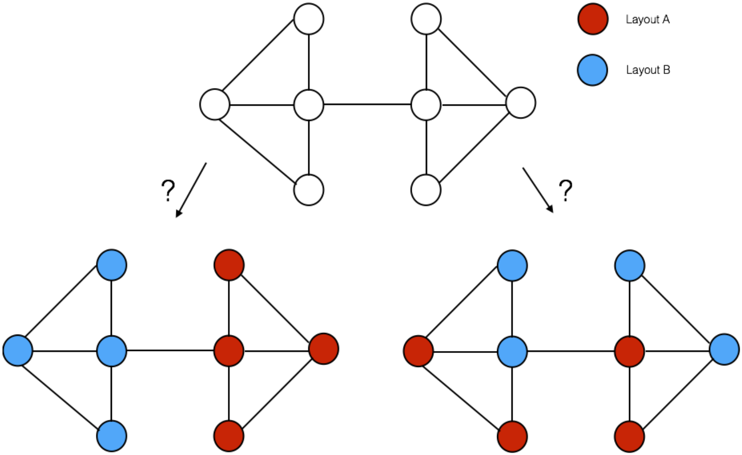
Optimal Design of Experiments with Network-Correlated Outcomes

Guillaume Basse and Edo Airoldi

Harvard University

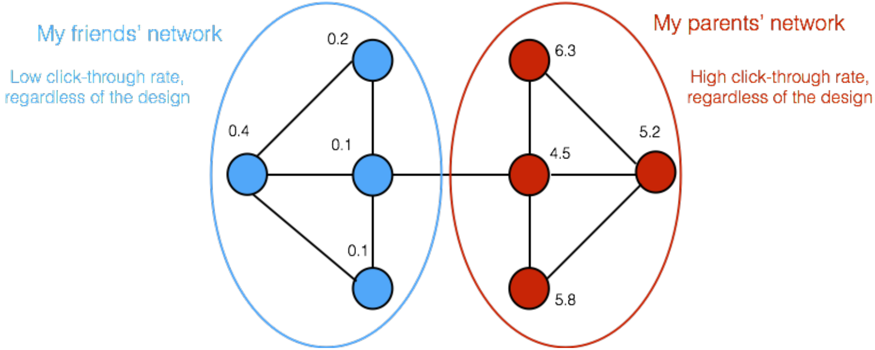
04/25/2013

A Simple Example - bias



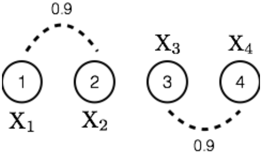
Task: Assign each node to group A or B to estimate relative effect of these designs on clickthrough rate.

A Simple Example - bias



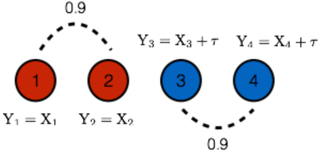
Network Homophily $\Rightarrow$ Inference on average effect is likely to be biased.

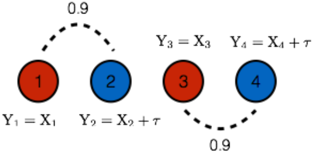
A Simple Example - variance



$$X_i \sim [\mu, \sigma^2]$$
$$Cor(X_1, X2) = Cor(X_3, X_4) = 0.9$$
$$Y_1, Y_2 \perp Y_3, Y_4$$

A Simple Example - variance





$$\hat{\tau}_1 = \frac{1}{2}(Y_1 + Y_2) - \frac{1}{2}(Y_3 + Y_4)$$
$$E[\hat{\tau}_1] = \tau$$
$$Var(\hat{\tau}_1) = \sigma^2 + \frac{3.6\sigma^2}{4} = 1.9\sigma^2$$

$$\hat{\tau}_2 = \frac{1}{2}(Y_1 + Y_3) - \frac{1}{2}(Y_2 + Y_4)$$
$$E[\hat{\tau}_2] = \tau$$
$$Var(\hat{\tau}_2) = \sigma^2 - \frac{3.6\sigma^2}{4} = 0.1\sigma^2$$

Conclusion

$$\frac{Var(\hat{\tau}_1)}{Var(\hat{\tau}_2)} = 19$$

Both estimators are unbiased but we can achieve a large variance reduction.

Informal observation:

Informally, we understand that grouping together units that are tightly connected is a bad idea:

- Bias due to homophily (clusters are confounding variables)
- Large variance has deeper roots (to be explored later)

$\Rightarrow$ We try to formalize these ideas.

Our Goal:

Ideal Goal:

Given a social network and covariates.

We want to test effect of design A and B on click through rate.

How to assign individuals to A and B to maximize precision of estimation of effect?

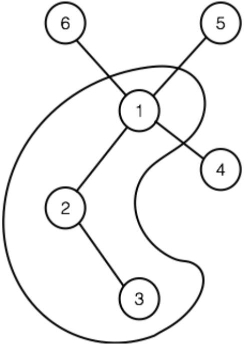
Realistic Goal:

Given a social network **without covariates**.

Assuming additive causal effect.

How to assign individuals to A and B to maximize precision of estimation of effect?

Notation



Augmented Adjacency Matrix:

$$A = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Closed Neighborhood:

$\mathcal{N}_2 = \{1, 2, 3\}$

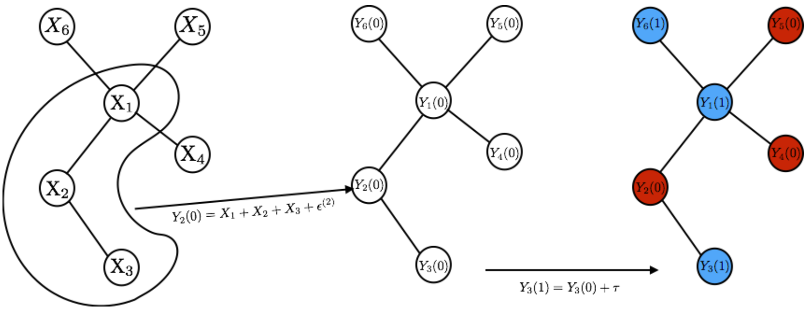
$|\mathcal{N}_2| = 3$

$\mathcal{N}_2 \cap \mathcal{N}_1 = \{1, 2\}$

Assignment Vector: $W = (W_i)_{i=1,..,6}$
where:

$$W_i = \begin{cases} = 1 & \text{if unit i} \in \text{treatment (Layout A)} \\ = 0 & \text{if unit i} \in \text{control (Layout B)} \end{cases}$$

A simple model



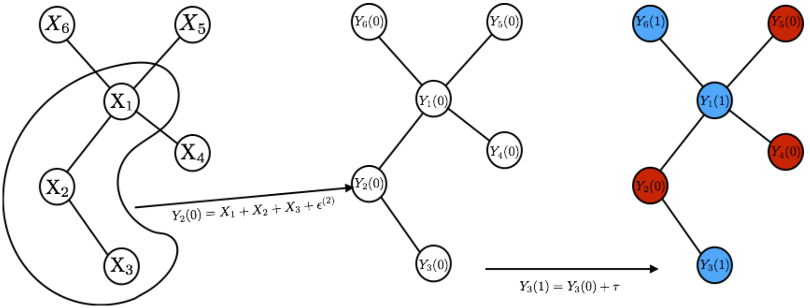
Data Generating Process & Estimator

$$Y_i(1) = Y_i(0) + \tau \qquad \epsilon^{(i)} \overset{ind}{\sim} \mathcal{N}(0, \gamma^2)$$

$$Y_i(0) = \sum_{j \in \mathcal{N}_i} X_j + \epsilon^{(i)} \qquad X_j \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$$

$$\hat{\tau} = \overline{Y(1)}^{obs} - \overline{Y(0)}^{obs}$$

Mean Squared Error



Mean Squared Error

$$\mathrm{mse}_W(\hat{\tau}, \tau) = \mu^2 (Z^*(W))^2 + (Z(W)^t (\gamma^2 I + \sigma^2 A^t A) Z(W))$$

where: $Z^*(W) = \frac{\sum_{W_i=1} \mathcal{N}_i}{N_1} - \frac{\sum_{W_i=0} \mathcal{N}_i}{N_0}$ and $Z(W) = \frac{W}{N_1} - \frac{1-W}{N_0}$

Optimal Design- Integrated MSE

We follow a Bayesian experimental design approach. Starting with:

$$\mathrm{mse}_{W,\theta}(\hat{\tau},\tau)$$

we consider:

$$imse_W(\hat{\tau},\tau) = \int \pi(\theta) mse_{W,\theta}(\hat{\tau},\tau)\, d\theta$$

and in most cases:

$$\widehat{imse}_W(\hat{\tau},\tau) = \sum_{t=1}^{N} \pi(\theta^t) mse_{W,\theta^t}(\hat{\tau},\tau)$$

where $\theta^t \sim \pi(\theta)$.

So we are interested in:

$$\boxed{\tilde{W}^{(opt)} = \operatorname*{argmin}_W \left\{ \widehat{imse}_W(\hat{\tau},\tau) \right\}}$$

Closer look at the MSE of the normal model

$$\text{mse}_W(\hat{\tau}, \tau) = \underbrace{\mu^2 (Z^*(W))^2}_{\text{bias}^2} + \underbrace{\gamma^2 Z(W)^t Z(W) + \sigma^2 Z(W)^t A^t A Z(W)}_{\text{variance}}$$

$\mu^2 (Z^*(W))^2 \Rightarrow$ penalizes imbalance in average size of neighborhoods in each group

$\gamma^2 Z(W)^t Z(W) \Rightarrow$ penalizes imbalance in group sizes

$\sigma^2 Z(W)^t A^t A Z(W) \Rightarrow$ encourages shared neighbors between groups

Bias term

$$\mu^2(Z^*(W))^2 = \mu^2(\overline{|\mathcal{N}(1)|} - \overline{|\mathcal{N}(0)|})^2$$

$\Rightarrow$ Is minimized when nodes on each treatment group have same average number of neighbors

$\Rightarrow$ Bias disappears in k-regular graphs

Variance term - 1

$$\gamma^2 Z(W)^t Z(W) = \gamma^2 (\frac{1}{N_1} + \frac{1}{N_0})$$

⇒ Is minimized when both treatment groups have same size

⇒ Similar to MSE in iid case

Variance term - 2

$$\sigma^2 Z(W)^t A^t A Z(W) = \frac{\sigma^2}{N_1^2} \sum_{i,j/W_i=W_j=1} |\mathcal{N}_i \cap \mathcal{N}_j| \quad (T_1)$$
$$+ \frac{\sigma^2}{N_0^2} \sum_{i,j/W_i=W_j=0} |\mathcal{N}_i \cap \mathcal{N}_j| \quad (T_2)$$
$$- \frac{2\sigma^2}{N_1 N_0} \sum_{i,j/W_i=1\&W_j=0} |\mathcal{N}_i \cap \mathcal{N}_j| \quad (T_3)$$

where the T_1, T_2, T_3 are the average number of shared neighbors:

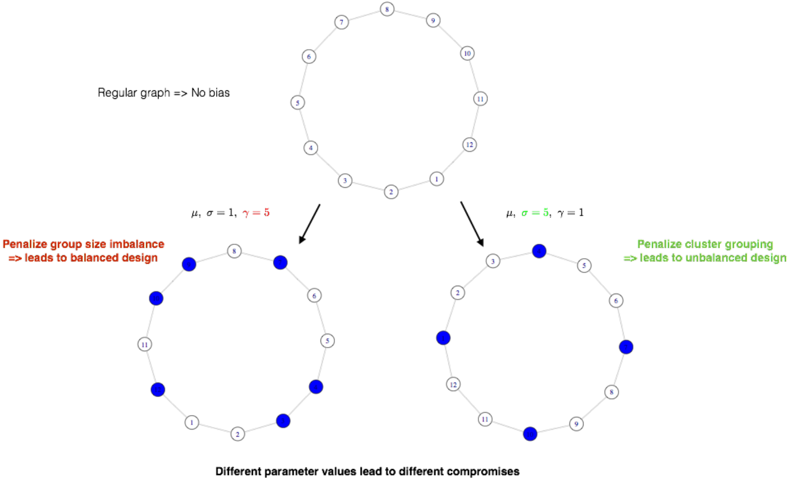
(T_1) within treatment group

(T_2) within control group

(T_3) across groups

$\Rightarrow$ Penalizes cluster-like structures

Simple Network - cycle



Conclusion

What this talk was not about:

- Estimating parameters of the model
- Randomization
- Peer effect (aka interactions)

Thanks!

The Rubin Causal Framework (classical version)

i	Y(0)	Y(1)
1	2	4
2	1	5
3	2	5
4	2.5	4
5	0	3
6	1	4

Estimand: $\tau = \overline{Y(1)} - \overline{Y(0)}$

Estimator: $\hat{\tau} = \overline{Y(1)}^{obs} - \overline{Y(0)}^{obs}$

i	Y(0)	Y(1)	W
1	?	4	1
2	?	5	1
3	?	5	1
4	2.5	?	0
5	0	?	0
6	1	?	0

$\hat{\tau} = 3.17$

i	Y(0)	Y(1)	W
1	?	4	1
2	1	?	0
3	?	5	1
4	2.5	?	0
5	?	3	1
6	1	?	0

$\hat{\tau} = 2.5$

i	Y(0)	Y(1)	W
1	2	?	0
2	1	?	0
3	?	5	1
4	2.5	?	0
5	?	3	1
6	?	4	1

$\hat{\tau} = 2.17$

The Rubin Causal Framework (model-based version)

i	Y(0)	Y(1)
1	2	4
2	2.2	4.2
3	1.5	3.5
4	3	5
5	0.9	2.9
6	1	3

↓

i	Y(0)	Y(1)	W
1	2	?	0
2	2.2	?	0
3	1.5	?	0
4	?	5	1
5	?	2.9	1
6	?	3	1

↓

$\hat{\tau} = 1.7$

i	Y(0)	Y(1)
1	3	5
2	2.9	4.9
3	1	3
4	2.5	4.5
5	2	4
6	1.9	3.9

↓

i	Y(0)	Y(1)	W
1	3	?	0
2	2.9	?	0
3	1	?	0
4	?	4.5	1
5	?	4	1
6	?	3.9	1

↓

$\hat{\tau} = 1.8$

i	Y(0)	Y(1)
1	1	3
2	1.2	3.2
3	3	5
4	1	3
5	2.5	4.5
6	2.5	4.5

↓

i	Y(0)	Y(1)	W
1	1	?	0
2	1.2	?	0
3	3	?	0
4	?	3	1
5	?	4.5	1
6	?	4.5	1

↓

$\hat{\tau} = 2.26$

Model: $(Y(0), Y(1)) \sim F$

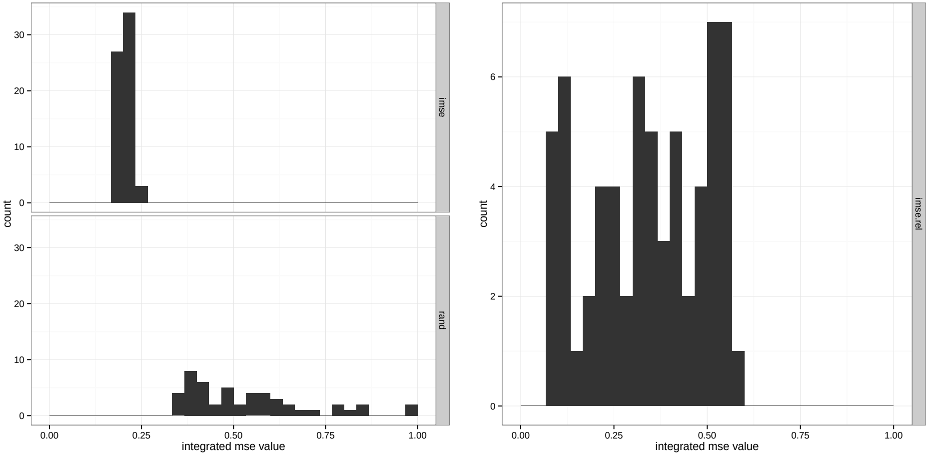
Estimand: $\tau = g(Y(0), Y(1))$ (random variable)

We will consider the following:

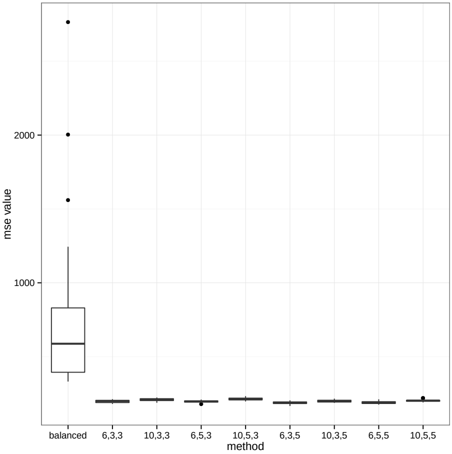
$Y(0) \sim F$

$Y_i(1) = Y_i(0) + \tau \ \forall i$

Comparison with designs ignoring the network structure



Robustness



True parameters: $(\mu, \sigma, \gamma) = (20, 20, 20)$

Design obtained with wrong parameters:
$\mu = 6, 10,\ \sigma = 3, 5,\ \gamma = 3, 5$

Conclusion

Doing something - anything - is better than ignoring the network structure.