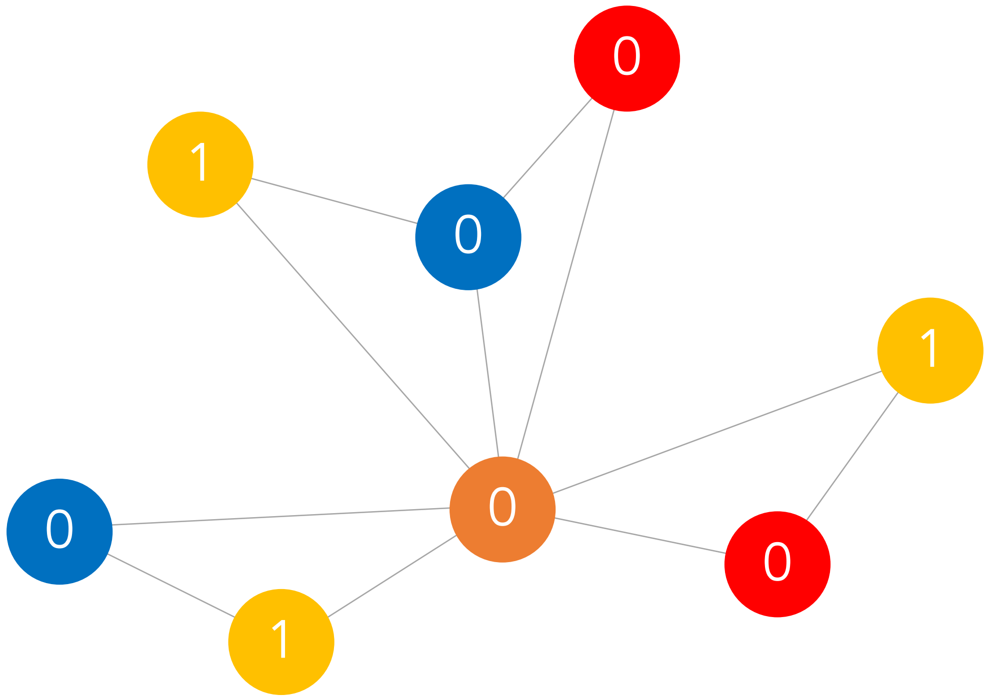
Step 3: For the first $\bar{T}'$ steps, assign 1 to orange units and 0 to all others



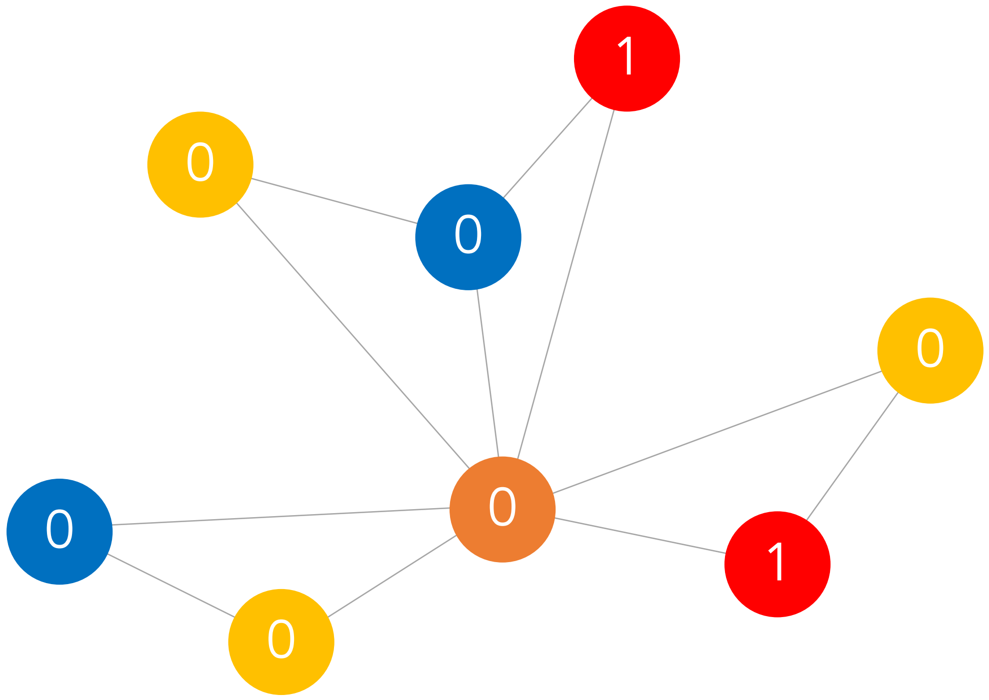
Step 4: For the first $\bar{T}'$ steps, assign 1 to blue units and 0 to all others



Step 5: For the first $\bar{T}'$ steps, assign 1 to yellow units and 0 to all others



Step 6: For the first $\bar{T}'$ steps, assign 1 to red units and 0 to all others



Sample complexity

Lemma. Proposed experiment requires $O(d^2)$ training samples.

degree of graph

Proposition (informal). Under Gaussian latent factors, consistency and asymptotic normality hold with $O(\mathrm{poly(d)}/\varepsilon^4)$ samples.

Significantly better than $O(2^d)$ samples required without latent factor model or experiment design.

Summary

Spillover is common. We model as network interference.

NSI produces point estimates and confidence interval.

Using a latent-factor model & some assumptions, NSI is shown to give **consistent, asymptotically normal estimates**.

We provide two validity tests.

We provide experiment design that has low sample complexity.

Future work: Interventions on graph, test on real data.

Thank you!

shcen@mit.edu