Causal Inference Under Approximate Neighborhood Interference*

Michael P. Leung†

June 25, 2020

Abstract. This paper studies causal inference in randomized experiments under network interference. Most of the literature assumes a model of interference under which treatments assigned to alters beyond a certain network distance from the ego have no effect on the ego's response. However, many models of social interactions do not satisfy this assumption. This paper proposes a substantially weaker model of "approximate neighborhood interference" (ANI), under which treatments assigned to alters further from the ego have a smaller, but potentially nonzero, impact on the ego's response. We show that ANI is satisfied in well-known models of social interactions. We also prove that, under ANI, standard inverse-probability weighting estimators can consistently estimate useful exposure effects and are asymptotically normal under asymptotics taking the network size large. For inference, we consider a network HAC variance estimator. Under a finite population model, we show the estimator is biased but that the bias can be interpreted as the variance of unit-level exposure effects. This generalizes Neyman's well-known result on conservative variance estimation to settings with interference.

JEL Codes: C22, C31, C57
Keywords: causal inference, network interference, social networks

*This research is supported by National Science Foundation grant SES-1755100. I thank Guido Imbens for suggestions that led to a simpler formulation of Assumption 4. I also thank Dean Eckles, Fredrik Sävje, and seminar audiences at Harvard, MIT, Northwestern, Stanford, and the 2019 SAMSI Causal Inference Workshop for useful feedback.

†Department of Economics, University of Southern California. E-mail: leungm@usc.edu.

1 Introduction

Randomized experiments in settings with network interference have seen increasing use in economics and the social sciences.[1] This paper develops methods for causal inference in experiments under weak restrictions on interference. We consider a finite population of n units connected through a network $\boldsymbol{A}$. Let $Y_i(\boldsymbol{d})$ denote the potential outcome of unit i under the counterfactual that the network is assigned treatment vector $\boldsymbol{d} = (d_i)_{i=1}^n \in \{0,1\}^n$. The dependence of $Y_i(\boldsymbol{d})$ on the entire vector of assignments allows for "interference" or "spillovers" and distinguishes this setup from the standard potential outcomes model. Interest centers on "exposure effects," defined below, that summarize how outcomes change in response to manipulations of $\boldsymbol{d}$.

The main inferential challenge is that, with a single network, the econometrician observes only one realization of the treatment assignment vector $\boldsymbol{D} = (D_i)_{i=1}^n$, where $D_i \in \{0,1\}$ denotes unit i's realized assignment. Identification of exposure effects is therefore impossible without restrictions on the manner in which $Y_i(\cdot)$ varies with $\boldsymbol{d}$. The predominant approach in the literature is to assume interference operates through a low-dimensional vector of sufficient statistics.[2] That is, $Y_i(\boldsymbol{D})$ is only a function of $\boldsymbol{D}$ through a vector-valued *exposure mapping* of fixed dimension

$$T_i \equiv T(i, \boldsymbol{D}, \boldsymbol{A}).$$

For example, Cai et al. (2015) run an experiment to study the effect of attending information sessions explaining the benefits of weather insurance on farmers' take-up of insurance. Spillover effects are of interest, since a farmer who attends a session may pass the information to friends. The authors estimate linear versions of the model

$$Y_i(\boldsymbol{D}) = \tilde{Y}_i(T_i), \quad \text{where} \quad T_i = \left(D_i, \frac{\sum_j A_{ij} D_j}{\sum_j A_{ij}} \right) \tag{1}$$

and A_{ij} is an indicator for whether farmers i and j are friends. Here the exposure mapping T_i is two-dimensional. Variation in its first component identifies the direct effect of the intervention, while variation in the second identifies a spillover effect.

[1] E.g. Bandiera et al. (2009), Bond et al. (2012), Bursztyn et al. (2014), Miguel and Kremer (2004), Paluck et al. (2016).

[2] E.g. Aronow and Samii (2017), Basse et al. (2019), Forastiere et al. (2016), Manski (2013), Toulis and Kao (2013).

More generally, exposure mappings, if correctly specified, substantially reduce the dimensionality of the model, allowing us to reparameterize potential outcomes as

$$Y_i(\boldsymbol{d}) = \tilde{Y}_i(t) \quad \text{for} \quad t \in \mathcal{T}, \tag{2}$$

where $\mathcal{T}$ is the range of $T(\cdot)$. Interest then centers on "exposure effects"

$$\frac{1}{n}\sum_{i=1}^{n}\left(\tilde{Y}_i(t) - \tilde{Y}_i(t')\right) \tag{3}$$

for $t, t' \in \mathcal{T}$, which measure the average in change in potential outcomes in response to counterfactual manipulations of the exposure mapping.

The literature predominantly studies model (2) under the assumption that the exposure mapping T_i only depends on treatments assigned to the *K-neighborhood* of i for some small K.[3] However, this imposes strong structural restrictions on the underlying outcome process that are incompatible with a variety of models of social interactions and contagion of interest in the networks literature (Guilbeault et al., 2018; Jackson, 2010; Manski, 1993). In these models, interference can arise from units outside of the ego's K-neighborhood, for any K. This is the case for models with endogenous peer effects, where outcomes are functions of the outcomes of neighbors (Eckles et al., 2017). Another example is the Cai et al. (2015) setting, where under a simple diffusion model, information obtained by treated units can eventually diffuse to arbitrarily distant alters, which violates (1).

Important recent work studies *misspecified* exposure mappings (Chin, 2018; Sävje et al., 2017; Sävje, 2019). The insight of this literature is that standard estimators for (3) unbiasedly estimate meaningful exposure effects even without imposing (2) to restrict interference, which indicates a certain robustness to more general patterns of interference. However, under what conditions inference can be made similarly robust is a more challenging question. Without (2), potential outcomes can depend arbitrarily on the entire assignment vector $\boldsymbol{D}$, which makes large-sample inference impossible. These papers accordingly propose a variety of high-level conditions weaker than (2) that implicitly restrict interference in order to obtain large-sample results. Unfortunately, the connection between these conditions and the literature on contagion and social interactions remains unclear. It is an open question whether models in the

[3]This is the set of units whose network distance from i is at most K, formally defined in §2.

networks literature violating (2) (e.g. models with endogenous peer effects or other diffusion models) can satisfy these high-level conditions.

Our Contribution. We study inference on exposure effects with misspecified exposure mappings under a new restriction on interference. We propose a model of *approximate neighborhood interference* (ANI), which requires treatments assigned to units far from the ego to have a small, but *potentially nonzero*, impact on the ego's response. Unlike the existing literature, we formally verify ANI in well-known models of social interactions. We then show that, under ANI, the data satisfies ψ-*weak dependence*, a recently proposed notion of network weak dependence. This enables us to apply limit theorems due to Kojevnikov et al. (2019) to establish that standard inverse probability weighting (IPW) estimators are consistent for certain exposure effects and asymptotically normal.

For inference, we consider a network HAC (heteroskedasticity and autocorrelation consistent) variance estimator and characterize its asymptotic bias under a finite population model. We show the bias can be interpreted as the variance of unit-level exposure effects, which is fundamentally unidentified even under no interference. This generalizes the well-known result on conservativeness of the standard variance estimator of the difference-in-means estimate under no interference (e.g. Imbens and Rubin, 2015, Ch. 6.4.2) to settings with dependence due to interference.

Finally, we propose a novel bandwidth for the network HAC estimator based on the average path length (APL) of the observed network. For a given bandwidth rule, the HAC estimator has different rates of convergence depending on whether the average K-neighborhood size in $\boldsymbol{A}$ grows exponentially or polynomially with K. The utility of the APL is that its magnitude adapts to the neighborhood growth rate.

Kojevnikov et al. (2019) and Kojevnikov (2019) respectively provide consistency results for network HAC and bootstrap variance estimators for ψ-weakly dependent network data. Their results pertain to settings in which the data is mean-homogeneous, which often holds in superpopulation models. We extend their results to settings with mean-heterogeneous data, as is the case in finite population models.

Choi (2017) and Choi (2018) study causal inference without imposing an exposure mapping model. These papers focus on different estimands than ours and assume treatment responses satisfy a monotonicity condition, which we do not require. There is also work on testing for interference, which can be used to test for correct specifi-

cation of exposure models (e.g. Athey et al., 2018).

Several papers in econometrics study causal inference under interference (Baird et al., 2018; He and Song, 2018; Lazzati, 2015; Leung, 2020; Vazquez-Bare, 2017; Viviano, 2019). The second paper studies a dynamic setting, whereas the others focus on a static setting like ours but under correctly specified exposure mappings. Many assume the special case of *stratified interference* whereby the data consists of many clusters and interference only operates within clusters. We instead consider a single large cluster with known network structure.

In the next section, we define the estimand and estimators and state basic assumptions. Then §3 presents our model of interference and large-sample results. We discuss variance estimation in §4. In §5, we illustrate the performance of our methods in an empirical application and simulation study. Finally, §6 concludes. All proofs are given in Appendix C.

2 Setup

We consider a finite population model in which the only random quantity is $\boldsymbol{D}$, which captures *design uncertainty* (Abadie et al., 2020; Imbens and Menzel, 2018). In the special case of no interference, this corresponds to the well-known Neyman causal model. The setup is advantageous because it allows for arbitrary dependence between the network and potential outcomes. For example, links may form at higher rates between units with similar unobservables, which corresponds to unobserved homophily, a well-known hindrance to identifying social interactions (Shalizi and Thomas, 2011).

Let $\mathcal{N}_n = \{1, \ldots, n\}$ denote the set of units. We assume $\boldsymbol{A}$ is an undirected and unweighted network with no self-links, represented as an adjacency matrix with ijth entry $A_{ij} \in \{0, 1\}$ denoting a potential link between units i and j. Treatments $\boldsymbol{D}$ are independent across units but not necessarily identically distributed, which allows for assignment based on unit covariates and network position. For example, treatment may be assigned "optimally" according to these characteristics (e.g. Viviano, 2019), or randomization may be stratified, as in the empirical application in §5.1.

Recall from §1 the definition of an exposure mapping $T(\cdot)$. Let $\mathcal{T} \subseteq \mathbb{R}^{d_T}$ be the range of $T(\cdot)$, and assume that $\mathcal{T}$ is discrete and d_T does not depend on n. Most of the literature assumes $T(\cdot)$ is *correctly specified* in the sense that (2) holds, which enables a simple definition of exposure effects (3). We instead follow the literature on

misspecified exposure mappings and employ $T(\cdot)$ only to define useful estimands that summarize treatment and spillover effects but not to restrict the true interference structure. This is a reasonable solution to the task of parsimoniously summarizing the causal effect of a high-dimensional vector $\boldsymbol{D}$ on potential outcomes.

Specifically, define the *unit-level exposure effect*

$$\tau_i(t,t') = \mu_i(t) - \mu_i(t'), \quad \text{where} \quad \mu_i(t) = \sum_{\boldsymbol{d}\in\{0,1\}^n} Y_i(\boldsymbol{d})\mathbf{P}(\boldsymbol{D} = \boldsymbol{d} \mid T_i = t)$$

and $t, t' \in \mathcal{T}$. This contrasts the expected response of unit i under two different values of the exposure mapping. The estimand of interest is the average effect

$$\tau(t,t') = \mu(t) - \mu(t'), \quad \text{where} \quad \mu(t) = \frac{1}{n}\sum_{i=1}^n \mu_i(t), \tag{4}$$

what Sävje (2019) calls a "misspecification-robust exposure effect." This is analogous to estimands proposed by Hudgens and Halloran (2008) but generalized to allow for an incomplete network. We refer to these references for more detailed discussion of interpretation, but the idea is analogous to (3), which is to compare the average outcomes of units under two different values of the exposure mapping.

Recall $T_i \equiv T(i, \boldsymbol{D}, \boldsymbol{A})$, and define the generalized propensity score

$$\pi_i(t) = \mathbf{E}[\mathbf{1}_i(t)], \quad \text{where} \quad \mathbf{1}_i(t) = \mathbf{1}\{T_i = t\}.^4$$

We estimate $\tau(t,t')$ using the standard IPW estimator, which is unbiased:

$$\hat{\tau}(t,t') = \hat{\mu}(t) - \hat{\mu}(t'), \quad \text{where} \quad \hat{\mu}(t) = \frac{1}{n}\sum_{i=1}^n Y_i \frac{\mathbf{1}_i(t)}{\pi_i(t)}.$$

As discussed in §1, most of the literature focuses on *K-neighborhood exposure mappings*. This requires $T(\cdot)$ to only be a function of $\boldsymbol{d}, \boldsymbol{A}$ through i's *K-neighborhood*, denoted $\mathcal{N}_{\boldsymbol{A}}(i,K) = \{j \in \mathcal{N}_n \colon \ell_{\boldsymbol{A}}(i,j) \leqslant K\}$, where $\ell_{\boldsymbol{A}}(i,j)$ is the path distance between i, j.[5] To formalize this requirement, for any $\boldsymbol{d} \in \{0,1\}^n$, define $d_{\mathcal{N}_{\boldsymbol{A}}(i,K)} =$

[4] Computation of $\pi_i(t)$ depends on the exposure mapping and design. In §5.1 where treatments are block randomized, we discuss use of the hypergeometric distribution.

[5] A *path* between i, j is a sequence of links $A_{k_1k_2}, A_{k_2k_3}, \ldots, A_{k_{m-1}k_m} = 1$ such that $k_1 = i$, $k_m = j$, and $k_a \neq k_b$ for all $a, b \in \{1, \ldots, m\}$. The *length* of this path is $m - 1$. The *path distance* between i, j is the length of the shortest path between them, defined as ∞ if no path exists.

$(d_j \colon j \in \mathcal{N}_{\boldsymbol{A}}(i,K))$ and $A_{\mathcal{N}_{\boldsymbol{A}}(i,K)} = (A_{kl} \colon k,l \in \mathcal{N}_{\boldsymbol{A}}(i,K))$, respectively the subvector of $\boldsymbol{d}$ and subnetwork of $\boldsymbol{A}$ on $\mathcal{N}_{\boldsymbol{A}}(i,K)$. Let $\mathcal{A}_n$ denote the set of networks on n units.

Assumption 1 (Exposure Mappings)**.** *There exists $K \in \mathbb{N}$ such that, for any $n \in \mathbb{N}$ and $i \in \mathcal{N}_n$, $T(i,\boldsymbol{d},\boldsymbol{A}) = T(i,\boldsymbol{d}',\boldsymbol{A}')$ for all $\boldsymbol{d},\boldsymbol{d}' \in \{0,1\}^n$ and $\boldsymbol{A},\boldsymbol{A}' \in \mathcal{A}_n$ such that $A_{\mathcal{N}_{\boldsymbol{A}}(i,K)} = A'_{\mathcal{N}_{\boldsymbol{A}'}(i,K)}$ and $d_{\mathcal{N}_{\boldsymbol{A}}(i,K)} = d'_{\mathcal{N}_{\boldsymbol{A}'}(i,K)}$.*

This is a weak restriction satisfied by most exposure mappings of interest in the literature. See for example (1) or the estimands in §5, where $K = 1$.

Assumption 2 (Overlap)**.** $\pi_i(t) \in [\underline{\pi},\overline{\pi}] \subset (0,1)$ $\forall\, n \in \mathbb{N}$, $i \in \mathcal{N}_n$, $\boldsymbol{A} \in \mathcal{A}_n$, $t \in \mathcal{T}$.

While overlap is standard, it can be restrictive. For instance, if $T_i = \mathbf{1}\{\sum_j A_{ij} D_j > 0\}$ and treatments are i.i.d., then $\pi_i(1)$ can be close to one if i's degree $\sum_j A_{ij}$ is large. However, overlap can be restored if we instead randomize treatment only to a smaller subset of "eligible" units and restrict the sample to units with an eligible neighbor (which changes the estimand). This is done in the empirical application in §5.1. Hence, the choice of design and estimand are important for overlap.

Finally, we assume for convenience that potential outcomes are uniformly bounded.

Assumption 3 (Bounded Outcomes)**.** $\sup_{n\in\mathbb{N},i\in\mathcal{N}_n,\boldsymbol{d}\in\{0,1\}^n,\boldsymbol{A}\in\mathcal{A}_n} |Y_i(\boldsymbol{d})| < \infty$.

3 Approximate Neighborhood Interference

We next present our model of interference. Recall that $d_{\mathcal{N}_{\boldsymbol{A}}(i,s)}$ is the subvector of $\boldsymbol{d}$ on i's s-neighborhood. Define

$$\Delta_{n,i}(s) = \max\big\{|Y_i(\boldsymbol{d}) - Y_i(\boldsymbol{d}')| \colon \boldsymbol{d},\boldsymbol{d}' \in \{0,1\}^n, d_{\mathcal{N}_{\boldsymbol{A}}(i,s)} = d'_{\mathcal{N}_{\boldsymbol{A}}(i,s)}\big\}.$$

Assumption 4 (ANI)**.** *For $\theta_{n,s} \equiv \max_{i\in\mathcal{N}_n} \Delta_{n,i}(s)$, $\sup_n \theta_{n,s} \to 0$ as $s \to \infty$.*[6]

The quantity $\Delta_{n,i}(s)$ measures the largest perturbation in i's potential outcome due

[6]This can be relaxed to $\theta_{n,s} \equiv \max_{i\in\mathcal{N}_n} \mathbf{E}[|Y_i(\boldsymbol{D}) - Y_i(\boldsymbol{D}')|]$ for random assignments $\boldsymbol{D},\boldsymbol{D}'$ such that $D_{\mathcal{N}_{\boldsymbol{A}}(i,s)} = D'_{\mathcal{N}_{\boldsymbol{A}}(i,s)}$ and $D_{-\mathcal{N}_{\boldsymbol{A}}(i,s)} = (D_j \colon j \notin \mathcal{N}_{\boldsymbol{A}}(i,s))$ is an independent copy of $D'_{-\mathcal{N}_{\boldsymbol{A}}(i,s)}$. The assumption in Proposition 2 below can then be weakened by redefining $\sigma_j = \mathbf{P}(0 < \phi(D_j,\varepsilon_j) \leqslant \beta)$ in the definition of $\boldsymbol{G}$.

to interference induced by changes to the treatment assignments of alters outside of i's s-neighborhood. ANI requires the maximal such perturbation $\theta_{n,s}$ to decay to zero with s. In the special case of correct specification (2), there is *no* interference from units outside i's K-neighborhood, so $\Delta_{n,i}(s) = 0$ for all i and $s \geqslant K$. In contrast, ANI generally allows $\Delta_{n,i}(s) \neq 0$ for *all* s but requires $\Delta_{n,i}(s)$ to decay with s. This means interference from outside of a unit's s-neighborhood is increasingly negligible as one expands the neighborhood radius s. Hence, ANI says that a unit's response is primarily, but not entirely, determined by the assignments of units close to it.[7]

The existence of interference is determined by $\boldsymbol{A}$ as well as the response model $Y_i(\cdot)$. Hence, ANI pertains to both quantities. For instance, if $\boldsymbol{A}$ is empty, then for $s > 0$, $\Delta_{n,i}(s) = 0$ *for any* $Y_i(\cdot)$, since no interference is possible without neighbors. On the other hand, if $Y_i(\boldsymbol{d}) = \tilde{Y}_i(d_i)$ for all $\boldsymbol{d}$, then for $s > 0$, $\Delta_{n,i}(s) = 0$ *for any* $\boldsymbol{A}$.

3.1 Social Interactions Models

We next verify ANI for arbitrary $\boldsymbol{A}$ under two models of social interactions whose reduced forms characterize $Y_i(\cdot)$. It may be possible to verify in other models, for example the SIR model, but we leave this to future work. Our results yield uniform bounds on $\theta_{n,s}$ that decay exponentially with s under restrictions on the strength of social interactions.

Linear-in-Means Model. Consider a network version of the Manski (1993) model

$$Y_i = \alpha + \beta \frac{\sum_j A_{ij} Y_j}{\sum_j A_{ij}} + D_i \gamma + \varepsilon_i, \tag{5}$$

where $\{\varepsilon_i\}_{i=1}^n$ is uniformly bounded (for Assumption 3). As usual, to ensure model (5) is coherent, we assume

$$|\beta| < 1. \tag{6}$$

The model defines potential outcomes $Y_i(\boldsymbol{D})$ through its reduced form

$$\boldsymbol{Y} = \frac{\alpha}{1-\beta}\boldsymbol{1} + \boldsymbol{D}\gamma + \gamma\beta \sum_{k=0}^{\infty} \beta^k \tilde{\boldsymbol{A}}^{k+1} \boldsymbol{D} + \sum_{k=0}^{\infty} \beta^k \tilde{\boldsymbol{A}}^k \boldsymbol{\varepsilon}$$

[7] Assumption 4 has some similarities with Assumption 6 of Chin (2018) in bounding the effect of manipulations of treatment assignments of distant units, although Chin's assumption also implicitly restricts the design. For a CLT, we do not need a high-level condition analogous to his Assumption 5, which requires correlations between $\{Y_i\}_{i=1}^n$ to be sufficiently weak.

(e.g. Bramoullé et al., 2009, eq. (6), assuming $\boldsymbol{A}$ is connected), where $\boldsymbol{Y} = (Y_i)_{i=1}^n$, $\boldsymbol{\varepsilon}$ is similarly defined, and $\tilde{\boldsymbol{A}}$ is the row-normalized version of $\boldsymbol{A}$ (divide each row by its sum). The third term roughly says that the impact of treatments assigned to k-neighbors is exponentially down-weighted by β^k. This suggests the following result.

Proposition 1. *If responses are realized according to the linear-in-means model, then there exists $C > 0$ such that Assumption 4 holds with $\theta_{n,s} \leqslant C|\beta|^s$ for all n.*

Complex Contagion. We next consider a model of "complex contagion," which has been widely studied in the networks literature.[8] Initialize a dynamic discrete-time process at period 0 at some binary outcome vector $\boldsymbol{Y}^0 \in \{0,1\}^n$. At each period t, units update their outcomes according to

$$Y_i^t = \mathbf{1}\left\{\beta \frac{\sum_j A_{ij} Y_j^{t-1}}{\sum_j A_{ij}} \geqslant \phi(D_i, \varepsilon_i)\right\}, \tag{7}$$

for some real-valued function $\phi(\cdot)$, to obtain their new outcomes $\boldsymbol{Y}^t$ from last period's outcomes $\boldsymbol{Y}^{t-1} = (Y_i^{t-1})_{i=1}^n$. The rule says that unit i chooses outcome 1 over 0 if and only if the fraction of neighbors choosing response 1 in the previous period is large enough relative to the heterogeneous threshold $\phi(D_i, \varepsilon_i)$. The parameter β controls the extent to which social interactions matter.

Because the setup in §2 is static, we consider running the dynamic process until the first period T such that $\boldsymbol{Y}^T = \boldsymbol{Y}^{T-1}$. To ensure such a T exists for any $\boldsymbol{Y}^0$, we assume $\beta \geqslant 0$ (Milgrom and Roberts, 1990). We then take $\boldsymbol{Y}^T$ as the vector of responses $\boldsymbol{Y}$ observed in the data. This process implicitly defines potential outcomes $(Y_i(\boldsymbol{D}))_{i=1}^n$. Note that different starting values $\boldsymbol{Y}^0$ can yield different outcomes $\boldsymbol{Y}$.

To verify Assumption 4, we need a condition analogous to (6), which will be more complicated to state, since the model is nonlinear. Define a weighted directed network $\boldsymbol{G}$ on $\mathcal{N}_n$ with ijth entry $G_{ij} = A_{ij}\sigma_j$ for $\sigma_j = \max_{d\in\{0,1\}} \mathbf{1}\{0 < \phi(d, \varepsilon_j) \leqslant \beta\}$. Let

$$\rho_n(\bar{s}) = \sup_{s>\bar{s}} \|\boldsymbol{G}^s\|_\infty^{1/s} = \sup_{s>\bar{s}} \left(\max_{i\in\mathcal{N}_n} \sum_{j=1}^n (\boldsymbol{G}^s)_{ij}\right)^{1/s}$$

for any $\bar{s} > 0$, where $\|\cdot\|_\infty$ is the matrix norm induced by the vector ∞-norm.

[8] E.g. Granovetter (1978), Guilbeault et al. (2018), Jackson (2010), Montanari and Saberi (2010).

Proposition 2. *Let $\alpha_{n,s} = 2\rho_n(\bar{s})^s$ for $s > \bar{s}$ and $\alpha_{n,s} = 2$ for $s \leqslant \bar{s}$. Suppose responses are realized according to the complex contagion model. If $\sup_n \rho_n(\bar{s}) < 1$ for some $\bar{s} > 0$, then Assumption 4 holds with $\theta_{n,s} \leqslant \alpha_{n,s}$.*

We next discuss the interpretation of $\sup_n \rho_n(\bar{s}) < 1$ in relation to (6). Let $\boldsymbol{I}$ be the $n \times n$ identity matrix. For the linear-in-means model to be coherent, we need $\boldsymbol{I} - \beta\tilde{\boldsymbol{A}}$ to be invertible, which is true provided

$$\lambda_{\max}(\beta\tilde{\boldsymbol{A}}) < 1, \tag{8}$$

where $\lambda_{\max}(\cdot)$ is the spectral radius. This is equivalent to $|\beta| < 1$, since $\tilde{\boldsymbol{A}}$ is row-normalized (Bramoullé et al., 2009). On the other hand, $\|\boldsymbol{G}^s\|_\infty^{1/s} \overset{s\to\infty}{\longrightarrow} \lambda_{\max}(\boldsymbol{G})$ by Gelfand's formula, so for any $\epsilon > 0$, we can choose $\bar{s}$ large enough such that

$$\sup_n \lambda_{\max}(\boldsymbol{G}) < 1 - \epsilon \quad \text{implies} \quad \sup_n \rho_n(\bar{s}) < 1 \tag{9}$$

(Xu and Lee, 2015). The left-hand side is clearly analogous to (8). The difference is that, in $\boldsymbol{G}$, we weight each potential link A_{ij} by σ_j, whereas in $\beta\tilde{\boldsymbol{A}}$, the weight is $\beta/\sum_k A_{ik}$. However, both weights are monotonically increasing in β, so both (8) and (9) implicitly restrict social interactions, as measured by β.

3.2 Weak Dependence

Define $Z_i = (\mathbf{1}_i(t)\pi_i(t)^{-1} - \mathbf{1}_i(t')\pi_i(t')^{-1})Y_i$, so that $\hat{\tau}(t,t') = n^{-1}\sum_{i=1}^n Z_i$. For large-sample inference, we would like the data $\{Z_i\}_{i=1}^n$ to be weakly dependent. Recall that $\ell_{\boldsymbol{A}}(i,j)$ is the path distance between i,j in $\boldsymbol{A}$. Since treatments are independent, the indicators have a local dependence structure: $\mathbf{1}_i(t) \perp\!\!\!\perp \mathbf{1}_j(t)$ if $\ell_{\boldsymbol{A}}(i,j) > 2K$ by Assumption 1. However, $\{Y_i\}_{i=1}^n$ may be strongly dependent, since Y_i can depend on the entire vector $\boldsymbol{D}$ under Assumption 4. Nonetheless, we expect this dependence to be sufficiently weak, since the assumption states that each Y_i depends primarily on the treatments of units nearby. We next formalize this idea.

Abusing notation, for any $H, H' \subseteq \mathcal{N}_n$, define $\ell_{\boldsymbol{A}}(H,H') = \min\{\ell_{\boldsymbol{A}}(i,j)\colon i \in H, j \in H'\}$. Let $Z_H = (Z_i\colon i \in H)$, $\mathcal{L}_d$ denote the set of bounded real-valued Lipschitz

functions on $\mathbb{R}^d$, and

$$\mathcal{P}_n(h, h'; s) = \{H, H' \subseteq \mathcal{N}_n \colon |H| = h, |H'| = h', \ell_{\boldsymbol{A}}(H, H') \geqslant s\}.$$

Definition 1. A triangular array $\{Z_i\}_{i=1}^n$ is *ψ-weakly dependent* if there exist (a) uniformly bounded constants $\{\tilde{\theta}_{n,s}\}_{s,n\in\mathbb{N}}$ such that $\sup_n \tilde{\theta}_{n,s} \to 0$ as $s \to \infty$, and (b) a collection of functionals $\{\psi_{h,h'}(\cdot,\cdot)\}_{h,h'\in\mathbb{N}}$ with $\psi_{h,h'} \colon \mathcal{L}_h \times \mathcal{L}_{h'} \to [0, \infty)$ such that

$$|\mathrm{Cov}(f(Z_H), f'(Z_{H'}))| \leqslant \psi_{h,h'}(f, f')\, \tilde{\theta}_{n,s}$$

for all $f \in \mathcal{L}_h$, $f' \in \mathcal{L}_{h'}$, and $H, H' \in \mathcal{P}_n(h, h'; 2s + 2)$ with $s > 0$.

This is essentially Definition 2.2 of Kojevnikov et al. (2019). It is analogous to a notion of weak dependence proposed by Doukhan and Louhichi (1999) for time series data but using path in place of temporal distance. The concept says two sets of observations Z_H and $Z_{H'}$ have small covariance if they are sufficiently distant.

For any Lipschitz $f(\cdot)$, let $\|f\|_\infty$ be its sup norm and $\mathrm{Lip}(f)$ its Lipschitz constant.

Theorem 1 (Weak Dependence)**.** *Under Assumptions 1–4, $\{Z_i\}_{i=1}^n$ is ψ-weakly dependent with (a) $\tilde{\theta}_{n,s} = \theta_{n,s}$ for all $n \in \mathbb{N}$ and $s > 0$, and (b) for some $C > 0$,*

$$\psi_{h,h'}(f, f') = C\left(\|f\|_\infty \|f'\|_\infty + h\|f'\|_\infty Lip(f) + h'\|f\|_\infty Lip(f')\right)$$

for all $h, h' \in \mathbb{N}$, $f \in \mathcal{L}_h$, and $f' \in \mathcal{L}_{h'}$.

3.3 Large-Sample Theory

Having established that the data $\{Z_i\}_{i=1}^n$ is ψ-weakly dependent, we can apply the large-sample theory developed in Kojevnikov et al. (2019) to show that $\hat{\tau}(t, t')$ is consistent and asymptotically normal. Their results require $\theta_{n,s}$ to decay to zero fast enough. How fast depends on the network topology, in particular the growth rate of s-neighborhood sizes. Intuitively, ANI says that Z_i depends primarily on the observations associated with other units in $\mathcal{N}_{\boldsymbol{A}}(i, s)$. Hence, if the size of this neighborhood grows rapidly with s, then for weak dependence, this needs to be counterbalanced by having the covariances (controlled by $\theta_{n,s}$) decay to zero faster with s.

We next present assumptions that formalize these ideas. To help clarify their

content, in §A.1, we verify them for networks with polynomial and exponential neighborhood growth rates when $\theta_{n,s}$ decays exponentially with s, as in the examples in §3.1. Let $\mathcal{N}_{\boldsymbol{A}}^{\partial}(i,s) = \{j \in \mathcal{N}_n \colon \ell_{\boldsymbol{A}}(i,j) = s\}$ be the *s-neighborhood boundary* of i, the set of units exactly path distance s from i and $\delta_n^{\partial}(s) = n^{-1}\sum_{i=1}^n |\mathcal{N}_{\boldsymbol{A}}^{\partial}(i,s)|$.

Assumption 5 (Weak Dependence for LLN)**.** $\sum_{s=0}^n \delta_n^{\partial}(s)\theta_{n,s} = o(n)$.

This is Assumption 3.2 of Kojevnikov et al. (2019). Since the average s-neighborhood boundary size grows with s, $\theta_{n,s}$ must decay to zero faster for the sum to be $o(n)$.

It is useful to compare this to its analog for α-mixing spatial processes. Consider, for example, Assumption 3(b) of Jenish and Prucha (2009), which essentially requires $\sum_{s=1}^{\infty} s^{d-1}\alpha(s) < \infty$, where d is the dimension of the underlying space and $\alpha(s)$ is the α-mixing coefficient, which measures dependence between sets of observations at spatial distance s apart. In the spatial setting, the s-neighborhood boundary of i is the set of units at any distance $h \in [s, s+1)$ from i. By their Lemma A.1(iii), the size of this set is $O(s^{d-1})$. Thus, we have a similar trade-off between the sizes of spatial s-neighborhood boundaries and the rate of decay of the mixing coefficient.

Theorem 2 (Consistency)**.** *Under Assumptions 1–5,* $|\hat{\tau}(t,t') - \tau(t,t')| \xrightarrow{p} 0$.

The proof of the theorem indicates that we can sharpen the result to $|\hat{\tau}(t,t') - \tau(t,t')| = O_p(n^{-1/2})$ if we strengthen Assumption 5 to $\sum_{s=0}^n \delta_n^{\partial}(s)\theta_{n,s} = O(1)$.

Asymptotic normality requires a stronger version of Assumption 5. Let $\delta_n(s,k) = n^{-1}\sum_{i=1}^n |\mathcal{N}_{\boldsymbol{A}}(i,s)|^k$, the kth moment of the s-neighborhood size, and

$$\mathcal{H}_n(s,m) = \left\{(i,j,k,l) \in \mathcal{N}_n^4 \colon k \in \mathcal{N}_{\boldsymbol{A}}(i,m), l \in \mathcal{N}_{\boldsymbol{A}}(j,m), \ell_{\boldsymbol{A}}(\{i,k\},\{j,l\}) = s\right\}.$$

This is the set of paired couples (i,j) and (k,l) such that the units within each couple are at most path distance m apart from one another, and the two pairs are exactly path distance s apart. Define $\sigma_n^2 = \mathrm{Var}(n^{-1/2}\sum_{i=1}^n Z_i)$.

Assumption 6 (Weak Dependence for CLT)**.** *There exist $\epsilon > 0$ and a sequence of*

positive constants $\{m_n\}_{n\in\mathbb{N}}$ *such that* $m_n \to \infty$ *and*

$$\max\left\{\sigma_n^{-4}\frac{1}{n^2}\sum_{s=0}^{n}|\mathcal{H}_n(s,m_n)|\theta_{n,s}^{1-\epsilon},\quad \sigma_n^{-3}n^{-1/2}\delta_n(m_n,2),\quad \sigma_n^{-1}n^{3/2}\theta_{n,m_n}^{1-\epsilon}\right\}\to 0. \quad (10)$$

This is essentially Assumption 3.4 of Kojevnikov et al. (2019). The first term in (10) is key. Similar to Assumption 5, it requires $\theta_{n,s}$ to decay to zero fast enough relative to s-neighborhood sizes. The second term restricts s-neighborhood growth rates, while the third requires sufficiently fast decay of $\theta_{n,s}$.

Theorem 3 (Asymptotic Normality)**.** *Under Assumptions 1–4 and 6,*

$$\frac{\sqrt{n}\big(\hat{\tau}(t,t')-\tau(t,t')\big)}{Var(\sqrt{n}\hat{\tau}(t,t'))^{1/2}}\xrightarrow{d}\mathcal{N}(0,1).$$

In the typical case where the variance is asymptotically non-degenerate, meaning $\liminf_{n\to\infty}\sigma_n^2>0$, the rate of convergence is $\sqrt{n}$.

4 Variance Estimation

For large-sample inference, we propose the variance estimator

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(Z_i-\hat{\tau}(t,t'))(Z_j-\hat{\tau}(t,t'))\mathbf{1}\{\ell_{\boldsymbol{A}}(i,j)\leqslant b_n\}, \quad (11)$$

where $b_n\geqslant 0$ is a bandwidth parameter discussed below. When $b_n=0$, this reduces to the sample variance of the Z_i's, which is only a valid estimator under no interference. Choosing $b_n>0$ places positive weight on pairs at most b_n apart in the network, which accounts for possible autocorrelation.[9]

In the special case of correctly specified exposure mappings, we can choose $b_n=2K$, since $\mathbf{1}_i(t)\perp\!\!\!\perp\mathbf{1}_j(t)$ if $\ell_{\boldsymbol{A}}(i,j)>2K$. This corresponds to the estimator of Leung (2020). For misspecified exposure mappings, we need b_n to grow with the sample size, and its rate of growth depends on the network topology. In this case, (11)

[9]The estimator is easy in practice to compute. First calculate the path distance matrix $(\ell_{\boldsymbol{A}}(i,j))_{i,j\in\mathcal{N}_n}$, which can be done very efficiently for sparse networks using Dijkstra's algorithm (e.g. the Python function `dijkstra` in the `scipy.sparse.csgraph` module). Then for adjacency matrix $\boldsymbol{G}=(\mathbf{1}\{\ell_{\boldsymbol{A}}(i,j)\leqslant b_n\})_{i,j\in\mathcal{N}_n}$ and $\tilde{\boldsymbol{Z}}=(n^{-1/2}(Z_i-\hat{\tau}(t,t')))_{i=1}^n$, we have $\hat{\sigma}^2=\tilde{\boldsymbol{Z}}'\boldsymbol{G}\tilde{\boldsymbol{Z}}$.

corresponds to a network version of a HAC estimator, which has been previously used in practice (e.g. Acemoglu et al., 2015) and whose formal properties have been studied by Kojevnikov et al. (2019) in a superpopulation setting. Our result below characterizes its behavior in a finite population model.

Remark 1. Kojevnikov et al. (2019) considers more general kernel functions that include the uniform kernel in (11). Kojevnikov (2019) proposes alternative weights motivated by network bootstrap procedures, which are guaranteed to be positive semidefinite (PSD) in finite sample, unlike kernel-based weights. A previous version of this paper (Leung, 2019a) considered a modification of his wild bootstrap weights, which are also guaranteed PSD and can be used as an alternative in instances where (11) is not PSD (which never occurred in our simulations in §5.2). While these weights all have the same asymptotic properties, in simulation experiments, we have found that uniform weights better control size in smaller sample sizes than weights that decay with distance.

Choice of Bandwidth. For $\hat{\sigma}^2$ to have good large-sample properties, we need to restrict the rate at which $|\mathcal{N}_{\boldsymbol{A}}(i, b_n)|$ diverges with n (see Assumption 7). Hence, how fast b_n can diverge depends on how rapidly s-neighborhoods expand with s. In spatial settings, K-neighborhood sizes grow polynomially with K, so b_n is allowed to diverge at a polynomial rate. (A faster rate is better for bias but worse for variance.) However, in network settings, K-neighborhood sizes can also grow exponentially with K. Based on the analysis in §A.2, we suggest b_n be chosen as follows. Let $\delta(\boldsymbol{A}) = n^{-1}\sum_{i,j} A_{ij}$ denote the average degree and $\Delta(\boldsymbol{A})$ the *average path length* (APL).[10] We propose

$$b_n = \lfloor \max\{\tilde{b}_n, 2K\} \rceil \quad \text{for} \quad \tilde{b}_n = \begin{cases} \frac{1}{2}\Delta(\boldsymbol{A}) & \text{if } \Delta(\boldsymbol{A}) > 2\frac{\log n}{\log \delta(\boldsymbol{A})}, \\ \Delta(\boldsymbol{A})^{1/3} & \text{otherwise,} \end{cases} \tag{12}$$

where $\lfloor \cdot \rceil$ means round to the nearest integer. We suggest in practice the researcher report results for several bandwidths in a neighborhood of (12).

As discussed above, b_n is at least $2K$ to account for correlation in $\{\mathbf{1}_i(t)\}_{i=1}^n$. The fractions 1/2 and 1/3 come from verifying Assumption 7 below (see §A.2). The pur-

[10] The APL is the average value of $\ell_{\boldsymbol{A}}(i, j)$ over all pairs in the largest component of $\boldsymbol{A}$. A *component* of a network is a connected subnetwork such that all units in the subnetwork are disconnected from those not in the subnetwork.

pose of comparing $\Delta(\boldsymbol{A})$ and $\log n / \log \delta(\boldsymbol{A})$ is to determine whether K-neighborhood sizes grow approximately exponentially or polynomially with K. As discussed in §A.2, in the exponential case, the difference between the two is expected to converge to zero, whereas in the polynomial case, $\Delta(\boldsymbol{A})$ is much larger, having polynomial order. See, for example, the simulations in §5.2, which show at least a four-fold difference in APL between the two regimes. Thus, (12) selects a bandwidth of logarithmic (polynomial) order when neighborhood growth rates are approximately exponential (polynomial).[11]

Bias of $\hat{\sigma}^2$. Define

$$\hat{\sigma}_*^2 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(Z_i - \tau_i(t,t'))(Z_j - \tau_j(t,t'))\mathbf{1}\{\ell_{\boldsymbol{A}}(i,j) \leqslant b_n\} \quad \text{and}$$

$$R_n = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(\tau_i(t,t') - \tau(t,t'))(\tau_j(t,t') - \tau(t,t'))\mathbf{1}\{\ell_{\boldsymbol{A}}(i,j) \leqslant b_n\},$$

where $\hat{\sigma}_*^2$ is an "oracle" version of $\hat{\sigma}^2$ that replaces $\hat{\tau}(t,t')$ with $\tau_i(t,t')$, while R_n is a bias term. Theorem 4 below establishes that

$$\hat{\sigma}^2 = \hat{\sigma}_*^2 + R_n + o_p(1) \quad \text{and} \tag{13}$$

$$|\hat{\sigma}_*^2 - \mathrm{Var}(\sqrt{n}\hat{\tau}(t,t'))| \xrightarrow{p} 0. \tag{14}$$

Equation (14) says that the oracle estimator is consistent for the variance, and (13) says that our estimator is biased. The source of bias is mean-heterogeneity: $\hat{\tau}(t,t')$ is consistent for $\tau(t,t')$ but not $\tau_i(t,t')$, which is heterogeneous across units.

The bias R_n has the form of a HAC estimate of the variance of the unit-level exposure effects. It is helpful to compare this to the case of no interference, where $T_i = D_i$, $t = 1$, and $t' = 0$, so that $\tau(t,t')$ is the usual average treatment effect (ATE). Knowing that units are independent, we can choose $b_n = 0$, in which case

$$R_n = \frac{1}{n}\sum_{i=1}^{n}\big(\tau_i(1,0) - \tau(1,0)\big)^2.$$

[11] In the exponential case, we need $b_n = O(\log n)$ for $\mathrm{Var}(\hat{\sigma}^2)$ to be small, which (12) accomplishes, since $\Delta(\boldsymbol{A})$ is expected to be $O(\log n)$ in this regime (see §A.2). In the polynomial case, a bandwidth of logarithmic order is also valid, but the bias will vanish at a very slow rate (see Kojevnikov et al., 2019, proof of Proposition 4.1). Our choice of $\Delta(\boldsymbol{A})^{1/3}$ substantially improves this rate, since $\Delta(\boldsymbol{A})$ is polynomial in n in this regime.

This is the well-known asymptotic bias of the standard variance estimator for the difference-in-means estimate of the ATE (e.g. Imbens and Rubin, 2015, Theorem 6.2). It measures the variance of the unit-level treatment effects and is generally impossible to estimate in the finite population setting, so the variance estimator is conservative. In the special case of homogeneous unit-level treatment effects, meaning $\tau_i(t,t')$ does not vary with i, the bias is zero, a property also shared by our R_n. Thus, (13) generalizes Neyman's well-known result on conservative variance estimation to a setting with interference. The additional covariance terms in R_n weighted by $\mathbf{1}\{\ell_{\boldsymbol{A}}(i,j) \leqslant b_n\}$ can be thought of as accounting for dependence due to interference.

Remark 2. In the Appendix B, we compare R_n with the bias of the Aronow and Samii (2017) variance estimator used in the existing literature for correctly specified exposure mappings. Simulation results there show that our bias is positive but can be notably smaller than theirs. More generally, the asymptotic behavior of R_n depends on the superpopulation model towards which our framework is agnostic. Since R_n has the form of a network HAC, we expect that it typically converges to the population variance of the unit-level exposure effects, although this requires additional weak dependence conditions on the distributions of $Y_i(\boldsymbol{d})$ and $\boldsymbol{A}$. Some such conditions are given in Theorem 4.2 of Leung (2019b).

To show consistency of $\hat{\sigma}^2$, define

$$\mathcal{J}_n(s,m) = \left\{(i,j,k,l) \in \mathcal{N}_n^4 \colon k \in \mathcal{N}_{\boldsymbol{A}}(i,m), l \in \mathcal{N}_{\boldsymbol{A}}(j,m), \ell_{\boldsymbol{A}}(i,j) = s\right\}.$$

This is similar to, and in fact contains, $\mathcal{H}_n(s,m)$ from Assumption 6.

Assumption 7 (Weak Dependence for $\hat{\sigma}^2$)**.** *For some $\epsilon > 0$, (a) $\sum_{s=0}^{n} \delta_n^{\partial}(s)\theta_{n,s}^{1-\epsilon} = O(1)$, (b) $\delta_n(b_n,1) = o(n^{1/2})$, (c) $\delta_n(b_n,2) = o(n)$, (d) $\sum_{s=0}^{n} |\mathcal{J}_n(s,b_n)|\theta_{n,s}^{1-\epsilon} = o(n^2)$.*

Part (a) strengthens Assumption 5. The main assumptions are parts (b)–(d), which regulate how fast b_n grows relative to neighborhood sizes. In §A.2, we use these to derive (12). Part (b) allows us to replace $\hat{\tau}(t,t')$ in $\hat{\sigma}^2$ with its expectation. Part (d) is used to derive the asymptotic bias. It is similar to the first requirement of Assumption 6, except we replace b_n and $\mathcal{J}_n(s,\cdot)$ with m_n and $\mathcal{H}_n(s,\cdot)$.

Theorem 4 (Variance Estimator). *If $b_n \to \infty$ as $n \to \infty$, then under Assumptions 1–4 and 7,* (13) *and* (14) *hold.*

5 Numerical Illustrations

5.1 Empirical Application

We revisit a network experiment analyzed in Paluck et al. (2016) and Aronow and Samii (2017) that studies the effect of an anti-conflict intervention on adolescent social norms for antagonistic behavior, including harassment, rumor mongering, social exclusion, and bullying. In the experimental design, 28 of 56 schools are first randomized into treatment. Then within treated schools, a subset of students are selected as eligible for treatment based on covariates, and half of eligibles are block randomized into treatment. Treated students are invited to participate in bi-monthly meetings that follow an anti-conflict curriculum designed in part by the researchers of the study. At these meetings, a trained adult leader helps students identify social conflicts at their school and design strategies to reduce conflict.

Aronow and Samii (2017) and part of the analysis of Paluck et al. (2016) examine the causal effect of the offer to participate on endorsement of anti-conflict norms. This is measured by self reports of wearing a wristband disseminated as part of the program as a reward to students observed engaging in conflict-mitigating behavior. Through the course of the experiment, over 2500 wristbands were disseminated and tracked. We follow the analysis of Aronow and Samii (2017) and study similar exposure effects. Unlike their analysis, we restrict the sample to the five largest treated schools to illustrate what can be learned from data on a few large networks. In each of our schools, the number of eligibles is exactly 64.

We estimate a treatment and a spillover effect. For the latter, the exposure mapping is $T_i = \mathbf{1}\{\sum_j A_{ij} D_j > 0\}$, an indicator for whether at least one friend is offered treatment. Following Aronow and Samii (2017), to ensure overlap, we restrict to the "spillover sample" consisting of students that have at least one eligible friend ($n = 1685$). For the treatment effect, the exposure mapping is $T_i = D_i$, and we restrict to the "treatment sample" consisting of students eligible for treatment ($n = 320$).

Networks are measured by asking students to name up to ten students at the school with "whom they chose to spend time with in the last few weeks, either in

school, out of school, or online." Consequently, $\boldsymbol{A}$ is directed. When computing the number of treated friends for the exposure mappings, we use the directionality of links. However, when computing network neighborhoods for our variance estimator, we ignore the directionality of links to conservatively define larger neighborhoods and avoid taking a stance on neighborhood definitions for directed networks.

We next provide some summary statistics. Within the treatment sample, the average outcome Y_i is 0.16 (SD 0.37), and by block randomization, exactly 50 percent are treated. Within the spillover sample, the average outcome is 0.11 (SD 0.32), and 58 percent (SD 0.49) have at least one treated friend. The data includes the blocks in which eligible students are block-randomized, so we can compute the propensity scores $\pi_i(t)$ for each student using the hypergeometric distribution. For the exposure mapping $T_i = \mathbf{1}\{\sum_j A_{ij}D_j > 0\}$, given N eligible neighbors, $\pi_i(0)$ is the chance of having 0 out of N successes when drawing without replacement. Then $n^{-1}\sum_{i=1}^n \pi_i(1) = 0.597$, which is assuringly very close to empirical proportion of 58 percent.

The average out-degree $n^{-1}\sum_{i,j} A_{ij}$ is 7.96. The APL is small, on average 3.37 across our five schools. Since there are $n = 3306$ students, $\log n / \log \delta(\boldsymbol{A}) = 3.96$, which is very close to 3.37. Thus, given $K = 1$, our suggested bandwidth (12) is $b_n = 2$. We report results for the range of bandwidths $\{0, \ldots, 3\}$, noting that 0 corresponds to i.i.d. standard errors.

Table 1: Estimates and SEs

	Treatment	Spillover
$\hat{\tau}(1,0)$	0.1500	0.0407
$\hat{\mu}(1)$	0.2375	0.1293
$\hat{\mu}(0)$	0.0875	0.0885
$b_n = 0$	0.0443	0.0167
$b_n = 1$	0.0460	0.0184
$b_n = 2$	0.0394	0.0205
$b_n = 3$	0.0470	0.0170

Columns display results for the treatment ($n = 320$) and spillover ($n = 1685$) effects. Rows "$b_n = k$" report SEs for the indicated bandwidths.

Table 1 presents the results. The first row is the IPW estimator for the indicated exposure effect, and the last four rows are standard errors for the indicated

bandwidths. We find a large treatment effect of 0.15, which is significant at the 5 percent level across all bandwidths. The spillover effect is smaller at 0.04, with larger standard errors, and is statistically insignificant for our suggested bandwidth $b_n = 2$ at the 5 percent level. The small spillover estimate is largely in line with the estimates implied by Figure 3C of Paluck et al. (2016). It does not contradict the overall message of their paper, since they find, for example, sizeable spillover effects when comparing treated and untreated schools. In contrast, our analysis above only makes comparisons within treated schools, using only a subsample of five schools.

To compare treated and untreated schools, note that in the latter, $Y_i = 0$ for all i by design. Then our estimate of $\hat{\mu}(0)$ for the treatment effect shows that even untreated units are 8.8 percentage points more likely to wear wristbands in treated compared to untreated schools. We compute standard errors for $\hat{\mu}(0)$ by replacing Z_i and $\hat{\tau}(t, t')$ in (11) with $\mathbf{1}_i(t)\pi_i(t)^{-1}Y_i$ and $\hat{\mu}(0)$. For $b_n = 0, \ldots, 3$, the standard errors range from 0.012 to 0.017, all of which indicate a statistically significant effect. Overall, these results indicate that, despite the potential conservativeness of our estimator due to the bias term R_n, they can still deliver reasonable standard errors.

5.2 Monte Carlo

To study the finite sample properties of our estimators, we simulate data from the two response models studied in §3.1 and two models of network formation calibrated to the school data from §5.1. For the linear-in-means model, $Y_i = V_i(\boldsymbol{D}, \boldsymbol{A}, \boldsymbol{\varepsilon})$ for

$$V_i(\boldsymbol{D}, \boldsymbol{A}, \boldsymbol{\varepsilon}) = \alpha + \beta \frac{\sum_j A_{ij} Y_j}{\sum_j A_{ij}} + \delta \frac{\sum_j A_{ij} D_j}{\sum_j A_{ij}} + D_i \gamma + \varepsilon_i$$

and $(\alpha, \beta, \delta, \gamma) = (-1, 0.8, 1, 1)$. For the complex contagion model, $Y_i = \mathbf{1}\{V_i(\boldsymbol{D}, \boldsymbol{A}, \boldsymbol{\varepsilon}) > 0\}$ for $(\alpha, \beta, \delta, \gamma) = (-1, 1.5, 1, 1)$. We simulate $\boldsymbol{A}$ from configuration and random geometric graph (RGG) models. The former is calibrated to the empirical out-degree sequence $(\sum_{j=1}^n A_{ij})_{i=1}^n$ of the schools used in §5.1. This model (approximately) draws an undirected network uniformly at random from the set of all networks with this degree sequence (e.g. Jackson, 2010, Ch. 4.1.4). An RGG is a spatial network where units only link with geographically close alters: $A_{ij} = \mathbf{1}\{\|\rho_i - \rho_j\| \leqslant r_n\}$ for $\rho_i \overset{iid}{\sim} \mathcal{U}([0, 1]^2)$ and $r_n = (\kappa/(\pi n))^2$. Since κ is the limiting expected degree of the model (Penrose, 2003), we set it equal to the average of the empirical out-degree sequence of the

schools. Let $\{\nu_i\}_{i=1}^n \overset{iid}{\sim} \mathcal{N}(0,1)$ be independent of $\boldsymbol{A}$. Under the configuration model, we take $\nu_i = \varepsilon_i$. Under the RGG, $\varepsilon_i = (\rho_{i1} - 0.5) + \nu_i$ adding the centered first component of i's "location" ρ_{i1}. Since units with similar ρ_{i1}'s are more likely to form links, this generates unobserved homophily.

Both network formation models have approximately the same average degree of about 8 and are therefore sparse. For our purposes, the main distinction between them is the APL. The configuration model is theoretically known to have a small APL of logarithmic order in the network size (Van Der Hofstad, 2016), while the RGG model generates a larger APL of polynomial order (Friedrich et al., 2013). We will therefore choose different bandwidths for the two according to (12). We also note that the value of the peer effect β in both outcome models is actually quite large. The calculations in Appendices A.1 and A.2 suggest that both outcome models actually fail to satisfy our weak dependence conditions under the configuration model (but not the RGG), since it generates a low APL (and hence approximately exponential neighborhood growth rates). Of course, these conditions are likely stronger than necessary, which may explain the good performance in both designs below.

To show different sample sizes, we report results using the largest, two largest, and four largest of the treated schools when calibrating the network models. In all cases, we pool the degree sequences across the schools to treat them as one single network. Following the design in §5.1, we randomly assign treatments to only units classified as eligible in the data with probability 0.5.

We compute estimates and standard errors for the spillover effect $\tau(1,0)$ in §5.1, whose exposure mapping is $T_i = \mathbf{1}\{\sum_j A_{ij} D_j > 0\}$, again restricting to the sample of units with an eligible neighbor. The standard errors use the bandwidth (12). Given the APL differences discussed above, which can also be seen in the tables below, the RGG uses $\tilde{b}_n = \Delta(\boldsymbol{A})^{1/3}$, while the configuration model uses $\tilde{b}_n = \Delta(\boldsymbol{A})/2$.

To illustrate the performance of our bandwidth rule across different networks, we redraw $\boldsymbol{D}$, $\boldsymbol{A}$, and $\boldsymbol{\varepsilon}$ for each of the 10k simulation draws. This corresponds to a superpopulation design, so we expect our standard errors to yield confidence intervals with coverage close to the nominal rate of 0.95. We report "oracle" standard errors, which correspond to $\mathrm{Var}(\hat{\tau}(1,0))^{1/2}$, approximated by taking the standard deviation of $\hat{\tau}(1,0)$ over 10k separate simulation draws. We also report "naive" i.i.d. standard errors to illustrate the degree of dependence in the data.

Tables 2 and 3 present results for the configuration and RGG models, respectively.

Table 2: Simulation Results: Configuration Model

	Linear-in-Means			Complex Contagion		
# Schools	1	2	4	1	2	4
$\hat{n}(1)$	199.3	399.7	804.2	199.3	399.7	804.2
$\hat{n}(0)$	151.5	293.1	570.9	151.5	293.1	570.9
$\hat{\tau}(1,0)$	0.311	0.317	0.313	0.077	0.077	0.080
Our SE	1.310	0.955	0.686	0.109	0.081	0.059
Oracle SE	1.402	0.986	0.694	0.117	0.084	0.060
Our Coverage	0.923	0.938	0.947	0.928	0.937	0.943
Oracle Coverage	0.947	0.950	0.954	0.948	0.950	0.951
Naive Coverage	0.544	0.549	0.563	0.729	0.725	0.730
APL	3.471	3.753	4.070	3.471	3.753	4.070
b_n	2.000	2.000	2.000	2.000	2.000	2.000
Network Size	805	1456	2725	805	1456	2725

Cells are averages over 10k simulations. $\hat{n}(t) = \sum_i \mathbf{1}_i(t)$ gives the effective sample size of $\hat{\mu}(t)$. "Coverage" rows display empirical coverage for 95% CIs. "Naive" and "Oracle" respectively correspond to i.i.d. and true standard errors.

In both tables $\hat{n}(t) = \sum_i \mathbf{1}_i(t)$, which gives the effective sample sizes for $\hat{\mu}(1)$ and $\hat{\mu}(0)$. Row "Our SE" displays the standard error obtained from our variance estimator, and "Our Coverage" displays the empirical coverage of the corresponding two-sided 95-percent confidence interval. "Network Size" is the number of units in the network, whereas the sample size is $\hat{n}(1) + \hat{n}(0)$, since, as in the empirical application, we only use units with an eligible neighbor for overlap.

The results show that the oracle coverage rates are all very close to 0.95, which illustrates the quality of the normal approximation. Our standard errors perform well, with coverage near 0.95. This is in spite of fairly large peer effects β, which can be seen in the magnitudes of the spillover effect estimates. By contrast, naive standard errors are severely anti-conservative.

6 Conclusion

The causal literature on interference typically assumes K-neighborhood exposure mappings are correctly specified. This implies a limited model of interference in which units further than distance K from the ego have no effect on the ego's response for

Table 3: Simulation Results: RGG Model

	Linear-in-Means			Complex Contagion		
# Schools	1	2	4	1	2	4
$\hat{n}(1)$	210.0	418.2	830.1	210.0	418.2	830.1
$\hat{n}(0)$	155.5	297.8	576.7	155.5	297.8	576.7
$\hat{\tau}(1,0)$	0.702	0.700	0.700	0.088	0.089	0.089
Our SE	1.523	1.107	0.805	0.144	0.105	0.076
Oracle SE	1.601	1.142	0.824	0.152	0.108	0.079
Our Coverage	0.927	0.936	0.936	0.935	0.941	0.943
Oracle Coverage	0.953	0.950	0.947	0.950	0.951	0.954
Naive Coverage	0.519	0.522	0.525	0.626	0.641	0.623
APL	13.996	18.444	24.927	13.996	18.444	24.927
b_n	2.007	3.000	3.000	2.007	3.000	3.000
Network Size	805	1456	2725	805	1456	2725

See table notes of Table 2.

some small, known K. Such a model is incompatible with those studied in the large theoretical and empirical literature on contagion and social interactions, in which units arbitrarily far from the ego can have an effect on the ego's outcome. This paper proposes a richer model of "approximate neighborhood interference" (ANI), under which the effect on the ego's outcome of treatment assigned to any alter is potentially nonzero but decreasing with network distance between ego and alter. Unlike with existing models of interference, we show that ANI is satisfied by well-known models of social interactions. We also show that our model is useful for large-sample inference, proving that, under ANI, standard IPW estimators are consistent for exposure effects and asymptotically normal.

For inference, we consider a network HAC variance estimator, which enables inference robust to misspecification of exposure mappings under ANI. We show the estimator is biased in a finite population setting. The bias term captures the variance of the unit-level treatment effects, which generalizes the well-known result on conservative variance estimation under no interference to settings with dependence due to interference. Finally, we propose a new bandwidth rule for the HAC estimator, which trades off bias and variance in manner that adapts to the growth rate of K-neighborhood sizes.

A Verifying Assumptions

A.1 Assumption 6

We verify Assumption 6 for networks with either polynomial or exponential neighborhood growth rates. We do so in the context of a model in which the variance is non-degenerate, so that $\sigma_n^{-2} = O(1)$, and $\theta_{n,s} = e^{-c(1-\epsilon)^{-1}s}$ for some $c \geqslant 0$ and $\epsilon > 0$, which holds for the examples in §3.1.

Polynomial Growth Rate. Suppose

$$\max_{i \in \mathcal{N}_n} |\mathcal{N}_{\boldsymbol{A}}(i,s)| = Cs^d \tag{15}$$

for s sufficiently large and some $C > 0$, $d \geqslant 1$. Polynomial rates appear to be a property of spatial networks, which are models in which link formation is less likely between spatially distant units.[12] Examples include latent space (Hoff et al., 2002) and RGG models (Penrose, 2003). Appendix A of Leung (2019b) shows that, for the RGG, path distance is of the same order as spatial distance for connected units. Since spatial s-neighborhoods grow polynomially with s, it follows that network s-neighborhoods also grow polynomially, with d equal to the underlying spatial dimension.

We next verify Assumption 6 under this setup for $m_n = n^{1/(\alpha d)}$, $\alpha > 4$. First consider the third term in (10). This is $O(n^{1.5}e^{-cm_n})$ and hence $o(1)$, since m_n is polynomial in n. The second term $n^{-3/2}\sum_{i=1}^n |\mathcal{N}_{\boldsymbol{A}}(i,m_n)|^2$ is order $n^{-1/2}m_n^{2d} = n^{2/\alpha - 0.5} = o(1)$. Finally, for the first term in (10), observe that we can bound

$$|\mathcal{H}_n(s,m_n)| \leqslant \sum_{i=1}^n \sum_{j \in \mathcal{N}_{\boldsymbol{A}}^{\partial}(i,s)} |\mathcal{N}_{\boldsymbol{A}}(i,m_n)|\,|\mathcal{N}_{\boldsymbol{A}}(j,m_n)|, \tag{16}$$

which is conservatively order $n\,m_n^{2d}s^d = n^{2/\alpha+1}s^d$. Hence, the first term in (10) is order $n^{2/\alpha-1}\sum_{s=0}^n s^d e^{-cs} = o(1)$.

Exponential Growth Rates. Suppose

$$\max_{i \in \mathcal{N}_n} |\mathcal{N}_{\boldsymbol{A}}(i,s)| = Ce^{\beta s} \tag{17}$$

[12] This may require replacing the max in (15) with something weaker like an average, but the max makes the verification argument far simpler.

for some $C, \beta > 0$ and s sufficiently large. In the extreme case of a k-regular tree network, the left-hand side is exactly k^s. Inhomogeneous random graphs, a large class of models that includes the Erdős-Rényi and stochastic block models, have a similar property (see Bollobás et al., 2007, proofs of Lemmas 14.2 and 14.3). The usual heuristic argument (Barabási, 2015, Ch. 3.8) is that the average number of units in a unit's 1-neighborhood is the average degree $\delta(\boldsymbol{A}) = n^{-1}\sum_{i,j} A_{ij}$, so the typical number in its s-neighborhood is $\approx \delta(\boldsymbol{A})^s$, in which case $\beta \approx \log \delta(\boldsymbol{A})$ and $C \approx 1$.

We verify Assumption 6 for $m_n = \alpha\beta^{-1}\log n$, $\alpha \in (1.5\beta c^{-1}, 0.5)$, with c from the definition of $\theta_{n,s}$ above. Such an α exists only if $c > 3\beta$, which requires $\theta_{n,s}$ to decay sufficiently fast relative to neighborhood growth rates. The third term in (10) is order $n^{1.5}e^{-cm_n} = n^{1.5-c\alpha\beta^{-1}} = o(1)$. The second term is order $n^{-1/2}e^{2\beta m_n} = n^{2\alpha-0.5} = o(1)$. Finally, using (16), the first term of (10) is conservatively order

$$n^{-1}e^{2\beta m_n}\sum_{s=0}^{n} e^{\beta s}e^{-cs} = n^{2\alpha-1}\sum_{s=0}^{n} e^{(\beta-c)s} = o(1).$$

A.2 Choice of Bandwidth

We use a mix of formal and heuristic arguments to show our proposed bandwidth (12) satisfies Assumption 7(b)–(d) under different neighborhood growth rates. As in §A.1, we suppose that $\sigma_n^{-2} = O(1)$ and $\theta_{n,s} = e^{-c(1-\epsilon)^{-1}s}$ for some $c \geqslant 0$ and $\epsilon > 0$.

Polynomial Growth Rates. We first consider the case in which s-neighborhood sizes grow polynomially with s in the sense of (15). Let $\nabla(\boldsymbol{A})$ be the diameter of $\boldsymbol{A}$, which is the maximum path length between pairs in the largest component. Then

$$\max_i |\mathcal{N}_{\boldsymbol{A}}(i, \nabla(\boldsymbol{A}))| = \alpha n, \tag{18}$$

where α is the fraction of units in the largest component of $\boldsymbol{A}$. Most real-world networks have a "giant component" (Barabási, 2015) meaning $\alpha > 0$ for n large. Then since the left-hand side of (18) is $C\nabla(\boldsymbol{A})^d$ under (15), $\nabla(\boldsymbol{A}) = O(n^{1/d})$. This is a well-known heuristic argument for the asymptotic behavior of the diameter or average path length (Barabási, 2015, Ch. 3.8).[13]

[13] See Friedrich et al. (2013) for a formal argument for the RGG.

Our bandwidth rule (12) is based on the APL $\Delta(\boldsymbol{A})$ rather than the diameter, since the former is considered a more robust measurement of network width.[14] While the heuristics above pertain to the diameter, in fact the derived rate is accurate for the APL. As written in Barabási (2015), Ch. 3.8, "for most networks [these heuristics offer] a better approximation to the average distance between two randomly chosen nodes, than to [the diameter]. This is because [the diameter] is often dominated by a few extreme paths, while [the APL] is averaged over all node pairs, a process that supresses [sic] the fluctuations." Hence, our calculations below will assume $\Delta(\boldsymbol{A}) \approx n^{1/d}$.

We can now verify Assumption 7(b)–(d). For (b), using our bandwidth choice $b_n = \Delta(\boldsymbol{A})^{1/3} \approx n^{1/(3d)}$, $n^{-1}\sum_{i=1}^n |\mathcal{N}_{\boldsymbol{A}}(i, b_n)| \approx b_n^d \approx n^{1/3} = o(n^{1/2})$. Part (c) is the same as the second part of (10) except with b_n in place of m_n. Then following the argument in §A.1, $n^{-1}\sum_{i=1}^n |\mathcal{N}_{\boldsymbol{A}}(i, b_n)|^2 \approx b_n^{2d} \approx n^{2/3} = o(n)$. Part (d) is the same as the first part of (10) but with b_n in place of m_n and $\mathcal{J}_n(s, \cdot)$ in place of $\mathcal{H}_n(s, \cdot)$. Note that (16) applies with $\mathcal{J}_n(s, \cdot)$ in place of $\mathcal{H}_n(s, \cdot)$, so

$$\frac{1}{n^2}\sum_{s=0}^{n} |\mathcal{J}_n(s, b_n)| \theta_{n,s}^{1-\epsilon} \approx n^{-1} b_n^{2d} \sum_{s=0}^{n} s^d e^{-cs} = O(n^{-1/3}).$$

Exponential Growth Rates. Now suppose (17), and take the typical case of $\beta = \log \delta(\boldsymbol{A})$ and $C = 1$ discussed after (17). The heuristics following (18) yield $\nabla(\boldsymbol{A}) = \log n / \log \delta(\boldsymbol{A})$, as in (3.18) of Barabási (2015). As discussed above, this heuristic is actually more accurate for the APL, so we next assume $\Delta(\boldsymbol{A}) \approx \log n / \log \delta(\boldsymbol{A})$.[15]

Choosing b_n according to (12) yields $b_n \approx 0.5 \log n / \log \delta(\boldsymbol{A})$. Strictly speaking, we will actually need $b_n \approx \alpha \log n / \log \delta(\boldsymbol{A})$ for some $\alpha < 0.5$, which we will assume next. In practice, since path lengths take discrete values, we just simplify to $\alpha = 0.5$ in (12). For Assumption 7(b), $n^{-1}\sum_{i=1}^n |\mathcal{N}_{\boldsymbol{A}}(i, b_n)| \leqslant e^{b_n \log \delta(\boldsymbol{A})} \approx e^{\alpha \log n} = o(n^{1/2})$, and (c) is similar. For (d), if $c > \beta = \log \delta(\boldsymbol{A})$ as in §A.1, then since (16) holds for $\mathcal{J}_n(s, \cdot)$ in place of $\mathcal{H}_n(s, \cdot)$,

$$\frac{1}{n^2}\sum_{s=0}^{n} |\mathcal{J}_n(s, b_n)| \theta_{n,s}^{1-\epsilon} \approx n^{-1} e^{2b_n \log \delta(\boldsymbol{A})} \sum_{s=0}^{n} e^{s \log \delta(\boldsymbol{A})} e^{-cs} = O(n^{2\alpha - 1}).$$

[14] I thank a referee for this suggestion.

[15] For a formal argument for inhomogeneous graphs, see Theorem 3.14 of Bollobás et al. (2007).

B AS Variance Estimator

We provide a theoretical comparison of our variance estimator and that of Aronow and Samii (2017) (henceforth AS) and provide simulation evidence on differences in conservativeness. Since their estimator is only valid under correctly specified exposure mappings, we now assume (2). The estimand is

$$\begin{aligned}
&\sigma_n^2 = \text{Var}(\sqrt{n}\hat{\tau}(t,t')) = V(t) + V(t') + 2C(t,t'), \quad \text{where} \\
&\Psi(t) = \frac{1}{n}\sum_{i=1}^n \tilde{Y}_i(t)^2 \frac{1-\pi_i(t)}{\pi_i(t)} + \frac{1}{n}\sum_{i=1}^n \sum_{j\neq i} \tilde{Y}_i(t)\tilde{Y}_j(t) \frac{\pi_{ij}(t) - \pi_i(t)\pi_j(t)}{\pi_i(t)\pi_j(t)}, \\
&C(t,t') = -\frac{1}{n}\sum_{i=1}^n \sum_{j=1}^n \tilde{Y}_i(t)\tilde{Y}_j(t') \frac{\pi_{ij}(t,t') - \pi_i(t)\pi_j(t')}{\pi_i(t)\pi_j(t')}.
\end{aligned}$$

Define $\pi_{ij}(t,t') = \mathbf{E}[\mathbf{1}_i(t)\mathbf{1}_j(t')]$. The AS estimator given in their equation (11) is

$$\begin{aligned}
&\hat{\sigma}_{AS}^2 = \hat{V}(t) + \hat{V}(t') + 2\hat{C}(t,t'), \quad \text{where} \\
&\hat{V}(t) = \frac{1}{n}\sum_{i=1}^n \frac{Y_i^2 \mathbf{1}_i(t)}{\pi_i(t)} \frac{1-\pi_i(t)}{\pi_i(t)} + \frac{1}{n}\sum_{i=1}^n \sum_{j\neq i} \frac{Y_i \mathbf{1}_i(t) Y_j \mathbf{1}_j(t)}{\pi_{ij}(t,t)} \frac{\pi_{ij}(t,t) - \pi_i(t)\pi_j(t)}{\pi_i(t)\pi_j(t)} \mathbf{1}\{\pi_{ij}(t,t) \neq 0\} \\
&\qquad + \frac{1}{n}\sum_{i=1}^n \sum_{j\neq i} \left(\frac{Y_i^2 \mathbf{1}_i(t)}{2\pi_i(t)} + \frac{Y_j^2 \mathbf{1}_j(t)}{2\pi_j(t)} \right) \mathbf{1}\{\pi_{ij}(t,t) = 0\}, \\
&\hat{C}(t,t') = -\frac{1}{n}\sum_{i=1}^n \sum_{j\neq i} \frac{Y_i \mathbf{1}_i(t) Y_j \mathbf{1}_j(t')}{\pi_{ij}(t,t')} \frac{\pi_{ij}(t,t') - \pi_i(t)\pi_j(t')}{\pi_i(t)\pi_j(t')} \mathbf{1}\{\pi_{ij}(t,t') \neq 0\} \\
&\qquad + \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^n \left(\frac{Y_i^2 \mathbf{1}_i(t)}{2\pi_i(t)} + \frac{Y_j^2 \mathbf{1}_j(t')}{2\pi_j(t')} \right) \mathbf{1}\{\pi_{ij}(t,t') = 0\}.
\end{aligned}$$

This is conservative for σ_n^2 with bias

$$\begin{aligned}
&R_{n,AS} \equiv \mathbf{E}[\hat{\sigma}_{AS}^2] - \sigma_n^2 = \frac{1}{n}\sum_{i=1}^n (\tilde{Y}_i(t) - \tilde{Y}_i(t'))^2 \\
&+ 0.5\frac{1}{n}\sum_{i=1}^n \sum_{j\neq i} \big((\tilde{Y}_i(t) + \tilde{Y}_j(t))^2 \mathbf{1}\{\pi_{ij}(t,t) = 0\} + 2(\tilde{Y}_i(t) - \tilde{Y}_j(t'))^2 \mathbf{1}\{\pi_{ij}(t,t') = 0\} \\
&\qquad + (\tilde{Y}_i(t') + \tilde{Y}_j(t'))^2 \mathbf{1}\{\pi_{ij}(t',t') = 0\} \big).
\end{aligned}$$

The asymptotic bias of our estimator $\hat{\sigma}^2$ is R_n given in (13) with $b_n = 2K$, since

we are assuming correct specification. It does not appear that R_n and $R_{n,AS}$ can generally be ordered. However, we can consider a few instructive special cases.

First consider **no interference** ($K = 0$). Then $R_n = n^{-1}\sum_{i=1}^n(\tau_i(t,t') - \tau(t,t'))^2$. For the AS estimator, since $\pi_{ij}(t,t') \neq 0$ for all $i \neq j$, $R_{n,AS} = n^{-1}\sum_{i=1}^n \tau_i(t,t')^2$, which is strictly larger than R_n whenever the treatment effect $\tau(t,t')$ is nonzero. Indeed $R_{n,AS}$ is larger by the *square* of the treatment effect, so that for large effects and relatively small sample sizes, $\hat{\sigma}^2_{AS}$ can be substantially more conservative than $\hat{\sigma}^2$. Next consider **homogeneous unit-level exposure effects**, which corresponds to $\tau_i(t,t') = \tau(t,t')$ for all i. In this case, $R_n = 0$, whereas $R_{n,AS}$ does not simplify in general and is strictly positive if, for example, $\tau_i(t,t') \neq 0$ for all i. Finally, consider the **has-treated-neighbor spillover effect** in §5, where $t = 1$, $t' = 0$, and $T_i = \mathbf{1}\{\sum_j A_{ij}D_j > 0\}$. Call a unit i "eligible" if $\mathbf{E}[D_i] > 0$. Let $\mathcal{E}_{ij}$ be the event that the set of i's eligible neighbors in $\boldsymbol{A}$ equals the set of eligible neighbors i has in common with j. Then $\pi_{ij}(1,0) = 0$ if and only if $\mathcal{E}_{ij}$ occurs, and $\pi_{ij}(s,s) \neq 0$ for all s and units i and j who have at least one eligible neighbor. Thus, assuming the population consists of all units with eligible neighbors,

$$R_{n,AS} = \frac{1}{n}\sum_{i=1}^n \tau_i(1,0)^2 + \frac{1}{n}\sum_{i=1}^n\sum_{j\neq i}(\tilde{Y}_i(1) - \tilde{Y}_j(0))^2\mathbf{1}\{\mathcal{E}_{ij}\}, \quad \text{whereas} \tag{19}$$

$$\begin{aligned} R_n = \frac{1}{n}\sum_{i=1}^n &(\tau_i(1,0) - \tau(1,0))^2 \\ &+ \frac{1}{n}\sum_{i=1}^n\sum_{j\neq i}(\tau_i(1,0) - \tau(1,0))(\tau_j(1,0) - \tau(1,0))\mathbf{1}\{\ell_{\boldsymbol{A}}(i,j) \leqslant 2\}. \end{aligned} \tag{20}$$

While the first terms of both expressions can be ordered as in the no-interference case, the remaining terms cannot. The remaining term in (20) is an estimate of the covariances of unit-level exposure effects, but the remainder in (19) does not have an obvious interpretation. However, we can provide simulation evidence on magnitudes.

We simulate networks calibrated to the data in the empirical application as in §5.2 but use a different outcome model to impose correct specification: $\tilde{Y}_i(0) = \varepsilon_i + \sum_j A_{ij}\varepsilon_j / \sum_j A_{ij}$ and $\tilde{Y}_i(1) = \beta_i + \tilde{Y}_i(0)$, where $\{\varepsilon_i\}_{i=1}^n \overset{iid}{\sim} \mathcal{N}(0,1)$ is independent of $\{\beta_i\}_{i=1}^n \overset{iid}{\sim} \mathcal{N}(1,1)$. The fraction in $\tilde{Y}_i(0)$ can be interpreted as exogenous peer effects in unobservables. It serves to generate network autocorrelation in baseline outcomes. As in §5.2, the sample is the set of units with eligible neighbors. The left half of

Table 4 reports the average values of $R_{n,AS}$ and R_n across 5000 simulation draws. We see that $R_{n,AS}$ is 3–5 times larger than R_n on average.

In the previous design, $\tau_i(1,0)$ is independent across i. To introduce network autocorrelation in unit-level exposure effects, we instead take $\tilde{Y}_i(1) = \beta_i + \sum_j A_{ij}\beta_j / \sum_j A_{ij} + \tilde{Y}_i(0)$. The right half of Table 4 reports the results. For the configuration model, $R_{n,AS}$ is twice as large as R_n on average, and for the RGG model, it is 25 percent larger.

Table 4: Comparison of Average Bias

	Independent Effects				Autocorrelated Effects			
	Configuration		RGG Model		Configuration		RGG Model	
# Schools	1	2	1	2	1	2	1	2
$R_{n,AS}$	9.4	10.2	7.1	7.6	154.6	158.7	170.0	180.0
R_n	1.9	1.9	2.0	2.0	67.6	73.8	136.6	144.1
n	350.8	692.8	365.5	716.1	350.8	692.8	365.5	716.1

Cells are averages over 5k simulations. n = # units with eligible neighbors.

C Proofs

Proof of Proposition 1. The reduced form of the linear-in-means model is

$$Y_i(\boldsymbol{d}) = \alpha + d_i\gamma + \varepsilon_i + \mathbf{1}\left\{\sum_j A_{ij} > 0\right\} \times \left(\frac{\alpha\beta}{1-\beta} + \gamma\beta \sum_{k=0}^{\infty} \beta^k \left(\sum_{j_1=1}^{n} \frac{A_{ij_1}}{\sum_\ell A_{i\ell}} \sum_{j_2=1}^{n} \frac{A_{j_1 j_2}}{\sum_\ell A_{j_2\ell}} \cdots \sum_{j_{k+1}=1}^{n} \frac{A_{j_k j_{k+1}}}{\sum_\ell A_{j_{k+1}\ell}} d_{j_{k+1}}\right) + \sum_{k=1}^{\infty} \beta^k \left(\sum_{j_1=1}^{n} \frac{A_{ij_1}}{\sum_\ell A_{i\ell}} \sum_{j_2=1}^{n} \frac{A_{j_1 j_2}}{\sum_\ell A_{j_2\ell}} \cdots \sum_{j_k=1}^{n} \frac{A_{j_{k-1} j_k}}{\sum_\ell A_{j_k\ell}} \varepsilon_{j_k}\right)\right).$$

for $\boldsymbol{d} = (d_i)_{i=1}^n \in \{0,1\}^n$. Consider a counterfactual linear-in-means model in which the network is $A_{\mathcal{N}_{\boldsymbol{A}}(i,s)}$ rather than $\boldsymbol{A}$. This means that the set of units is $\mathcal{N}_{\boldsymbol{A}}(i,s)$ rather than $\mathcal{N}_n$, and outcomes are realized according to (5) but with primitives $(d_{\mathcal{N}_{\boldsymbol{A}}(i,s)}, A_{\mathcal{N}_{\boldsymbol{A}}(i,s)}, \varepsilon_{\mathcal{N}_{\boldsymbol{A}}(i,s)})$ rather than $(\boldsymbol{d}, \boldsymbol{A}, \boldsymbol{\varepsilon})$, where $\boldsymbol{\varepsilon} = (\varepsilon_i)_{i=1}^n$ and $\varepsilon_{\mathcal{N}_{\boldsymbol{A}}(i,s)} = (\varepsilon_i \colon i \in \mathcal{N}_{\boldsymbol{A}}(i,s))$. Let $Y_i^{(s)}(\boldsymbol{d})$ be unit i's outcome in this counterfactual model. To compare $Y_i^{(s)}(\boldsymbol{d})$ with $Y_i(\boldsymbol{d})$, consider the kth term in the first series of the previous

equation (the term multiplying $\gamma\beta \cdot \beta^k$). We can rewrite this as $\sum_{j=1}^n \omega_{ij}^k d_j$, where ω_{ij}^k is the following weighted sum of all walks of length $k+1$ from i to j:

$$\omega_{ij}^k = \sum_{j_1=1}^n \sum_{j_2=1}^n \cdots \sum_{j_{k-1}=1}^n \frac{A_{ij_1}}{\sum_\ell A_{i\ell}} \frac{A_{j_1 j_2}}{\sum_\ell A_{j_2\ell}} \cdots \frac{A_{j_k j}}{\sum_\ell A_{j\ell}} \leqslant 1.$$

Thus, in going from $Y_i(\boldsymbol{d})$ to $Y_i^{(s)}(\boldsymbol{d})$, we lose terms in ω_{ij}^k involving walks that traverse paths of length greater than s. We can conservatively bound this loss by including *all* walks with length exceeding s in the loss. Then since $\omega_{ij}^k \leqslant 1$ and d_j is binary,

$$|Y_i(\boldsymbol{d}) - Y_i^{(s)}(\boldsymbol{d})| \leqslant \gamma\beta \sum_{k=s+1}^\infty |\beta|^k + \sup_i |\varepsilon_i| \sum_{k=s+1}^\infty |\beta|^k$$

for $s \geqslant 1$. Since ε_i is uniformly bounded, the right-hand side is bounded by a constant times $|\beta|^s$. Furthermore, since $\boldsymbol{d}$ is arbitrary, everything above holds replacing $\boldsymbol{d}$ with $\boldsymbol{d}'$ such that $d_{\mathcal{N}_{\boldsymbol{A}}(i,s)} = d'_{\mathcal{N}_{\boldsymbol{A}}(i,s)}$. Since $Y_i^{(s)}(\boldsymbol{d}) = Y_i^{(s)}(\boldsymbol{d}')$, the result follows from the triangle inequality. ∎

Proof of Proposition 2. The arguments that follow borrow ideas from the proof of Proposition 1 of Xu and Lee (2015) and that of Theorem 6.1 of Leung (2019b). Let C_i be the set of units in the strongly connected component of $\boldsymbol{G}$ containing i and

$$C_i^+ = C_i \cup \{k \in \mathcal{N}_n \colon \exists j \in C_i \text{ such that } A_{jk}(1-\sigma_i) = 1\}.$$

Let $Y_i(\boldsymbol{d})$ be i's outcome under the complex contagion model when the network is assigned treatments according to $\boldsymbol{d} = (d_i)_{i=1}^n$. As in the proof of Proposition 1, consider a counterfactual model in which the network is $A_{C_i^+}$ rather than $\boldsymbol{A}$. This means that the set of units is C_i^+ rather than $\mathcal{N}_n$, and outcomes are realized according to the same model but with primitives $(d_{C_i^+}, A_{C_i^+}, \varepsilon_{C_i^+})$ rather than $(\boldsymbol{d}, \boldsymbol{A}, \boldsymbol{\varepsilon})$. Let $Y_i^{(s)}(\boldsymbol{d})$ be i's outcome in this counterfactual model. Key to our argument is the fact

$$Y_i(\boldsymbol{d}) = Y_i^{(s)}(\boldsymbol{d}). \tag{21}$$

To see why, note that if $\sigma_i = 0$, then i has a dominant strategy $Y_i^*(\boldsymbol{d})$, and as soon as $Y_i^t(\boldsymbol{d}) = Y_i^*(\boldsymbol{d})$ at some period t of the dynamic process, $Y_i^{t+s}(\boldsymbol{d}) = Y_i^t(\boldsymbol{d})$ for all $s > 0$. If instead $\sigma_i = 1$, then i's outcome may potentially change at any period t in

the process, depending on the outcomes of neighboring units at $t-1$.

Generalizing this logic, consider any path in $\boldsymbol{A}$ connecting units i and j. If $\sigma_k = 1$ for all units k along that path, then unit i's outcome may change at any period t in the dynamic process, depending on the outcome of j at some prior period. However, if for all such paths, there exists some unit k along that path such that $\sigma_k = 0$, then unit i's outcome will never be affected by unit's j outcome at any past period. Now, if $j \notin C_i^+$, then by construction, there exists such a unit k along any path connecting i and j. Therefore, unit i's eventual outcome $Y_i(\boldsymbol{d})$ is invariant to the removal of units $\mathcal{N}_n \backslash C_i^+$ from the network, and (21) follows. This is the same logic as the decentralized selection mechanism assumption in Appendix A of Leung (2019b). Thus,

$$\begin{aligned} |Y_i(\boldsymbol{d}) - Y_i^{(s)}(\boldsymbol{d})| &\leqslant \mathbf{1}\{Y_i(\boldsymbol{d}) \neq Y_i^{(s)}(\boldsymbol{d})\} \leqslant \mathbf{1}\{A_{C_i^+} \nsubseteq A_{\mathcal{N}_{\boldsymbol{A}}(i,s)}\} \\ &\leqslant \sum_{j_1 \neq \cdots \neq j_{s-2}} A_{ij_1}\sigma_{j_1} A_{j_1 j_2}\sigma_{j_2} \cdot \cdots \cdot A_{j_{s-3}j_{s-2}}\sigma_{j_{s-2}} = \sum_j (\boldsymbol{G}^{s-2})_{ij}. \quad (22) \end{aligned}$$

The first two inequalities use (21). The third uses the union bound and the fact that $A_{C_i^+} \nsubseteq A_{\mathcal{N}_{\boldsymbol{A}}(i,s)}$ implies there exists a path of length $s-2$ from i to $j_{s-2} \in \mathcal{N}_{\boldsymbol{A}}(i,s)$ such that $\sigma_k = 1$ for all units k on that path.

For all s, $|Y_i(\boldsymbol{d}) - Y_i^{(s)}(\boldsymbol{d})| \leqslant 1$. For $s - 2 > \bar{s}$, (22) $\leqslant \rho_n(\bar{s})^s$. Since $\boldsymbol{d}$ is arbitrary, everything above holds replacing $\boldsymbol{d}$ with $\boldsymbol{d}'$ such that $d_{\mathcal{N}_{\boldsymbol{A}}(i,s)} = d'_{\mathcal{N}_{\boldsymbol{A}}(i,s)}$. Since $Y_i^{(s)}(\boldsymbol{d}) = Y_i^{(s)}(\boldsymbol{d}')$, the result follows from the triangle inequality. ∎

Proof of Theorem 1. Let $h, h' \in \mathbb{N}$, $s > 0$, $H, H' \in \mathcal{P}_n(h, h'; 2s+2)$, $f \in \mathcal{L}_h$, and $f' \in \mathcal{L}_{h'}$. Define $\xi = f(Z_H)$ and $\zeta = f'(Z_{H'})$. Let $\boldsymbol{D}', \boldsymbol{D}''$ each be independent copies of $\boldsymbol{D}$, $\boldsymbol{D}^{(s,\xi)} = (\boldsymbol{D}_{\mathcal{N}_{\boldsymbol{A}}(i,s)}, \boldsymbol{D}'_{-\mathcal{N}_{\boldsymbol{A}}(i,s)})$, and $\boldsymbol{D}^{(s,\zeta)} = (\boldsymbol{D}_{\mathcal{N}_{\boldsymbol{A}}(i,s)}, \boldsymbol{D}''_{-\mathcal{N}_{\boldsymbol{A}}(i,s)})$. Define

$$\begin{aligned} Z_i^{(s,\xi)} &= Y_i(\boldsymbol{D}^{(s,\xi)}) \left(\frac{\mathbf{1}\{T(i, \boldsymbol{D}^{(s,\xi)}, \boldsymbol{A}) = t\}}{\pi_i(t)} - \frac{\mathbf{1}\{T(i, \boldsymbol{D}^{(s,\xi)}, \boldsymbol{A}) = t'\}}{\pi_i(t')} \right), \\ Z_i^{(s,\zeta)} &= Y_i(\boldsymbol{D}^{(s,\zeta)}) \left(\frac{\mathbf{1}\{T(i, \boldsymbol{D}^{(s,\xi)}, \boldsymbol{A}) = t\}}{\pi_i(t)} - \frac{\mathbf{1}\{T(i, \boldsymbol{D}^{(s,\zeta)}, \boldsymbol{A}) = t'\}}{\pi_i(t')} \right). \end{aligned}$$

Finally, let $\xi^{(s)} = f((Z_i^{(s,\xi)} : i \in H))$ and $\zeta^{(s)} = f'((Z_i^{(s,\zeta)} : i \in H'))$. Recall that $T(\cdot)$ is a K-neighborhood exposure mapping (Assumption 1). Since Z_i is uniformly bounded by Assumptions 2 and 3, $|\mathrm{Cov}(\xi, \zeta)| \leqslant 2\|f\|_\infty \|f'\|_\infty$. Thus, we can choose C in the theorem large enough such that $|\mathrm{Cov}(\xi, \zeta)| \leqslant \psi_{h,h'}(f, f')\theta_{n,s}$ for any $s \leqslant K-1$.

Now, fix any $s > K-1$, so that $2s+2 > 2K$. By Assumption 1, $\ell_{\boldsymbol{A}}(H, H') > 2K$ implies $(Z_i^{(s,\xi)} : i \in H) \perp\!\!\!\perp (Z_i^{(s,\zeta)} : i \in H)$. Then there exists $C > 0$ such that

$$\begin{aligned}
|\mathrm{Cov}(\xi, \zeta)| &\leqslant |\mathrm{Cov}(\xi - \xi^{(s)}, \zeta)| + |\mathrm{Cov}(\xi^{(s)}, \zeta - \zeta^{(s)})| \\
&\leqslant C\|f'\|_\infty \mathbf{E}[|\xi - \xi^{(s)}|] + C\|f\|_\infty \mathbf{E}[|\zeta - \zeta^{(s)}|] \\
&\leqslant C\left(h\|f'\|_\infty \mathrm{Lip}(f) + h'\|f\|_\infty \mathrm{Lip}(f')\right)\theta_{n,s}.
\end{aligned}$$

The last line follows because under Assumption 1,

$$\frac{\mathbf{1}\{T(i, \boldsymbol{D}^{(s,\xi)}, \boldsymbol{A}) = t\}}{\pi_i(t)} = \frac{\mathbf{1}\{T(i, \boldsymbol{D}^{(s,\zeta)}, \boldsymbol{A}) = t\}}{\pi_i(t)} = \frac{\mathbf{1}\{T(i, \boldsymbol{D}, \boldsymbol{A}) = t\}}{\pi_i(t)}$$

for any $t \in \mathcal{T}$, and Assumption 4 implies $\max_{i \in \mathcal{N}_n} \mathbf{E}[|Y_i - Y_i(\boldsymbol{D}^{(s,\xi)})|] \leqslant \theta_{n,s}$. ∎

Proof of Theorem 2. Since $\mathbf{E}[\hat{\tau}(t,t')] = \tau(t,t')$, we only need to show that $\mathrm{Var}(\hat{\tau}(t,t')) = o(1)$. Since treatments are independent across units, by Assumption 4, $\mathrm{Cov}(Z_i, Z_j) = 0$ if (i,j) are not connected in $\boldsymbol{A}$. Hence,

$$\mathrm{Var}(\hat{\tau}(t,t')) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(Z_i) + \sum_{s=1}^{n-1} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j \neq i} \mathbf{1}\{\ell_{\boldsymbol{A}}(i,j) = s\} \mathrm{Cov}(Z_i, Z_j).$$

Using Theorem 1 and uniform boundedness of Z_i, the right-hand side is bounded above by $C(n^{-1} + n^{-2}\sum_{s=1}^{n-1} \theta_{n,s} \sum_{i=1}^{n} |\mathcal{N}_{\boldsymbol{A}}^{\partial}(i,s)|)$ for some universal positive constant C, and this is $o(1)$ by Assumption 5. ∎

Proof of Theorem 3. Apply Theorem 3.2 of Kojevnikov et al. (2019). ∎

Proof of Theorem 4. Observe that (14) follows from Proposition 4.1 of Kojevnikov et al. (2019), since our Assumption 7 implies their Assumption 4.1. Define

$$\begin{aligned}
r_n = \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} (Z_i - \tau(t,t'))(\tau(t,t') - \hat{\tau}(t,t'))\mathbf{1}\{\ell_{\boldsymbol{A}}(i,j) \leqslant b_n\} \\
+ (\tau(t,t') - \hat{\tau}(t,t'))^2 \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{1}\{\ell_{\boldsymbol{A}}(i,j) \leqslant b_n\}.
\end{aligned}$$

To establish (13), note that there are two parts of alleged $o_p(1)$ term in (13). The

first is due to replacing $\hat{\tau}(t,t')$ with $\tau(t,t')$ in the formula for $\hat{\sigma}^2$. This replacement creates the remainder term r_n. Since Z_i is uniformly bounded, for some $C > 0$,

$$|r_n| \leqslant C\|\tau(t,t') - \hat{\tau}(t,t')\| \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{1}\{\ell_{\boldsymbol{A}}(i,j) \leqslant b_n\}.$$

The summation term is equal to $\delta_n(b_n, 1)$. The norm term is $O_p(n^{-1/2})$, since Assumption 7(a) implies that $\mathrm{Var}(\hat{\tau}(t,t')) = O(n^{-1})$ (see the proof of Theorem 2). Hence, the previous display is $o_p(1)$ by Assumption 7(b).

The remaining parts of the alleged $o_p(1)$ term in (13) are the cross terms

$$\frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(Z_i - \tau_i(t,t'))(\tau_j(t,t') - \tau(t,t'))\mathbf{1}\{\ell_{\boldsymbol{A}}(i,j) \leqslant b_n\}.$$

We show this is $o_p(1)$. For $W_i = \sum_{j=1}^{n}(\tau_j(t,t') - \tau(t,t'))\mathbf{1}\{\ell_{\boldsymbol{A}}(i,j) \leqslant b_n\}$,

$$\begin{aligned}
&\mathbf{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(Z_i - \tau_i(t,t'))(\tau_j(t,t') - \tau(t,t'))\mathbf{1}\{\ell_{\boldsymbol{A}}(i,j) \leqslant b_n\}\right|\right] \\
&\qquad \leqslant \frac{1}{n}\mathbf{E}\left[\left(\sum_{i=1}^{n}(Z_i - \tau_i(t,t'))W_i\right)^2\right]^{1/2} \\
&\qquad \leqslant \left(\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(Z_i)W_i^2 + \bar{\psi}\frac{1}{n^2}\sum_{s=0}^{n}\theta_{n,s}\sum_{i=1}^{n}\sum_{j\neq i}\mathbf{1}\{\ell_{\boldsymbol{A}}(i,j) = s\}|W_iW_j|\right)^{1/2},
\end{aligned}$$

for some $\bar{\psi} > 0$ by Theorem 1. Since Z_i is uniformly bounded, for some $C > 0$, $n^{-2}\sum_{i=1}^{n}\mathrm{Var}(Z_i)W_i^2 \leqslant Cn^{-1}\delta_n(b_n, 2)$, which is $o(1)$ by Assumption 7(c). Likewise,

$$\frac{1}{n^2}\sum_{s=0}^{n}\theta_{n,s}\sum_{i=1}^{n}\sum_{j\neq i}\mathbf{1}\{\ell_{\boldsymbol{A}}(i,j) = s\}|W_iW_j| \leqslant \frac{C}{n^2}\sum_{s=0}^{n}\theta_{n,s}\mathcal{J}_n(s, b_n)$$

for some $C > 0$, and this is $o(1)$ by Assumption 7(d). ∎

References

Abadie, A., S. Athey, G. Imbens, and J. Wooldridge, "Sampling-Based vs. Design-Based Uncertainty in Regression Analysis," *Econometrica*, 2020, *88* (1),

265–296.

Acemoglu, D., C. Garcia-Jimeno, and J. Robinson, "State Capacity and Economic Development: A Network Approach," *American Economic Review*, 2015, *105* (8), 2364–2409.

Aronow, P. and C. Samii, "Estimating Average Causal Effects Under General Interference, with Application to a Social Network Experiment," *Annals of Applied Statistics*, 2017, *11* (4), 1912–1947.

Athey, Susan, D. Eckles, and G. Imbens, "Exact P-Values for Network Interference," *Journal of the American Statistical Association*, 2018, *113* (521), 230–240.

Baird, S., J. Bohren, C. McIntosh, and B. Özler, "Optimal Design of Experiments in the Presence of Interference," *Review of Economics and Statistics*, 2018, *100* (5), 844–860.

Bandiera, O., I. Barankay, and I. Rasul, "Social Connections and Incentives in the Workplace: Evidence from Personnel Data," *Econometrica*, 2009, *77* (4), 1047–1094.

Barabási, A., *Network Science*, Cambridge University Press, 2015.

Basse, G., A. Feller, and P. Toulis, "Randomization Tests of Causal Effects Under Interference," *Biometrika*, 2019, *106* (2), 487–494.

Bollobás, B., S. Janson, and O. Riordan, "The Phase Transition in Inhomogeneous Random Graphs," *Random Structures and Algorithms*, 2007, *31* (1), 3–122.

Bond, R., C. Fariss, J. Jones, A. Kramer, C. Marlow, Jaime S., and James F., "A 61-Million-Person Experiment in Social Influence and Political Mobilization," *Nature*, 2012, *489* (7415), 295–298.

Bramoullé, Y., H. Djebbari, and B. Fortin, "Identification of Peer Effects Through Social Networks," *Journal of Econometrics*, 2009, *150* (1), 41–55.

Bursztyn, L., F. Ederer, B. Ferman, and N. Yuchtman, "Understanding Mechanisms Underlying Peer Effects: Evidence From a Field Experiment on Financial Decisions," *Econometrica*, 2014, *82* (4), 1273–1301.

Cai, J., A. De Janvry, and E. Sadoulet, "Social Networks and the Decision to Insure," *American Economic Journal: Applied Economics*, 2015, *7* (2), 81–108.

Chin, A., "Central Limit Theorems via Stein's Method for Randomized Experiments Under Interference," *arXiv preprint arXiv:1804.03105*, 2018.

Choi, D., "Estimation of Monotone Treatment Effects in Network Experiments," *Journal of the American Statistical Association*, 2017, *112* (519), 1147–1155.

_ , "Using Exposure Mappings as Side Information in Experiments with Interference," *arXiv preprint arXiv:1806.11219*, 2018.

Doukhan, P. and S. Louhichi, "A New Weak Dependence Condition and Applications to Moment Inequalities," *Stochastic Processes and their Applications*, 1999, *84* (2), 313–342.

Eckles, D., B. Karrer, and J. Ugander, "Design and Analysis of Experiments in Networks: Reducing Bias from Interference," *Journal of Causal Inference*, 2017, *5* (1).

Forastiere, L., E. Airoldi, and F. Mealli, "Identification and Estimation of Treatment and Interference Effects in Observational Studies on Networks," *arXiv preprint arXiv:1609.06245*, 2016.

Friedrich, T., T. Sauerwald, and A. Stauffer, "Diameter and Broadcast Time of Random Geometric Graphs in Arbitrary Dimensions," *Algorithmica*, 2013, *67* (1), 65–88.

Granovetter, M., "Threshold Models of Collective Behavior," *American Journal of Sociology*, 1978, *83* (6), 1420–1443.

Guilbeault, D., J. Becker, and D. Centola, "Complex Contagions: A Decade in Review," in "Complex Spreading Phenomena in Social Systems," Springer, 2018, pp. 3–25.

He, X. and K. Song, "Measuring Diffusion over a Large Network," *arXiv preprint arXiv:1812.04195*, 2018.

Hoff, P., A. Raftery, and M. Handcock, "Latent Space Approaches to Social Network Analysis," *Journal of the American Statistical Association*, 2002, *97* (460), 1090–1098.

Hofstad, R. Van Der, *Random Graphs and Complex Networks*, Vol. 1, Cambridge University Press, 2016.

Hudgens, M. and M. Halloran, "Toward Causal Inference with Interference," *Journal of the American Statistical Association*, 2008, *103* (482), 832–842.

Imbens, G and D. Rubin, *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*, Cambridge University Press, 2015.

Imbens, G. and K. Menzel, "A Causal Bootstrap," *NBER No. w24833*, 2018.

Jackson, M., *Social and Economic Networks*, Princeton University Press, 2010.

Jenish, N. and I. Prucha, "Central Limit Theorems and Uniform Laws of Large Numbers for Arrays of Random Fields," *Journal of Econometrics*, 2009, *150* (1), 86–98.

Kojevnikov, D., "The Bootstrap for Network Dependent Processes," *University of British Columbia working paper*, 2019.

_ , V. Marmer, and K. Song, "Limit Theorems for Network Dependent Random Variables," *arXiv preprint arXiv:1903.01059*, 2019.

Lazzati, N., "Treatment Response with Social Interactions: Partial Identification via Monotone Comparative Statics," *Quantitative Economics*, 2015, *6* (1), 49–83.

Leung, M., "Causal Inference Under Approximate Neighborhood Interference," *arXiv preprint arXiv:1911.07085v1*, 2019.

_ , "Inference in Models of Discrete Choice with Social Interactions Using Network Data," *arXiv preprint arXiv:1911.07106*, 2019.

_ , "Treatment and Spillover Effects Under Network Interference," *Review of Economics and Statistics*, 2020, *102*, 368–380.

Manski, C., "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 1993, *60* (3), 531–542.

__, "Identification of Treatment Response with Social Interactions," *The Econometrics Journal*, 2013, *16* (1), S1–S23.

Miguel, E. and M. Kremer, "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica*, 2004, *72* (1), 159–217.

Milgrom, P. and J. Roberts, "Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities," *Econometrica*, 1990, pp. 1255–1277.

Montanari, A. and A. Saberi, "The Spread of Innovations in Social Networks," *Proceedings of the National Academy of Sciences*, 2010, *107* (47), 20196–20201.

Paluck, E., H. Shepherd, and P. Aronow, "Changing Climates of Conflict: A Social Network Experiment in 56 Schools," *Proceedings of the National Academy of Sciences*, 2016, *113* (3), 566–571.

Penrose, M., *Random Geometric Graphs*, Oxford University Press, 2003.

Sävje, F., "Causal Inference with Misspecified Exposure Mappings," *Yale University working paper*, 2019.

Sävje, F., P. Aronow, and M. Hudgens, "Average Treatment Effects in the Presence of Unknown Interference," *arXiv preprint arXiv:1711.06399*, 2017.

Shalizi, C. and A. Thomas, "Homophily and Contagion Are Generically Confounded in Observational Social Network Studies," *Sociological Methods and Research*, 2011, *40* (2), 211–239.

Toulis, P. and E. Kao, "Estimation of Causal Peer Influence Effects," in "International Conference on Machine Learning" 2013, pp. 1489–1497.

Vazquez-Bare, G., "Identification and Estimation of Spillover Effects in Randomized Experiments," *arXiv preprint arXiv:1711.02745*, 2017.

Viviano, D., "Policy Targeting under Network Interference," *arXiv preprint arXiv:1906.10258*, 2019.

Xu, X. and L. Lee, "Estimation of a Binary Choice Game Model with Network Links," *Ohio State University working paper*, 2015.