Toward Evaluation of Dissemination of Non-Randomized HIV Prevention Interventions Among Observed Networks of People who Inject Drugs

Ashley Buchanan, DrPH, MS

Assistant Professor, Biostatistics
College of Pharmacy
University of Rhode Island

TRIPODS *Models and Machine Learning for Causal Inference and Decision Making in Health Research* 2019

Outline

1. Introduction
2. Evaluation of Disseminated Effects in Networks
3. Future Directions and Collaborations

1 Introduction

2 Evaluation of Disseminated Effects in Networks

3 Future Directions and Collaborations

Definition

Implementation Science

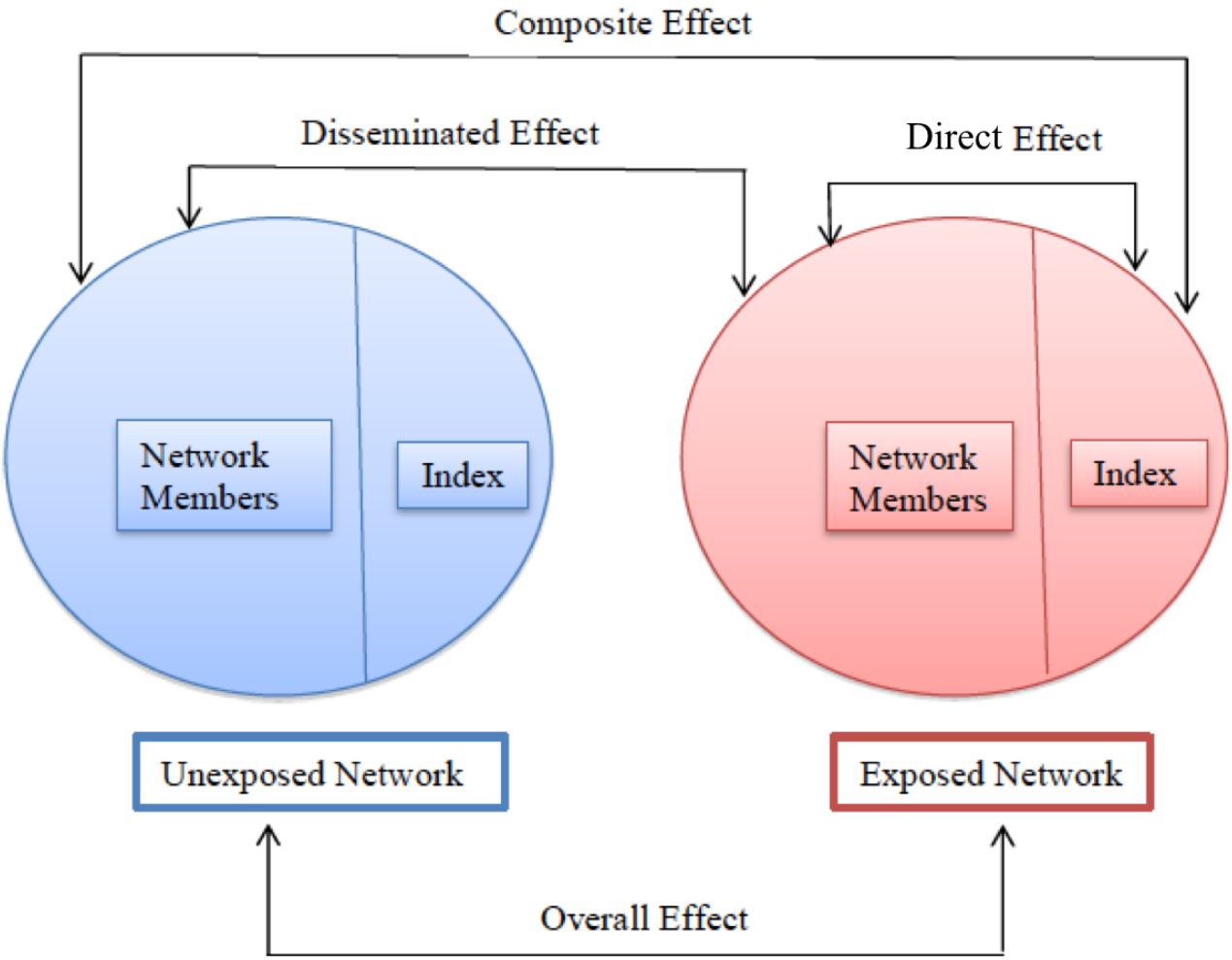
Translation and scale-up of research evidence into practice (Madon et al., 2007; Padian et al., 2011)

- Natural clustering by social network or community
- Biological and social influence in networks
- Inform public health practice and policy

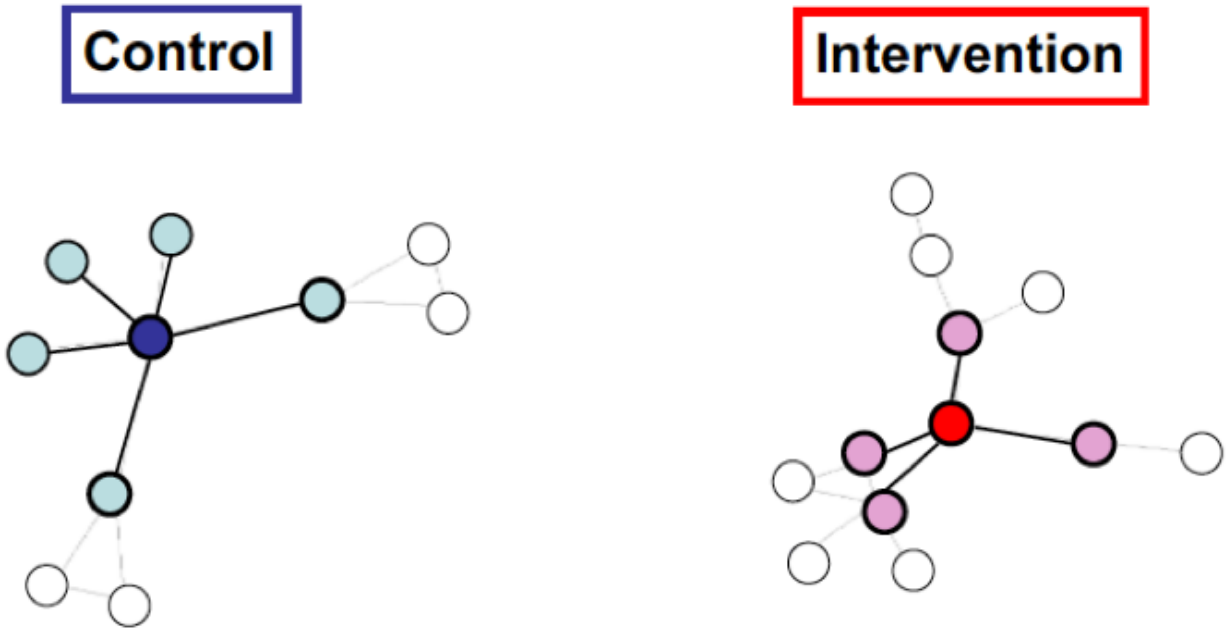
Causal Inference

- A Potential Outcome (i.e., Counterfactual)
  - $Y(0)$: Response that would have been seen if (possibly contrary to fact) the participant were not exposed
  - $Y(1)$: Response that would have been seen if (possibly contrary to fact) the participant were exposed
- Assumptions: Consistency, No Interference, Positivity, Exchangeability (Cole and Frangakis, 2009; Rubin, 1980)
- *With associations, we can predict the future. With causation, we can change the future.*

Cluster-Randomized Trial



Network-Based Study



- Connections: Shared HIV risk (injection or sexual)
- Index shaded blue or red nodes
- Nearest neighbors outlined nodes

1 Introduction

2 Evaluation of Disseminated Effects in Networks

3 Future Directions and Collaborations

Study Motivation

- Injection drug use increases HIV risk through sharing equipment (e.g. syringes, needles, etc.) and often correlates with risky sexual behaviors
- PWIDs are not only at high risk of HIV infection, but also face unique barriers along the HIV treatment cascade (Ghosh et al., 2017)

Primary Motivation

PWIDs are embedded in HIV/AIDS risk network and such network structure can support and sustain **positive behavioral change** via interventions that leverage network structure.

Motivating Example

The Social Factors and HIV Risk Study (SFHR)

- Sociometric network study conducted between 1991 and 1993 in Bushwick, Brooklyn, New York among *street-recruited injection drug users*
- Investigated how HIV/AIDS infection spread through shared sexual and injection risk behaviors.
- 767 participants along with 3,162 dyadic relationships (i.e. a connection b/w two people).
- Connections were shared **risk behaviors** (i.e. inject drug together and/or having sexual intercourse) within 30 days before the interview.

full network

Goal of the Study

- Assess attitudes toward HIV/AIDS risk among PWIDs and their effect on health-seeking behaviors.
- Employ **causal inference methods** in the presence of *dissemination or interference* in an observational study.
- Exposures were (1) **locus of control** and (2) **blame attribute** of each participants.
- Outcomes were (1) **receipt of study-based HIV testing result** and (2) **a recent medical visit** within the past year.

SFHR PWIDs Network for Analysis

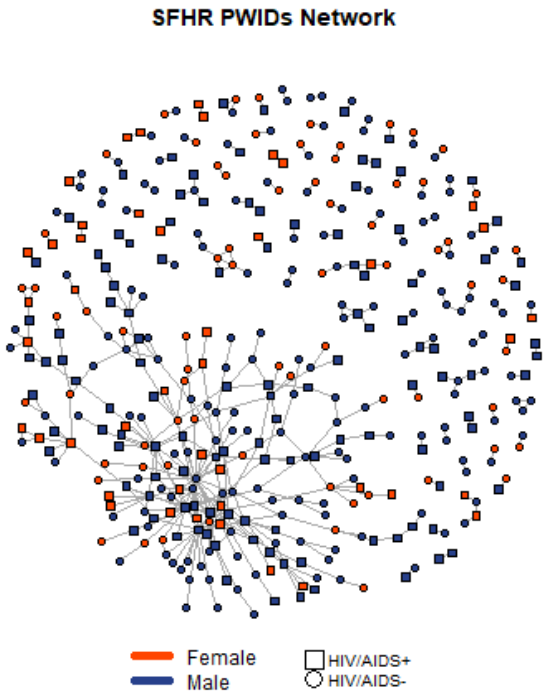


Figure 1: SFHR PWIDs Network with 402 vertices and 403 edges.

A network (or graph) G is defined as a collection of vertices (or nodes) (V) and edges (or links)(E), $G = (V, E)$. Here, G is the SFHR PWIDs network.

SFHR PWIDs Network for Analysis

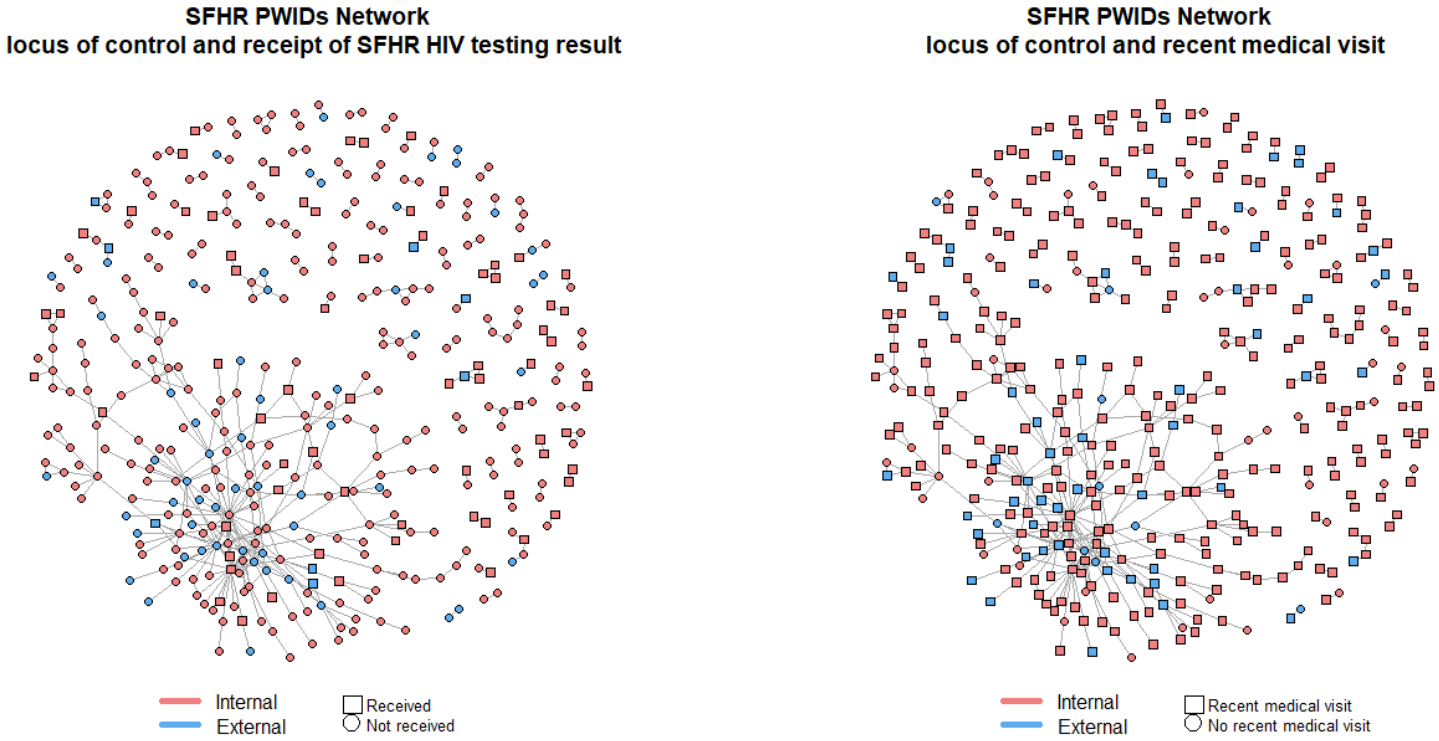


Figure 2: SFHR PWIDs Network: Locus of control (internal = red, external = blue) and receipt of HIV testing result (left) and recent medical visit (right).

Causal Parameters

$$\textit{Direct}: \quad \overline{DE}(\alpha) = \overline{Y}(a=0;\alpha) - \overline{Y}(a=1;\alpha)$$

$$\textit{Disseminated/Indirect}: \quad \overline{IE}(\alpha,\alpha') = \overline{Y}(a=0;\alpha) - \overline{Y}(a=0;\alpha')$$

$$\textit{Composite/Total}: \quad \overline{TE}(\alpha,\alpha') = \overline{Y}(a=0;\alpha) - \overline{Y}(a=1;\alpha')$$

$$\textit{Overall}: \quad \overline{OE}(\alpha,\alpha') = \overline{Y}(\alpha) - \overline{Y}(\alpha')$$

coverage: $\alpha < \alpha'$

Causal Inference with Observational Study

- **Inverse probability weighting (IPW)** method to replicate two-stage randomized design with observational study (Tchetgen and VanderWeele, 2012)

APPROACH

1. Split the SFHR network into smaller subnetworks/ communities of PWIDs.
2. Calculate group-level propensity score (i.e. probability of having internal/external HIV locus of control) for each subnetwork based on individual-level covariates of sex, race, education, age and their pairwise interactions.
3. Use the inverse of propensity score as weights to compute IPW estimators for contrasts of potential outcomes.

Assumptions

(1) **Partial interference**: Allow interference within a subnetwork/community, but not between subnetworks.

(2) **Stratified interference**: Individual's potential outcome is dependent only on own exposure and the proportion of exposed in the community.

(3) **Bernoulli individual group allocation strategy**: The distribution of treatment A is assumed to be a Bernoulli distribution and used to define the potential outcomes Y^a.

(4) **No homophily**: Assume individuals in the network are not connected based on some unobservable variables associated with their health-seeking behaviors.

(5) **Well-defined interventions**: Locus of control is well defined and there is no other version of locus of control that would affect the potential outcomes.

(6) **Positivity**: Probability of exposure is positive given each level of covariates.

(7) **Conditional exchangeability**: Assume that conditioning on a set of pre-exposure covariates is sufficient to control confounding.

Community Detection

- **Community**: A set of vertices densely connected, with only sparser tie to vertices that belong to other groups or communities.
- **Hierarchical clustering**: Common methods for community detection where the closest or most similar vertices are combined to form communities with *a measure of similarity* or *connection strength* between vertices based on the network structure.
- As the measure of similarity, we use **modularity** (Kolaczyk, 2009; Newman, 2006)

modularity

IPW estimation: Group-Level Propensity Score

Group-level propensity score can be calculated by adjusting with individual level covariates who belong in the community.

$$f_{A_i|X_i}(A_i|X_i;\theta_x,\theta_s) = \int \prod_{j=1}^{n_i} h_{ij}(b_i;\theta_x)^{A_{ij}}\{1-h_{ij}(b_i;\theta_x)\}^{1-A_{ij}} f_b(b_i;\theta_s)db_i$$

where
$h_{ij}(b_i;\theta_x) = Pr(a_{ij}=1|X_{ij}, b-i, \theta_x) = logit^{-1}(X_{ij}\theta_x + b_i)$ is a propensity score for individual j in community i and $f_b(\cdot;\theta_s)$ is the density of community specific random effect $b_i \sim N(0,\theta_s)$.

IPW estimation: IPW estimator

IPW estimator for group-level potential outcome:

$$\hat{Y}_i^{ipw}(a,\alpha) = \frac{\sum_{j=1}^{n_i} \pi_i(A_{i,-j};\alpha) I(A_{ij}=\alpha) Y_{ij}}{n_i f_{A_i|X_i}(A_i|X_i;\hat{\theta})}$$

Marginal potential outcome:

$$\hat{Y}_i^{ipw}(\alpha) = \frac{\sum_{j=1}^{n_i} \pi_i(A_i;\alpha) I(A_i=\alpha) Y_{ij}}{n_i f_{A_i|X_i}(A_i|X_i;\hat{\theta})}$$

Population-level IPW estimators

$$\widehat{DE}(\alpha) = \hat{Y}^{ipw}(a = 0; \alpha) - \hat{Y}^{ipw}(a = 1; \alpha)$$

$$\widehat{IE}(\alpha, \alpha') = \hat{Y}^{ipw}(a = 0; \alpha) - \hat{Y}^{ipw}(a = 0; \alpha')$$

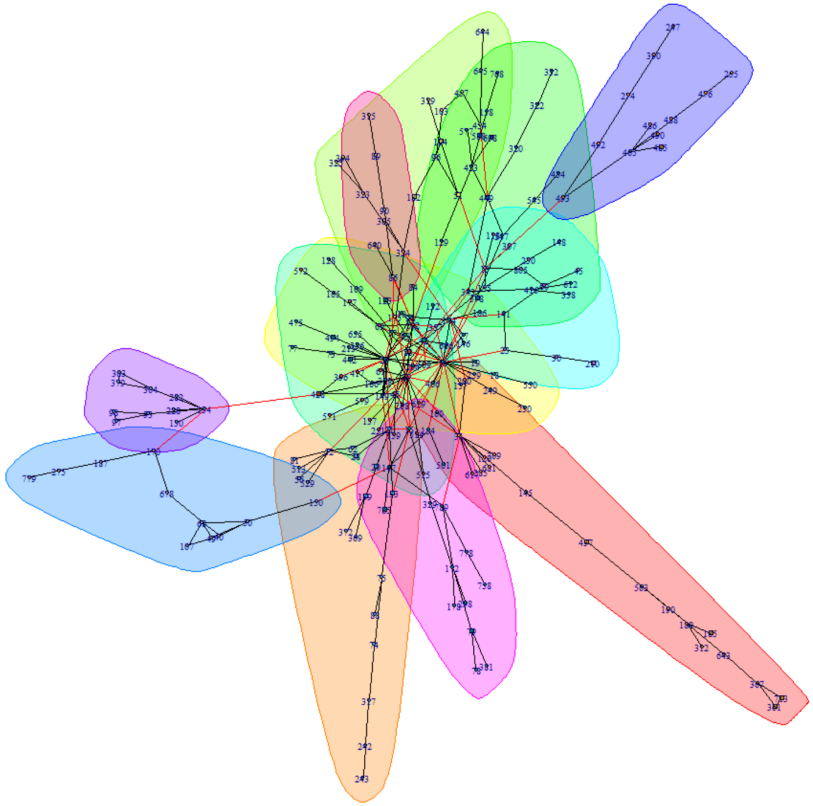
$$\widehat{TE}(\alpha, \alpha') = \hat{Y}^{ipw}(a = 0; \alpha) - \hat{Y}^{ipw}(a = 1; \alpha')$$

$$\widehat{OE}(\alpha, \alpha') = \hat{Y}^{ipw}(\alpha) - \hat{Y}^{ipw}(\alpha')$$

coverage: $\alpha < \alpha'$

Modularity-Based Community Detection

85 connected components and one is GC with 199 participants. 12 subnetworks based on the community detection against the GC. In total, **96 communities** in the SFHR network for analysis.



Descriptive Statistics

Table 1: Individual-level Associations between Locus of Control and Health-Seeking Behaviors in SFHR

	Odds Ratio (95% CI)		
	Giant	Not Giant	Total
Received HIV Test (Yes vs. Not) Internal vs. External	1.94 (0.55, 6.85)	1.64 (0.59, 4.57)	1.87 (0.85, 4.11)
Medical visit (Yes vs. No) Internal vs. External	1.32 (0.49, 3.56)	0.88 (0.28, 2.72)	1.07 (0.51, 2.24)

Causal Inference: IPW estimation - Model 1

Table 2: Estimated risk differences (RDs) with 95% CIs for the effects of **locus of control** (external vs. internal) on the likelihood of **receiving HIV test results** in SFHR (α = coverage of internal)

		Unadjusted		Adjusted with interactions	
Effect	Coverage (α, α')	RD	95% CI	RD	95% CI
Direct	(50%, 50%)	**-0.148**	(-0.230, -0.065)	**-0.160**	(-0.265, -0.055)
Direct	(70%, 70%)	**-0.142**	(-0.246, -0.038)	**-0.162**	(-0.268, -0.055)
Direct	(99%, 99%)	-0.101	(-0.258, 0.056)	-0.130	(-0.268, 0.008)
Indirect	(50%, 70%)	**-0.041**	(-0.071, -0.012)	**-0.031**	(-0.054, -0.008)
Indirect	(50%, 99%)	-0.070	(-0.156, 0.019)	-0.062	(-0.123, 0.000)
Indirect	(70%, 99%)	-0.029	(-0.098, 0.040)	-0.030	(-0.072, 0.011)
Total	(50%, 70%)	**-0.183**	(-0.271, -0.096)	**-0.193**	(-0.286, -0.100)
Total	(50%, 99%)	**-0.172**	(-0.278, -0.066)	**-0.192**	(-0.291, -0.093)
Total	(70%, 99%)	**-0.130**	(-0.254, -0.006)	**-0.161**	(-0.272, -0.049)
Overall	(50%, 70%)	**-0.067**	(-0.096, -0.038)	**-0.065**	(-0.089, -0.041)
Overall	(50%, 99%)	**-0.097**	(-0.183, -0.010)	**-0.111**	(-0.190, -0.032)
Overall	(70%, 99%)	-0.030	(-0.095, 0.035)	-0.046	(-0.105, 0.013)

Predictive Analysis: IPW estimation - Model 2

Table 3: Estimated risk differences (RDs) with 95% CIs of the associations of **locus of control** (external vs. internal) with the likelihood of **a recent medical visit** in SFHR (α = coverage of internal)

		Unadjusted		Adjusted with interactions	
Effect	Coverage (α, α')	RD	95% CI	RD	95% CI
Direct	(50%, 50%)	0.211	(-0.286, 0.708)	0.090	(-0.271, 0.451)
Direct	(70%, 70%)	0.003	(-0.296, 0.301)	-0.111	(-0.346, 0.123)
Direct	(99%, 99%)	-0.227	(-0.463, 0.009)	**-0.280**	(-0.470, -0.089)
Indirect	(50%, 70%)	-0.001	(-0.260, 0.257)	-0.008	(-0.181, 0.165)
Indirect	(50%, 99%)	0.208	(-0.298, 0.715)	0.077	(-0.311, 0.464)
Indirect	(70%, 99%)	0.210	(-0.063, 0.482)	0.085	(-0.136, 0.305)
Total	(50%, 70%)	0.001	(-0.537, 0.539)	-0.119	(-0.496, 0.258)
Total	(50%, 99%)	-0.019	(-0.491, 0.453)	-0.203	(-0.559, 0.153)
Total	(70%, 99%)	-0.017	(-0.255, 0.220)	-0.195	(-0.409, 0.018)
Overall	(50%, 70%)	-0.105	(-0.316, 0.106)	-0.131	(-0.265, 0.003)
Overall	(50%, 99%)	-0.122	(-0.358, 0.114)	**-0.246**	(-0.430, -0.061)
Overall	(70%, 99%)	-0.017	(-0.082, 0.048)	**-0.114**	(-0.181, -0.047)

Discussion

- Additional benefit to reporting internal locus beyond being around those who have internal for likelihood of receipt of HIV test result
- Among those with external locus, having more community members with internal increased likelihood of receipt of HIV test
- Protective overall association of internal locus on recent medical visit and additional benefit for those with internal among 99% coverage networks
- Attitudes are likely an important determinant of health-seeking behavior among PWIDs
- Future interventions could consider this influence in the network to increase and sustain impact

1 Introduction

2 Evaluation of Disseminated Effects in Networks

3 Future Directions and Collaborations

Future Research Directions

- More realistic allocation strategies by allowing the correlation in exposure selection strategy (Barkley et al., 2017; Papadogeorgou et al., 2017)
- Allow community detection to depend on important covariates (Huang and Feng, 2018)
- Relax the partial interference assumption (Tchetgen et al., 2017)
- Methods to evaluate homophily (McPherson et al., 2001)
- Alternative estimators that account for the size of the subnetwork (Basse and Feller, 2018)
- **New collaborations to apply these methods to important public health settings**

Acknowledgements

- Thank you to my mentor, Dr. Spiegelman, and collaborators, Ayako Miura, MS, and Drs. Katenka, Friedman, and Kogut, and recent graduates and students, Benjamin Skov, Hilary Aroke, and Tianyu Sun.
- We thank Dr. Friedman for access to the SFHR data. Data from the SFHR study were obtained with support from the National Institute on Drug Abuse (NIDA), under National Institutes of Health (NIH) grant 1R01DA006723.
- This work was supported by NIH Avenir grant 1DP2DA046856-01, Advance Clinical and Translational Research (Advance-CTR) U54GM115677, and the RI Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or RI Foundation.

Job Opportunity at URI

Postdoctoral Fellowship in Biostatistics
NIDA Avenir Project at URI

buchanan@uri.edu

Retrieved from: https://www.narragansettri.gov/

References

Brian G Barkley, Michael G Hudgens, John D Clemens, Mohammad Ali, and Michael E Emch. Causal inference from observational studies with clustered interference. *arXiv preprint arXiv:1711.04834*, 2017.

Guillaume Basse and Avi Feller. Analyzing two-stage experiments in the presence of interference. *Journal of the American Statistical Association*, 113(521):41–55, 2018.

Stephen R Cole and Constantine E Frangakis. The consistency statement in causal inference: A definition or an assumption? *Epidemiology*, 20(1):3–5, 2009.

Debarchana Ghosh, Archana Krishnan, Britton Gibson, Shan-Estelle Brown, Carl A Latkin, and Frederick L Altice. Social network strategies to address hiv prevention and treatment continuum of care among at-risk and hiv-infected substance users: a systematic scoping review. *AIDS and Behavior*, 21(4):1183–1207, 2017.

Sihan Huang and Yang Feng. Pairwise covariates-adjusted block model for community detection. *arXiv preprint arXiv:1807.03469*, 2018.

Eric D Kolaczyk. *Statistical Analysis of Network Data*. Springer, 2009.

Temina Madon, Karen J Hofman, Linda Kupfer, and Roger I Glass. Implementation science. *Science*, 318(5857):1728–1729, 2007.

Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444, 2001.

Mark EJ Newman. Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23):8577–8582, 2006.

Nancy S Padian, Charles B Holmes, Sandra I McCoy, Rob Lyerla, Paul D Bouey, and Eric P Goosby. Implementation science for the US President's Emergency Plan for AIDS Relief (PEPFAR), 2011.

Georgia Papadogeorgou, Fabrizia Mealli, and Corwin Zigler. Causal inference for interfering units with cluster and population level treatment allocation programs. *arXiv preprint arXiv:1711.01280*, 2017.

Donald B Rubin. Randomization analysis of experimental data: The Fisher randomization test comment. *Journal of the American Statistical Association*, 75(371):591–593, 1980.

Eric J Tchetgen Tchetgen and Tyler J VanderWeele. On causal inference in the presence of interference. *Statistical methods in medical research*, 21(1):55–75, 2012.

Eric J Tchetgen Tchetgen, Isabel Fulcher, and Ilya Shpitser. Auto-g-computation of causal effects on a network. *arXiv preprint arXiv:1709.01577*, 2017.

Original PWIDs Network in SFHR

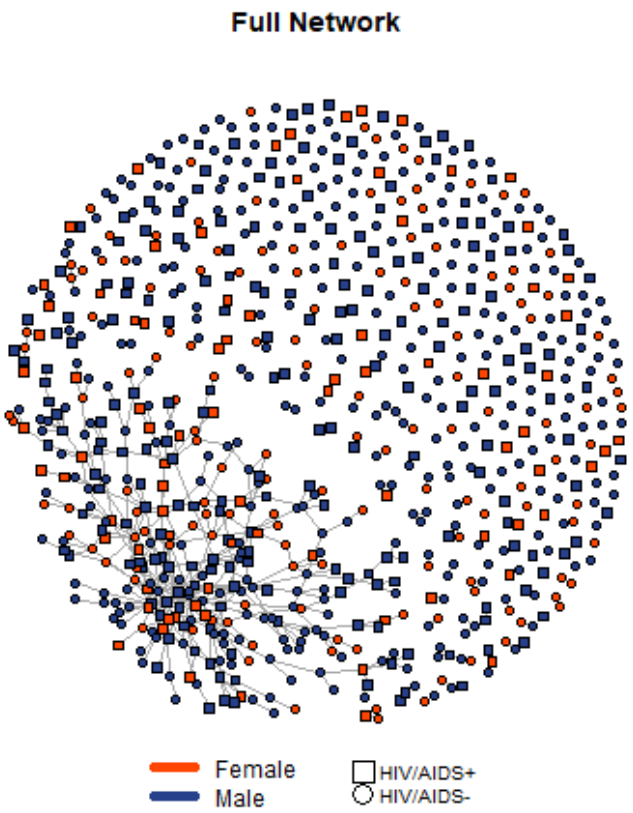


Figure 3: Full Network.

back

Modularity

Modularity is defined as following:

- Assume there are $C = \{C_1, C_2, \cdots, C_K\}$ candidate of K communities in an observed network G.
- We also define $f_{ij} = f_{ij}(C)$ as the fraction of edges in the original network that connect vertices in cluster i with vertices in cluster j, $i \neq j$.
- Given this,

$$mod(C) = \sum_{k=1}^{K} [f_{kk}(C) - f_{kk}^*]^2 \tag{1}$$

where f_{kk} is the fraction of edges which connect vertices within the same cluster k in G, and f_{kk}^* is the expected value of f_{kk} under some model of random edge assignment.

back