Estimating average causal effects under general interference between units

Peter M. Aronow and Cyrus Samii

Yale University and New York University

May 23, 2012

- Randomized experiments often involve treatments that may induce "interference between units"
- Interference: the outcome for unit i depends on the treatment assigned to unit j. If we administer a treatment to unit j, what are the effects on unit i?
- Recent work in non-parametric inference focuses on hypothesis testing or estimation in hierarchical (i.e., multilevel) interference settings.
- We develop a theory of estimation under *general* forms of interference.

- We provide a nonparametric *design-based* (c.f. Neyman 1923) method for estimating average causal effects, including, but not limited to:
  - Direct effect of assigning a unit to treatment
  - Indirect effects of, e.g., a unit's peer being assigned to treatment
  - More complex effects (e.g., effect of having a majority of proximal peers treated)
- In so doing, we highlight how equal probability of treatment assignment does not imply equal probability of indirect exposure to treatment (e.g., proximity to treated units)
- We develop our main results drawing on classical sampling theory, though model-assisted refinements are possible

Method summary:

- Design information gives probability distribution for treatment,

$$\mathbf{Z} \text{ s.t. } \text{supp}(\mathbf{Z}) = \Omega.$$

- Specify an exposure model that converts assigned treatment vectors $\mathbf{z} \in \Omega$ to exposures based on unit attributes (e.g., network degree),

$$f(\mathbf{Z}, \theta_i) \equiv D_i$$

- Implies the *exact* probabilities of exposure:

$$\pi_i(d_k) = \sum_{\mathbf{z} \in \Omega} p_z \mathbf{I}(f(\mathbf{z}, \theta_i) = d_k)$$

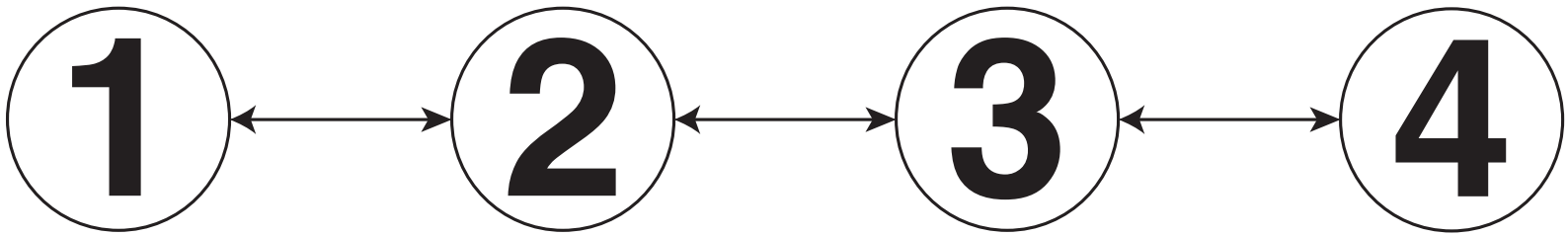
- Average causal effects are the average difference between the potential outcomes under exposure d_k vs. those under d_l.
- Estimate average causal effects accounting for varying probability of exposures (via some variant of inverse probability weighting).

Roadmap:

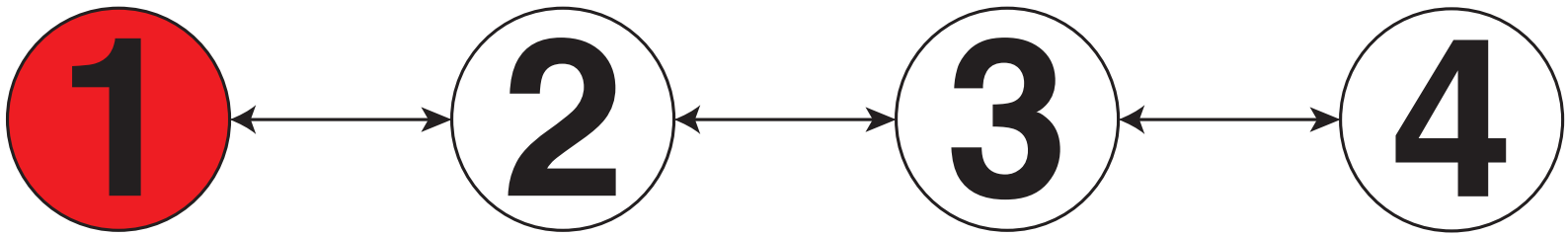
- Simple running example.
- Some technical details.
- Application.
- Anticipating some concerns.

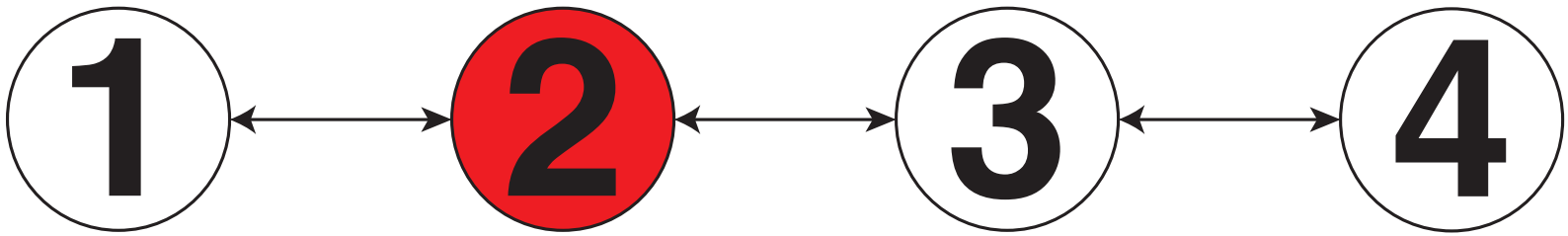
Simple running example.

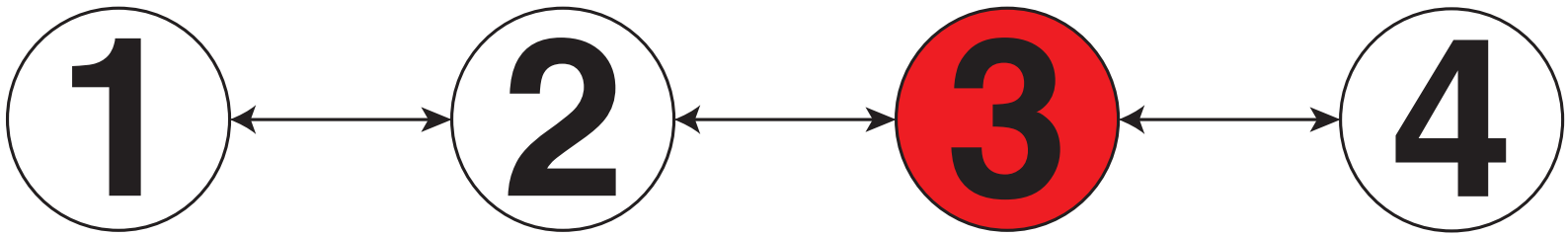
- Consider a randomized experiment performed on a finite population of four units in a simple, fixed network:

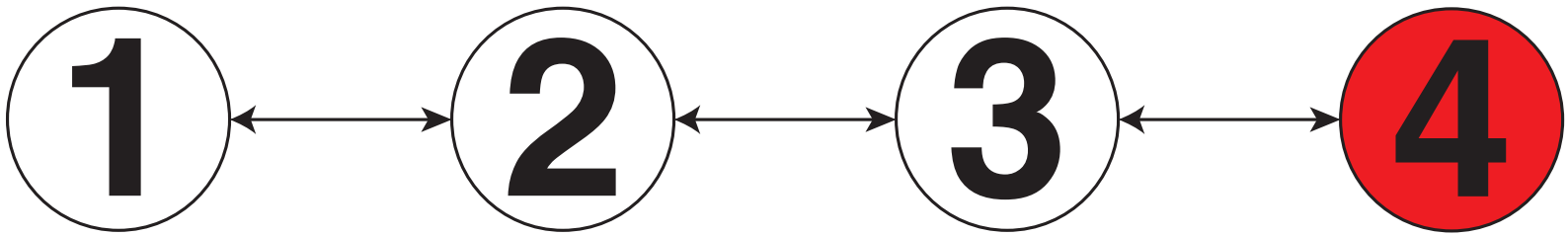

1
2
3
4

- One of these units is assigned to receive an advertisement and the other three are assigned to control, equal probability
- We want to estimate the effects of advertising on opinion
- There are four possible randomizations $\mathbf{z}$:


1
2
3
4


1
2
3
4


1
2
3
4

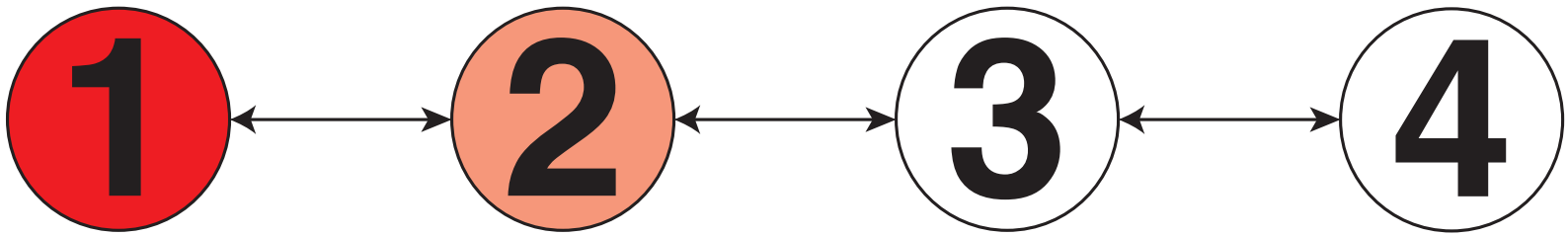

1
2
3
4

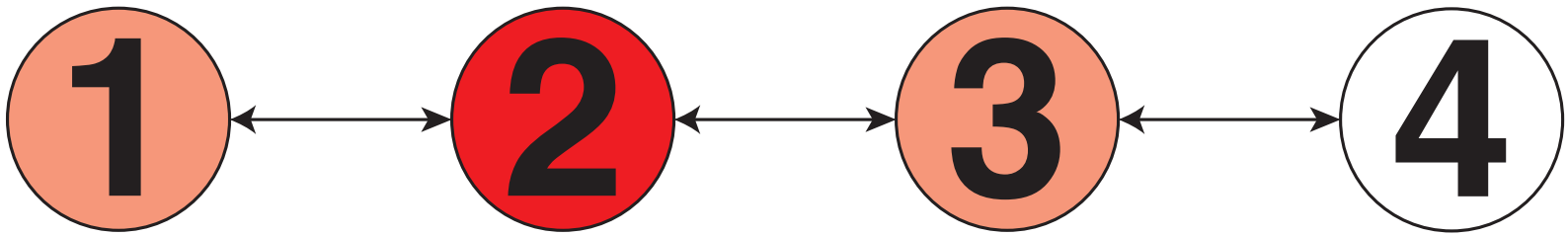
- So we have exact knowledge of the randomization scheme.
- But what of the exposure model? This requires researcher discretion. How do we model exposure to a treatment?
- One example.

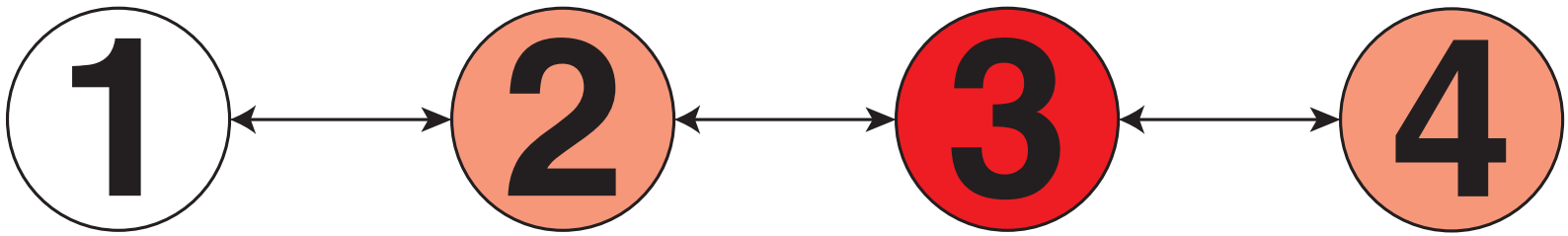
- Direct exposure means that you have been treated.
- Indirect exposure means that a peer has been treated.

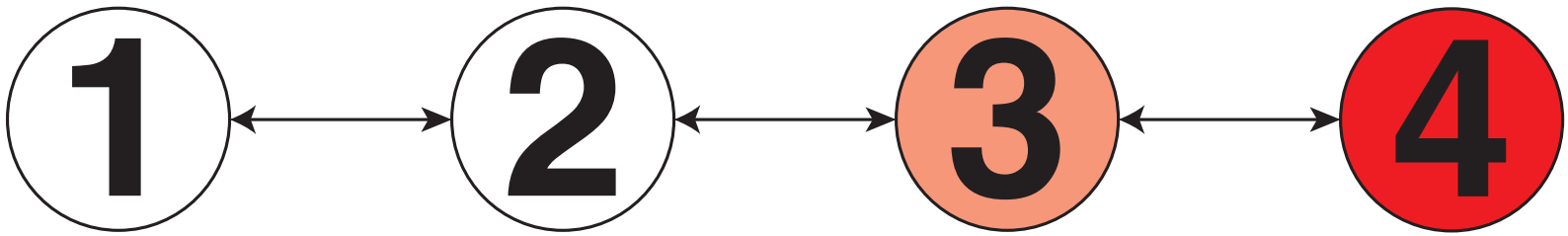
$$D_i = \begin{cases} \text{Di(rect)}: & z_i = 1 \\ \text{In(direct)}: & z_{i\pm1} = 1 \\ \text{Co(ntrol)}: & z_i = Z_{i\pm1} = 0. \end{cases}$$

- There is nothing particularly special about this model, except for its parsimony. Arbitrarily complex exposure models are possible.
- Let's visualize this.


1
2
3
4


1
2
3
4


1
2
3
4


1
2
3
4

Summarizing:

		Unit #			
		1	2	3	4
Rand. #	1	1	0	0	0
	2	0	1	0	0
	3	0	0	1	0
	4	0	0	0	1
Design Z_i					

$\longrightarrow$

		Unit #			
		1	2	3	4
Rand. #	1	Di	In	Co	Co
	2	In	Di	In	Co
	3	Co	In	Di	In
	4	Co	Co	In	Di
Exposure D_i					

We can figure out the exact probabilities that each of the four units would be in each of the exposure conditions:

		Unit #			
		1	2	3	4
Rand. #	1	Di	In	Co	Co
	2	In	Di	In	Co
	3	Co	In	Di	In
	4	Co	Co	In	Di
Exposure D_i					

	Unit #			
	1	2	3	4
Direct	0.25	0.25	0.25	0.25
Indirect	0.25	0.50	0.50	0.25
Control	0.50	0.25	0.25	0.50
Probabilties $\pi_i(D_i)$				

Let's make up some potential outcomes associated with each exposure:

	Unit #				
	1	2	3	4	Mean
Direct	5	10	10	3	7
Indirect	0	3	3	2	2
Control	1	3	6	2	3
Potential outcomes $Y_i(D_i)$					

- Average causal effect: $\tau(d_k, d_l) = \frac{1}{N} \sum_{i=1}^{N} [Y_i(d_k) - Y_i(d_l)]$.
- E.g., $\tau(Direct, Control) = \frac{1}{N} \sum_{i=1}^{N} [Y_i(Direct) - Y_i(Control)] = 4$.

Unequal probability design provides a natural and design-unbiased estimator.

- Assuming $\pi_i(d_k) > 0$ and $\pi_i(d_l) > 0$, the Horvitz-Thompson (HT) estimator:

$$\hat{\tau}_{HT}(d_k, d_l) = \frac{1}{N} \sum_{i=1}^{N} \left[\frac{\mathbf{I}(D_i = d_k)}{\pi_i(d_k)} Y_i(d_k) - \frac{\mathbf{I}(D_i = d_l)}{\pi_i(d_l)} Y_i(d_l) \right]$$

- Unbiasedness follows from $\mathrm{E}\left[\mathbf{I}(D_i = d_k)\right] = \pi_i(d_k)$.
- Note: when, for some i, $\pi_i(d_k) = 0$ or $\pi_i(d_j) = 0$, $\tau(d_k, d_l)$ must be estimated only for units with some probability of receiving both exposures.

Applying estimators to this setup:

		Diff. in Means		OLS w/ cov. adj.		$\widehat{\tau_{HT}}(d_k, d_l)$	
Rand. #	1	1.00	-1.00	3.00	-3.00	-2.00	-5.50
	2	8.00	-0.50	5.00	-2.00	9.00	0.50
	3	9.00	1.50	8.00	1.00	9.50	3.00
	4	1.00	1.00	2.00	-5.44	-0.50	-2.00
	E[.]	4.75	0.25	4.50	-1.00	4.00	-1.00
	Bias	0.75	1.25	0.50	0.00	0.00	0.00
		$\tau(Di, Co)$	$\tau(In, Co)$	$\tau(Di, Co)$	$\tau(In, Co)$	$\tau(Di, Co)$	$\tau(In, Co)$

- Other approaches are biased and inconsistent (i.e., this is not just a small sample problem).
- Bias can go any number of ways depending on nature of confounding and effect heterogeneity.
- Another crucial point is that the *variance* of HT estimator is straightforward. We cannot rely on standard methods to compute standard errors or confidence intervals:

Exact variance:

$$\mathrm{Var}\left(\widehat{\tau_{HT}}(d_k, d_l)\right) = \frac{1}{N^2}\Big\{\mathrm{Var}\left[\widehat{Y_{HT}^T}(d_k)\right] + \mathrm{Var}\left[\widehat{Y_{HT}^T}(d_l)\right] - 2\mathrm{Cov}\left[\widehat{Y_{HT}^T}(d_k), \widehat{Y_{HT}^T}(d_l)\right]\Big\},$$

where

$$\mathrm{Var}\left[\widehat{Y_{HT}^T}(d_k)\right] = \sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{Cov}\left[\mathbf{I}(D_i = d_k), \mathbf{I}(D_j = d_k)\right]\frac{Y_i(d_k)}{\pi_i(d_k)}\frac{Y_j(d_k)}{\pi_j(d_k)}$$

$$\mathrm{Cov}\left[\widehat{Y_{HT}^T}(d_k), \widehat{Y_{HT}^T}(d_l)\right] = \sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{Cov}\left[\mathbf{I}(D_i = d_k), \mathbf{I}(D_j = d_l)\right]\frac{Y_i(d_k)}{\pi_i(d_k)}\frac{Y_j(d_l)}{\pi_j(d_l)}$$

Conservative variance estimator:

Via Young's inequality (c.f., Aronow and Samii 2012), given $\pi_{ij}(d_k, d_l) > 0, \forall i \neq j$,

$$\begin{aligned}
\widehat{\mathrm{Var}}\left[\widehat{\tau_{HT}}(d_k, d_l)\right] = \frac{1}{N^2} \Bigg\{ & \sum_{i \in U} \mathbf{I}(D_i = d_k)[1 - \pi_i(d_k)] \left[\frac{Y_i(d_k)}{\pi_i(d_k)}\right]^2 \\
& + \sum_{i \in U} \sum_{j \in U \setminus i} \mathbf{I}(D_i = d_k)\mathbf{I}(D_j = d_k) \frac{\pi_{ij}(d_k) - \pi_i(d_k)\pi_j(d_k)}{\pi_{ij}(d_k)} \frac{Y_i(d_k)}{\pi_i(d_k)} \frac{Y_j(d_k)}{\pi_j(d_k)} \Bigg\} \widehat{\mathrm{Var}}\left[\widehat{\mu}_{HT}(d_l)\right] \\
& + \sum_{i \in U} \mathbf{I}(D_i = d_l)[1 - \pi_i(d_l)] \left[\frac{Y_i(d_l)}{\pi_i(d_l)}\right]^2 \\
& + \sum_{i \in U} \sum_{j \in U \setminus i} \mathbf{I}(D_i = d_l)\mathbf{I}(D_j = d_l) \frac{\pi_{ij}(d_l) - \pi_i(d_l)\pi_j(d_l)}{\pi_{ij}(d_l)} \frac{Y_i(d_l)}{\pi_i(d_l)} \frac{Y_j(d_l)}{\pi_j(d_l)} \Bigg\} \widehat{\mathrm{Var}}\left[\widehat{\mu}_{HT}(d_k)\right] \\
& - 2 \sum_{i \in U} \sum_{j \in U \setminus i} \frac{\mathbf{I}(D_i = d_k)\mathbf{I}(D_j = d_l)}{\pi_{ij}(d_k, d_l)} \frac{Y_i(d_k)}{\pi_i(d_k)} \frac{Y_j(d_l)}{\pi_j(d_l)} \\
& + 2 \sum_{i \in U} \left[\frac{\mathbf{I}(D_i = d_k) Y_i(d_k)^2}{2\pi_i(d_k)} + \frac{\mathbf{I}(D_i = d_l) Y_i(d_l)^2}{2\pi_i(d_l)}\right] \Bigg\} \Bigg\} - 2\widehat{\mathrm{Cov}}_C\left[\widehat{\mu}_{HT}(d_l), \widehat{\mu}_{HT}(d_k)\right].
\end{aligned}$$

Unbiased under sharp null hypothesis of no effect, given $\pi_{ij}(d_k, d_l) > 0$.
(More) conservative variance estimator when $\exists i, j, k, l$ s.t. $\pi_{ij}(d_k, d_l) = 0$.

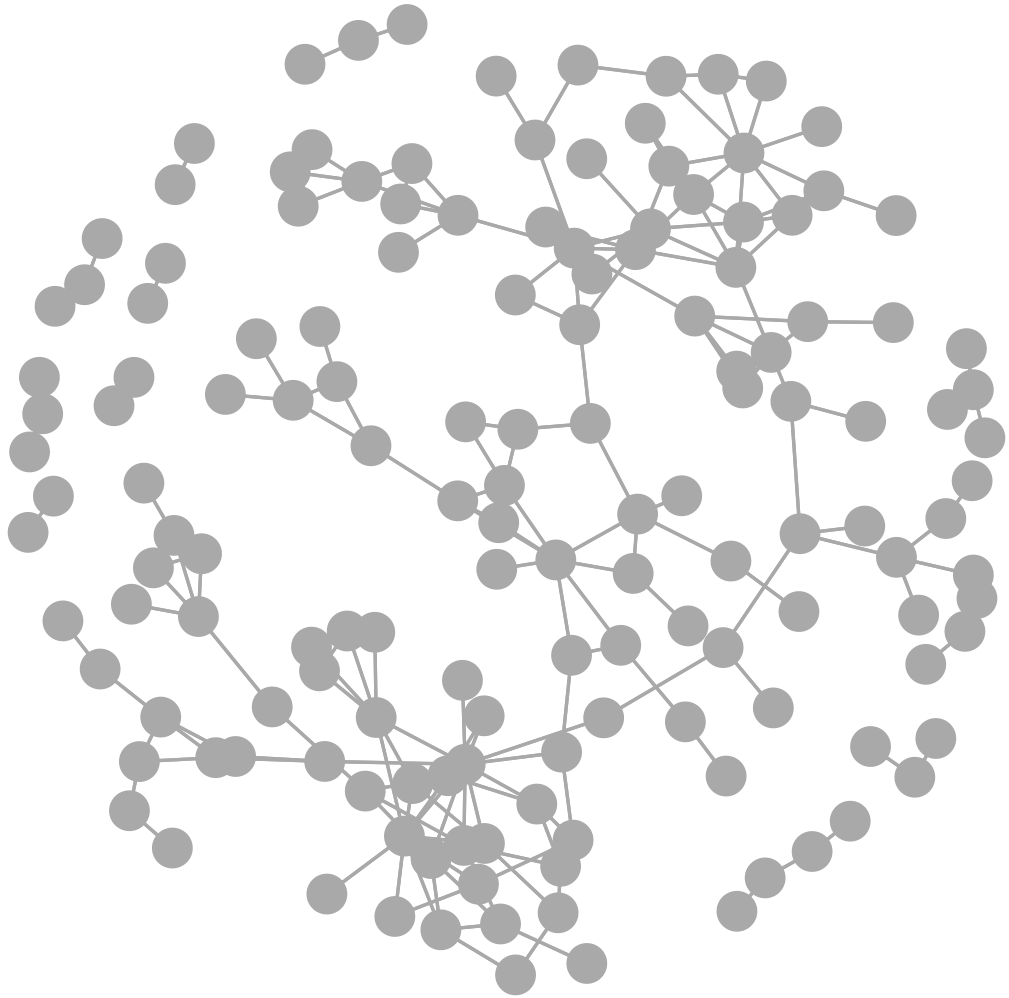
Asymptotics and intervals:

- We adopt Brewer (1979)'s large sample scaling, analogous to obtaining estimates by aggregating results from repeated experimentation on a fixed finite population.
- Consistency and asymptotic normality of $\widehat{\tau_{HT}}(d_k, d_l)$ follow from the WLLN and classical CLT respectively. By the WLLN, $N\widehat{\mathrm{Var}}\left[\widehat{\tau_{HT}}(d_k, d_l)\right] \xrightarrow{p} N\mathrm{Var}\left[\widehat{\tau_{HT}}(d_k, d_l)\right] + c_1$, where $c_1 \geq 0$. Then $\left(\widehat{\tau_{HT}}(d_k, d_l) - \tau_{HT}(d_k, d_l)\right) / \sqrt{\widehat{\mathrm{Var}}\left[\widehat{\tau_{HT}}(d_k, d_l)\right]} \xrightarrow{d} \mathrm{N}\left(0, 1 - c_2\right)$, where $0 \leq c_2 < 1$. Intervals constructed as $\widehat{\tau_{HT}}(d_k, d_l) \pm z_{1-\alpha/2}\sqrt{\widehat{\mathrm{Var}}\left[\widehat{\tau_{HT}}(d_k, d_l)\right]}$ will asymptotically cover $\tau_{HT}(d_k, d_l)$ at least $100(1-\alpha)\%$ of the time.
- We've also proven consistency of estimators and variance under a generalized m-dependence set-up. Restrictions on clustering are key.

- Paper proposes refinements for covariate adjustment, weight stabilization, and variance approximation under a constant effect assumption.
- Further refinements include modeling outcomes based on determinants of exposure probabilities, using HT results to determine appropriate variance approximation.
- Regardless of the method used, the implied inverse probability weights are *fundamental* for the consistency of any estimator of average causal effects.
- Under proper specification, this weighting can be reproduced by regression estimators (in particular, interaction with centered fixed effects for all unique values of probability of exposure) in the limit.

- Let's consider a richer example.
- Goal is to estimate direct and indirect effects of a treatment offered to a randomly selected set of individuals on a complex, undirected network (e.g., an anti-prejudice curriculum in schools – Paluck and Shepherd 2012)

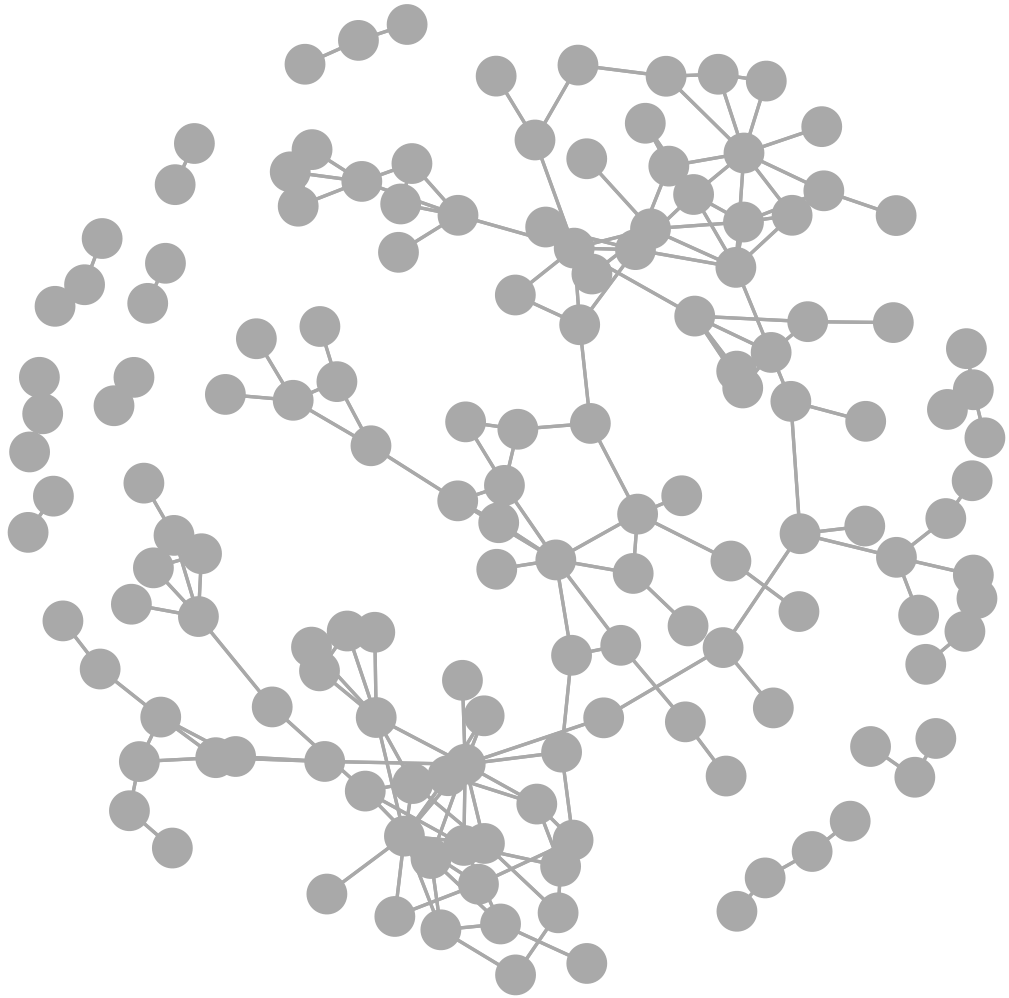
Network



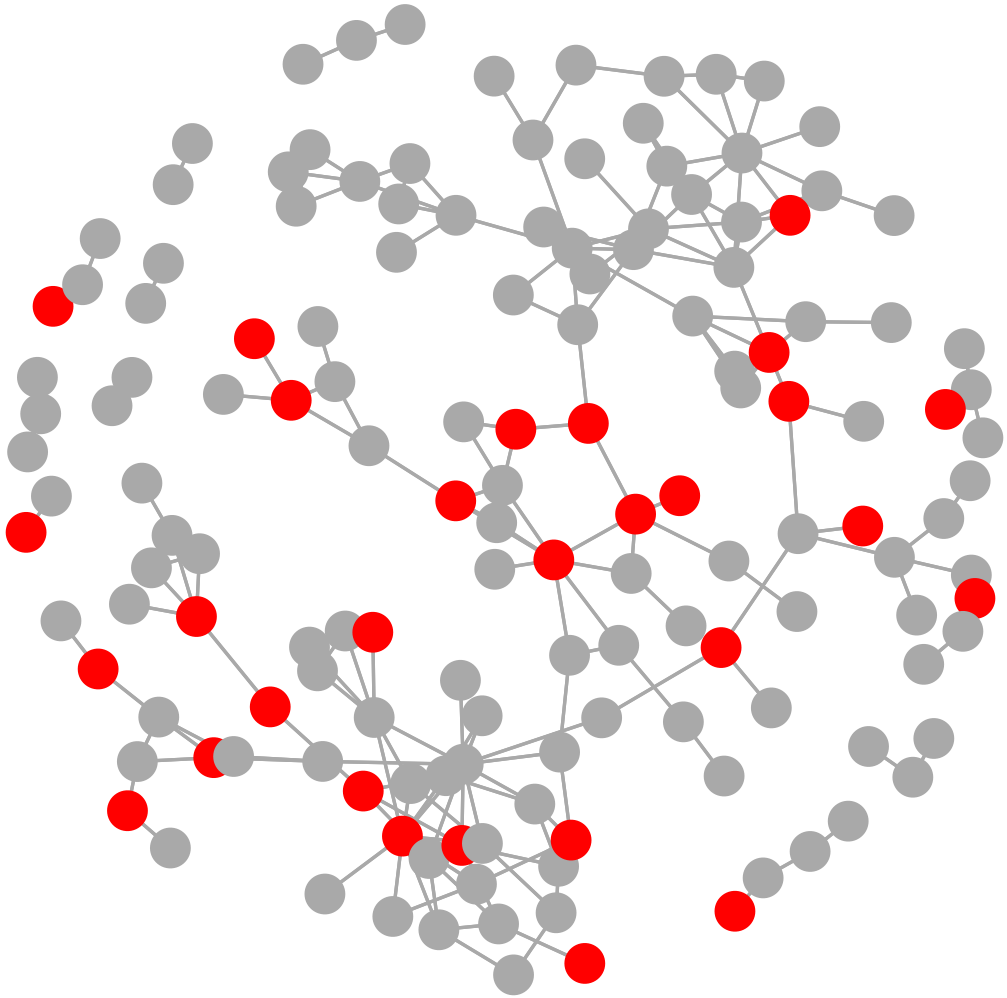
- Suppose complete random assignment of $M = .2N$ units to treatment.
- Design implies $\mathbf{Z}$ has uniform probability over Ω, an $N \times \binom{N}{M}$ indicator matrix, where $\mathbf{z}$ is a realization a $\mathbf{Z}$, e.g.,

$$\mathbf{z} = (z_1, z_2, z_3, ..., z_{N-1}, z_N)' = (0, 1, 0, ..., 1, 0)'.$$

Network



Treatment Assignment



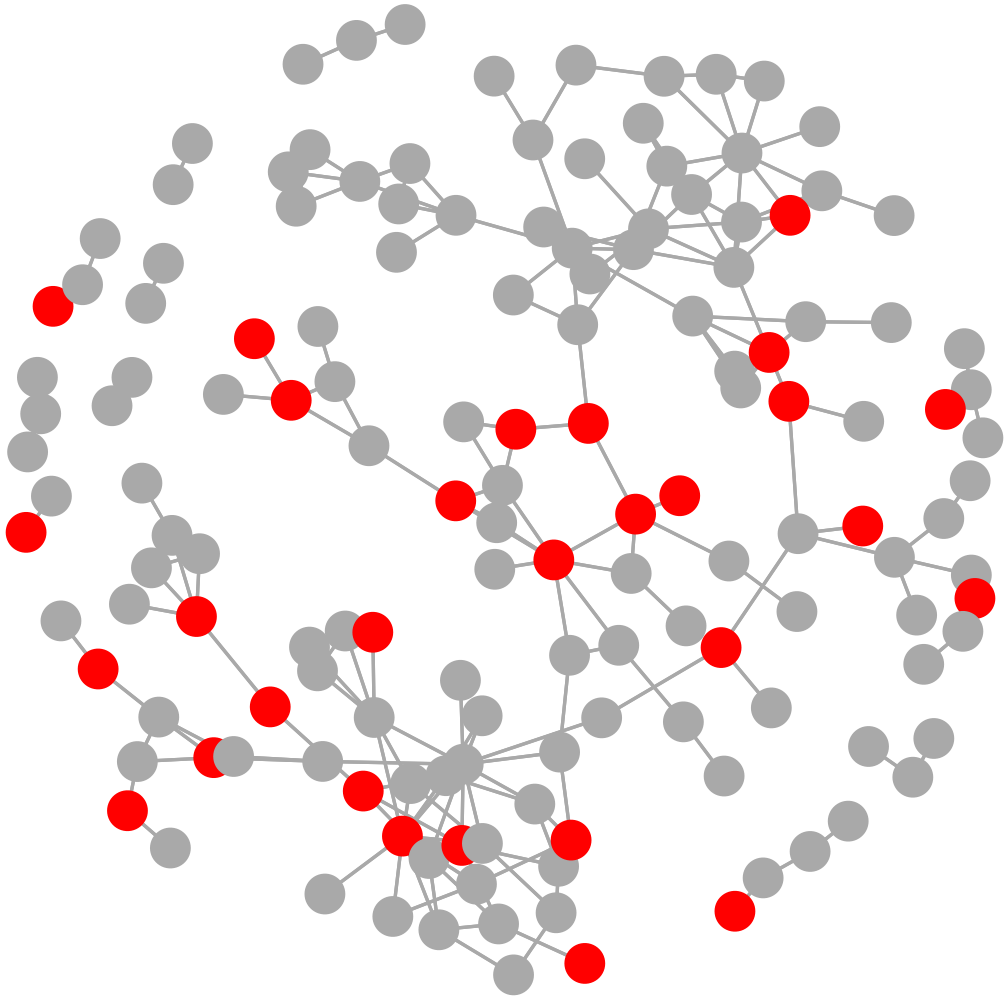
- Let θ_i be i's row in the adjacency matrix (with diagonal zeroed out):

```
> adj
   1 2 5 8 9 11 13 14 15 16 17 18 19 21 22 23 25
1  0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  1
2  0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0
5  0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0
8  0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0
9  0 0 0 0 0  0  0  0  0  0  0  0  1  0  0  0  0
11 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0
13 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0
14 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0
15 0 0 0 0 0  0  0  0  0  0  0  0  0  1  0  0
16 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0
17 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0
18 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0
19 0 0 0 0 1  0  0  0  0  0  0  0  0  0  0  0  0
21 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0
```

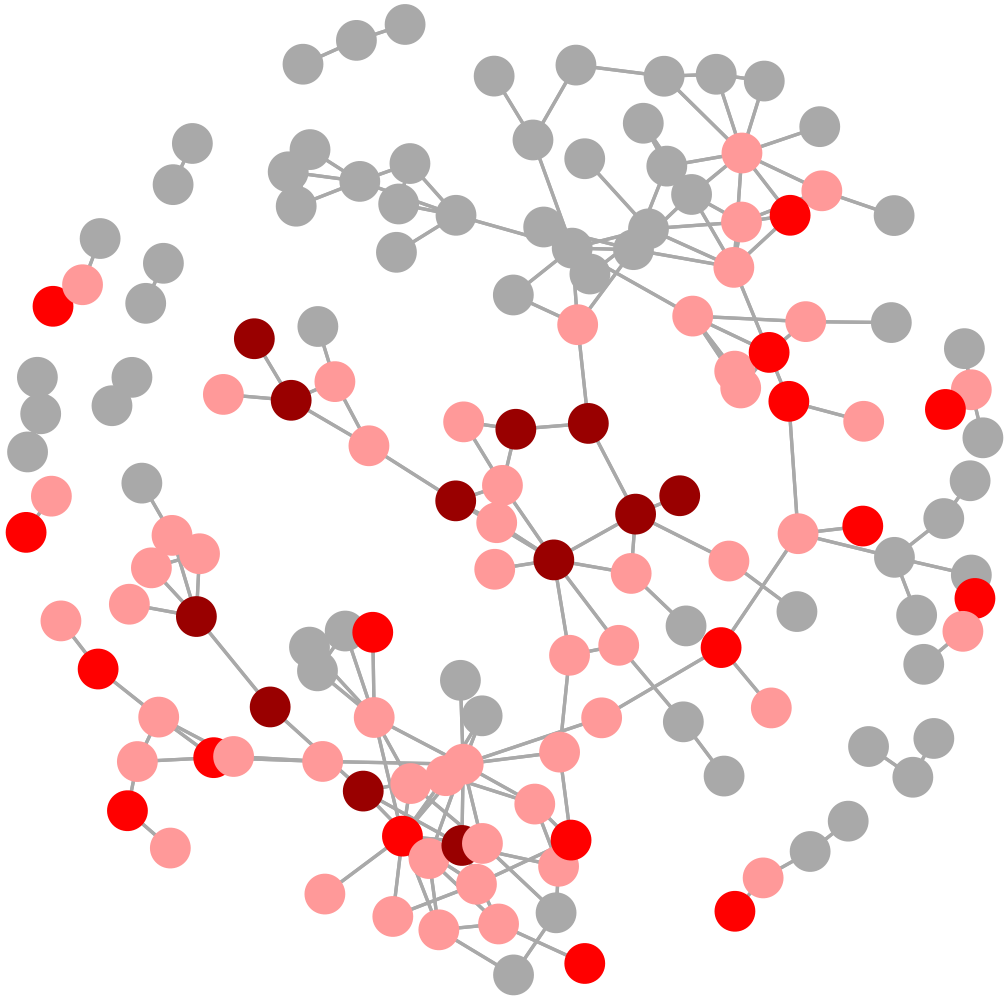
Define an exposure model corresponding to our substantive interests:

$$f(\mathbf{z}, \theta_i) = \begin{pmatrix} z_i(\mathbf{I}(\mathbf{z}'\theta_i = 0)) \\ (1 - z_i)\mathbf{I}(\mathbf{z}'\theta_i > 0) \\ z_i\mathbf{I}(\mathbf{z}'\theta_i > 0) \\ (1 - z_i)\mathbf{I}(\mathbf{z}'\theta_i = 0) \end{pmatrix} = \begin{pmatrix} \text{Isolated Direct} \\ \text{Indirect} \\ \text{Direct \& Indirect} \\ \text{Control} \end{pmatrix},$$

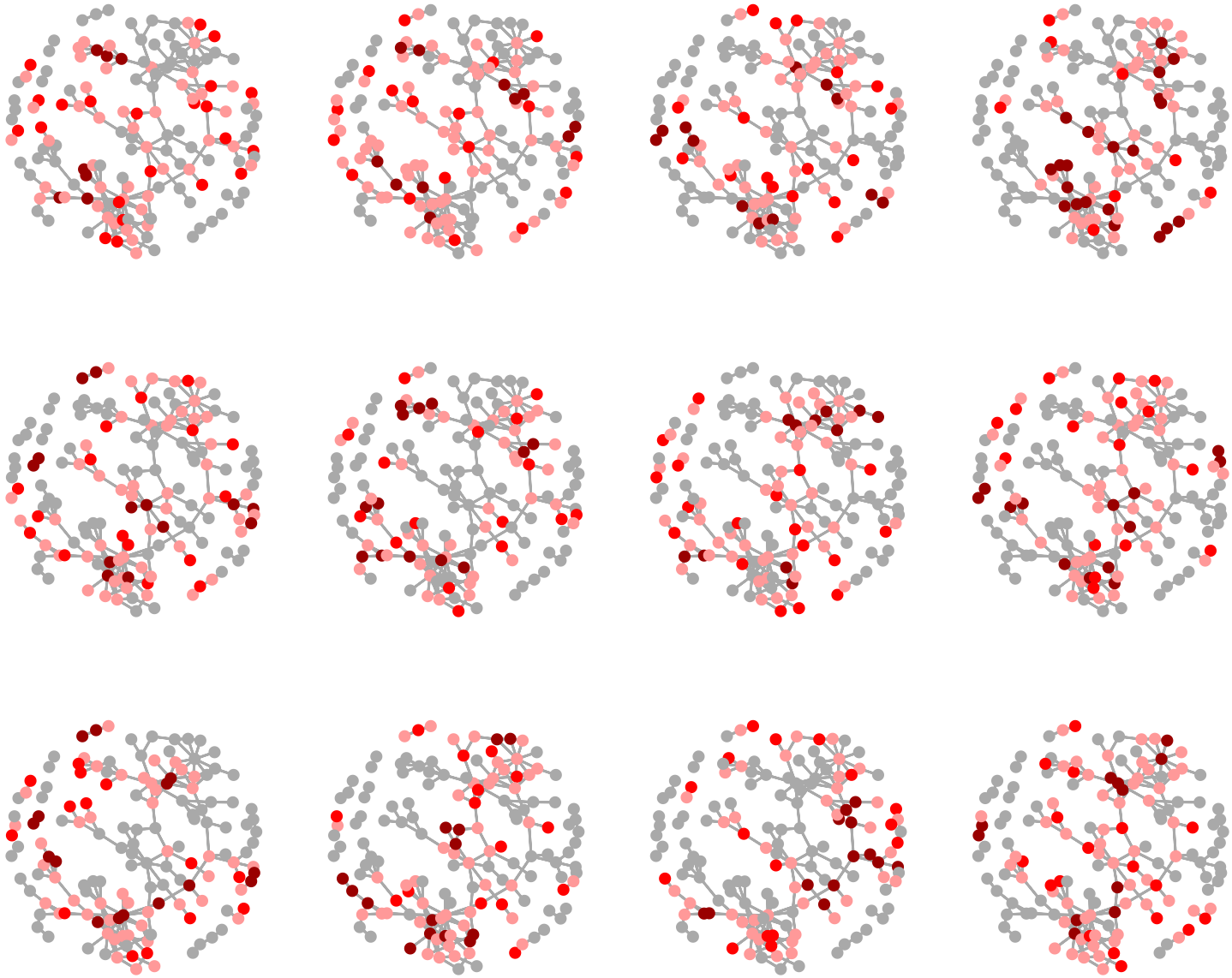
Treatment Assignment



Exposure Conditions



And all possible randomizations...



This yields a matrix of indicators for exposure k associated with each randomization $\mathbf{z}$:

$$\mathbf{I}_k = [\mathbf{I}(f(\mathbf{z}, \theta_i) = d_k)]_{\substack{\mathbf{z}\in\Omega \\ i=1,\ldots,N}} =$$

$$\begin{bmatrix} \mathbf{I}(f(\mathbf{z}_1, \theta_1) = d_k) & \mathbf{I}(f(\mathbf{z}_2, \theta_1) = d_k) & \ldots & \mathbf{I}(f(\mathbf{z}_N, \theta_1) = d_k) \\ \mathbf{I}(f(\mathbf{z}_1, \theta_2) = d_k) & \mathbf{I}(f(\mathbf{z}_2, \theta_2) = d_k) & \ldots & \mathbf{I}(f(\mathbf{z}_N, \theta_2) = d_k) \\ \vdots & \vdots & \ddots & \\ \mathbf{I}(f(\mathbf{z}_1, \theta_N) = d_k) & \mathbf{I}(f(\mathbf{z}_2, \theta_N) = d_k) & & \mathbf{I}(f(\mathbf{z}_N, \theta_N) = d_k) \end{bmatrix}.$$

Then for exposure k, first and second-order exposure probabilities are,

$$\frac{\mathbf{I}_k \mathbf{I}_k'}{|\Omega|} = \begin{bmatrix} \pi_1(d_k) & \pi_{12}(d_k) & \ldots & \pi_{N1}(d_k) \\ \pi_{12}(d_k) & \pi_2(d_k) & \ldots & \pi_{N2}(d_k) \\ \vdots & \vdots & \ddots & \\ \pi_{N1}(d_k) & \pi_{N2}(d_k) & & \pi_N(d_k) \end{bmatrix},$$

Cross exposure probabilities computed analogously.

A real application along these lines: data snippet courtesy of Paluck and Shepherd (2012)

Exposure	Naive (Diff-in-Means)	Cov. Adj. (Fixed Effects)	HT (Proposed)
Direct	-0.775	-0.752	-1.400
(SE)	(0.793)	(0.927)	(1.133)
Indirect	-0.382	-0.648	-0.607
(SE)	(0.434)	(0.596)	(1.106)
Combined	-1.331	-1.663	-1.792
(SE)	(0.956)	(1.220)	(1.617)

Anticipating some concerns.

$$f(\mathbf{Z}, \theta_i)$$

Concern: "What if you don't believe the exposure model?!"

- We *always* specify an exposure model to define causal effects.
- But! The framework permits exposure models of arbitrary generality.
- By definition, there is a finite (but potentially large) set of exposure models that may be associated with any randomization scheme.
- These models can be nested.

Concern: "What if you don't really know θ?!"

- We can model the θ and then use available data to estimate a probability distribution over θ's.
- Then, we can marginalize conditional estimates.

$$\int_{\Phi} \tau \begin{pmatrix} f(\mathbf{Z}, \theta_1(\phi)) \\ \vdots \\ f(\mathbf{Z}, \theta_N(\phi)) \end{pmatrix} dF(\phi)$$

- E.g., graph models can use covariate data to predict possible adjacency matrices. Impute 1,000 possible adjacency matrices (ϕ) based on $F(\phi)$, estimate causal effects on each (τ), and then average.

Some other thoughts / extensions:

- Design implications?
  - Basic results from survey sampling suggest minimizing variation in exposure probability vectors.
  - Variance expression suggests limiting clustering in exposures.
  - Possible to construct maximum entropy designs or minimum risk designs given bounded potential outcomes – we are currently at work on this ("solved" via brute-force optimization, but...)
- Observational studies?
  - If we can *estimate* the treatment assignment mechanism, then simple enough to specify an exposure model again.

Thank you!

You can find our paper on my website:
http://j.mp/paronow