Optimal treatment assignment rules on networked populations

Abhishek Ananth, Department of Economics, Cornell University

Acknowledgements: The author is grateful for all the helpful comments and suggestions from advisor, Francesca Molinari and committee members, Larry Blume, David Easley and Jörg Stoye. All errors that remain are author's own, likely despite warnings by the aforementioned.

Preliminary draft, please do not circulate!

Abstract

My job market paper explores the optimal distribution of a limited number of preventative treatments (eg: vaccines) for a deadly communicable disease such as malaria or Ebola among individuals on a network. While the literature has considered the optimal distributions of treatments in the presence of heterogeneous treatment effects, the main contribution of this paper lies in accommodating for spillovers in treatments. This paper explicitly models disease propagation on the contact network for diseases such as Ebola or malaria. I extend the empirical welfare maximization (EWM) procedure considered in Kitagawa, Tetenov (2019) to estimate an optimal treatment assignment rule using data from a randomized control trial (RCT). Like Kitagawa, Tetenov (2019), I provide a finite sample bound for the effectiveness (measured in terms of uniform regret) of the proposed procedure. This is the first paper to establish theoretical guarantees for the treatment assignments under restriction on the spillovers and the network structure. I demonstrate that the welfare attained by EWM converges to the maximum attainable welfare as the size of the RCT grows. I also show that there can exist no other statistical procedure with a faster rate of convergence.

1 Introduction

According to a 2019 situation report[1] by *Doctors without Borders*, during the recent Ebola outbreak in the Democratic Republic of Congo, WHO faced with the possibility of shortage of vaccines, rationed its limited supply[2]. The report argues that this ad-hoc rationing of vaccines prolonged the adversity resulting from the outbreak. Recognizing the need to effectively distribute limited treatments, especially during an epidemic, this paper provides a method to estimate an optimal rationing rule in the presence of heterogeneous treatment effects and spillovers. Optimality here is defined as maximizing a population level notion of welfare.

The presence of heterogeneous treatment effects, whereby individuals respond differently to the same treatment, has important implications for the optimal allocation of a limited number of treatments. As an example, a treatment may be highly effective among elderly males, while being ineffective among young females. A policy-maker could leverage this fact to efficiently distribute the treatments within the population. Suppose the treatment is for a deadly communicable disease; examples include vaccine for Ebola, insecticide treated bed-net for malaria, and PrEP for HIV. Then, the treatment not only directly benefits those treated but also subsequently restricts the spread of the disease to the untreated, which economists term as spillovers. These spillovers are often of large enough magnitude to warrant careful consideration by the policy-maker, particularly in the context of communicable diseases. Especially when supplies are limited, policy-makers or public health officials can maximize welfare by strategically distributing treatments to simultaneously account for the heterogeneity in treatments as well as the spillovers to the untreated. An example of welfare is the expected number of uninfected members of the populations one year after the distribution of treatments. In this paper, I extend the notion of Empirical Welfare Maximization (EWM) to allow for spillovers. I am also able to demonstrate useful properties of the treatment rule implied by EWM.

This paper models the problem as one of contagion on a network. I posit that individuals (or households) are arranged on a contact network where links between individuals represent interactions through which the disease may be transmitted. A notable similarity between

[1] Doctors without Borders. (2019, September 23). WHO rationing Ebola vaccines as outbreak still not under control in Democratic Republic of Congo. `shorturl.at/jpBGN`

[2] "WHO is restricting the availability of the vaccine in the field and the eligibility criteria and their application for reasons that are unclear ... We think that upping the pace of vaccination is necessary and feasible. At least 2,000-2,500 people could be vaccinated each day, instead of 500-1,000 as is currently the case." - Dr. Isabelle Defourny, director of operations, *Doctors without Borders*.

Ebola, malaria and HIV is that their respective contact networks can be assumed to have a bounded degree distribution[3]. While this assumption of bounded degree distribution restricts the generality of the model, it provides enough structure for my results to go through.

The novelty of this paper lies in accommodating for spillovers, i.e. the outcome of any individual depends on own treatment as well as the treatments of others in the network. In the language of Neyman-Rubin causal models, while each particular individual has a binary treatment status (either assigned a treatment or not), each individual has more than two potential outcomes. Owing to numerous ways in which the treatments of others in the network might effect an individual, such problems are intractable without further structure. Manski (2013) lists a set of useful restrictions that facilitate a dimension reduction on the set of potential outcomes. The reduced set of potential outcomes resulting from such restrictions are defined as effective treatments. In two models of disease propagation and spillovers compatible with the spread of either malaria or Ebola separately, I define the implied effective treatments. This dimension reduction allows for the estimation of potential outcomes when treatments are experimentally assigned.

In the case of malaria, households share an edge on the contact network if they are in close proximity to each other as well as lush vegetation where mosquitoes live[4]. This is because, if one household has malaria, the mosquitoes (living in lush vegetation nearby) can spread this disease to neighboring households. For the case of malaria, I assume the effective treatments satisfy local spillovers and exchangeability. In conjunction, these restrictions imply that the treatment effect in any household is a function of its own treatment as well as the number of treated neighboring households.

Ebola is transmitted among humans through close physical contact with individuals who are infected. Additionally, those infected show symptoms early on and are often incapacitated by the resulting symptoms. Consequently, Ebola typically spreads among family, friends and caretakers of infected individuals. In particular, nurses and doctors at health care facilities are especially vulnerable [5]. Therefore, in the context of Ebola, individuals are assumed to share an edge on the contact network if they are family, or close friends, or health care workers assigned to the individual. I describe a structural model of disease propagation

[3]The maximum number of individuals to which the disease may be transmitted by a particular individual is small.

[4]Cowman, Healer, Marapana, Marsh (2016); American Mosquito Control Association.

[5]Chowell, Nishiura (2014)

and treatment response. I adopt the SIR (susceptible - infected - recovered) model[6] from the epidemiology literature[7] and augment it to allow for heterogeneity. I then derive the effective treatment implied by this model.

Working with these frameworks, the main contribution of this paper is to construct a statistically valid estimator for the optimal treatment assignment rule in each settings. A treatment assignment rule specifies the treatment assignment to each individual in the network given the network topology as well as covariates of individuals (or households). The optimal treatment assignment rule maximizes some notion of welfare subject to constraints. For example, a planner might want to strategically assign 20 vaccines to maximize the number of non-infected individuals in a village with 100 individuals. The planner faces an additional statistical challenge: it needs to estimate the treatment effects as well as spillovers. To do so, the planner has access to data from a pilot study conducted in a different village where the treatments are randomly assigned on an observed network.

I extend the empirical welfare maximization procedure, based on constructing an empirical analogue of the welfare function (in the above example, the estimated number of non-infected individuals) to account for spillovers. The optimal rule is then estimated by maximizing this empirical welfare function. I provide a finite sample bound on the uniform regret associated with the proposed method, i.e. the difference between welfare implied by the proposed rule and the highest welfare that can be attained in a non-arbitrarily complex class of treatment rules. I am able to demonstrate that under the bounded degree distribution in the contact network assumption, the root$-n$ rate of convergence of the uniform regret is preserved, i.e. the same rate as observed in the literature without spillovers is stablished here. This is done separately for the two distinct frameworks considered above. Finally, a lower bound is also provided for the uniform regret in the malaria case. This demonstrates the tightness[8] of the rate of convergence obtained by the proposed empirical welfare maximization procedure. Currently, a similar lower bound is being worked out for the Ebola case.

Econometricians have successfully built an arsenal of techniques to estimate treatment effects, whether by conducting experiments or using quasi-experimental methods and observational data[9]. An active area of research in economics explores the estimation of optimal treatment rules that informs policy-makers which members of a population to treat.

[6] Hethcote (1994)

[7] Ball, Britton, Leung, Sirl (2019)

[8] A tight bound means the bound is close to binding

[9] For examples, refer to Heckman, Vytlacil (2001) and Imbens, Angrist, Rubin (1996)

However, much of this literature operates under the Stable Unit Treatment Value Assumption (SUTVA). This assumption requires that treating one individual does not impose any spillovers on other individuals. The above example of vaccinating people against communicable diseases violates this assumption. The first important contribution of this paper is that it extends the literature, particularly the empirical welfare maximization procedure suggested in Kitagawa, Tetenov (2018), to allow for spillovers.

A second important contribution of this paper is that it provide results for the single growing network as well as the many networks paradigm. Research on estimation on networks has largely studied settings where many networks are observed, with only a few recent examples of studies on a single large network. The many networks paradigm requires that the pilot study contain information on a large number of components. For example, a pilot study conducted over many villages with no links across villages. Recognizing that this is often very expensive, the single growing network paradigm posits that the experiment is conducted over one village instead.

Outside of economics, computer scientists and network epidemiologists have considered the problem of optimal vaccine distribution on a network. However, in these studies, heterogeneity in treatment response is suppressed in favor of analytical tractability. Further, the process governing the spread of disease as well as the effectiveness of the treatment are assumed to be known by the planner. This paper allows the planner to estimate treatment effects and spillovers through an experiment conducted on a separate network. Presently, the only other paper to allow for spillovers in this setting is Viviano (2019 wp). It considers spillovers restricted to satisfy local spillovers and exchangeability. However, in contrast to the paradigm presented in this paper, Viviano (2019) considers localized treatment rules rather than network wide treatment rules. This distinction allows the planner in my context to use the full network information. Consequently, the paradigm presented in this paper is able to account for global constraints on the number of treatments available. Further, the framework presented here is better able to accommodate the heterogeneity in local network topology across different parts of the same network. These benefits come at the cost of increased computation complexity.

2 Literature Review

This paper falls at the intersection of numerous literatures.

Firstly, the literature on optimal statistical treatment rules. Within economics, the study of optimal treatment rules is not new, at least in settings where treating one individual does not generate spillovers on other individuals [10]. Manski (2000) discusses un-dominated treatment rules in the event of ambiguity in the planner's problem and thus forges a link with the statistical decision theory literature beginning with Wald (1950) and Savage (1951). Later, Manski (2004) provided bounds for minimax regret for the Conditional Empirical Success Treatment rules under specific experimental designs. Hirano, Porter (2009) provides results on asymptotic validity of empirical welfare maximization procedures. Bhattacharya, Dupas (2012) demonstrate the use of such empirical welfare maximizing techniques on problems where the planner faces constraints. Recently, Kitagawa, Tetenov (2019) use tools from empirical process theory to demonstrate that the empirical welfare maximizer is rate-optimal among a broad class of statistical treatment rules that are not arbitrarily complex.

Secondly, the literature on identification and estimation of network spillover effects. This studies identification and estimation of treatment effects with spillovers (i.e. dropping SUTVA) using data from a single growing network. Manski (2013) discusses identification of treatment effects under a variety of spillover assumptions. Leung (2020) derives consistent estimators and standard errors for treatment and spillover effects under the restrictions of *local spillovers* and *exchangeability*. There is also a literature on estimating games of network formation using data from a single growing network, for examples see Graham (2017), De Paula, Richards-Shubik, Tamer (2018) and Menzel (2020). [11] A related literature, initiated by Ballester, Calvó-Armengol, Zenou (2006), identifies key players in network. They do so by studying games on networks with linear-quadratic utilities. They demonstrate conditions under which the key player (player whose removal leads to the greatest change in aggregate activity) is the player with the highest Katz-Bonacich centrality. Subsequently, an active literature estimating these models has emerged. There are two key distinctions between these papers and the present work. Firstly, actions are endogenous and jointly determined (i.e. agents play a game) in the games while actions are not endogenous in my setting (though, they are jointly determined). Secondly, while this literature identifies the key player to remove from the network, I am interested in finding the optimal subgroup to treat when treatments themselves induce heterogenous responses (both in terms of treatment effects as well as spillovers).

[10] This assumption is usually referred to as- Stable Unit Treatment Value Assumptions (SUTVA) and sometimes by the more descriptive- Individualistic Treatment Response (ITR).

[11] Leung's work considers flexible networks but very restrictive interference structures while Vazquez-Bare (2020) looks at flexible interference structures on restricted networks (clusters). Cleary there is a trade-off between the two in ensuring identification.

Thirdly, the network epidemiological literature. Scholarly work in epidemiology recognized the benefits of considering a 'network approach' in studying the spread of STDs (see Klovdahl (1985) for an example involving AIDS). Social connections play a vital role in facilitating the spread of the HIV virus, which spreads from an infected individual to a susceptible individual through intimate physical contact (for example exchange of body fluids). Recognizing the role of such social connections in facilitating the spread of diseases, network epidemiologists have suggested 'targeted immunization protocols' to minimize the probability of an epidemic. This research has mainly focused on the two situations. Firstly, showing that for general scale-free networks with random immunization of even unrealistically high density of individuals does not lead to global immunity. Pastor, Satorras, Vespignani (2002) - also show that vaccinating based on connectivity hierarchy (targeting more connected nodes) sharply lowers vulnerability to epidemics. Secondly, comparing immunization protocols under local information. Holme (2004) demonstrates the benefits of targeting nodes sequentially where we treat most connected neighbors of already treated in neighbors in successive iterations. It is worth noting that much of this literature explores treatments that make the treated completely immune to the disease (analogous to node removal). To the contrary, I explore more general notions of treatment. Additionally, in my setting, the planner needs to first estimate treatment effects and spillovers.

Finally, the literature on influence maximization. This literature prescribes model of disease propagation and analytically detects nodes with the highest influence on the propagation of disease in the network. As discussed above, in this literature, often heterogeneity in treatment response is suppressed in favor of analytical tractability. Examples of this include Domingos, Richardson (2001), Kempe et al. (2003) , Banerjee et al. (2014) and Jackson, Storms (2019).

Apart from these, this paper builds on technical results in empirical process theory as well as models of disease propagation in network epidemiology. These will be highlighted in-text.

3 Setup

In this section, I state technically, the main problem addressed in this paper. Suppose there is a network consisting on J nodes whose links (undirected and unweighted) are encoded in adjacency matrix $\mathbf{A}$. Each node is endowed with a vector of pre-treatment covariates

$X_j, \ \forall 1 \leq j \leq J$. Each node's treatment status is recorded in $T_j \in \{0,1\}, \ \forall 1 \leq j \leq J$. This can easily be generalized to the case with many treatments. Following the assignment of treatments to the entire network, for each node an outcome is measured $Y_j, \ \forall 1 \leq j \leq J$. Let $\mathbf{X}, \mathbf{T}, \mathbf{Y}$ record covariates, treatment status and outcomes for each node in the network respectively. On account of the spillovers allowed, the set of potential outcomes given network adjacency matrix $\mathbf{A}$ is:

$$\big\{Y_j\big(\mathbf{T}|\mathbf{A}\big) : \mathbf{T} \in \{0,1\}^J\big\}$$

Consider a planner who observes the network adjacency matrix $\mathbf{A}$ and covariates for each node $\mathbf{X}$. The planner must allocate treatment vector $\mathbf{T}$ so as to maximize some notion of welfare for the nodes on this network. In this paper, I consider the case of utilitarian welfare. Thus, the planner chooses a treatment rule (Rule : $\mathbf{A}, \mathbf{X} \mapsto \mathbf{T}$) so as to maximize this welfare. That is, suppose the planner knew the distribution $\big\{\big(\mathbf{Y}(\mathbf{T})\big)_{\mathbf{T}}, \mathbf{X}, \mathbf{A}\big\} \sim \mathbf{P}$, then the planner's *idealized* problem can be written as:

$$\max_{\text{Rule}} \mathbb{E}_{\mathbf{P}}\left[\frac{1}{J}\sum_{j=1}^{J}\left[\sum_{\mathbf{T}\in\{0,1\}^J} Y_j(\mathbf{T}|\mathbf{A})1\{\text{Rule}(\mathbf{A},\mathbf{X}) = \mathbf{T}\}\right]\right]$$

The literature on statistical decision theory explores the case where the planner does not know this distribution $\mathbf{P}$. Instead the planner has access to a dataset from which features of $\mathbf{P}$ may be learnt. This is elaborated on in the following section.

3.1 Statistical Treatment Rules

Suppose the planner has access to B treatments to distribute among the nodes of a network of size $J \geq B$. The *treatment set* of the planner, the set of feasible treatment allocations, is denoted by

$$\mathscr{T} \equiv \big\{\mathbf{T} \in \{0,1\}^J : \sum_{j=1}^{J} T_j \leq B\big\}$$

Suppose that elements of $\mathscr{T}$ can be indexed by $q \in \{1, \cdots, Q\}$. Using *law of iterated expectations*, the planner's problem may be equivalently written as:

$$\max_{\text{Rule}} \mathbb{E}_{\mathbf{P}}\left[\frac{1}{J}\sum_{j=1}^{J}\left[\sum_{\mathbf{T}\in\{0,1\}^J} \mathbb{E}\big[Y_j\big|\mathbf{A},\mathbf{X}\big]1\{\text{Rule}(\mathbf{A},\mathbf{X}) = \mathbf{T}\}\right]\right]$$

While the planner does not know the distribution $\mathbf{P}$, suppose she (planner) is able to estimate relevant features of $\mathbf{P}$, she can then construct a *statistical treatment rule*:

$$\hat{Z}(\mathbf{A}, \mathbf{X}|\text{Data}) = \arg\max_{Z} \ \hat{\mathbb{E}}_{\mathbf{P}}\left[\frac{1}{J}\sum_{j=1}^{J}\left[\sum_{\mathbf{T}\in\{0,1\}^J} \hat{\mathbb{E}}\big[Y_j\big|\mathbf{A}, \mathbf{X}\big]1\{Z(\mathbf{A}, \mathbf{X}|\text{Data}) = \mathbf{T}\}\right]\right]$$

Manski (2000, 2004) point out the need to verify the performance of $\hat{Z}$ (for the simpler problem without spillovers). To do so, he extends concepts from *statistical decision theory* initiated by Wald (1950) and Savage (1951). In the present setup, his prescription would amount to defining a *regret* function which is then used to compare between different treatment rules. Necessary definitions are formalized below.

Definition 1. *Space of network characteristics.*

$$\mathcal{S} = \mathscr{A}_J \times \mathscr{X}^J$$

where

- $\mathscr{A}_J$ as the set of all pairs of $J \times J$ adjacency matrices that are undirected, unweighted.
- $\mathscr{X}$ is the support of covariate X

Definition 2. *Statistical Treatment Rule.*
A statistical treatment rule is a mapping estimated from data that assigns to each condition $(\mathbf{A}, \mathbf{X}) \in \mathcal{S}$, a treatment vector $\mathbf{T} \in \mathscr{T}$.

$$Z(\cdot|\text{Data}): \ \mathcal{S} \to \mathscr{T}$$

We will index these rules by the *decision sets* they generate.

Definition 3. *Decision Sets.*
$G_1, \cdots, G_Q$ is a *partition* of $\mathcal{S}$ such that:

$$\forall q, \ \{(\mathbf{A}, \mathbf{X}) \in \mathcal{S}: \ Z(\mathbf{A}, \mathbf{X}) = \mathbf{T}_q\} = G_q$$

Let $\mathcal{G}_{Full}$ be the set of all partitions of $\mathcal{S}$.

Definition 4. *Relevant probability measures.*
Define $\mathcal{P}^o$ to be the set of all probability measures satisfying the some structure. This structure depends on the problem being considered, examples for the two cases explored in

this paper are provided below

$$\mathcal{P}^o \equiv \{P : P \text{ satisfies problem specific structure; } (\mathbf{A}, \mathbf{X}) \sim F(\cdot; J)\ \}$$

where F is the known distribution of $(\mathbf{A}, \mathbf{X})$ and depends on network population size, J.

Definition 5. *Welfare.*
The welfare for any treatment rule Z given distribution $\mathbf{P} \in \mathcal{P}^o$ is define as

$$W(Z; \mathbf{P}) = \mathbb{E}_{\mathbf{P}}\left[\frac{1}{J}\sum_{j=1}^{J}\left[\sum_{\mathbf{T}\in\{0,1\}^J} \hat{\mathbb{E}}\big[Y_j \big| \mathbf{A}, \mathbf{X}\big] 1\{Z(\mathbf{A}, \mathbf{X}) = \mathbf{T}\}\right]\right]$$

Definition 6. *Regret.*
To compare between different statistical treatment rules, I define the regret associated with any rule $Z(\cdot|\text{Data}$. To do so, first

$$R(Z; \mathbf{P}) = \mathbb{E}_{\text{Data}}\left[\max_{Z} W(Z; \mathbf{P}) - W(Z(\cdot|\text{Data}); \mathbf{P})\right]$$

Definition 7. *Uniform Regret.*

$$R(Z) = \sup_{P\in\mathcal{P}^o} R(Z; P)$$

At this point, it is useful to introduce relevant concepts from *empirical process theory.*

3.1.1 Empirical Process Theory

Suppose that the distribution of $\mathbf{A}, \mathbf{X} \sim F(\cdot; J)$ is known, one candidate for the minimizer of regret (in the many villages problem) is to solve:

$$\max_{Z(\cdot|\text{Data}):\mathcal{S}\to\mathcal{T}} \sum_{q=1}^{Q} F\big(Z(\mathbf{A}, \mathbf{X}|\text{Data}) = \mathbf{T}_q; J\big) \cdot \left(\frac{1}{J}\sum_{j=1}^{J} \hat{\mathbb{E}}\big[Y_j(\mathbf{T}_q)|Z(\mathbf{A}, \mathbf{X}|\text{Data}) = \mathbf{T}_q\big]\right) \equiv W_S^*$$

In the event that $\mathcal{S}$ in infinite, there are an infinite number of statistical treatment rules to choose from. This presents a new challenge:

$$\max_{Z(\cdot|\text{Data}):\mathcal{S}\to\mathcal{T}} \hat{\mathbb{E}}\big[Y_j(\mathbf{T}_q)|Z(\mathbf{A}, \mathbf{X}|\text{Data}) = \mathbf{T}_q\big] \overset{\mathbf{P}}{\nrightarrow} \max_{Z(\cdot|\text{Data}):\mathcal{S}\to\mathcal{T}} \mathbb{E}\big[Y_j(\mathbf{T}_q)|Z(\mathbf{A}, \mathbf{X}|\text{Data}) = \mathbf{T}_q\big]$$

Essentially, the convergence is killed by the fact that one can always find a probability distribution where each term individually converges but the supremum does not converge. It is at this point, chaining methods provide a restriction on the complexity of of the class of statistical treatment rules considered such that the uniform convergence goes through. The logic here is that by ensuring the class of the rules considered is *simple*, one can still ensure uniform convergence. Here, the notion of simplicity used is that of *finite VC dimension.* This definition is presented in terms of the *decision sets* below:

Definition 8. *VC dimension.*
Suppose $\mathcal{G}$ is a collection of subsets of $\mathcal{S}$, we say $\mathcal{G}$ shatters a (finite) collection of point $S \equiv \{x_1, ..., x_m\}$ in $\mathcal{S}$ if

$$|\{S \cap G : G \in \mathcal{G}\}| = 2^m$$

i.e. all 2^m combinations are recoverable.
The *VC dimension* of $\mathcal{G}$ is the cardinality of the largest collection of points that are shattered by $\mathbb{G}$.

This is a key insight exploited by Kitagawa, Tetenov (2018) in their work on estimating optimal treatment rules under SUTVA. My paper builds on their method but has been modified to accommodate violations of SUTVA. In this paper, I suggest the following statistical treatment rule:

$$\max_{G_1, \cdots, G_Q} \sum_{q=1}^{Q} F\big(\mathbf{A}, \mathbf{X} \in G_q; J\big) \cdot \left(\frac{1}{J}\sum_{j=1}^{J} \hat{\mathbb{E}}\big[Y_j(\mathbf{T}_q)|\mathbf{A}, \mathbf{X} \in G_q\big]\right)$$

The main contribution of this paper is to provide finite sample bounds on the *uniform regret* corresponding to this *empirical welfare maximizing treatment rule.*

3.2 Latent Space Configuration

I now clarify the structure of the data required for this exercise. The two types of data structures that are relevant here are:

1. Increasing number of nodes asymptotics

2. Increasing number of networks asymptotics

While results presented for both cases are novel, the first case is more technically demanding and involves considerable advances over and above the second case. For this reason, more attention is paid to these results.

In the first case, I also elaborate on assumptions of network formation. Particularly, I assume that the network formed can be expressed as a random graph embedded on some latent space. Note that these assumptions are not required for the second case.

3.2.1 Increasing number of nodes asymptotics

Here, the idea is that planner observes one large network (with all its nodes). To help formalize the notion of increasing number of nodes asymptotics, consider a sequence of networks indexed by the number of nodes they contain, n. Assume that the nodes are arranged on some underlying latent space. This space is referred to as an *Index Set* and is formalized below.

Assumption 1. *Index set.*
Assume that there exists a d-dimensional integer lattice, $\mathbb{Z}^d$ on which potential nodes are located.

Definition 9. *Metric on the index set.*
$\forall$ $\mathbf{k}, \mathbf{l} \in \mathbb{Z}^d$,

$$\rho(\mathbf{k}, \mathbf{l}) = \max_{1 \le b \le d} |k_b - l_b|$$

where k_b and l_b corresponds to the b-th component of $\mathbf{k}$ and $\mathbf{l}$ respectively.

Definition 10. *Network of size n.*
In this paper, a network population is a realization of $\mathcal{D}_n \subset \mathbb{Z}_n$ such that $|\mathcal{D}_n| = n$. Thus, $\mathcal{D}_n$ is a random variable which takes values in $\{\mathcal{D} : \mathcal{D} \subset \mathbb{Z}_n,\ |\mathcal{D}| = n\}$. Suppose $(\Omega, \mathscr{S})$ is a measurable space. Then, $\mathcal{D}_n : \Omega \to \{\mathcal{D} : \mathcal{D} \subset \mathbb{Z}_n,\ |\mathcal{D}| = n\}$ is a set-valued random variable. i.e.

$$\forall D \subset \mathbb{Z}_n, \text{ such that } D \text{ is open}, \ \mathcal{D}_n^{-1}(D) \equiv \{\omega \in \Omega : \mathcal{D}_n(\omega) \cap D \neq \emptyset\} \in \mathscr{S}$$

These latent *locations* are not observed by the planner. Instead, a planner observe arbitrary identifiers for each node, i.e. in a network of size n, the planner labels nodes $\{1, \cdots, n\}$.

Definition 11. *Location mapping function for network of size n.*
$loc_n : \{1, \cdots, n\} \mapsto \mathbb{Z}^d$ maps each node with arbitrarily assigned identifier i to a location on the index set when node i is in network of size n. Thus, this location function (or n-dimensional vector) is itself a random variable, since it depends on $\mathcal{D}_n$.

The planner only observes the identity (i) of the nodes, their covariates X_i, as well as *all* the connections between the nodes. Additional assumptions are now made concerning the formation of connections of these connections. First, note that, under definitions above, one can define distances between nodes i and j in a network of size n:

$$\rho_{ij}^n = \rho(loc_n(i), loc_n(j)) = \max_{1 \leq b \leq d} |(loc_n(i))_b - (loc_n(j))_b|$$

In any network of size n, this is a random function.

This paper postulates that nodes share an edge on the network if and only if they are close on this latent space. This is formalized below.

Assumption 2. *Edge formation on contact network.*
For a network of size n, assume that nodes i and j are connected iff $\rho_{ij}^n \leq \bar{\rho}$ for some $\bar{\rho} > 1$.

$$\mathbf{A}_{ij} = 1\{\rho_{ij}^n \leq \bar{\rho}\} \quad w.p.\ 1$$

Assumption 3. *Super-population Sampling Frame.*
For any network of size n, $\mathcal{D}_n$ is the complete population within that network.

Thus, an observation of a network of size n corresponds to observing the entire set of links for that network. Unlike the case of snowball sampling where an increase in n corresponds to observation of a larger number of nodes on the same network, here, an increase in n corresponds to a larger network. Note that this can be weakened to fit with an analogue of snowball sampling. For brevity, this paper abstracts away from such extensions.

Implicit Assumption in 2. *Increasing Domain Asymptotics.*
For any $\mathbf{l} \neq \mathbf{k} \in \mathbb{Z}^d$, $\rho(\mathbf{l}, \mathbf{k}) \geq \rho_0 = 1$.

Lemma 1. *Sparsity implied by the latent space characterization.*
Under assumptions 1, 2 and 3, for any village of arbitrary size n, the maximum degree of any node in the contact network is bounded above by:

$$(1 + 2\bar{\rho})^d - 1$$

Proof: To prove the claim above, the above lemma is restated in terms of the notation established. Fix any $\mathbf{l} \in \mathbb{Z}^d$, where $\forall \mathbf{l} \neq \mathbf{k} \in \mathbb{Z}^d$, $\rho(\mathbf{l}, \mathbf{k}) \geq 1$. Then, given assumption #, the above lemma states that

$$|\{\mathbf{k} \in \mathcal{D}_n : \mathbf{k} \neq \mathbf{l},\ \rho(\mathbf{k}, \mathbf{l}) \leq \bar{\rho}\}| \leq (1 + 2\bar{\rho})^d - 1 \text{ with probability 1.}$$

Define a lattice L around $\mathbf{l} \in \mathbb{Z}^d$ as follows.

$$L = \{\mathbf{k} \in \mathbb{Z}^d : \forall 1 \leq b \leq d,\ k_b \in \{l_b, l_b + 1, l_b - 1, \cdots, l_b + \bar{\rho}, l_b - \bar{\rho}\}$$

Under assumptions made,

$$\{\mathbf{k} \in \mathcal{D}_n : \mathbf{k} \neq \mathbf{l},\ \rho(\mathbf{k}, \mathbf{l}) \leq \bar{\rho}\} \subseteq L \setminus \{\mathbf{l}\}$$

Thus,

$$|\{\mathbf{k} \in \mathcal{D}_n : \mathbf{k} \neq \mathbf{l},\ \rho(\mathbf{k}, \mathbf{l}) \leq \bar{\rho}\}| \leq |L| - 1 = (1 + 2\rho^*)^d - 1 \ \text{ with probability 1.} \ \blacksquare$$

Based on the assumptions made thus far, for a village of size n, given $\mathcal{D}_n$, $(\mathbf{A}, \mathbf{I}^0, \mathbf{X})$ are determined. Thus, we assume that there exists a measure on measurable space $(\Omega, \mathscr{S})$, μ_n such that the implications of $\mathcal{D}_n$, $(\mathbf{A}(\mathcal{D}_n), \mathbf{I}^0(\mathcal{D}_n), \mathbf{X}(\mathcal{D}_n))$ agrees with $F(\cdot; n)$ [12]. Here, I need to make a strong assumption of stationarity to ensure that villages of different sizes are not themselves different.

Assumption 4. *Strong stationarity.*
For any $m \leq n$, define $\mathcal{D}_m(n) = \mathcal{D}_m | \mathcal{D}_m \subset \mathcal{D}_n$, then with probability one, $\mu_{m|n}$, which is the distribution of $\mathcal{D}_m(n)$ is equal to μ_m.

3.2.2 Increasing number of networks asymptotics

Here, the idea is that planner observes one many networks (each network is observed completely). None of the assumptions made above are necessary. In this paradigm, the planner observes the adjacency matrix, covariate vector and treatment vector for n networks $\left(\tilde{\mathbf{A}}_i, \tilde{\mathbf{X}}_i, \tilde{\mathbf{T}}_i\right)_{i=1}^n$.

Assumption 5. *Isolated networks.*
There are no links between nodes of different villages.

Assumption 6. *IID networks.*

[12] With the notational simplicity that $F(\cdot; J) \equiv F(\cdot)$

Each network is IID distributed, $\big(\tilde{\mathbf{A}}_i, \tilde{\mathbf{X}}_i, \tilde{\mathbf{T}}_i\big)_{i=1}^n$ in independent across i and $\big(\tilde{\mathbf{A}}_i, \tilde{\mathbf{X}}_i, \tilde{\mathbf{T}}_i\big) \sim F(:, J)$ for all i. Each network observed contains J nodes.

Later, I make assumptions on the dependence structure between node within a network. An important assumption maintained throughout this paper has to do with the evolution of networks.

Assumption 7. *No endogenous response.*
Distribution of treatment does not induce any endogenous response on the relevant network. Further, this paper also assumes that the networks are static and do not evolve over time.

3.3 Effective Treatments

I now adapt the definition of effective treatments suggested by Manski (2013) to the two cases I consider in this paper. I start by defining relevant concepts.

Definition 12. *Effective treatment.*
Define $r_j(\cdot; \mathbf{A}) : \{0,1\}^J \to \mathcal{R}_j(\mathbf{A})$ to be a function that maps the effective treatment implied by treatment vector $(\mathbf{T})$, given the network adjacency matrix $(\mathbf{A})$ into implied effective treatment for agent j. The set $\mathcal{R}_j(\mathbf{A})$ is the set of possible effective treatments for node j given the network characteristics. Effective treatments satisfy the property that for any nodes j and k, for any $(\mathbf{T}, \mathbf{A}) \neq (\mathbf{T}', \mathbf{A}')$ such that $r_j(\mathbf{T}; \mathbf{A}) = r_k(\mathbf{T}'; \mathbf{A}')$ implies $Y_j\big(\mathbf{T}\big|\mathbf{A}\big) \overset{d}{=} Y_k\big(\mathbf{T}'\big|\mathbf{A}'\big)$.

Example 1: *With SUTVA - no covariates.*
Under the common interference assumption of SUTVA (as in Kitagawa, Tetenov (2018)), the set of effective treatments are $\mathcal{C}_j(\mathbf{A}) = \mathcal{C} = \{0,1\}$ and the effective treatment mapping is $c_j(\mathbf{T}|\mathbf{A}) = T_j$.

Example 2: *Without SUTVA - no covariates.*
Without making use of the structure introduced in section 3.1, for a village with J individuals, the profile of treatments across all member could affect outcome of individual j. So, $\mathcal{C}_j(\mathbf{A}) = \{0,1\}^J$, implying that each profile of treatment distributions across the village corresponds to a different treatment from node j perspective. Consequently, $c_j(\mathbf{T}|\mathbf{A}) = \mathbf{T}$.

Definition 13. *Network characteristics.*

The network characteristics corresponds to the network adjacency matrix ($\mathbf{A}$)and covariate vector ($\mathbf{X}$) in a network.

Definition 14. *Homogenized effective treatment.*
Define $c_j(\cdot;\mathbf{A},\mathbf{X}) : \mathscr{T} \to \mathcal{C}_j(\mathbf{A},\mathbf{X})$ a function that maps the effective treatment implied by treatment vector ($\mathbf{T}$), given the network characteristics ($\mathbf{A},\mathbf{X}$) into implied effective treatment for agent j. The set $\mathcal{C}_j(\mathbf{A},\mathbf{X})$ is the set of possible effective treatments for node j given the network characteristics. Effective treatments satisfy the property that for any nodes j and k, for any $(\mathbf{T},\mathbf{A},\mathbf{X}) \neq (\mathbf{T}',\mathbf{A}',\mathbf{X}')$ such that $c_j(\mathbf{T};\mathbf{A},\mathbf{X}) = c_k(\mathbf{T}';\mathbf{A}',\mathbf{X}')$ implies $Y_j\big(\mathbf{T}\big|\mathbf{A}\big)\big|\mathbf{X} \overset{d}{=} Y_k\big(\mathbf{T}'\big|\mathbf{A}'\big)\big|\mathbf{X}'$.

3.3.1 Local spillovers and exchangeability

Assumption 8: *Local spillover and exchangeability*
Outcome of node j depends only on treatment assigned to itself as well as the number of treatments assigned among its *neighbor*. Then,

$$\forall j,\ \mathcal{R}_j(\mathbf{A}) = \{\{0,1\} \times \{0,\cdots,\sum_i \mathbf{A}_{ij}\}\}$$

and

$$\forall j,\ r_j(\mathbf{T}|\mathbf{A}) = \big(T_j, \sum_i \mathbf{A}_{ij}T_i,, \sum_i \mathbf{A}_{ij}\big)$$

This particular assumption has received a lot of attention in the literature, owing to its tractability. Consider the following example that economically motivates this assumption. Suppose the planner wants to distribute textbooks in classrooms across schools in an attempt to raise test scores. The planner is aware of study groups that form within these classes and has access to this data in the form of an adjacency matrix ($\mathbf{A}$). Further, the planner also observes covariates for each student in each classroom ($\mathbf{X}$). Here, the above assumption implies that the outcome (test score) of node j only depends on whether or not node j receives a textbook, the number of study partners of node j that receive the textbook and the number of study partners of node j. Thus, this assumption implies that the identity of the recipients of the textbook are inconsequential.

Assumption 9: *No contextual heterogeneity in treatment effects*

For any j and for any $\mathbf{A}$, assume that

$$\left\{Y_j(\mathbf{T}|\mathbf{A}) - Y_j(\mathbf{T}'|\mathbf{A}) : \mathbf{T} \neq \mathbf{T}'\right\} \perp\!\!\!\perp \mathbf{X}_{-j}\big|X_j$$

3.3.2 Network epidemiology

In this subsection, I formalize the model of propagation of HIV-AIDS by adapting a popular model from network epidemiology to incorporate *heterogenous treatment effects* and *spillovers*. Using this structure, I define *effective treatments* á la Manski (2013).

Suppose that a village contains J people (nodes) arranged on a contact network. Each edge on this node indicates the presence of an intimate physical contact, which allows the spread of the HIV virus. This paper models the propagation of the disease as a variant of the SIR model which has been studied extensively in the network epidemiology literature (for example, see Rusu (2015)). The SIR model posits each node to be in one of three states (catagories): *Susceptible*- nodes that are currently healthy, but may get infected upon coming in contact with infected nodes; *Infected*- nodes that are currently infected and transmit the disease to those in contact with them; *Recovered*- nodes that have been infected in the past and can now no longer receive nor pass on the disease. There have been several modifications to this specification, including SIS (Susceptible-Infected-Susceptible), where an infected node returns to being susceptible once (s)he 'recovers'. This is a more reasonable specification to model flu and flu-like diseases. In the context of HIV-AIDS, we consider the SID variant where the *Recovered* state is replaced with *Deceased*. This paper considers treatments as drugs that limit the spread of the disease. Specifically, the treatment is assumed to reduce the probability that a node gets infected on coming in contact with other infected nodes. Recently, one such drug has become available in the market, namely Pre-Exposure Prophylaxis (PrEP). These drugs are not perfectly effective and their effectiveness depends on the characteristics of the node.

Suppose that the contact network is represented by network *adjacency matrix* $\mathbf{A}$, with the usual convention that $\mathbf{A}_{jj} = 0$, $\forall j$. Throughout this paper, the network is restricted to be undirected and unweighted, implying the *adjacency matrix* is symmetric and only contain elements 0 and 1. Define the neighbors of node j to be the nodes with which node j shares an edge, i.e. $\{k : \mathbf{A}_{jk} = 1\}$. Each node is also endowed with some pre-treatment covariate $X_j \in \mathcal{X} \subset \mathbb{R}^{\text{dim}}$ for some fixed dim $\in \mathbb{N}$. Denote the treatment status of individual j by T_j, where $T_j = 1$ if individual j is treated and $T_j = 0$ otherwise. $\mathbf{X}$ and $\mathbf{T}$ corresponds to the vectors of covariates and treatments for each node in the network respectively. Suppose $s_j^{(t)} \in \{S, I, D\}$ denotes the state of the j^{th} node at time t. $\mathbf{s}^{(t)}$ is the $J \times 1$ vector of states of all nodes at time t. It will be convenient to denote by $I_j^t = 1\{s_j^{(t)} = I\}$. I assume that the first component of X_j corresponds to the initial state $s_j^{(0)}$. However, to ease notation, I will often write $\mathbf{X}, \mathbf{I}^{(0)}$ separately.

Details of the SID model of HIV-AIDS propagation are now presented. For any node j:

$$\mathbf{P}(s_j^{(t+1)} = D \,|\mathbf{s}^{(t)}, s_j^{(t)} = I, T_i) = 1 - \theta; \quad \mathbf{P}(s_j^{(t+1)} = I \,|\mathbf{s}^{(t)}, s_j^{(t)} = I, T_i) = \theta$$

This states that regardless of the treatment status, an infected node dies in the following period with probability $1-\theta$ or remains infected with probability θ. Further,

$$\mathbf{P}(s_j^{(t+1)} = D \,|\mathbf{s}^{(t)}, s_j^{(t)} = D, T_i) = 1; \quad \mathbf{P}(s_j^{(t+1)} \in \{S, I\} \,|\mathbf{s}^{(t)}, s_j^{(t)} = I, T_i) = 0$$

implying that *Death* is an absorbing state. Now, to consider the transmission of the infection,

$$\mathbf{P}(s_j^{(t+1)} = I \,|\mathbf{s}^{(t)}, s_j^{(t)} = S, T_j, X_j) \equiv \zeta_j^{(t)} = \begin{cases} \frac{\sum_k \mathbf{A}_{jk} 1\{s_k^{(t)} = I\}}{\sum_k \mathbf{A}_{jk}} \cdot [p(X_j) - T_j q(X_j)], \text{ if } \sum_k \mathbf{A}_{jk} \neq 0 \\ 0, \text{ if } \sum_k \mathbf{A}_{jk} = 0 \end{cases}$$

with the restriction that $0 \leq q(x) \leq p(x)$ for all x. $p(\cdot)$ corresponds to the susceptibility of the treatment which is the probability that a node with only infected neighbors gets infected. $q(\cdot)$ corresponds to the treatment response which is the reduction in the probability with which a node with only infected neighbors gets infected. This structure can be conveniently annotated as follows.

Assumption 10. *Structural model of disease propagation.*
The probability that node j goes to state $s_j^{(t+1)}$ at time $t+1$ given the entire state vector at t, $\mathbf{s}^{(t)}$ by:

$$\mathbf{P}(s_j^{(t+1)} | \mathbf{s}^{(t)}, T_i, \mathbf{A}, \mathbf{X}) = \begin{array}{c} \\ S \\ I \\ D \end{array} \begin{array}{c} \begin{array}{ccc} S & I & D \end{array} \\ \left\| \begin{array}{ccc} 1 - \zeta_j^{(t)} & \zeta_j^{(t)} & 0 \\ 0 & \theta & 1-\theta \\ 0 & 0 & 1 \end{array} \right\| \end{array}$$

The dependence of these transition probabilities of node j on the states of nodes other than j means that the evolution of states of each individual node can be thought of as a non-homogeneous Markov chain.

Assumption 11. *Independent state transitions.*
The evolution of states over time is independent across nodes.

$$\forall \mathbf{s}^{(t+1)} \in \{S, I, D\}^J, \mathbf{P}(\mathbf{s}^{(t+1)}|\mathbf{s}^{(t)}, \mathbf{T}, \mathbf{A}, \mathbf{X}) = \prod_{j \leq J} \mathbf{P}(s_j^{(t+1)}|\mathbf{s}^{(t)}, \mathbf{T}, \mathbf{A}, \mathbf{X})$$

Note here that this assumption does not restrict *susceptible* nodes who share the exact same neighbors from having probabilities of transitioning into the *infected* state. However, the event that one of the nodes transitions into *infected* state at time t is independent of the transition of the other node at time t.

Definition 15. $g(i, d; \mathbf{A})$ *- paths.*
Define $g(i, d; \mathbf{A})$ on network $\mathbf{A}$ to be the set of paths in to i with path length d.

$$\mathbf{g}(i, d; \mathbf{A}) = \{(k_1, \cdots, k_d) : \ \mathbf{A}_{ik_1} = \mathbf{A}_{k_1 k_2} = \cdots = \mathbf{A}_{k_{d-1} k_d} = 1\}$$

Definition 16. *k degree neighborhood: $\mathcal{N}_i^k$.*
A node j is said to be a k *degree neighbor* of node i if they are connected by a path on length k.

$$\text{i.e. } \exists i_1, \cdots i_{k-1} : \ \mathbf{A}_{ii_1} = 1, \mathbf{A}_{i_{k-1}j} = 1 \text{ and } \mathbf{A}_{i_l i_{l+1}} = 1, \forall \ l = 1, \cdots, k-2$$

The set of all k degree neighbors of node i is the k degree neighborhood of node i, $\mathcal{N}_i^k$.

In the following lemma, I derive the effective treatment vector for the network epidemiology model in section 3.1 for a pre-specified and fixed $\kappa < \infty$.

Lemma 2. *Homogenized effective treatments without additional assumptions.*
Under assumptions 1 to 4, the effective treatments can be characterized as follows:

$$c_i(\mathbf{T}; \mathbf{A}, \mathbf{I}^0, \mathbf{X}) = \Big(F_i, \big(F_{k_1} : k_1 \in g(i, 1; \mathbf{A})\big), \big((F_{k_1}, F_{k_2}) : (k_1, k_2) \in g(i, 2; \mathbf{A})\big), \cdots, \\ \big((F_{k_1}, \cdots, F_{k_{\kappa-1}}) : (k_1, \cdots, k_{\kappa-1}) \in g(i, \kappa-1; \mathbf{A})\big)\Big)$$

where, $\forall j$,

$$F_j = \Big(\sum_k \mathbf{A}_{jk}, \sum_k \mathbf{A}_{jk} I_k^0, X_j, T_j, I_j^0, 1\{s_j^{(0)} = D\}\Big)$$

■ **Proof:** The proof demonstrates that $c_i(\mathbf{T}; \mathbf{A}, \mathbf{I}^0, \mathbf{X})$ completely characterizes the distribution of $Y_i(\mathbf{T})|\mathbf{A}, \mathbf{I}^0, \mathbf{X}$ given parameters $(\theta, p(\cdot), q(\cdot))$. In the event, $\sum_k \mathbf{A}_{ik} > 0$:

$$\begin{aligned}\mathbf{P}(s_i^{(1)} = a|\mathbf{T}, \mathbf{A}, \mathbf{I}^0, \mathbf{X}) = 1\{a = D\}&\Big[(1-\theta)I_i^0 + 1\{s_i^{(0)} = D\}\Big] \\ + 1\{a = I\}&\Big[\theta I_i^0 + \frac{\sum_k \mathbf{A}_{ik} I_k^0}{\sum_k \mathbf{A}_{ik}}\big(p(X_i) - T_i q(X_i)\big)1\{s_i^{(0)} = S\}\Big] \\ + 1\{a = S\}&\Big[\Big(1 - \frac{\sum_k \mathbf{A}_{ik} I_k^0}{\sum_k \mathbf{A}_{ik}}\big(p(X_i) - T_i q(X_i)\big)\Big)1\{s_i^{(0)} = S\}\Big]\end{aligned}$$

When, $\sum_k \mathbf{A}_{ik} = 0$:

$$\begin{aligned}\mathbf{P}(s_i^{(1)} = a|\mathbf{T}, \mathbf{A}, \mathbf{I}^0, \mathbf{X}) = 1\{a = D\}&\Big[(1-\theta)I_i^0 + 1\{s_i^{(0)} = D\}\Big] \\ + 1\{a = I\}&\Big[\theta I_i^0\Big] + 1\{a = S\}\Big[1\{s_i^{(0)} = S\}\Big]\end{aligned}$$

Given parameters, $\mathbf{P}(s_i^{(1)} = a|\mathbf{T}, \mathbf{A}, \mathbf{I}^0, \mathbf{X})$ is completely characterized by

$$\Big(\sum_k \mathbf{A}_{jk}, \sum_k \mathbf{A}_{jk} I_k^0, X_j, T_j, I_j^0, 1\{s_j^{(0)} = D\}\Big)$$

For κ period transmissions:

$$\begin{aligned}\mathbf{P}\Big(s_j^{(\kappa)} = b\Big|\mathbf{T}, \mathbf{A}, \mathbf{I}^0, \mathbf{X}\Big) = &\sum_{\mathbf{a}^{(\kappa-1)}} \mathbf{P}\Big(s_j^{(\kappa)} = b\Big|\mathbf{T}, \mathbf{A}, \mathbf{X}, \mathbf{s}^{(\kappa-1)} = \mathbf{a}^{(\kappa-1)}\Big) \times \\ &\sum_{\mathbf{a}^{(\kappa-2)}} \mathbf{P}\Big(\mathbf{s}^{(\kappa-1)} = \mathbf{a}^{(\kappa-1)}\Big|\mathbf{T}, \mathbf{A}, \mathbf{X}, \mathbf{s}^{(\kappa-2)} = \mathbf{a}^{(\kappa-2)}\Big) \times \\ &\dots \\ &\sum_{\mathbf{a}^{(1)}} \mathbf{P}\Big(\mathbf{s}^{(2)} = \mathbf{a}^{(2)}\Big|\mathbf{T}, \mathbf{A}, \mathbf{X}, \mathbf{s}^{(1)} = \mathbf{a}^{(1)}\Big) \times \mathbf{P}\Big(\mathbf{s}^{(1)} = \mathbf{a}^{(1)}\Big|\mathbf{T}, \mathbf{A}, \mathbf{I}^0, \mathbf{X}\Big)\end{aligned}$$

where $b \in \{S, I, D\}$ and $\mathbf{a}^{(t)} \in \{S, I, D\}^J$ for all $1 \leq t \leq \kappa$. Assumption 1 implies that for any node, each one step transmission is only governed by the infection states of immediate neighbors (and not the rest of the network). Thus, the above equality can be written as:

$$\begin{aligned}\mathbf{P}\Big(s_j^{(\kappa)} = b\Big|\mathbf{T}, \mathbf{A}, \mathbf{I}^0, \mathbf{X}\Big) = &\sum_{\mathbf{a}_{\mathcal{N}_i^1}^{(\kappa-1)}} \mathbf{P}\Big(s_j^{(\kappa)} = b\Big|\mathbf{T}, \mathbf{A}, \mathbf{X}, \mathbf{s}_{\mathcal{N}_i^1}^{(\kappa-1)} = \mathbf{a}_{\mathcal{N}_i^1}^{(\kappa-1)}\Big)\times \\ &\sum_{\mathbf{a}_{\mathcal{N}_i^2}^{(\kappa-2)}} \mathbf{P}\Big(\mathbf{s}_{\mathcal{N}_i^1}^{(\kappa-1)} = \mathbf{a}_{\mathcal{N}_i^1}^{(\kappa-1)}\Big|\mathbf{T}, \mathbf{A}, \mathbf{X}, \mathbf{s}_{\mathcal{N}_i^2}^{(\kappa-2)} = \mathbf{a}_{\mathcal{N}_i^2}^{(\kappa-2)}\Big)\times \\ &\cdots \\ &\sum_{\mathbf{a}_{\mathcal{N}_i^{\kappa-1}}^{(1)}} \mathbf{P}\Big(\mathbf{s}_{\mathcal{N}_i^{\kappa-2}}^{(2)} = \mathbf{a}_{\mathcal{N}_i^{\kappa-2}}^{(2)}\Big|\mathbf{T}, \mathbf{A}, \mathbf{X}, \mathbf{s}_{\mathcal{N}_i^{\kappa-1}}^{(1)} = \mathbf{a}_{\mathcal{N}_i^{\kappa-1}}^{(1)}\Big)\times \\ &\mathbf{P}\Big(\mathbf{s}_{\mathcal{N}_i^{\kappa-1}}^{(1)} = \mathbf{a}_{\mathcal{N}_i^{\kappa-1}}^{(1)}\Big|\mathbf{T}, \mathbf{A}, \mathbf{I}^0, \mathbf{X}\Big)\end{aligned}$$

where $\mathbf{a}_{\mathcal{N}_i^k}^{(t)} = (\mathbf{a}_j^{(t)} : j \in \mathcal{N}_i^k)$ and similarly $\mathbf{s}_{\mathcal{N}_i^k}^{(t)} = (\mathbf{s}_j^{(t)} : j \in \mathcal{N}_i^k)$. Further, assumption 2 allows characterizing the joint distribution of each one step transmission as a product of marginals.

$$\mathbf{P}\Big(\mathbf{s}_{\mathcal{N}_i^k}^{(t)} = \mathbf{a}_{\mathcal{N}_i^k}^{(t)}\Big|\mathbf{T}, \mathbf{A}, \mathbf{X}, \mathbf{s}_{\mathcal{N}_i^{k+1}}^{(t-1)} = \mathbf{a}_{\mathcal{N}_i^{k+1}}^{(t-1)}\Big) = \prod_{j\in\mathcal{N}_i^k} \mathbf{P}\Big(\mathbf{s}_j^{(t)} = \mathbf{a}_j^{(t)}\Big|\mathbf{T}, \mathbf{A}, \mathbf{X}, \mathbf{s}_{\mathcal{N}_i^{k+1}}^{(t-1)} = \mathbf{a}_{\mathcal{N}_i^{k+1}}^{(t-1)}\Big)$$

Thus,

$$\begin{aligned}c_i(\mathbf{T}; \mathbf{A}, \mathbf{I}^0, \mathbf{X}) = \Big(&F_i, \big(F_{k_1} : k_1 \in g(i, 1; \mathbf{A})\big), \big((F_{k_1}, F_{k_2}) : (k_1, k_2) \in g(i, 2; \mathbf{A})\big), \cdots, \\ &\big((F_{k_1}, \cdots, F_{k_{\kappa-1}}) : (k_1, \cdots, k_{\kappa-1}) \in g(i, \kappa-1; \mathbf{A})\big)\Big)\blacksquare\end{aligned}$$

4 Data/Experiment

In this section, I summarize assumptions on the experimental protocol that generates the data used by the planner in building the statistical treatment rule. Owing to technical as well as notational differences, these assumptions are laid out separately for the two distinct asymptotic frameworks.

4.1 Increasing number of networks asymptotics

4.1.1 Pilot Experiment

The planner has access to data from a single network of size n. That is, the planner observes the *entire* network in the village, recorded in adjacency matrix $\tilde{\mathbf{A}}$. For each node, $1 \leq i \leq n$, covariates $\tilde{X}_i$, as well as experimental treatment assignment $\tilde{D}_i$ are recorded. The planner then observes outcome Y_i for each node in the sampled network.

4.1.2 Identification

The main object of interest are functionals of the joint distribution,

$$\mathbb{E}\big[Y_j(\mathbf{T}|\mathbf{A})\big|(\mathbf{A},\mathbf{X})\in G\big]$$

for each $1 \le j \le J$, $\mathbf{T} \in \mathscr{T}$ and $G \subseteq \mathcal{S}$. This object is well defined under the following assumption.

Assumption 12. *No measure 0 treatment sets.* The distribution $F(\cdot;\cdot)$ is known and satisfies:

$$\forall G \in \mathcal{S}, \quad F\big((\mathbf{A},\mathbf{X})\in G; J\big) > 0$$

A sufficient condition would be for $F(\cdot; J)$ to be absolutely continuous.

No I make necessary assumptions on the experiment to ensure that this quantity can be estimated from the data.

Assumption 13. *Unconfoundedness*

$$\left\{\tilde{Y}_j(\mathbf{T}|\tilde{\mathbf{A}}),\ \mathbf{T}\in\{0,1\}^J\right\} \perp\!\!\!\perp \tilde{\mathbf{D}} \ \Big|\ \tilde{\mathbf{A}},\tilde{\mathbf{X}}$$

Definition 16. $\mathcal{P}$.
Denote the set of probability measure on $\big(\tilde{Y}_i,\tilde{X}_i,\tilde{D}_i|\mathcal{D}_n\big)_{1\le i\le n}$ that satisfy assumption of $\mathcal{P}^o$ and assumptions 12 and 13 by $\mathcal{P}$.

4.1.3 Estimation

This paper proposes a cell-mean estimator under known experimental protocol, i.e. known distribution of $\tilde{\mathbf{D}}|\tilde{\mathbf{A}},\tilde{\mathbf{X}}$. In now propose separate estimators for the two different cases of spillovers discussed.

Case 1: Local spillovers and exchangeability

Assumption 14 (a). *Strict Overlap on Effective Treatments*
For any $\mathbf{T} \in \mathscr{T}$ and for any $\mathbf{A}, \mathbf{X}$, there exists a $\bar{\gamma} \in (0, 0.5)$ such that:

$$\mathbf{P}_{\sim}\big(r_i(\tilde{\mathbf{D}}; \tilde{\mathbf{A}}) \in D_j^r(\mathbf{T}, G),\ \tilde{X}_i \in G|_{X_j}\big) \in (\bar{\gamma}, 1 - \bar{\gamma})$$

where $D_j^r(\mathbf{T}, G) \equiv \{r_j(\mathbf{T}|\mathbf{A}) : (\mathbf{A}, \mathbf{X}) \in G\}$, $G|_{X_j}$ is defined as $\{z_j : \exists \mathbf{X}, \mathbf{A}$ with $X_j = z_j$ and $\mathbf{A}, \mathbf{X} \in G\}$, and $\mathbf{P}_{\sim}$ corresponds to the probability distribution governing the experimental sample.

No simple sufficient condition has been worked out as of yet for this assumption.

The cell-mean estimator in this context is:

$$\hat{\mathbb{E}}\big[Y_j(\mathbf{T}|\mathbf{A})|(\mathbf{A}, \mathbf{X}) \in G\big] = \sum_{i=1}^{n} \frac{\tilde{Y}_i 1\{r_i(\tilde{\mathbf{D}}; \tilde{\mathbf{A}}) \in D_j^r(\mathbf{T}, G), \tilde{X}_i \in G|_{X_j}\}}{\mathbf{P}_{\sim}[r_i(\tilde{\mathbf{D}}; \tilde{\mathbf{A}}) \in D_j^r(\mathbf{T}, G), \tilde{X}_i \in G|_{X_j}]}$$

Lemma 5. *Unbiasedness of cell-mean estimator*
Under assumptions 1-4, 8, 9 , 13 and 14 (a), the above estimator is unbiased.

Proof: Define the following useful shorthands:

$$\begin{aligned} \{r_i \in D_j^r\} &\equiv \{r_i(\tilde{\mathbf{D}}; \tilde{\mathbf{A}}) \in D_j^r(\mathbf{T}, G)\} \\ \{G|_{X_j}\} &\equiv \{\tilde{X}_i \in G|_{X_j}\} \end{aligned}$$

In the notation of this shorthand, this estimator can be expressed as:

$$\hat{\theta} = \hat{\mathbb{E}}\big[Y_j(\mathbf{T}|\mathbf{A})|(\mathbf{A}, \mathbf{X}) \in G\big] = \sum_{i=1}^{n} \frac{\tilde{Y}_i 1\{r_i \in D_j^r\} 1\{G|_{X_j}\}}{\mathbf{P}\big(\{r_i \in D_j^r\}, \{G|_{X_j}\}\big)}$$

Now, taking expectations with respect to data $(\tilde{\mathbf{Y}}, \tilde{\mathbf{D}}, \tilde{\mathbf{X}}, \tilde{\mathbf{A}})$,

$$
\begin{aligned}
\mathbb{E}_{\sim}\big(\hat{\theta}\big) &= \sum_{i=1}^{n} \frac{1}{\mathbf{P}_{\sim}\big(\{r_i \in D_j^r\}, \{G|_{X_j}\}\big)} \mathbb{E}_{\sim}\big(\tilde{Y}_i 1\{r_i \in D_j^r\} 1\{G|_{X_j}\}\big) \\
&= \sum_{i=1}^{n} \frac{1}{\mathbf{P}_{\sim}\big(\{r_i \in D_j^r\}, \{G|_{X_j}\}\big)} \mathbb{E}_{\tilde{\mathbf{D}},\tilde{\mathbf{X}},\tilde{\mathbf{A}}}\big(1\{r_i \in D_j^r\} 1\{G|_{X_j}\} \mathbb{E}_{\tilde{\mathbf{Y}}}(\tilde{Y}_i|\tilde{\mathbf{D}}, \tilde{\mathbf{X}}, \tilde{\mathbf{A}})\big)
\end{aligned}
$$

The second equality follows from the law of iterated expectations. Unconfoundedness implies that $\mathbb{E}_{\tilde{\mathbf{Y}}}(\tilde{Y}_i|\tilde{\mathbf{D}}, \tilde{\mathbf{X}}, \tilde{\mathbf{A}}) = \mathbb{E}_{\tilde{\mathbf{Y}}}(\tilde{Y}_i|\tilde{\mathbf{X}}, \tilde{\mathbf{A}})$. Now, by definition of effective treatments,

$$
1\{r_i \in D_j^r\} 1\{G|_{X_j}\} \mathbb{E}_{\tilde{\mathbf{Y}}}\big(\tilde{Y}_i|\tilde{\mathbf{X}}, \tilde{\mathbf{A}}\big) = 1\{r_i \in D_j^r\} 1\{G|_{X_j}\} \mathbb{E}\big(Y_j(\mathbf{T}|\mathbf{A})\big|(\mathbf{A}, \mathbf{X}) \in G\big)
$$

Substituting this into the above equation:

$$
\begin{aligned}
\mathbb{E}_{\sim}\big(\hat{\theta}\big) &= \mathbb{E}\big(Y_j(\mathbf{T}|\mathbf{A})\big|(\mathbf{A}, \mathbf{X}) \in G\big) \sum_{i=1}^{n} \frac{\mathbb{E}_{\tilde{\mathbf{D}},\tilde{\mathbf{X}},\tilde{\mathbf{A}}}\big(1\{r_i \in D_j^r\} 1\{G|_{X_j}\}}{\mathbf{P}_{\sim}\big(\{r_i \in D_j^r\}, \{G|_{X_j}\}\big)} \\
&= \mathbb{E}\big(Y_j(\mathbf{T}|\mathbf{A})\big|(\mathbf{A}, \mathbf{X}) \in G\big) \blacksquare
\end{aligned}
$$

Thus, the estimator is unbiased.

Case 2: Network epidemiology

Assumption 14 (b). *Strict Overlap on Homogenized effective Treatments*
For any $\mathbf{T} \in \mathscr{T}$ and for any G, there exists a $\bar{\gamma} \in (0, 0.5)$ such that:

$$
\mathbf{P}_{\sim}\big(c_i(\tilde{\mathbf{D}}; \tilde{\mathbf{A}}, \tilde{\mathbf{X}}) \in D_j^c(\mathbf{T}, G)\big) \in (\bar{\gamma}, 1 - \bar{\gamma})
$$

where $D_j^c(\mathbf{T}, G) \equiv \{c_j(\mathbf{T}; \mathbf{A}, \mathbf{X}) : (\mathbf{A}, \mathbf{X}) \in G\}$ and $\mathbf{P}_{\sim}$ corresponds to the probability distribution governing the experimental sample.

No simple sufficient condition has been worked out as of yet for this assumption.

The cell-mean estimator in this context is:

$$
\hat{\mathbb{E}}\big[Y_j(\mathbf{T}|\mathbf{A})|(\mathbf{A}, \mathbf{X}) \in G\big] = \sum_{i=1}^{n} \frac{\tilde{Y}_i 1\{c_i(\tilde{\mathbf{D}}; \tilde{\mathbf{A}}, \tilde{\mathbf{X}}) \in D_j^c(\mathbf{T}, G)\}}{\mathbf{P}_{\sim}[c_i(\tilde{\mathbf{D}}; \tilde{\mathbf{A}}, \tilde{\mathbf{X}}) \in D_j^c(\mathbf{T}, G)]}
$$

Lemma 6. *Unbiasedness of cell-mean estimator*
Under assumptions 1-4, 10, 11, 13 and 14 (b), the above estimator is unbiased.

Proof: Define $\boldsymbol{\{c \in D_j^c\}} \equiv \big(\{c_i(\tilde{\mathbf{D}}; \tilde{\mathbf{A}}, \tilde{\mathbf{X}}) \in D_j^c(\mathbf{T}, G)\}\big)_{i=1,\cdots,n}$.

$$\begin{aligned}
&\mathbb{E}_{\sim}\Big[\hat{\mathbb{E}}\big[Y_j(\mathbf{T}|\mathbf{A})|(\mathbf{A},\mathbf{X}) \in G\big]\Big] \\
&= \mathbb{E}_{\tilde{\mathbf{A}},\tilde{\mathbf{X}}}\left\{\mathbb{E}_{\tilde{Y}_i,\tilde{\mathbf{D}}}\left[\sum_{i=1}^{n} \frac{\tilde{Y}_i 1\{c_i(\tilde{\mathbf{D}}; \tilde{\mathbf{A}}, \tilde{\mathbf{X}}) \in D_j^c(\mathbf{T}, G)\}}{\mathbf{P}[c_i(\tilde{\mathbf{D}}; \tilde{\mathbf{A}}, \tilde{\mathbf{X}}) \in D_j^c(\mathbf{T}, G)]}\middle| \boldsymbol{\{c \in D_j^c\}}\right]\right\} \\
&= \mathbb{E}_{\tilde{\mathbf{A}},\tilde{\mathbf{X}}}\left\{\mathbb{E}_{\tilde{\mathbf{D}}}\left[\sum_{i=1}^{n} \frac{\mathbb{E}_{\tilde{Y}_i}[\tilde{Y}_i|\mathbf{1}\boldsymbol{\{c \in D_j^c\}}] 1\{c_i(\tilde{\mathbf{D}}; \tilde{\mathbf{A}}, \tilde{\mathbf{X}}) \in D_j^c(\mathbf{T}, G)\}}{\mathbf{P}[c_i(\tilde{\mathbf{D}}; \tilde{\mathbf{A}}, \tilde{\mathbf{X}}) \in D_j^c(\mathbf{T}, G)]}\middle| \boldsymbol{\{c \in D_j^c\}}\right]\right. \\
&= \mathbb{E}_{\sim}\left[\sum_{i=1}^{n} \frac{\mathbb{E}[Y_j(\mathbf{T}|\mathbf{A})|(\mathbf{A},\mathbf{X}) \in G] 1\{c_i(\tilde{\mathbf{D}}; \tilde{\mathbf{A}}, \tilde{\mathbf{X}}) \in D_j^c(\mathbf{T}, G)\}}{\mathbf{P}[c_i(\tilde{\mathbf{D}}; \tilde{\mathbf{A}}, \tilde{\mathbf{X}}) \in D_j^c(\mathbf{T}, G)]}\right] \\
&= \mathbb{E}[Y_j(\mathbf{T}|\mathbf{A})|(\mathbf{A},\mathbf{X}) \in G] \sum_{i=1}^{n} \frac{\mathbb{E}_{\sim}\big[1\{c_i(\tilde{\mathbf{D}}; \tilde{\mathbf{A}}, \tilde{\mathbf{X}}) \in D_j^c(\mathbf{T}, G)\}\big]}{\mathbf{P}[c_i(\tilde{\mathbf{D}}; \tilde{\mathbf{A}}, \tilde{\mathbf{X}}) \in D_j^c(\mathbf{T}, G)]} \\
&= \mathbb{E}[Y_j(\mathbf{T}|\mathbf{A})|(\mathbf{A},\mathbf{X}) \in G]
\end{aligned}$$

The first equality follows from law of iterated expectations, the second equality follows from the unconfoundedness, the third follows from the definition of Homogenized effective treatments. ■

4.1.4 Covariance

A consequence of the network spillovers is that $\tilde{Y}_i|\tilde{\mathbf{D}}, \tilde{\mathbf{A}}, \tilde{\mathbf{X}}$ are no longer IID across nodes i. For example, in the network epidemiology example, for each one period transmission, the infection status and hence $\tilde{Y}_j$ of nodes in the one degree neighborhood matters, i.e. $j \in \mathcal{N}_i^1$. Since the planner only observes outcomes after κ-periods, $\tilde{Y}_i|\tilde{\mathbf{D}}, \tilde{\mathbf{A}}, \tilde{\mathbf{I}}^0, \tilde{\mathbf{X}}$ is correlated with $\tilde{Y}_j|\tilde{\mathbf{D}}, \tilde{\mathbf{A}}, \tilde{\mathbf{I}}^0, \tilde{\mathbf{X}}$ for all $j \in \mathcal{N}_i^\kappa$ but independent of outcomes among nodes outside this κ degree neighborhood. This decaying dependence in distance (measured in path length) brings to mind notions of strong mixing from the literature on time series analysis. Consequently similar tools are used here. In particular, I make the following assumption on the correlation structure.

Assumption 15. *κ-uncorrelatedness*
There exists a $1 \leq \kappa < \infty$ that is known by the planner such that nodes on the network with a path length larger than κ are uncorrelated. Using the metric defined on the integer

lattice, the following correlation structure is assumed:

$$\mathbf{k} \notin B_\rho(\mathbf{l}, \kappa) \Rightarrow corr(Y(\mathbf{k}), Y(\mathbf{l})) = 0$$

where $B_\rho(\mathbf{l}, \kappa) = \{\mathbf{k} : \rho(\mathbf{k}, \mathbf{l}) \leq \kappa\bar{\rho}\}$.

4.2 Increasing number of networks asymptotics

In this section, I present the setup for the large number of network asymptotic framework. With access to data from many networks, I can drop several of the assumptions maintained thus far. Particularly, none of the assumptions regarding the model of disease propagation (in section 3) are required. $Z_{j,(i)}$ denotes the observations of random variable Z for node j in network i and $\mathbf{Z}_{(i)}$ denotes the vector of Z realization for all nodes within the village. As before, I continue to operate in the setting where the propensity score is known.

4.3 Estimation

Here I directly provide an estimator for the welfare function.

$$W_n(G_1, \cdots, G_Q) = \frac{1}{N}\sum_{i=1}^{N} \frac{\sum_{q=1}^{Q} \mathcal{Y}_{(i)} \cdot 1\big\{\tilde{\mathbf{D}}_{(i)} = \mathbf{T}_q\big\} \cdot 1\big\{(\tilde{\mathbf{A}}_{(i)}, \tilde{b}X_{(i)}) \in G_q\big\}}{\mathbf{P}\big(\tilde{\mathbf{D}}_{(i)} = \mathbf{T}_q \big| \tilde{\mathbf{A}}_{(i)}, \tilde{\mathbf{X}}_{(i)}\big)}$$

where

$$\mathcal{Y}_{(i)} = \frac{1}{J}\sum_{j=1}^{J} \tilde{Y}_{j,(i)}$$

Assumption 16. *Unconfoundedness*
For all $1 \leq i \leq N$,

$$\Big\{\tilde{\mathbf{Y}}_{(i)}(\mathbf{T}),\ \mathbf{T} \in \{0,1\}^J\Big\} \perp\!\!\!\perp \tilde{\mathbf{D}}_{(i)} \,\Big|\, \tilde{\mathbf{A}}_{(i)}, \tilde{\mathbf{X}}_{(i)}$$

Assumption 17. *Strict Overlap on Treatments*
For any $\tilde{\mathbf{T}} \in \{0,1\}^J$ and for any $\tilde{\mathbf{A}}_{(i)}, \tilde{\mathbf{X}}_{(i)}$, there exists a $\bar{\gamma} \in (0, 0.5)$ such that:

$$\mathbf{P}_\sim\big(\tilde{\mathbf{D}}_{(i)} = \mathbf{T} \big| \tilde{\mathbf{A}}_{(i)}, \tilde{\mathbf{X}}_{(i)}\big) \in (\bar{\gamma}, 1 - \bar{\gamma})$$

where $\mathbf{P}_\sim$ corresponds to the probability distribution governing the experimental sample.

Under the above assumptions as well as assumptions 5-7, it is trivial to establish unbiasedness of the proposed estimator.

5 Empirical Welfare Maximization

I now set up the proposed empirical welfare maximization procedure.

5.1 Increasing number of nodes asymptotics

As suggested earlier, the paper proposes the empirical welfare maximizer

$$\max_{G_1,\cdots,G_Q\in\mathcal{G}} \sum_{q=1}^{Q} F\big((\mathbf{A},\mathbf{X})\in G_q;J\big)\cdot\left(\frac{1}{J}\sum_{j=1}^{J}\hat{\mathbb{E}}\big[Y_j(\mathbf{T}_q|\mathbf{A})|(\mathbf{A},\mathbf{X})\in G_q\big]\right)$$

Suppose this is solved by some partition $\hat{\mathbf{G}}_{EWM}\in\mathcal{G}$. Denote the welfare associated with this statistical treatment rule by $W(\hat{\mathbf{G}}_{EWM})$. In fact, for any $\mathbf{G}=\{G_1,\cdots,G_Q\}\in\mathcal{G}$, write

$$W(G_1,\cdots,G_Q)=\sum_{q=1}^{Q} F\big((\mathbf{A},\mathbf{X})\in G_q;J\big)\cdot\left(\frac{1}{J}\sum_{j=1}^{J}\mathbb{E}\big[Y_j(\mathbf{T}_q|\mathbf{A})|(\mathbf{A},\mathbf{X})\in G_q\big]\right)$$

Let $W^*_{\mathcal{G}}$ denote the maximum welfare attainable in $\mathcal{G}$.

Lemma 7. Under assumptions 1-4, 10-14,

$$\sup_{\mathbf{P}\in\mathcal{P}}\mathbb{E}_{Data\sim(\mathbf{P},F)}\left[W^*_{\mathcal{G}}-W(\hat{\mathbf{G}}_{EWM})\right]\leq 2\sum_{q=1}^{Q}\sup_{\mathbf{P}\in\mathcal{P}}\mathbb{E}_{Data\sim(\mathbf{P},F)}\left[\sup_{G\in\mathbb{G}}|W^q_n(G)-W^q(G)|\right]$$

Proof: Observe that:

$$W^*_{\mathcal{G}}-W(\hat{\mathbf{G}}_{EWM})\geq 0$$

and that $\forall\ \mathbf{G}=\{G_1,\cdots,G_Q\}\in\mathcal{G}$, define, where

$$W_n(\mathbf{G})=\sum_{q=1}^{Q} F\big((\mathbf{A},\mathbf{X})\in G_q\big)\cdot\left(\frac{1}{J}\sum_{j=1}^{J}\hat{\mathbb{E}}\big[Y_j(\mathbf{T}_q|\mathbf{A})|(\mathbf{A},\mathbf{X})\in G_q\big]\right)$$

Then, write:

$$\begin{aligned}
&W(\mathbf{G}) - W(\hat{\mathbf{G}}_{EWM}) \\
&= W(\mathbf{G}) - W_n(\hat{\mathbf{G}}_{EWM}) - W(\hat{\mathbf{G}}_{EWM}) + W_n(\hat{\mathbf{G}}_{EWM}) \\
&\leq W(\mathbf{G}) - W_n(\mathbf{G}) + \sup_{\mathbf{G}\in\mathcal{G}} |W_n(\mathbf{G}) - W(\mathbf{G})| \qquad \text{(Since } W_n(\hat{\mathbf{G}}_{EWM}) \geq W_n(\mathbf{G})\text{)} \\
&\leq 2 \sup_{\mathbf{G}\in\mathcal{G}} |W_n(\mathbf{G}) - W(\mathbf{G})|
\end{aligned}$$

By application of triangle inequality,

$$\sup_{\mathbf{G}\in\mathcal{G}} |W_n(\mathbf{G}) - W(\mathbf{G})| \leq \sum_{q=1}^{Q} \sup_{G\in\mathbb{G}} |W_n^q(G) - W^q(G)|$$

where

$$W^q(G) = F\big((\mathbf{A}, \mathbf{X}) \in G\big)\frac{1}{J}\sum_{j=1}^{J}\mathbb{E}\big[Y_j(\mathbf{T}_q|\mathbf{A})|(\mathbf{A}, \mathbf{X}) \in G\big]$$

and

$$W_n^q(G) = F\big((\mathbf{A}, \mathbf{X}) \in G\big)\frac{1}{J}\sum_{j=1}^{J}\hat{\mathbb{E}}\big[Y_j(\mathbf{T}_q|\mathbf{A})|(\mathbf{A}, \mathbf{X}) \in G\big]$$

Thus, one can bound the uniform regret associated with the *empirical welfare maximization* rule as follows:

$$\sup_{\mathbf{P}\in\mathcal{P}} \mathbb{E}_{Data\sim(\mathbf{P},F)}\left[W_{\mathcal{G}}^* - W(\hat{\mathbf{G}}_{EWM})\right] \leq 2\sum_{q=1}^{Q} \sup_{\mathbf{P}\in\mathcal{P}} \mathbb{E}_{Data\sim(\mathbf{P},F)}\left[\sup_{G\in\mathbb{G}} |W_n^q(G) - W^q(G)|\right] \blacksquare$$

5.2 Increasing number of networks asymptotics

In this case the empirical welfare maximization corresponds to:

$$\max_{G_1,\cdots,G_Q\in\mathcal{G}} \frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{q=1}^{Q}\mathcal{Y}_{(i)} \cdot 1\big\{\tilde{\mathbf{D}}_{(i)} = \mathbf{T}_q\big\} \cdot 1\big\{(\tilde{\mathbf{A}}_{(i)}, \tilde{b}X_{(i)}) \in G_q\big\}}{\mathbf{P}\big(\tilde{\mathbf{D}}_{(i)} = \mathbf{T}_q\big|\tilde{\mathbf{A}}_{(i)}, \tilde{\mathbf{X}}_{(i)}\big)}$$

Suppose this is solved by some partition $\hat{\mathbf{G}}_{EWM} \in \mathcal{G}$. Denote the welfare associated with this statistical treatment rule by $W(\hat{\mathbf{G}}_{EWM})$. In fact, for any $\mathbf{G} = \{G_1, \cdots, G_Q\} \in \mathcal{G}$, write

$$W(G_1, \cdots, G_Q) = \sum_{q=1}^{Q} \mathbb{E}\left[\frac{1}{J}\sum_{j=1}^{J} Y_j\big(\mathbf{T}_q|\mathbf{A}\big) \cdot 1\big\{(\mathbf{A}, \mathbf{X}) \in G_q\big\}\right]$$

Let $W^*_{\mathcal{G}}$ denote the maximum welfare attainable in $\mathcal{G}$.

Lemma 8. Under assumptions 5-7, 16 and 17

$$\mathbb{E}_{Data\sim(\mathbf{P},F)}\left[\sup_{G\in\mathbb{G}} |W^q_N(G) - W^q(G)|\right] \leq \mathbb{E}_{Data\sim(\mathbf{P},F)}\left[\sup_{f\in\mathcal{F}} |\mathbb{E}_n(f) - \mathbb{E}(f)|\right]$$

Proof: Observe that:

$$W^*_{\mathcal{G}} - W(\hat{\mathbf{G}}_{EWM}) \geq 0$$

and that $\forall\ \mathbf{G} = \{G_1, \cdots, G_Q\} \in \mathcal{G}$, define, where

$$W_N(\mathbf{G}) = \frac{1}{N}\sum_{i=1}^{N} \frac{\sum_{q=1}^{Q} \mathcal{Y}_{(i)} \cdot 1\big\{\tilde{\mathbf{D}}_{(i)} = \mathbf{T}_q\big\} \cdot 1\big\{(\tilde{\mathbf{A}}_{(i)}, \tilde{b}X_{(i)}) \in G_q\big\}}{\mathbf{P}\big(\tilde{\mathbf{D}}_{(i)} = \mathbf{T}_q\big|\tilde{\mathbf{A}}_{(i)}, \tilde{\mathbf{X}}_{(i)}\big)}$$

Then, write:

$$\begin{aligned}
&W(\mathbf{G}) - W(\hat{\mathbf{G}}_{EWM}) \\
&= W(\mathbf{G}) - W_N(\hat{\mathbf{G}}_{EWM}) - W(\hat{\mathbf{G}}_{EWM}) + W_N(\hat{\mathbf{G}}_{EWM}) \\
&\leq W(\mathbf{G}) - W_N(\mathbf{G}) + \sup_{\mathbf{G}\in\mathcal{G}} |W_N(\mathbf{G}) - W(\mathbf{G})| \qquad \text{(Since } W_N(\hat{\mathbf{G}}_{EWM}) \geq W_N(\mathbf{G})\text{)} \\
&\leq 2 \sup_{\mathbf{G}\in\mathcal{G}} |W_N(\mathbf{G}) - W(\mathbf{G})|
\end{aligned}$$

By application of triangle inequality,

$$\sup_{\mathbf{G}\in\mathcal{G}} |W_N(\mathbf{G}) - W(\mathbf{G})| \leq \sum_{q=1}^{Q} \sup_{G\in\mathbb{G}} |W^q_N(G) - W^q(G)|$$

where

$$W^q(G) = \mathbb{E}\left[\frac{1}{J}\sum_{j=1}^{J} Y_j\big(\mathbf{T}_q|\mathbf{A}\big) \cdot 1\big\{(\mathbf{A}, \mathbf{X}) \in G_q\big\}\right]$$

and

$$W^q_N(G) = \frac{1}{N}\sum_{i=1}^{N} \frac{\mathcal{Y}_{(i)} \cdot 1\big\{\tilde{\mathbf{D}}_{(i)} = \mathbf{T}_q\big\} \cdot 1\big\{(\tilde{\mathbf{A}}_{(i)}, \tilde{\mathbf{X}}_{(i)}) \in G_q\big\}}{\mathbf{P}\big(\tilde{\mathbf{D}}_{(i)} = \mathbf{T}_q\big|\tilde{\mathbf{A}}_{(i)}, \tilde{\mathbf{X}}_{(i)}\big)}$$

Thus, one can bound the uniform regret associated with the *empirical welfare maximization* rule as follows:

$$\sup_{\mathbf{P}\in\mathcal{P}} \mathbb{E}_{Data\sim(\mathbf{P},F)}\left[W^*_{\mathcal{G}} - W(\hat{\mathbf{G}}_{EWM})\right] \leq 2\sum_{q=1}^{Q} \sup_{\mathbf{P}\in\mathcal{P}} \mathbb{E}_{Data\sim(\mathbf{P},F)}\left[\sup_{G\in\mathbb{G}} |W^q_N(G) - W^q(G)|\right]$$

I now provide a bound for $\mathbb{E}_{Data\sim(\mathbf{P},F)}\left[\sup_{G\in\mathbb{G}} |W^q_N(G) - W^q(G)|\right]$ that is independent of $\mathbf{P}$ and hence applies uniformly over $\mathcal{P}$. To that end, define $\mathbf{Z}_i \equiv (\tilde{\mathbf{Y}}_i, \tilde{\mathbf{D}}_i, \tilde{\mathbf{A}}_i, \tilde{\mathbf{X}}_i)$ and

$$W^q_N(G) = \frac{1}{n}\sum_{i=1}^{n} f(Z_i; G)$$

Then, from the unbiasedness results of the cell-mean estimator:

$$\mathbb{E}_n\big(f(\cdot;G)\big) \equiv \frac{1}{n}\sum_{i=1}^{n} f(U_i;G) = W^q_N(G) \text{ and } \mathbb{E}_{Data\sim(\mathbf{P},F)}\big(f(\cdot;G)\big) = \mathbb{E}_{Data\sim(\mathbf{P},F)}[f(U_i;G)] = W^q(G)$$

Denote by $\mathcal{F} \equiv \{f(\cdot;G) : G \in \mathbb{G}\}$ the collection of all such functions generated by decision sets. By Lemma A.1 of Kitagawa-Tetenov (2018), $\mathcal{F}$ is a VC-subgraph class of functions with VC dimension less than or equal to v, when $\mathbb{G}$ has VC dimension v. Thus, one can write

$$\mathbb{E}_{Data\sim(\mathbf{P},F)}\left[\sup_{G\in\mathbb{G}} |W^q_N(G) - W^q(G)|\right] \leq \mathbb{E}_{Data\sim(\mathbf{P},F)}\left[\sup_{f\in\mathcal{F}} |\mathbb{E}_n(f) - \mathbb{E}(f)|\right] \quad \blacksquare$$

5.3 Restriction on statistical treatment rules

Earlier, I suggest that in order to ensure good properties on the regret, $\mathcal{G}$ cannot be arbitrarily complex. Here, I specify the exact assumptions required to ensure this is so.

Assumption 17. *Partition with elements having finite VC dimension.*
Any partition $\mathbf{G} = \{G_1, \cdots, G_Q\} \in \mathcal{G}$ can be written as:

$$\mathcal{G} \equiv \{G_1, \cdots, G_Q : \forall 1 \leq q \leq Q, G_q \in \mathbb{G}; \bigcup_{1\leq q\leq Q} G_q = \mathcal{S}; \forall 1 \leq q \neq r \leq Q, G_q \cap G_r = \emptyset\}$$

where $\mathbb{G}$ has finite VC dimension v and is countable.

Consider the following example.

Example 4. *Linear rule on pre-specified network statistics.*
This example makes two simplifying assumptions:

1. The treatment rules depend on the network through L pre-specified network statistics, $\Psi_1(.), \cdots, \Psi_L(.)$, where $\Psi_i : \mathscr{A}_J \to \mathbb{R}^J$. Eg: Katz centrality, Bonacich centrality, alpha centrality, etc.
$$Z(\mathbf{A}, \mathbf{I}^0, \mathbf{X}) = Z(\Psi_1(\mathbf{A}), \cdots, \Psi_L(\mathbf{A}), \mathbf{I}^0, \mathbf{X})$$

2. Linear treatment rules, i.e.
$$Z(\mathbf{A}, \mathbf{I}^0, \mathbf{X}) = \begin{cases} \mathbf{T}_1, \text{ if } \boldsymbol{\alpha}_1 \boldsymbol{\Psi}(\mathbf{A}) + \boldsymbol{\beta}_1(\mathbf{I}^0) + \boldsymbol{\gamma}_1(\mathbf{X}) > \max\limits_{q \neq 1} \big[\boldsymbol{\alpha}_q \boldsymbol{\Psi}(\mathbf{A}) + \boldsymbol{\beta}_q(\mathbf{I}^0) + \boldsymbol{\gamma}_q(\mathbf{X})\big] \\ \cdots \\ \mathbf{T}_Q, \text{ if } \boldsymbol{\alpha}_Q \boldsymbol{\Psi}(\mathbf{A}) + \boldsymbol{\beta}_Q(\mathbf{I}^0) + \boldsymbol{\gamma}_Q(\mathbf{X}) > \max\limits_{q \neq Q} \big[\boldsymbol{\alpha}_q \boldsymbol{\Psi}(\mathbf{A}) + \boldsymbol{\beta}_q(\mathbf{I}^0) + \boldsymbol{\gamma}_q(\mathbf{X})\big] \end{cases}$$
where each component G_q is chosen from
$$\mathbb{G} \equiv \Big\{\mathbf{A}, \mathbf{s^{(0)}}, \mathbf{X} : \boldsymbol{\alpha}_1 \boldsymbol{\Psi}(\mathbf{A}) + \boldsymbol{\beta}_1(\boldsymbol{I}) + \boldsymbol{\gamma}_1(\mathbf{X}) > \max_{q \neq 1} \big[\boldsymbol{\alpha}_q \boldsymbol{\Psi}(\mathbf{A}) + \boldsymbol{\beta}_q(\boldsymbol{I}) + \boldsymbol{\gamma}_q(\mathbf{X})\big] \\ \text{for some } \alpha_1, \cdots, \alpha_Q, \beta_1, \cdots, \beta_Q, \gamma_1, \cdots, \gamma_Q \Big\}$$
and the partitions are constructed from elements in $\mathbb{G}$.

Since these linear rules yield decision sets that are themselves intersections of half space, we find that for any partition $G_1, \cdots, G_Q$ constructed from the above linear rules, have finite VC dimension due to the following proposition.

Proposition 1. *Blumer* et al. *(1989).*
The set of all intersection of k half spaces in $\mathbb{R}^p$ has VC dimension, v finite. Moreover,
$$v = O(pk \log k)$$

6 Results

I start by proving two technical lemmas that separately handle the case of the two different asymptotic frameworks.

Lemma 9: Under $\mathcal{D}_n$ coming from assumptions 1-4 and 7, $Z_{1:n}$ satisfies assumption 15. Let $\mathcal{F}$ be a class of uniformly bounded functions with $\|f\|_\infty \leq \bar{F}$ for all $f \in \mathcal{F}$. Further, assume that $\mathcal{F}$ is countable and has finite VC dimension. Then, there exists a universal constant C_1 such that

$$\mathbb{E}_{\mathcal{D}_n}\left[\mathbb{E}_{Z_{1:n}|\mathcal{D}_n}\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n} f(Z_i) - \mathbb{E}f\right|\right]\right] \leq 2C_1\left[2^d(1+2\kappa\bar{\rho})^d\right]3/2\bar{F}\sqrt{\frac{v}{n}}$$

Proof: The quantity of interest is:

$$\mathbb{E}_{\mathcal{D}_n}\left[\mathbb{E}_{Z_{1:n}|\mathcal{D}_n}\left[\sup_{f\in\mathcal{F}}\left|\mathbb{E}_n f - \mathbb{E}f\right|\right]\right]$$

Using the metric defined on the integer lattice, the following correlation structure is assumed:

$$\mathbf{k} \notin B_\rho(\mathbf{l},\kappa) \Rightarrow corr(Z(\mathbf{k}), Z(\mathbf{l})) = 0$$

where $B_\rho(\mathbf{l},\kappa) = \{\mathbf{k} : \rho(\mathbf{k},\mathbf{l}) \leq \kappa\bar{\rho}\}$.

Now, I decompose the integer lattice into 2^d categories and use these to then construct Bernstein hypercubes.

For any directional vector $\mathbf{b} \in \{0,1\}^d$, define:

$$C_{\mathbf{b}} \equiv \left\{\{B_\rho(\mathbf{k},\kappa) : \text{ where } \forall 1 \leq i \leq d, k_i = (2a_i + b_i)\kappa\bar{\rho}\} \text{ where } \mathbf{a} \in \mathbb{Z}^d\right\}$$

Now, the idea is to define a Bernstein hypercube with the following feature:

1. Each Bernstein hypercubes contains exactly one elements from each category.

Thus, I define, $\forall \mathbf{m} \in \mathbb{Z}^d$,

$$B_{\mathbf{m}} \equiv \bigcup_{\mathbf{b}\in\{0,1\}^d} \{\mathbf{l} \in B_\rho(\mathbf{k},\kappa) : \text{ where } \forall 1 \leq i \leq d, k_i = (2a_i + b_i)\kappa\bar{\rho}\}$$

Now, notice that for any fixed $\mathcal{D}_n$, there exists at most finite number of Bernstein hypercubes which are non-empty. Consequently, these non-empty elements can be numbered: $B(1), B(2), \cdots, B(\mu_n)$, where $\mu_n = \left|\{\mathbf{m} : B_{\mathbf{m}}(\mathcal{D}_n) \neq \emptyset\}\right|$. Note that by the network formation assumption, for any $1 \leq i \leq \mu_n$, $a_n^j \equiv \left|B(i)\right| \leq (1+2\kappa\bar{\rho})^d 2^d$. Thus, by definition $n = \sum_{j=1}^{\mu_n} a_n^j$.

Define, $\bar{a}_n = \frac{\sum_{j=1}^{\mu_n} a_n^j}{\mu_n}$, the average density of a Bernstein hypercube. Note here that is stochastic since it depends on $\mathcal{D}_n$. With this set up, I can make progress on the object of interest.

$$\mathbb{E}_{\mathcal{D}_n}\left[\mathbb{E}_{Z_{1:n}|\mathcal{D}_n}\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n} f(Z_i) - \mathbb{E}f\right|\right]\right]$$
$$= \mathbb{E}_{\mathcal{D}_n}\left[\mathbb{E}_{Z_{1:n}|\mathcal{D}_n}\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{\mu_n}\sum_{j=1}^{\mu_n}\frac{1}{\bar{a}_n}\sum_{\mathbf{b}\in\{0,1\}^d}\left(\sum_{\mathbf{l}\in D_n\cap C_{\mathbf{b}}\cap B(j)} f\big(Z(\mathbf{l})\big)\right) - \mathbb{E}f\right|\right]\right]$$

By triangle inequality,

$$\leq \mathbb{E}_{\mathcal{D}_n}\left[\mathbb{E}_{Z_{1:n}|\mathcal{D}_n}\left[\sup_{f\in\mathcal{F}}\sum_{\mathbf{b}\in\{0,1\}^d}\left|\frac{1}{\mu_n}\sum_{j=1}^{\mu_n}\frac{1}{\bar{a}_n}\left(\sum_{\mathbf{l}\in D_n\cap C_{\mathbf{b}}\cap B(j)} f\big(Z(\mathbf{l})\big)\right) - \mathbb{E}f\right|\right]\right]$$

By splitting the sup and using linearity of $\mathbb{E}$,

$$\leq \sum_{\mathbf{b}\in\{0,1\}^d}\mathbb{E}_{\mathcal{D}_n}\left[\mathbb{E}_{Z_{1:n}|\mathcal{D}_n}\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{\mu_n}\sum_{j=1}^{\mu_n}\frac{1}{\bar{a}_n}\left(\sum_{\mathbf{l}\in D_n\cap C_{\mathbf{b}}\cap B(j)} f\big(Z(\mathbf{l})\big)\right) - \mathbb{E}f\right|\right]\right]$$

By the strong stationarity imposed in assumption 4, this can be equivalently written as

$$= 2^d\mathbb{E}_{\mathcal{D}_n}\left[\mathbb{E}_{Z_{1:n}|\mathcal{D}_n}\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{\mu_n}\sum_{j=1}^{\mu_n}\frac{1}{\bar{a}_n}\left(\sum_{\mathbf{l}\in D_n\cap C_{\mathbf{b}}\cap B(j)} f\big(Z(\mathbf{l})\big)\right) - \mathbb{E}f\right|\right]\right]$$

Notice that for any $\mathbf{b} \in \{0,1\}^d$, elements in $D_n \cap C_{\mathbf{b}} \cap B(j)$ and $D_n \cap C_{\mathbf{b}} \cap B(k)$ are uncorrelated for $j \neq k$. Now, I start the process of symmetrization. In order to do so, create an independent and identically distributed copy $Z'_{1:n}$ of $Z_{1:n}$.

$$
\begin{aligned}
&= 2^d \mathbb{E}_{\mathcal{D}_n}\Bigg[\mathbb{E}_{Z_{1:n}|\mathcal{D}_n}\Bigg[\sup_{f\in\mathcal{F}}\Bigg|\frac{1}{\mu_n}\sum_{j=1}^{\mu_n}\frac{1}{\bar{a}_n}\Bigg(\sum_{\mathbf{l}\in D_n\cap C_{\mathbf{b}}\cap B(j)} f\big(Z(\mathbf{l})\big)\Bigg) \\
&\qquad - \mathbb{E}_{Z'_{1:n}|\mathcal{D}_n}\frac{1}{\mu_n}\sum_{j=1}^{\mu_n}\frac{1}{\bar{a}_n}\Bigg(\sum_{\mathbf{l}\in D_n\cap C_{\mathbf{b}}\cap B(j)} f\big(Z'(\mathbf{l})\big)\Bigg)\Bigg|\Bigg]\Bigg] \\
&\leq 2^d \mathbb{E}_{\mathcal{D}_n}\Bigg[\mathbb{E}_{Z_{1:n},Z'_{1:n}|\mathcal{D}_n}\Bigg[\sup_{f\in\mathcal{F}}\Bigg|\frac{1}{\mu_n}\sum_{j=1}^{\mu_n}\frac{1}{\bar{a}_n}\Bigg(\sum_{\mathbf{l}\in D_n\cap C_{\mathbf{b}}\cap B(j)} f\big(Z(\mathbf{l})\big)\Bigg) \\
&\qquad - \frac{1}{\mu_n}\sum_{j=1}^{\mu_n}\frac{1}{\bar{a}_n}\Bigg(\sum_{\mathbf{l}\in D_n\cap C_{\mathbf{b}}\cap B(j)} f\big(Z'(\mathbf{l})\big)\Bigg)\Bigg|\Bigg]\Bigg]
\end{aligned}
$$

The inequality follows from the application of Jensen's inequality. Now, define a Rademacher sequence $\sigma_{1:\mu_n}$ that is independent of both $Z'_{1:n}$ and $Z_{1:n}$. Then, by triangle inequality

$$
\leq 2\cdot 2^d \mathbb{E}_{\mathcal{D}_n}\Bigg[\frac{1}{\mu_n}\mathbb{E}_{Z_{1:n},\sigma_{1:\mu_n}|\mathcal{D}_n}\Bigg[\sup_{f\in\mathcal{F}}\Bigg|\sum_{j=1}^{\mu_n}\frac{\sigma_j}{\bar{a}_n}\sum_{\mathbf{l}\in D_n\cap C_{\mathbf{b}}\cap B(j)} f\big(Z(\mathbf{l})\big)\Bigg|\Bigg]\Bigg]
$$

This is known as the Rademacher complexity bound. Fix any $\mathbf{b}$, $\mathcal{D}_n$ and $Z_{1:n}$, then define a μ_n dimensional vector

$$
\mathbf{f} \equiv \Bigg(\frac{1}{\bar{a}_n}\sum_{\mathbf{l}\in D_n\cap C_{\mathbf{b}}\cap B(j)} f\big(Z(\mathbf{l})\big)\Bigg)_{j=1}^{\mu_n} \equiv \Big(f_{C_n^j}\Big)_{j=1}^{\mu_n}
$$

Notice that for any $1 \leq j \leq \mu_n$, $f_{C_n^j} \leq (1+2\kappa\bar{\rho})^d \bar{F}$.

Thus, define: $\mathbb{F}_n \equiv \big\{\mathbf{f} : \|\mathbf{f}\|_\infty \leq (1+2\kappa\bar{\rho})^d\bar{F}\big\} \subset \mathbb{R}^{\mu_n}$. Now, define the following Euclidean norm to $\mathbb{F}_n$:

$$
d_n(\mathbf{f},\mathbf{g}) \equiv \Bigg[\frac{1}{\mu_n}\sum_{j=1}^{\mu_n}\Big(f_{C_n^j} - g_{C_n^j}\Big)^2\Bigg]^{1/2}
$$

Also define the Rademacher complexity maximizing element,

$$
\mathbf{f}_n^* \equiv \big(f_{n,j}^*\big)_{j=1}^{\mu_n} \in \arg\max_{\mathbf{f}\in\mathbb{F}_n}\Bigg|\sum_{j=1}^{\mu_n}\sigma_j f_{C_n^j}\Bigg|
$$

Next, I define a sequence of minimal covers of $\mathbb{F}_n$. In order to do so, start by defining

$$
\mathbf{f}_n^{(0)} \equiv (0,\cdots 0)
$$

Then, for any $k \geq 1$, denote a minimal cover of radius $2^{-k}(2\kappa\bar{\rho}+1)^d\bar{F}$ by M_k^n. Since $\mathbb{F}_n$ is totally bounded, there exists a $\bar{K}$ such that $M_{\bar{K}}^n = \mathbb{F}_n$. Now, within each minimal cover, define the Rademacher complexity maximizing element as

$$\forall 1 \leq k \leq \bar{K},\ \mathbf{f}_n^{(k)} \in \arg\max_{\mathbf{f}\in M_k^n}\left|\sum_{j=1}^{\mu_n}\sigma_j f_{C_n^j}\right|$$

A straightforward application of triangle inequality yields:

$$d_n\big(\mathbf{f}_n^{(k)}, \mathbf{f}_n^{(k-1)}\big) \leq 3\cdot 2^{-k}(1+2\kappa\bar{\rho})^d\bar{F}$$

Consequently, using the telescopic sum representation:

$$\begin{aligned}
&\mathbb{E}_{\sigma_{1:\mu_n}}\left|\sum_{j=1}^{\mu_n}\sigma_j f^*_{C_n^j}\right| \\
&\leq \sum_{k=1}^{\bar{K}}\mathbb{E}_{\sigma_{1:\mu_n}}\left|\sum_{j=1}^{\mu_n}\sigma_j\left(f^{(k)}_{C_n^j} - f^{(k-1)}_{C_n^j}\right)\right| \\
&\leq \sum_{k=1}^{\bar{K}}\mathbb{E}_{\sigma_{1:\mu_n}}\max_{\mathbf{f}\in M_k^n,\mathbf{g}\in M_{k-1}^n: d_n(\mathbf{f},\mathbf{g})\leq\cdot 2^{-k}(1+2\kappa\bar{\rho})^d\bar{F}}\left|\sum_{j=1}^{\mu_n}\sigma_j\left(f_{C_n^j} - g_{C_n^j}\right)\right|
\end{aligned}$$

Now, I'd like to use the maximal inequality to provide a bound for the above. In order to do, I fist verify that an exponential bound holds. For this, I use Hoeffding's inequality:

$$\begin{aligned}
\mathbb{E}_{\sigma_{1:\mu_n}}\left(e^{s\sum_{j=1}^{\mu_n}\sigma_j(f_{C_n^j}-g_{C_n^j})}\right) &= \prod_{j=1}^{\mu_n}\mathbb{E}_{\sigma_j}\left(e^{s\sigma_j(f_{C_n^j}-g_{C_n^j})}\right) \\
&\leq \prod_{j=1}^{\mu_n} e^{s^2(f_{C_n^j}-g_{C_n^j})^2/2} \\
&= e^{s^2\mu_n d_n(\mathbf{f},\mathbf{g})^2/2} \\
&\leq e^{s^2\mu_n(3\cdot 2^{-k}(1+2\kappa\bar{\rho})^d\bar{F})^2/2}
\end{aligned}$$

This exponential bound allows us to write the maximal inequality bound:

$$
\begin{aligned}
&\mathbb{E}_{\sigma_{1:\mu_n}} \max_{\mathbf{f}\in M_k^n, \mathbf{g}\in M_{k-1}^n : d_n(\mathbf{f},\mathbf{g})\leq \cdot 2^{-k}(1+2\kappa\bar{\rho})^d \bar{F}} \left| \sum_{j=1}^{\mu_n} \sigma_j \left(f_{C_n^j} - g_{C_n^j} \right) \right| \\
&\leq 3 \cdot 2^{-k}(1+2\kappa\bar{\rho})^d \bar{F} \sqrt{\mu_n} \sqrt{2\ln(2|M_k^n|^2)} \\
&= 6 \cdot 2^{-k}(1+2\kappa\bar{\rho})^d \bar{F} \sqrt{\mu_n} \sqrt{\ln\left(\sqrt{2} N(2^{-k}(2\kappa\bar{\rho}+1)^d \bar{F}, \mathbb{F}_n, d_n)\right)}
\end{aligned}
$$

where $N(2^{-k}(2\kappa\bar{\rho}+1)^d\bar{F}, \mathbb{F}_n, d_n)$ is the covering number. This allows us to provide an entropy bound on the Rademacher complexity:

$$
\begin{aligned}
&\mathbb{E}_{\sigma_{1:\mu_n}} \left| \sum_{j=1}^{\mu_n} \sigma_j f^*_{C_n^j} \right| \\
&\leq \sum_{k=1}^{\bar{K}} 6 \cdot 2^{-k}(1+2\kappa\bar{\rho})^d \bar{F} \sqrt{\mu_n} \sqrt{\ln\left(\sqrt{2} N(2^{-k}(2\kappa\bar{\rho}+1)^d \bar{F}, \mathbb{F}_n, d_n)\right)} \\
&\leq 12(1+2\kappa\bar{\rho})^d \bar{F} \sqrt{\mu_n} \int_0^1 \sqrt{\ln\left(\sqrt{2} N(\epsilon(2\kappa\bar{\rho}+1)^d \bar{F}, \mathbb{F}_n, d_n)\right)} d\epsilon
\end{aligned}
$$

A result from van der Vaart, Wellner (1996) provides a bound for the covering number in terms of the VC dimension.

$$
N(\epsilon(2\kappa\bar{\rho}+1)^d \bar{F}, \mathbb{F}_n, d_n)) \leq C_3(v+1)(16e)^{v+1}\left(\frac{1}{\epsilon}\right)^{2v}
$$

where C_3 is a universal constant. Then,

$$
\mathbb{E}_{\sigma_{1:\mu_n}} \left| \sum_{j=1}^{\mu_n} \sigma_j f^*_{C_n^j} \right| \leq C_2(1+2\kappa\bar{\rho})^d \bar{F} \sqrt{\mu_n v}
$$

Consequently, the main object of interest can be bounded:

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{D}_n}\left[\mathbb{E}_{Z_{1:n}|\mathcal{D}_n}\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^n f(Z_i) - \mathbb{E}f\right|\right]\right] \\
&\leq 2^{d+1}\mathbb{E}_{\mathcal{D}_n}\left[C_2(1+2\kappa\bar{\rho})^d \bar{F}\sqrt{\frac{v}{\mu_n}}\right]
\end{aligned}
$$

Now, writing $\mu_n = \frac{n}{\bar{a}_n}$ and subsequently using Jensen's inequality

$$\begin{aligned}
&\mathbb{E}_{\mathcal{D}_n}\left[\mathbb{E}_{Z_{1:n}|\mathcal{D}_n}\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n} f(Z_i)-\mathbb{E}f\right|\right]\right]\\
&\leq C_2(1+2\kappa\bar{\rho})^d\bar{F}\sqrt{\frac{v}{n}}\sqrt{\mathbb{E}_{\mathcal{D}_n}\bar{a}_n}\\
&\leq 2C_1\left[2^d(1+2\kappa\bar{\rho})^d\right]3/2\bar{F}\sqrt{\frac{v}{n}}
\end{aligned}$$

The last inequality follows from the fact that $a_n^j \leq 2^d(1+2\kappa\bar{\rho})^d$ for all j with probability 1. ■

Lemma 10: Under $\mathbf{Z}_{1:N}$ satisfying assumptions 5-7. Let $\mathcal{F}$ be a class of uniformly bounded functions with $\|f\|_\infty \leq \bar{F}$ for all $f \in \mathcal{F}$. Further, assume that $\mathcal{F}$ is countable and has finite VC dimension. Then, there exists a universal constant C_2' such that:

$$\mathbb{E}_{\mathbf{Z}_{1:N}}\left[\sup_{f\in\mathcal{F}}\left|\mathbb{E}_N f-\mathbb{E}f\right|\right] \leq 2\left[C_2'\bar{F}\sqrt{\frac{v}{N}}\right]$$

Proof: The quantity of interest is:

$$\mathbb{E}_{\mathbf{Z}_{1:N}}\left[\sup_{f\in\mathcal{F}}\left|\mathbb{E}_N f-\mathbb{E}f\right|\right]$$

The steps involved in this proof mirror Lemma A.4 in Kitagawa, Tetenov (2019). The proof is included here for completeness.

In order to symmetrize the process, I create an independent and identically distributed copy $\mathbf{Z}'_{1:N}$ of $\mathbf{Z}_{1:N}$.

$$\mathbb{E}_{\mathbf{Z}_{1:N}}\left[\sup_{f\in\mathcal{F}}\left|\mathbb{E}_N f-\mathbb{E}f\right|\right] = \mathbb{E}_{\mathbf{Z}_{1:N}}\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{N}\sum_{i=1}^{N} f\big(\mathbf{Z}_i\big)-\mathbb{E}_{\mathbf{Z}'_{1:N}}\frac{1}{N}\sum_{i=1}^{N} f\big(\mathbf{Z}'_i\big)\right|\right]$$

An application on Jensen's inequality yields that:

$$\begin{aligned}\mathbb{E}_{\mathbf{Z}_{1:N}}\left[\sup_{f\in\mathcal{F}}\Big|\mathbb{E}_N f - \mathbb{E}f\Big|\right] &\leq \mathbb{E}_{\mathbf{Z}_{1:N},\mathbf{Z}'_{1:N}}\left[\sup_{f\in\mathcal{F}}\Big|\frac{1}{N}\sum_{i=1}^{N} f\big(\mathbf{Z}_i\big) - \frac{1}{N}\sum_{i=1}^{N} f\big(\mathbf{Z}'_i\big)\Big|\right] \\ &= \mathbb{E}_{\mathbf{Z}_{1:N},\mathbf{Z}'_{1:N}}\left[\frac{1}{N}\sup_{f\in\mathcal{F}}\Big|\sum_{i=1}^{N}\Big(f\big(\mathbf{Z}_i\big) - f\big(\mathbf{Z}'_i\big)\Big)\Big|\right]\end{aligned}$$

Now, define a Rademacher sequence $\sigma_{1:N}$ that is independent of both $\mathbf{Z}'_{1:n}$ and $\mathbf{Z}_{1:n}$. Then, $\sigma_i\big(f(\mathbf{Z}_i) - f(\mathbf{Z}'_i)\big) \sim \big(f(\mathbf{Z}_i) - f(\mathbf{Z}'_i)\big)$, and so

$$\begin{aligned}\mathbb{E}_{\mathbf{Z}_{1:N}}\left[\sup_{f\in\mathcal{F}}\Big|\mathbb{E}_N f - \mathbb{E}f\Big|\right] &\leq \mathbb{E}_{\mathbf{Z}_{1:N},\mathbf{Z}'_{1:N},\sigma_{1:N}}\left[\frac{1}{N}\sup_{f\in\mathcal{F}}\Big|\sum_{i=1}^{N}\sigma_i\Big(f\big(\mathbf{Z}_i\big) - f\big(\mathbf{Z}'_i\big)\Big)\Big|\right] \\ &\leq \mathbb{E}_{\mathbf{Z}_{1:N},\mathbf{Z}'_{1:N},\sigma_{1:N}}\left[\frac{1}{N}\sup_{f\in\mathcal{F}}\Big|\sum_{i=1}^{N}\sigma_i f\big(\mathbf{Z}_i\big)\Big| + \Big|\sum_{i=1}^{N}\sigma_i f\big(\mathbf{Z}'_i\big)\Big|\right] \\ &\leq \frac{2}{N}\mathbb{E}_{\mathbf{Z}_{1:N},\sigma_{1:N}}\left[\sup_{f\in\mathcal{F}}\Big|\sum_{i=1}^{N}\sigma_i f\big(\mathbf{Z}_i\big)\Big|\right]\end{aligned}$$

This is known as the Rademacher complexity bound. Fix any $\mathbf{Z}_{1:N}$, then define a N dimensional vector

$$\mathbf{f} \equiv \big(f(\mathbf{Z}_i)\big)_{i=1}^{N}$$

By definition, for any $1 \leq i \leq N$, $f_i \leq \bar{F}$.

Thus, define: $\mathbb{F} \equiv \big\{\mathbf{f} : \|\mathbf{f}\|_\infty \leq \bar{F}\big\} \subset \mathbb{R}^{\mu_n}$. Now, introduce the following Euclidean norm to $\mathbb{F}$:

$$d(\mathbf{f},\mathbf{g}) \equiv \left[\frac{1}{N}\sum_{i=1}^{N}\Big(f_i - g_i\Big)^2\right]^{1/2}$$

Also define the Rademacher complexity maximizing element,

$$\mathbf{f}^* \equiv \big(f_i^*\big)_{i=1}^{N} \in \arg\max_{\mathbf{f}\in\mathbb{F}}\Big|\sum_{i=1}^{N}\sigma_i f_i\Big|$$

Next, I define a sequence of minimal covers of $\mathbb{F}_n$. In order to do so, start by defining

$$\mathbf{f}^{(0)} \equiv (0, \cdots 0)$$

Then, for any $k \geq 1$, denote a minimal cover of radius $2^{-k}\bar{F}$ by M_k. Since $\mathbb{F}$ is totally bounded, there exists a $\bar{K}$ such that $M_{\bar{K}} = \mathbb{F}$. Now, within each minimal cover, define the

Rademacher complexity maximizing element as

$$\forall 1 \leq k \leq \bar{K},\ \mathbf{f}^{(k)} \in \arg\max_{\mathbf{f}\in M_k} \left| \sum_{i=1}^{N} \sigma_i f_i \right|$$

A straightforward application of triangle inequality yields:

$$d\Big(\mathbf{f}^{(k)}, \mathbf{f}^{(k-1)}\Big) \leq 3 \cdot 2^{-k} \bar{F}$$

Consequently, using the telescopic sum representation:

$$\begin{aligned}
&\mathbb{E}_{\sigma_{1:N}} \left| \sum_{i=1}^{N} \sigma_i f_i^* \right| \\
&\leq \sum_{k=1}^{\bar{K}} \mathbb{E}_{\sigma_{1:N}} \left| \sum_{i=1}^{N} \sigma_i \left(f_i^{(k)} - f_i^{(k-1)} \right) \right| \\
&\leq \sum_{k=1}^{\bar{K}} \mathbb{E}_{\sigma_{1:N}} \max_{\mathbf{f}\in M_k, \mathbf{g}\in M_{k-1} : d(\mathbf{f},\mathbf{g}) \leq \cdot 2^{-k}\bar{F}} \left| \sum_{i=1}^{N} \sigma_i \Big(f_i - g_i \Big) \right|
\end{aligned}$$

Now, I'd like to use the maximal inequality to provide a bound for the above. In order to do, I fist verify that an exponential bound holds. For this, I use Hoeffding's inequality:

$$\begin{aligned}
\mathbb{E}_{\sigma_{1:N}} \left(e^{s \sum_{i=1}^{N} \sigma_j (f_i - g_i)} \right) &= \prod_{i=1}^{N} \mathbb{E}_{\sigma_i} \left(e^{s\sigma_i (f_i - g_i)} \right) \\
&\leq \prod_{i=1}^{N} e^{s^2 (f_i - g_i)^2/2} \\
&= e^{s^2 N d(\mathbf{f},\mathbf{g})^2/2} \\
&\leq e^{s^2 N \cdot (3 \cdot 2^{-k} \bar{F})^2/2}
\end{aligned}$$

This exponential bound allows us to write the maximal inequality bound:

$$\begin{aligned}
&\mathbb{E}_{\sigma_{1:N}} \max_{\mathbf{f}\in M_k,\mathbf{g}\in M_{k-1}:d(\mathbf{f},\mathbf{g})\leq \cdot 2^{-k}\bar{F}} \left| \sum_{i=1}^{N} \sigma_i \Big(f_i - g_i \Big) \right| \\
&\leq 3 \cdot 2^{-k} \bar{F} \sqrt{N} \sqrt{2 \ln(2|M_k|^2)} \\
&= 6 \cdot 2^{-k} \bar{F} \sqrt{N} \sqrt{\ln \big(\sqrt{2} N(2^{-k}\bar{F}, \mathbb{F}, d)\big)}
\end{aligned}$$

where $N(2^{-k}\bar{F}, \mathbb{F}, d)$ is the covering number. This allows us to provide an entropy bound on the Rademacher complexity:

$$\begin{aligned}
&\mathbb{E}_{\sigma_{1:N}} \left| \sum_{i=1}^{N} \sigma_i f_i^* \right| \\
&\leq 6 \cdot 2^{-k} \bar{F} \sqrt{N} \sqrt{\ln \big(\sqrt{2} N(2^{-k}\bar{F}, \mathbb{F}, d)\big)} \\
&\leq 12 \bar{F} \sqrt{N} \int_0^1 \sqrt{\ln \big(\sqrt{2} N(\epsilon\bar{F}, \mathbb{F}, d)\big)} d\epsilon
\end{aligned}$$

A result from Van Der Vaart, Wellner (1996) provides a bound for the covering number in terms of the VC dimension.

$$N(\epsilon\bar{F}, \mathbb{F}, d)\big) \leq C_3'(v+1)(16e)^{v+1} \left(\frac{1}{\epsilon}\right)^{2v}$$

where C_3 is a universal constant. Then,

$$\mathbb{E}_{\sigma_{1:N}} \left| \sum_{i=1}^{N} \sigma_i f_i^* \right| \leq C_2' \bar{F} \sqrt{Nv}$$

Consequently, the main object of interest can be bounded:

$$\mathbb{E}_{\mathbf{Z}_{1:N}} \left[\sup_{f\in\mathcal{F}} \Big| \mathbb{E}_N f - \mathbb{E} f \Big| \right] \leq 2 \left[C_2' \bar{F} \sqrt{\frac{v}{N}} \right] \quad \blacksquare$$

6.1 Theorem 1: Upper bound for local spillovers and exchangeability under increasing number of nodes asymptotics

Assumption 18. *Bounded outcome space*
Suppose that $Y_i \in \left[\frac{-M}{2}, \frac{+M}{2}\right]$.

Theorem 1. *Finite sample bound for Empirical Welfare Maximization.* Under assumptions in Lemma 5, Lemma 9, as well as assumption 18:

$$\sup_{\mathbf{P}\in\mathcal{P}} \mathbb{E}_{Data\sim(\mathbf{P},F)}\left[W^*_{\mathcal{G}} - W(\hat{\mathbf{G}}_{EWM})\right] \leq 4QC_1\left[2^d(1+2\kappa\bar{\rho})^d\right]3/2\frac{M}{2\bar{\gamma}}\sqrt{\frac{v}{n}}$$

Proof: I now provide a bound for $\mathbb{E}_{Data\sim(\mathbf{P},F)}\left[\sup_{G\in\mathbb{G}} |W^q_n(G) - W^q(G)|\right]$ that is independent of $\mathbf{P}$ and hence applies uniformly over $\mathcal{P}$. To that end, define $Z_i \equiv \tilde{Y}_i, r_i(\tilde{\mathbf{D}};\tilde{\mathbf{A}}), \tilde{X}_i$ and

$$f(Z_i;G) \equiv \frac{\tilde{Y}_i 1\{r_i(\tilde{\mathbf{D}};\tilde{\mathbf{A}}) \in D^r_j(\mathbf{T},G), \tilde{X}_i \in G|_{X_j}\}}{\mathbf{P}_\sim[r_i(\tilde{\mathbf{D}};\tilde{\mathbf{A}}) \in D^r_j(\mathbf{T},G), \tilde{X}_i \in G|_{X_j}]}$$

Then, from the unbiasedness results of the cell-mean estimator:

$$\mathbb{E}_n\big(f(\cdot;G)\big) \equiv \frac{1}{n}\sum_{i=1}^n f(U_i;G) = W^q_n(G) \text{ and } \mathbb{E}_{Data\sim(\mathbf{P},F)}\big(f(\cdot;G)\big) = \mathbb{E}_{Data\sim(\mathbf{P},F)}[f(U_i;G)] = W^q(G)$$

Denote by $\mathcal{F} \equiv \{f(\cdot;G) : G \in \mathbb{G}\}$ the collection of all such functions generated by decision sets. By Lemma A.1 of Kitagawa-Tetenov (2018), $\mathcal{F}$ is a VC-subgraph class of functions with VC dimension less than or equal to v, when $\mathbb{G}$ has VC dimension v. Thus, one can write

$$\mathbb{E}_{Data\sim(\mathbf{P},F)}\left[\sup_{G\in\mathbb{G}} |W^q_n(G) - W^q(G)|\right] \leq \mathbb{E}_{Data\sim(\mathbf{P},F)}\left[\sup_{f\in\mathcal{F}} |\mathbb{E}_n(f) - \mathbb{E}(f)|\right]$$

Using lemma 9, with $\bar{F} = \frac{M}{2\gamma}$, I get the following bound:

$$\sup_{\mathbf{P}\in\mathcal{P}} \mathbb{E}_{Data\sim(\mathbf{P},F)}\left[W^*_{\mathcal{G}} - W(\hat{\mathbf{G}}_{EWM})\right] \leq 4QC_1\left[2^d(1+2\kappa\bar{\rho})^d\right]3/2\frac{M}{2\bar{\gamma}}\sqrt{\frac{v}{n}}$$

6.2 Theorem 2: Upper bound for local spillovers and exchangeability under increasing number of networks asymptotics

Theorem 2. *Finite sample bound for Empirical Welfare Maximization.* Under assumptions in Lemma 6 and Lemma 9:

$$\sup_{\mathbf{P}\in\mathcal{P}} \mathbb{E}_{Data\sim(\mathbf{P},F)}\left[W^*_{\mathcal{G}} - W(\hat{\mathbf{G}}_{EWM})\right] \leq 4QC_1\left[2^d(1+2\kappa\bar{\rho})^d\right]3/2\frac{1}{\bar{\gamma}}\sqrt{\frac{v}{n}}$$

Proof: I now provide a bound for $\mathbb{E}_{Data\sim(\mathbf{P},F)}\left[\sup_{G\in\mathbb{G}}|W^q_n(G) - W^q(G)|\right]$ that is independent of $\mathbf{P}$ and hence applies uniformly over $\mathcal{P}$. To that end, define $Z_i \equiv \tilde{Y}_i, c_i(\tilde{\mathbf{D}}; \tilde{\mathbf{A}}, \tilde{\mathbf{X}})$ and

$$f(Z_i;G) \equiv \frac{\tilde{Y}_i 1\{c_i(\tilde{\mathbf{D}};\tilde{\mathbf{A}},\tilde{\mathbf{X}}) \in D^c_j(\mathbf{T},G)\}}{\mathbf{P}_\sim[c_i(\tilde{\mathbf{D}};\tilde{\mathbf{A}},\tilde{\mathbf{X}}) \in D^c_j(\mathbf{T},G)]}$$

Then, from the unbiasedness results of the cell-mean estimator:

$$\mathbb{E}_n\big(f(\cdot;G)\big) \equiv \frac{1}{n}\sum_{i=1}^{n} f(U_i;G) = W^q_n(G) \text{ and } \mathbb{E}_{Data\sim(\mathbf{P},F)}\big(f(\cdot;G)\big) = \mathbb{E}_{Data\sim(\mathbf{P},F)}[f(U_i;G)] = W^q(G)$$

Denote by $\mathcal{F} \equiv \{f(\cdot;G) : G \in \mathbb{G}\}$ the collection of all such functions generated by decision sets. By Lemma A.1 of Kitagawa-Tetenov (2018), $\mathcal{F}$ is a VC-subgraph class of functions with VC dimension less than or equal to v, when $\mathbb{G}$ has VC dimension v. Thus, one can write

$$\mathbb{E}_{Data\sim(\mathbf{P},F)}\left[\sup_{G\in\mathbb{G}}|W^q_n(G) - W^q(G)|\right] \leq \mathbb{E}_{Data\sim(\mathbf{P},F)}\left[\sup_{f\in\mathcal{F}}|\mathbb{E}_n(f) - \mathbb{E}(f)|\right]$$

Using lemma 9, with $\bar{F} = \frac{1}{1\gamma}$, I get the following bound:

$$\sup_{\mathbf{P}\in\mathcal{P}} \mathbb{E}_{Data\sim(\mathbf{P},F)}\left[W^*_{\mathcal{G}} - W(\hat{\mathbf{G}}_{EWM})\right] \leq 4QC_1\left[2^d(1+2\kappa\bar{\rho})^d\right]3/2\frac{1}{1\bar{\gamma}}\sqrt{\frac{v}{n}} \quad \blacksquare$$

6.3 Theorem 3: Upper bound under increasing number of nodes asymptotics

Note that in the increasing number of nodes case, since the planner needs to remain agnostic about the nature of the spillover, the same bound applies for both cases.

Theorem 3. *Finite sample bound for Empirical Welfare Maximization.* Under assumptions in Lemma 8 and Lemma 10:

$$\sup_{\mathbf{P}\in\mathcal{P}} \mathbb{E}_{Data\sim(\mathbf{P},F)}\left[W^*_{\mathcal{G}} - W(\hat{\mathbf{G}}_{EWM})\right] \leq 4QC_1\frac{M}{2\bar{\gamma}}\sqrt{\frac{v}{N}}$$

Proof: I now provide a bound for $\mathbb{E}_{Data\sim(\mathbf{P},F)}\left[\sup_{G\in\mathbb{G}} |W^q_n(G) - W^q(G)|\right]$ that is independent of $\mathbf{P}$ and hence applies uniformly over $\mathcal{P}$. To that end, define $Z_i \equiv \tilde{Y}_i, c_i(\tilde{\mathbf{D}}; \tilde{\mathbf{A}}, \tilde{\mathbf{X}})$ and

$$f(Z_i; G) \equiv \frac{\tilde{Y}_i 1\{c_i(\tilde{\mathbf{D}}; \tilde{\mathbf{A}}, \tilde{\mathbf{X}}) \in D^c_j(\mathbf{T}, G)\}}{\mathbf{P}_{\sim}[c_i(\tilde{\mathbf{D}}; \tilde{\mathbf{A}}, \tilde{\mathbf{X}}) \in D^c_j(\mathbf{T}, G)]}$$

Then, from the unbiasedness results of the cell-mean estimator:

$$\mathbb{E}_n\big(f(\cdot; G)\big) \equiv \frac{1}{n}\sum_{i=1}^{n} f(U_i; G) = W^q_n(G) \text{ and } \mathbb{E}_{Data\sim(\mathbf{P},F)}\big(f(\cdot; G)\big) = \mathbb{E}_{Data\sim(\mathbf{P},F)}[f(U_i; G)] = W^q(G)$$

Denote by $\mathcal{F} \equiv \{f(\cdot; G) : G \in \mathbb{G}\}$ the collection of all such functions generated by decision sets. By Lemma A.1 of Kitagawa-Tetenov (2018), $\mathcal{F}$ is a VC-subgraph class of functions with VC dimension less than or equal to v, when $\mathbb{G}$ has VC dimension v. Thus, one can write

$$\mathbb{E}_{Data\sim(\mathbf{P},F)}\left[\sup_{G\in\mathbb{G}} |W^q_n(G) - W^q(G)|\right] \leq \mathbb{E}_{Data\sim(\mathbf{P},F)}\left[\sup_{f\in\mathcal{F}} |\mathbb{E}_n(f) - \mathbb{E}(f)|\right]$$

Using lemma 9, with $\bar{F} = \frac{1}{1\gamma}$, I get the following bound:

$$\sup_{\mathbf{P}\in\mathcal{P}} \mathbb{E}_{Data\sim(\mathbf{P},F)}\left[W^*_{\mathcal{G}} - W(\hat{\mathbf{G}}_{EWM})\right] \leq 4QC_1\frac{\max\{1, M\}}{2\bar{\gamma}}\sqrt{\frac{v}{N}} \quad \blacksquare$$

6.4 Theorem 4: Lower bound for local spillovers and exchangeability under increasing number of networks asymptotics

This section is work in progress. The theorem should states that no other procedure should be able to attain a faster rate of convergence (measured in terms of regret). Intuitively, this result is useful because it expresses that the bound attained is tight (at least in terms of the rate).

Theorem 4: Under assumptions 5 and 9, for any statistical treatment rule $\hat{\mathcal{Z}}$, it holds that

$$\sup_{\mathbf{P}\in\mathcal{P}} \mathbb{E}_{Data\sim(\mathbf{P},F)}\Big[W^*_{\mathcal{G}} - W(\hat{\mathbf{G}}_{EWM})\Big] \geq \frac{1}{2}\sqrt{\frac{v}{N}}e^{-2\sqrt{2}}$$

as long as $J^2N \geq 16v$.

Proof: Recall that

$$\mathcal{Z}(\cdot|\text{Data}) : \mathcal{S} \to \{0,1\}^J$$

Thus, the treatment assignment rules are characterized by the 2^J-way partition on $\mathcal{S}$, $G^{1:Q} \equiv G^1, \cdots, G^Q$ where $Q = 2^J$ and each $G^q \in \mathcal{G}$ for all $1 \leq q \leq Q$. Thus, the village level welfare can be written as:

$$W(G^{1:Q}) = \sum_{j=1}^{J} \mathbb{E}\big[\sum_{q=1}^{Q} Y_j(\mathbf{T}_q|\mathbf{A})1\big\{(\mathbf{X},\mathbf{A}) \in G^q\big\}\big]$$

where $\mathbf{X} = (X_1, \cdots, X_J)$ and $\big\{Y_j(\mathbf{T}|\mathbf{A}) : \mathbf{T} \in \{0,1\}^J, \mathbf{A} \in \mathscr{A}\big\}$ is the exhaustive list of potential outcomes. Under the assumptions of local spillovers and exchangeablity, the set of effective treatments $\big(r_j(\mathbf{T};\mathbf{A}) = (T_j, \sum_k \mathbf{A}_{jk}T_k\big)$ are reduced to:

$$\mathcal{R} \equiv \big\{(t,s) : t \in \{0,1\}, s \in \mathbb{N}\big\}$$

As a result, the set of potential outcomes now reduces to:

$$\{Y(r) : r \in \mathcal{R}\}$$

6.4.1 Step 1: construct $\mathcal{P}^*$

Since $\mathcal{G}$ is assumed to have finite VC dimension, there exists $\{u_1, \cdots, u_v\} \subset \mathcal{S}$ that is shattered by $\mathcal{G}$.

1. $\forall \mathbf{P} \in \mathcal{P}^*$, $(\mathbf{X},\mathbf{A})$ has uniform support over $\{u_1, \cdots, u_v\}$.

2. $\forall \mathbf{P} \in \mathcal{P}^*$, $D_j \perp\!\!\!\perp \{Y(r) : r \in \mathcal{R}\}, \mathbf{X}, \mathbf{A}$ and D_j are IID across j with $\mathbf{P}(D_j = 1) = \mathbf{P}(D_j = 0) = 1/2$.

3. In this version, we characterize the outcome distributions to be without any spillovers.

   - $Y(0,s) = 0$ w.p.1 for all $s \in \mathbb{N}$.

- Elements of $\mathcal{P}^*$ are characterized by bit vector $\mathbf{b} \in \{0,1\}^v$. For any $s \in \mathbb{N}$,

$$\big(Y(1,s)\big|(\mathbf{X},\mathbf{A}) = u_l\big) = \begin{cases} +1/2, \text{ w.p. } 1/2 + \gamma b_l - \gamma(1-b_l) \\ -1/2, \text{ w.p. } 1/2 - \gamma b_l + \gamma(1-b_l) \end{cases}$$

4. Y_j are IID across j.

In this iteration, I also consider the many village framework where $\mathbf{Z}_{1:N}$ should be interpreted as data from N IID villages with J individuals each.

6.4.2 Step 2: characterizing optimal rule given b

Under the allowed distributions in $\mathcal{P}^*$, the representation of the welfare function can be simplified further:

$$\begin{aligned} W(G^{1:Q}) &= \sum_{q=1}^{Q}\sum_{j=1}^{J}\mathbb{E}_{\mathbf{b}}\big[Y_j(\mathbf{T}_q(j),0)\cdot 1\big\{(\mathbf{X},\mathbf{A}) \in G_q\big\}\big] \\ &= \sum_{q=1}^{Q}\sum_{j=1}^{J}\frac{1}{v}\sum_{l=1}^{v}\big[\mathbf{T}_q(j)\cdot\big(\gamma b_l - \gamma(1-b_l)\big)\cdot 1\big\{u_l \in G_q\big\}\big] \\ &= \sum_{q=1}^{Q}\frac{1}{v}\sum_{l=1}^{v}\big[\big(\gamma b_l - \gamma(1-b_l)\big)\cdot 1\big\{u_l \in G_q\big\}\big]\cdot\big[\sum_{j=1}^{J}\mathbf{T}_q(j)\big] \end{aligned}$$

Thus, if $b_l = 1$, it is optimal to treat everyone, i.e. $\mathbf{T} = \mathbf{1}$. While, when $b_l = 0$, treating no body is optimal, i.e. $\mathbf{T} = \mathbf{0}$. Suppose G^0 corresponds to the set which receives $\mathbf{T} = \mathbf{0}$ and G^Q is the set which receives $\mathbf{T} = \mathbf{1}$. Then, I can write:

$$\begin{aligned} G^{*1}_{\mathbf{b}} &= \{u_l : b_l = 1, 1 \leq l \leq v\} \\ G^{*Q}_{\mathbf{b}} &= \{u_l : b_l = 0, 1 \leq l \leq v\} \\ \forall 1 < q < Q,\ G^{*q}_{\mathbf{b}} &= \emptyset \end{aligned}$$

Now, I can compute the welfare under the optimal assignment noting that by construction, this partition is feasible within $\mathcal{G}$:

$$W^*_{\mathbf{b}} = \frac{1}{v}\sum_{l=1}^{v} J\gamma b_l = \frac{J\gamma}{v}\sum_{l=1}^{v} b_l$$

6.4.3 Step 3: Bayes Risk Minimization

Start with noting that $\hat{G}^{1:Q} = (\hat{G}^1, \cdots, \hat{G}^Q)$. I can write:

$$W^*_{\mathbf{b}} - W_{\mathbf{b}}(G^{\hat{1:Q}}) = \sum_{q=1}^{Q} \frac{1}{v} \sum_{l=1}^{v} \sum_{j=1}^{J} \mathbf{T}_q(j) \big[\big(\gamma b_l - \gamma(1-b_l) \big) \big(1\{u_l \in G^{*q}_{\mathbf{b}}\} - 1\{u_l \in \hat{G}^q\ \} \big) \big]$$

The Bayes Risk Minimization problem is:

$$\begin{aligned} \mathbb{E}_{\mathbf{b}} \mathbb{E}_{Z_{1:n}} \big[W^*_{\mathbf{b}} - W_{\mathbf{b}}(G^{\hat{1:Q}}) \big] &= \sum_{q=2}^{Q} \gamma \big(\sum_{j=1}^{J} \mathbf{T}_q(j) \big) \mathbb{E}_{\mathbf{b}} \mathbb{E}_{Z_{1:N}} \big[\mathbf{P}(G^{*1} \cap \hat{G}^q) \big] \\ &+ \sum_{q=1}^{Q-1} \gamma \big(J - \sum_{j=1}^{J} \mathbf{T}_q(j) \big) \mathbb{E}_{\mathbf{b}} \mathbb{E}_{Z_{1:N}} \big[\mathbf{P}(G^{*Q} \cap \hat{G}^q) \big] \end{aligned}$$

For any u_l observe the loss associated with:

$$\begin{aligned} u_l \in \hat{G}^1 &\mapsto J\gamma\pi(b_l = 1|Z_{1:N}) \\ u_l \in \hat{G}^Q &\mapsto J\gamma\big(1 - \pi(b_l = 1|Z_{1:N})\big) \\ u_l \in \hat{G}^q &\mapsto \gamma\big(J - \sum_{j=1}^{J} \mathbf{T}_q(j)\big)\pi(b_l = 1|Z_{1:N}) + \gamma \sum_{j=1}^{J} \mathbf{T}_q(j)\big(1 - \pi(b_l = 1|Z_{1:N})\big) \end{aligned}$$

There are two cases to consider here:

1. When $\pi(b_l = 1|\mathbf{Z}_{1:N}) \leq 1/2$:
   First of all, it is easy to see:

$$J\gamma\pi(b_l = 1|\mathbf{Z}_{1:N}) \leq J\gamma\big(1 - \pi(b_l = 1|\mathbf{Z}_{1:N})\big)$$

   Next,

$$\begin{aligned} &\gamma\big(J - \sum_{j=1}^{J} \mathbf{T}_q(j)\big)\pi(b_l = 1|\mathbf{Z}_{1:N}) + \gamma \sum_{j=1}^{J} \mathbf{T}_q(j)\big(1 - \pi(b_l = 1|\mathbf{Z}_{1:N})\big) \\ &\geq \gamma\big(J - \sum_{j=1}^{J} \mathbf{T}_q(j)\big)\pi(b_l = 1|\mathbf{Z}_{1:N}) + \gamma \sum_{j=1}^{J} \mathbf{T}_q(j)\pi(b_l = 1|\mathbf{Z}_{1:N}) \\ &= J\gamma\pi(b_l = 1|\mathbf{Z}_{1:N}) \end{aligned}$$

Thus, $u_l \in \hat{G}^1$ is the optimal assignment.

2. When $\pi(b_l = 1|\mathbf{Z}_{1:N}) > 1/2$:
   First of all, it is easy to see:

$$J\gamma\big(1 - \pi(b_l = 1|\mathbf{Z}_{1:N})\big) < J\gamma\pi(b_l = 1|\mathbf{Z}_{1:N})$$

   Next,

$$\begin{aligned}
&\gamma\big(J - \sum_{j=1}^{J}\mathbf{T}_q(j)\big)\pi(b_l = 1|\mathbf{Z}_{1:N}) + \gamma\sum_{j=1}^{J}\mathbf{T}_q(j)\big(1 - \pi(b_l = 1|\mathbf{Z}_{1:N})\big)\\
&< \gamma\big(J - \sum_{j=1}^{J}\mathbf{T}_q(j)\big)\big(1 - \pi(b_l = 1|\mathbf{Z}_{1:N})\big) + \gamma\sum_{j=1}^{J}\mathbf{T}_q(j)\big(1 - \pi(b_l = 1|\mathbf{Z}_{1:N})\big)\\
&= J\gamma\big(1 - \pi(b_l = 1|\mathbf{Z}_{1:N})\big)
\end{aligned}$$

   Thus, $u_l \in \hat{G}^Q$ is the optimal assignment.

Now, I can move on to computing the Minimized Bayes Risk (MBR):

$$MBR = J\gamma\mathbb{E}_{\mathbf{b}}\mathbb{E}_{\mathbf{Z}_{1:N}}\big[\min\big\{\pi(b_l = 1|\mathbf{Z}_{1:N}), 1 - \pi(b_l = 1|\mathbf{Z}_{1:N})\big\}\big]$$

I evaluate $\pi(b_l = 1|\mathbf{Z}_{1:N})$ in a setting where the data comes from n villages of size J each.

$$\pi(b_l = 1|\mathbf{Z}_{1:N}) = \frac{(1/2+\gamma)^{\bar{k}_l^+}(1/2-\gamma)^{\bar{k}_l^-}}{(1/2+\gamma)^{\bar{k}_l^+}(1/2-\gamma)^{\bar{k}_l^-}(1/2-\gamma)^{\bar{k}_l^+}(1/2+\gamma)^{\bar{k}_l^-}}$$

where

$$\begin{aligned}
\bar{k}_l^+ &= \sum_{i=1}^{N} 1\{(\mathbf{X}^{(i)}, \mathbf{A}^{(i)}) = u_l\} \cdot \big|\{j : Y_j^{(i)} D_j^{(i)} = +1/2\}\big|\\
\bar{k}_l^- &= \sum_{i=1}^{N} 1\{(\mathbf{X}^{(i)}, \mathbf{A}^{(i)}) = u_l\} \cdot \big|\{j : Y_j^{(i)} D_j^{(i)} = -1/2\}\big|
\end{aligned}$$

Now, I follow similar steps to that in KT to obtain a lower bound for the MBR.

$$MBR = \frac{J\gamma}{v}\sum_{l=1}^{v}\mathbb{E}\left[\frac{1}{1 + a^{|\bar{k}_l^+ - \bar{k}_l^-|}}\right]$$

where $a = \frac{1+2\gamma}{1-2\gamma} > 1$ and $|\bar{k}_l^+ - \bar{k}_l^-| = \left|\sum_{i:(\mathbf{X}^{(i)},\mathbf{A}^{(i)})=u_l}\sum_{j=1}^{J} 2Y_j^{(i)}D_j^{(i)}\right|$. Since $a > 1$ then

$1 + a^{|}x| > 2$ for any x. So,

$$MBR \geq \frac{J\gamma}{2v}\sum_{l=1}^{v}\mathbb{E}\left[\frac{1}{a^{|\bar{k}_l^+ - \bar{k}_l^-|}}\right]$$

Using Jensen's inequality with $f(x) = 1/a^x$, for any $x \geq 0$. Here, f is convex, so

$$MBR \geq \frac{J\gamma}{2v}\sum_{l=1}^{v} a^{-\mathbb{E}|\bar{k}_l^+ - \bar{k}_l^-|}$$

I now evaluate $\mathbb{E}|\bar{k}_l^+ - \bar{k}_l^-|$.

$$\begin{aligned}
\mathbb{E}|\bar{k}_l^+ - \bar{k}_l^-| &= \mathbb{E}\left|\sum_{i:(\mathbf{X}^{(i)},\mathbf{A}^{(i)})=u_l}\sum_{j=1}^{J} 2Y_j^{(i)}D_j^{(i)}\right| \\
&\leq \sum_{j=1}^{J}\mathbb{E}\left|\sum_{i:(\mathbf{X}^{(i)},\mathbf{A}^{(i)})=u_l} 2Y_j^{(i)}D_j^{(i)}\right|
\end{aligned}$$

The inequality follows from triangular inequality. Now, using that for any i, j, $\mathbf{P}(D_j^{(i)} = 1) = \mathbf{P}(D_j^{(i)} = 0) = 1/2$ and is IID across i, j,

$$\begin{aligned}
\mathbb{E}\left|\sum_{i:(\mathbf{X}^{(i)},\mathbf{A}^{(i)})=u_l} 2Y_j^{(i)}D_j^{(i)}\right| &= \sum_{k=0}^{n}\binom{N}{k}\left(\frac{1}{2v}\right)^k\left(1-\frac{1}{2v}\right)^{N-k}\mathbb{E}\big|Bin(k,1/2) - k/2\big| \\
&\leq \sum_{k=0}^{n}\binom{N}{k}\left(\frac{1}{2v}\right)^k\left(1-\frac{1}{2v}\right)^{N-k}\mathbb{E}\sqrt{k/4} \\
&\leq \mathbb{E}\sqrt{\frac{Bin(N,1/2)}{4}} \\
&\leq \sqrt{\frac{\mathbb{E}[Bin(N,1/2)]}{4}} \\
&= \sqrt{\frac{N}{8v}}
\end{aligned}$$

The last inequality follows by Jensen's inequality. Thus,

$$\mathbb{E}|\bar{k}_l^+ - \bar{k}_l^-| \leq J\sqrt{\frac{N}{8v}}$$

Then:

$$MBR \geq \frac{J\gamma}{2}a^{-J\sqrt{\frac{N}{8v}}}$$

Using that $1 + x \leq e^x$, $\forall x$:

$$MBR \geq \frac{J\gamma}{2} e^{-J\sqrt{\frac{N}{8v}}\frac{4\gamma}{1-2\gamma}}$$

Set $\gamma = \frac{1}{J}\sqrt{\frac{v}{N}}$:

$$\begin{aligned} MBR &\geq \frac{1}{2}\sqrt{\frac{v}{N}} e^{-\frac{4}{1-2\gamma}\frac{1}{2\sqrt{2}}} \\ &= \frac{1}{2}\sqrt{\frac{v}{N}} e^{-\frac{\sqrt{2}}{1-2\gamma}} \end{aligned}$$

If $1 - 2\gamma \geq 1/2 \equiv N \geq \frac{16v}{J^2}$, then

$$MBR \geq \frac{1}{2}\sqrt{\frac{v}{N}} e^{-2\sqrt{2}}$$

This bound is identical to Kitagawa, Tetenov. ■

7 Broader Applications

This paper is primarily concerned with establishing theoretical properties of the Empirical Welfare Maximization procedure when extended to accommodate for spillovers in treatments. While the results have been established with deadly communicable diseases such as Ebola, HIV and malaria in mind, the tools developed herein have wider applicability. In this section, I suggest some broader applications.

7.1 Saving Amazon Rainforests

In the past three decades, deforestation has taken a dramatic toll on the Amazon rainforests. As a consequence, in 2008, the Brazilian government released a Priority List of 36 municipalities with high levels of deforestation. These municipalities were subject to rigorous monitoring and stricter penalty to deforestation.

Stricter monitoring in one municipality is believed to discourage illegal deforestation in neighboring municipalities too. Thus, municipalities of the Amazon rainforest can be thought of as vertices of a network. Two municipalities are said to be linked by an edge if they are geographic neighbors. Assunção, McMillan, Murphy, Souza-Rodrigues (2019) estimate the treatment effect and the spillovers associated with this policy. They then use these estimates to compute the optimal set of municipalities to target in the Priority List.

The results in this paper can be extended to establish theoretical guarantees for Empirical Welfare Maximization in optimally targeting municipalities.

7.2 Marketing

In a different paper[13], the authors explore an example from marketing. MS Office is a licensed software offers a suite of document, spreadsheet and presentation editors. It also allows users to share documents and work collaboratively between themselves. MS Office recently launched Office Lite, a web based service with limited functionality. The introduction of Office Lite has two opposing effects on the overall purchase of MS Office licenses. On the one hand, the introduction of a free limited-feature product cannibalized the existing product. On the other hand, offering a free version allows for a larger collaborator network. This might induce positive externalities on users causing them to upgrade to the full feature version of MS Office.

The paper quantifies this trade-off and suggests a profit maximizing roll out of MS Office Lite to selected parts of the collaborator network. The results of this paper, when combined with a model of product choice and network formation can help inform theoretical properties of the Empirical Welfare Maximization procedure in optimal marketing on the collaborator network.

8 Applicability to COVID 19

The on-going COVID-19 pandemic has resulted in an elevated interest around such questions. While a vaccine for COVID-19 is yet to be approved, it is becoming increasingly clear that ramping up supply will be challenging, particularly in its early days. Consequently, governments across the globe will need to ration the limited stocks of the vaccine. This rationing may be done according to one of several objective functions. On example is maximizing the economics activity subject to a tolerable number of infections. In this section, I discuss some adaptations to the presented framework that are required before the results may be transferred to the study of COVID-19.

The first key distinction involves the contact network. This paper assumes that the degree distribution of the contact network is bounded. Depending on the compliance to social distancing and the use of masks in public places, this may or may not be a suitable assump-

[13] Ananth, Molinari, Peng (2020)

tion for COVID-19. Consider first the extreme case where nobody follows social distancing or the use of masks in public settings. Since, the disease is believed to be transmitted primarily through droplets that are expelled by an infected individual coughing, sneezing or talking[14], the implied contact network could be very dense. Consider the example of an asymptomatic but infected individual passing through a crowded area such a public transit or football game[15]. The degree of such an individual on the contact network can be very large. In general, I do not expect the contact network to satisfy the bounded degree assumption is this setting. At the other extreme, consider now a situation where all members of society are perfectly adhering to social distancing requirements and the appropriate use of masks. In such an environment, the bounded degree assumption is likely to hold. In a realistic model of COVID-19, individuals' decisions to comply with social distancing norms would need to be endogenized, adding to its complexity.

A second distinction lies in the appropriate model of disease transmission on the contact network. While the models presented here might be a good fit, much remains to be known about the transmission of COVID-19 as of date. Finally, the objective functions considered by the planner also needs special attention. In this paper, I consider the population objective functions to be the aggregate outcome across members of the population. This would be consistent with the welfare maximization by a utilitarian social planner who only cares about the survival of the members of the population. In the COVID-19 case, governments may have a host of other objective functions which might include, at least in part, maximizing economic activity. Extending the results presented here to such welfare functions would require additional work[16].

9 Simulation Design

This section *will* detail findings from simulation studies that I am yet to conduct. In a series of simulations, I *will* compare the treatment assignment and regret associated with the Empirical Welfare Maximization presented in this paper with some popular immunization protocols suggested in the epidemiology literature.

For a target population with 10 nodes and pilot experiment with 50 nodes:

[14] Center for Disease Control FAQs on Coronavirus Disease 2019 (COVID-19)

[15] Robinson, J. (2020), The Soccer Match that Kicked Off Italy's Coronavirus Disaster, Wall Street Journal, 1 April. Available at: https://www.wsj.com/articles/the-soccer-match-that-kicked-off-italys-coronavirus-disaster-11585752012

[16] See Kitagawa, Tetenov (2019b) for equality minded welfare functions.

Parameter values: (will use the local spillovers + exchangeability framework)

9.1 Empirical welfare maximization

9.2 Treatment at random

9.3 Treating neighbors of randomly selected nodes

9.4 Treating most central nodes

10 Conclusion

This paper provides an estimator for the optimal treatment assignment on networks for two different models of disease propagation. These distinct frameworks formalize different assumptions on the diffusion of the disease and the nature of spillovers to the treatment, thus allowing me to model different communicable diseases. The first framework makes assumptions consistent with the context of malaria while the second details a structural model of disease propagation consistent with Ebola.

References

[1] J. D. Angrist, G. W. Imbens, and D. B. Rubin, *Identification of causal effects using instrumental variables*, Journal of the American statistical Association, 91 (1996), pp. 444–455.

[2] J. Assunção, R. McMillan, J. Murphy, and E. Souza-Rodrigues, *Optimal environmental targeting in the amazon rainforest*, tech. rep., National Bureau of Economic Research, 2019.

[3] F. Ball, T. Britton, K. Y. Leung, and D. Sirl, *A stochastic sir network epidemic model with preventive dropping of edges*, Journal of Mathematical Biology, 78 (2019), pp. 1875–1951.

[4] C. Ballester, A. Calvó-Armengol, and Y. Zenou, *Who's who in networks. wanted: The key player*, Econometrica, 74 (2006), pp. 1403–1417.

[5] A. Banerjee, A. G. Chandrasekhar, E. Duflo, and M. O. Jackson, *The diffusion of microfinance*, Science, 341 (2013).

[6] D. Bhattacharya and P. Dupas, *Inferring welfare maximizing treatment assignment under budget constraints*, Journal of Econometrics, 167 (2012), pp. 168–196.

[7] G. Chowell and H. Nishiura, *Transmission dynamics and control of ebola virus disease (evd): a review*, BMC medicine, 12 (2014), p. 196.

[8] A. F. Cowman, J. Healer, D. Marapana, and K. Marsh, *Malaria: biology and disease*, Cell, 167 (2016), pp. 610–624.

[9] Á. De Paula, S. Richards-Shubik, and E. Tamer, *Identifying preferences in networks with bounded degree*, Econometrica, 86 (2018), pp. 263–288.

[10] P. Domingos and M. Richardson, *Mining the network value of customers*, in Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, 2001, pp. 57–66.

[11] B. S. Graham, *An econometric model of network formation with degree heterogeneity*, Econometrica, 85 (2017), pp. 1033–1063.

[12] J. J. Heckman and E. Vytlacil, *Policy-relevant treatment effects*, American Economic Review, 91 (2001), pp. 107–111.

[13] H. W. Hethcote, *A thousand and one epidemic models*, in Frontiers in mathematical biology, Springer, 1994, pp. 504–515.

[14] K. Hirano and J. R. Porter, *Asymptotics for statistical treatment rules*, Econometrica, 77 (2009), pp. 1683–1701.

[15] P. Holme, *Efficient local strategies for vaccination and network attack*, EPL (Europhysics Letters), 68 (2004), p. 908.

[16] M. O. Jackson and E. Storms, *Behavioral communities and the atomic structure of networks*, Available at SSRN 3049748, (2019).

[17] D. Kempe, J. Kleinberg, and É. Tardos, *Maximizing the spread of influence through a social network*, in Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, 2003, pp. 137–146.

[18] T. Kitagawa and A. Tetenov, *Who should be treated? empirical welfare maximization methods for treatment choice*, Econometrica, 86 (2018), pp. 591–616.

[19] ——, *Equality-minded treatment choice*, Journal of Business & Economic Statistics, (2019), pp. 1–14.

[20] A. S. Klovdahl, *Social networks and the spread of infectious diseases: the aids example*, Social science & medicine, 21 (1985), pp. 1203–1216.

[21] M. P. Leung, *Treatment and spillover effects under network interference*, Review of Economics and Statistics, 102 (2020), pp. 368–380.

[22] C. F. Manski, *Economic analysis of social interactions*, Journal of economic perspectives, 14 (2000), pp. 115–136.

[23] ——, *Statistical treatment rules for heterogeneous populations*, Econometrica, 72 (2004), pp. 1221–1246.

[24] ——, *Identification of treatment response with social interactions*, The Econometrics Journal, 16 (2013), pp. S1–S23.

[25] K. Menzel, *Many player asymptotics for large network formation problems*, in The Econometric Analysis of Network Data, Elsevier, 2020, pp. 183–223.

[26] R. Pastor-Satorras and A. Vespignani, *Immunization of complex networks*, Physical review E, 65 (2002), p. 036104.

[27] L. J. Savage, *The theory of statistical decision*, Journal of the American Statistical association, 46 (1951), pp. 55–67.

[28] A. W. Van Der Vaart and J. A. Wellner, *Weak convergence*, in Weak convergence and empirical processes, Springer, 1996, pp. 16–28.

[29] G. Vazquez-Bare, *Causal spillover effects using instrumental variables*, arXiv preprint arXiv:2003.06023, (2020).

[30] D. Viviano, *Policy targeting under network interference*, 2019.

[31] A. Wald, *Statistical decision functions.*, (1950).