The Local Approach to Causal Inference under Network Interference

Eric Auerbach* Max Tabord-Meehan†

May 13, 2021

Abstract

We propose a new unified framework for causal inference when outcomes depend on how agents are linked in a social or economic network. Such network interference describes a large literature on treatment spillovers, social interactions, social learning, information diffusion, social capital formation, and more. Our approach works by first characterizing how an agent is linked in the network using the configuration of other agents and connections nearby as measured by path distance. The impact of a policy or treatment assignment is then learned by pooling outcome data across similarly configured agents. In the paper, we propose a new nonparametric modeling approach and consider two applications to causal inference. The first application is to testing policy irrelevance/no treatment effects. The second application is to estimating policy effects/treatment response. We conclude by evaluating the finite-sample properties of our estimation and inference procedures via simulation.

1 Introduction

Economists are often tasked with predicting agent outcomes under some counterfactual policy or treatment assignment. In many cases, the counterfactual depends on how the agents are linked in a social or economic network. A diverse literature on treatment spillovers, social interactions, social learning, information diffusion, social capital formation, and more has approached this problem from a variety of specialized, often highly-parametric frameworks (see for instance reviews by Athey and Imbens 2017; Jackson et al. 2017). In this paper, we

*Department of Economics, Northwestern University. E-mail: eric.auerbach@northwestern.edu.

†Department of Economics, University of Chicago. E-mail: maxtm@uchicago.edu. This paper supersedes the earlier working paper *A Nonparametric Network Regression.*

propose a unified framework for causal inference that accommodates many such examples of network interference.

Our main innovation is a new nonparametric modeling approach for sparse network data based on local configurations. Informally, a local configuration refers to the features of the network (the agents, their characteristics, treatment statuses, and how they are connected) nearby a focal agent as measured by path distance. The idea is that these local configurations index the different ways in which a policy or treatment assignment can impact a focal agent's outcome under network interference.

This modeling approach generalizes a developed literature on spillovers and social interactions in which the researcher specifies reference groups or an exposure map that details exactly how agents influence each other (see for instance Hudgens and Halloran 2008; Manski 2013; Aronow and Samii 2017). One potential limitation of this literature is that the results are often sensitive to exactly how the researcher models this dependence. For example, in the spillovers literature it is often assumed that agents respond to the average treatment of their peers, while in the diffusion literature agents may be informed or infected by any peer. And in the social learning literature agents may be influenced more by those peers that are also central in the network. When the researcher is uncertain as to exactly how agents influence each other, misspecification can lead to inaccurate estimates and invalid inferences.

Another potential limitation of this literature is that it does not generally consider policies that change the structure of the network. Network-altering policies are becoming increasingly relevant to economic research. Examples include those that add or remove agents, or connections between agents, from the community (see for instance, broadly, Ballester et al. 2006; Azoulay et al. 2010; Donaldson and Hornbeck 2016). Such policies may be difficult to evaluate using standard frameworks, which focus mostly on the reassignment of treatment statuses to agents while keeping the network structure fixed.

Our methodology addresses these limitations by using local configurations to model network interference. Intuitively, we use the space of local configurations as a "network sieve" that indexes the ways in which an agent may be affected by a given policy or treatment assignment. A key feature of local configurations is that they also naturally describe policies that alter network structure. A contribution of our work is to formalize this local approach and apply it to causal inference under network interference.

The use of local configurations in economics was pioneered by de Paula et al. (2018) (see also Anderson and Richards-Shubik 2019). In their work, local configurations (which they call network types) index moment conditions that partially identify the parameters of a strategic network formation model. The researcher has the flexibility to choose the configurations used in this task and can restrict attention to those that occur frequently in the data. In our setting, local configurations correspond to fixed counterfactual policies. It is usually the case that no exact instances of a given policy appear in the data and so we substitute outcomes associated with similar but not exactly the same configurations. Characterizing the resulting bias-variance trade-off requires additional machinery, which we introduce and formalize below by extending ideas from Benjamini and Schramm (2001).

We consider two causal inference problems. In both problems the researcher starts with a status-quo policy as described by one local configuration and is tasked with evaluating an alternative policy as described by a different local configuration. The researcher also has access to data from a many-networks experiment. Many-networks experimental designs are common in education, industrial organization, labor, and development economics where the researcher may collect network data on multiple independent schools, markets, firms, or villages. We argue that our framework can also be applied to other settings (for example, data on one large network), but leave formal extensions of our results to future work.

The first problem is a test of policy irrelevance (e.g. no treatment effects). For instance, the status-quo policy may be given by a particular social network structure where no agents are treated and the new policy may keep the same connections between agents but have every agent treated. The hypothesis to be tested is that both policies are associated with the same distribution of outcomes for one or more agents. We propose an asymptotically valid permutation test for this hypothesis building on work by Canay and Kamat (2018).

The second problem is estimating policy effects (e.g. average treatment response). For instance, the status-quo policy may be given by a particular social network structure and the new policy may be one in which a key agent is removed. The policy effect to be estimated is the expected change in outcomes for one or more agents. We propose a k-nearest-neighbors estimator for the policy effect and provide non-asymptotic bounds on mean-squared error building on work by Döring et al. (2017).

The remainder of this paper is organized as follows. Section 2 specifies a general model of network interference. We use this model to motivate our local approach, which is formally introduced in Section 3. Section 4 applies the local approach to test policy irrelevance and estimate policy effects. Section 5 contains simulation results and Section 6 concludes. Proof of claims and other details can be found in an appendix.

2 Setup and motivation

We introduce a general model of treatment response under network interference. This model is used to motivate our local approach in Section 3.

2.1 Terminology and notation

A countable population of agents is indexed by $\mathcal{I} \subseteq \mathbb{N}$. The population of agents are linked in a weighted and directed network. The weight of a link from agent i to j is given by $D_{ij} \in \mathbb{Z}_+ \cup \{\infty\}$. The (potentially infinite-dimensional) matrix D indexed by $\mathcal{I} \times \mathcal{I}$ with D_{ij} in the ijth entry is called the adjacency matrix. We take the convention that larger values of D_{ij} correspond to weaker relationships between i and j. For instance, D_{ij} might measure the physical distance between agents i and j. We suppose that $D_{ij} = 0$ if and only if $i = j$. When the network is unweighted (agent pairs are either linked or not) $D_{ij} = 1$ denotes a link and $D_{ij} = \infty$ denotes no link from agent i to j with $i \neq j$. We denote the set of all such matrices D by $\mathcal{D}$.

A path from agent i to j is a finite ordered subset of $\mathbb{N}$ denoted $\{t_1, ..., t_L\}$ with $t_1 = i$, $t_L = j$, and $L \in \mathbb{N}$. The length of the path $\{t_1, ..., t_L\}$ is given by $\sum_{s=1}^{L-1} D_{t_s t_{s+1}}$. The path distance from agent i to j, $\rho(i, j)$, is the length of the shortest path from i to j. That is,

$$\rho(i, j) := \inf_{\{t_1,...,t_L\} \subset \mathbb{N}} \left\{ \sum_{s=1}^{L-1} D_{t_s t_{s+1}} : t_1 = i, t_L = j \right\}.$$

For any $i \in \mathcal{I}$ and $r \in \mathbb{Z}_+$, agent i's r-neighborhood $\mathcal{N}_i(r) := \{j \in \mathcal{I} : \rho(i, j) \leq r\}$ is the collection of agents within path distance r of i. $N_i(r) := |\mathcal{N}_i(r)|$ is the size of agent

i's r-neighborhood (i.e. the number of agents in $\mathcal{N}_i(r)$). For any agent-specific variable $\mathbf{W} := \{W_i\}_{i\in\mathcal{I}}$ (see Section 2.2 below), $W_i(r) = \sum_{j\in\mathcal{I}} W_j \mathbb{1}\{\rho(i,j) \leq r\}$ is the r-neighborhood count of W for agent i. It describes the partial sum of W for the agents in $\mathcal{N}_i(r)$. Since $D_{ij} = 0$ if and only if $i = j$, $\mathcal{N}_i(0) = \{i\}$, $N_i(0) = 1$, and $W_i(0) = W_i$.

We assume that the network is *locally finite*. That is, for every $i \in \mathcal{I}$ and $r \in \mathbb{Z}_+$, $N_i(r) < \infty$ (see also de Paula et al. 2018). In words, the assumption is that every r-neighborhood of every agent contains only a finite number of agents. The assumption is implicit in much of the literature on network interference (including the examples below) where the researcher observes all of the relevant dependencies between agents in finite data.

2.2 General outcome model

Each agent $i \in \mathcal{I}$ has an outcome $Y_i \in \mathbb{R}$ and treatment assignment $T_i \in \mathbb{R}$. We call these quantities agent-specific variables and denote the corresponding population vectors $\mathbf{Y} = \{Y_i\}_{i\in\mathcal{I}}$, $\mathbf{T} = \{T_i\}_{i\in\mathcal{I}}$, etc. A general outcome model under network interference is

$$Y_i = f_i(\mathbf{T}, D, U_i) \ ,$$

where f_i is a real-valued function and $U_i \in \mathcal{U}$ represents unobserved agent-specific policy-invariant heterogeneity. We model Y_i as a function of $\mathbf{T}$ and D, so that individual i's outcome may vary with any of the treatments or network connections in the population. Let $\mathcal{T}$ denote the set of all population treatment vectors $\mathbf{T}$. The counterfactual outcome for agent i at a fixed policy choice $(\mathbf{t}, d) \in \mathcal{T} \times \mathcal{D}$ is $f_i(\mathbf{t}, d, U_i)$.

A limitation of this general model is that without further assumptions it is not informative about policy effects that rely on counterfactual outcomes (see Manski 2013, Section 1.2). This is because the model does not specify how data from one policy can be used to learn about the outcomes associated with a counterfactual policy. A common solution to this problem is to impose what Manski (2013) calls a *constant treatment response* (CTR) assumption (see also Aronow and Samii 2017; Basse et al. 2019; Hudgens and Halloran 2008; Vazquez-Bare 2017; Leung 2019; Sävje 2021, and others). Let $\lambda_i : \mathcal{T} \times \mathcal{D} \to \mathcal{G}$ be an *exposure map* that maps treatment and network information into an *effective treatment* for agent i, where $\mathcal{G}$

denotes the set of all effective treatments. The CTR assumption is that for $(\mathbf{t}, d)$ and $(\mathbf{t}', d')$ such that $\lambda_i(\mathbf{t}, d) = \lambda_i(\mathbf{t}', d')$

$$f_i(\mathbf{t}, d, u) = f_i(\mathbf{t}', d', u) \ ,$$

for all $u \in \mathcal{U}$. In words, the CTR assumption states that for a fixed $u \in \mathcal{U}$, all policies $(\mathbf{t}, d)$ associated with the same effective treatment $\lambda_i(\mathbf{t}, d)$ generate the same outcome $f_i(\mathbf{t}, d, u)$.

Under the CTR assumption, we define the function $h : \mathcal{G} \times \mathcal{U} \to \mathbb{R}$ such that $h(\lambda_i(\mathbf{t}, d), u) = f_i(\mathbf{t}, d, u)$, and rewrite our outcome model as

$$Y_i = h(G_i, U_i) \ ,$$

where $G_i = \lambda_i(\mathbf{T}, D)$. Agent i's outcome now only depends on the policy $(\mathbf{T}, D)$ through their effective treatment G_i.

The CTR assumption may help identify the policy effect of interest. A fixed policy implies a collection of effective treatments, one for each individual. If the effective treatments associated with a counterfactual policy are also observed in the data, then the researcher can potentially use those outcomes to characterize the counterfactual policy. For example, if $\lambda_i(\mathbf{T}, D) = T_i$, then agent i's outcome depends on only their own treatment status. The counterfactual outcome associated with treating an untreated agent might then be learned by looking at the outcomes of treated agents in the data.

Some common choices of λ_i and $\mathcal{G}$ are illustrated in Examples 2.1, 2.2, and 2.3 below. These choices are however based on strong modeling assumptions that when wrong may lead to inaccurate estimates and invalid inferences. Our local approach instead fixes a specific but flexible choice of $\mathcal{G}$ called the space of *rooted networks* which generalizes the notion of a local configuration and can approximate a large class of effective treatments. Under appropriate continuity assumptions on h, we propose estimating and inferring causal effects by pooling outcome data associated with effective treatments that are close to but not exactly the effective treatment of interest. Such extrapolation plays a key role in Example 2.3, where the effective treatment is not low-dimensional and so finding instances of a given effective treatment in the data is rare.

2.3 Examples

Example 2.1. (Neighborhood spillovers): Agents are assigned to either treatment or control status with $T_i = 1$ if i is treated and $T_i = 0$ if i is not. Agent i's potential outcome depends on their treatment status, and the number of treated agents proximate to i. A common neighborhood spillovers model is

$$Y_i = Y\left(T_i, S_i(r), U_i\right)$$

where $S_i(r) = \sum_{j \in \mathbb{N}} T_j \mathbb{1}\{\rho(i,j) \leq r\}$ are the neighborhood counts of treatment assigned within radius r of i. See for instance Cai et al. (2015); Leung (2016); Viviano (2019). A potential policy of interest is the effect of treating every agent in the community versus treating no one, holding the network connections fixed. In this example an effective treatment for i is given by $\lambda_i(\mathbf{T}, D) = (T_i, S_i(r))$. It only depends on the network connections and treatment statuses within path distance r of i.

Example 2.2. (Social capital formation): Agents leverage their position in the network to garner favors, loans, etc. Jackson et al. (2012) specify a model in which agents are linked in an unweighted and undirected network. Two agents exchange favors if there exists a third agent linked to both agents that can monitor the exchange. The number of favors exchanged by agent i is given by

$$Y_i = \left(\sum_{j \in \mathcal{I}} \mathbb{1}\{\mathcal{N}_i(1) \cap \mathcal{N}_j(1) \neq \emptyset\}\right) \cdot U_i \ .$$

A potential policy of interest in this literature is the effect of expanding agent i's 1-neighborhood to include an additional agent. In this example an effective treatment for i is given by $\lambda_i(\mathbf{T}, D) = \sum_{j \in \mathcal{I}} \mathbb{1}\{\mathcal{N}_i(1) \cap \mathcal{N}_j(1) \neq \emptyset\}$. It only depends on the network connections within path distance 2 of agent i. Karlan et al. (2009); Cruz et al. (2017) consider related models of social capital.

Example 2.3. (Social interactions): In the linear-in-means model, agent i's outcome depends on the total or average outcomes and treatment statuses of their peers. For instance

$$Y_i = T_i\beta + T_i^*(1)\gamma + \delta Y_i^*(1) + V_i$$

where $T_i^*(1) = T_i(1)/N_i(1)$. Outcomes are observed in equilibrium: agents first draw (T, V) and then coordinate on Y. Examples include Manski (1993); Lee (2007); Bramoullé et al. (2009); Blume et al. (2010); De Giorgi et al. (2010); Lee et al. (2010); Goldsmith-Pinkham and Imbens (2013). A potential policy effect of interest is the average effect of removing a particular agent from the community.

In this example we derive the effective treatment following Bramoullé et al. (2009). They show that under certain conditions the linear-in-means model admits a unique equilibrium

$$\begin{aligned} Y_i &= \left[(I - \delta A^*(D))^{-1}\left(\mathbf{T}\beta + \mathbf{T}^*(1)\gamma + \mathbf{V}\right)\right]_i \\ &= \lim_{S\to\infty}\sum_{s=0}^{S}\left[\delta^s A^*(D)^s\left(\mathbf{T}\beta + \mathbf{T}^*(1)\gamma + \mathbf{V}\right)\right]_i \ , \end{aligned}$$

where $[\cdot]_i$ is the ith entry of a vector, $\delta\rho < 1$, ρ is the spectral radius of $A^*(D)$, and the ijth entry of $A^*(D)$ is $A_{ij}^*(D) = \mathbb{1}\{0 < D_{ij} \leq 1\}/N_i(1)$. It follows that an effective treatment for i is

$$\lambda_i(\mathbf{T}, D) = \{[\delta^s A^*(D)^s]_{i\cdot}, [\delta^s A^*(D)^s\left(\mathbf{T}\beta + \mathbf{T}^*(1)\gamma\right)]_i\}_{s=0}^{\infty}$$

and $U_i = \mathbf{V}$ where $[\cdot]_{i\cdot}$ denotes the ith row of a matrix. The second element of the effective treatment is closely related to common measures of network centrality such as eigenvector, Katz-Bonacich, or diffusion centrality (see for instance Ballester et al. 2006; Calvó-Armengol et al. 2009; Banerjee et al. 2013)

One key difference between Example 2.3 and Examples 2.1 and 2.2 is that the effective treatment for i depends on all of the treatment statuses and network connections of the agents that are path connected to i. Leung (2019) shows that this is a common feature of many economic models of network interference. Our local approach will also be able to accommodate such dependence.

3 The local approach

We start with an informal description of the local configurations that form the basis of our approach, see also de Paula et al. (2018), Section 4. We then specify the model, extending ideas of Benjamini and Schramm (2001), and show how it generalizes the Section 2 examples.

3.1 Informal description

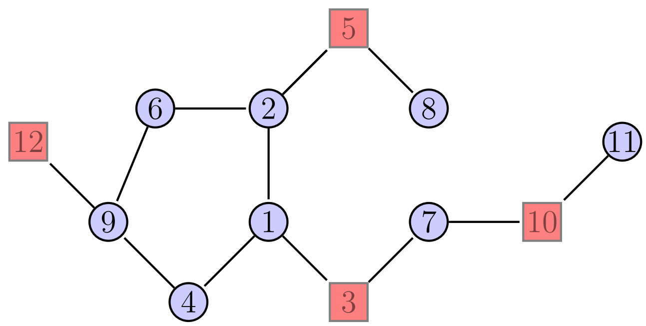
Intuitively, agent i's local configuration refers to the agents within path distance r of i, their characteristics, and how they are connected. Larger values of r are associated with more complicated configurations, which give a more precise picture of how the agent is connected in the network. This idea is illustrated in Figure 1.

Panel (a) depicts twelve agents connected in an unweighted and undirected network with binary treatment. Agents are either assigned to treatment (red square nodes) or control (blue circle nodes). Panel (b) depicts the local configurations of radius 1 for agents 1 and 2. They are both equivalent to a wheel with the focal untreated agent in the center and three other agents on the periphery, one of which is treated. Panel (c) depicts the local configurations of radius 2 for agents 1 and 2. They are both equivalent to a ring between five untreated agents (one of which is the focal agent). The focal agent is also connected to a treated agent who is connected to an untreated agent. Another agent in the ring adjacent to the focal agent is connected to a treated agent. Panel (d) depicts the local configurations of radius 3 for agents 1 and 2. They are not equivalent because the local configuration for agent 1 contains four treated agents while the local configuration for agent 2 contains only three treated agents.

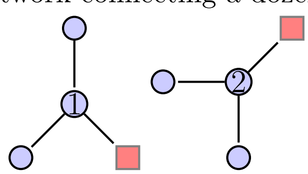
In this way, one can describe the local configurations for any choice of agent and radius. Since the diameter of the network (the maximum path distance between any two agents) is 7, any local configuration of radius greater than 7 will be equal to the local configuration of radius 7. However, for networks defined on large connected populations, increasing the radius of the local configuration typically reveals a more complicated network structure.

Benjamini and Schramm (2001) call the infinite-radius local configuration a rooted network (graph). Our local approach is based on the observation that, for many models with

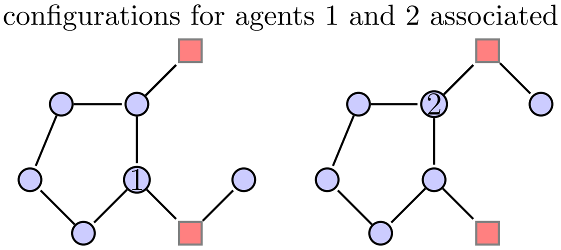
Figure 1: Illustration of local configurations.



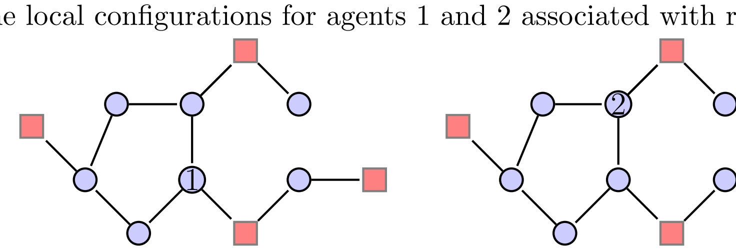
(a) A network connecting a dozen agents.



(b) The local configurations for agents 1 and 2 associated with radius 1.



(c) The local configurations for agents 1 and 2 associated with radius 2.



(d) The local configurations for agents 1 and 2 associated with radius 3.

network interference, an agent's effective treatment is determined by their rooted network.

For example, in the spillovers model of Example 2.1, an agent is influenced by their treatment status and the treatment statuses of their r-neighbors. As a result, the agent's effective treatment is determined by their local configuration of radius r. In the linear-in-means model of Example 2.3, an agent's effective treatment is not necessarily determined by any local configuration of finite radius. This is because an agent is influenced by all of the other agents to which they are path connected. However, the agent's effective treatment can be arbitrarily well-approximated by a local configuration of finite radius, and so the agent's effective treatment is determined by their rooted network. We show this in Example 3.3 below.

This observation motivates the main idea of this paper, which is to model network interference using rooted networks. We now formalize this idea.

3.2 Model specification

We first define the space of rooted networks which builds on Benjamini and Schramm (2001) (see also Aldous and Steele 2004, Section 2) and generalizes the idea of a local configuration from Section 3. We then show how rooted networks characterize the effective treatments from the Section 2 examples.

3.2.1 Rooted networks

We maintain the assumptions of Section 2, but introduce some new notation. For an adjacency matrix $D \in \mathcal{D}$ and treatment assignment vector $\mathbf{T} \in \mathcal{T}$, a network is the triple $(V, E, \mathbf{T})$ where V is the vertex set (the population of agents represented by $\mathcal{I}$) and E is the weighted edge set (represented by the adjacency matrix D). A rooted network $G_i = ((V, E, \mathbf{T}), i)$ is the triple $(V, E, \mathbf{T})$ with a focal agent $i \in V$ called the root. Informally, G_i is the network $(V, E, \mathbf{T})$ "from the point of view" of i.

For any $r \in \mathbb{Z}_+$, G_i^r is the subnetwork of $(V, E, \mathbf{T})$ rooted at i and induced by the agents within path distance r of i as measured by path distance ρ. Formally, $G_i^r := ((V_i^r, E_i^r, T_i^r), i)$, where $V_i^r := \mathcal{N}_i(r) := \{j \in V : \rho(i, j) \leq r\}$, $E_i^r := \{e_{jk} \in E : j, k \in V_i^r\}$, and $T_i^r := \{T_j \in$

$\mathbf{T} : j \in V_i^r\}$. We say G_i^r is the rooted network G_i truncated at radius r. In Section 3.1 we called this object agent i's local configuration of radius r.

For any $\varepsilon \geq 0$, two rooted networks G_{i_1} and G_{i_2} are ε-isomorphic (denoted $G_{i_1} \simeq_\varepsilon G_{i_2}$) if all of their r-neighborhoods are equivalent up to a relabeling of the non-rooted agents, but where treatments are allowed to disagree up to a tolerance of ε. Formally, $G_{i_1} \simeq_\varepsilon G_{i_2}$ if for any $r \in \mathbb{Z}_+$ there exists a bijection $f : V_{i_1}^r \leftrightarrow V_{i_2}^r$ such that $f(i_1) = i_2$, $e_{jk} = e_{f(j)f(k)}$, for any $j, k \in V_{i_1}^r$, and $|T_j - T_{f(j)}| \leq \varepsilon$ for any $j \in V_{i_1}^r$.

Two rooted networks that are not 0-isomorphic are assigned a strictly positive distance inversely related to the largest r and smallest ϵ such that they have ϵ-isomorphic r-neighborhoods. Formally, let $\zeta : \mathbb{R}_+ \to \mathbb{R}_{++}$ be strictly decreasing with $\lim_{x\to\infty} \zeta(x) = 0$ and $\zeta(0) = 1$ (for example, $\zeta(x) = (1 + x)^{-1}$). Then the following distance defines a pseudo-metric on the set of rooted networks:

$$d(G_{i_1}, G_{i_2}) := \min \left\{ \inf_{r\in\mathbb{Z}_+, \varepsilon\in\mathbb{R}_{++}} \{\zeta(r) + \varepsilon : G_{i_1}^r \simeq_\varepsilon G_{i_2}^r\}, \;\; 2 \right\} .$$

We demonstrate that $d(\cdot, \cdot)$ is in fact a pseudo-metric in Appendix A.1. The outer minimization is not essential to our analysis, but taking d to be bounded simplifies the exposition of Section 4.

Let $\mathcal{G}$ denote the set of equivalence classes of all possible rooted networks under d. We demonstrate that $(\mathcal{G}, d)$ is a separable and complete metric space in Appendix A.1. Following Aldous and Steele (2004), we call the topology on $\mathcal{G}$ induced by d the local topology, and more broadly call modeling on $\mathcal{G}$ the local approach.

3.2.2 Model of outcomes

Recall the model of outcomes derived in Section 2

$$Y_i = h(G_i, U_i) \;\; ,$$

where $G_i \in \mathcal{G}$ is a collection of effective treatments. We propose taking $\mathcal{G}$ to be the space of rooted networks, so that $G_i \in \mathcal{G}$ is the rooted network of agent i. In Section 3.3 below we revisit the examples of Section 2 and show that in each of the examples below G_i is an effective treatment for i.

We endow $\mathcal{G} \times \mathcal{U}$ with the usual product topology and define probability measures on the corresponding Borel sigma-algebra. We associate a stochastic rooted network and error pair (G_i, U_i) with a probability measure μ. For now we take μ as arbitrary and fixed by the researcher. We fix a specific choice of μ in the context of a many-networks experimental design in Section 4.2 below.

Our main objects of interest are the average structural function (ASF) or distributional structural function (DSF) that describe the outcome for i associated with a policy that sets the rooted network to some $g \in \mathcal{G}$. That is,

$$h(g) = E\left[h(g, U_i)\right] \quad \text{and}$$

$$h_y(g) = E\left[\mathbb{1}\{h(g, U_i) \leq y\}\right] \ ,$$

where the expectation refers to the marginal distribution of U_i under μ (see generally Blundell and Powell 2003). These functions can be used to estimate and conduct inferences about many causal effects of interest under network interference. For instance the average treatment effect (ATE) associated with a policy that alters agent i's rooted network from g to g' is described by $h(g') - h(g)$. This object is analogous to the usual ATE recovered from an experiment that describes the impact of assigning i to treatment or control status (see for instance Heckman and Robb 1985; Manski 1990; Imbens and Angrist 1994). Other causal effects can be defined analogously. A key message of our paper is that under appropriate continuity assumptions, $h(g)$ and $h_y(g)$ can be approximated by studying effective treatments which are close to but not exactly equal to g under d. We demonstrate this idea in two concrete applications involving a many-networks experimental design in Sections 4.3 and 4.4 below.

3.3 Examples Revisited

We revisit the examples of Section 2 and show that for each model rooted networks serve the role of an effective treatment.

Example 3.1. In the treatment spillovers model of Example 2.1

$$Y_i = Y\left(T_i, S_i(r), U_i\right)$$

with $S_i(r) = \sum_{j \in \mathbb{N}} T_j \mathbb{1}\{\rho(i,j) \leq r\}$. An effective treatment is $\lambda_i(\mathbf{T}, D) = (T_i, S_i(r))$. Let G_i be agent i's rooted network. T_i is associated with agent i and so is determined by G_i^0. $S_i(r)$ counts the number of treated agents within distance r of i and so is determined by G_i^r. As a result, $\lambda_i(\mathbf{T}, D)$ is determined by G_i. It follows that the outcome for agent i can be written as $Y_i = h(G_i, U_i)$ for some h.

Example 3.2. In the social capital formation model of Example 2.2

$$Y_i = \left(\sum_{j \in \mathcal{I}} \mathbb{1}\{\mathcal{N}_i(1) \cap \mathcal{N}_j(1) \neq \emptyset\}\right) \cdot U_i \ .$$

An effective treatment is $\lambda_i(\mathbf{T}, D) = \sum_{j \in \mathcal{I}} \mathbb{1}\{\mathcal{N}_i(1) \cap \mathcal{N}_j(1) \neq \emptyset\}$. Let G_i be i's rooted network. Here agent i's effective treatment is a function of the number of agents that are of path distance 2 from i, which is determined by G_i^2. It follows that the outcome of agent i can be written as $Y_i = h(G_i, U_i)$ for some h.

Example 3.3. In the social interactions model of Example 2.3, equilibrium outcomes are described by the model

$$Y_i = \lim_{S \to \infty} \sum_{s=0}^{S} \left[\delta^s A^*(D)^s \left(\mathbf{T}\beta + \mathbf{T}^*(1)\gamma + \mathbf{V}\right)\right]_i \ .$$

An effective treatment is $\lambda_i(\mathbf{T}, D) = \{[\delta^s A^*(D)^s]_{i\cdot}, [\delta^s A^*(D)^s \left(\mathbf{T}\beta + \mathbf{T}^*(1)\gamma\right)]_i\}_{s=0}^{\infty}$ and the structural error is $U_i = \mathbf{V}$. Let G_i be agent i 's rooted network. The ith row of $\delta^s A^*(D)^s$ and ith entry of $\delta^s A^*(D)^s \left(\mathbf{T}\beta + \mathbf{T}^*(1)\gamma\right)$ only depend on the treatment statuses and connections of agents within path-distance s of i and so is determined by G_i^s. It follows that

$$Y_i = \lim_{S \to \infty} \sum_{s=1}^{S} h_s(G_i^s, U_i) = h(G_i, U_i)$$

for some functions h_s and h.

4 Inference and estimation of causal effects

We apply the local approach framework of Section 3 to two causal inference problems. In both problems, the researcher begins with a status-quo policy (as described by an adjacency matrix and treatment assignment pair) and is tasked with evaluating the impact of an alternative. We assume that the researcher has access to data on outcomes and policies from a randomized experiment on many independent communities or clusters such as schools, villages, firms, or markets. This data structure, described in Section 4.1, is not crucial to our methodology, but simplifies the analysis. The study of alternative experimental designs, for example with data from one large network, or endogenous policies is left to future work.

Our first application, described in Section 4.2, is to testing policy irrelevance. Specifically, we test the hypothesis that the two policies are associated with the same distribution of outcomes. Our second application, described in Section 4.3, is to estimating policy effects. Specifically, we construct a k-nearest-neighbors estimator for the policy function of Section 3.2.2 and provide non-asymptotic bounds on estimation error.

These applications contrast a developed literature that studies the magnitude of any potential spillover effects or tests the hypothesis of no spillovers (see for instance Aronow 2012; Athey et al. 2018; Hu et al. 2021; Sävje et al. 2021). A recent literature also considers the estimation and inference of exposure effects where the estimand depends on both a true model of dependence and a potentially misspecified exposure map (see Leung 2019; Sävje 2021).

4.1 Many-networks setting

A random sample of communities or clusters are indexed by $t \in [T] := 1, \dots T$. Each $t \in [T]$ is associated with a finite collection of m_t observations $\{W_{it}\}_{i \in [m_t]}$, where m_t is a positive integer-valued random variable, $W_{it} := (Y_{it}, G_{it})$, and $Y_{it} := h(G_{it}, U_{it})$ for some unobserved error U_{it}. Intuitively, each community t is represented by an initial network connecting m_t agents, and the m_t rooted networks refer to this initial network rooted at each of agents in the community. Let $W_t := \{W_{it}\}_{i \in [m_t]}$. We impose the following assumptions on $\{(G_{it}, U_{it})\}_{i \in [m_t], t \in [T]}$.

Assumption 4.1.

(i) $\{W_t\}_{t\in[T]}$ *is independent and identically distributed (across communities).*

(ii) $\{U_{it}\}_{i\in[m_t],t\in[T]}$ *is identically distributed (within and across communities).* U *is an independent copy of* U_{it} *(that is,* U *has the same marginal distribution as* U_{it} *but is independent of* $\{W_t\}_{t\in[T]}$*).*

(iii) For any measurable f, $i \in [m_t]$, *and* $t \in [T]$,

$$E\left[f(G_{it}, U_{it})|G_{1t}, ..., G_{m_t t}\right] = E\left[f(G_{it}, U)|G_{it}\right] \ ,$$

assuming that the conditional expectations exist.

Assumption 4.1 (i) is what makes our analysis "many-networks." It states that the networks and errors are independent and identically distributed across communities. We exploit independence across communities to characterize the statistical properties of our test procedure and estimator below. We do not make any restrictions on the dependence structure between observations within a community. Weakening the independence assumption (for instance, considering dependent data from one large community) would require additional assumptions characterizing the intra-community dependence structure, which we leave to future work.

Assumption 4.1 (ii) fixes the marginal distribution of the errors. It is used to define the policy function and effects of interest. Specifically, an average policy effect is defined as the average outcome over the homogeneous marginal distribution of U_{it} for a fixed rooted network. This assumption can be dropped by defining the expectation to be with respect to the (mixture) distribution of $U_{\iota_t t}$ generated by drawing ι_t uniformly at random from $[m_t]$.

Assumption 4.1 (iii) states that the rooted networks are exogenous (i.e. the errors are policy-irrelevant). Specifically, we require that the conditional distribution of (G_{it}, U_{it}) given $G_{1t}, ..., G_{m_t t}$ is equal to the conditional distribution of (G_{it}, U) given G_{it}, where U is an independent copy of U_{it}. Exogeneity is a strong assumption, but allows us to approximate the unknown policy functions using sample averages (see below). It is also often assumed in the literature cited in Section 2.3. The study of endogenous rooted networks where the policy $(\mathbf{T}, D)$ potentially depends on the errors $\mathbf{U}$ is left to future work.

4.2 Testing policy irrelevance

The policy maker begins with a status-quo community policy described by some treatment and network pair $(\mathbf{t}, d)$, and proposes an alternative $(\mathbf{t}', d')$. The researcher is tasked with testing whether the two policies are associated with the same distribution of outcomes for agents whose effective treatment under the status-quo is described by the rooted network $g \in \mathcal{G}$ and whose effective treatment under the alternative is described by the rooted network $g' \in \mathcal{G}$. Following Section 3.2, the potential outcomes under treatments g and g' are given by $h(g, U)$ and $h(g', U)$ respectively for some error U. The null hypothesis of policy irrelevance is

$$H_0 : h_y(g) = h_y(g') \text{ for every } y \in \mathbb{R}, \tag{1}$$

where $h_y(g) := E\left[\mathbb{1}\{h(g,U)\} \leq y\right]$. Under Assumption 4.1 (iii), $h_y(g)$ describes the conditional distribution of Y_i given $G_i \simeq g$.

The proposed test is described in Section 4.2.2. Intuitively, it compares the empirical distribution of outcomes for the agents in the data whose rooted networks are most similar to g or g'. Asymptotic validity of the test relies on the following assumptions.

4.2.1 Assumptions

We impose smoothness conditions on the model parameters and suppose that the rooted networks g and g' are "supported" in the data. We do not believe these conditions to be restrictive in practice. Let $\psi_{\tilde{g}}(\ell) := P\left(\min_{i \in [m_t]} d(G_{it}, \tilde{g}) \leq \ell\right)$.

Assumption 4.2. *For every $\ell > 0$ and $g_0 \in \{g, g'\}$, $\psi_{g_0}(\ell) > 0$.*

The function $\psi_g(\ell)$ measures the probability that there exists an agent from a randomly drawn community whose rooted network is within distance ℓ of g. It partly determines the statistical properties of our test in this section and the estimator in Section 4.3.

Assumption 4.2 states that the nearest neighbor of g or g' from a randomly drawn community has a positive probability of being within distance ℓ of g or g'. The condition implies that as the number of communities T grows, the researcher will eventually observe agents whose rooted networks are arbitrarily close to g or g'.

Assumption 4.3. *For every $\tilde{g} \in \mathcal{G}$, the distribution of $h(\tilde{g}, U)$ is continuous, and for every $y \in \mathbb{R}$, $h_y(\tilde{g})$ is continuous at $\tilde{g}$.*

Assumption 4.3 states that the distribution of outcomes associated with agents whose rooted networks are similar to $\tilde{g}$ approximate the distribution of outcomes at $\tilde{g}$. Recall that this continuity assumption is satisfied by the three examples of Section 2.3.

4.2.2 Test procedure

We propose an approximate permutation test for H_0 building on Canay et al. (2017); Canay and Kamat (2018). The test procedure is described in Algorithm 4.1. We assume that T is even to simplify notation. When determining the closest agent in Step 2 or reordering the vectors in Step 3, ties are broken uniformly at random.

Algorithm 4.1. Input: data $\{W_t\}_{t\in[T]}$ and parameters $q \in [T/2]$, $\alpha \in [0,1]$. Output: a rejection decision.

1. Randomly partition the communities $[T]$ into two sets of size $T/2$, labelled $\mathcal{D}_1$ and $\mathcal{D}_2$.

2. For every $t \in \mathcal{D}_1$, let $i_t(g) := \text{argmin}_{i\in[m_t]}\{d(G_{it}, g)\}$ be the agent whose value of G_{it} is closest to g and $W_t(g) := (Y_t(g), G_t(g)) := (Y_{i_t(g)t}, G_{i_t(g)t})$ be the $i_t(g)$th outcome and rooted network. Similarly define $W_t(g') := (Y_t(g'), G_t(g'))$ for every $t \in \mathcal{D}_2$.

3. Reorder $\{W_t(g)\}_{t\in\mathcal{D}_1}$ and $\{W_t(g')\}_{t\in\mathcal{D}_2}$ so that the entries are increasing in $d(G_t(g), g)$ and $d(G_t(g')g')$ respectively. Denote the first q elements of the reordered $\{W_t(g)\}_{t\in\mathcal{D}_1}$

$$\begin{aligned} W^*(g) &:= W_1^*(g), W_2^*(g), ..., W_q^*(g) \\ &:= (Y_1^*(g), G_1^*(g)), (Y_2^*(g), G_2^*(g)), ..., (Y_q^*(g), G_q^*(g)). \end{aligned}$$

   Similarly define $W^*(g')$. Collect the $2q$ outcomes of $W^*(g)$ and $W^*(g')$ into the vector

$$S_T := (S_{T,1}, ..., S_{T,2q}) := \left(Y_1^*(g), \dots, Y_q^*(g), Y_1^*(g'), \dots, Y_q^*(g')\right) \ .$$

4. Define the Cramer von Mises test statistic

$$R(S_T) = \frac{1}{2q}\sum_{j=1}^{2q}\left(\hat{F}_1(S_{T,j};S_T) - \hat{F}_2(S_{T,j};S_T)\right)^2 ,$$

where

$$\hat{F}_1(y;S_T) = \frac{1}{q}\sum_{j=1}^{q}\mathbb{1}\{S_{T,j} \le y\} \text{ and } \hat{F}_2(y;S_T) = \frac{1}{q}\sum_{j=q+1}^{2q}\mathbb{1}\{S_{T,j} \le y\} .$$

5. Let $\mathbf{G}$ be the set of all permutations $\pi = (\pi(1), \ldots, \pi(2q))$ of $\{1, ..., 2q\}$ and

$$S_T^{\pi} = (S_{T,\pi(1)}, ..., S_{T,\pi(2q)}) .$$

Reject H_0 if the p-value $p \le \alpha$ where $p := \frac{1}{|\mathbf{G}|}\sum_{\pi\in\mathbf{G}}\mathbb{1}\{R(S_T^{\pi}) \ge R(S_T)\}$.

In some cases the p-value in Step 5 may be difficult to compute exactly because the set $\mathbf{G}$ is intractably large. It can be shown that the result below continues to hold if $\mathbf{G}$ is replaced by $\hat{\mathbf{G}}$ where $\hat{\mathbf{G}} = \{\pi_1, ..., \pi_B\}$, π_1 is the identity permutation, and $\pi_2, ..., \pi_B$ are drawn independently and uniformly at random from $\mathbf{G}$. This sampling is standard in the literature, see also Canay and Kamat (2018), Remark 3.2.

The test presented in Algorithm 4.1 is a non-randomized version of the permutation test, in the sense that the decision to reject the null hypothesis is a deterministic function of the data. This leads to a test which is potentially conservative. We could alternatively consider a non-conservative version of this test which is randomized (see for example Lehmann and Romano 2006, Section 15.2).

4.2.3 Asymptotic validity

If the entries of $Y^*(g)$ and $Y^*(g')$ were identically distributed to $h(g, U)$ and $h(g', U)$ respectively, then the test described in Algorithm 4.1 would control size in finite samples following standard arguments (see for instance Lehmann and Romano 2006, Theorem 15.2.1). However, since the entries of $G^*(g)$ and $G^*(g')$ are not exactly g and g', this condition is generally false.

Our assumptions instead imply that in an asymptotic regime where q is fixed and $T \to \infty$, the entries of $Y^*(g)$ and $Y^*(g')$ are approximately equal in distribution to $h(g, U)$ and $h(g', U)$. A fixed q rule is appropriate in our setting because the quality of the nearest neighbors $G^*(g)$ and $G^*(g')$ in terms similarity to g or g' may degrade rapidly with q. This is shown via simulation in Section 5.4 below.

We demonstrate that the test procedure in Algorithm 4.1 is asymptotically valid using the framework of Canay et al. (2017); Canay and Kamat (2018). Finite-sample behavior is examined via simulation in Section 5.

Theorem 4.1. *Under Assumptions 4.1, 4.2 and 4.3, the test described in Algorithm 4.1 is asymptotically ($T \to \infty$, q fixed) level α.*

4.3 Estimating Policy Effects

The policy maker begins with a status-quo community policy described by some treatment and network pair $(\mathbf{t}, d)$, and proposes an alternative $(\mathbf{t}', d')$. The researcher is tasked with estimating the average effect of the policy change for agents whose effective treatment under the status-quo is described by the rooted network $g \in \mathcal{G}$ and whose effective treatment under the alternative is described by the rooted network $g' \in \mathcal{G}$. Following Section 3.2, the potential outcomes under policies g and g' are described by $h(g, U)$ and $h(g', U)$ respectively for some error U. The average policy effect of interest is

$$h(g') - h(g)$$

where $h(g) = E\left[h(g, U)\right]$. Alternative effects (e.g. distributional effects) based on other features of the conditional distribution of outcomes can be estimated analogously.

The proposed estimator for the policy effect is described in Section 4.3.2. Intuitively, the estimator averages the outcomes of agents whose rooted networks are most similar to g or g'. Our bounds on estimation error rely on the following assumptions.

4.3.1 Assumptions

We impose smoothness conditions on the model parameters and bound the variance of the outcome. We do not believe these assumptions to be restrictive in practice. Let $\psi_{\tilde{g}}(\ell) := P\left(\min_{i\in[m_t]} d(G_{it}, \tilde{g}) \leq \ell\right)$.

Assumption 4.4. *For every $\ell \geq 0$ and $g_0 \in \{g, g'\}$, $\psi_{\tilde{g}}(\ell)$ is continuous at g_0.*

Recall from Section 4.2 that the function $\psi_{\tilde{g}}(\ell)$ measures the probability that there exists an agent from a randomly drawn community whose rooted network is within distance ℓ of $\tilde{g}$. Intuitively, $\psi_{\tilde{g}}$ measures the amount of regularity in the community. It plays a key role in determining the statistical properties of our estimator.

Assumption 4.4 states that $\psi_{\tilde{g}}(\ell)$ is continuous at g and g'. It justifies a probability integral transform used to characterize the bias of the estimator. If the data are such that Assumption 4.4 does not hold, then we can recover the assumption by simply adding a randomizing component to the metric (see the discussion following equation (19) in Györfi and Weiss 2020, for details).

Assumption 4.5. *For each $g_0 \in \{g, g'\}$ there exists an increasing $\phi_{g_0} : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\phi_{g_0}(x) \to \phi_{g_0}(0) = 0$ as $x \to 0$ and for every $\tilde{g} \in \mathcal{G}$*

$$|h(g_0) - h(\tilde{g})| \leq \phi_{g_0}\left(d(g_0, \tilde{g})\right) \ .$$

Assumption 4.5 states that the regression function h has a modulus of continuity ϕ_g at g. This restriction is analogous to smoothness assumptions on the conditional mean function that is common in the nonparametric estimation literature. A model of network interference often implies a specific choice of ϕ_g. See Section 4.3.4 below.

Assumption 4.6. *For every $\tilde{g} \in \mathcal{G}$, $\sigma^2(\tilde{g}) := E\left[(h(\tilde{g}, U_{it}) - h(\tilde{g}))^2\right] \leq \sigma^2$.*

Assumption 4.6 bounds the variance of the outcome variable.

4.3.2 Estimator

Let $W_t(g)$ be the observation from $\{W_{it}\}_{1\leq i\leq m}$ whose value of G_{it} is closest to g (see also Algorithm 4.1, Step 2). Order the elements of $\{W_t(g)\}_{1\leq t\leq T}$ to be increasing in $d(G_t(g), g)$, and denote the k closest values

$$W_1^*(g), W_2^*(g), \ldots, W_k^*(g).$$

The proposed estimator for $h(g)$ is the sample mean of the outcomes $\{Y_j^*(g)\}_{j=1}^k$ associated with $\{W_j^*(g)\}_{j=1}^k$

$$\hat{h}(g) := \frac{1}{k}\sum_{j=1}^{k} Y_j^*(g)$$

and the estimator for the average policy effect $h(g') - h(g)$ is the difference

$$\hat{h}(g') - \hat{h}(g) := \frac{1}{k}\sum_{j=1}^{k} \left(Y_j^*(g') - Y_j^*(g)\right).$$

4.3.3 Bound on estimation error

We derive a finite-sample bound on the mean-squared error of $\hat{h}(g)$ using the framework of Biau and Devroye (2015); Döring et al. (2017); Györfi and Weiss (2020).

Theorem 4.2. *Under Assumptions 4.1, 4.4, 4.5, and 4.6,*

$$E\left[\left(\hat{h}(g) - h(g)\right)^2\right] \leq \frac{\sigma^2}{k} + E\left[\varphi_g(U_{(k,T)})^2\right] ,$$

where $\varphi_g(x) = \phi_g \circ \psi_g^\dagger(x)$, $\psi_g^\dagger : [0,1] \to \mathbb{R}_+$ *refers to the upper generalized inverse*

$$\psi_g^\dagger(x) = \sup\{\ell \in \mathbb{R}_+ : \psi_g(\ell) \leq x\} ,$$

and $U_{(k,T)}$ *is distributed* $Beta(k, T-k+1)$.

The bound in Theorem 4.2 features a familiar bias-variance decomposition. The variance component σ^2/k is standard and decreases as k grows large. In contrast, the bias component $E[\varphi_g(U_{(k,T)})^2]$ and its relationship with k are difficult to characterize without further information about ϕ_g and ψ_g. We discuss how the bound depends on features of these parameters in Section 4.3.4 below.

Theorem 4.2 has the immediate corollary

Corollary 4.1. *Suppose the hypothesis of Theorem 4.2. Then*

$$E\left[\left(\hat{h}(g') - \hat{h}(g) - (h(g') - h(g))\right)^2\right] \leq \frac{4\sigma^2}{k} + 4E\left[\varphi_{g\vee g'}(U_{(k,T)})^2\right]$$

where $\varphi_{g\vee g'} := \varphi_g \vee \varphi_{g'}$.

One can potentially improve this bound via sample splitting. This is left to future work.

4.3.4 Interpreting the bias

Theorem 4.2 establishes consistency for our estimator of the policy function in well-behaved settings. For example if (for g fixed) $\varphi_g(\cdot)$ is bounded, continuous at zero, $\varphi_g(0) = 0$, and $k \to \infty$, $k/T \to 0$ as $T \to \infty$ then

$$E[\varphi_g(U_{(k,T)})^2] = E\left[\left(\phi_g \circ \psi_g^{\dagger}(U_{(k,T)})\right)^2\right] \to 0 \ .$$

Further characterization of the bias requires additional information about the components ϕ_g and ψ_g. Intuitively, the first controls the smoothness of the regression function h_g and the second controls the quality of the nearest-neighbors that make up $\hat{h}(g)$ in terms of proximity to g. This subsection provides an analytical discussion. Supporting simulation evidence can be found in Section 5.

The continuity parameter ϕ_g is often relatively easy to characterize because many models of network interference give an explicit bound. In particular, for our three examples in Section 2.3, $\phi_g(x)$ quickly converges to 0 with x for any $g \in \mathcal{G}$. In the neighborhood spillovers model of Example 2.1 with binary treatments and uniformly bounded expected outcomes, $\phi_g(x) \leq C\mathbb{1}\{x > \zeta(r)\}$ where $C = \sup_{g\in\mathcal{G}} 2h(g)$, because if $d(g, \tilde{g}) \leq \zeta(r)$ then $h(g) = h(\tilde{g})$. Similarly, in the social capital model of Example 2.2 with uniformly bounded 2-neighborhoods, $\phi_g(x) \leq C\mathbb{1}\{x > \zeta(2)\}$ where C bounds the number of agents within path distance 2 of the root agent.

In the linear-in-means peer effects model of Example 2.3 with binary treatments and uniformly bounded 1-neighborhood treatment counts, $\phi_g(x) \leq \frac{C(\delta\rho)^{\zeta^{\dagger}(x)}}{1-\delta\rho}$ where ρ bounds the spectral radius of $A^*(D)$, $|\delta\rho| < 1$ by assumption, and $C = \sup_{i\in V(\tilde{g}),\tilde{g}\in\mathcal{G}} 2\,|T_i\beta + T_i^*(1)\gamma|$.

This is because if $d(g, \tilde{g}) \leq \zeta(r)$ then $h^s(g) = h^s(\tilde{g})$ for every $s \leq r$ and the remainder term in the policy function $\left|\sum_{s=r+1}^{\infty}[\delta^s A^*(D)^s(\mathbf{T}\beta + \mathbf{T}^*(1)\gamma)]_i\right| \leq C\sum_{s=r+1}^{\infty}\delta^s\rho^s \leq \frac{C(\delta\rho)^r}{1-\delta\rho}$ for any $g \in \mathcal{G}$. See also Leung (2019).

In contrast to these explicit bounds on ϕ_g, we do not know of any convenient analytical way to characterize the regularity parameter ψ_g even for relatively simple models of link formation. If the network is sparse or has a predictable structure (agents interact in small groups or on a regular lattice), then the rooted network variable may essentially act like a discrete random variable, and so $\psi_g(\ell)$ may be uniformly bounded away from 0 (for $\ell > 0$ fixed). Unfortunately, such regularity rarely describes social or economic network data.

Irregular network formation models are more common. For example, a large literature considers models of network formation in which connections between agents are conditionally independent across agent-pairs. Examples include the Erdös-Renyi model, latent space model, and stochastic blockmodel (see generally Graham 2019). For such models, the implicit ψ_g function can change dramatically with the model parameters. We demonstrate some example ψ_g functions for the special case of the Erdos-Renyi model in Section 5 below.

5 Simulation evidence

Section 5.1 describes the simulation design. Section 5.2 gives results for the first application testing policy irrelevance. Section 5.3 gives results for the second application estimating policy effects.

5.1 Simulation design

We simulate data from T communities, where T is specified below. Each community contains 20 agents. Links between agents are drawn from an Erdös-Renyi model with parameter 0.1. That is, links between agent-pairs are independent and identically distributed $Bernoulli(.1)$ random variables. The Erdös-Renyi model is chosen not because it generates realistic-looking network data (see Jackson and Rogers 2007) but because a large class of rooted networks occur with non-trivial probability. This design choice is unfavorable to our method, which prefers models that reliably generate a small number of rooted network motifs (see for instance de Paula et al. 2018). One can instead think of the Erdös-Renyi model as the policy maker assigning network policies to communities at random. We discuss this model in more detail in Section 5.4.

In this relatively simple design the metric presented in Section 3 also simplifies. The distance between two rooted networks is simply $\zeta(r)$ where r is the largest radius such that the networks truncated at radius r are 0-isomorphic. In everything that follows we take $\zeta(r) = (1+r)^{-1}$.

Outcomes are generated for each agent $i \in [20]$ in each community $t \in [T]$ according to the model

$$Y_{it} = \alpha_1 f(G^1_{it}) + \alpha_2 f(G^2_{it}) + U_{it} \ ,$$

where $\alpha = (\alpha_1, \alpha_2) \in \mathbb{R}^2$ is specified below,

$$f(g) := deg(g) + 2clust(g) \ ,$$

$$deg(g) := \frac{1}{|V(g)|} \sum_{i \in V(g)} \sum_{j \in V(g)} \mathbb{1}\{ij \in E(g)\}$$

measures the average degree of the network $(V(g), E(g))$ and

$$clust(g) := \frac{1}{|V(g)|} \sum_{i \in V(g)} \frac{\sum_{j \in V(g)} \sum_{k \in V(g)} \mathbb{1}\{ij, ik, jk \in E(g)\}}{\sum_{j \in V(g)} \sum_{k \in V(g)} \mathbb{1}\{ik, jk \in E(g)\}}$$

measures the average clustering of the network $(V(g), E(g))$, and $U_{it} \sim U[-5, 5]$ independent of G_{it}. Our focus on degree and clustering statistics is meant to mimic the first two Examples of Section 2.3. That is, these network statistics are determined by the rooted network truncated at the first or second neighborhood.

5.2 Testing policy irrelevance

We first evaluate how the test procedure outlined in Algorithm 4.1 of Section 4.2.2 controls size when the null hypothesis of policy irrelevance is true. The choice of α we consider is $(0, 2)$. The choice of rooted networks (policies) we consider is represented by g_5 and g_6 in Figure 2. These networks are chosen because under the model in Section 5.1 the condi-

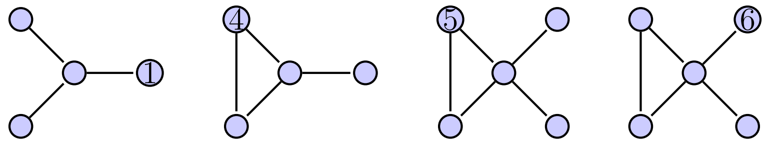


Figure 2: Four rooted networks truncated at radius 2, labeled g_1 and g_4 to g_6.

tional distribution of outcomes associated with g_5 and g_6 are the same, but the conditional distribution of outcomes associated with g_5^1 and g_6^1 are very different. The idea of this design is to illustrate the potential size distortion due to the fact that the permutation test is approximate. This can be seen in Table 1.

Columns 2-4 of Table 1 depict the results of 1000 simulations for $T \in \{20, 50, 100, 200\}$ communities and test parameters $q \in \{5, 8, 10\}$ and $\alpha = 0.05$. The results show that the test rejects the null hypothesis with probability approximately equal to α when q is small ($q = 5$) or T is large ($T = 200$). Size distortion occurs when q is large and T is small ($q \geq 8$ and $T \leq 50$).

To evaluate the power properties of the test procedure, we consider two rooted networks g_1 and g_4 that are associated with two different conditional distributions of outcomes under the model in Section 5.1. These two networks are shown in Figure 2.

Columns 5-7 depict the results for the same simulations as Columns 2-4 but for the test based on g_1 and g_4 instead of g_5 and g_6. The results show that the test correctly rejects the null hypothesis with probability greater than α. The probability of rejection generally increases with q (except for $T = 20$) at the cost of potential size distortions. Overall, our results suggest that unless a researcher has additional information about the structure of network interference relative to the quality of matches, they should not choose q to be too large.

5.3 Estimating policy effects

We study the mean-squared error of the k-nearest-neighbor estimator for the policy function h given in Section 4.3.2. for four rooted networks and $\alpha \in \{(1, 0), (1, 1/2)\}$. Under $\alpha = (1, 0)$ the distribution of outcomes depends on the features of the network within radius 1 of the root. Under $\alpha = (1, 1/2)$ the distribution of outcomes also depends on the features of the

Table 1: Rejection probabilities: $\alpha = 0.05$ ($1,000$ Monte Carlo iterations)

	$H_0 : g_5 =_d g_6$			$H_0 : g_1 =_d g_4$		
	q			q		
T	5	8	10	5	8	10
20	6.6	8.0	7.0	13.8	10.9	9.9
50	4.9	8.5	10.9	20.4	29.9	33.8
100	4.8	6.3	7.5	20.8	32.3	41.1
200	3.3	4.5	4.9	21.1	34.3	44.5

network within radius 2 of the root.

The choice of rooted networks we consider is represented by g_1 to g_4 in Figure 3. Networks g_1 and g_2 depict two wheels with the rooted agent on the periphery. These networks have moderate average degree and no average clustering: $(3/2, 0)$ and $(5/3, 0)$ respectively. Network g_3 depicts a closed triangle connected to a wheel with the rooted agent both on the periphery of the wheel and part of the triangle. This network has moderate average degree and average clustering $(2, 1/3)$. Finally, g_4 depicts a closed triangle connected to a single agent. This network has moderate average degree and high average clustering $(2, 7/12)$.

The results of the simulation are shown in Table 2. As suggested by Theorem 4.2, mean-squared error is generally decreasing with T for a fixed choice of k. In addition, mean-square error is generally smaller for the $\alpha = (1, 0)$ experiment than it is for the $\alpha = (1, 0.5)$ experiment for a fixed choice of T and k. The effect is more pronounced for the networks g_3 and g_4, for which we typically observe fewer good matches in the data compared to g_1 and g_2 (we quantify this observation by estimating ψ_g for each of the four networks in Section 5.4). This is also consistent with Theorem 4.2.

Fixing T and comparing across k, we expect a potential bias-variance trade-off. For networks g_1 and g_2, there is no meaningful bias in the estimated policy function because $f(\tilde{g}^1)$ and $f(\tilde{g}^2)$ are similar and $\psi_{\tilde{g}}(1) \approx 1$ for $\tilde{g} = g_1, g_2$ (see Section 5.4 below). As a result, it is optimal to use the nearest-neighbor from every community in $[T]$ (i.e. choose $k = T$). In contrast for g_3 and g_4 the rooted networks of the nearest neighbors in each community may be very different from the relevant policies, and so setting $k = T$ can lead to an inflated mean-squared error.

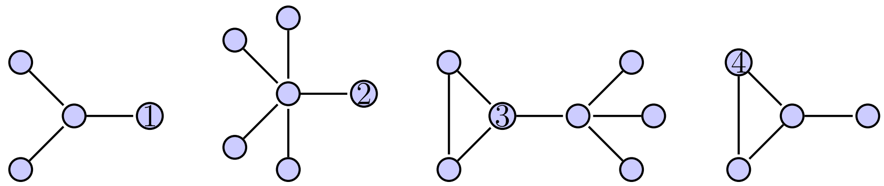


Figure 3: Four rooted networks truncated at radius 2, labeled g_1 to g_4.

We conclude that unless the researcher has additional information about the structure of network interference or the density of the policies of interest, k should not be large relative to the sample size. This is also consistent with our findings for the test of policy irrelevance in Section 5.2.

Table 2: Estimated MSEs (1,000 Monte Carlo iterations)

		$T=20$ k			$T=50$ k				$T=100$ k					
α	g_ι	5	10	20	5	10	20	50	5	10	20	50	75	100
$(1,0)$	g_1	1.71	0.84	0.41	1.69	0.86	0.43	0.18	1.63	0.81	0.41	0.17	0.11	0.08
	g_2	1.7	0.84	0.42	1.75	0.84	0.44	0.17	1.66	0.85	0.42	0.17	0.12	0.09
	g_3	1.77	0.98	1.41	1.6	0.81	0.42	1.17	1.74	0.84	0.44	0.2	0.59	1.09
	g_4	1.77	1.53	3.27	1.65	0.89	0.55	3.02	1.64	0.84	0.42	0.64	1.92	2.97
$(1,0.5)$	g_1	1.71	0.84	0.41	1.69	0.86	0.43	0.18	1.63	0.81	0.41	0.17	0.11	0.09
	g_2	1.7	0.84	0.42	1.75	0.84	0.44	0.17	1.66	0.86	0.42	0.17	0.12	0.09
	g_3	1.79	1.05	2.07	1.6	0.81	0.42	1.81	1.75	0.84	0.44	0.21	0.87	1.73
	g_4	1.85	2.15	5.37	1.66	0.9	0.72	5.09	1.65	0.86	0.44	1.1	3.32	5.04

5.4 Measuring network regularity

In Section 4.3 we identified the function $\psi_g(\ell)$ as a key determinant of the estimation bias in Theorem 4.2. This function measures the probability that the nearest-neighbor of g from a randomly drawn network is within distance ℓ of g. Recall that in our simulations we used $\zeta(r) = (1+r)^{-1}$ in our definition of the metric.

Figure 4 displays estimates of the ψ_g function for the four rooted networks considered in the simulation design of Section 5.3. The figures were constructed by generating 3000 Erdös-Renyi(0.1) random graphs with 20 nodes each and recording the distances of the nearest neighbor to g in each graph.

The results indicate that the nearest neighbor of g_1 always matches at least at a radius of 1, and often matches at a radius of 2. In contrast g_4 is only matched at a radius of 1

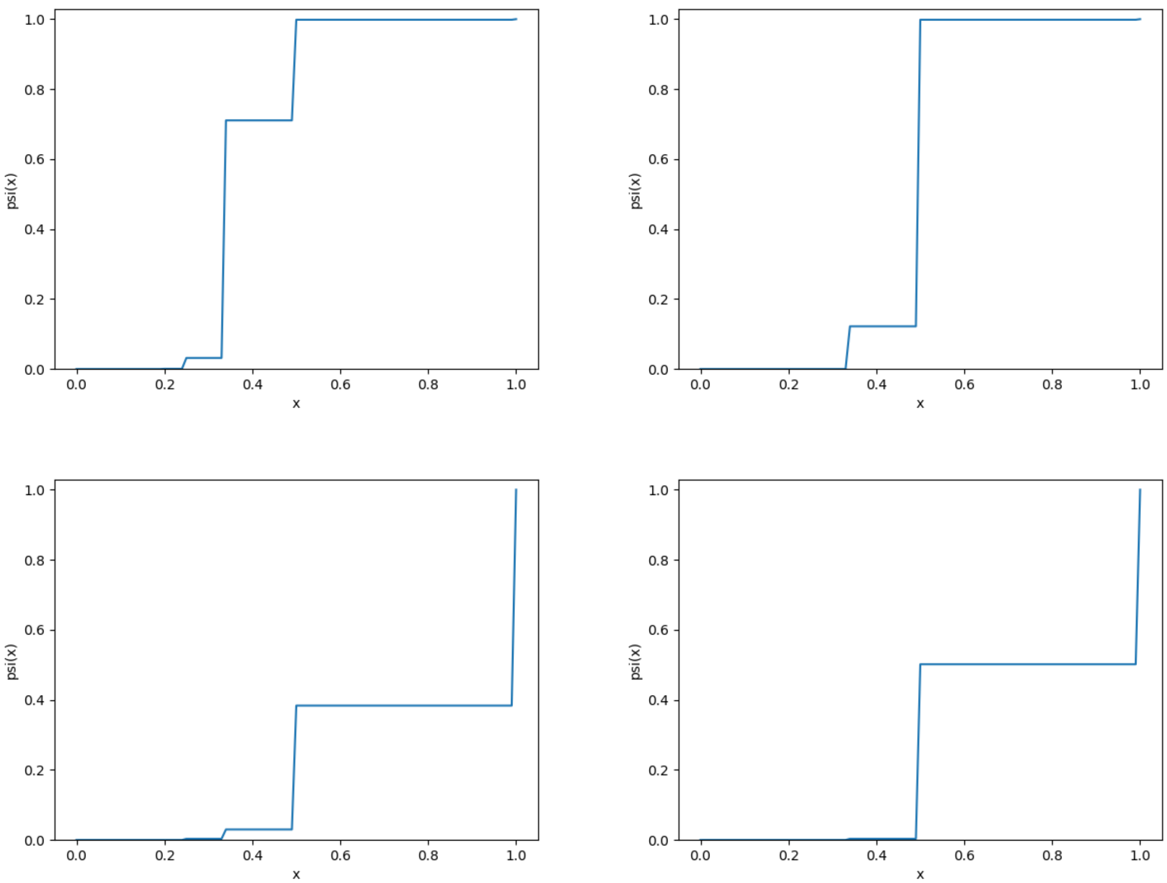


Figure 4: Estimated $\psi_g(\cdot)$ for g_1 to g_4 (from left to right, top to bottom).

in 40% of networks, and is almost never matched at a radius of 2. Intuitively, networks g_3 and g_4 are rare because triadic closure is uncommon under the random graph model. The network g_2 ia also relatively rare because the coincidence of five agents linked to a common agent is uncommon for such a sparse random network.

We remark that these results are the extreme case of a model of pure statistical noise. Strategic interaction between agents may potentially discipline the regularity of the network, particularly if only a small number of configurations are stable. Characterizing ψ_g for such strategic network formation models is an important area for future work.

6 Conclusion

This paper proposes a new unified framework for causal inference under network interference. We propose a model in which the impact of the policy on individual outcomes is indexed by rooted networks (also local configurations or network types). The model generalizes several popular specifications from the literature on treatment spillovers, social interactions, social learning, information diffusion, and social capital formation. We use the model to construct a test for policy irrelevance and estimates of policy effects. Some finite sample properties of the test are illustrated by simulation.

Much work remains to be done. A particularly interesting direction for future work would be to apply our methodology to the problem of policy learning under network interference (see for example Viviano 2019; Kitagawa and Wang 2020, Ananth 2020).

References

Aldous, David and J Michael Steele, "The objective method: probabilistic combinatorial optimization and local weak convergence," in "Probability on discrete structures," Springer, 2004, pp. 1–72.

Anderson, KA and S Richards-Shubik, "Collaborative production in science: An empirical analysis of coauthorships in economics," *CMU and Lehigh University Working Paper*, 2019.

Aronow, Peter M, "A general method for detecting interference between units in randomized experiments," *Sociological Methods & Research*, 2012, *41* (1), 3–16.

_ **and Cyrus Samii**, "Estimating average causal effects under general interference, with application to a social network experiment," *The Annals of Applied Statistics*, 2017, *11* (4), 1912–1947.

Athey, Susan and Guido W Imbens, "The state of applied econometrics: Causality and policy evaluation," *Journal of Economic Perspectives*, 2017, *31* (2), 3–32.

— , **Dean Eckles, and Guido W Imbens**, "Exact p-values for network interference," *Journal of the American Statistical Association*, 2018, *113* (521), 230–240.

Azoulay, Pierre, Joshua S Graff Zivin, and Jialan Wang, "Superstar Extinction," *The Quarterly Journal of Economics*, 2010, *125* (2), 549–589.

Ballester, Coralio, Antoni Calvó-Armengol, and Yves Zenou, "Who's who in networks. wanted: the key player," *Econometrica*, 2006, *74* (5), 1403–1417.

Banerjee, Abhijit, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson, "The diffusion of microfinance," *Science*, 2013, *341* (6144), 1236498.

Basse, Guillaume W, Avi Feller, and Panos Toulis, "Randomization tests of causal effects under interference," *Biometrika*, 2019, *106* (2), 487–494.

Benjamini, Itai and Oded Schramm, "Recurrence of Distributional Limits of Finite Planar Graphs," *Electronic Journal of Probability*, 2001, *6.*

Biau, Gérard and Luc Devroye, *Lectures on the nearest neighbor method*, Vol. 246, Springer, 2015.

Blume, Lawrence E, William A Brock, Steven N Durlauf, and Yannis M Ioannides, "Identification of social interactions," *Available at SSRN 1660002*, 2010.

Blundell, Richard and James L Powell, "Endogeneity in nonparametric and semiparametric regression models," *Econometric Society Monographs*, 2003, *36*, 312–357.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin, "Identification of peer effects through social networks," *Journal of econometrics*, 2009, *150* (1), 41–55.

Cai, Jing, Alain De Janvry, and Elisabeth Sadoulet, "Social networks and the decision to insure," *American Economic Journal: Applied Economics*, 2015, *7* (2), 81–108.

Calvó-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou, "Peer effects and social networks in education," *The Review of Economic Studies*, 2009, *76* (4), 1239–1267.

Canay, Ivan A and Vishal Kamat, "Approximate permutation tests and induced order statistics in the regression discontinuity design," *The Review of Economic Studies*, 2018, *85* (3), 1577–1608.

_ **, Joseph P Romano, and Azeem M Shaikh**, "Randomization tests under an approximate symmetry assumption," *Econometrica*, 2017, *85* (3), 1013–1030.

Cruz, Cesi, Julien Labonne, and Pablo Querubin, "Politician family networks and electoral outcomes: Evidence from the Philippines," *American Economic Review*, 2017, *107* (10), 3006–37.

de Paula, Áureo, Seth Richards-Shubik, and Elie Tamer, "Identifying preferences in networks with bounded degree," *Econometrica*, 2018, *86* (1), 263–288.

Donaldson, Dave and Richard Hornbeck, "Railroads and American economic growth: A "market access" approach," *The Quarterly Journal of Economics*, 2016, *131* (2), 799–858.

Döring, Maik, László Györfi, and Harro Walk, "Rate of convergence of k-nearest-neighbor classification rule," *The Journal of Machine Learning Research*, 2017, *18* (1), 8485–8500.

fei Lee, Lung, "Identification and estimation of econometric models with group interactions, contextual factors and fixed effects," *Journal of Econometrics*, 2007, *140* (2), 333–374.

_ **, Xiaodong Liu, and Xu Lin**, "Specification and estimation of social interaction models with network structures," *The Econometrics Journal*, 2010, *13* (2), 145–176.

Giorgi, Giacomo De, Michele Pellizzari, and Silvia Redaelli, "Identification of social interactions through partially overlapping peer groups," *American Economic Journal: Applied Economics*, 2010, *2* (2), 241–75.

Goldsmith-Pinkham, Paul and Guido W Imbens, "Social networks and the identification of peer effects," *Journal of Business & Economic Statistics*, 2013, *31* (3), 253–264.

Graham, Bryan S, "Network data," Technical Report, National Bureau of Economic Research 2019.

Györfi, László and Roi Weiss, "Universal consistency and rates of convergence of multiclass prototype algorithms in metric spaces," *arXiv preprint arXiv:2010.00636*, 2020.

Heckman, James J and Richard Robb, "Alternative methods for evaluating the impact of interventions: An overview," *Journal of econometrics*, 1985, *30* (1-2), 239–267.

Hu, Yuchen, Shuangning Li, and Stefan Wager, "Average Treatment Effects in the Presence of Interference," *arXiv preprint arXiv:2104.03802*, 2021.

Hudgens, Michael G and M Elizabeth Halloran, "Toward causal inference with interference," *Journal of the American Statistical Association*, 2008, *103* (482), 832–842.

Imbens, Guido W and Joshua D Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica: Journal of the Econometric Society*, 1994, pp. 467–475.

Jackson, Matthew O and Brian W Rogers, "Meeting strangers and friends of friends: How random are social networks?," *American Economic Review*, 2007, *97* (3), 890–915.

_ , _ , and Yves Zenou, "The economic consequences of social-network structure," *Journal of Economic Literature*, 2017, *55* (1), 49–95.

_ , Tomas Rodriguez-Barraquer, and Xu Tan, "Social capital and social quilts: Network patterns of favor exchange," *The American Economic Review*, 2012, *102* (5), 1857–1897.

Karlan, Dean, Markus Mobius, Tanya Rosenblat, and Adam Szeidl, "Trust and social collateral," *The Quarterly Journal of Economics*, 2009, *124* (3), 1307–1361.

Kitagawa, Toru and Guanyi Wang, "Who Should Get Vaccinated? Individualized Allocation of Vaccines Over SIR Network," *arXiv preprint arXiv:2012.04055*, 2020.

Lehmann, Erich L and Joseph P Romano, *Testing statistical hypotheses*, Springer Science & Business Media, 2006.

Leung, Michael, "Treatment and Spillover Effects under Network Interference," *Available at SSRN 2757313*, 2016.

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

Manski, Charles F, "Nonparametric bounds on treatment effects," *The American Economic Review*, 1990, *80* (2), 319–323.

_ , "Identification of endogenous social effects: The reflection problem," *The review of economic studies*, 1993, *60* (3), 531–542.

_ , "Identification of treatment response with social interactions," *The Econometrics Journal*, 2013, *16* (1), S1–S23.

Sävje, Fredrik, "Causal inference with misspecified exposure mappings," *arXiv preprint arXiv:2103.06471*, 2021.

_ **, Peter M Aronow, and Michael G Hudgens**, "Average treatment effects in the presence of unknown interference," *The Annals of Statistics*, 2021, *49* (2), 673–701.

Vazquez-Bare, Gonzalo, "Identification and estimation of spillover effects in randomized experiments," *arXiv preprint arXiv:1711.02745*, 2017.

Viviano, Davide, "Policy targeting under network interference," *arXiv preprint arXiv:1906.10258*, 2019.

A Proof of claims

A.1 Properties of $(\mathcal{G}, d)$

Note: For the proofs in this section, we consider the space of rooted networks $\mathcal{G}$ with individual-level covariates given by $C_i \in \mathbb{R}^P$ ipth entry given by c_{ip} as opposed to just treatment-information T_i. Accordingly, the definition of the metric in Section 3.2.1 is modified so that $|c_{jp} - c_{f(j)p}| < \epsilon$ for all $p \in [P]$.

Proposition A.1. *d is a pseudo-metric on the set of rooted networks.*

Sketch of Proof: To be a pseudo-metric on the set of rooted networks, d must satisfy (i) the identity condition $d(g,g) = 0$, (ii) the symmetry condition $d(g,g') = d(g',g)$, and (iii) the triangle inequality $d(g,g'') \leq d(g,g') + d(g',g'')$ for any three rooted networks g, g', g''. Conditions (i) and (ii) follow immediately from the definition of d. We demonstrate (iii).

Fix an arbitrary g, g', g'' and denote the vertex sets, edge sets, and agent-specific variables associated with g by $(V(g), E(g), C(g))$. Suppose $d(g,g') + d(g',g'') = \eta$. Then by the definition of d, for every $\nu > 0$ there exist $\varepsilon, \varepsilon' \in \mathbb{R}_+$ and $r, r' \in \mathbb{Z}_+$ with $\zeta(r) + \varepsilon + \zeta(r') + \varepsilon' < \eta + \nu$ such that for every $x, x' \in \mathbb{Z}_+$ there exists root-preserving bijections $f_x : V(g^r)^x \leftrightarrow V(g'^r)^x$, $f'_{x'} : V(g'^{r'})^{x'} \leftrightarrow V(g''^{r'})^{x'}$, (where we explicitly index the bijections by x and x', respectively) such that $e_{jk}(g^r) = e_{f_x(j)f_x(k)}(g'^r)$, $|c_{jp}(g^r) - c_{f_x(j)p}(g'^r)| \leq \varepsilon$, for every $j,k \in V(g^r)^x$, $p \in [P]$, and $e_{jk}(g'^{r'}) = e_{f'_{x'}(j)f'_{x'}(k)}(g''^{r'})$, $|c_{jp}(g'^{r'}) - c_{f'_{x'}(j)p}(g''^{r'})| \leq \varepsilon$, for every $j,k \in V(g'^{r'})^{x'}$, $p \in [P]$.

Let $r'' = \min(r, r')$. We show that for all $x'' \in \mathbb{Z}_+$, there exists a root-preserving bijection $f''_{x''} : V(g^{r''})^{x''} \leftrightarrow V(g''^{r''})^{x''}$ such that $e_{jk}(g^{r''}) = e_{f''_{x''}(j)f''_{x''}(k)}(g''^{r''})$ and $|c_{jp}(g^{r''}) - c_{f''_{x''}(j)p}(g''^{r''})| \leq \varepsilon + \varepsilon'$ for every $j,k \in V(g^{r''})^{x''}$, $p \in [P]$. It follows by the definition of d, $d(g,g'') \leq \zeta(r'') + \varepsilon + \varepsilon'$. Since $\zeta(\cdot)$ is nonnegative, $\zeta(r'') \leq \zeta(r) + \zeta(r')$ and so $d(g,g'') \leq \zeta(r) + \zeta(r') + \varepsilon + \varepsilon' < \eta + \nu$. Since $\nu > 0$ was arbitrary, (iii) follows.

It remains to show that there exist such a bijection $f''_{x''}$. To that end, note that for any rooted network g and $a, b \in \mathbb{R}_+$, $V(g^a)^b = V(g^{\min(a,b)})$. We use this fact repeatedly. First, it implies that it suffices to prove the above claim for all $x'' \leq r''$, since for every $x'' > r''$, $V(g^{r''})^{x''} = V(g^{r''})^{r''}$ and $V(g''^{r''})^{x''} = V(g''^{r''})^{r''}$. Second, this also implies that for $x'' \leq r''$, $V(g^r)^{x''} = V(g^{r''})^{x''}$, $V(g'^r)^{x''} = V(g'^{r'})^{x''} = V(g'^{r''})^{x''}$, and $V(g''^{r'})^{x''} = V(g''^{r''})^{x''}$. These equalities allow us to define the root-preserving bijection $f''_{x''} : V(g^{r''})^{x''} \leftrightarrow V(g''^{r''})^{x''}$ by $f''_{x''} = f'_{x''} \circ f_{x''}$. Moreover, we obtain immediately that $e_{jk}(g^{r''}) = e_{f''_{x''}(j)f''_{x''}(k)}(g''^{r''})$ for every $j,k \in V(g^{r''})^{x''}$, and by the triangle inequality for $|\cdot|$ we obtain $|c_{jp}(g^{r''}) - c_{f''_{x''}(j)p}(g''^{r''})| \leq \varepsilon + \varepsilon'$ for every $j \in V(g^{r''})^{x''}$ and $p \in [P]$. The claim follows. □

Proposition A.2. *$(\mathcal{G}, d)$ is a separable metric space.*

Sketch of Proof: To be separable, $\mathcal{G}$ must have a countable dense subset. Let $\tilde{\mathcal{G}}$ be the subset of rooted networks with finite vertex sets and rational-valued covariates (up to 0-isomorphism). The fact that $\tilde{\mathcal{G}}$ is countable follows immediately from the fact that it can be constructed as a countable union of countable sets. We show that $\tilde{\mathcal{G}}$ is dense in $\mathcal{G}$. The claim follows.

To demonstrate that $\tilde{\mathcal{G}}$ is dense in $\mathcal{G}$, we fix an arbitrary rooted network $g \in \mathcal{G}$ and $\eta > 0$, and show that there exists a rooted network $g' \in \tilde{\mathcal{G}}$ such that $d(g, g') \leq \eta$. Choose any $r \in \mathbb{Z}_+$, $\varepsilon \in \mathbb{R}_{++}$ such that $\zeta(r) + \varepsilon \leq \eta$. Then define g' to be the rooted network on the same set of agents and edge weights as g^r, but with rational-valued covariates chosen to be uniformly within ε of their analogues in g. Since g is locally finite, g^r has finitely many vertices, and so $g' \in \tilde{\mathcal{G}}$. Furthermore g' and g^r are ε-isomorphic by construction and so $d(g, g') \leq \zeta(r) + \varepsilon \leq \eta$. The claim follows. □

Proposition A.3. $(\mathcal{G}, d)$ *is a complete metric space.*

Sketch of proof: To be complete, every Cauchy sequence in $\mathcal{G}$ must have a limit that is also in $\mathcal{G}$. Let g_n refer to a Cauchy sequence in $\mathcal{G}$. That is, for every $\eta > 0$ there exists an $m \in \mathbb{N}$ such that $d(g_{m'}, g_{m''}) \leq \eta$ for every $m', m'' \in \mathbb{N}$ with $m', m'' \geq m$.

Fix an arbitrary $r \in \mathbb{Z}_+$ and consider g_n^r, the sequence formed by taking g_n and replacing each element of the sequence with its truncation at radius r. Since $\mathcal{G}$ is locally finite, the vertex set of each element of g_n^r is finite. Let $N_n^r := |V(g_n^r)|$ denote the sequence of vertex set sizes. Since g_n^r is a Cauchy sequence, the entries of N_n^r must also be uniformly bounded. Let $N^r := \limsup_{n \to \infty} N_n^r < \infty$.

We choose an infinite subsequence of g_n, denoted $g_{s(n)}$ such that $N_{s(n)}^r = V$ for some $V \in \mathbb{N}$. Such a subseqence and choice of V must exist because N_n^r takes only finitely many values. Since g_n is a Cauchy sequence, so too is $g_{s(n)}$, and so for every $\eta > 0$ there exists an $m \in \mathbb{N}$ such that for every $m', m'' \in \mathbb{N}$ with $m', m'' \geq m$ there exists a bijection $f : V \leftrightarrow V$ with $e_{jk}(g_{s(m')}) = e_{f(j)f(k)}(g_{s(m'')})$ for every $j, k \in [V]$ and $|c_{jp}(g_{s(m')}) - c_{f(j)p}(g_{s(m'')})| \leq \eta$ for every $j \in [V]$ and $p \in [P]$.

Take the "canonical" rooted network $\tilde{g}_{s(n)}$ from each equivalence class $g_{s(n)}$ to be such that for any $j \in [V]$ and $p \in [P]$ the corresponding sequences of covariates $c_{jp}(g_{t(n)})$ are Cauchy sequences in $\mathbb{R}$. The resulting sequence of covariate matrices are Cauchy sequences with respect to the max-norm in $\mathbb{R}^{V\times P}$, and so converge to a unique limit C because finite-dimensional Euclidean space is complete. Let E be the edge set corresponding to $\tilde{g}_{s(n)}$ for all n (note from above that the edge set is constant along this sequence).

Let g^r_∞ be the rooted network with vertex set $[V]$, edge set E and covariate set C. Then g^r_∞ is a limit for the sequence $g^r_{s(n)}$, and thus g^r_n because g^r_n is a Cauchy sequence. Since $r \in \mathbb{Z}_+$ was arbitrary, g^r_n converges to $g^r_\infty \in \mathcal{G}$ for any $r \in \mathbb{Z}_+$. Define g to be the rooted network such that $g^r = g^r_\infty$ for every $r \in \mathbb{Z}_+$. Then g is the limit of g_n and the claim follows. □

A.2 Theorem 4.1

We verify the hypothesis of Theorem 4.2 (Assumption 4.5) in Canay and Kamat (2018), focusing on the case where $h(g, U)$ is continuously distributed. Parts (i) and (ii) follow from our Lemma A.2 below. Part (iii) follows from our Assumption 4.3. Part (iv) follows from our choice of test statistic in Step 4 of Algorithm 4.1.

Lemma A.1. *Let Assumption 4.1 hold and for any measurable $f : \mathbb{R}\times\mathcal{G} \to \mathbb{R}$ define*

$$r(g) = E[f(h(g, U_{it}), g)] \ .$$

Then for any $g \in \mathcal{G}$,

$$\big(W^*_1(g), \ldots, W^*_q(g)\big)$$

are independent conditional on $G_1(g), \ldots, G_T(g)$ and for every $1 \le j \le q$

$$E[f(W^*_j(g))|G_1(g), \ldots G_T(g)] = r(G^*_j(g)) \ .$$

Proof. Proposition 8.1 of Biau and Devroye (2015) directly implies the first claim that the elements of $\big(W^*_1(g), \ldots, W^*_q(g)\big)$ are independent conditional on $G_1(g), \ldots, G_T(g)$. For the second claim, note that by a similar argument to the second statement in Proposition 8.1 of Biau and Devroye (2015),

$$E[f(W^*_j(g))|G_1(g), \ldots, G_T(g)] = \tilde{r}_g(G^*_j(g)) \ ,$$

where $\tilde{r}_g(\tilde{g}) = E[f(W_t(g))|G_t(g) \simeq_0 \tilde{g}]$. We show that $\tilde{r}_g(\tilde{g}) = r(\tilde{g})$ which demonstrates the claim.

To that end, let τ be the random index such that $W_{\tau,t} = W_t(g)$, then

$$\begin{aligned} E[f(W_t(g))|G_{1t}, \ldots, G_{m_t t}] &= \sum_{i=1}^{m_t} \mathbf{1}\{\tau = i\} E[f(W_{i,t})|G_{1t}, ..., G_{m_t t}] \\ &= \sum_{i=1}^{m_t} \mathbf{1}\{\tau = i\} E[f(h(G_{it}, U_{it}), G_{it})|G_{1t}, \ldots, G_{m_t t}] \\ &= \sum_{i=1}^{m_t} \mathbf{1}\{\tau = i\} E[f(h(G_{it}, U), G_{it})|G_{it}] \\ &= \sum_{i=1}^{m_t} \mathbf{1}\{\tau = i\} r(G_{it}) \\ &= r(G_t(g)) \ , \end{aligned}$$

where the first equality follows from the fact that τ is a function of $G_{1t}, ...G_{m_t t}$, the second follows from the definition of W_{it}, and the third and fourth equalities follow from Assumption 4.1(iii). The result $\tilde{r}_g(\tilde{g}) = r(\tilde{g})$ follows from the law of iterated expectations. □

Lemma A.2. *Under Assumptions 4.1, 4.2 and 4.3 and the null hypothesis (1),*

$$S_T \xrightarrow{d} S \ ,$$

where $S = (S_1, ..., S_{2q})$ is a continuously distributed random vector with independent and identically distributed entries equal in distribution to $h(g, U)$. For any permutation $\pi \in \mathbf{G}$,

$$S^\pi \stackrel{d}{=} S \ ,$$

satisfying Assumption 4.5 (i) and (ii) of Canay and Kamat (2018).

Proof. Lemma A.1 implies that the entries of

$$(W^*(g), W^*(g')) := \big(W_1^*(g), \ldots, W_q^*(g), W_1^*(g'), \ldots, W_q^*(g')\big)$$

are independent conditional on $\{G_t(g)\}_{t\in\mathcal{D}_1}$ and $\{G_t(g')\}_{t\in\mathcal{D}_2}$, and that the conditional distribution functions of $Y_j^*(g)$ and $Y_j^*(g')$ are given by $h_y(G_j^*(g))$ and $h_y(G_j^*(g'))$ respectively. It follows by the law of iterated expectations that

$$P\left(Y_1^*(g) \leq y_1, ..., Y_q^*(g) \leq y_q, Y_1^*(g') \leq y_{q+1}, \ldots, Y_q^*(g') \leq y_{2q}\right) \\ = E\left[\prod_{j=1}^{q} h_{y_j}(G_j^*(g)) \prod_{j=1}^{q} h_{y_{j+q}}(G_j^*(g'))\right] .$$

We first show that Assumption 4.2 implies that $G_j^*(g) \xrightarrow{p} g$ and $G_j^*(g') \xrightarrow{p} g'$ for every j as $T \to \infty$. To see this, fix $\epsilon > 0$ and write

$$\left\{d\left(G_j^*(g), g\right) > \epsilon\right\} = \left\{\frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\{G_t^*(g) \in B_{g,\epsilon}\} < \frac{j}{T}\right\} ,$$

where $B_{g,\epsilon} = \{\tilde{g} \in \mathcal{G} : d(g, \tilde{g}) \leq \epsilon\}$. By the law of large numbers and Assumption 4.2,

$$\frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\{G_t^*(g) \in B_{g,\epsilon}\} \xrightarrow{p} P\left(G_t^*(g) \in B_{g,\epsilon}\right) > 0 ,$$

and since $j/T \leq q/T \to 0$, it follows that $P\left(d\left(G_j^*(g), g\right) > \epsilon\right) \to 0$ as $T \to \infty$.

Assumption 4.3 and the continuous mapping theorem imply that

$$h_y(G_j^*(g)) \xrightarrow{p} h_y(g) \text{ and } h_y(G_j^*(g')) \xrightarrow{p} h_y(g') ,$$

for every j and $y \in \mathbb{R}$, and so it follows from the dominated convergence theorem that

$$E\left[\prod_{j=1}^{q} h_{y_j}(G_j^*(g)) \prod_{j=1}^{q} h_{y_{j+q}}(G_j^*(g'))\right] \to \prod_{j=1}^{q} h_{y_j}(g) \prod_{j=1}^{q} h_{y_{j+q}}(g') .$$

The claim follows from the fact that $h_y(g) = h_y(g')$ under the null hypothesis (1). □

A.3 Theorem 4.2

Our proof follows that of Döring et al. (2017), Theorem 6. Let $G_t(g)$ be the rooted network corresponding to $W_t(g)$. By Lemma A.1, the entries of

$$(W_1^*(g), \dots W_k^*(g)) \ ,$$

are independent conditional on $G_1(g), \dots, G_T(g)$ with

$$E[Y_j^*(g)|G_1(g), \dots G_T(g)] = h(G_j^*(g)) \ .$$

Let $\bar{h}(g) = E[\hat{h}(g)|G_1(g), \dots G_T(g)] = \frac{1}{k}\sum_{j=1}^k h(G_j^*(g))$. Decomposing the mean-squared error gives

$$E[(\hat{h}(g) - h(g))^2] = E[(\hat{h}(g) - \bar{h}(g))^2] + E[(\bar{h}(g) - h(g))^2] \ ,$$

and the claim follows by bounding $E[(\hat{h}(g) - \bar{h}(g))^2] \leq \sigma^2/k$ and $E[(\bar{h}(g) - h(g))^2] \leq E[\varphi_g(U_{(k,T)})^2]$.

To bound the first term, write

$$\begin{aligned} E[(\hat{h}(g) - \bar{h}(g))^2|G_1(g), \dots G_T(g)] &= E\left[\left(\frac{1}{k}\sum_{j=1}^k \left(Y_j^*(g) - h(G_j^*(g))\right)\right)^2 \middle| G_1(g), \dots G_T(g)\right] \\ &= \frac{1}{k^2}\sum_{j=1}^k \sigma^2(G_j^*(g)) \ , \end{aligned}$$

where $\sigma^2(\tilde{g}) = E\left[(h(\tilde{g}, U_{it}) - h(\tilde{g}))^2\right]$ and the second equality follows from Lemma A.1 because $\{Y_j^*(g) - h(G_j^*(g))\}_{j=1}^k$ are independent and mean zero conditional on $\{G_t(g)\}_{t=1}^T$ and $E\left[(Y_j^*(g) - h(G_j^*(g)))^2|G_1(g), \dots G_T(g)\right] = \sigma^2(G_j^*(g))$. By Assumption 4.6, $\sigma^2(G_j^*(g)) \leq \sigma^2$, and so

$$E[(\hat{h}(g) - \bar{h}(g))^2|G_1(g), \dots G_T(g)] \leq \sigma^2/k \ .$$

$E[(\hat{h}(g) - \bar{h}(g))^2] \leq \sigma^2/k$ follows from the law of iterated expectations.

To bound the second term, Assumption 4.5 and the definition of the upper generalized inverse implies

$$|h(g) - h(\tilde{g})| \leq \varphi_g \left(\psi_g(d(g, \tilde{g}))\right) \ ,$$

and so

$$\begin{aligned}
(\bar{h}(g) - h(g))^2 &\leq \left(\frac{1}{k} \sum_{j=1}^{k} \left| h(g) - h(G_j^*(g)) \right| \right)^2 \\
&\leq \left(\frac{1}{k} \sum_{j=1}^{k} \varphi_g \left(\psi_g(d(g, G_j^*(g))) \right) \right)^2 \\
&\leq \varphi_g \left(\psi_g(d(g, G_k^*(g))) \right)^2 \ .
\end{aligned}$$

Under Assumption 4.4, the probability integral transform implies

$$\psi_g(d(g, G_k^*(g))) \stackrel{d}{=} U_{(k,T)} \ ,$$

where $U_{(k,T)}$ is the kth order statistic from a sequence of T iid standard uniform random variables, and so is distributed $Beta(k, T+1-k)$ (see Biau and Devroye 2015, Section 1.2). As a result

$$\varphi_g \left(\psi_g(d(g, G_k^*(g))) \right)^2 \stackrel{d}{=} \varphi_g(U_{(k,T)})^2$$

and so

$$E[(\bar{h}(g) - h(g))^2] \leq E[\varphi_g(U_{(k,T)})^2] \ .$$

The claim follows. $\square$