Statistical and causal inference in social networks

Dean Eckles
MIT Sloan School of Management

WAW 2015 School

Thanks to Johan Ugander for some materials in this tutorial.

Schedule

1. Processes on networks
2. Causal inference & randomization inference
3. Randomization inference in networks
4. Reducing bias from interference
5. Other experimental designs and non-experimental studies

With breaks :-)

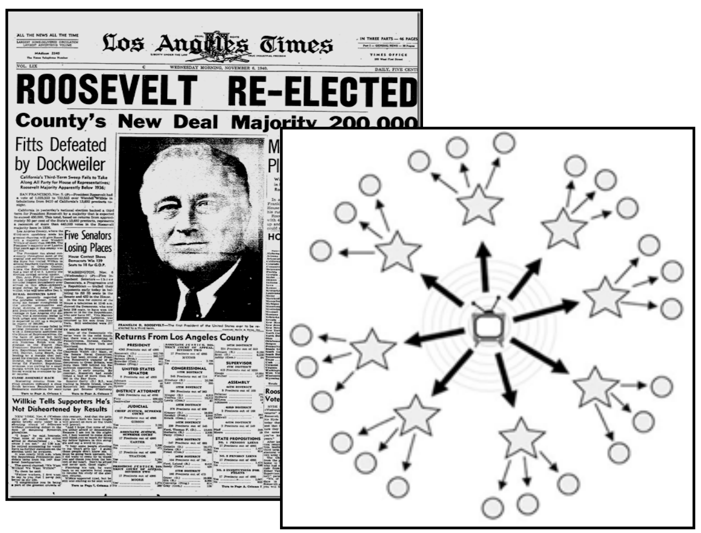
Themes

- Many important questions are causal questions, including questions about processes in networks

- Interesting graph theory and computational problems in causal inference in networks

Processes on social networks

1940 election
two-step theory of
opinion leaders

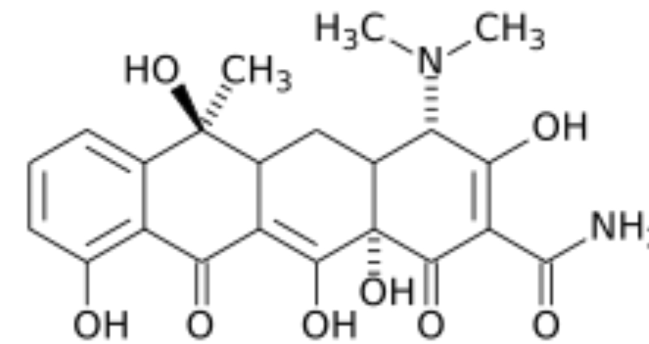


Lazarsfeld et al. '55
Watts-Dodds '07

Hybrid seed corn

Ryan-Gross '43

Tetracycline



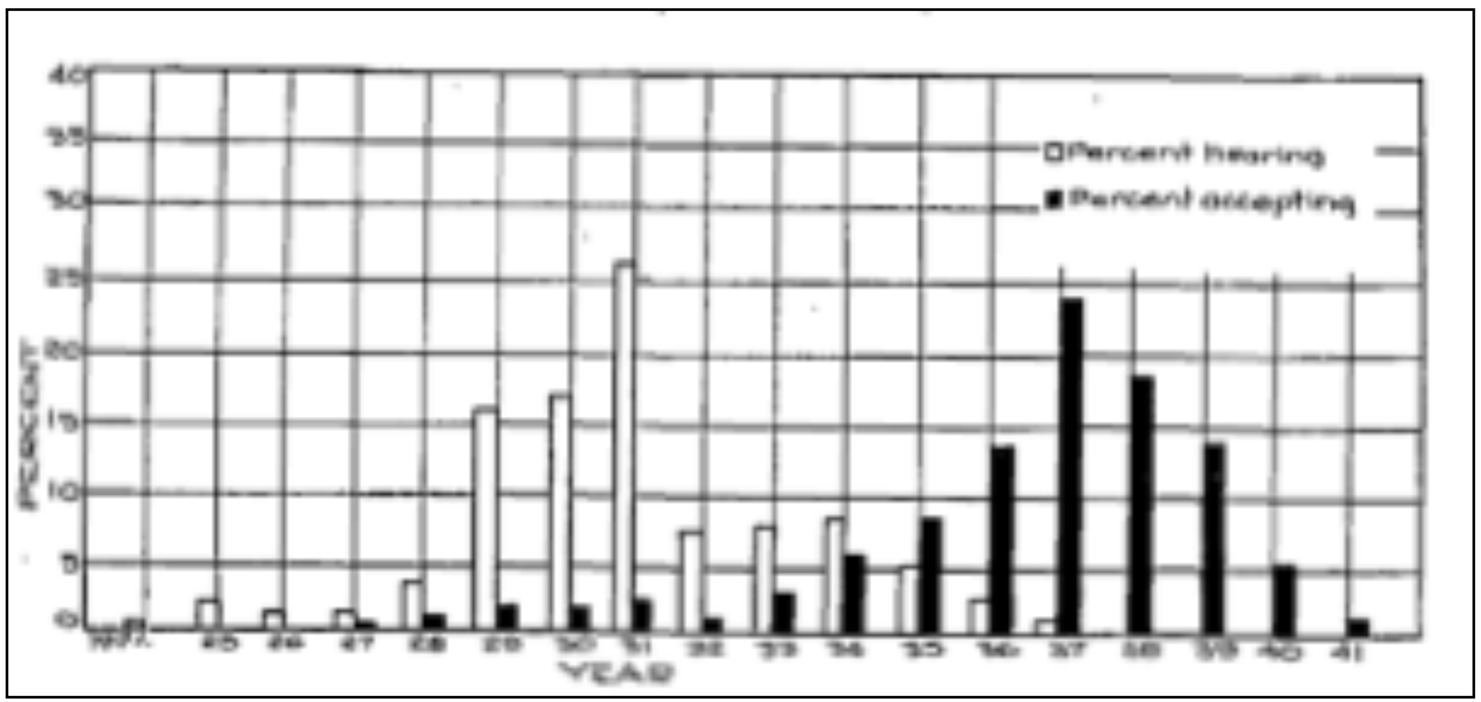
Coleman-Katz-Menzel '57

- B Ryan, N Gross (1943) "The diffusion of hybrid seed corn in two Iowa communities", Rural sociology.
- P Lazarsfeld; B Berelson, H Gaudet (1948) "The People's Choice. How the Voter Makes up His Mind in a Presidential Campaign".
- E Katz, P Lazarsfeld (1955) "Personal Influence, The part played by people in the flow of mass communications".
- E Katz (1957) "The Two-Step Flow of Communication: An Up-To-Date Report on an Hypothesis". Political Opinion Quarterly.
- J Coleman, E Katz, H Menzel (1957) "The diffusion of an innovation among physicians", Sociometry.
- D Watts, P Dodds (2007) "Influentials, Networks, and Public Opinion Formation" Journal of Consumer Research.

Processes on social networks

Hybrid seed corn (Ryan-Gross):

5 stages: awareness, interest, evaluation, trial, adoption

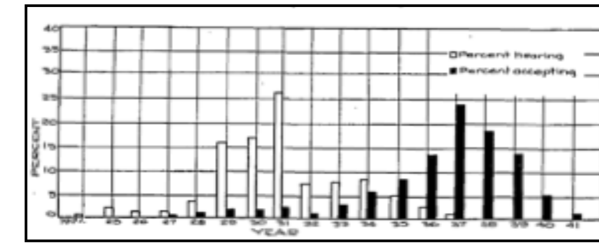


Survey of n=259 farmers

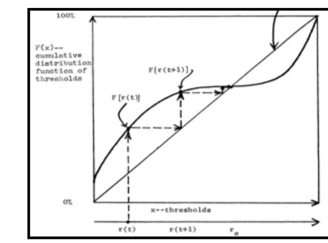
- B Ryan, N Gross (1943) “The diffusion of hybrid seed corn in two Iowa communities”, Rural sociology.
- P Lazarsfeld; B Berelson, H Gaudet (1948) “The People's Choice. How the Voter Makes up His Mind in a Presidential Campaign”.
- E Katz, P Lazarsfeld (1955) “Personal Influence, The part played by people in the flow of mass communications”.
- E Katz (1957) “The Two-Step Flow of Communication: An Up-To-Date Report on an Hypothesis”. Political Opinion Quarterly.
- J Coleman, E Katz, H Menzel (1957) “The diffusion of an innovation among physicians”, Sociometry.
- D Watts, P Dodds (2007) “Influentials, Networks, and Public Opinion Formation” Journal of Consumer Research.

Timeline

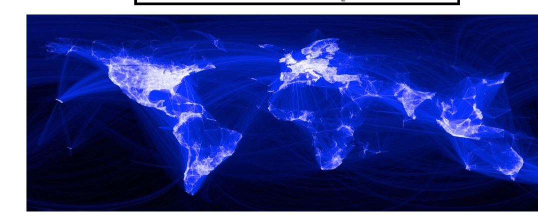
- 1940s-50s: Early theories, early data



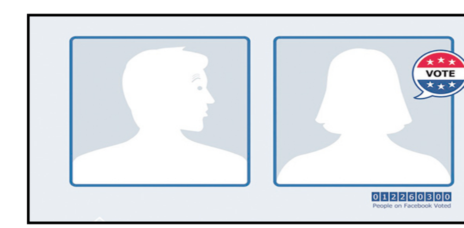
- 1960s-90s: Theory refinement/testing



- 2000s: Large-scale data



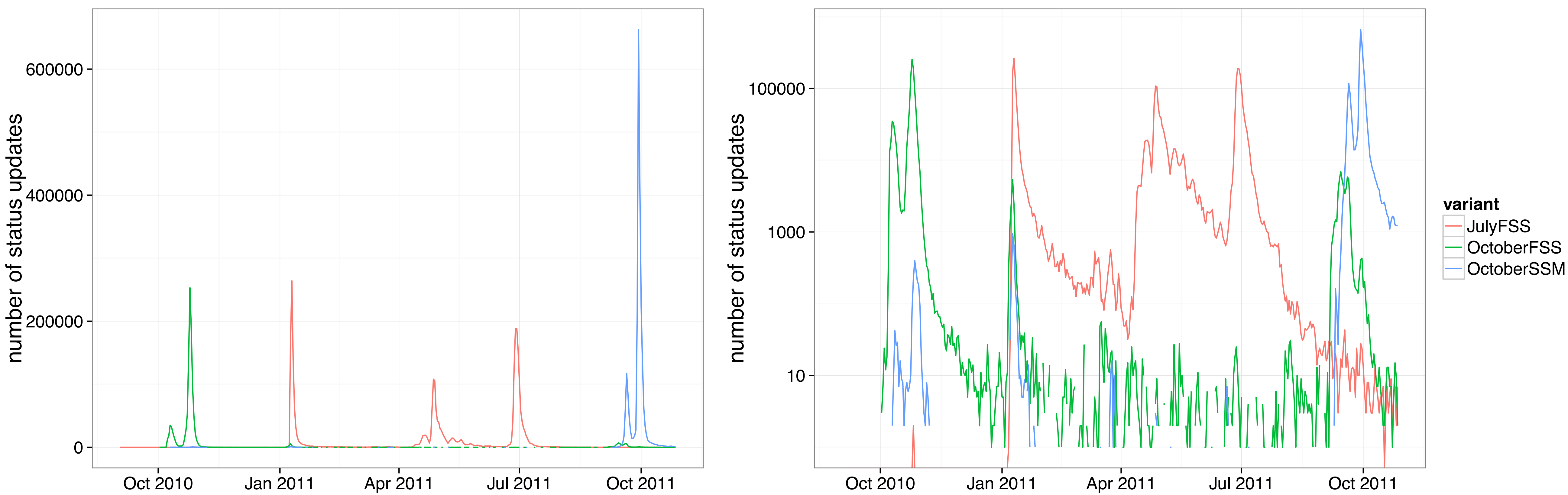
- 2010s: Large-scale experiments



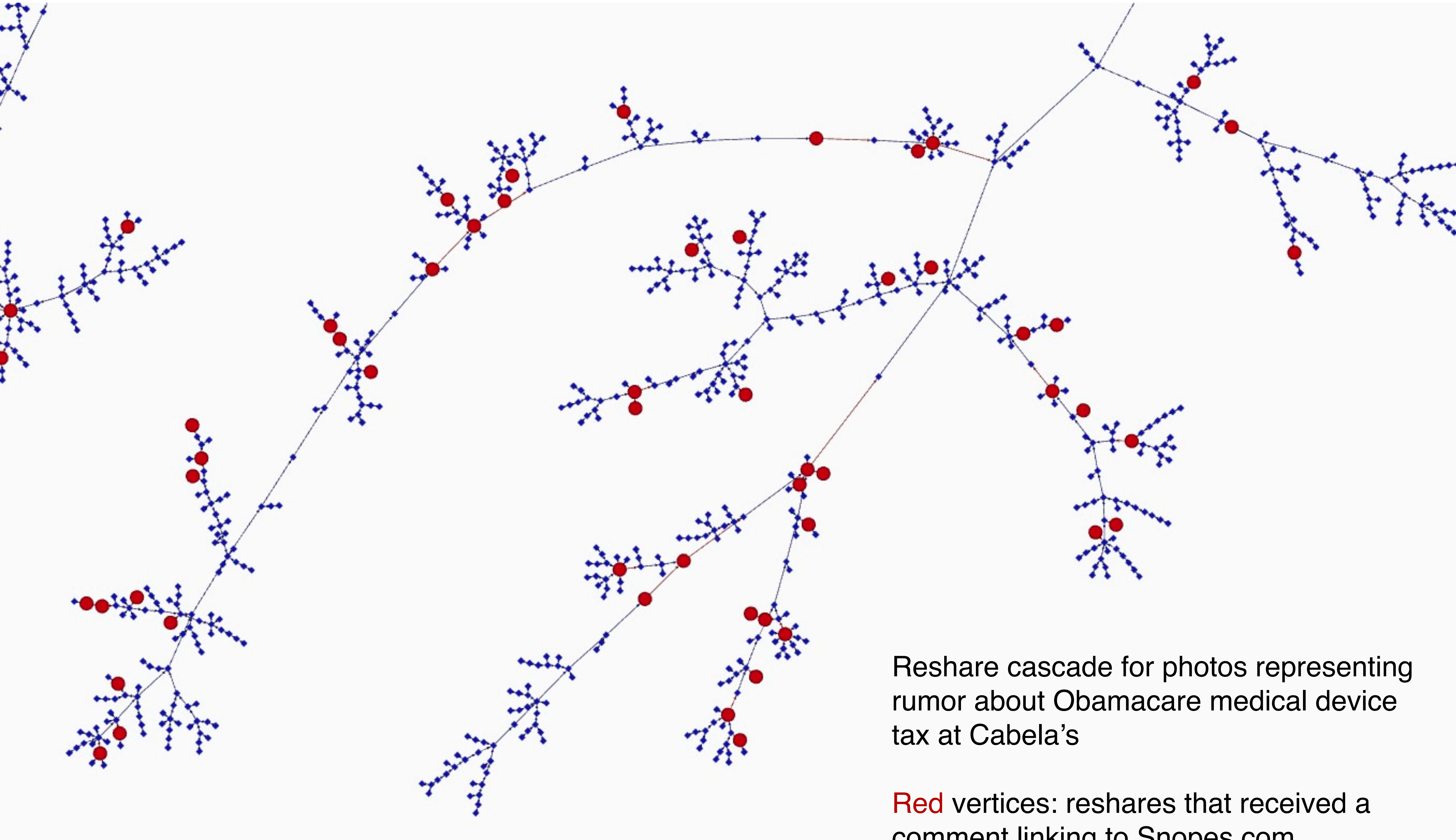
Designing/analyzing experiments to develop/test network theories = Big opportunity

Variations on a copy-paste meme

- Memes seem to appear, disappear, reappear
  - Even the same variant does this: how does it come back?



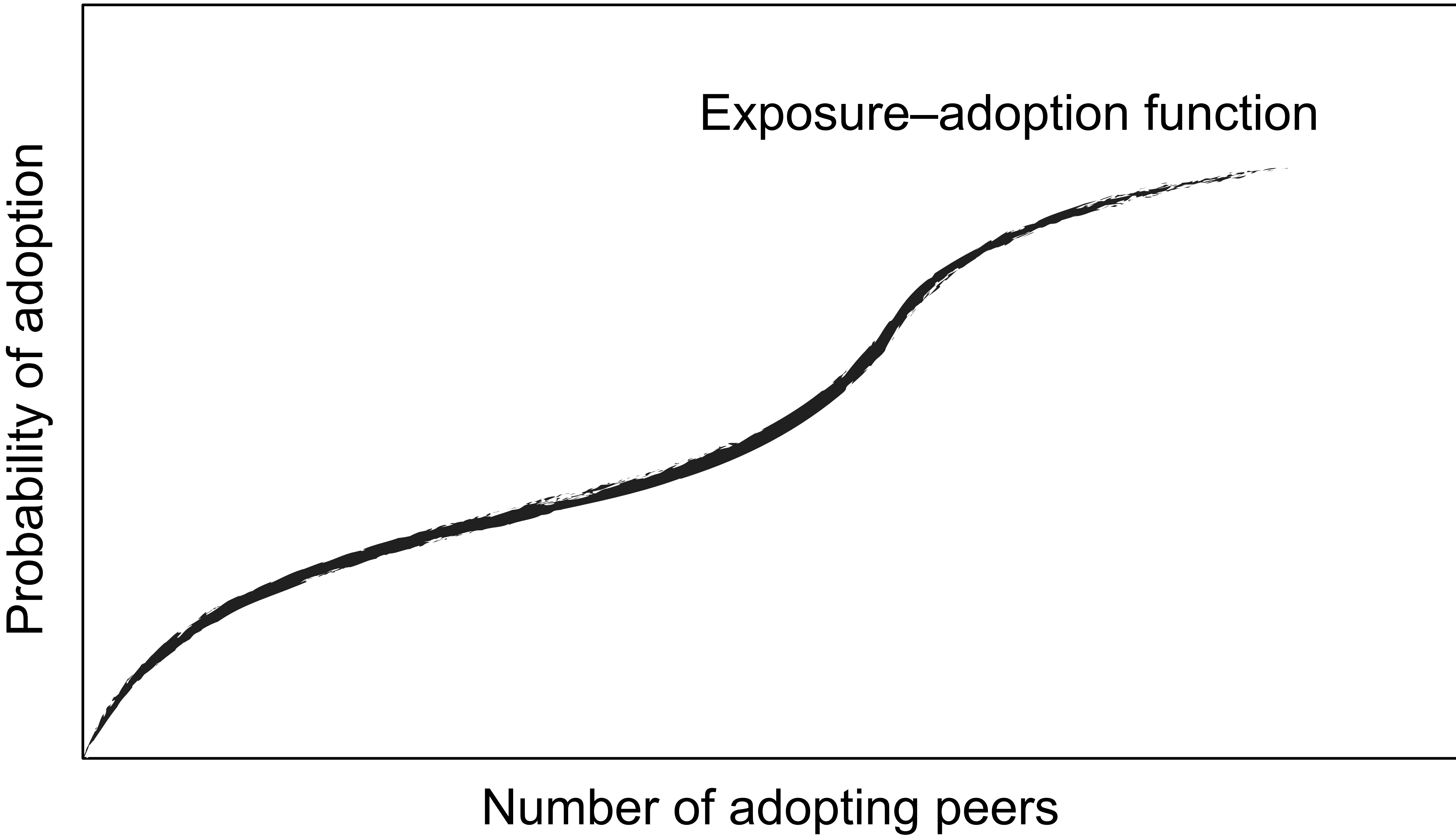
Friggeri, A., Adamic, L., Eckles, D., & Cheng, J. (2014) Rumor cascades. In *Proceedings of ICWSM*. AAAI.

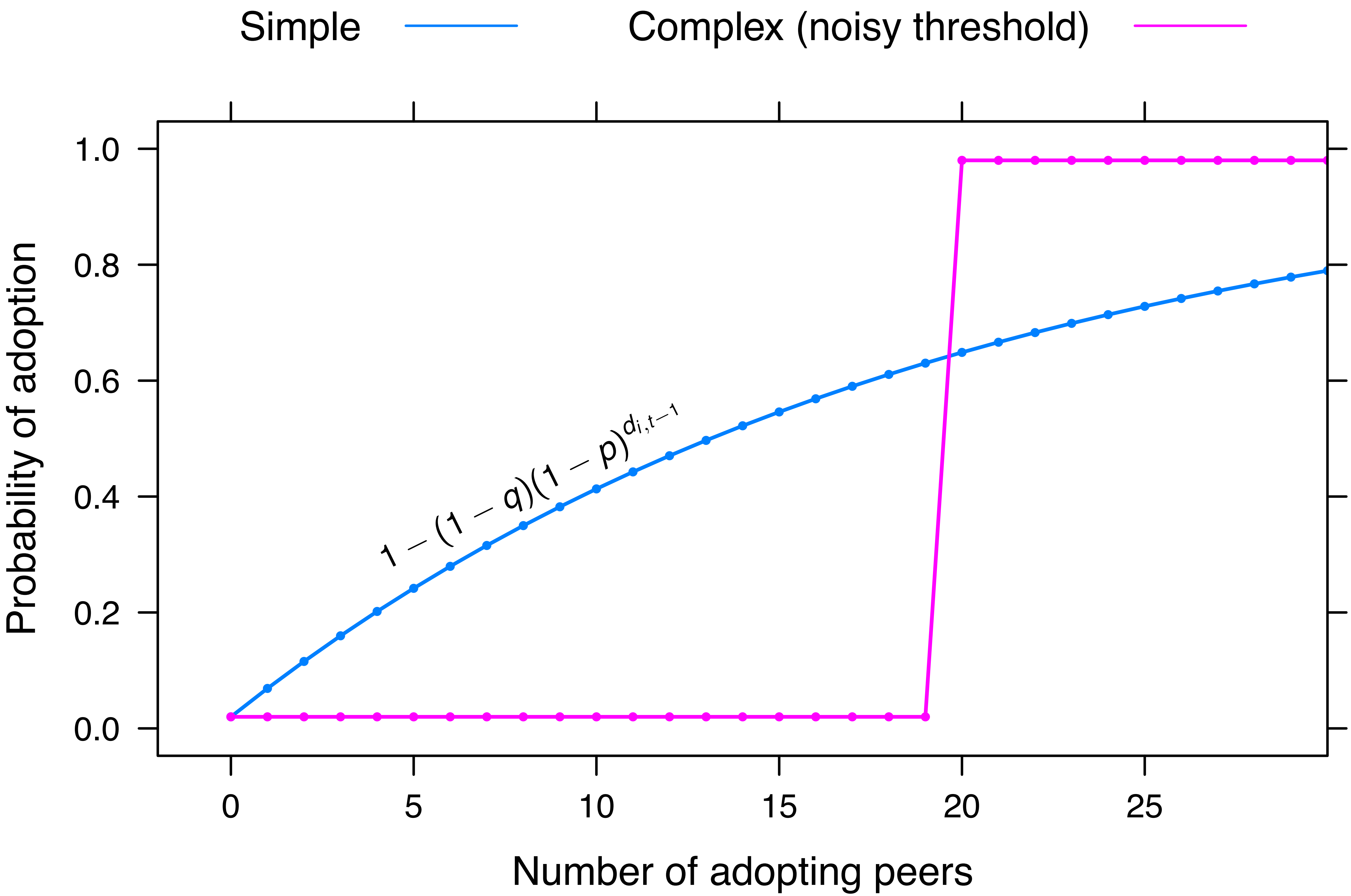


Reshare cascade for photos representing rumor about Obamacare medical device tax at Cabela's

Red vertices: reshares that received a comment linking to Snopes.com

Friggeri, A., Adamic, L., Eckles, D., & Cheng, J. (2014) Rumor cascades. In *Proceedings of ICWSM*. AAAI.


Exposure–adoption function
Probability of adoption
Number of adopting peers



Example exposure–adoption function under simple and complex contagion models

Models of peer effects

Epidemic, or “simple contagion”, models

- Each infected peer has independent probability p of infecting ego
- In discrete time, the probability of node with n infected peers at $t-1$ becoming infected at t is

$$P(Y_{i,t} = 1 | Y_{i,t-1} = 0) = 1 - (1 - q)(1 - p)^{d_{i,t-1}}$$

with $d_{i,t-1} = A_i Y_{t-1}$.

  - For very small p, approximated by linear probability model for all realistic values of n
  - For larger p, diminishing returns from larger n

Models of peer effects

Semi-anonymous graphical games

- A SAGG has *strategic complements* if for all degrees *n* and $d \geq d'$

$$u(1, d, n) - u(0, d, n) \geq u(1, d', n) - u(0, d', n)$$

- Payoff-maximizing behavior is exhausted by a threshold for adoption:

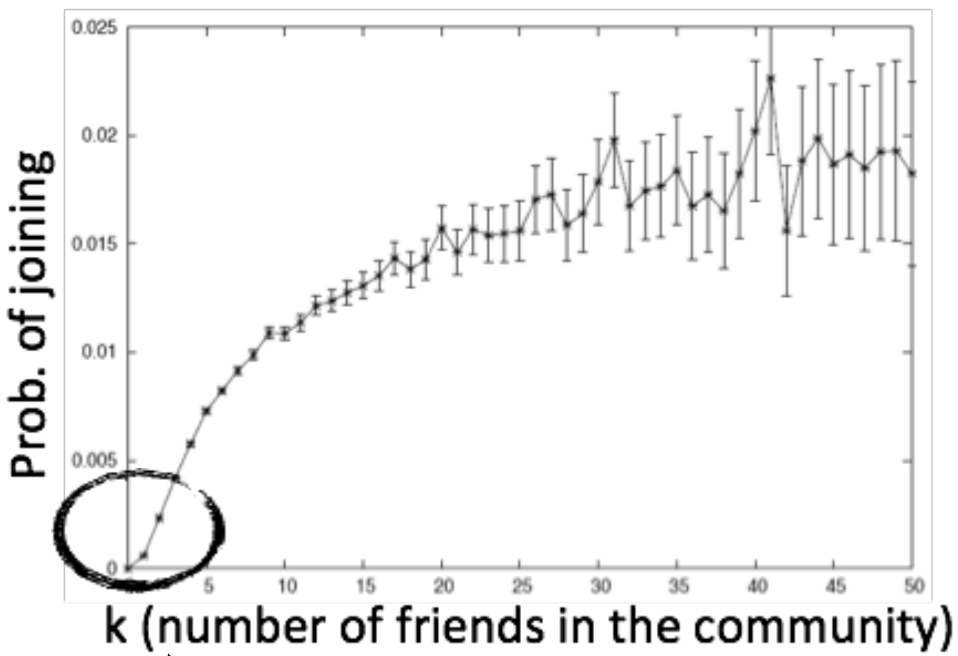
$$Y_{i,t} \leftarrow 1\{d_{i,t-1} \geq q_i\}$$

(Jackson 2008, for a review)

- This is one form of “complex contagion” (Centola & Macy, 2007)

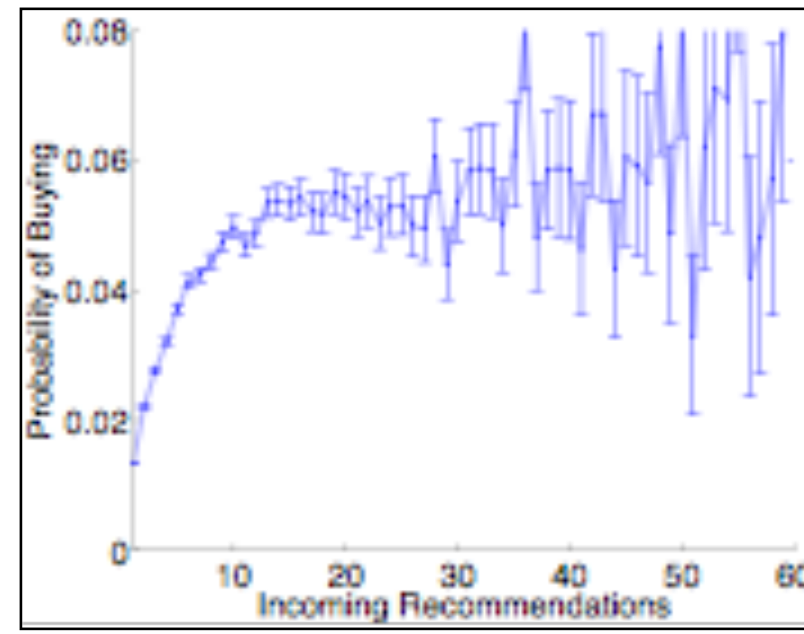
Observational estimates of influence

Backstrom et al. 2006: Probability of joining LiveJournal group



Complex contagion?

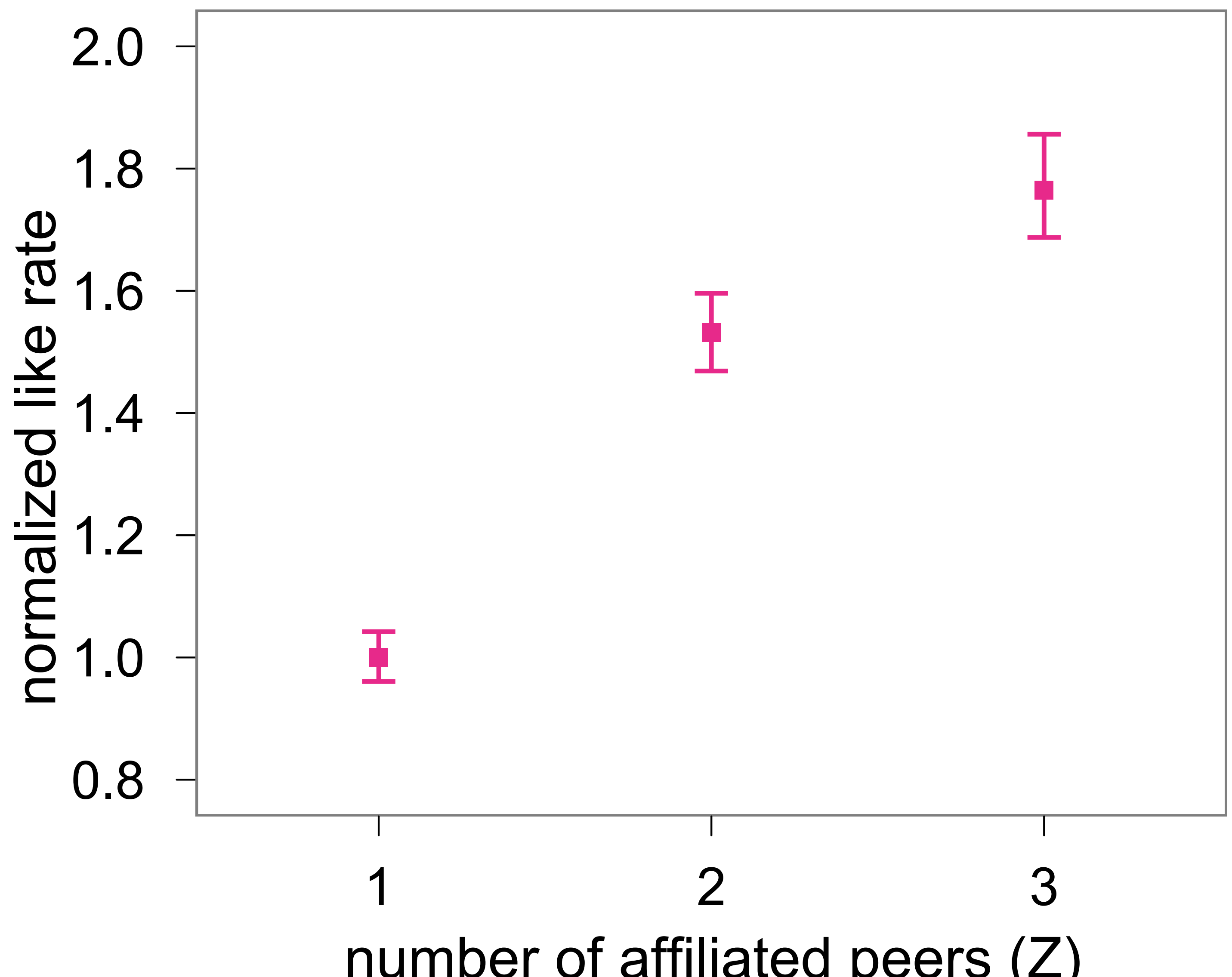
Leskovec et al. 2006: Probability of buying a DVD



- L Backstrom, D Huttenlocher, J Kleinberg, X Lan (2006) "Group formation in large social networks: membership, growth, and evo
- J Leskovec, LA Adamic, BA Huberman (2006) "The dynamics of viral marketing," EC.
- D Centola, V Eguiluz, M Macy (2007) "Cascade dynamics of complex propagation," Physica A.
- D Centola, M Macy (2007) "Complex contagions and the weakness of long ties" American Journal Sociology.

Observational estimates of influence

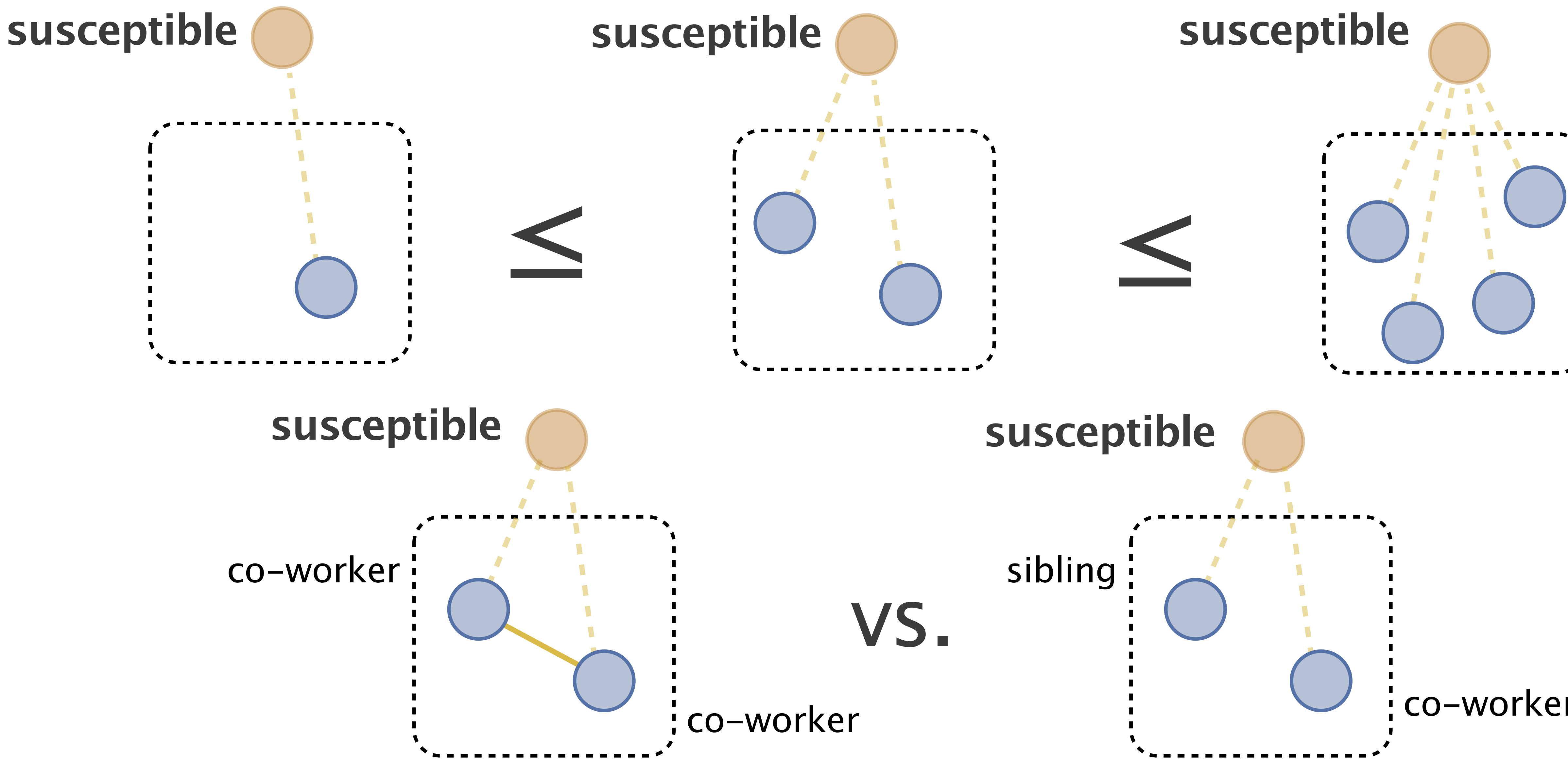
Like rates for ads featuring 1, 2, or 3 friends



Bakshy, E., Eckles, D., Yan, R., & Rosenn, I. (2012). Social influence in social advertising: Evidence from field experiments. In: *EC 2012: Proceedings of the ACM Conference on Electronic Commerce*. ACM. http://arxiv.org/abs/1206.4327

Influence and graph structure

Adoption as a function of 'contact neighborhood' size

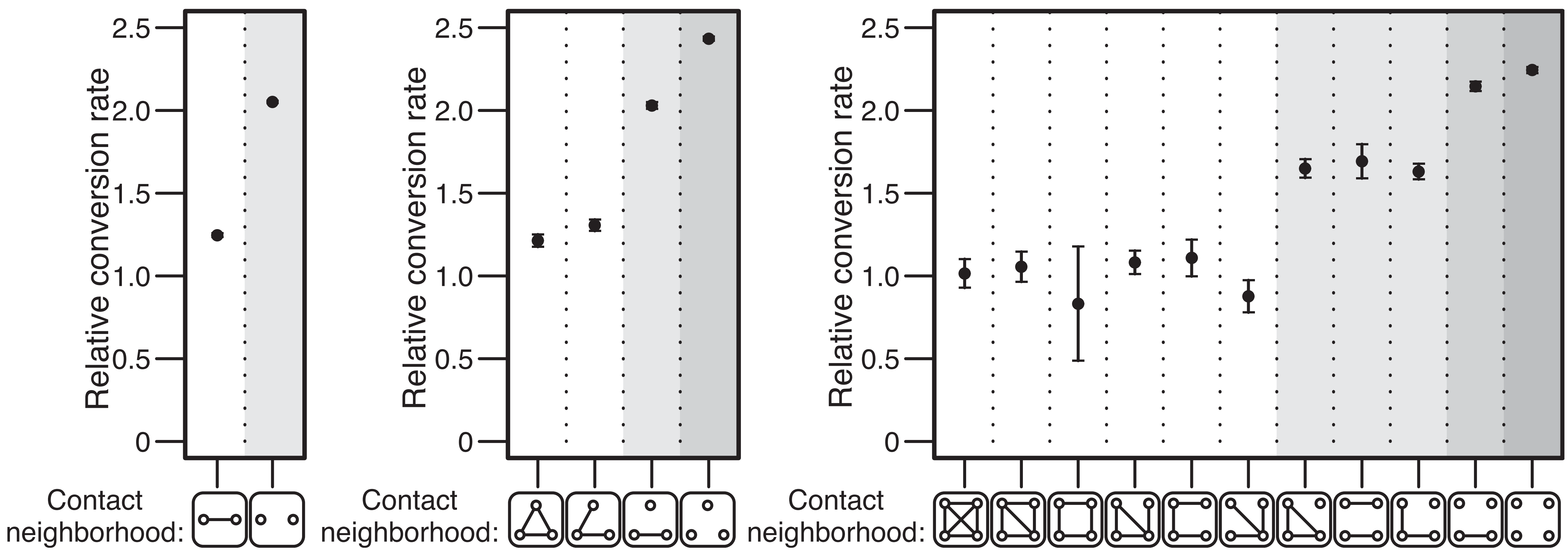


- J Ugander, L Backstrom, C Marlow, J Kleinberg (2012) "Structural diversity in social contagion," PNAS.

Observational estimates of influence

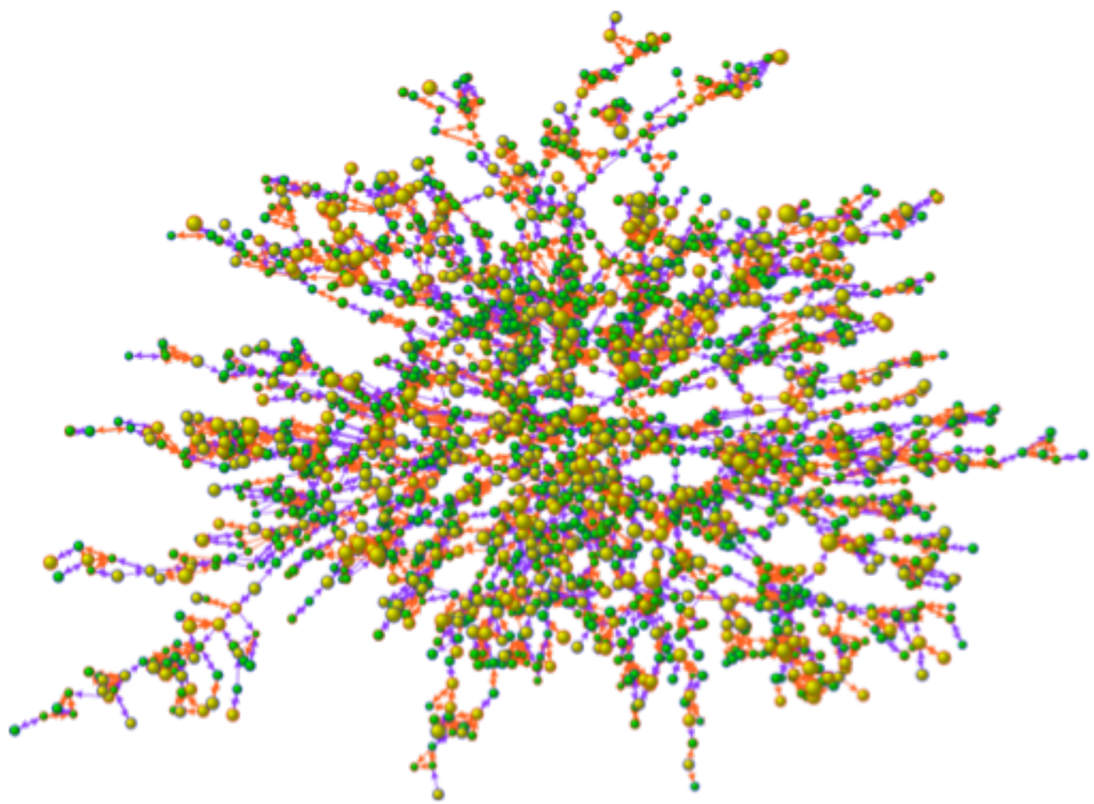
Structural diversity

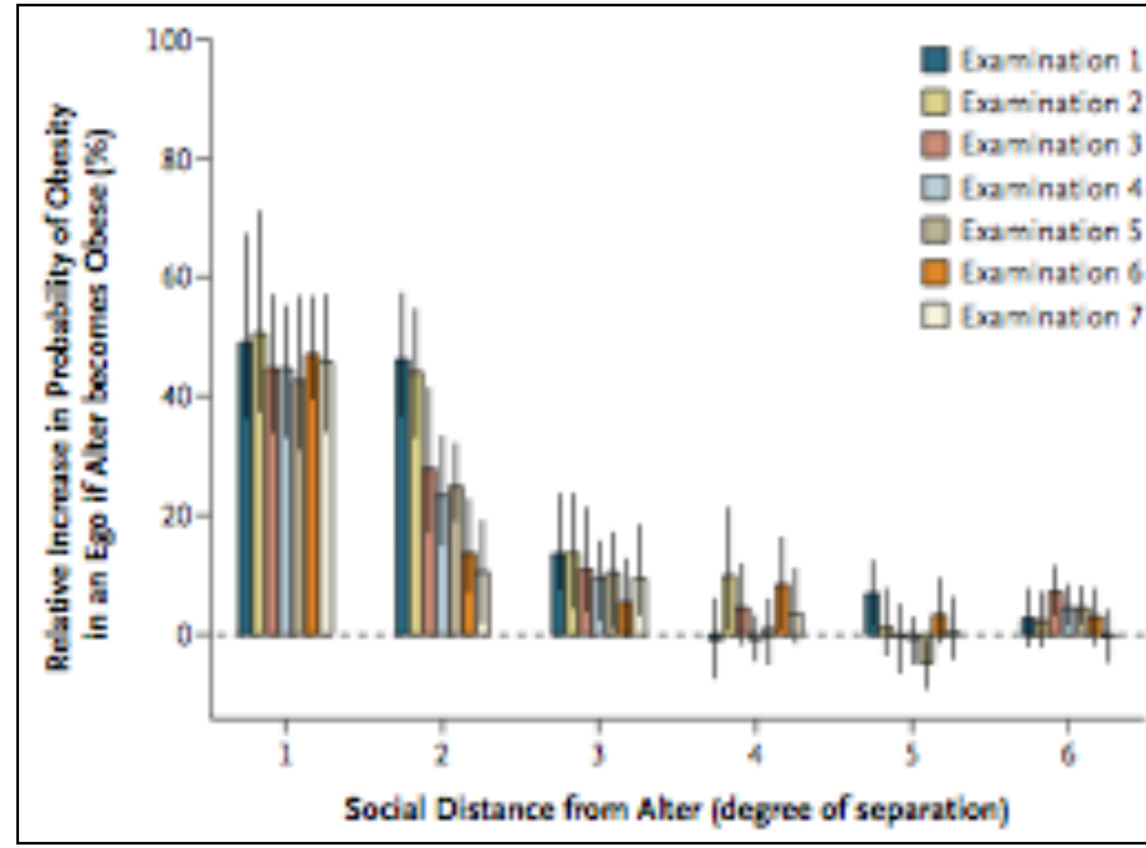
- Joining Facebook given different contact neighborhoods



J Ugander, L Backstrom, C Marlow, J Kleinberg (2012) “Structural diversity in social contagion,” PNAS.

Is obesity contagious?





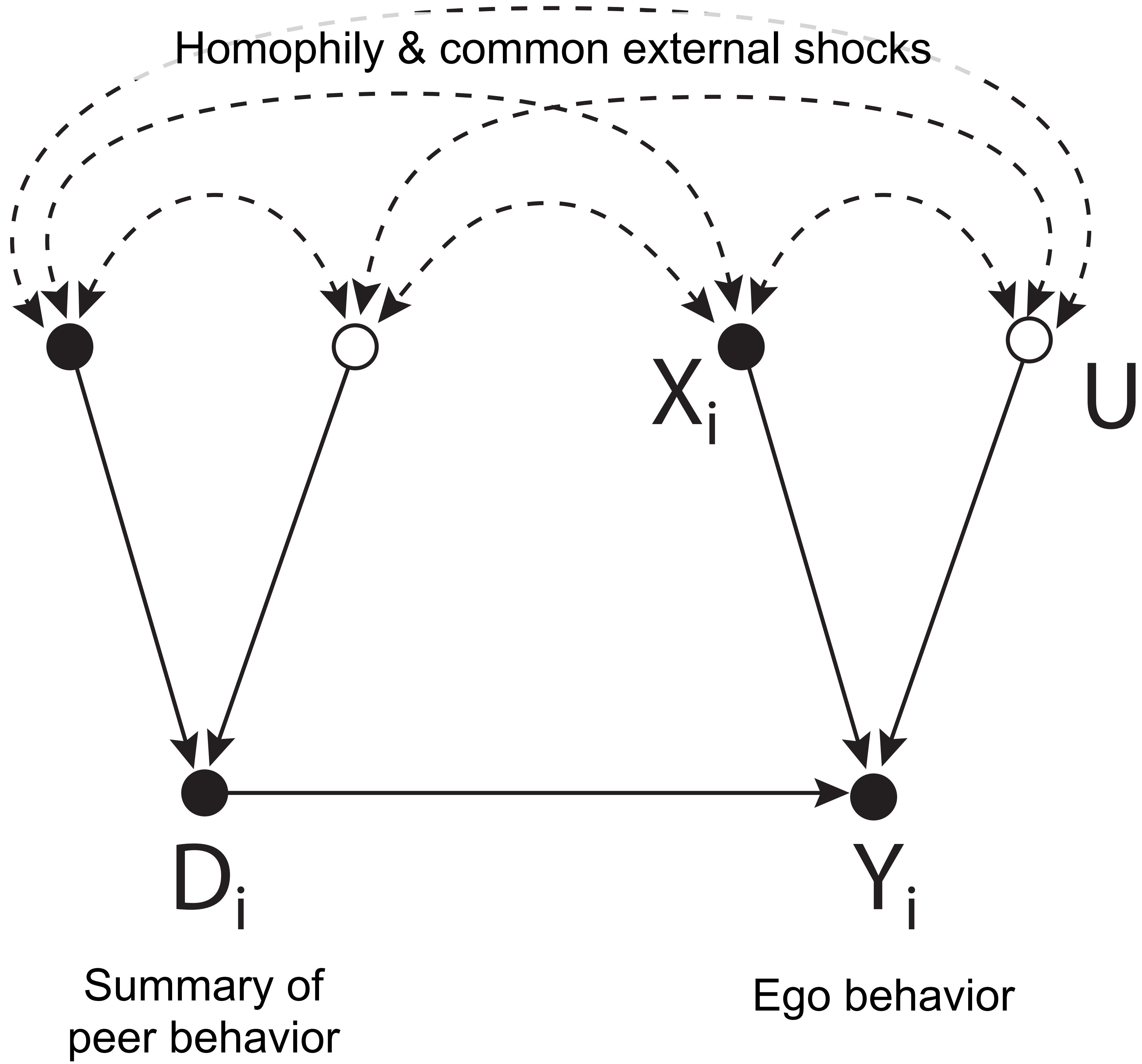
“comparing the conditional probability of obesity in the observed network with the probability of obesity in identical networks (with topology preserved) in which the same number of obese persons is randomly distributed”

N Christakis, J Fowler (2007) "The Spread of Obesity in a Large Social Network over 32 Years," New England J of Medicine.
C Shalizi, A Thomas (2011) "Homophily and contagion are generically confounded in observational social network studies," Sociological Methods & Research.

Why causal inference?

- The central concepts here are inherently causal
- The theories make claims about causal processes, not mere correlations
- The policy response depends on the causal relationships


$Y_j\,(t)$
$Y_i\,(t+1)$
Peer behavior
Ego behavior


Homophily & common external shocks
X_i
U_i
D_i
Y_i
Summary of peer behavior
Ego behavior

When is causal inference possible?

- Can we identify the effect of D on Y by conditioning on available covariates X? That is, in the case of discrete X, do we have that

$$\mathrm{P}(Y|\mathrm{do}(D=d)) = \sum_{x \in \mathcal{X}} \mathrm{P}(Y|D=d, X=x)\,\mathrm{P}(X=x)$$

- Depending on who you're talking to, this may be called
  - Selection on observables
  - Conditional ignorability
  - Conditional unconfoundedness
  - Weak exogeneity

Criteria for identification

- Can we identify the effect of D on Y by conditioning on available covariates X? That is, in the case of discrete X, do we have that

$$\mathrm{P}(Y|\mathrm{do}(D=d)) = \sum_{x \in \mathcal{X}} \mathrm{P}(Y|D=d, X=x)\,\mathrm{P}(X=x)$$

- Sufficient condition: back-door criterion
- A set of variables X satisfies the back-door criterion for identification of the effect of D on Y if
  a. X blocks every path from D to Y that has an arrow into D
  b. D is not a parent of any member of X

Goals

Test for or estimate peer effects

- How does a marginal peer adopting affect your adoption?

Test for or estimate spillovers

- How does treating a peer affect your outcome?

Estimate effects of global treatment

- What would happen if we gave everyone the treatment?

Simplest case: Causal inference for spillovers

- How can we tell if my outcome is affected by my peers' treatment?

- Plenty of jargon for this:
  - Spillovers
  - Exogenous peer effects (Manski)
  - Interference (Cox)

- If the treatments are randomly assigned, this might seem easy

Even simpler case

Effects of randomly assigned treatment on same unit

- How can we tell whether a unit's outcome is affected by its own random assignment?

Causal quantities

Difference in potential outcomes

- What would a unit have done under treatment and under control?
- The **(causal) effect** for a unit is the difference between what the ego *would have done* under different peer behaviors

$$\Delta_i = Y_i^{(1)} - Y_i^{(0)}$$

$Y_i^{(0)}$ potential outcome for unit *i* if assigned to control

$Y_i^{(1)}$ potential outcome for unit *i* if assigned to treatment

Will sometimes use more verbose notation, e.g., $Y_i(Z_i = 1)$ or $Y_i(W_i = 1)$

Causal inference

- Fundamental problem of statistical inference:
  We only observe data for some sample of units.

- Fundamental problem of causal inference (Holland 1988):
  We can only observe one potential outcome for each unit.

- Additional assumptions (e.g., time and order don’t matter) and design (e.g., within-subjects designs), we could observe multiple potential outcomes per unit

Randomization inference

Basic case

Table 5.5: First Six Observations from Honey Study with missing potential outcomes in brackets filled in under the null hypothesis of no effect

Unit	Potential Outcomes $Y_i(0)$	$Y_i(1)$	Observed Variables Treatment W_i^{obs}	X_i	Y_i^{obs}	$\text{rank}(Y_i^{\text{obs}})$
1	(3)	3	1	4	3	4
2	(5)	5	1	6	5	6
3	(0)	0	1	4	0	1.5
4	4	(4)	0	4	4	5
5	0	(0)	0	1	0	1.5
6	1	(1)	0	5	1	3

from Imbens & Rubin (2015, Ch. 5)

Randomization inference

Basic case: Null hypothesis of no treatment effect

Consider the null hypothesis that the treatment has no effects.

Null hypothesis of no treatment effects:

$$Y_i(W_i = 1) = Y_i(W_i = 0) \text{ for all } i$$

This is a *sharp null hypothesis*: we can infer all potential outcomes from observed outcomes.

Note we are also assuming no interference between units, so

$$Y_i(\boldsymbol{W}) = Y_i(\boldsymbol{W'})$$

Randomization inference

Basic case: Null hypothesis of no treatment effect

For fixed $\mathbf{Y}(0)$ and $\mathbf{Y}(1)$ with $Y_i(0) = Y_i(1)$, the distribution of

$$T(\mathbf{W}, \mathbf{Y}(\mathbf{W})) = \frac{1}{N/2} \sum_{i:W_i=1} Y_i^{\text{obs}} - \frac{1}{N/2} \sum_{i:W_i=0} Y_i^{\text{obs}}$$

under the randomization distribution is known.

The exact (finite sample) p-value associated with the null hypothesis is

$$\text{p-value} = \text{pr}(|T(\mathbf{W})| \geq |T^{\text{obs}}|)$$

Randomization inference recipe

1. Choose a test statistic $T(\boldsymbol{Y}, \boldsymbol{W})$ that is sensitive to expected effects
   - Difference in means, difference in logged means, etc.
2. Compute observed value of test statistic $T(\boldsymbol{Y}^{obs}, \boldsymbol{W}^{obs})$
3. Note that under sharp null, $\boldsymbol{Y}(\boldsymbol{W}) = \boldsymbol{Y}(\boldsymbol{W}') = \mathbf{Y}^{\text{obs}}$
4. Draw permuted treatment vector $\boldsymbol{W}^*$, consistent with original randomization method
5. Compute value of test statistic with observed outcomes and permuted treatment $T(\boldsymbol{Y}^{obs}, \boldsymbol{W}^*)$
6. Repeat 4 & 5 for R times. Compare observed and null test statistics.

Example

Jupyter notebook: Randomization inference

https://github.com/deaneckles/randomization_inference

Causal inference for spillovers

- How can we tell if my outcome is affected by my peers' treatment?

- Simple case:
  Population consists of isolated dyads
  - At least we know the dyads are independent of each other

- More interesting case:
  Population is a single connected component
  - Each unit might be affected by all other units' treatments

Causal inference for spillovers

Difference in potential outcomes

- The **spillover* effect** for an ego is the difference between what the ego *would have done* under different peer treatments
  - e.g., has peer who performs behavior (treatment) vs. does not have peer who performs behavior (control)

$$\Delta_i = Y_i^{(1)} - Y_i^{(0)}$$

$Y_i^{(0)}$ potential outcome for unit *i* if assigned to have non-adopter peer

$Y_i^{(1)}$ potential outcome for unit *i* if assigned to have adopter peer

* Or “total indirect effect”, “interference effect”, “exogenous peer effect” …

Setup

Exact p-values for network interference.
Susan Athey, Dean Eckles & Guido W. Imbens
NBER Working Paper No. 21313. (and arXiv)

We have a finite population $\mathbb{P}$ with N units. These units may be linked through a network with adjacency matrix $\mathbf{A}$. We also measure covariates on the individuals, with $\mathbf{X}$ the matrix of covariates.

The units are exposed to a treatment $\mathbf{W}$, where $\mathbf{W}$ is an N-vector with ith element W_i. $\mathbf{W}$ takes on values in $\mathbb{W}$.

For each unit there is a set of potential outcomes $Y_i(\mathbf{w})$, one for each $\mathbf{w} \in \mathbb{W}$. We observe $Y_i^{\text{obs}} = Y_i(\mathbf{W})$.

Causal effects are comparisons $Y_i(\mathbf{w}) - Y_i(\mathbf{w}')$ for any pair $\mathbf{w} \neq \mathbf{w}' \in \mathbb{W}$

Example null hypotheses of interest

No treatment effects:

$Y_i(\mathbf{w}) = Y_i(\mathbf{w}')$ for all units i, and all pairs of assignments $\mathbf{w}, \mathbf{w}' \in \mathbb{W}$.
(straightforward because this hypothesis is sharp)

No spillover effects: (but own treatment effects)

$Y_i(\mathbf{w}) = Y_i(\mathbf{w}')$ for all units i, and all pairs of assignment vectors $\mathbf{w}, \mathbf{w}' \in \mathbb{W}$ such that $w_i = w_i'$.

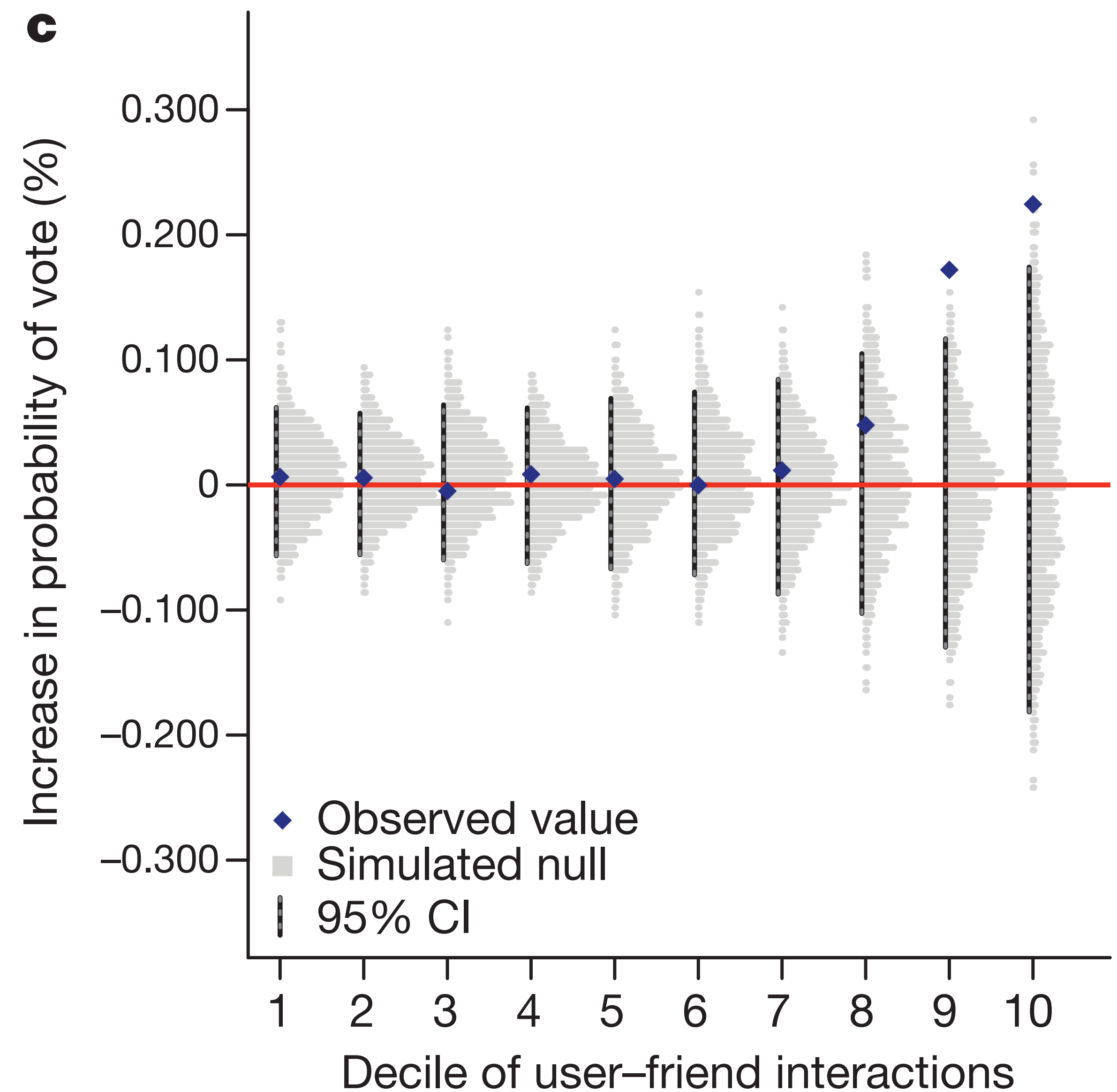
No higher order effects: (but effects of own treatment and friends' treatment)

$Y_i(\mathbf{w}) = Y_i(\mathbf{w}')$ for all units i, and for all pairs of assignment vectors $\mathbf{w}, \mathbf{w}' \in \mathbb{W}$ such that $w_j = w_j'$ for all units j such that $d(i, j) < 2$ (distance in network).

Naïve randomization inference for spillovers

Bond, Fariss, Jones, Kramer, Marlow, Settle and Fowler ("A 61-million-person experiment ...", *Nature* 2012) write:

"The messages not only influenced the users who received them but also the users' friends, and friends of friends."



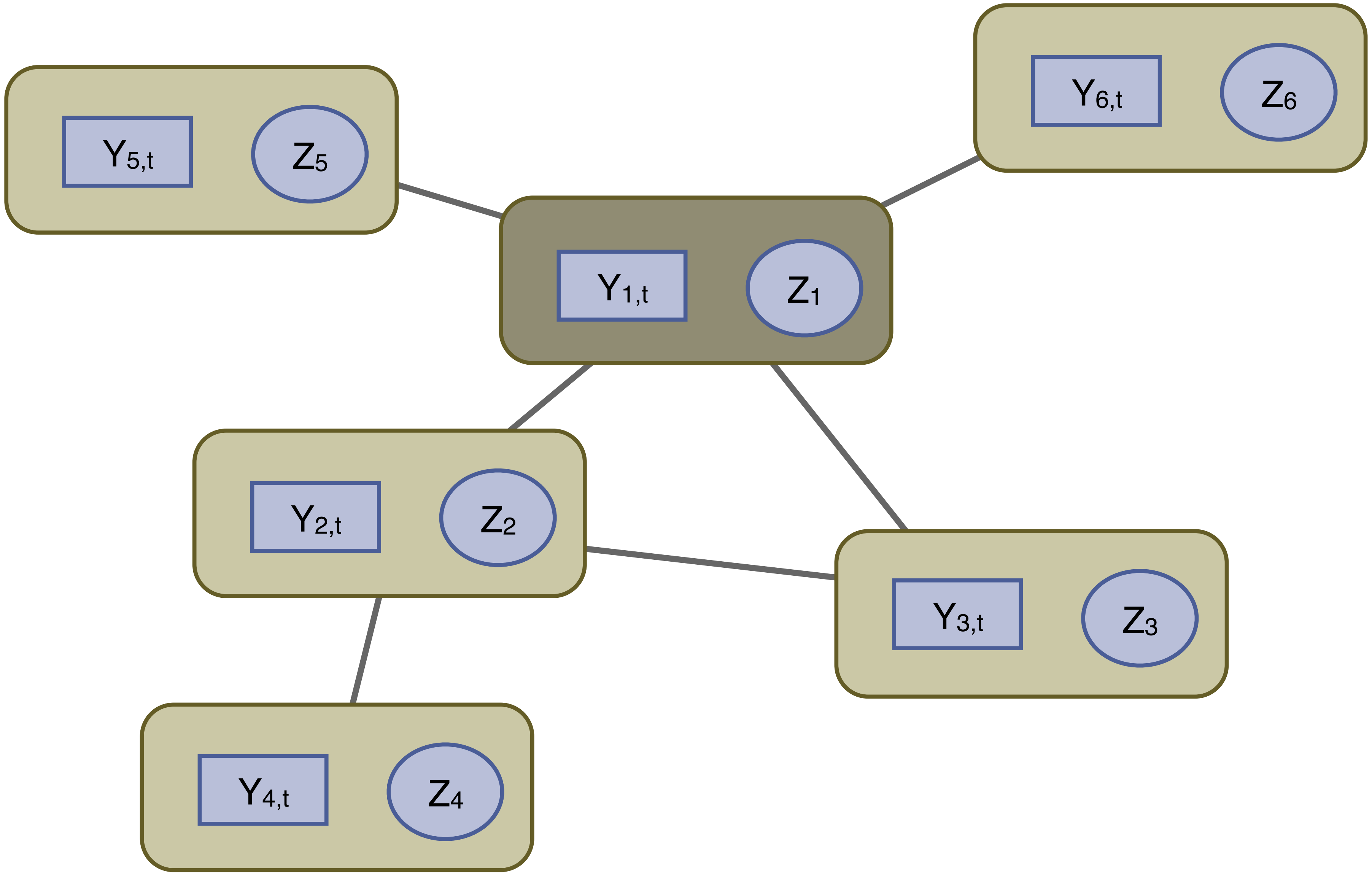
Naïve randomization inference for spillovers

1. Choose a test statistic $T(\boldsymbol{Y}, \boldsymbol{W})$ that obviously measures interference.
   - Coefficient for regression of outcome on number of treated peers
   - Edge-level contrast between edges with treated and control peer
2. Compute observed value of test statistic $T(\boldsymbol{Y}^{obs}, \boldsymbol{W}^{obs})$
3. Draw permuted treatment vector $\boldsymbol{W}^*$, consistent with original randomization method
4. Compute value of test statistic with observed outcomes and permuted treatment $T(\boldsymbol{Y}^{obs}, \boldsymbol{W}^*)$
5. Repeat 4 & 5 for R times. Compare observed and null test statistics.

This recipe can result in Type I error rates 2 times too large!

Randomization inference for spillovers

1. Select set of focal units F for which you will examine outcomes
2. Choose any test statistic $T(\boldsymbol{Y}_F, \boldsymbol{W})$ that is a function of treatments and only focal units' outcomes
3. Compute observed value of test statistic $T(\boldsymbol{Y}_F^{obs}, \boldsymbol{W}^{obs})$
4. Draw permuted treatment vector $\boldsymbol{W}^*$ such that all focal units get the same treatment as observed – i.e. $W_i^* = W_i^{obs}$ for all i in F
5. Compute value of test statistic with observed outcomes and permuted treatment $T(\boldsymbol{Y}_F^{obs}, \boldsymbol{W}^*)$
6. Repeat 4 & 5 for R times. Compare observed and null test statistics.



Z_i = treatment
$Y_{i,t}$ = response at time t

Conditional randomization inference

“Artificial experiments”

- In basic independent randomized experiment
  - Condition on how many units are treated N_1
  - Permute (keeping N_1 fixed) rather than re-randomize independently

- General case (including networks)
  - Condition on some units getting same treatments as we observed, such that null hypothesis is now “sharp”
  - Draw from randomization distribution conditional on this

Test statistics for interference

Score statistic

- Use linear regression coefficient as test statistic

$$Y_i^{\text{obs}} = \alpha_0 + \alpha_W \cdot W_i + \alpha_y \cdot \overline{Y}_{(i)}^{\text{obs}} + \varepsilon_i$$

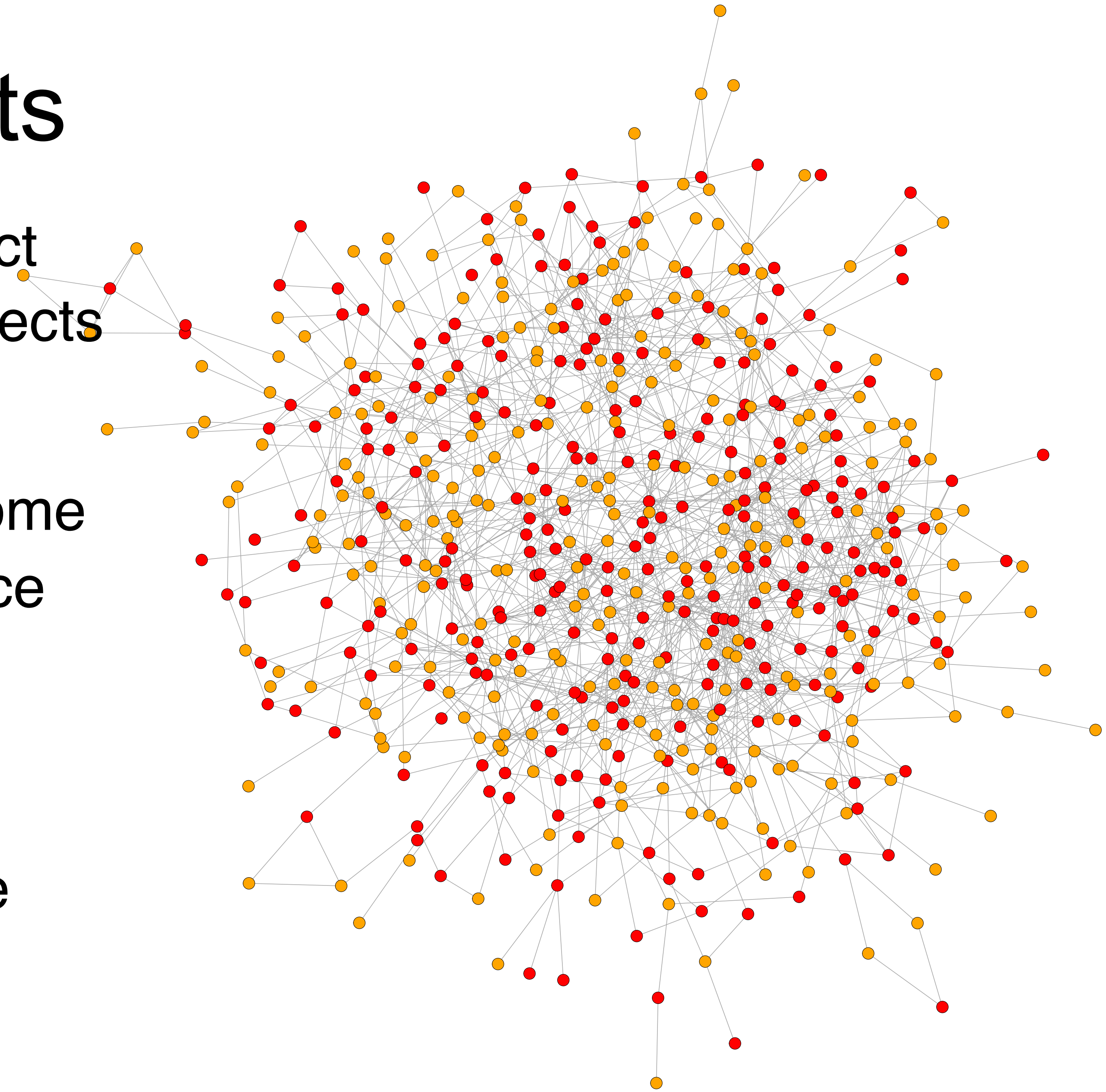
$\overline{\mathbf{G}}$ is the row-normalized adjacency matrix

$$T_{\text{score}} = \frac{1}{N_F} \sum_{i \in \mathbb{P}_F} \left\{ \left(Y_i^{\text{obs}} - \overline{Y}_{F,0}^{\text{obs}} - W_i \cdot \left(\overline{Y}_{F,1}^{\text{obs}} - \overline{Y}_{F,0}^{\text{obs}} \right) \right) \times \sum_{j=1}^{N} \left(\overline{G}_{ij} \cdot W_j - \overline{\overline{G} \cdot W} \right) \right\}$$

T_A is average of indicator of having at least one treated friend.

Selecting focal units

- Any choice is valid (i.e. correct Type I error rates), choice affects power
- If only some units have outcome data, that might simplify choice
- Otherwise, can use greedy heuristic methods – will come back to this



AddHealth network with example selection of focal and auxiliary units

Banner exposure

Intensity of social cues

Call to action: verb or noun?

News Feed exposure

Example

Jupyter notebook: Randomization inference in networks

https://github.com/deaneckles/randomization_inference

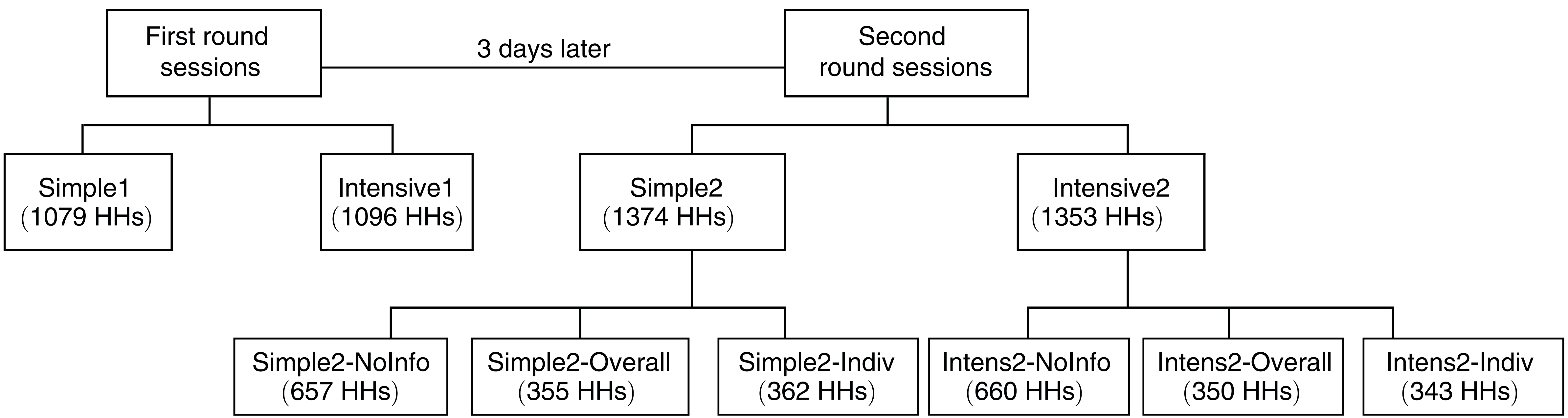
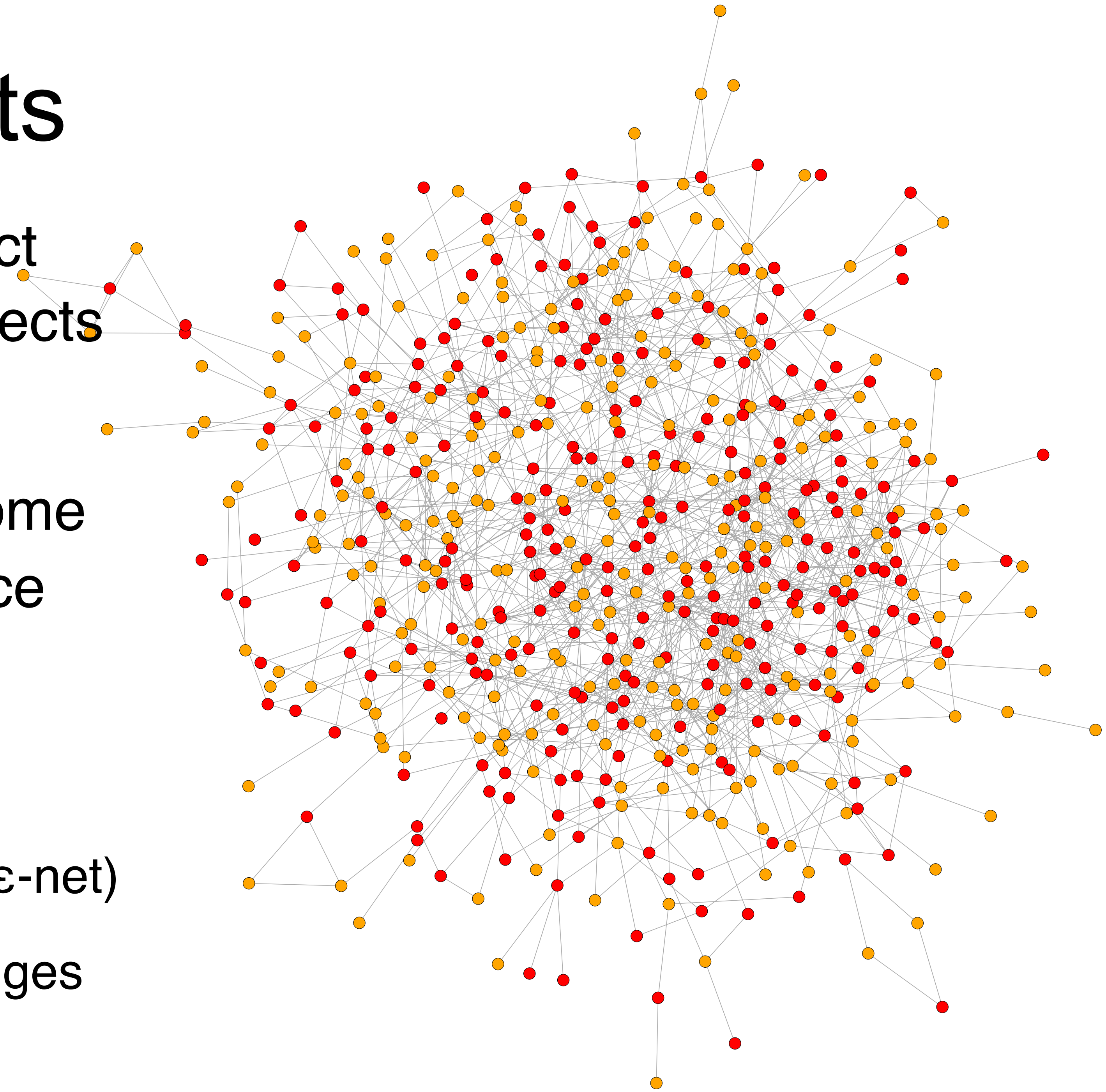


FIGURE 1.1. EXPERIMENTAL DESIGN: WITHIN-VILLAGE, HOUSEHOLD-LEVEL RANDOMIZATION

Cai, Jing, Alain De Janvry, and Elisabeth Sadoulet. 2015. "Social Networks and the Decision to Insure." American Economic Journal: Applied Economics, 7(2): 81-108.

Selecting focal units

- Any choice is valid (i.e. correct Type I error rates), choice affects power
- If only some units have outcome data, that might simplify choice
- Otherwise, can use greedy heuristic methods
  - Find maximal independent set (ε-net)
  - Select focal units maximizing edges between focal and auxiliary subgraphs ($\delta_{N,i}$)



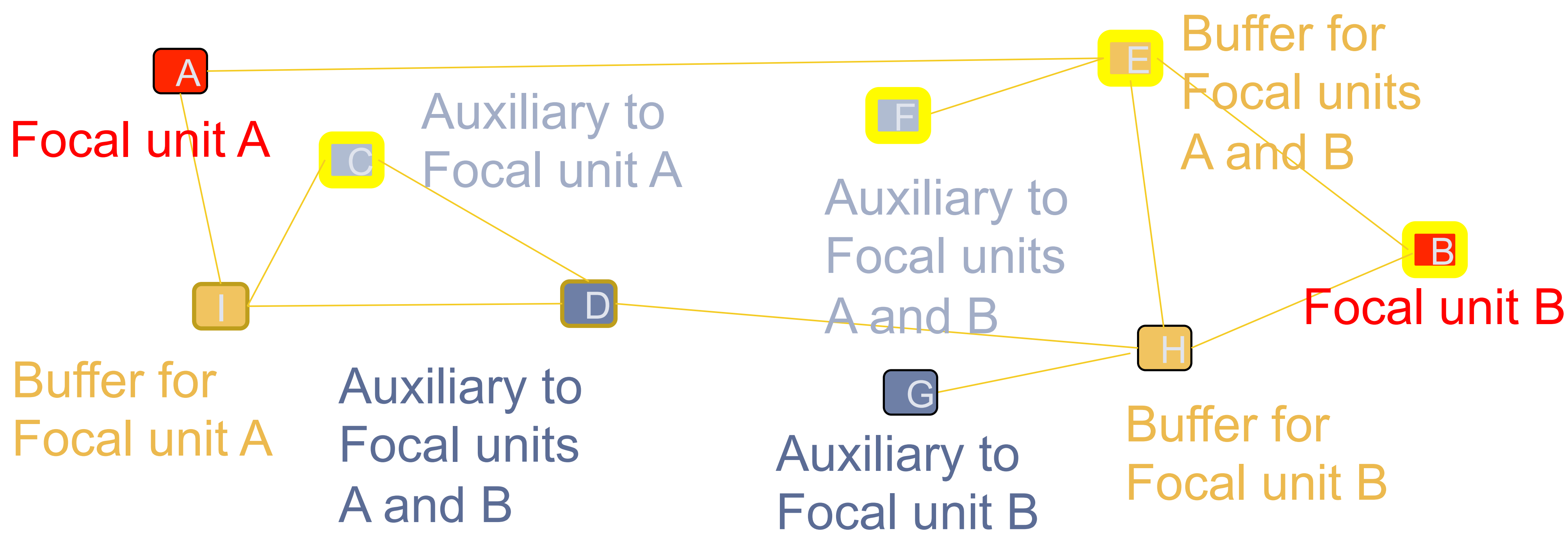
AddHealth network with example selection of focal and auxiliary units

Simulation results

Network	Statistic	Own Effect	Spillover Effect	Focal Node Selection: Random	ε-net	$\delta_{N,i}$
AddHealth	T_{score}	0	0	0.059	0.056	0.045
	T_{elc}	0	0	0.058	0.054	0.044
	T_A	0	0	0.059	0.039	0.046
	T_{score}	4	0	0.056	0.053	0.051
	T_{elc}	4	0	0.051	0.048	0.059
	T_A	4	0	0.050	0.053	0.051
	T_{score}	0	0.4	0.362	0.463	0.527
	T_{elc}	0	0.4	0.174	0.299	0.413
	T_A	0	0.4	0.141	0.296	0.327
	T_{score}	4	0.4	0.346	0.461	0.529
	T_{elc}	4	0.4	0.083	0.102	0.123
	T_A	4	0.4	0.069	0.088	0.116

Type I error rates are correct for 5% test

Power greatly affected by test statistic and choice of focal units



Aux FOF treat v. control:
A has C,F v. D
B has F v. D,G

Unit	Y_i(0 FOF*)	Y_i(>1 FOF*)	Aux Unit	Aux W_i	Alt. assignments of FOF W_i					
					1	2	3	4	5	6
A	3	3	C	1	1	0	1	0	0	1
B	2	2	D	0	1	0	0	1	1	0
	*Holding fixed own treatment and friends' treat		F	1	0	1	0	1	0	1
			G	0	0	1	1	0	1	0
			Probabilities		1/6	1/6	1/6	1/6	1/6	1/6
			Test statistic: Edge Level Contrast for FOF links between Focal and Auxiliary units	1/3	8/3-7/3 =1/3	7/3-8/3 =-1/3	5/2-5/2 =0	5/2-5/2 =0	7/3-8/3 =-1/3	8/3-7/3 =1/3

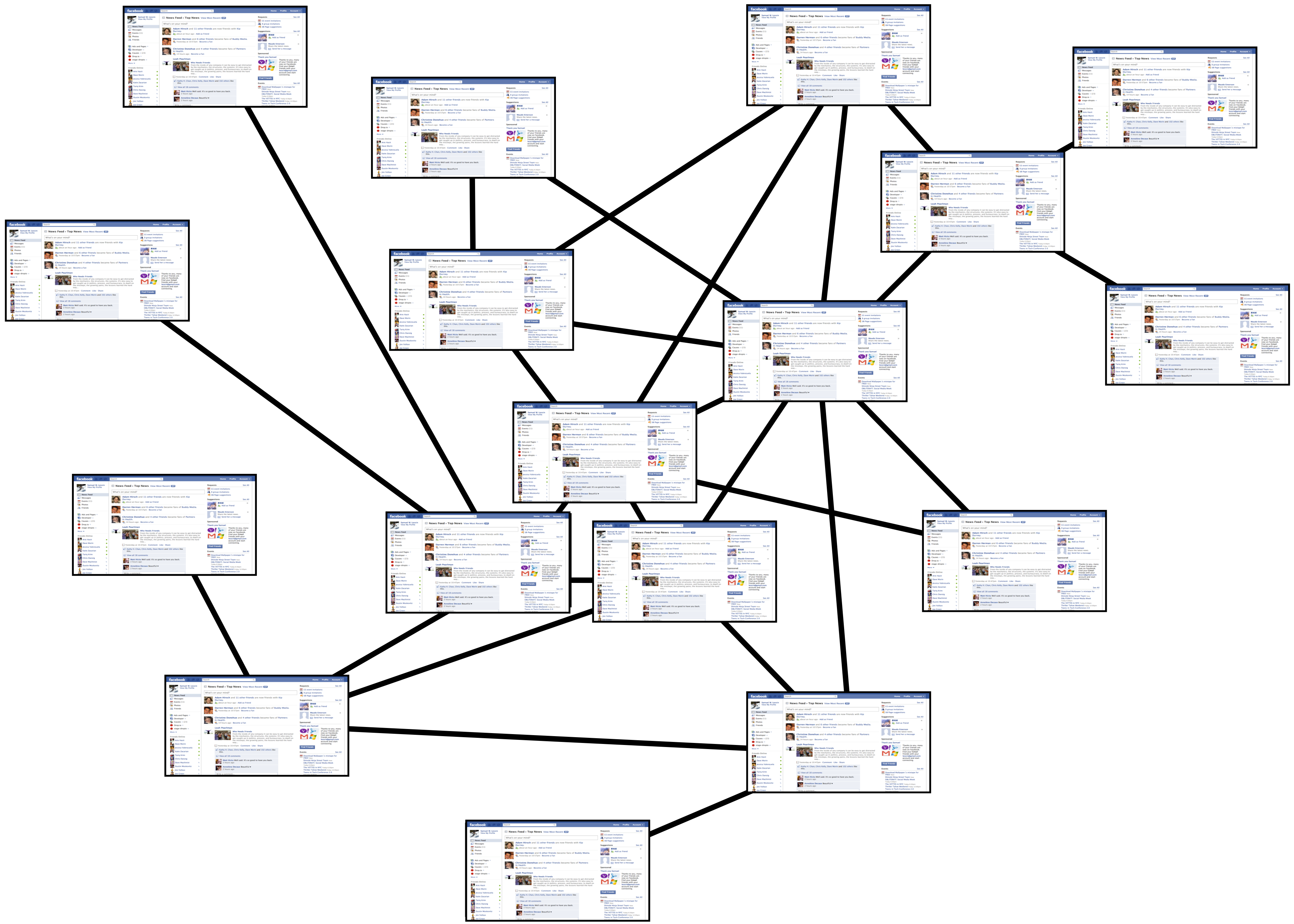
Discussion

Generality: Works for many null hypotheses

- No second-order spillovers
- With two measured networks for same units ,test for spillovers on one network, allowing for (e.g., first order) spillovers according to other network

Open directions

- Finding optimal set of focal units
- Computationally preferable approximations
- Asymptotic inference in networks



Goals

Test for or estimate peer effects

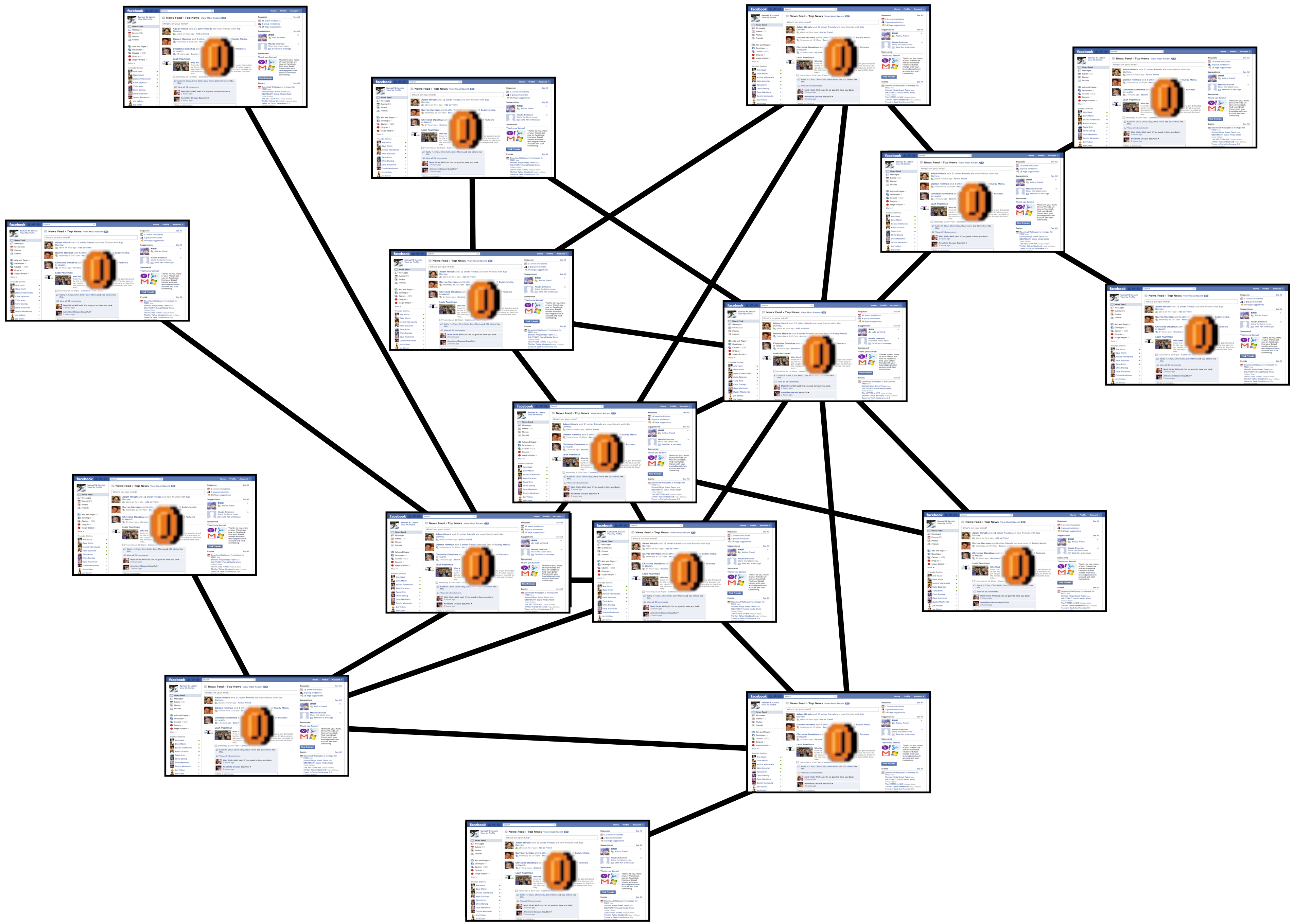
- How does a marginal peer adopting affect your adoption?
- Ideal experiment: *directly assign behaviors of existing peers*

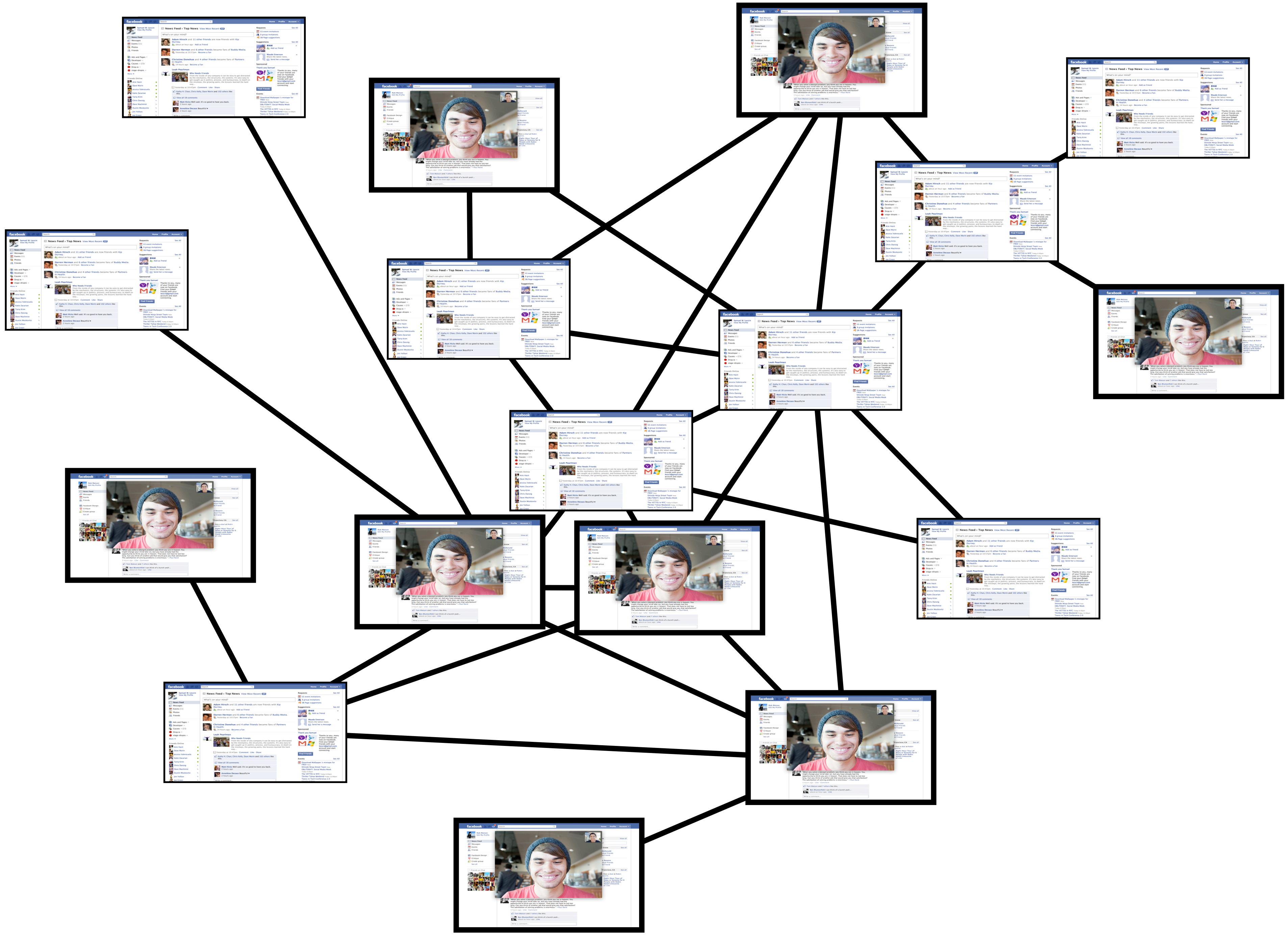
Test for or estimate spillovers

- How does treating a peer affect your outcome?

Estimate effects of global treatment

- What would happen if we gave everyone the treatment?





Goals

Test for or estimate peer effects

- How does a marginal peer adopting affect your adoption?
- Ideal experiment: *directly assign behaviors of existing peers*

Test for or estimate spillovers

- How does treating a peer affect your outcome?

Estimate effects of global treatment

- What would happen if we gave everyone the treatment?
- Ideal experiment: *assign connected components to treatments*

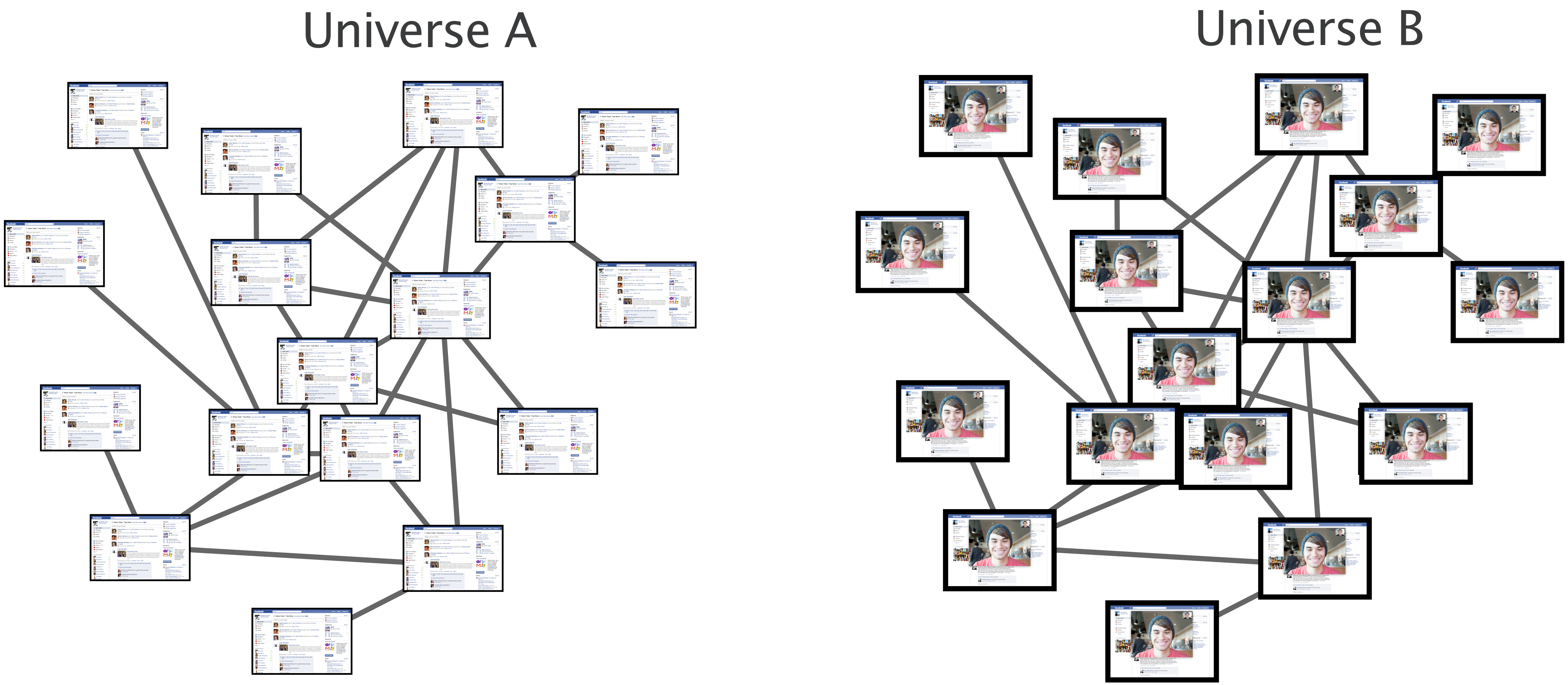
Related work on interference

Multiple non-interacting groups

- Most of literature, e.g., Sobel (2006), Hudgens & Halloran (2008), Tchetgen Tchetgen & VanderWeele (2012)

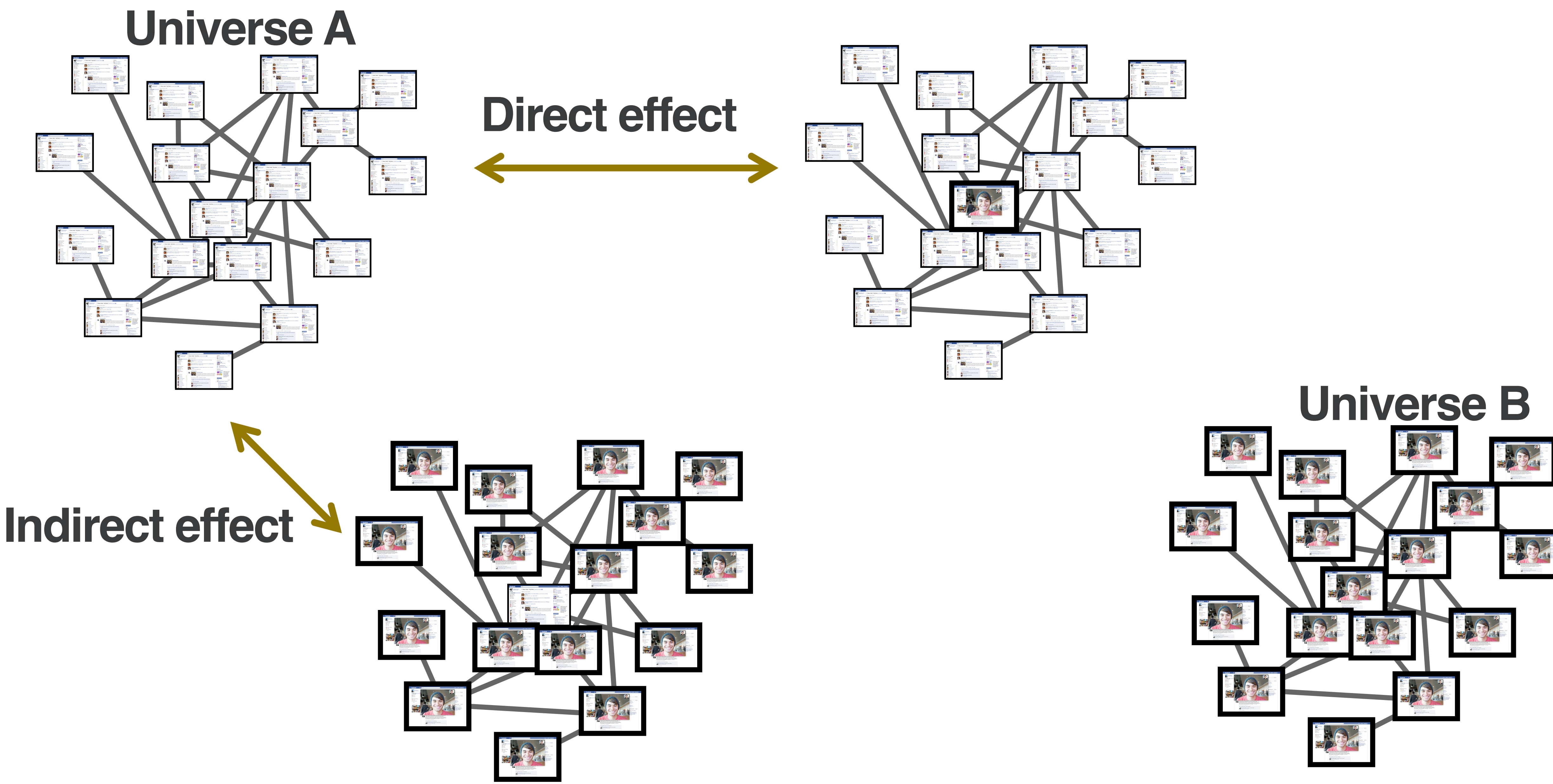
In a single network

- Assume some model of local interference
  i.e. my outcome only depends on my neighbors treatments
  Aronow & Samii (2012), Basse & Airoldi (2015), Manski (2014), Toulis & Kao (2013), Ugander et al. (2013)
- Alternatives: this paper, Choi (2014), van der Laan (2014)



Fundamental problem of causal inference (but worse):
Can only observe potential outcomes for a single global treatment assignment vector

J Ugander, B Karrer, L Backstrom, J Kleinberg (2013) "Graph Cluster Randomization: Network Exposure to Multiple Universes," KDD.
D Eckles, B Karrer, J Ugander (2014) "Design and analysis of experiments in networks: Reducing bias from interference," arXiv.
S Athey, D Eckles, G Imbens (2015) "Exact P-values for Network Interference," arXiv.

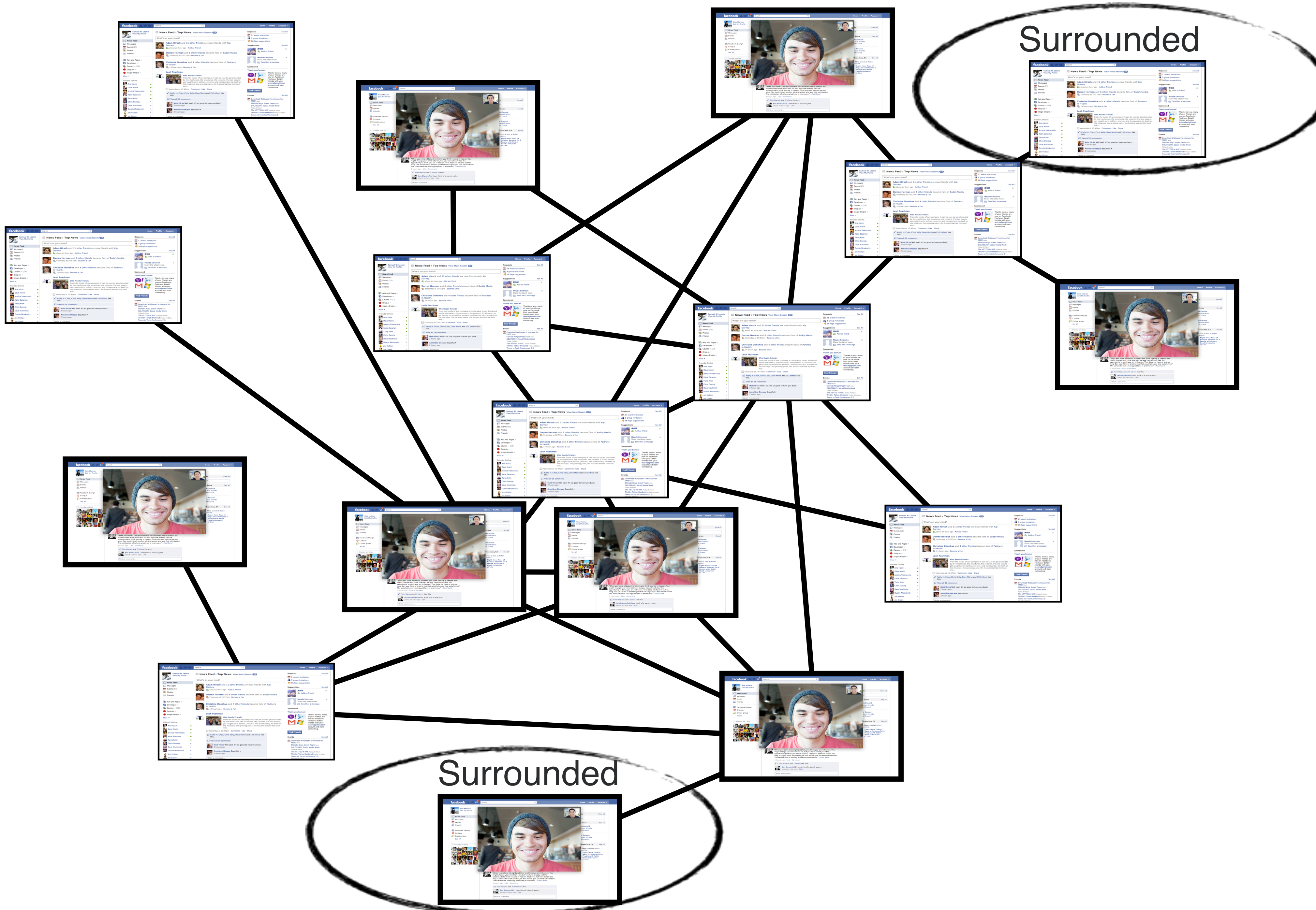


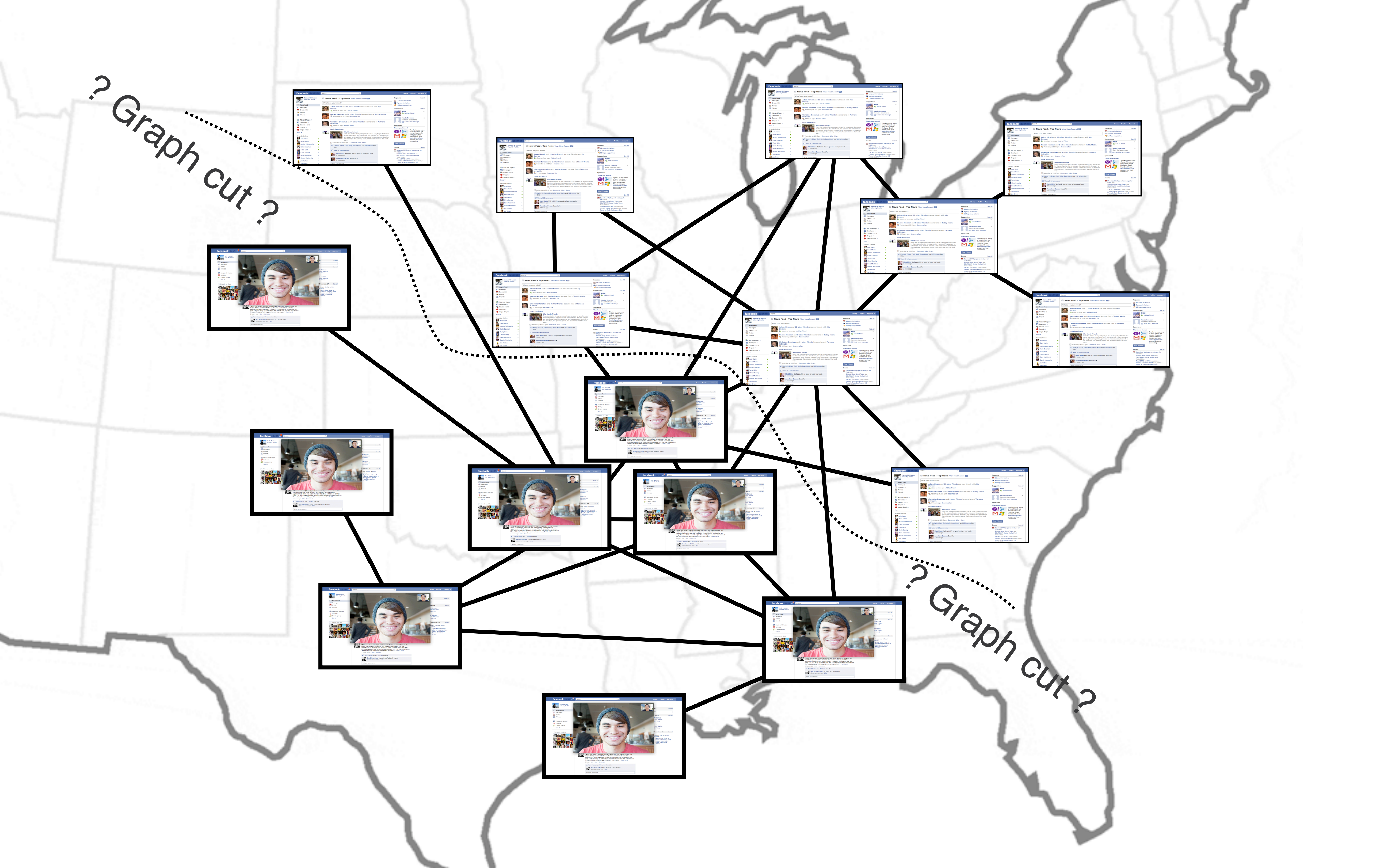
- P Aronow, C Samii (2013) "Estimating average causal effects under interference between units," arXiv.
- C Manski (2013) "Identification of treatment response with social interactions," The Econometrics Journal.

Goal 2: Estimate global effects

- Ideal experiment: *assign connected components to treatments*

- Alternatives
  - Ignore interference: independent assignment, standard analysis
  - Use known, large clusters: assign countries to treatments
  - Model peer effects using existing data; simulate intervention's effects
  - Use neighborhood-based definitions of effective treatments
  - *Assignment with network autocorrelation*


Surrounded
Surrounded


? Graph cut ?
? Graph cut ?

Effect of global treatment

- Have a new intervention, what would be the effects of assigning everyone to this treatment?
- Average treatment effect (ATE) of global treatment z_1 vs. z_0

$$\tau(z_1, z_0) = \frac{1}{N} \sum_i \mathrm{E}[Y_i(Z = z_1) - Y_i(Z = z_0)]$$

- We can't observe the whole network both ways
  - "Fundamental problem of causal inference" (Holland, 1988)
- Don't want to assume no interference, SUTVA, or something similar

Observed outcomes

In terms of global treatment assignment

- Observed outcomes $Y = f(Z, U)$ are a function from global treatment assignment and stochastic component

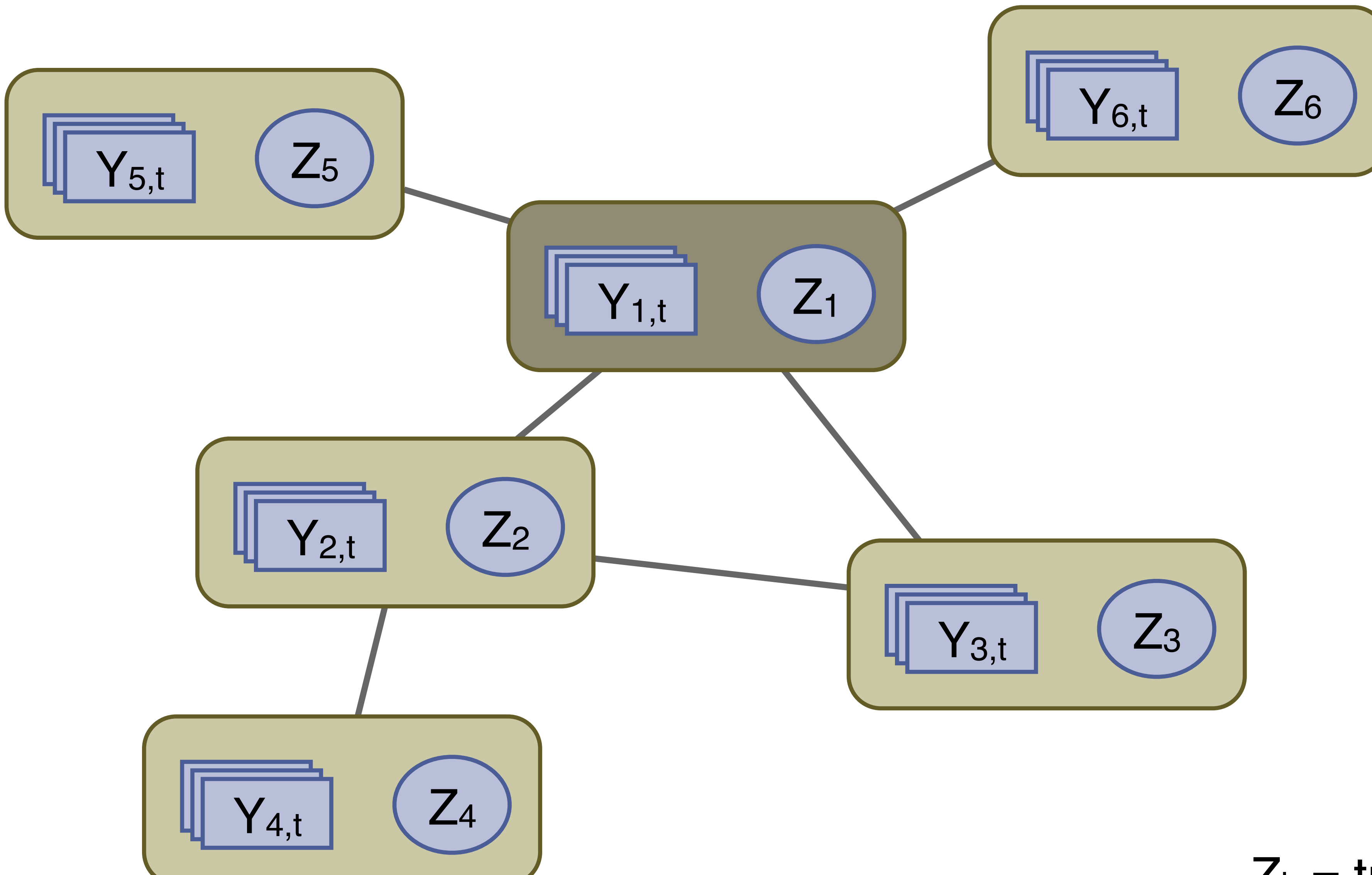
$$f_i(\cdot) : \mathbb{Z}^N \times \mathbb{U}^N \to \mathbb{Y}$$

- We can place restrictions on this function by specifying an *exposure model* (Aronow & Samii, 2012) or *effective treatments* (Manski, 2013)

$$c_i(z_1) = c_i(z_0) \Rightarrow f_i(z_1, u) = f_i(z_0, u)$$

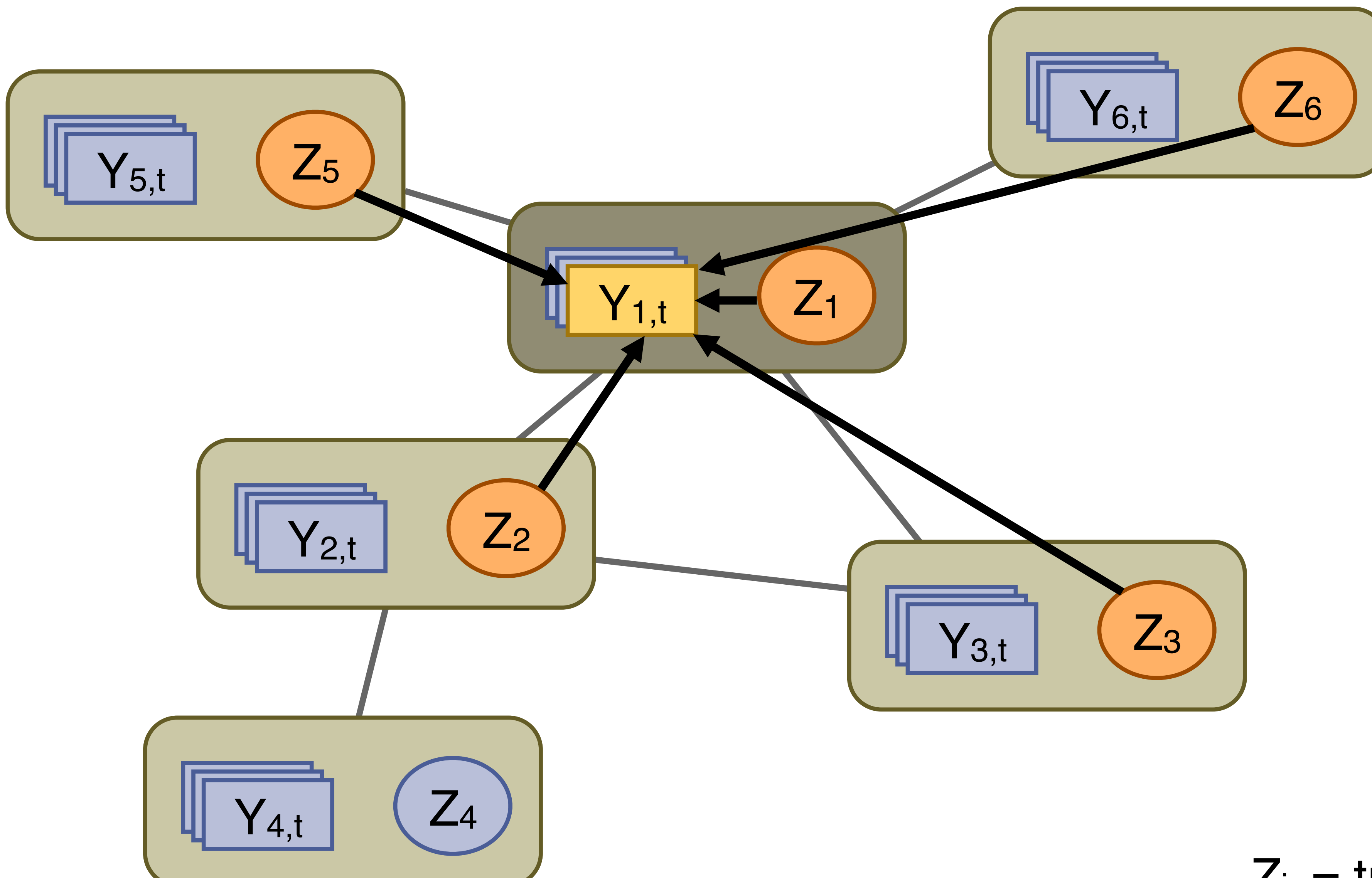
  - i.e., define levels sets, assume *local interference*

How does interference arise?



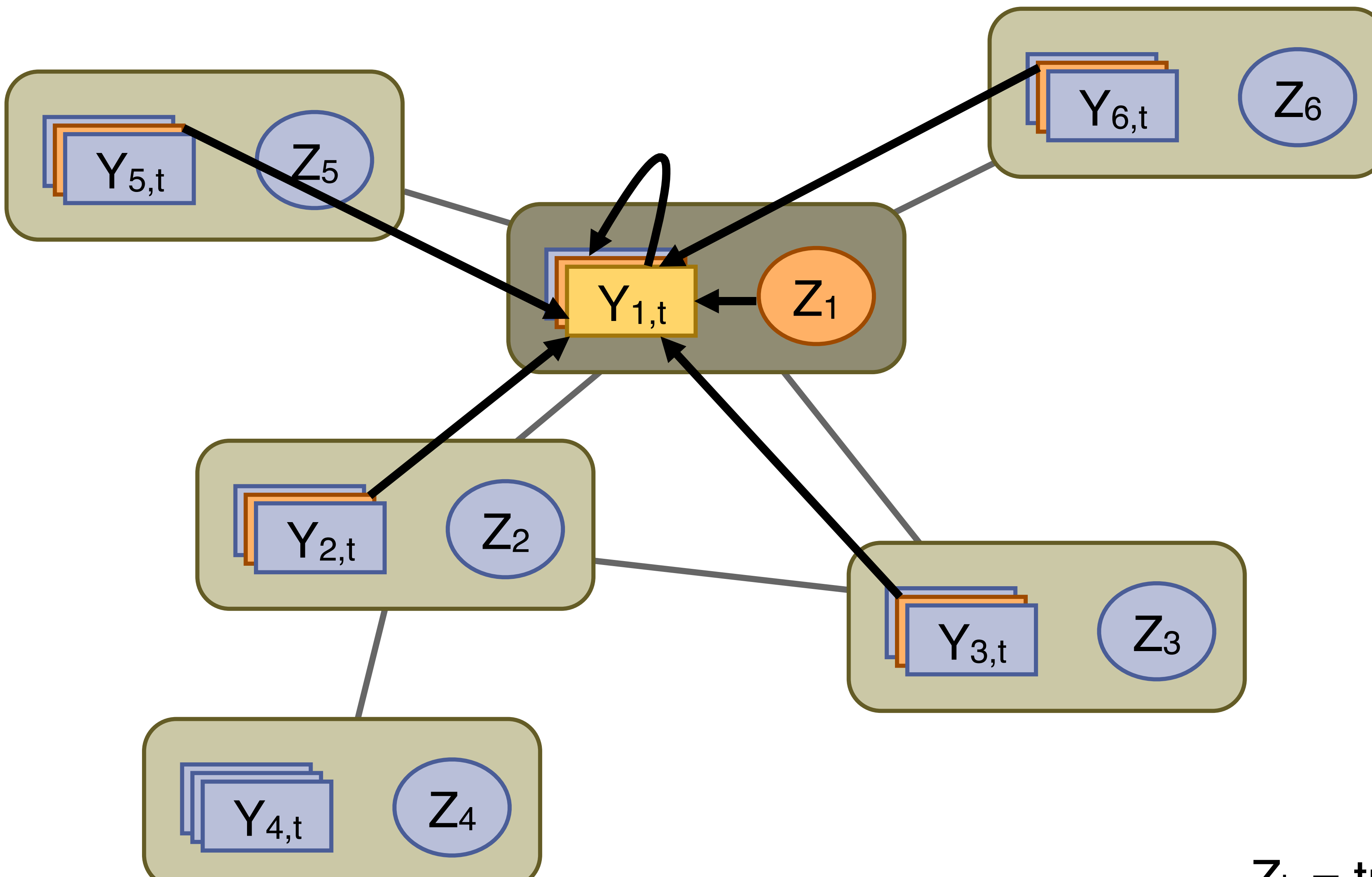
Z_i = treatment
$Y_{i,t}$ = response at time t

Dependence on neighbors' *treatments*



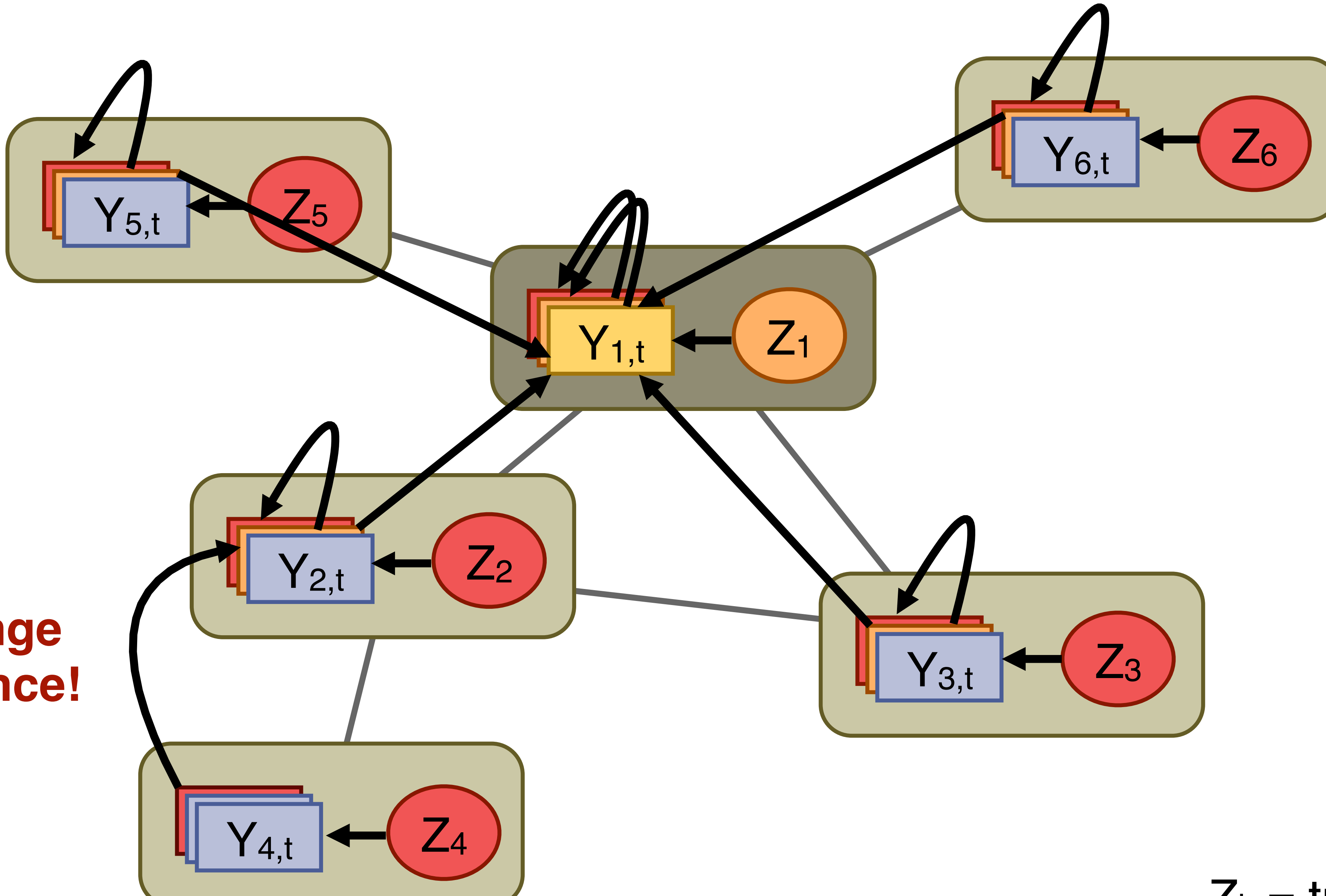
Z_i = treatment
$Y_{i,t}$ = response at time t

Dependence on neighbors' *behaviors*



Z_i = treatment
$Y_{i,t}$ = response at time t

Dependence on neighbors' *behaviors*



Long-range Dependence!

Z_i = treatment
$Y_{i,t}$ = response at time t

Observed outcomes

Implausibility of restrictions on *f*

- We expect interference because of peer effects, but peer effects make only local interference implausible
  - Except at a single discrete time step after treatment (i.e., $t = 2$)

- Can we motivate exposure model through model of peer effects?
  - Or at least evaluate exposure models using more realistic or "primary" models (cf. Manski, 2013)

Outcome generating process

- Nonparametric structural equation model for observed outcomes, where outcomes are a function of vertex *i*'s k_i neighbors' prior behavior: $\mathbb{Z} \times \mathbb{Y}^{k_i} \times \mathbb{U}^N \to \mathbb{Y}$

Example: Noisy best response model

- Latent utility is linear-in-means (probit model)

$$Y^*_{i,t} = \alpha + \beta Z_i + \gamma \frac{A_i' Y_{i,t-1}}{k_i} + U_{i,t}$$

$$Y_{i,t} = g\left(Y^*_{i,t}\right)$$

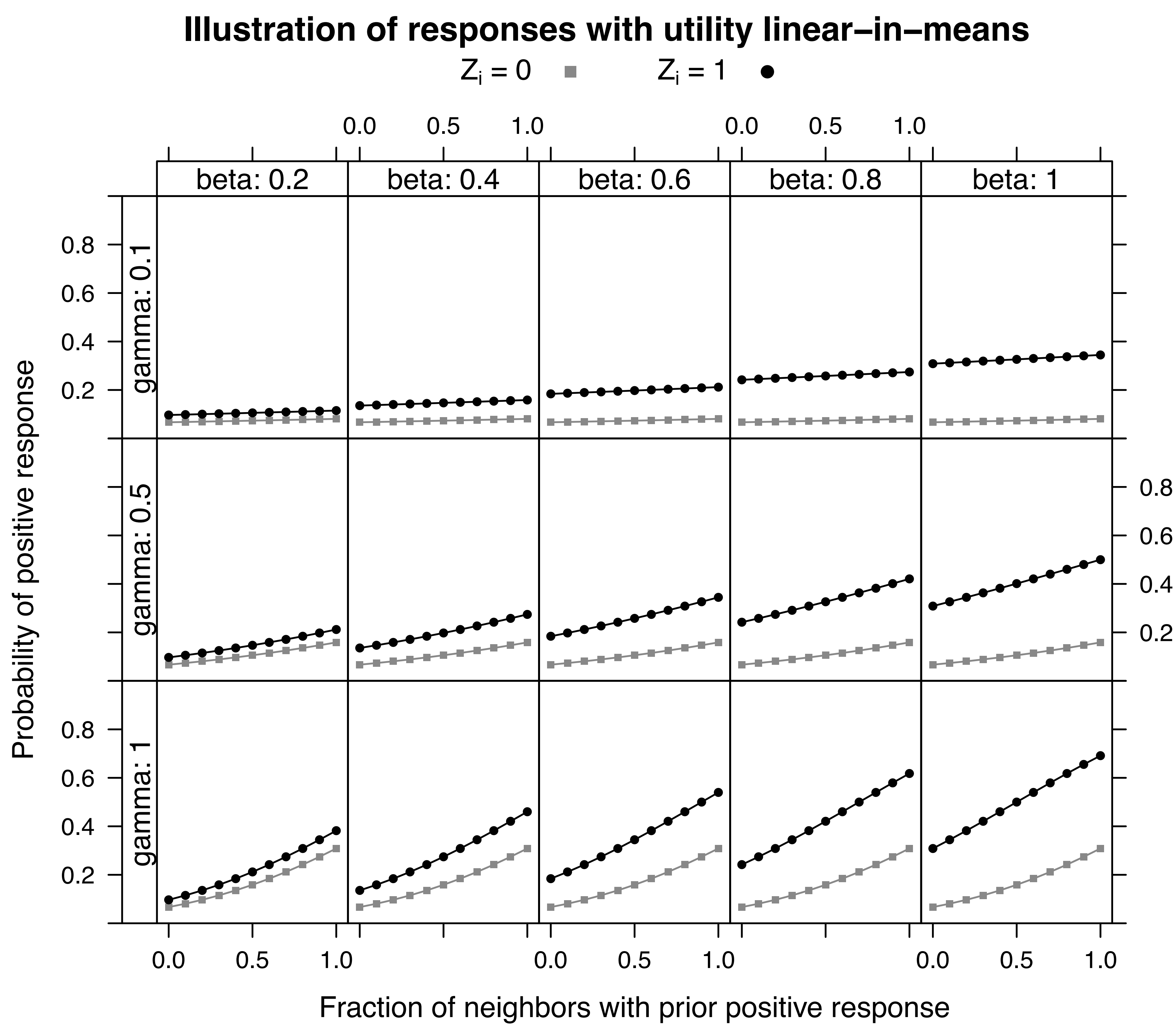
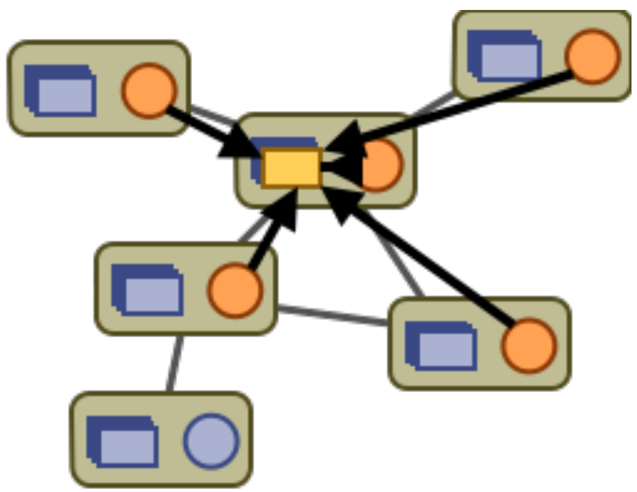


Illustration of responses with utility linear−in−means
Zi = 0
Zi = 1
beta: 0.2
beta: 0.4
beta: 0.6
beta: 0.8
beta: 1
gamma: 0.1
gamma: 0.5
gamma: 1
Probability of positive response
Fraction of neighbors with prior positive response
0.0
0.5
1.0
0.2
0.4
0.6
0.8

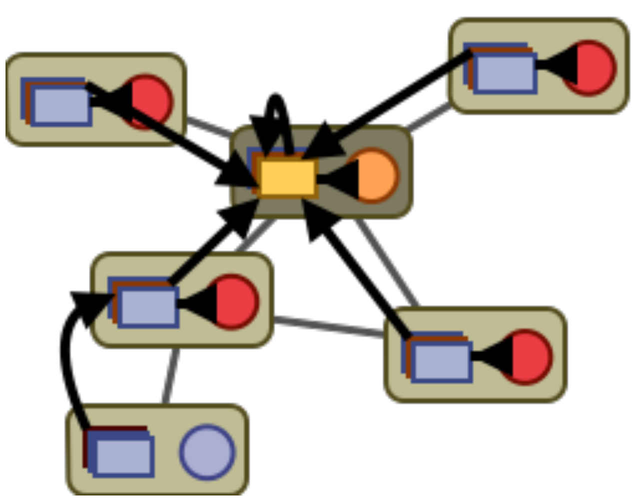
Ego behavior caused by…



Neighbors' treatments

- Local interference: No long range dependence
- Unbiased estimators available

(Aronow & Samii, Ugander et al.)

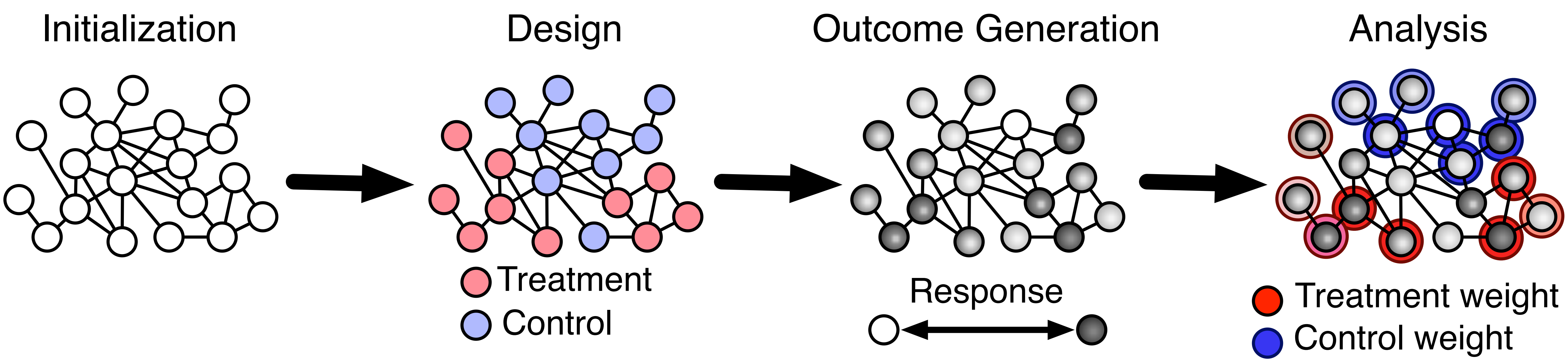


Neighbors' behaviors

- Global interference: long range dependence
- Bias difficult to eliminate, but can reduce it
- Much more realistic

Model of experiments in networks

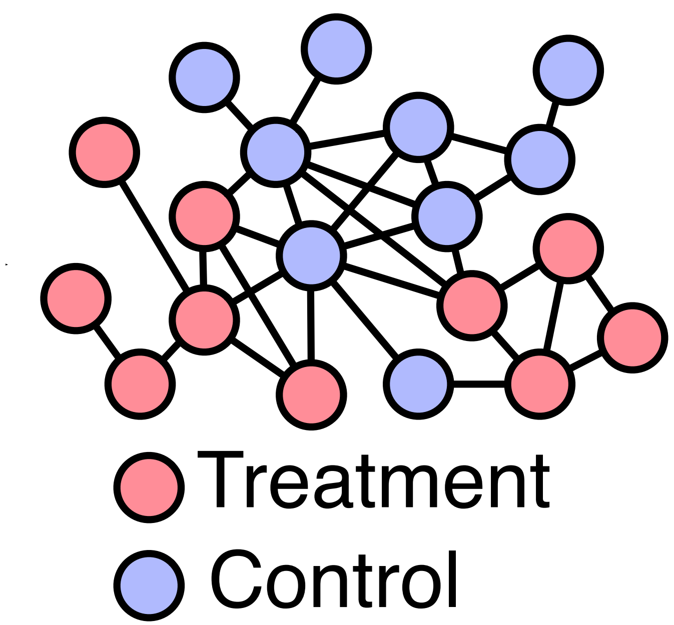
- How are units assigned to conditions?
- What is the true outcome generating process?
- What estimators are used?



Design

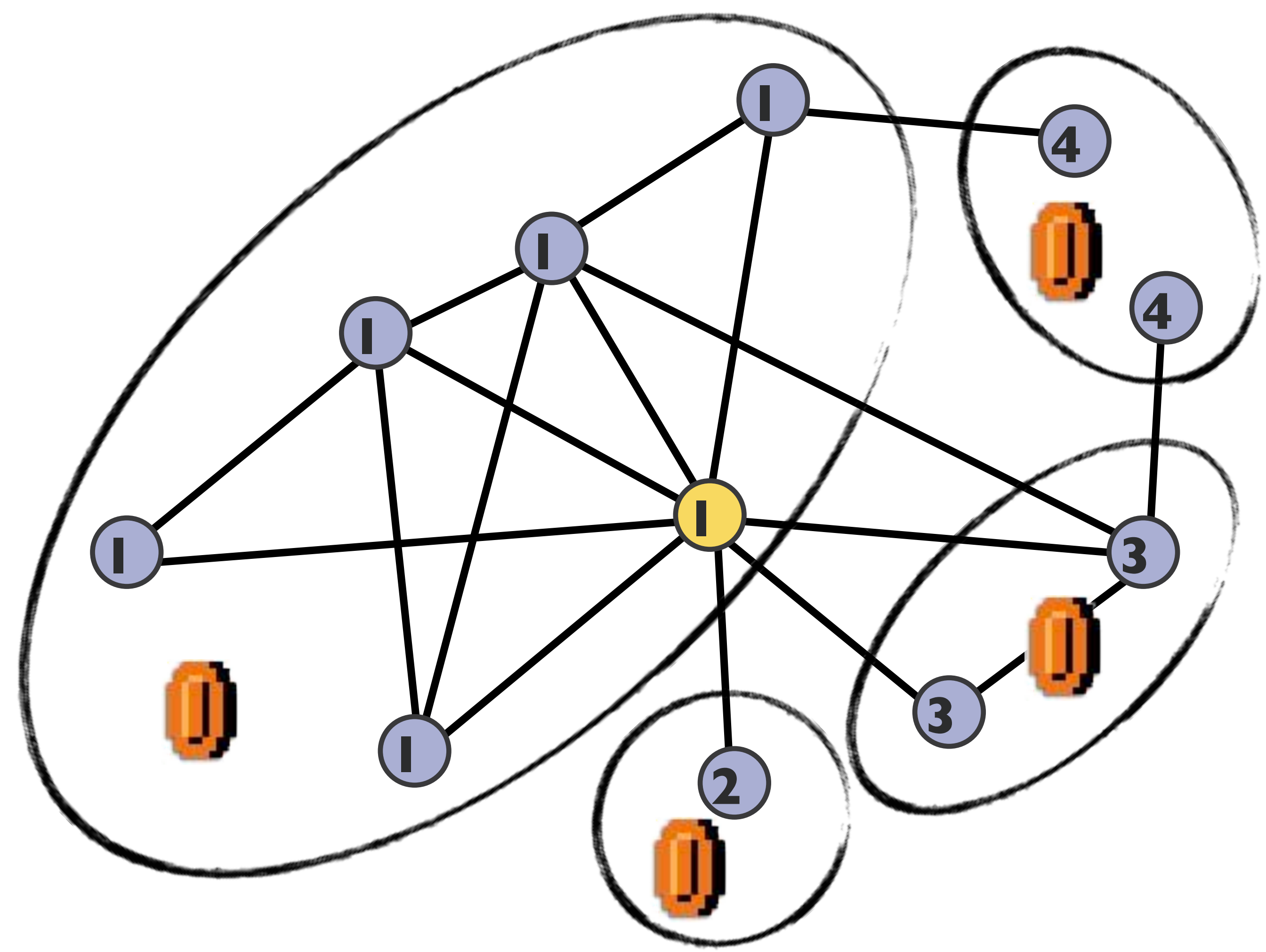
How to assign vertices to treatments?

- Independent random assignment
- Assignment with network autocorrelation
  - Many ways to do this
  - Many of which end up being producing uniform correlation (e.g. correlation = 1) between assignments of a set of vertices



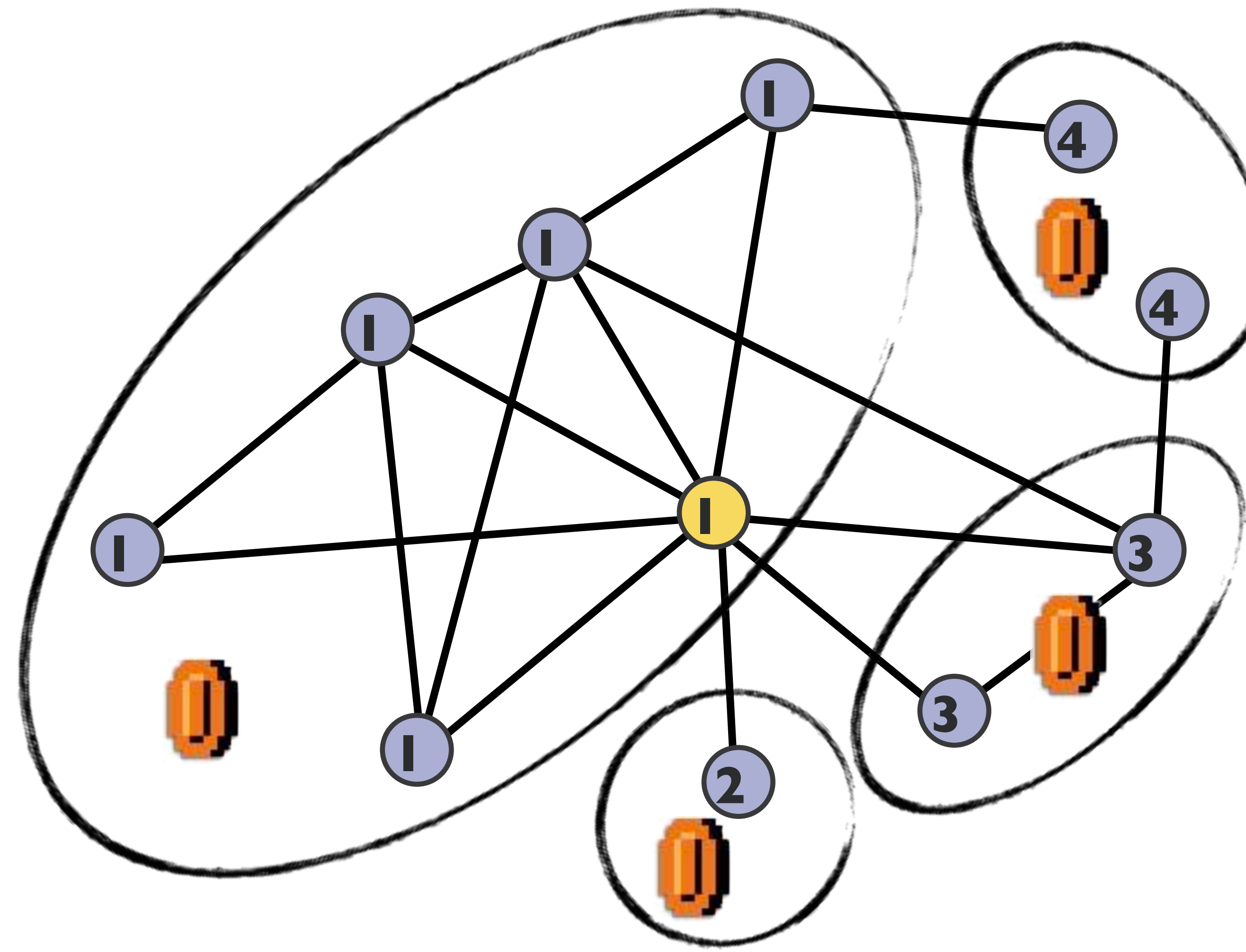
Graph cluster randomization

1. Partition graph into clusters
2. Assign each cluster to treatment with probability q
3. Assign all vertices to their cluster's treatment



Graph cluster randomization (hole punching)

1. Partition graph into clusters, so vertex *i* is in cluster *C(i)*
2. For each cluster *j*
   $V_j \sim$ Bernoulli(q)
3. For each node i
   $Z_j \sim$ Bernoulli($q_{C(i)}$)
   with $q_j = 1 - \eta$ *if* $V_j = 1$,
   $q_j = \eta$ *otherwise*




The Plumb-pudding in danger; — or — State Epicures taking un Petit Souper
"the great Globe itself, and all which it inherit", is too small to satisfy such insatiable appetites
OCEAN
EUROPE

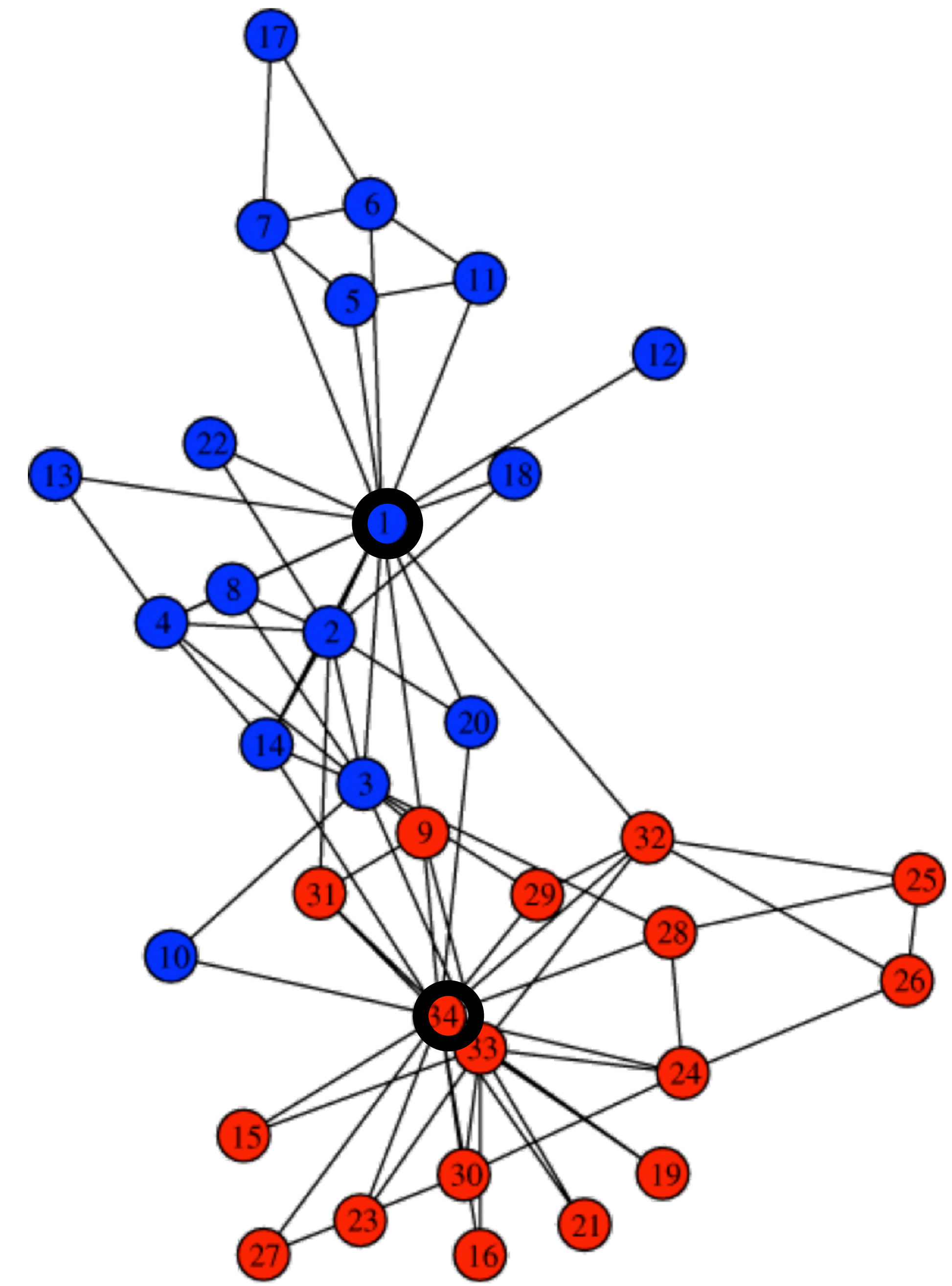

facebook
December 2010

Graph partitioning

- Facebook: 1B vertices, 100B+ edges (countries100M+ vertices)
- Some candidate, highly scalable methods:
  - Community detection, e.g., Louvain method (Blondel et al., 2008)
  - Label propagation (Zhu & Ghahramani, 2002, Raghavan et al., 2007, Ugander & Backstrom, 2013)
  - **ε-net clustering**
    - Greedy version: Pick a vertex. Put vertex and all vertices within distance ε - 1 into a cluster. Repeat.
    - Ugander et al. (2013) bound variance of ATE estimates when clustering with this method with ε = 3.

Zachary Karate Club

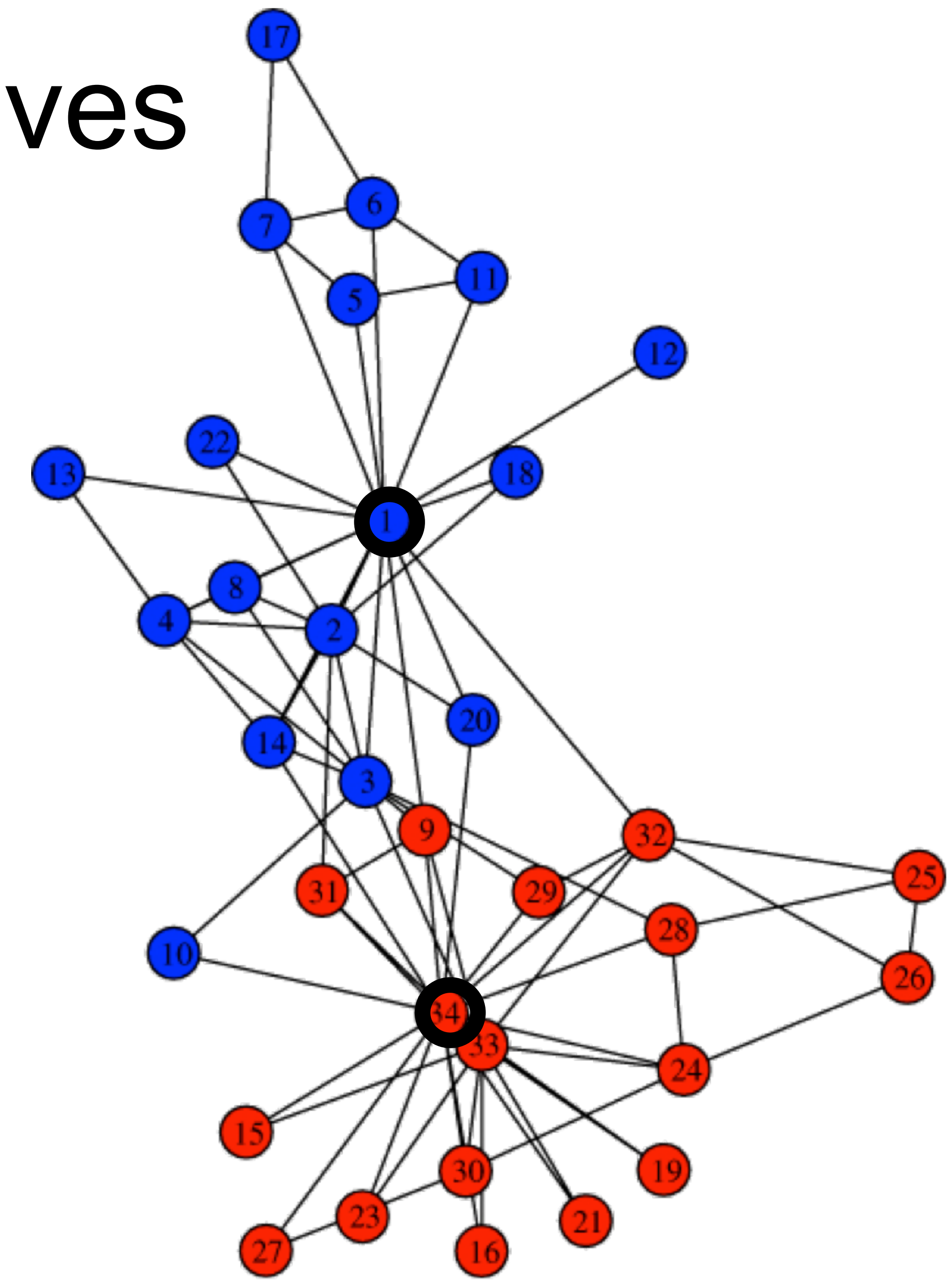
- Wayne Zachary, sociologist interested in group dynamics.
- Studied a karate club for 3 years ('70-'72)
- Club formed factions around instructor (1) and Club President (34).
- Zachary was interested in if faction structure could be predicted.
- Zachary (1977) applied Ford–Fulkerson, found group split was predicted by min-cut.



Zachary, W. W. (1977). An Information Flow Model for Conflict and Fission in Small Groups. *Journal of Anthropological Research, 33*(4), 452–473.

Community detection objectives

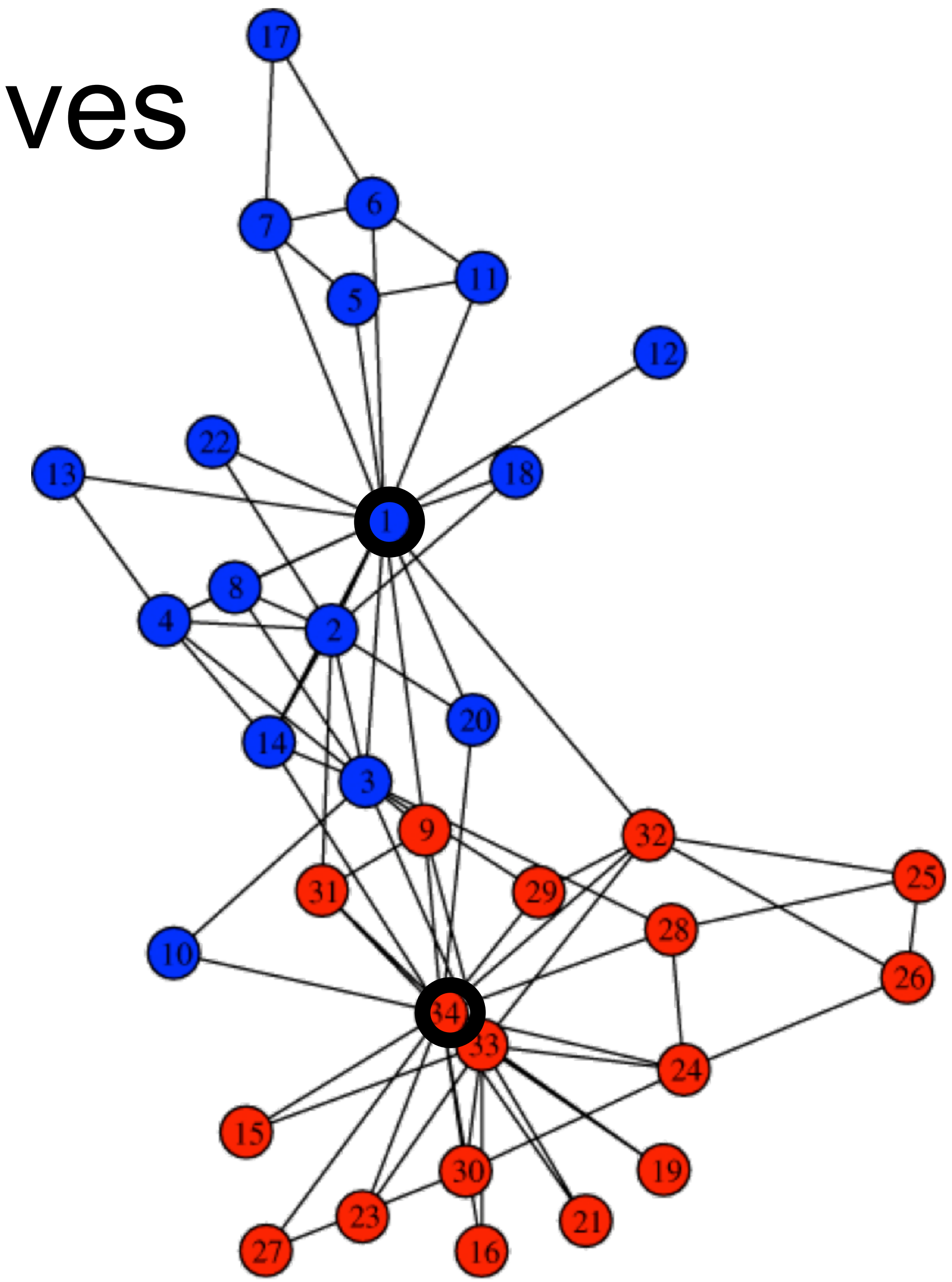
- In Zachary (1977): Predict how a group fissions when led by two rival leaders
- Similar objective function used for cases of labeled community data



- MEJ Newman, M Girvan (2004) "Finding and evaluating community structure in networks," Physical Rev E.
- S Fortunato, M Barthelemy (2007) "Resolution limit in community detection," PNAS.
- J Shi, J Malik (2000) "Normalized cuts and image segmentation," IEEE Trans Pattern Analysis and Machine Intelligence.
- E Mossel, J Neeman, A Sly (2012) "Stochastic block models and reconstruction"

Community detection objectives

- Modularity maximization:
  - Has “resolution limit”
- Conductance (normalized min-cut):
  - Produces balanced partitions; spectral guarantees
- Ability to recover Stochastic Block Model:
  - Stylized model in absence of ground truth data
- In network experiments: Minimize MSE of ATE estimate?



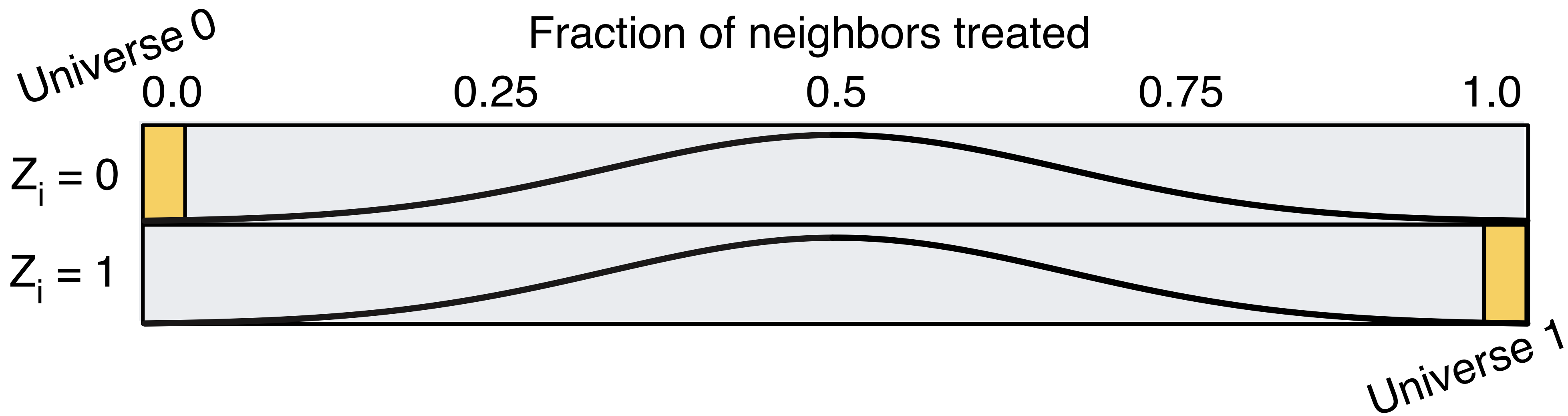
- MEJ Newman, M Girvan (2004) "Finding and evaluating community structure in networks,” Physical Rev E.
- S Fortunato, M Barthelemy (2007) "Resolution limit in community detection," PNAS.
- J Shi, J Malik (2000) “Normalized cuts and image segmentation,” IEEE Trans Pattern Analysis and Machine Intelligence.
- E Mossel, J Neeman, A Sly (2012) “Stochastic block models and reconstruction”

Graph partitioning

- Facebook: 1B vertices, 100B+ edges (countries100M+ vertices)
- Some candidate, highly scalable methods:
  - Community detection, e.g., Louvain method (Blondel et al., 2008)
  - Label propagation (Zhu & Ghahramani, 2002, Raghavan et al., 2007, Ugander & Backstrom, 2013)
  - **ε-net clustering**
    - Greedy version: Pick a vertex. Put vertex and all vertices within distance ε - 1 into a cluster. Repeat.
    - Ugander et al. (2013) bound variance of ATE estimates when clustering with this method with ε = 3.

Fraction of neighbors treated

Independent and (near ideal) correlated assignment



Bias reduction from design

Graph cluster randomization reduces bias

- Assume quite general model where outcomes are linear in the *global* treatment assignment vector:

$$\mathrm{E}_U[Y_i(z,U)] = a_i + \sum_{j \in V} B_{ij} z_j$$

  - This model has $N^2 + N$ parameters: 2 for each pair of vertices

  - Special case: Outcomes linear in prior peer behaviors for any t ≥ 0

$$\mathrm{E}^U[Y_{i,t}(z)] = \alpha + \beta z_i + \gamma \frac{A_i' \mathrm{E}^U[Y_{t-1}(z)]}{k_i}.$$

Bias reduction from design

Graph cluster randomization reduces bias

- Compare the simple difference-in-mean estimands under these two designs

- Under independent random assignment:

$$\tau_{\mathrm{ITR}}^{\mathrm{ind}}(1,0) = \frac{1}{N}\sum_{i=1}^{N} B_{ii}.$$

all coefficients cancel, except your own

- Under graph cluster randomization:

$$\tau_{\mathrm{ITR}}^{\mathrm{gcr}}(1,0) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} B_{ij}\mathbf{1}[C(i) = C(j)].$$

all coefficients cancel, except those in your cluster

Bias reduction from design

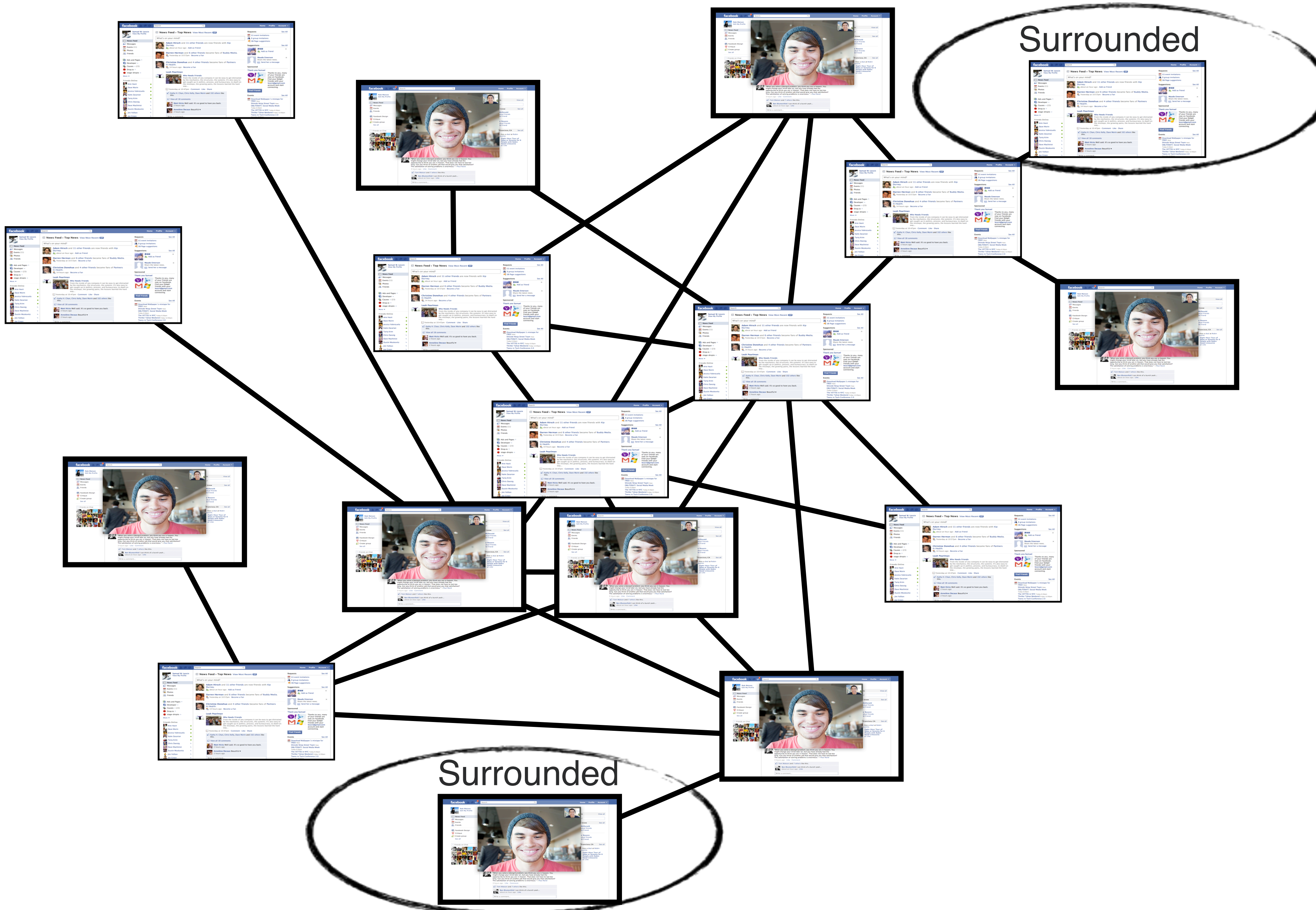
Graph cluster randomization reduces bias

Theorem 2.1. *Assume we have a linear outcome model for all vertices $i \in V$ such that*

$$\mathrm{E}_U[Y_i(z, U)] = a_i + \sum_{j \in V} B_{ij} z_j \tag{5}$$

and further assume that $Y_i(z, u)$ is monotonically increasing in z for every $u \in \mathbb{U}^N$ and vertex i such that $B_{ij} \geq 0$.

Then for any mapping of vertices to clusters, the absolute bias of $\tau^d_{\mathrm{ITR}}(1, 0)$ when the design d is graph cluster randomization is less than or equal to the absolute bias when d is independent assignment, with a fixed treatment probability p.



Surrounded
Surrounded

Bias reduction through analysis

Comparing only "surrounded" vertices

- Some units are more likely to end up surrounded: probability of "effective treatment" is not homogenous
  - e.g. with independent random assignment, high degree vertices have low probability of having all peers treated
  - If not accounted for, this can be a source of bias
- Can compute exact design-based propensity scores for each unit

$$\pi_i(z) = \Pr(g_i(Z) = g_i(z))$$

where $g_i(\cdot)$ is a function specifying effective treatments.

- For graph cluster randomization, this is a dynamic program (Ugander et al. 2013)

Bias reduction from analysis

Contrasting more surrounded vertices reduces bias

Consider functions $g_i(\cdot)$ such that $g_i(Z) = g_i(z)$ just implies that for some subset of vertices J_i we have that $\sum_{j \in J_i} 1\{Z_j = z_j\} \geq l_i$ and that $Z_i = z_i$.

Fractional neighborhood treatment response (FNTR) assumption: $J_i = \delta(i)$ and $l_i = \lceil \lambda k_i \rceil$, where k_i is vertex i's degree.

ITR and NTR are special cases with $\lambda = 0$ and $\lambda = 1$.

If we have two such functions $g_i^A(\cdot)$ and $g_i^B(\cdot)$ with the same J_i, and $g_i^A(z) = g_i^A(z')$ implies $g_i^B(z) = g_i^B(z')$, then we say that $g_i^A(\cdot)$ is *more restrictive* than $g_i^B(\cdot)$.

Bias reduction from analysis

Contrasting more surrounded vertices reduces bias

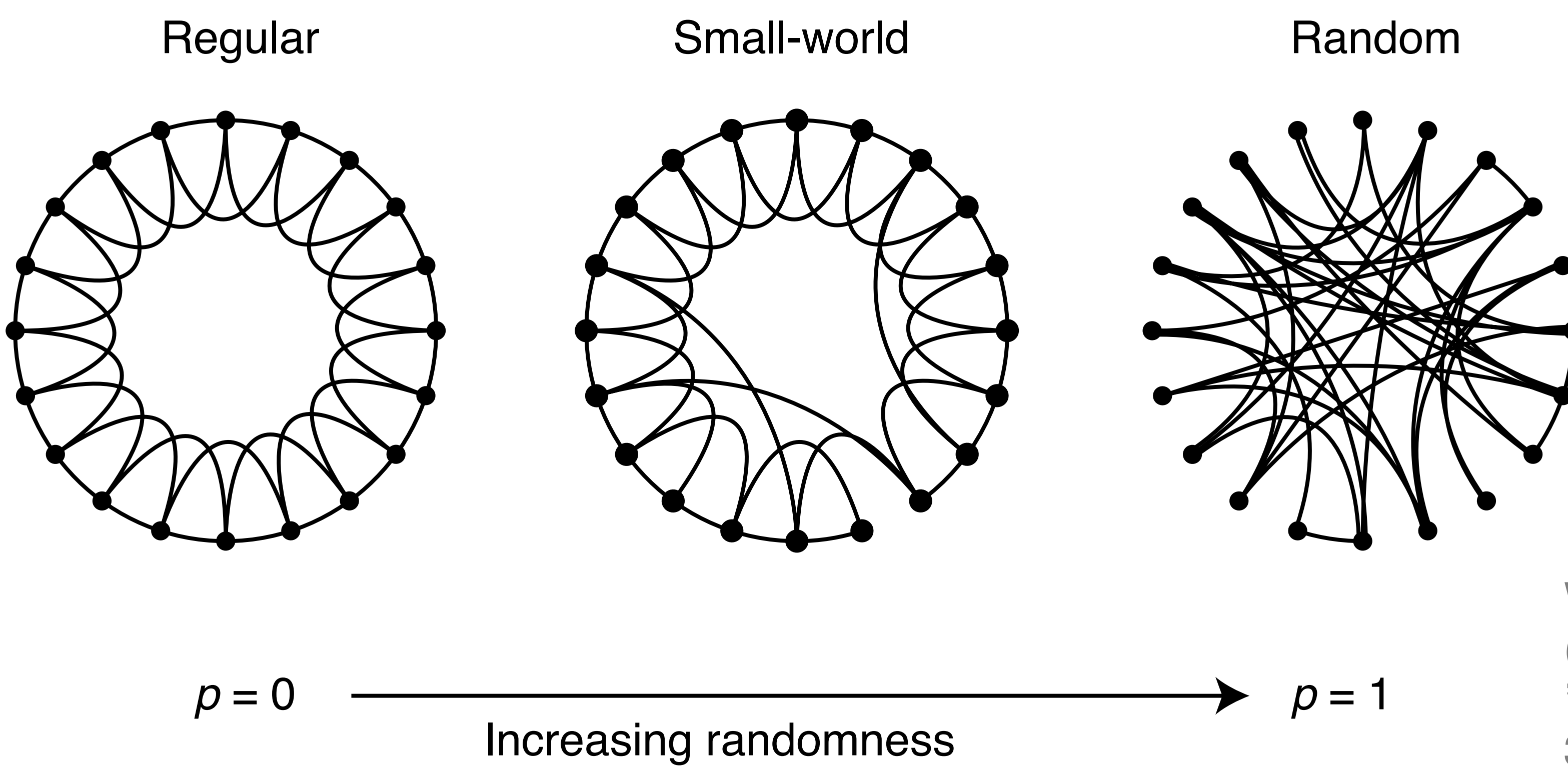
Theorem 2.2. *Let $g^A(\cdot)$ and $g^B(\cdot)$ be vectors of such functions where $g_i^A(\cdot)$ is more restrictive than $g_i^B(\cdot)$ for every vertex i, and let independent random assignment be the experimental design. A sufficient condition for estimand $\tau_{g^A}^{\text{ind}}(1,0)$ to have less than or equal absolute bias than $\tau_{g^B}^{\text{ind}}(1,0)$, where these estimands are defined by Equation 13, is that we have monotonically increasing responses or monotonically decreasing responses for every vertex with respect to z.*

Simulations

- Networks: small world networks
- Treatment assignment: 3-net clustered or independent
- Observed outcomes: utility linear-in-means
  - Vary direct and peer effects
- Estimators:
  - Simple difference-in-means
  - Difference-in-means with effective treatment from fractional neighborhood exposure model $\lambda = 0.75$
  - Additionally can weight using probabilities of assignment in Hajek estimator

Initialization: Networks

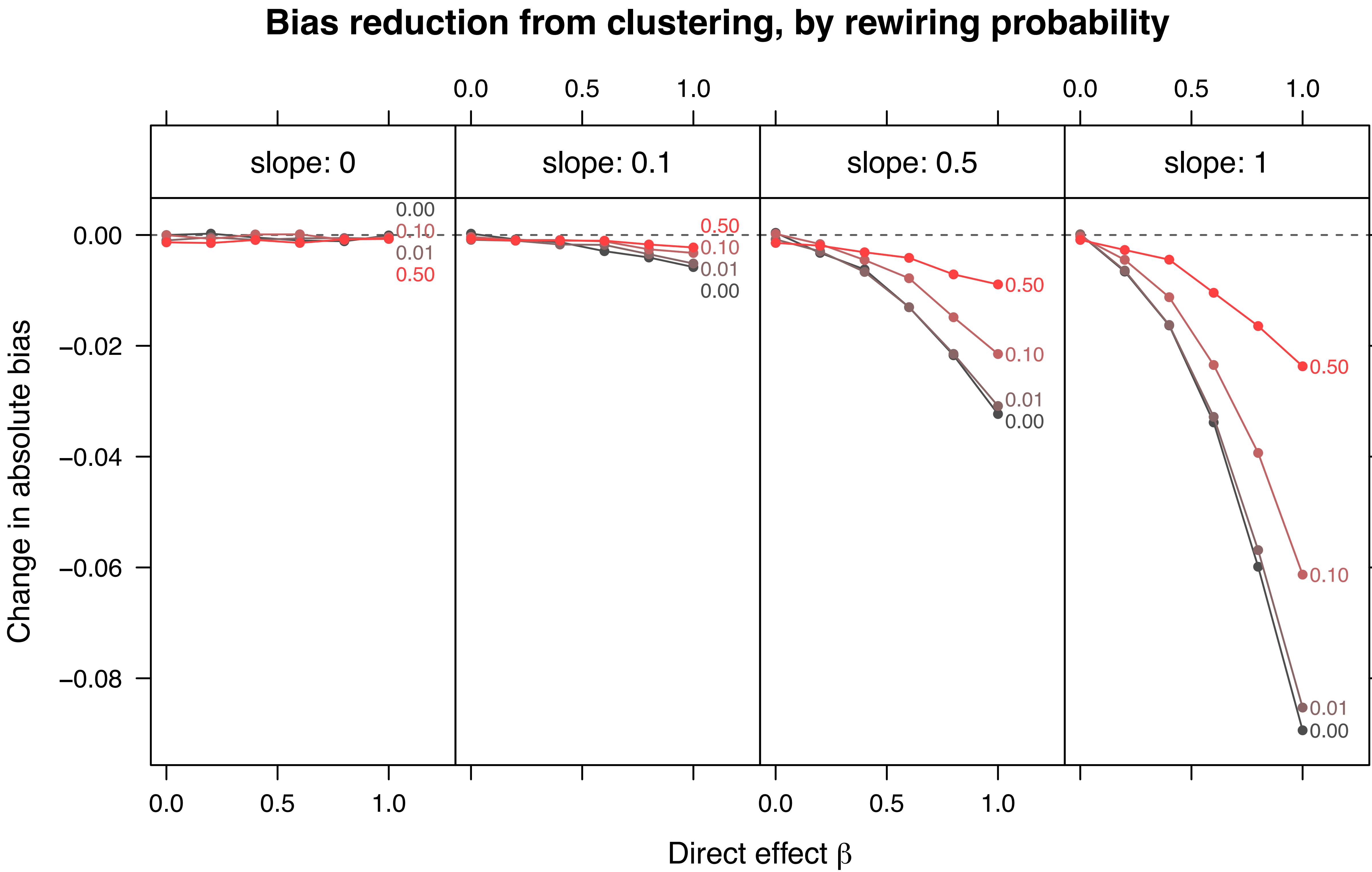
Small-world networks

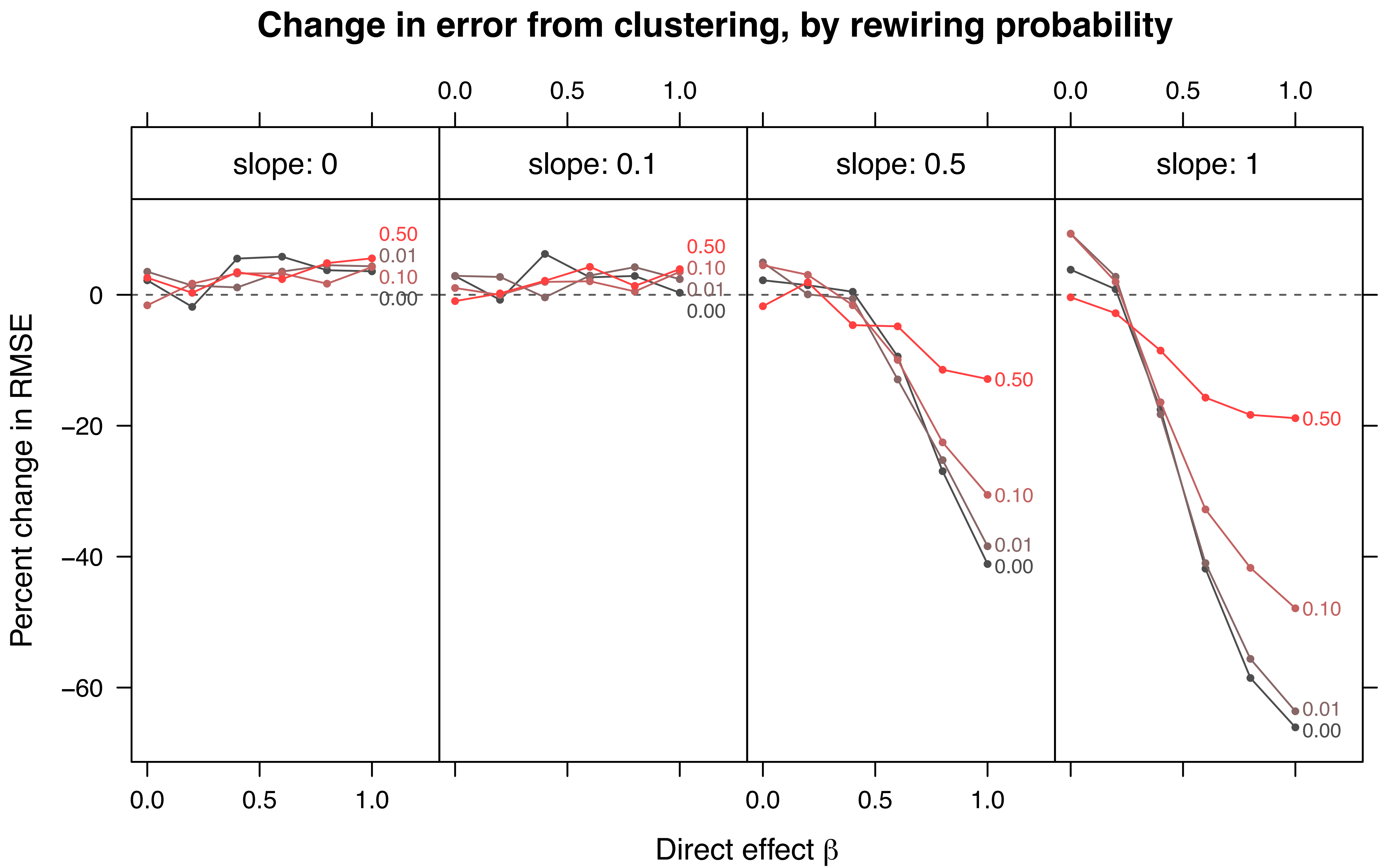


Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–2.

- N = 1000, k = 10. Vary rewiring probability p: 0.00, 0.01, 0.10, 0.50

Not today: Real networks, Degree-corrected block models


Bias reduction from clustering, by rewiring probability
slope: 0
slope: 0.1
slope: 0.5
slope: 1
Change in absolute bias
Direct effect β
0.00
−0.02
−0.04
−0.06
−0.08
0.0
0.5
1.0
0.00
0.10
0.01
0.50


Change in error from clustering, by rewiring probability
slope: 0
slope: 0.1
slope: 0.5
slope: 1
0.50
0.01
0.10
0.00
0.50
0.10
0.01
0.00
0.50
0.10
0.01
0.00
0.50
0.10
0.01
0.00
0
−20
−40
−60
Percent change in RMSE
0.0
0.5
1.0
Direct effect β

Effects of global interventions

Conclusions

- Clustered randomization reduces bias

    … under idiosyncratic global linear interference (Th 2.1)

    … under a range of other DGPs (simulations)

    … without adding “too much” variance (simulations)

- Comparing more “surrounded” individuals reduces bias (Th 2.2)

    … but often adds “too much” variance (simulations)

Open directions

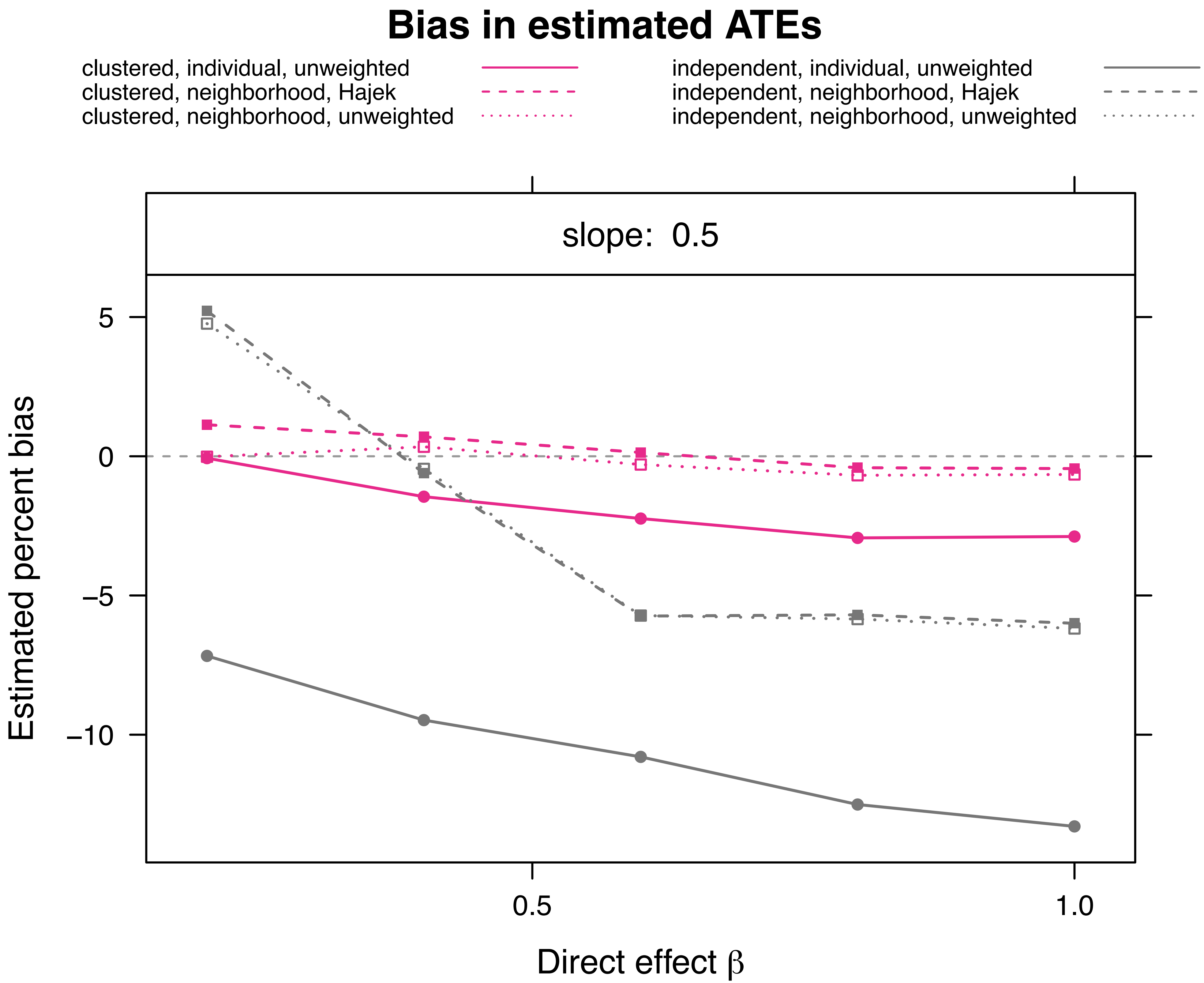
- Other ways to produce network autocorrelation
  - Many of these reduce to graph cluster randomization
  - Those that don’t may not have an easy way to compute propensity scores (probabilities of assignment)
- Find optimal design for a model of the network
  - Find the optimal design for stochastic block model approximation (Airoldi; Basse & Airoldi, 2015)
- Target global ATE via a model of peer effects
  - By combining estimates of direct effects and peer effects

Airoldi, E. Optimal design of experiments in the presence of network interference. “Soon on arXiv”
Basse, G. W., & Airoldi, E. (2015). Optimal design of experiments in the presence of network-correlated outcomes. arXiv.

Analysis

Choice of effective treatments

- Individual treatment response — compare vertices with $Z_i = 1$ to those with $Z_i = 0$
  - Leads to simple difference-in-means estimator
- Neighborhood treatment response
  - Full neighborhood treatment – vertices will all neighbors in same treatment
  - Fractional λ–neighborhood treatment – vertices with at least a fraction λ of neighbors in same treatment


Bias in estimated ATEs
clustered, individual, unweighted
clustered, neighborhood, Hajek
clustered, neighborhood, unweighted
independent, individual, unweighted
independent, neighborhood, Hajek
independent, neighborhood, unweighted
slope: 0.5
Estimated percent bias
5
0
−5
−10
0.5
1.0
Direct effect β

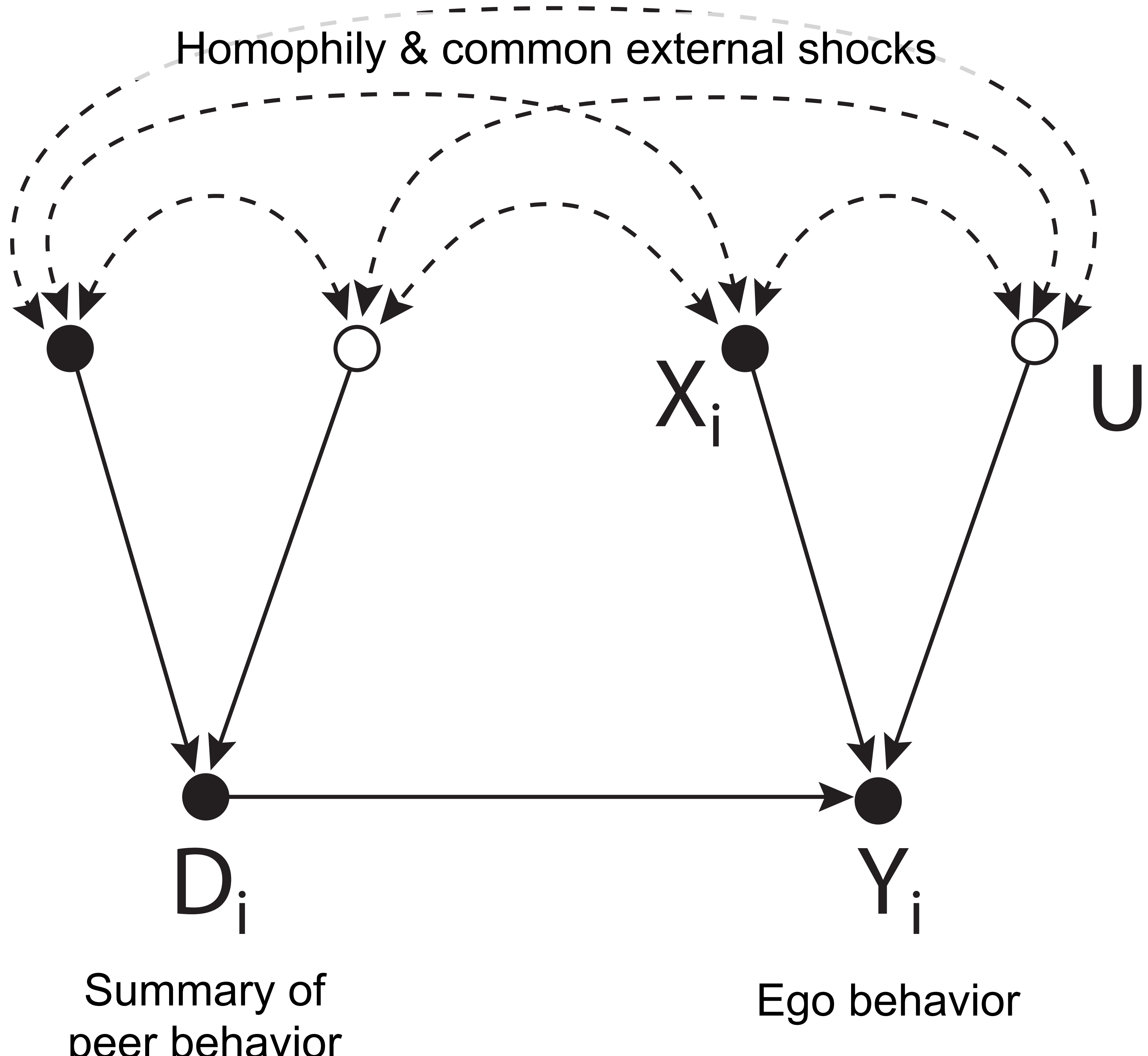
Goal: Estimate peer effects

- What would the ego do given different peer behaviors?
  - How does a marginal peer adopting effect ego adoption?
  - Which peers' adoptions are most influential on the ego?

- Ideal experiment:

Directly assign behaviors of existing peers

$Y_j(t)$
$Y_i(t+1)$
Peer behavior
Ego behavior


Homophily & common external shocks
X
i
U
i
D
i
Y
i
Summary of
peer behavior
Ego behavior

Goal: Estimate peer effects

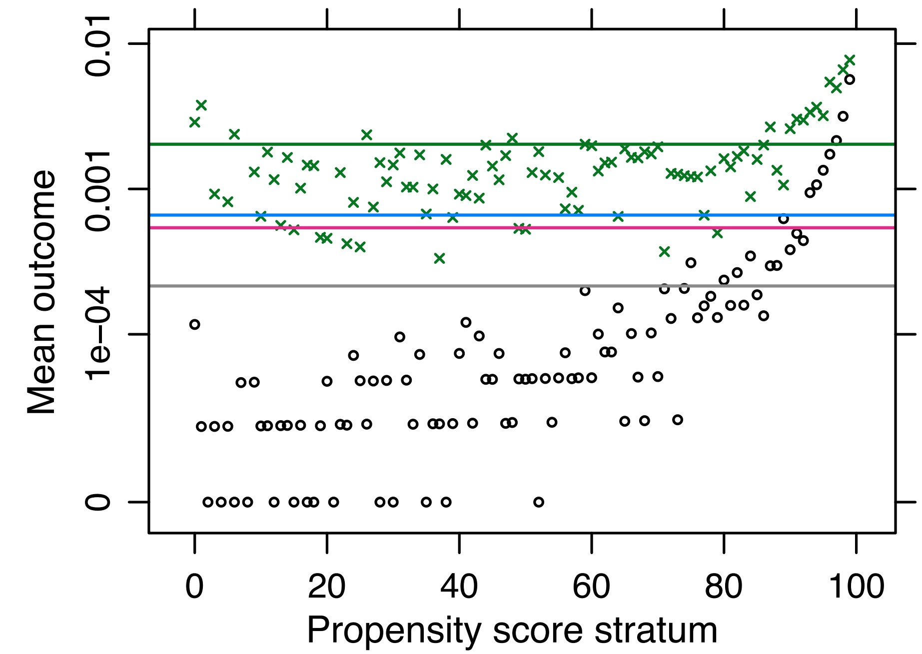
- Ideal experiment: *directly assign behaviors of existing peers*

- Alternatives
  1. Adjust for confounding in observational data
  2. Directly assign behaviors of *inauthentic* peers
  3. Assign individuals to *new* peers
  4. Modulate a *mechanism* by which peer behaviors have their effects
  5. *Indirectly* affect (i.e., encourage) behaviors of existing peers

Observational estimation of peer effects

High dimensional adjustment with propensity score stratification

1. Model probability of being exposed to a peer adoption using thousands of prior behaviors & other covariates
2. Compute mean outcome for exposed and unexposed units in each stratum of propensity score
3. Combine stratum-specific estimates, weighting by number of exposed cases



Early example: Mobile app adoption (Aral, Muchnik & Sundararajan, *PNAS*, 2009)

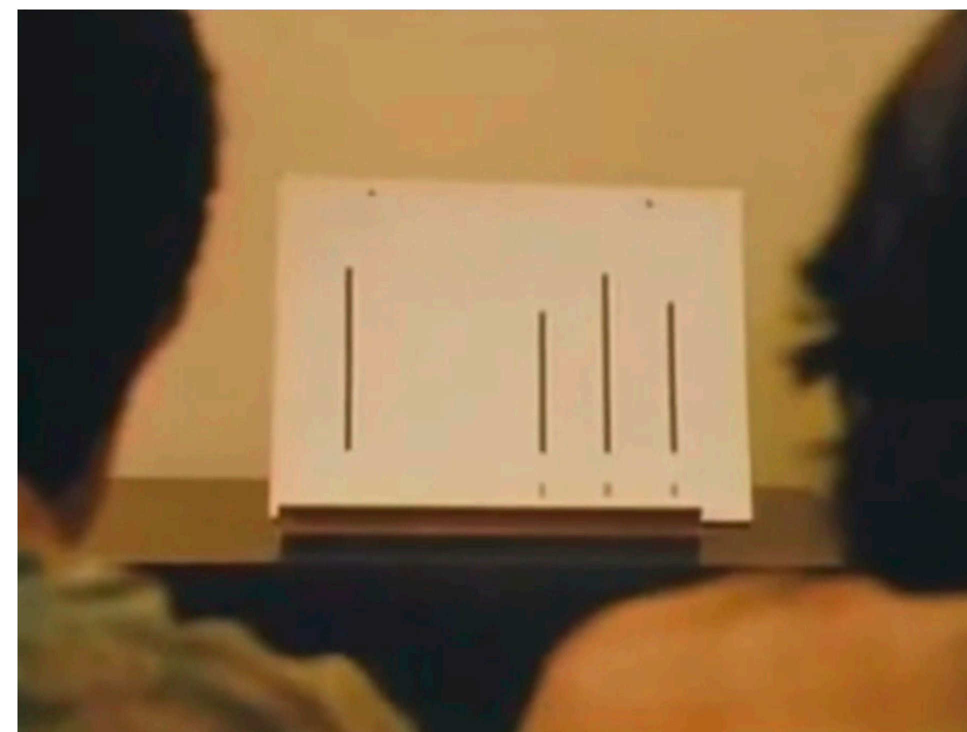
Evaluation with experimental gold standard: Link sharing on Facebook

- Naïve (unadjusted estimates) have **> 300% bias** for relative risk (28.5 vs. 6.8)
- Fully adjusted estimates have **< 30% bias** for relative risk (8.7 vs. 6.8)

Eckles, D. & Bakshy, E. Bias and high-dimensional adjustment in observational studies of peer effects.

Assign behaviors of *inauthentic* peers

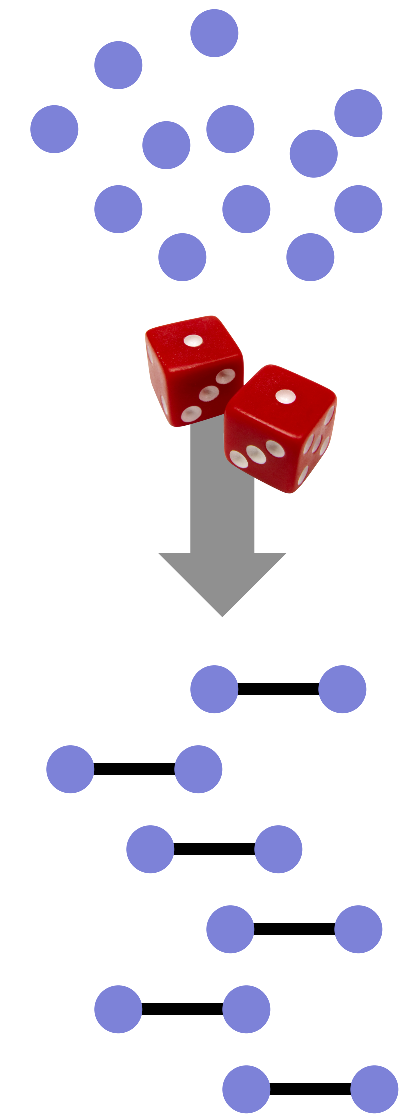
Individuals enter an artificial social environment with peers who are employed by the experimenter (i.e., confederates)



Asch, S. E. (1956) "Studies of independence and conformity: I. A minority of one against a unanimous majority." *Psychological Monographs: General and Applied* 70.9: 1-70.

Assign individuals to *new* peers

- Individuals are randomly assigned a position in a constructed network, group, or dyad
  - Freshmen assigned to college roommates (Sacerdote 2001)
  - Health network users assigned to random or clustered networks (Centola 2010)
  - Military academy freshmen assigned to squadron (Carrell, Hoekstra & West 2011)
  - Poor families given vouchers for rent in particular neighborhoods (Kling, Liebman, Katz 2007)
- Confounding of multiple peer behaviors and traits (Peers are complex bundles of traits and behaviors)



Modulate a *mechanism* of peer effects

- Peer behaviors often affect the ego via a small number of non-deterministic mechanisms
- Randomize whether peer adoption is communicated to ego

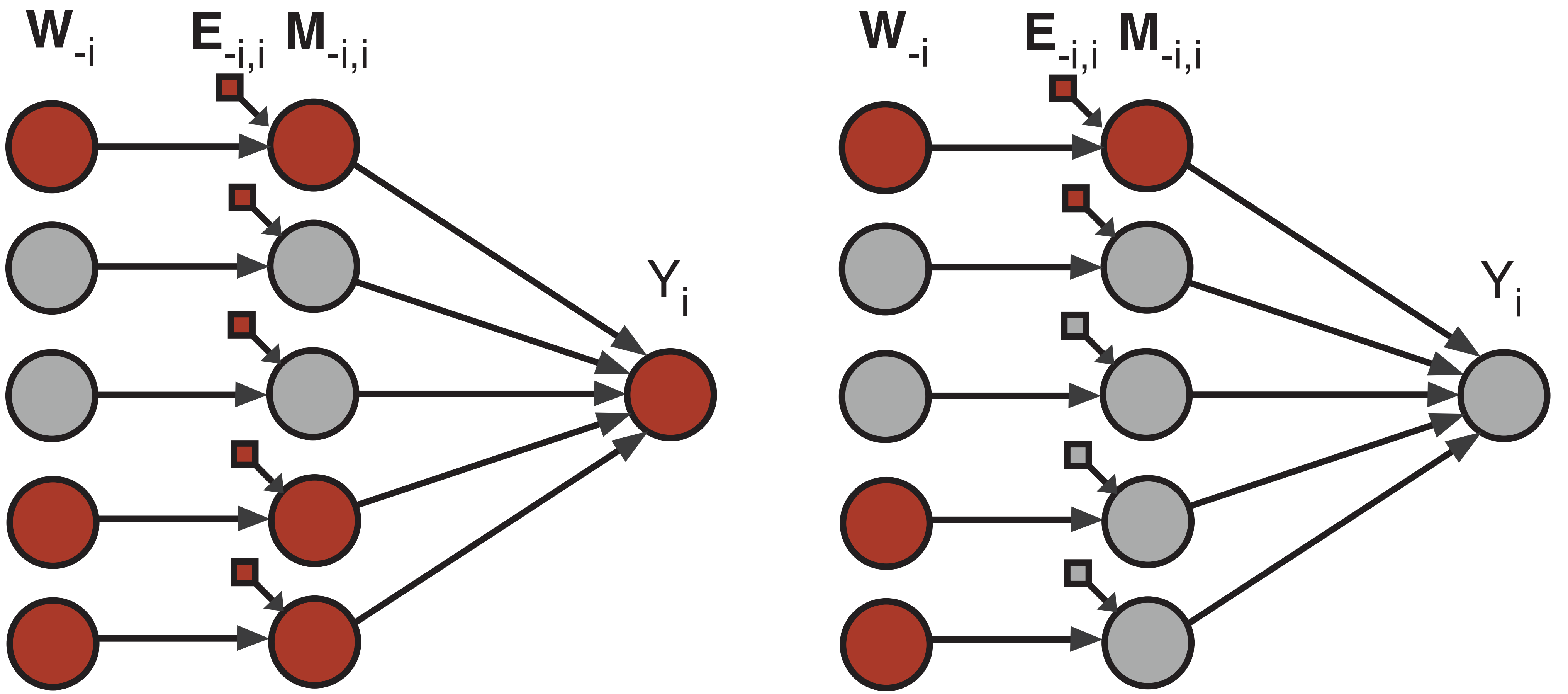
(Aral & Walker 2011; Bakshy et al. 2012a,b)

Do not show personalized social cue
D = 0

Show available personalized social cue
D = 1

Bakshy, E., Eckles, D., Yan, R., & Rosenn, I. (2012). Social influence in social advertising: Evidence from field experiments. In *Proc. of EC*. ACM. http://arxiv.org/abs/1206.4327

Mechanism experimental designs



All mechanisms enabled

Some mechanisms enabled

When mechanism is deterministic: $M_{j,i} = W_j E_{j,i}$

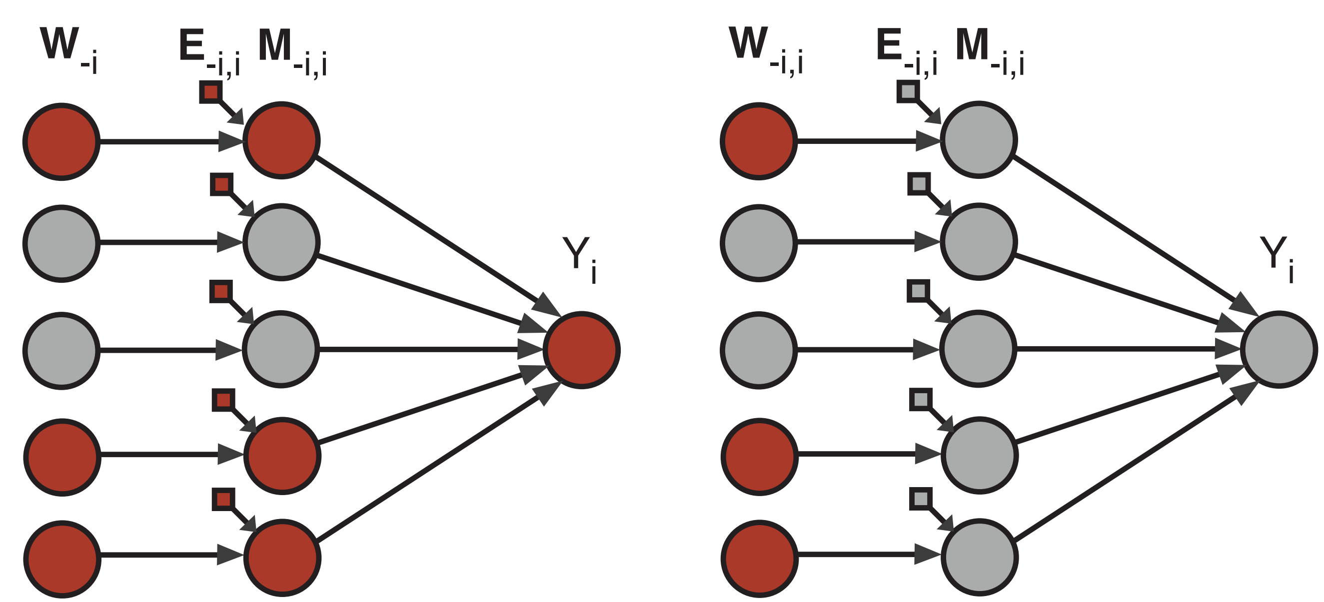
All variables represented by circles may have other common causes not shown. Variables represented by squares are root nodes.

Modulating a mechanism of peer effects

- Identifies the average treatment effect on the treated
  - If the mechanism is deterministic and exhaustive, then a mechanism experiment identifies the ATET
  - Binary case: $$p^{(0)} = \mathrm{P}(Y = 1 \mid M = 1, \mathrm{do}(M = 0))$$ $$p^{(1)} = \mathrm{P}(Y = 1 \mid M = 1)$$

$$\delta = p^{(1)} - p^{(0)}$$

  - Rarely actually deterministic and exhaustive…



All mechanisms enabled vs. All mechanisms disabled

Peer effects via a minimal social cue

- Experimental design
  - 1.4e8 user–ad pairs randomly assigned whether social ad unit includes minimal social cue mentioning affiliated peer
    - 5.7e6 distinct users, 1.2e6 distinct ads

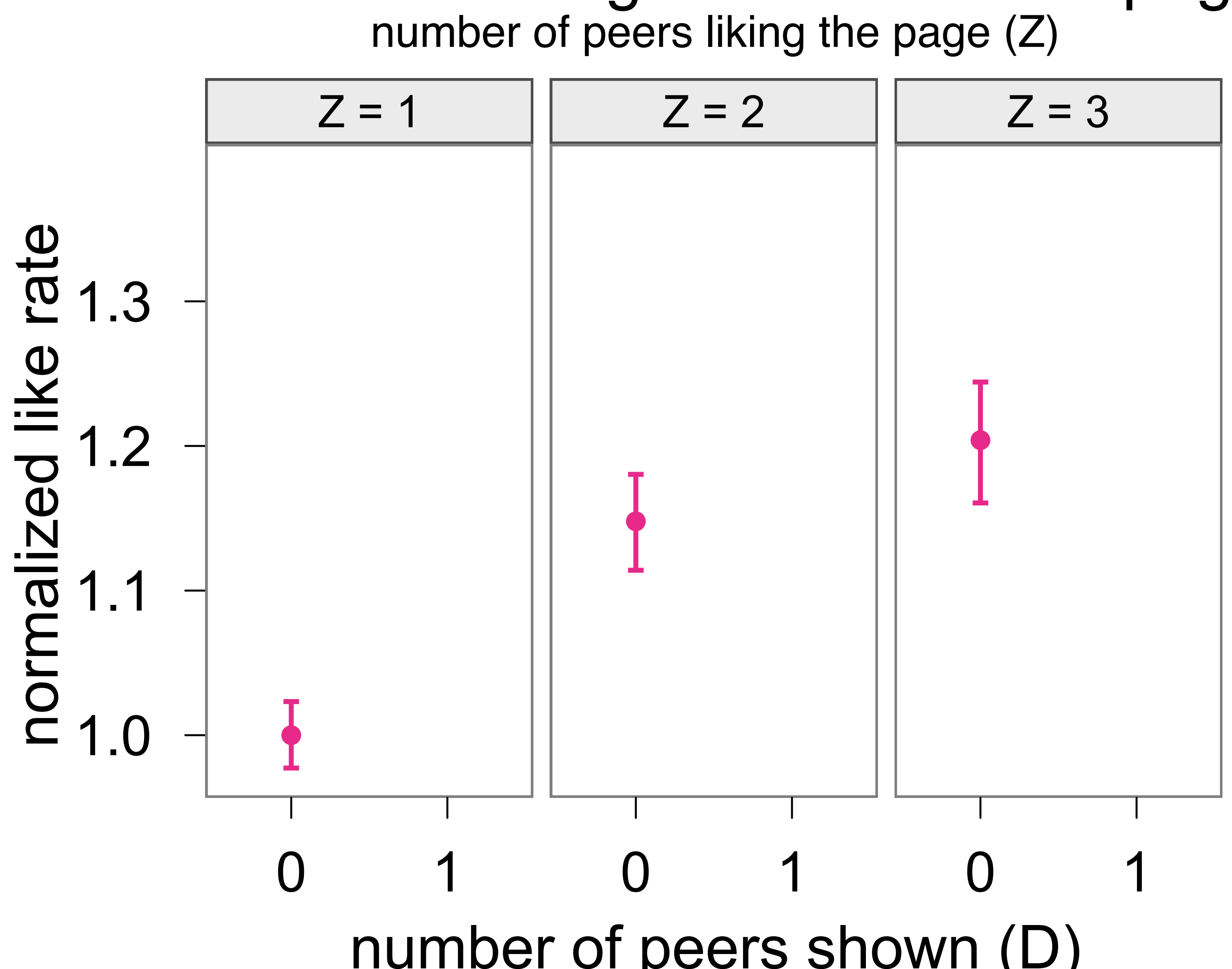
Make entire experiment consist of changes to small, light grey text on white background

D = 0

D = 1

Peer effects via a minimal social cue

- Relative increases in affiliating with advertised page



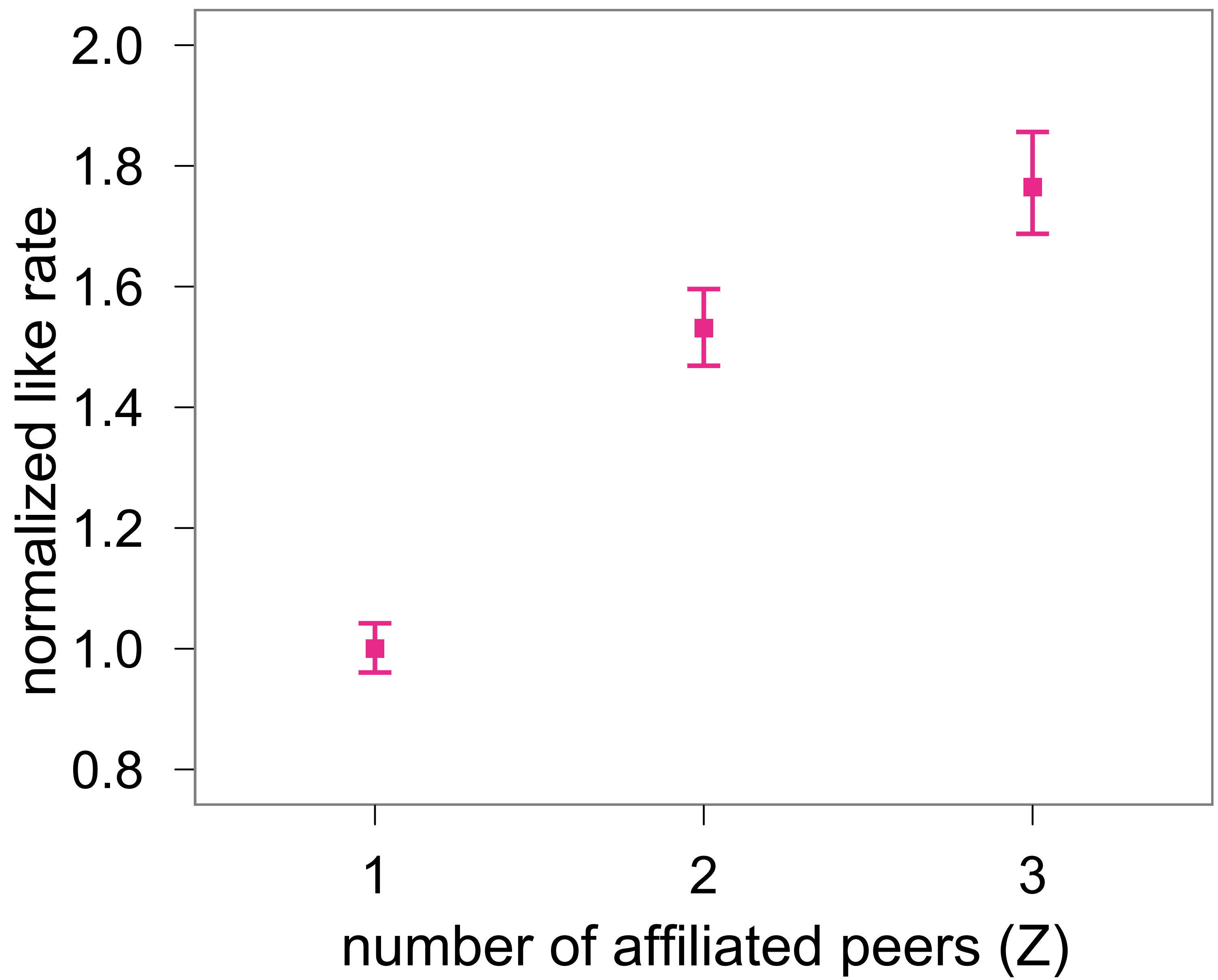
Bakshy, E., Eckles, D., Yan, R., & Rosenn, I. (2012). Social influence in social advertising: Evidence from field experiments. In *Proc. of EC.* ACM.

What is the marginal effect of social cues on an action?

Bakshy, E., Eckles, D., Yan, R., & Rosenn, I. (2012). Social influence in social advertising: Evidence from field experiments. In: *EC 2012: Proceedings of the ACM Conference on Electronic Commerce*. ACM. http://arxiv.org/abs/1206.4327

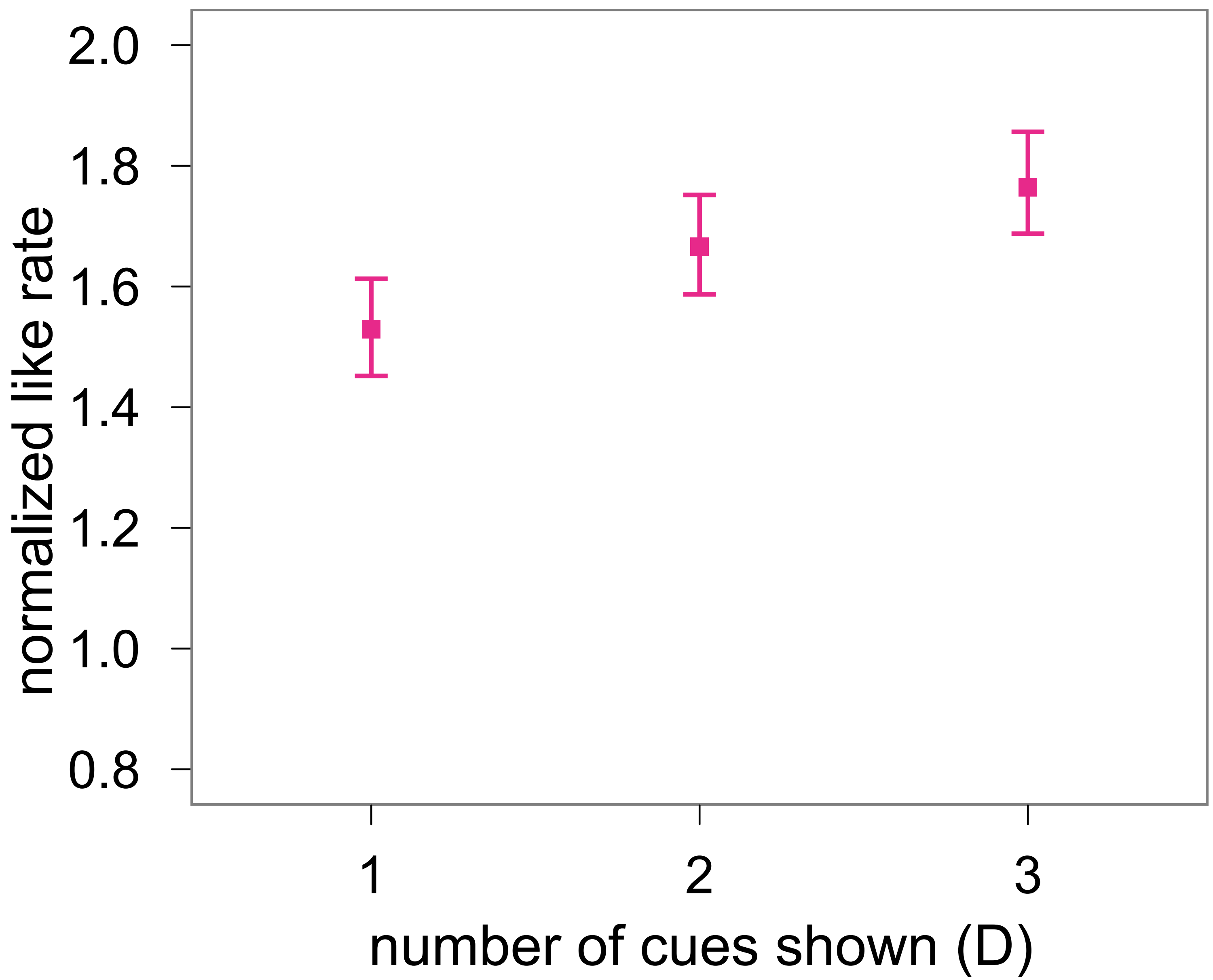
Average cue–response function

Naïve observational analysis from earlier



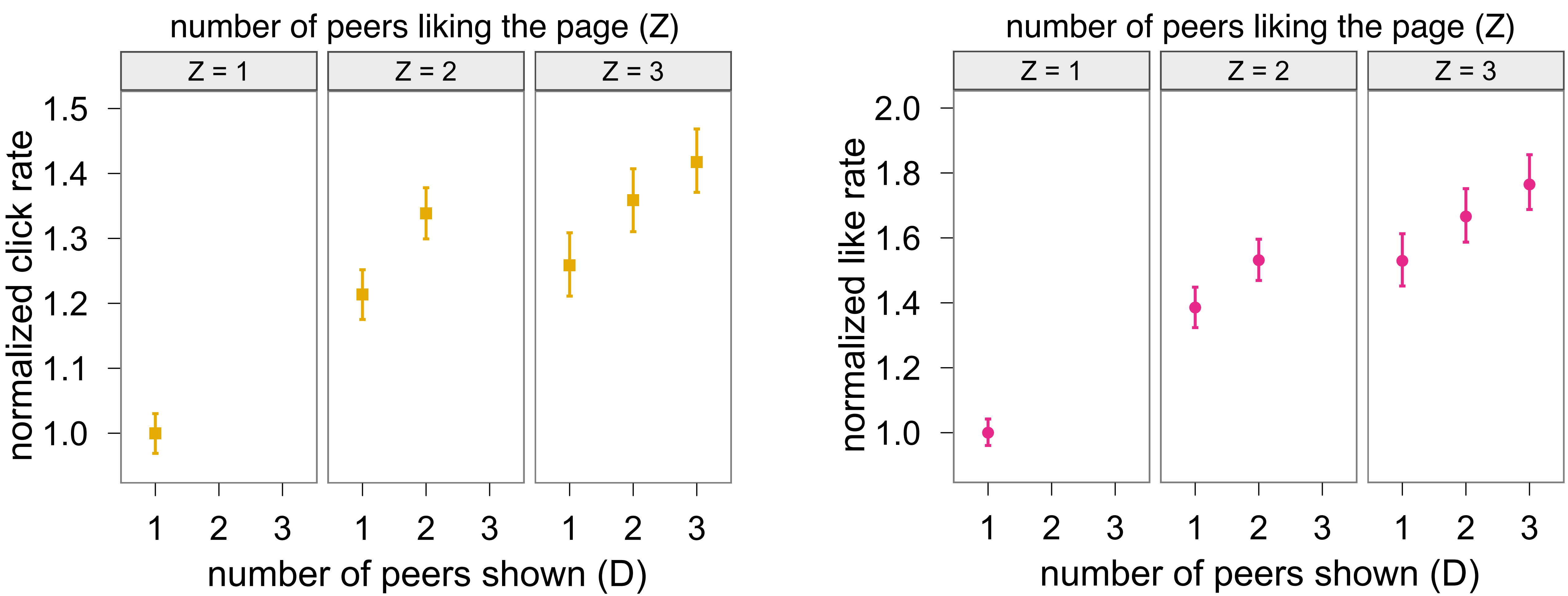
Average cue–response function

Experimental analysis for number of affiliated peers $Z = 3$



Average cue–response function

Experimental analysis for $Z = 3$



Bakshy, E., Eckles, D., Yan, R., & Rosenn, I. (2012). Social influence in social advertising: Evidence from field experiments. In: *EC 2012: Proceedings of the ACM Conference on Electronic Commerce*. ACM. http://arxiv.org/abs/1206.4327

Indirectly affect behaviors of existing peers

- Peer encouragement designs
  - Randomly assign vertices to encouragement to behavior of interest, and examine how this spills over to others
    - These designs, with groups, are used in development and labor economics (e.g., Angelucci & De Giorgi 2009, Duflo & Saez 2003, Miguel & Kremer 2004; cf. Moffitt 2001)
    - Similar analyses treating other non-randomly assigned variables as instrumental variables (e.g., Shriver et al. 2013, Tucker 2008)
    - Recent example on Twitter (Coppock et al. 2015)
  - Estimate effect of peer behaviors – not just the encouragement
  - Other view of this: assign peers to behaviors, see effect on ego

Encouragement designs

- Randomly assign units to encouragement Z to a focal behavior D
  - Randomly encourage students to study (Powers & Swinton 1984)
  - Assign to take drug or not (but they may not take it)
- Formal analysis using potential outcomes (Holland 1986, 1988)
  - Total effect of encouragement (intent-to-treat, ITT):

$$Y_i(Z_i = 1) - Y_i(Z_i = 0)$$

  - Effect of behavior (effect of D on Y):

$$Y_i(D_i = 1) - Y(D_i = 0)$$

  - Can we use Z to estimate the effect of D on Y?

Binary encouragement designs

- Four types of people by potential outcomes:
  - **Compliers** treatment if encouraged, control if not
  - **Always-takers** treatment whether encouraged or not
  - **Never-takers** control whether encouraged or not
  - **Defiers** control if encouraged, treatment if not

  - Not all of these may exist for a particular study
    - In a trial of a new drug or offering (removing) a new (existing) feature, there are neither always-takers nor defiers

Binary encouragement designs

- Latent types of units

		Treatment D	
		0	**1**
Assignment Z	**0**	Compliers & Never-takers	Defiers & Always-takers
	1	Defiers & Never-takers	Compliers & Always-takers

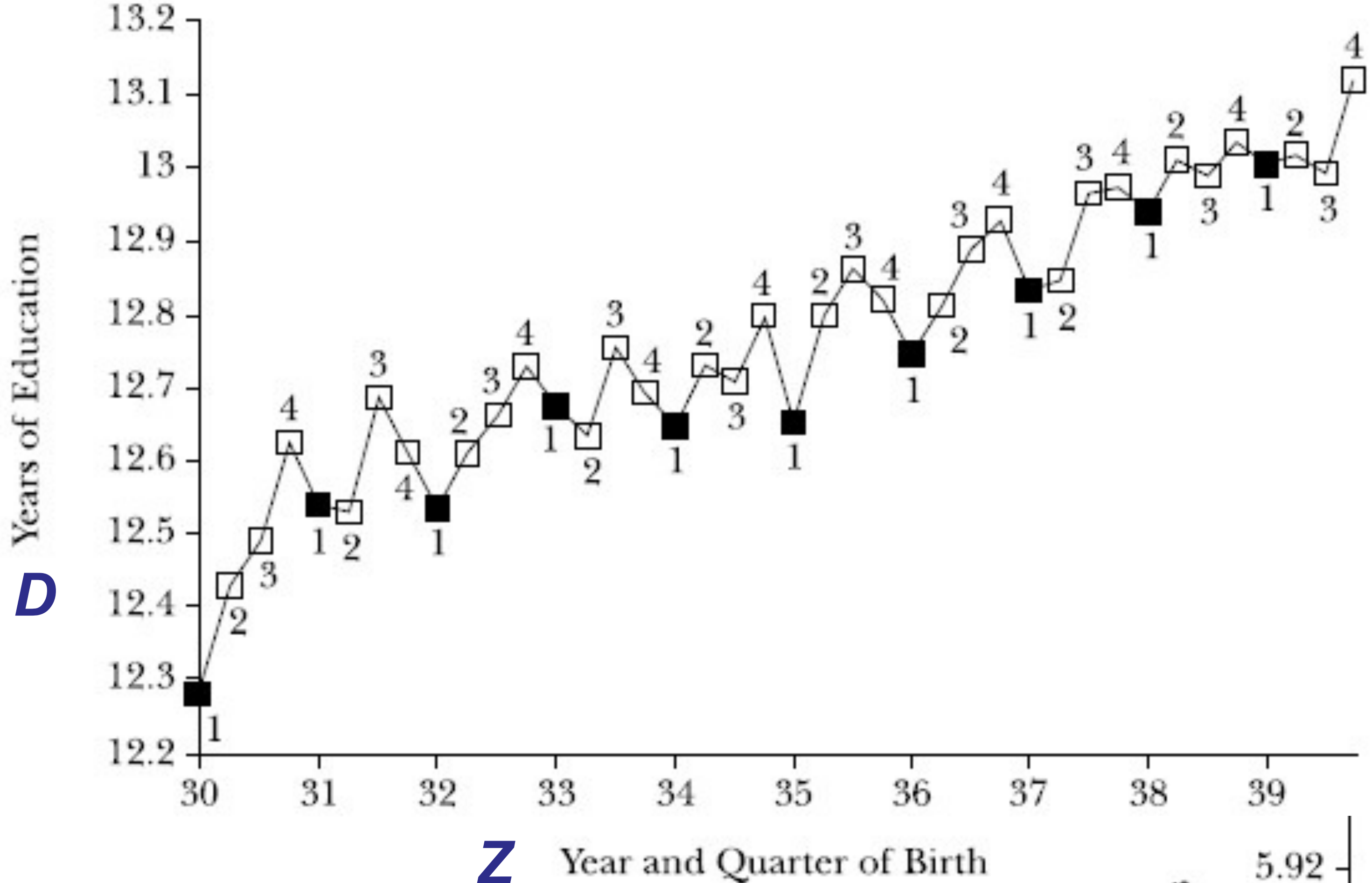
IV analysis of encouragement designs

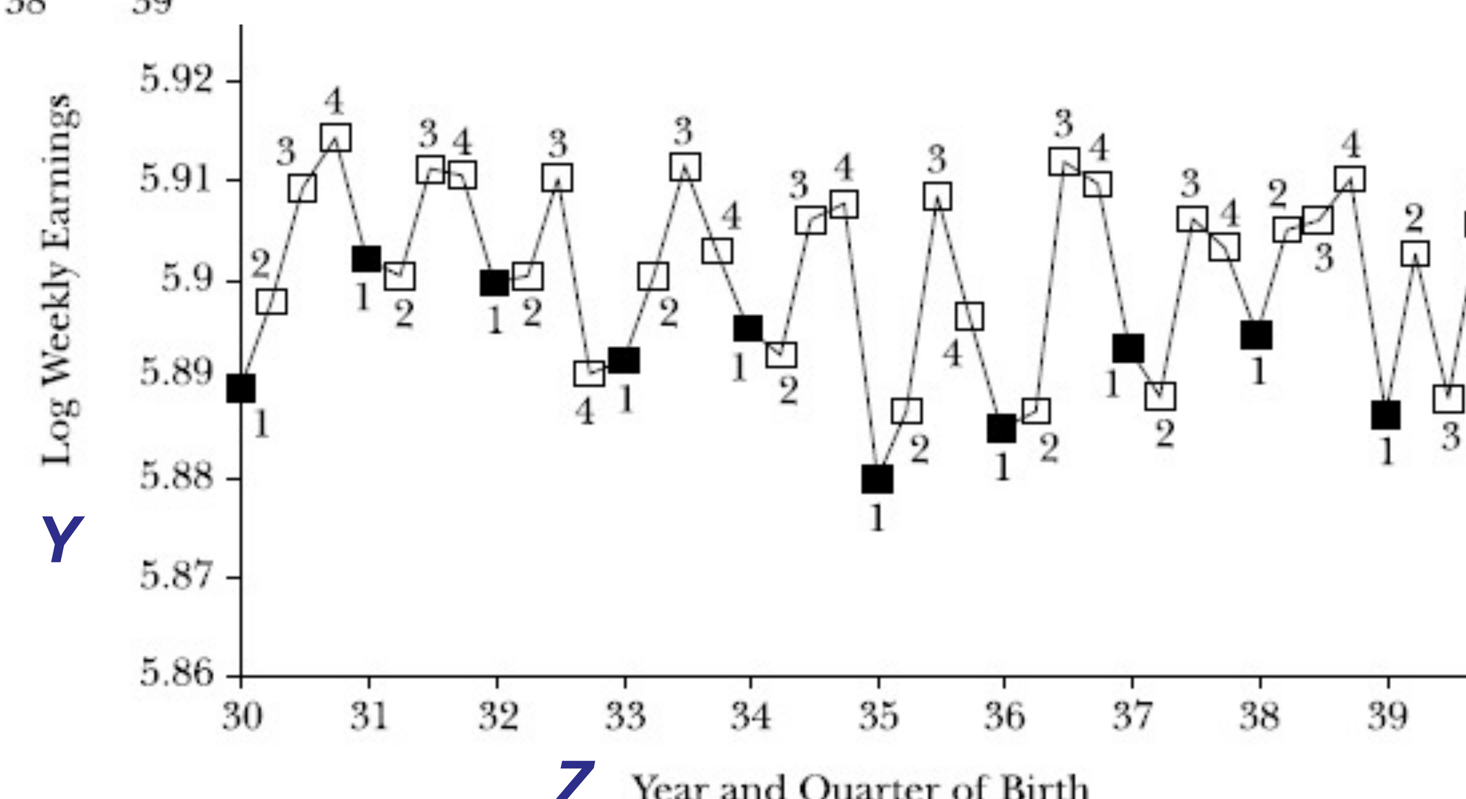
- with heterogeneous treatment effects

Monotonicity. With probability 1, $D_i^{(z)} \geq D_i^{(z')}$ for all $z \geq z'$ and all i.

  - Then local average treatment effect (LATE) is identified
    - In binary Z, D case, LATE is the average treatment effect for the population of compliers
  - Are we interested in the LATE?

(Angrist, Imbens & Rubin 1996)



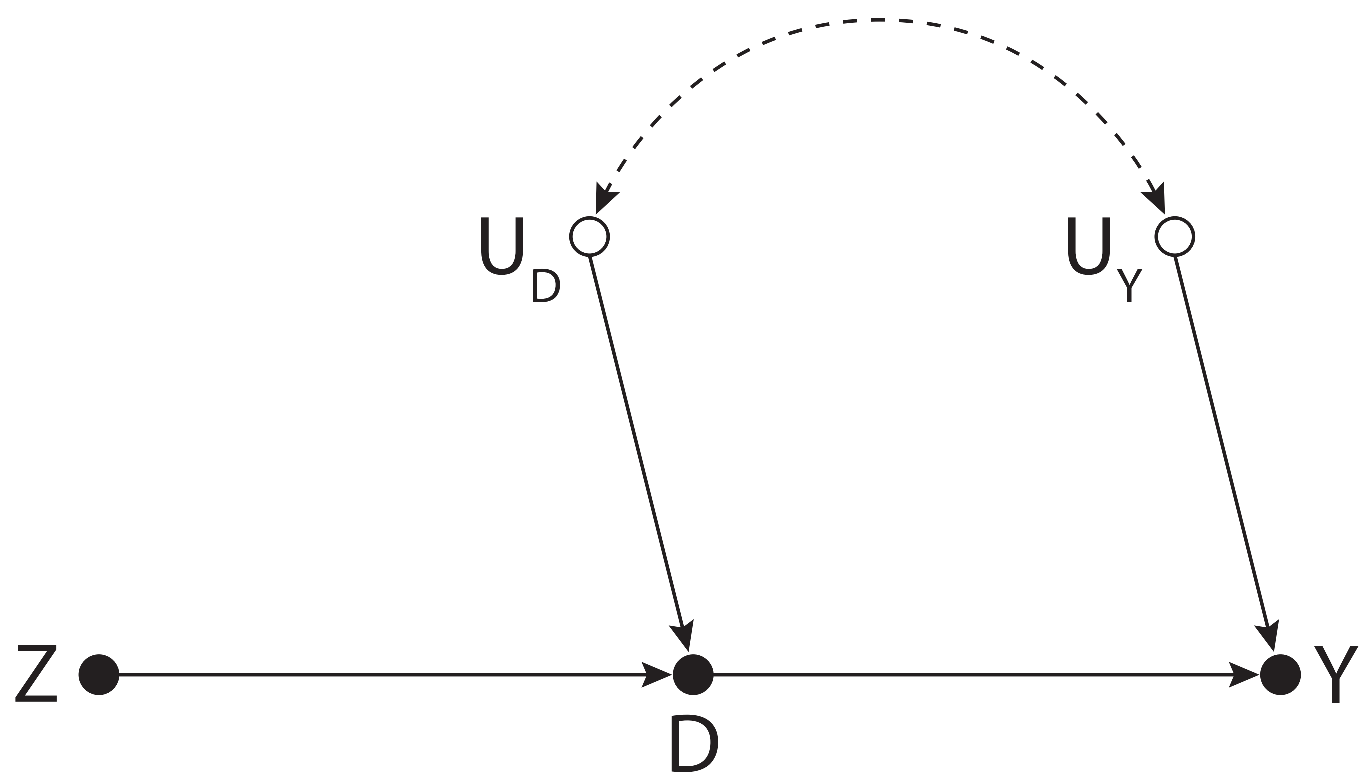


(Angrist & Krueger 1991)

Encouragement designs

"The best instrumental variables are randomly assigned"

- Find a variable *Z* that affects *D* but is otherwise unrelated to *Y*
- Use this exogenous variation in *D* to estimate effect of *D* on *Y*



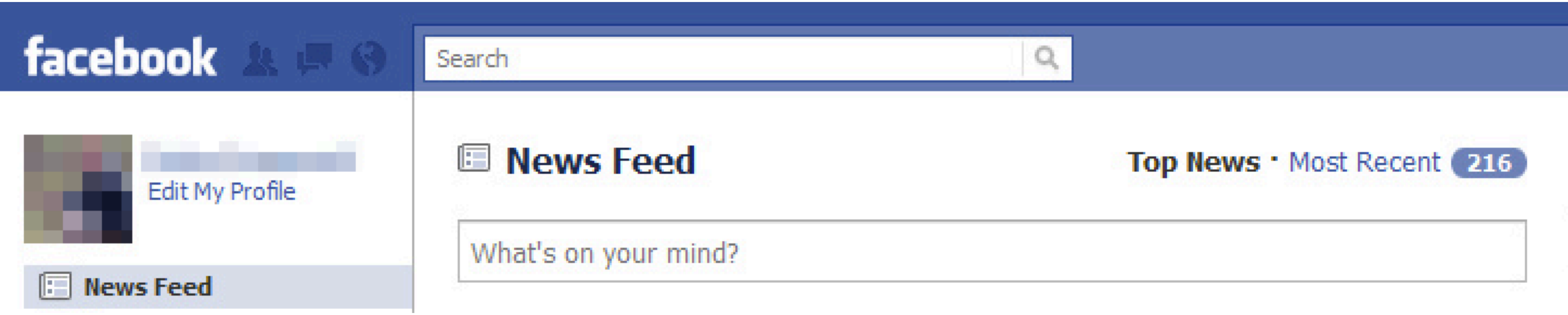
Causal DAG illustrating satisfaction of some necessary conditions

Peer encouragement designs

With a single behavior of interest

- Assign to encouragement to, e.g.,
  - Enroll in a retirement savings account (Duflo & Saez 2003)
  - Post a thankful status update on Thanksgiving Day
- Summarize peer assignments (e.g., number of peers assigned)
- Summarize peer behaviors (e.g., number of adopter peers)

… Compute average ego behaviors as a function of these



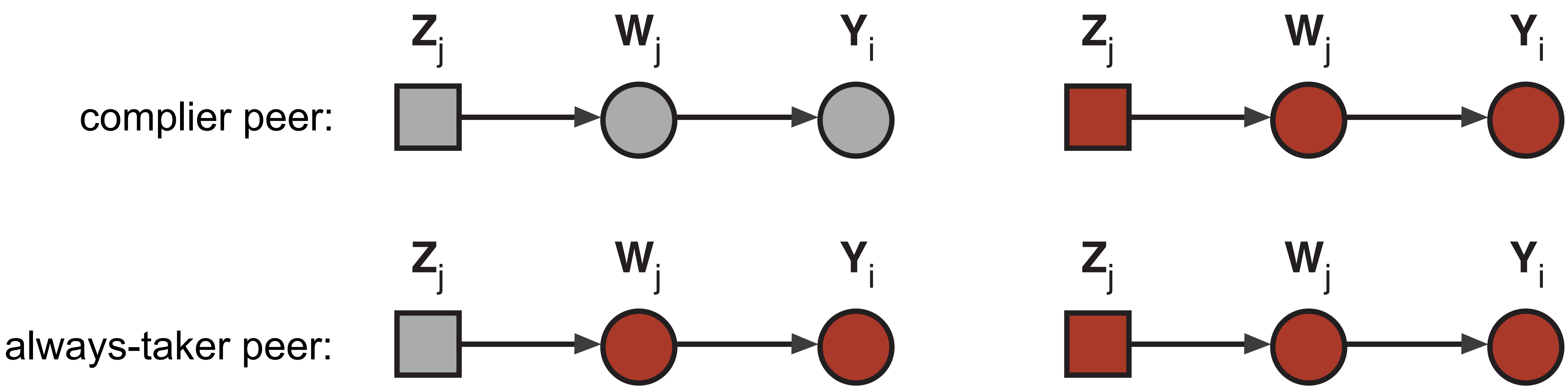
On Thanksgiving Day 2010, 1% of Facebook users in the United States who were using Facebook in American English were randomly assigned to be presented with an alternative prompt, “What are you thankful for?”

Top left of the Facebook homepage (2010)

Peer encouragement designs with dyads

- Binary encouragement, peer behavior, and ego outcome

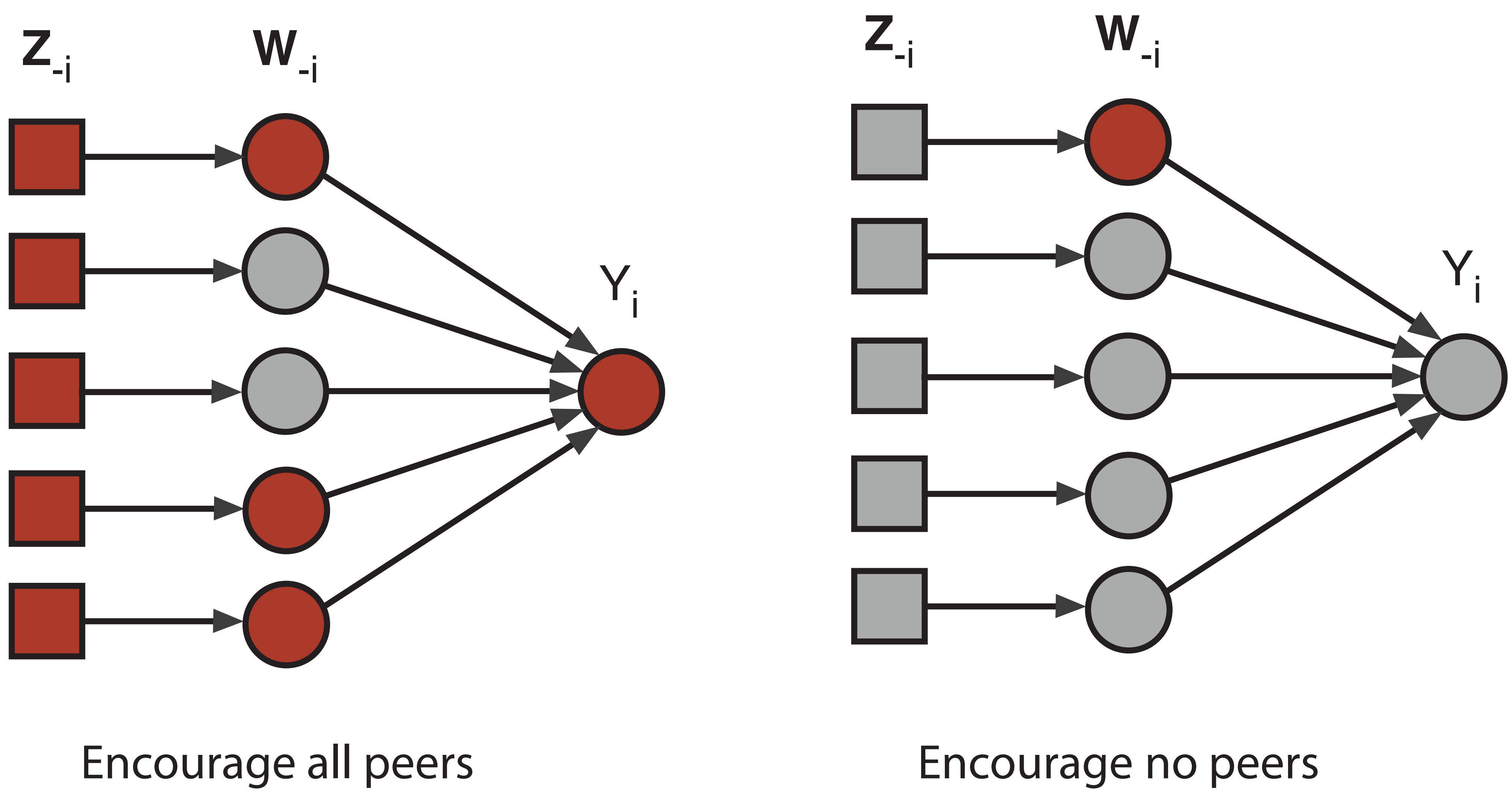
1. Randomly encourage j or not
2. Observe j's behavior (endogenous treatment for i)
3. Observe i's behavior (outcome)



All variables represented by circles may have other common causes not shown. Variables represented by squares are root nodes.

Peer encouragement designs

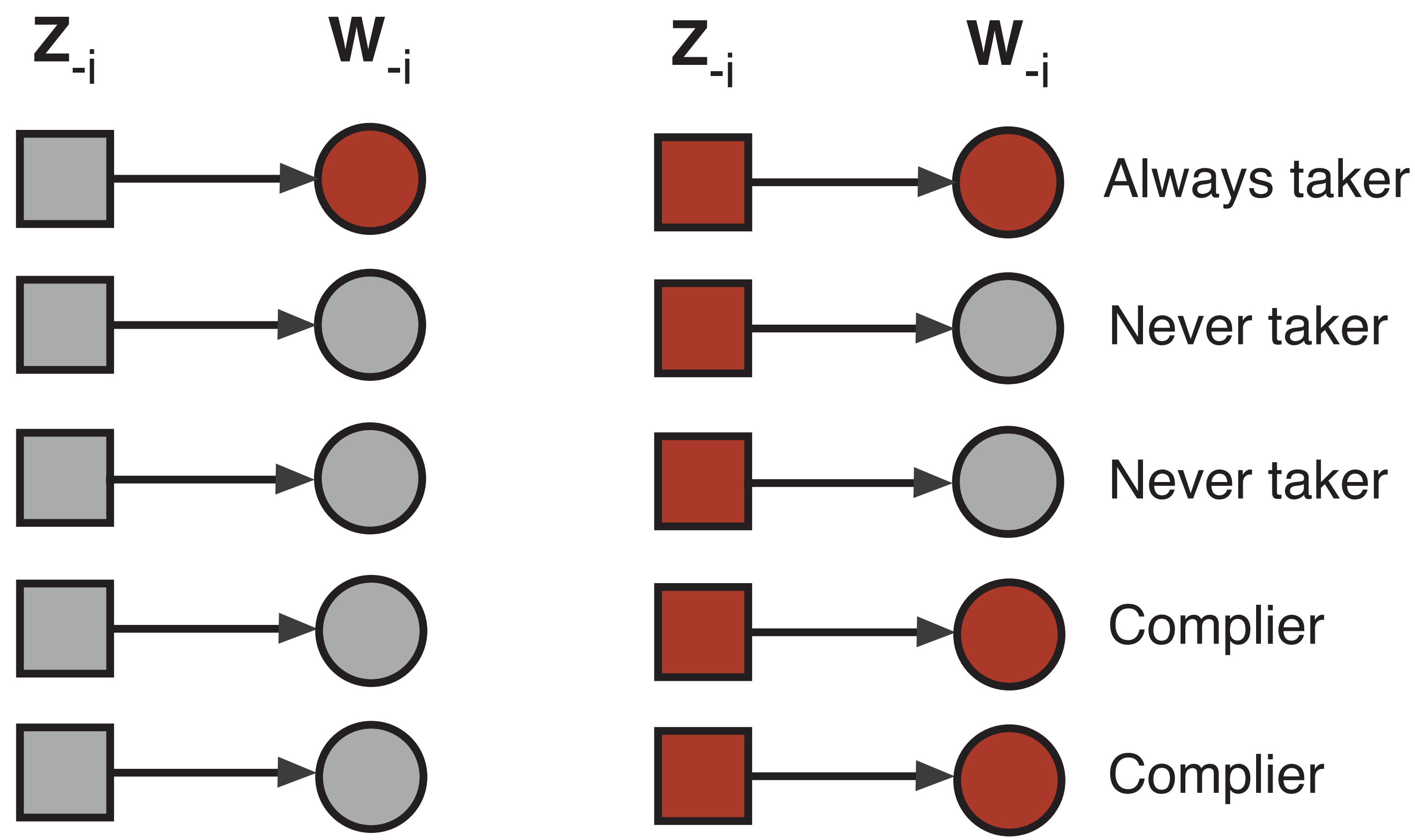
- With multiple peers



All variables represented by circles may have other common causes not shown. Variables represented by squares are root nodes.

Peer encouragement designs

- Noncompliance with multiple peers
  - An ego's peer may be a mix of compliance types



Analysis of peer encouragement designs

- Intent-to-treat
  - Put some structure on the potential outcomes

    $$Y_i(\vec{z}) = Y_i(\vec{z}') \text{ for all } \vec{z}, \vec{z}' \text{ s.t. } d = f_i(\vec{z}) = f_i(\vec{z}')$$

    - Analyze ego outcome as function of number of peers assigned
  - Nuisance issues:
    - Compute probabilities of assignment $\pi_i(\vec{Z})$
    - Use inverse probability weighting to estimate average outcomes

(Aronow & Samii 2012; Eckles, Karrer & Ugander 2014; Ugander, Karrer, Backstrom & Kleinberg 2013)

Analysis of peer encouragement designs

Instrumental variables analysis

- Ego’s outcome caused by own assignment and peer behaviors

- Homogeneous effects of different peers
(i.e., the number of adopter peers is what matters)

$$Y_{it_1}(z, \vec{y}_{it_0}) = Y_{it_1}(z, \vec{y}'_{it_0}) \text{ for all } \vec{y}_{it_0}, \vec{y}'_{it_0} \text{ s.t. } a = h_i(\vec{y}_{it_0}) = h_i(\vec{y}'_{it_0})$$

- Two-stage least squares (or other related methods)

Peer encouragement designs

Instrumental variables analysis

- Number of encouraged peers is an instrument for peer effects
  - *Complete mediation (aka exclusion restriction)* – Peer encouragement only affects ego behavior via peer behavior
  - Even if effects are heterogeneous, IV analysis of encouragement designs identifies average treatment effect of interest
    - *Local average treatment effect* or *average causal response*

      (Angrist, Imbens & Rubin 1996, Angrist & Imbens 1995)
    - Likely an advantage over other instruments that aren't encouragements

Effects of receiving feedback

Motivation

- When an individual shares content in social media, what are the effects of receiving additional feedback (likes & comments)?
  - Generating further conversation (e.g., ego replies to comments)
  - In-kind peer effects in giving feedback (including generalized reciprocity)
  - Creating and sharing more content in the future

- These posited virtuous cycles are critical to the adoption and continued use of communication technologies

Fact: Pre-expanding vs. not pre-expanding comment boxes modulates feedback

Expanded comment box

Higher interaction rate

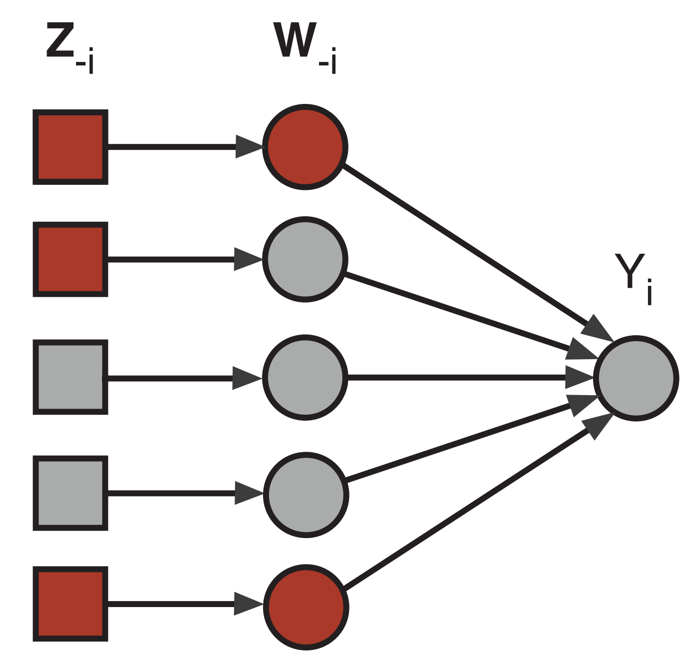
Unexpanded comment box

Lower interaction rate

Alternative experimental designs

Example: Effects of receiving feedback on posts

- Assign viewers to condition encouraging giving feedback on whatever they view
- Assign posters to condition encouraging their friends to give them feedback
- Assign directed edges to encouragement of feedback in that direction (from person A to person B)



Encourage some peers

Some interesting problems / things to try

Easier (using Cai et al. 2015 data)

- Alternative ways of selecting focal units, choosing test statistics
- Simulate graph clustered design data (Note: some other treatments are clustered at village already)

Harder

- Graph partitioning to minimize MSE of ATE estimate – how does this change the optimization problem?
- Optimal design under an approximation to the graph (e.g., a stochastic block model)

Further resources

Randomization inference in general / networks:

Exact p-values for network interference
Susan Athey, Dean Eckles & Guido W. Imbens
NBER Working Paper No. 21313. (and arXiv)

Field Experiments: Design and Analysis
Gerber and Green.

Clustered random assignment in networks:

Graph cluster randomization: Network exposure to multiple universes
Ugander, J., B. Karrer, L. Backstrom, and J. M. Kleinberg. KDD 2013.

Design and analysis of experiments in networks: Reducing bias from interference
Eckles, D., Karrer, B., & Ugander, J.
http://arxiv.org/abs/1404.7530

General causal inference references:

Causal Inference in Statistics, Social, and Biological Science
(with potential outcomes, including randomization inference)
Imbens & Rubin

Counterfactuals and Causal Inference
(with potential outcomes & causal DAGs)
Morgan & Winship

Causality
(mainly causal DAGs, also causal discovery)
Pearl

Worked examples on github:
https://github.com/deaneckles/randomization_inference