Journal of Machine Learning Research 23 (2022) 1-61 Submitted 10/18; Revised 4/21; Published 11/22

Estimating Causal Effects under Network Interference with Bayesian Generalized Propensity Scores

Laura Forastiere LAURA.FORASTIERE@YALE.EDU
Yale Institute for Network Science
Yale University, New Haven, CT 06520, USA

Fabrizia Mealli FABRIZIA.MEALLI@UNIFI.IT
Department of Statistics, Computer Science, Applications
University of Florence, Florence, 50134, Italy

Albert Wu ALBERTWU@G.HARVARD.EDU
Department of Statistics
Harvard University, Cambridge, MA 02138, USA

Edoardo M. Airoldi AIROLDI@TEMPLE.EDU
Department of Statistics, Operations and Data Science, Fox School of Business
Data Science Institute
Temple University, Philadelphia, PA 19122, USA

Editor: Jon D. McAuliffe

Abstract

Real-world systems are often comprised of interconnected units, and can be represented as networks, with nodes and edges. In a social system, for instance, individuals may have social ties and financial relationships. In these settings, when a node (the unit analysis) is exposed to a treatment, its effects may spill over to connected units; then estimating both the direct effect of the treatment and its spillover effects presents several challenges. First, assumptions about the mechanism through which spillover effects occur along the observed network are required. Second, in observational studies, where the treatment assignment has not been randomized, confounding and homophily are further potential threats to the identification and to the estimation of causal effects, on networks. Here, we make two structural assumptions: (i) neighborhood interference, which assumes interference operates only through a function of the immediate neighbors' treatments, and (ii) unconfoundedness of the individual and neighborhood treatment, which rules out the presence of unmeasured confounding variables, including those driving homophily. Under these assumptions we develop a new covariate-adjustment estimator for direct treatment and spillover effects in observational studies on networks. We proposed an estimation strategy based on a generalized propensity score that balances individual and neighborhood covariates across units under different levels of individual treatment and of exposure to neighbors' treatment. Adjustment for propensity score is performed using a penalized spline regression. Our inference strategy capitalizes on a three-step Bayesian procedure, which allows to take account for the uncertainty in the propensity score estimation, and avoids model feedback. The correlation among connected units is taken into account using a community detection algorithm, and incorporating random effects in the outcome model. All these sources of variability, including variability of treatment assignment, are accounted for in the posterior distribution of the finite-sample causal estimands we target. We design a simulation study

©2022 Laura Forastiere, Fabrizia Mealli, Albert Wu and Edoardo M. Airoldi.

License: CC-BY 4.0, see https://creativecommons.org/licenses/by/4.0/. Attribution requirements are provided at http://jmlr.org/papers/v23/18-711.html.

to assess the performance of the proposed estimator on different network topologies, both on synthetic networks and on real friendship network from the Add-Health study.

Keywords: causal inference, interference, spillover effect, Bayesian inference, network data

1. Introduction

In many areas of social, economic and medical sciences, researchers are interested in assessing not just the association but the causal relationship between two variables, i.e., the exposure to a condition and an outcome variable that is expected to be affected by the exposure. Causality plays a crucial role during the decision-making process, because the decision maker must understand the consequences of specific actions in order to decide which one to implement. Many studies conducted to assess the effect of the exposure to a certain observed condition, as well as many non-experimental or experimental studies designed to evaluate the impact of public policies and programs, are actually aiming at inferring causal effects. Causal conclusions are typically used for predicting causal consequences of a hypothetical intervention that manipulates the exposure or implements the public policy or program eventually by selecting or improving specific components.

Causal inference can be drawn using experimental or non-experimental studies. In the former, the main challenge lies in the design, which involves both the sampling strategy and the assignment of subjects to different experimental conditions. In the latter, the main challenge is the imbalance of observed and unobserved characteristics across individual in different conditions. Under the assumption that all confounders are measured, covariate-adjustment methods are needed to compare similar units in different treatment arms (Imbens & Rubin, 2015; Hernán & Robins, 2018).

Most estimators of causal effects rely on the assumption of no interference between units, that is, a unit's outcome is assumed to depend only on the treatment he received. When, instead, interference is present, common estimators fail to estimate the causal effect of the treatment. Interference mechanism are common in many fields, from economics to epidemiology. For example, in Angelucci & De Giorgi (2009), recipients of conditional cash transfers may share resources with or purchase goods and services from ineligible households who live in the same area; de-worming a group of children may also benefit untreated children by reducing disease transmission (Miguel & Kremer, 2004); incentives or trainings to promote healthy behaviors could have an effect that spreads in the population Kim et al. (2015); the use of bed nets for malaria prevention can also benefit unprotected individuals Forastiere et al. (2016).

In the past two decades the literature on causal inference in the presence of interference has rapidly started to grow, with increasingly rapid advances in both areas of experimental design (e.g., Hudgens & Halloran, 2008; Toulis & Kao, 2013; Ugander et al., 2013; Eckles et al., 2014; Basse & Airoldi, 2018a,b; Doudchenko et al., 2020; Cai et al., 2022) and statistical inference (e.g., Aronow & Samii, 2017; Karwa & Airoldi, 2018). The recently proposed approaches for the estimation of spillover effects can be categorized in the following situations: randomized experiments on clusters (e.g., Hudgens & Halloran, 2008; Perez-Heydrich et al., 2014; Pouget-Abadie et al., 2018), randomized experiments on networks (e.g., Rosenbaum, 2007; Bowers et al., 2013; Aronow, 2012; Athey et al., 2015; Aronow & Samii, 2017;

Sussman & Airoldi, 2018), randomized experiments on an unknown network (Lee Yu et al., 2022), observational studies on clusters (e.g.,Hong & Raudenbush 2006; Tchetgen Tchetgen & VanderWeele 2012; Liu et al. 2016), and observational studies on networks (Van der Laan, 2014; Sofrygin & van der Laan, 2017; Ogburn et al., 2017; Toulis et al., 2018; Forastiere et al., 2020). All the approaches dealing with clusters (e.g., villages or cities) rely on the assumption of 'partial interference', which rules out interference between clusters but assumes that all units belonging to the same cluster interfere with each other. As opposed to this cluster setting, in networks units directly interfere only with the set of units with whom they share a link (e.g. friends, coworkers, classmates). Most papers on spillover effects on networks only allow direct interference from direct network contacts, with few recent attempts to also allow the presence of indirect interference from indirect contacts (Tchetgen Tchetgen et al., 2017; Ogburn et al., 2018).

1.1 Contributions of the Paper

In this paper, we tackle the problem of interference on networks in observational studies. As common in the literature, we restrict interference to the direct neighborhood in the network and we assume that a unit's outcome is affected by a summary the neighbors' treatment status. Under this assumption, treatment and spillover effects are defined by comparisons of the potential outcomes under different values of the individual treatment and the neighborhood treatment. In observational studies, the estimation of these causal effects must face the problem of having both the individual treatment and the exposure to neighbors' treatment not randomized, which results in a covariate imbalance between groups of units observed under a different treatment status and exposed to different neighbors' treatments. As a covariate adjustment method, Van der Laan (2014) and its extensions Sofrygin & van der Laan (2017); Ogburn et al. (2017) proposed the use of a targeted maximum likelihood estimator (TMLE), while Forastiere et al. (2020) developed a generalized propensity score-based estimator. Building on Forastiere et al. (2020), here we extend the proposed generalized propensity score-based estimator to more flexible functional forms of the outcome model and to incorporate neighborhood outcome correlation using a Bayesian non-linear random effect model in combination with a Bayesian imputation approach. In particular, we capitalize on penalized splines to allow a non-linear dependency of the outcome with the neighborhood treatment and the individual and neighborhood propensity scores. In addition, as opposed to most literature on spillover effects, we recognize the possible presence of a correlation between the potential outcomes of connected units and we introduce a random effect model based on dense clusters identified by a community detection algorithm. Furthermore, we

develop a Bayesian estimation method for finite sample causal effects, which relies on a modular technique and the imputation approach to causal inference. As an alternative to the commonly used frequentist and Fisherian perspectives, this paper pioneers the use of Bayesian inference for the estimation of treatment and spillover effects in the presence of interference. The proposed Bayesian methodology allows flexible estimation of a large range of causal effects, incorporates different sources of uncertainty and allows taking into account correlation among neighbors using community random effects. Finally, because Bayesian inference for propensity score-based regression adjustment can incur into the problem of

model feedback, we developed a three-step modular approach approximating the posterior distribution while avoiding the propagation of the bias due to model misspecification.

The remainder of the paper is structured as follows. In Section 2 we review the potential outcome framework and the imputation-based and Bayesian approach to causal inference. We provide an overview of the propensity score-based methods for observational studies for both binary and continuous treatment. In Section 3 we introduce the problem of causal inference on networks and review the concepts of individual and neighborhood propensity scores. Section 4 is dedicated to our proposed Bayesian estimator based on the generalized propensity score and its performance is assessed with a simulation study reported in Section 5. Section 6 concludes the paper.

2. Potential Outcome Framework

The most widely used statistical framework for causal inference is the potential outcome framework (Rubin, 1974, 1978), also known as the Rubin Causal Model (RCM) (Holland, 1986). The first component of the RCM are potential outcomes defined as potential values of the outcome variable that each unit would experience under each level of the treatment condition. Causal effects are then defined as comparisons of potential outcomes under different treatment conditions for the same set of units. The concept of potential outcomes was first proposed by Neyman (1923) in the context of randomized experiments, and was extended to observational studies by Rubin (1974). The fundamental problem of causal inference under the RCM is that, in one study, for each unit at most one of the potential outcomes is observed —the one corresponding to the treatment the unit is actually exposed to— and the other potential outcomes are missing. Therefore, unit-level causal effects are not identifiable without further assumptions. The second component of the RCM is the assignment mechanism: the process that determines which units receive which treatments, hence which potential outcomes are realized and thus can be observed, and which are missing. In randomized experiments the assignment mechanism is under the control of the experimenter. In contrast, in observational studies the assignment mechanism is the unknown process underlying the observed distribution of treatment and in general depends on units' characteristics. Identification of causal effects relies on specific assumptions on the assignment mechanism. The last optional component of the RCM is a model for the potential outcomes and covariates. Incorporating scientific understanding in a probability distribution allows formal probability statements about the causal effects.

2.1 Setup and Notation

Consider a study where we observe a set of N units. Let i be the indicator of a unit in the sample, with $i = 1, \ldots, N$.

The variable the causal effect of which is under investigation can be the exposure to a certain condition (e.g., environmental exposure, socio-economic status, behavior) or a certain treatment or intervention. Throughout we will refer to this variable as treatment. Let $Z_i \in \mathcal{Z}$ be the treatment received by unit i and $Y_i \in \mathcal{Y}$ the observed outcome that we wish to estimate the effect on. For each unit we also collect a vector of baseline covariate $\mathbf{X}_i \in \mathcal{X}$. Let $\mathbf{O} = (\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ denote the observed data in the sample, where $\mathbf{X} = \{\mathbf{X}_i\}_i^N$ is

the collection of baseline covariate across all units, $\mathbf{Z} = \{Z_i\}_i^N$ is the treatment vector in the sample and $\mathbf{Y} = \{Y_i\}_i^N$ is the corresponding outcome vector.

In principle, we should define the potential outcome of each unit as the potential value of the outcome variable that the unit would experience under a specific assignment of the whole treatment vector $\mathbf{Z} = \mathbf{z}$. Under this general definition, a potential outcome is denoted by $Y_i(\mathbf{Z} = \mathbf{z})$ or simply $Y_i(\mathbf{z})$. We then have for each unit $|\mathcal{Z}|^N$ potential outcomes but we can only observe the one corresponding to the treatment vector that was actually observed, i.e., $Y_i(\mathbf{Z})$ $\forall i$. A dimensionality reduction is needed both for the definition and for the identification of causal estimands as comparisons between potential outcomes under different treatment conditions. Restrictions can be given by structural assumptions or by specific assignment mechanisms in randomized experiments.

2.2 SUTVA

The first basic assumption that is typically invoked is the stable unit treatment value assumption (SUTVA; Rubin, 1980). There are two components to this assumption. The first is that there is only one version of each treatment level possible for each unit (consistency). The second is the no interference assumption (Cox, 1958), that is, the treatment of one unit does not affect the potential outcomes of other units, formally:

Assumption 1 *Stable Unit Treatment Value (SUTVA)*

$$\textit{If} \quad Z_i = Z_i' \quad \textit{then} \quad Y_i(\boldsymbol{Z}) = Y_i(\boldsymbol{Z}') \ \forall \boldsymbol{Z}, \boldsymbol{Z}', \forall i$$

Under this assumption we can index potential outcomes of unit i only by the treatment received by unit i, i.e., $Y_i(Z_i = z)$ or simply $Y_i(z)$. Therefore, under SUTVA, for each unit there exists only one potential outcome for each treatment level, with the observed outcome $Y_i = Y_i(Z_i)$ or $Y_i = \sum_{z \in \mathcal{Z}} I(Z_i = z) Y_i(z)$.

2.3 Causal Estimands

Causal estimands are defined as comparisons of potential outcomes under different treatment levels. Unit-level estimands are comparisons at the unit level, while average estimands are average comparisons on the same set units. A common comparison is the average difference. For the (finite) population of units, under SUTVA, this is the SATE or simply ATE:

$$ATE(z, z') = \frac{1}{N} \sum_{i=1}^{N} (Y_i(z) - Y_i(z')).$$

If the set of N units is considered a sample from a larger (finite of size $M \geq N$ or infinite) population, then define population average causal effects, sometimes referred to as PATE:

$$PATE(z, z') = E[Y_i(z) - Y_i(z')].$$

2.4 Causal Inference as a Missing Data Problem

The problem of identifying and estimating causal estimands relates to the fundamental problem of causal inference (Holland, 1986), that is, the inability to simultaneously observe

all the potential outcomes of the same unit. In fact, for each unit, we can observe at most the potential outcome corresponding to the the treatment to which the unit is exposed, i.e., $Y_i = Y_i(Z_i)$, whereas all the other potential outcomes $Y_i(z)$, with $Z_i \neq z$, are missing. The missing potential outcomes are often referred to as 'counterfactuals' in the literature. Therefore, causal inference is inherently a missing data problem and causal effects are not identifiable without further assumptions, which, in general, allow extrapolations of information on missing potential outcomes. The central identifying assumptions concern the assignment mechanism.

Because potential outcomes of the same unit are never simultaneously observed, we cannot in general estimate individual contrasts in potential outcomes. We could estimate a marginal contrast in potential outcomes, e.g., $ATE(z, z')$, if we could recover the distribution of $Y_i(z)$ for units $Z_i \neq z$. This could be done by assuming that the distribution of potential outcomes is independent of the actual treatment received. This assumption is known as *unconfoundedness*, *ignorability*, *exchangeability* or *exogeneity*. Throughout we will use the term unconfoundedness. In the literature we can find alternative expressions of such assumption. Here we report a weak version, which assumes the marginal independence between each potential potential outcome and the treatment receipt. Formally, we have the following assumption.

Assumption 2 *Unconfoundedness of the Treatment*
Given an assignment mechanism, a treatment is unconfounded if it does not depend on the potential outcomes:

$$Y_i(z) \perp\!\!\!\perp Z_i \ \forall z, \forall i$$

This implies that for all individuals treatment assignment does not depend on the potential outcomes and, in turn, the distribution of the potential outcomes is independent of the treatment assignment. The unconfoundedness assumption holds by design in classical randomized experiments, where the assignment mechanism is known and the treatment is randomly assigned. Thus, the randomized assignment mechanism implies that those assigned to $Z_i = z$ and those assigned to $Z_i \neq z$ are exchangeable. The independence of the potential outcomes from the treatment assignment, is more intuitive in the way it is crucial to causal inference. In fact, unconfoundedness implies that $Pr(Y_i(z)|Z_i = z) = Pr(Y_i(z)|Z_i \neq z)$ and thus allows imputation of the missing potential outcomes $Y_i(z)$, with $Z_i \neq z$, from the distribution of observed outcomes in the treatment arm where the treatment received is actually z. This ensures the identifiability of causal effects from the observed data. Specifically, under Assumption 2 $ATE(z, z')$ is identified.

2.5 Imputation Approach to Causal Inference

Estimating causal effects requires properly handling the missing potential outcomes. As a matter of fact, all methods for causal inference can be viewed as imputation methods. In fact, because any causal estimand depends on missing potential outcomes, every estimator of causal estimands must explicitly or implicitly estimate or *impute* the missing potential outcomes.

The unconfoundedness assumption is crucial for this imputation: one can recover the distribution of missing potential outcomes from the distribution of observed outcomes of units in other treatment arms.

The imputation approach to causal inference explicitly imputes the missing potential outcomes for every unit in the sample. This can be done from the empirical distribution of observed outcomes of units in other treatment arms or from a parametric model.

A unit-level causal estimand is a comparison between potential outcomes corresponding to different treatment levels $\tau_i = \tau(\mathbf{X}_i, \{Y_i(z)\}_{z\in\mathcal{Z}})$. A finite sample causal estimand is a sample moment of the distribution of τ_i in the whole sample or in a sub-sample defined by observed treatment and covariates; it can be expressed by $\tau^S = \mu^S(\mathbf{X}, \mathbf{Z}, \boldsymbol{\tau})$. Similarly, a population level causal estimand is a population moment of the population distribution of τ_i unconditional or conditional on a treatment level and/or covariates; it can be expressed by $\tau^P = \mu^P(\mathbf{X}, \mathbf{Z}, \boldsymbol{\tau})$. An imputation-based method hinges on a stochastic model for all potential outcomes, both observed and missing:

$$p(\{\mathbf{Y}(z)\}_{z\in\mathcal{Z}}|\mathbf{X}, \mathbf{Z}; \boldsymbol{\theta}).$$

Such a model generally depends on some unknown parameters $\boldsymbol{\theta}$. The observed data $\mathcal{O} = (\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ are used to learn about these parameters. The postulated model with the estimated parameters is then used to impute the missing potential outcomes, given the observed data, and to conduct inference for the estimands of interest. Typically, potential outcomes of different units are assumed independent and identically distributed. In addition, potential outcomes of a unit corresponding to different treatment levels are considered independent. Therefore, under the unconfoundedness assumption, we posit a model for each potential outcome

$$p(Y_i(z)|\mathbf{X}_i; \boldsymbol{\theta}).$$

Once parameters are estimated and potential outcomes for all units are imputed, the distribution of causal estimands is drawn. Specifically, population average causal estimands τ^P are computed as a function of the parameters $\boldsymbol{\theta}$. In contrast, finite sample causal estimands τ^S are derived using the imputed potential outcomes: first unit-level estimands τ_i are computed from the imputed set of potential outcomes $\{Y_i(z)\}_{z\in\mathcal{Z}}$, then average causal estimands τ^S are determined by sample moments of the distribution of τ_i in the sample.

2.6 Bayesian Causal Inference

The outline of Bayesian inference for causal effects was first proposed in Rubin (1978). From the Bayesian perspective, the observed outcomes are considered to be realizations of random variables and the unobserved potential outcomes are unobserved random variables. Thus, missing potential outcomes are considered as unknown parameters. Bayesian inference for the causal estimands is obtained from the posterior predictive distribution of the missing potential outcomes, given a model for potential outcomes and a prior distribution for the parameters.

Let $\mathbf{Y}_i^{mis} = \{Y_i(z)\}_{z\neq Z_i\in\mathcal{Z}}$ be the vector of missing potential outcomes for each unit and $\mathbf{Y}^{mis}$ the corresponding vector in the whole sample. Given the prior distribution $p(\boldsymbol{\theta})$, under unconfoundedness and common independence assumptions, the posterior predictive

distribution of these missing potential outcomes can be written as

$$p(\mathbf{Y}^{mis}|\mathbf{X},\mathbf{Z},\mathbf{Y}) = \int \prod_{i}^{N} \prod_{z \neq Z_i \in \mathcal{Z}} p(Y_i(z)|\mathbf{X}_i,\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{X},\mathbf{Z},\mathbf{Y}) d\boldsymbol{\theta}. \quad (1)$$

The integrand of (1) is the likelihood function. Equation (1) suggests that, under unconfoundedness and iid units, one only needs to specify the potential outcome distribution $p(Y_i(z)|\mathbf{X}_i;\boldsymbol{\theta})$ and the prior distribution $p(\boldsymbol{\theta})$ to conduct Bayesian inference for causal effects. Closed-form posterior distribution of the causal estimand is generally not available. Instead, we can use Markov Chain Monte Carlo (MCMC) methods such as Gibbs sampling (Gelfand & Smith, 1990); specifically, one can simulate the joint posterior distribution of all unknown random variables, $p(\mathbf{Y}^{mis},\boldsymbol{\theta}|\mathbf{X},\mathbf{Z},\mathbf{Y})$, by iteratively drawing from the two conditional distributions, $p(\mathbf{Y}^{mis}|\mathbf{X},\mathbf{Z},\mathbf{Y},\boldsymbol{\theta})$ and $p(\boldsymbol{\theta}|\mathbf{X},\mathbf{Z},\mathbf{Y},\mathbf{Y}^{mis})$. After obtaining the posterior draws of $\mathbf{Y}^{mis}$ and of the parameters $\boldsymbol{\theta}$, it is straightforward to extract any posterior summary, e.g. mean, variance, quantiles, of any causal estimand. The posterior distribution of population average estimands τ^P is derived from posterior draws of the parameters or functions of the parameters. For finite-sample average estimands τ^S, one would instead obtain the posterior draws of the missing potential outcomes $\mathbf{Y}^{mis}$ and then for each draw compute the sample average of interest.

2.7 Unconfoundedness in Observational Studies

In randomized experiments Assumption 2 holds by design. Conversely, in observational studies the investigator does not control the assignment of treatments and cannot ensure that similar subjects receive different levels of treatment. Therefore, Assumption 2 cannot be ensured by design. Moreover, such an assumption cannot be directly validated by the data because it involves missing quantities, that is, the distribution of a potential outcome $Y_i(z)$ cannot be observed in the treatment arms where $Z_i \neq z$ and, hence, we are unable to say whether it would be the same across all treatment arms. In an observational setting we typically relax the unconditional unconfoundedness assumption by assuming exchangeability conditional on a set of observed covariates. The set of covariates that can make the unconfoundedness assumption more plausible must be chosen based on subject matter knowledge.

The conditional unconfoundedness assumption can be formally stated as follows.

Assumption 3 *Conditional Unconfoundedness of the Treatment*
Given an assignment mechanism, the treatment is unconfounded if it does not depend on potential outcomes conditional on an observed covariate set $\boldsymbol{X}$*:*

$$Y_i(z) \perp\!\!\!\perp Z_i|\boldsymbol{X}_i \ \forall z \in \mathcal{Z}, \forall i$$

Assumption 3 implies that the treatment is as good as randomized among units with the same value of the observed covariates. In other words, we can say that the treatment is unconfounded conditional on an observed set of covariates if there are no other unmeasured confounders, that is, unobserved characteristics that —by affecting both the exposure to treatment and the potential outcomes— are said to confound the relationship between the treatment and the outcome.

Under conditional unconfoundedness, identification of causal effects also requires a nonzero probability of all levels of the treatment for all covariate combinations, also known as the positivity assumption (Hernán & Robins, 2018).

2.8 Propensity Score for Binary Treatment

Much of causal inference literature is concerned with settings with a binary treatment $Z_i \in \{0, 1\}$. In this case, for each unit there are only two potential outcomes $Y_i(0)$ and $Y_i(1)$, one observed and one missing.

Several covariate-adjustment methods, relying on the conditional unconfoundedness assumption, have been proposed. When the set of covariates needed to make the unconfoundedness assumption more plausible is large and includes continuous covariates, a simple stratification in groups where units share the same values of covariates is not possible. In such settings, propensity score-based methods are particularly useful. For each unit, the propensity score, denoted by $\phi(\mathbf{X}_i)$, is defined as the probability of receiving the active treatment ($Z_i = 1$) given the unit's covariates (Rosenbaum & Rubin, 1983). Formally, we can write

$$\phi(\mathbf{X}_i) = Pr(Z_i = 1|\mathbf{X}_i). \tag{2}$$

The propensity score $\phi(\mathbf{X}_i)$ has two important properties: (i) it is a balancing score, that is, it satisfies $\mathbf{X}_i \perp\!\!\!\perp Z_i|\phi(\mathbf{X}_i)$; (ii) if the treatment assignment is unconfounded given $\mathbf{X}_i$, then it is also unconfounded given $\phi(\mathbf{X}_i)$, that is, $Y_i(z) \perp\!\!\!\perp Z_i|\phi(\mathbf{X}_i)$.

Therefore, under unconfoundedness, adjusting for the difference in the propensity scores between treated and control units would remove all biases due to covariate imbalance, i.e., we can compare the observed outcomes between treated and control units within groups defined by the value of the propensity score rather than by the value of covariates. Because the propensity score can be viewed as a scalar summary of the multivariate covariates, the use of the propensity score for covariate adjustment is often more convenient.

In randomized experiments, the propensity score is known and in classical randomized experiments it does not depend on covariates. On the contrary, in observational studies, the propensity score is typically unknown and must be estimated. It has been shown that a consistent estimate of the propensity score leads to more efficient estimation of the ATE than the true propensity score (Rosenbaum, 1987; Rubin & Thomas, 1996; Hahn , 1998; Hirano et al., 2003).

Propensity score methods usually involve two stages: in the first stage ('GPS stage'), the propensity scores $\phi(\mathbf{X}_i)$ are determined by estimating the parameters of a model, usually through a logistic regression ($\text{logit}(Pr(Z_i = 1|\mathbf{X}_i)) = \alpha + \beta_Z^T\mathbf{X}_i$), and then by computing the individual probabilities given covariates; in the second stage ('outcome stage'), estimate the causal effects based on the estimated propensity scores, through three main alternative strategies: matching, stratification, weighting or combination of these methods (for a review, see Stuart 2010 or Austin 2011).

For matching, arguably the most popular causal inference method, the propensity score is used as the distance metric for finding matched pairs of treated and control units and the causal effects are estimated by the average within-pair difference in the observed outcome. The most common method for propensity score matching is one-to-one or pair matching, in which pairs of treated and control subjects are formed such that matched subjects have

similar values of the propensity score. There are different sets of options for forming propensity score matched sets. The first choice is between matching with or without replacement. Matching with replacement allows a given subject to be included in more than one matched set, resulting in closer pairs but higher variance due to the fact that less information is used (Rosenbaum, 2002). On the contrary, in matching without replacement each unit is included in at most one matched set. In this case, one must choose between greedy and optimal matching (Rosenbaum, 2002). The greedy matching method is a sequential process where at each step the nearest control unit is selected for matching to a given treated unit, even if that control unit would better serve as a match for a subsequent treated unit. An alternative to greedy matching is optimal matching, in which matches are formed so as to minimize the total within-pair difference of the propensity score. Different optimization algorithms can be used (e.g., Rosenbaum 2002; Hansen 2004).

When there are large numbers of control individuals, it is sometimes possible to get multiple good matches for each treated individual, called ratio matching (e.g., Rubin & Thomas 1996).

For stratification, one stratifies units into K (usually 5 or 6) subclasses based on quantiles of the propensity scores so that treated and control units in each subclass have similar covariate distribution, and then calculates the difference in the average outcome between treated and control units in each subclass. The stratification estimator is the average of these within-subclass causal estimates weighted by the size of each subclass (Imbens & Rubin, 2015).

For weighting, the basis is an analogy of the Horvitz-Thompson estimator in survey sampling (Horvitz and Thompson, 1952): Therefore, one can define an inverse-probability weight (IPW) for each unit: $\phi(\mathbf{X}_i)$ for treated units and $1-\phi(\mathbf{X}_i)$ for control units, and then the difference of the weighted average outcome between treatment groups is an unbiased estimator of the average causal effect (e.g., Hernán & Robins 2018).

In general, variance estimation for these propensity score-based estimators relies on the common structure of weighted estimators (Imbens & Rubin, 2015). In addition, different empirical formulas have been proposed to account for correlation within matched sets as well as for the uncertainty in the estimation of the propensity score (e.g., Abadie & Imbens 2006, 2009). When researchers want to account for the uncertainty in the propensity score, a bootstrap procedure has been found to outperform other methods (Hill & Reiter, 2006), but standard bootstrap methods can yield invalid inferences when applied to matching estimators (Abadie & Imbens, 2008).

2.9 Propensity Score for Continuous Treatment

In many studies the treatment is not binary but units may receive different treatment levels. In such settings, where the treatment is discrete or continuous, propensity score methods for binary treatment cannot be used. In fact, when the treatment is multivalued there is a propensity score for each treatment level. Thus, adjusting for the difference in the propensity score corresponding to one specific treatment level $z^\star$ does not yield unbiased estimate of causal estimands comparing potential outcomes under different treatment levels, i.e., $ATE(z, z')$ with $z, z \neq z^\star$.

In addition, when the treatment is discrete or continuous, we are usually not only interested in comparing specific treatment levels, instead the focus is on assessing the heterogeneity of treatment effects arising from different amounts of treatment exposure, that is, on estimating an average dose-response function.

Over the last years, propensity score methods have been generalized to the case of discrete treatments (e.g., Imbens 2000; Lechner 2001) and, more recently, continuous treatments and arbitrary treatment regimes (e.g., Hirano & Imbens 2004; Imai & Van Dik 2004; Flores et al. 2012; Kluve et al. 2012; Bia et al. 2014). Here we review the work by Imbens (2000) and Hirano & Imbens (2004), who introduced the concept of the Generalized Propensity Score (GPS) and use it to estimate an average dose-response function (aDRF) $\mu(z) = E[Y_i(z)]$.

2.9.1 Generalized Propensity Score (GPS)

The GPS, denoted by $\lambda(z; \mathbf{x})$, is the conditional density of the treatment given the covariates: $\lambda(z; \mathbf{X}_i) = p(z|\mathbf{X}_i)$. Each unit is then characterized by a different density of the treatment. For each unit the GPS corresponding to the actual treatment to which the unit is exposed, $\Lambda_i = \lambda(Z_i; \mathbf{x})$, is the probability for that unit of receiving the treatment he actually received given his characteristics $\mathbf{X}_i$. As the classic propensity score, the GPS is a balancing score. In the continuous case, this means that, within strata with the same value of $\lambda(z; \mathbf{X}_i)$, the probability that $Z_i = z$ does not depend on the value of $\mathbf{X}_i$. In other words, for units with the same value of the GPS corresponding to a specific treatment level z, the distribution of covariates is the same between the arm assigned to treatment z and all the other arms. Furthermore, the unconfoundedness assumption, combined with the balancing score property, implies that the treatment is unconfounded given the GPS. Formally,

$$Y_i(z) \perp\!\!\!\perp Z_i|\lambda(z; \mathbf{X}_i) \ \forall z \in \mathcal{Z}, \forall i \tag{3}$$

Thus, any bias caused by covariate imbalance across groups with different treatment levels can be removed adjusting for the difference in the GPS. Under unconfoundedness of the treatment given $\mathbf{X}_i$, then

$$\mu(z) = E[Y_i(z)] = E[E[Y_i(z)|, \lambda(z, \mathbf{X}_i)]] = E[E[Y_i|Z_i = z, \lambda(z, \mathbf{X}_i)]] \tag{4}$$

According to Equation (4), for each treatment level z, within strata with same GPS at level z, $\lambda(z, \mathbf{X}_i)$, the average potential outcome corresponding to z, $Y_i(z)$, can be estimated using the average outcome observed for units who were actually exposed to level z of the treatment. Given the continuous nature of the treatment and the GPS, such stratification is unfeasible.

Hirano & Imbens (2004) estimate the DRF using the estimated GPS by employing a parametric partial mean approach. Specifically, we posit a model for the treatment Z_i

$$\mathrm{h}^{(Z)}\big(Z_i\big) \sim f^{(Z)}\left(q^{(Z)}\big(\mathbf{X}_i; \boldsymbol{\beta}^{(Z)}\big), \nu^{(Z)}\right) \tag{5}$$

where $h^{(Z)}(\cdot)$ is a link function, $f^{(Z)}$ is a probability density function (pdf), $q^{(Z)}(\cdot)$ is a flexible function of the covariates depending on a vector of parameters $\boldsymbol{\beta}^{(Z)}$, and $\nu^{(Z)}$ is

a scale parameter. For instance, $h^{(Z)}(\cdot)$ can be the identity function, f is the normal pdf, $q^{(Z)}(\cdot)$ is a linear combination of the covariates, i.e., $q^{(Z)}\big(\mathbf{X}_i;\boldsymbol{\beta}^{(Z)}\big) = \beta_0^{(Z)} + \boldsymbol{\beta}_X^{(Z)T}\mathbf{X}_i$, and $\nu^{(Z)}$ is the standard deviation of Z. We also postulate a model for the potential outcomes given the GPS:

$$\mathrm{h}^{(Y)}\big(Y_i(z)\big) \sim f^{(Y)}\left(q^{(Y)}\big(z,\lambda(z;\mathbf{X}_i);\boldsymbol{\beta}^{(Y)}\big),\nu^{(Y)}\right) \tag{6}$$

where $h^{(Y)}(\cdot)$ is a link function, $f^{(Y)}$ is a probability density function (pdf), $q^{(Y)}(\cdot)$ is a flexible function of the treatment and the GPS depending on a vector of parameters $\boldsymbol{\beta}^{(Y)}$, and $\nu^{(Y)}$ is a scale parameter. In Hirano & Imbens (2004) $q^{(Y)}(\cdot)$ is a cubic polynomial function of the treatment z and the generalized propensity score $\lambda(z;\mathbf{X}_i)$, including the interaction term.

Bia et al. (2014) replace the parametric approach with a semi-parametric estimator based on penalized spline techniques. In particular, they use penalized bivariate splines, with radial basis functions of the form $C||(z,\lambda)-(k,k')||$.

Relying on the two models for the treatment (Equation (5)) and the outcome (Equation(6)), the average DRF is derived using a two-step estimator is used.

Estimating Procedure based on the GPS

We describe here the two-step estimator proposed by Hirano & Imbens (2004) for the estimation of the average DRF based on the generalized propensity score. The first step involves parametrically modeling and estimating the GPS. The second step consists of estimating the average DRF.

Algorithm 1: Generalized Propensity Score for Multivalued Treatment

Input: Data set $\mathcal{D}^\star$, PS model, outcome model
Output: Average DRF $\mu(z)^\star, z \in \mathcal{Z}$

GPS Stage:

Estimate the parameters $\boldsymbol{\beta}^{(\boldsymbol{Z})}$ and $\nu^{(Z)}$ of the GPS model
for $i = 1$ *to* N **do**
  Predict $\widehat{\Lambda}_i = \lambda(Z_i;\mathbf{X}_i)$
end

Outcome Stage:

Estimate the parameters $\boldsymbol{\beta}^{(\boldsymbol{Y})}$ and $\nu^{(Y)}$ of the outcome model, given the data $(\mathbf{Z}^\star,\mathbf{Y}^\star)$ and $\widehat{\boldsymbol{\Lambda}}$
Impute potential outcomes and Compute average DRF:
for $z \in \mathcal{Z}$ **do**
  for $i = 1$ *to* N **do**
    Impute the potential outcome $\widehat{Y}_i(z)$:
      a. Predict the GPS $\phi(z;\mathbf{X}_i)$
      b. Predict $\widehat{Y}_i(z)$, given z, $\lambda(z;\mathbf{X}_i)$, and the estimated parameters $\widehat{\boldsymbol{\beta}^{(\boldsymbol{Y})}}$ and $\widehat{\nu^{(Y)}}$
  end
  Average the potential outcomes over all units: $\widehat{\mu}(z) = \frac{1}{N}\sum_{i=1}^{N}\widehat{Y}_i(z)$
end

Statistical Inference for Population Average DRF

In Hirano & Imbens (2004), as well as in Bia et al. (2014), the target causal estimand is the population average dose-response function $\mu(z)$. In order to assess the sampling variability

of the GPS estimator with respect to the population average DRF, standard errors and 95% confidence intervals are derived using bootstrap methods.

Algorithm 2: Bootstrap for Generalized Propensity Score for Multivalued Treatment

Input: Data set $\mathcal{D}$, number of iterations M
Output: Distribution of average DRF $\mu(z), z \in \mathcal{Z}$

for $m = 1$ *to* M **do**
1 | **for** $k = 1$ *to* N **do**
| | Sample $s_k \sim \mathcal{U}(1, N)$
| **end**
2 | Data set $\mathcal{D}^m = \{(\mathbf{X}_{s_k})_{k=1}^N, (Z_{s_k})_{k=1}^N, (Y_{s_k})_{k=1}^N\}$
3 | Run Algorithm 1 with $\mathcal{D}^m$
| **return** average DRF $\mu(z)^m, z \in \mathcal{Z}$
end

2.10 Bayesian Propensity Score Adjustment

After Rubin (1985) reflected on the usefulness of propensity scores for Bayesian inference, only recently has Bayesian estimation of causal effects been combined with propensity score methods. (Hoshino, 2008; McCandless et al., 2009; An, 2010; McCandless et al., 2010; Kaplan & Chen, 2012; Zigler et al., 2013; McCandless et al., 2012; Zigler & Dominici, 2014). The first advantage of the Bayesian propensity score approach is that it allows embedding propensity score adjustment within broader Bayesian modeling strategies, incorporating prior information as well as complex models for hierarchical data, measurement error or missing data (Rubin, 1985). The second major motivation for using Bayesian inference for propensity scores is the propagation of propensity score uncertainty in the estimation of causal effects. In fact, the traditional frequentist approach accommodates the two-stage nature of propensity score methods with a separate and sequential estimation: the estimated model and the predicted propensity scores from the first stage are treated as fixed and known in the outcome stage. A limitation of this sequential approach is that confidence intervals for the treatment effect estimate are usually calculated without acknowledging uncertainty in the estimated propensity scores. On the contrary, Bayesian methods offer a natural strategy for modeling uncertainty in the propensity scores.

McCandless et al. (2009) proposed to model the joint distribution of the data and parameters with the propensity score as a latent variable. Let $\boldsymbol{\theta}^{(Z)}$ and $\boldsymbol{\theta}^{(Y)}$ be the vectors of parameters of the propensity score and the outcome model, respectively. Markov chain Monte Carlo (MCMC) methods allow to draw from the posterior distribution for model parameters,

$$p(\boldsymbol{\theta}^{(Z)}, \boldsymbol{\theta}^{(Y)} | \mathbf{Y}, \mathbf{Z}, \mathbf{X}) \propto p(\mathbf{Y}, \mathbf{Z}, \mathbf{X} | \boldsymbol{\theta}^{(Z)}, \boldsymbol{\theta}^{(Y)}) p(\boldsymbol{\theta}^{(Z)}) p(\boldsymbol{\theta}^{(Y)})$$

by successively drawing from the full conditional distributions

$$p(\boldsymbol{\theta}^{(Z)} | \mathbf{Y}, \mathbf{Z}, \mathbf{X}, \boldsymbol{\theta}^{(Y)}) \propto \prod_i^N p(Y_i, Z_i, \mathbf{X}_i | \boldsymbol{\theta}^{(Z)}, \boldsymbol{\theta}^{(Y)}) p(\boldsymbol{\theta}^{(Z)})$$

and

$$p(\boldsymbol{\theta}^{(Y)}|\mathbf{Y},\mathbf{Z},\mathbf{X},\boldsymbol{\theta}^{(Z)}) \propto \prod_i^N p(Y_i,Z_i,\mathbf{X}_i|\boldsymbol{\theta}^{(Z)},\boldsymbol{\theta}^{(Y)})p(\boldsymbol{\theta}^{(Y)}).$$

In McCandless et al. (2009), as well as in the whole literature on Bayesian propensity scores, the focus is on binary treatment.

Within the Bayesian framework, we denote the propensity score of a binary treatment by $\phi(\mathbf{X}_i;\boldsymbol{\theta}^{(Z)})$ to highlight the dependence on the parameters. With propensity score adjustment, we do not directly model the dependence of the outcome of covariates, but rather we adjust for covariates by modeling parametrically or semi-parametrically the propensity score $\phi(\mathbf{X}_i;\boldsymbol{\theta}^{(Z)})$. As a consequence, the outcome model indirectly depends on all parameters, including the set of parameters of the propensity score models, i.e., $\boldsymbol{\theta}^{(Z)}$, and, thus, the likelihood cannot be factorized into two parts, $p(Y_i|Z_i,\mathbf{X}_i,\boldsymbol{\theta}^{(Y)})p(Z_i|\mathbf{X}_i,\boldsymbol{\theta}^{(Z)})$, that separately depend on different sets of parameters. Therefore, the posterior distribution of the parameters $\boldsymbol{\theta}^{(Z)}$ of the PS stage are in part informed by the outcome stage. Because of this phenomenon, referred to as 'model feedback' (e.g., McCandless et al. 2010; Zigler et al. 2013), the joint Bayesian PS estimation has raised some concerns. First, since the propensity score adjustment is meant to approximate the design stage in a randomized experiment it should be done without access to the outcome data (Rubin, 2007, 2008). Furthermore, a practical consequence is the propagation of error due to model misspecification. In fact, when the model for the relationship between the outcome and the propensity score is misspecified, then the joint Bayesian approach was shown to provide invalid inferences for $\boldsymbol{\theta}^{(Z)}$, which distorts the balancing property of the propensity score and yields incorrect estimates of the treatment effect (Zigler et al., 2013).

Various methods described as 'two-step Bayesian' have been recently proposed to 'cut the feedback' between the propensity score and outcome stages (Hoshino, 2008; McCandless et al., 2010; Kaplan & Chen, 2012). These methods represent a special case of the so-called 'modularization' in Bayesian inference (Liu et al., 2009). To limit feedback, the general idea is based on an approximate Bayesian technique that uses the posterior distribution of the PS model as an input when fitting the outcome model. Specifically, the posterior distribution of the parameters $\boldsymbol{\theta}^{(Z)}$ of the PS model are not updated from the full conditional, but rather from the approximate conditional distribution

$$p(\boldsymbol{\theta}^{(Z)}|\mathbf{Y},\mathbf{Z},\mathbf{X},\boldsymbol{\theta}^{(Y)}) \propto \prod_i^N p(Z_i|\mathbf{X}_i,\boldsymbol{\theta}^{(Z)})p(\boldsymbol{\theta}^{(Z)})$$

which ignores the likelihood contribution from the outcome. This restricts the flow of information between models during MCMC computation, and is similar in spirit to two-stage estimation (Lunn et al., 2009).

Algorithm 3: Bayesian Two-Stage Propensity Score for Binary Treatment

Input: Data set $\mathcal{D}$, PS model, outcome model, priors
Output: Posterior distribution of causal estimand

for $m = 1$ *to* M **do**
  PS Stage:
1 Draw the parameters $\boldsymbol{\theta}^{(\boldsymbol{Z})} \sim p(\boldsymbol{\beta}^{(\boldsymbol{Z})}|\mathbf{X}, \mathbf{Z})$
2 **for** $i = 1$ *to* N **do**
    Predict $\widehat{\Phi}_i = \phi(\mathbf{X}_i; \boldsymbol{\theta}^{(\boldsymbol{Z})})$
  end
  Outcome Stage:
3 Draw the parameters of the outcome model $\mathrm{h}^{(Y)}\big(Y_i(z)\big) \sim f^{(Y)}\Big(q^{(Y)}\big(z, \widehat{\Phi}_i; \boldsymbol{\theta}^{(Y)}\big), \nu^{(Y)}\Big)$:
   $\boldsymbol{\beta}^{(\boldsymbol{Y})} \sim p(\boldsymbol{\beta}^{(\boldsymbol{Y})}|\mathbf{Y}, \mathbf{Z}, \mathbf{X}, \widehat{\boldsymbol{\Phi}})$ and $\nu^{(Y)} \sim p(\nu^{(Y)}|\mathbf{Y}, \mathbf{Z}, \mathbf{X}, \widehat{\boldsymbol{\Phi}})$
4 **for** $i = 1$ *to* N **do**
    Impute potential outcomes $Y_i^{mis} = Y_i(1)(1 - Z_i) + Y_i(0)Z_i$ from the posterior predictive distribution $p(Y_i^{mis}|Z_i, \mathbf{X}_i, \widehat{\Phi}_i, \boldsymbol{\beta}^{(\boldsymbol{Y})}, \nu^{(Y)})$
  end
5 Compute the causal estimand $ATE = \frac{1}{N}\sum_{i=1}^{N}(Y_i - Y_i^{mis})(2Z_i - 1)$
end

3. Causal Inference under Interference on Network Data

In this section we describe the problem of interference on network data in observational studies. The presence of interference on networks poses two major challenges: i) spillover effects of a unit's treatment on other units' outcomes, including those through contagion mechanisms, and ii) homophily, the tendency of units with similar characteristics to form ties, which creates a structure of dependence among interacting units in both pre-treatment characteristics and the outcome.

Also, within observational studies, where treatment is not randomized, homophily can generate correlation among the neighbors' treatment due to propensity similarities in taking the treatment given similar covariates, as well as peer influence in the treatment uptake.

We will first define the problem of interference on networks and review common assumptions of interference on neighborhood. Then, we will define causal estimands of interest, review identifying assumptions, and finally propose a novel Bayesian estimator.

3.1 Network Data

Consider a network $\mathcal{N}$ of N units, indexed by i, with adjacency matrix $\mathbf{A}$, where element $A_{ij} > 0$ represents the presence of a tie between unit i and unit j. Ties are assumed to be fixed and known. Recall that for each unit we measure a vector of covariates $\mathbf{X}_i$, a treatment variable Z_i, and an outcome variable Y_i. Here we focus on binary treatments $Z_i \in \{0, 1\}$. The adjacency matrix $\mathbf{A}$ defines for each unit i the set of units that have a direct tie with i. We refer to this set as neighborhood of unit i, denoted by $\mathcal{N}_i = \{j : A_{ij} > 0\}$, and to the units belonging to this set as neighbors of unit i. In real-world applications, neighbors can be geographical neighbors, or friends, partners or collaborators. Let $N_i = \sum_{j \neq i} I(A_{ij} > 0)$ denote the number of neighbors of unit i, referred to as *degree* of unit i. The complement of $\mathcal{N}_i$ in $\mathcal{N}$, excluding i, is denoted by $\mathcal{N}_{-i} \backslash \mathcal{N}_i$.

For each unit i, the partition $(i, \mathcal{N}_i, \mathcal{N}_{-i} \backslash \mathcal{N}_i)$ defines the following partitions of the treatment and outcome vectors: $(Z_i, \mathbf{Z}_{\mathcal{N}_i}, \mathbf{Z}_{\mathcal{N}_{-i} \backslash \mathcal{N}_i})$ and $(Y_i, \mathbf{Y}_{\mathcal{N}_i}, \mathbf{Y}_{\mathcal{N}_{-i} \backslash \mathcal{N}_i})$. With non-network data, $\mathbf{X}_i$ typically includes individual characteristics or cluster-level characteristics representing contextual factors (e.g., demographic or socio-economic factors) or contextual covariates (e.g., geographical factors or presence of infrastructures). On the contrary, in network data $\mathbf{X}_i$ might also include variables describing the network. In particular, it might contain variables representing of the neighborhood $\mathcal{N}_i$, including the topology but also the distribution of individual-level characteristics, and it can contain network properties at node-level representing the position of unit's neighborhood in the graph (e.g., centrality, betweenness, number of shared neighbors, ...).

3.2 Neighborhood Interference

In principle, the potential outcome for unit i could depend on the entire treatment assignment vector $\mathbf{Z}$, i.e., $Y_i(\mathbf{Z} = \mathbf{z})$. The no interference assumption, or SUTVA, restricts the dependency to only the treatment received by unit i, i.e., $Y_i(Z_i = z)$. On the contrary, under interference the potential outcome for unit i depends on the treatment received by other units. However, if each outcome depends on the whole treatment vector, then for each treatment vector $\mathbf{Z}$ each unit would be observed under the same treatment $\mathbf{Z}$. Therefore, the data would not provide any information on missing potential outcomes under different treatment conditions. It is clear that the definition of causal effects as well as their estimation requires assumptions that restrict the number of potential outcomes for each unit. Depending on the type of interference Y_i can depend on the treatment received by a specific group of units. Hudgens & Halloran (2008) consider the 'partial interference' assumption, that deals with settings where units are organized in separate groups (e.g. villages, households, schools, workplaces,...) , i.e. clusters, and allows units to be affected only by the treatment received by units belonging to the same clusters.

In network data, the adjacency matrix provides information on the interaction between units and thus on the flow of the treatment effect. Oftentimes, we have reasons to assume that the outcome of a unit only depends on the treatment received by the neighbors, that is, by the units that individual is in direct contact with. In such case, the potential outcome could be written as $Y_i(\mathbf{Z}) = Y_i(\mathbf{Z}_{\mathcal{N}_i} = \mathbf{z}_{\mathcal{N}_i})$. This assumption excludes spillover effects of the treatment received by higher order connections. Nevertheless, when the number of neighbors is substantial then under each treatment vector $\mathbf{Z}$ the variability on the treatment vector $\mathbf{Z}_{\mathcal{N}_i}$ across units is still very low, and information on the $2^{N_i} - 1$ missing potential outcomes would be hard to extrapolate. For this reason, we introduce the concept of *exposure mapping*. In general terms, we define an exposure mapping as a function that maps a treatment vector $\mathbf{z}$, the adjacency matrix $\mathbf{A}$ and unit-level characteristics $\mathbf{X}$ to an exposure value denoted by G_i: $G_i = g(\mathbf{z}, \mathbf{A}, \mathbf{X}_i, \mathbf{X}_{-i})$, with $g : \mathcal{Z}^N \times \mathcal{A}^{N^2} \times \mathcal{X}^N \to \mathcal{G}_i$. In network data, a special case is the function $g_{\mathcal{N}}(\mathbf{z}_{\mathcal{N}_i}, \mathbf{A}_i, \mathbf{X}_i, \mathbf{X}_{\mathcal{N}_i})$, which receives as input only the treatment vector in the neighborhood, the unit's row of the adjacency matrix, the unit's covariates, and the covariates of units in the neighborhood. Given this definition, we can formalize the neighborhood interference assumption.

Assumption 4 (Neighborhood Interference) *Given a function* $g_{\mathcal{N}} : \mathcal{Z}^{N_i} \times \mathcal{A}^{N_i} \times \mathcal{X}^{N_i+1} \rightarrow \mathcal{G}_i$, $\forall i \in \mathcal{N}$, $\forall\, \boldsymbol{Z}_{\mathcal{N}_{-i}}, \boldsymbol{Z}'_{\mathcal{N}_{-i}}$ *and* $\forall\, \boldsymbol{Z}, \boldsymbol{Z}' \in \mathcal{Z}^N : g_{\mathcal{N}}(\boldsymbol{Z}_{\mathcal{N}_i}, \mathbf{A}_i, \boldsymbol{X}_i, \boldsymbol{X}_{\mathcal{N}_i}) = g_{\mathcal{N}}(\boldsymbol{Z}'_{\mathcal{N}_i}, \mathbf{A}_i, \boldsymbol{X}_i, \boldsymbol{X}_{\mathcal{N}_i})$, *then* $Y_i(\boldsymbol{Z}) = Y_i(\boldsymbol{Z}')$.

Assumption 4 rules out the dependence of the potential outcomes of unit i from the treatment received by units outside his neighborhood, i.e., $\mathbf{Z}_{\mathcal{N}_{-i} \setminus \mathcal{N}_i}$, but allows Y_i to depend on the treatment received by his neighbors, i.e., $\mathbf{Z}_{\mathcal{N}_i}$. Moreover, this dependence is assumed to be through a specific exposure mapping function $g_{\mathcal{N}}(\cdot)$. This formulation is similar to the 'exposure mapping' introduced by Aronow & Samii (2017) and the one in Van der Laan (2014). In Assumption 4 the function $g_{\mathcal{N}}(\cdot)$ is assumed to be known and well-specified. When analyzing network data, we must use substantive knowledge of the subject matter and judgment about the mechanism of interference to fix the exposure mapping function. We refer to $G_i = g_{\mathcal{N}}(\mathbf{Z}_{\mathcal{N}_i}, \mathbf{A}_i, \mathbf{X}_i, \mathbf{X}_{\mathcal{N}_i})$ as the *neighborhood treatment*, and, by contrast, we refer to Z_i as the *individual treatment*. In general, we can write $G_i = \sum_{j \in \mathcal{N}_i} A_{ij} w(\mathbf{X}_i, \mathbf{X}_j) Z_j$, where A_{ij} is the element of the adjacency matrix, which could be binary or weighted, and $w(\mathbf{X}_i, \mathbf{X}_j)$ is a function of unit-level characteristics of the interacting units. This means that we assume that the extent to which the treatment of unit j affects the outcome of unit i depends on the level on interaction between the two units, encoded in A_{ij}, and on their similarity in terms of characteristics. In the simplest case, G_i can be the number or the proportion of treated neighbors, i.e., $G_i = \sum_{j \in \mathcal{N}_i} Z_j$ or $G_i = \frac{\sum_{j \in \mathcal{N}_i} Z_j}{N_i}$, respectively. The domain of G_i depends on how the function $g_{\mathcal{N}}(\cdot)$ is defined. For example, if we consider the simple number of treated neighbors, then $\mathcal{G}_i = \{0, 1, \ldots, N_i\}$. We denote the overall domain by $\mathcal{G} = \bigcup_i \mathcal{G}_i$.

Under Assumption 4, potential outcomes can be indexed just by the the individual treatment and the neighborhood treatment, i.e., $Y_i(Z_i = z, G_i = g)$, which can be simplified to $Y_i(z, g)$. The potential outcome $Y_i(z, g)$ represents the outcome that unit i would exhibit under individual treatment $Z_i = z$ and if exposed to the value g of a function $g_{\mathcal{N}}(\cdot)$ of the treatment vector of his neighbors, $\mathbf{Z}_{\mathcal{N}_i}$. [1]

A potential outcome $Y_i(z, g)$ is defined only for a subset of nodes where G_i can take on value g. We denote this subset by $V_g = \{i : g \in \mathcal{G}_i\}$. For instance, in the case where G_i is the number of treated neighbors, V_g is the set of nodes with degree $N_i \geq g$, that is, with at least g neighbors. It is worth noting that each unit can belong to different subsets V_g, depending on the cardinality of $\mathcal{G}_i$.

3.3 Causal Estimands: Treatment and Spillover Effects

We define here the causal estimands of interest under the neighborhood interference. We focus on finite-sample causal estimands, that is, estimands that are defined on the network

1. In principle, our framework could accommodate higher-order spillover effects. One way would be to define G_i as a weighted function of the treatment received by both neighbors and non-neighbors up to a certain distance, with weights decreasing with distance. In this way, we could estimate the direct effect of the treatment by switching the individual treatment, controlling for spillover effects and cascading effects from the neighborhoods. In addition, the higher-order spillover effect, defined by changing the value of the exposure to the network treatment status G_i, would measure the average effect of changing the treatment status in the network. The definition of G_i as a weighted summary of the treatment status in the unit's neighborhoods up to a certain distance does not require a change in our estimation procedure.

at hand. The advantage of this type of estimand is that its definition does not require the specification of the sampling mechanism from a larger population, which can be difficult to conceptualize in network settings. We first define the average potential outcome $Y_i(z, g)$ in a set of units V as:

$$\mu(z, g; V) = \frac{1}{|V|} \sum_{i \in V} Y_i(z, g) \qquad z \in \{0, 1\}, g \in \mathcal{G} \tag{7}$$

V is a set of units, possibly defined by covariates, including individual or network characteristics. In order for the potential outcomes to be well-defined, V must be the set (or a subset of the set) of units where g is a possible value, i.e., $V \subseteq V_g$. We can view $\mu(z, g; V)$ as an average dose-response function (ADRF), which measures the heterogeneity of potential outcomes arising from a variation in the bivariate treatment, i.e., the binary individual treatment, and the neighborhood treatment, which is a discrete or continuous variable.

We can now define causal estimands as comparisons between average potential outcomes. We define the average (individual) *treatment effect* at neighborhood level $g \in \mathcal{G}$ by

$$\tau(g; V) = \mu(1, g; V) - \mu(0, g; V) \tag{8}$$

which denotes the average causal effect of the individual treatment when the neighborhood treatment is set to level g. Again here $V \subseteq V_g$. Instead of fixing the neighborhood treatment, we can consider a hypothetical intervention that assigns the neighborhood treatment to unit i based on a probability distribution $\pi^\star(g; \mathbf{X}_{\mathcal{N}_i})$. Thus, we define the average *treatment effect* $\tau(\pi^\star, V)$ given the neighborhood treatment assignment $\pi^\star(g; \mathbf{X}_{\mathcal{N}_i})$ by the average effect of the individual treatment marginalized over the probability distribution of the neighborhood treatment, that is

$$\tau(\pi^\star, V) = \int \Big(\mu(1, g; V) - \mu(0, g; V)\Big) \pi^\star(g; \mathbf{X}_{\mathcal{N}_i}) dg. \tag{9}$$

We now define the causal effects of the neighborhood treatment, often referred to as spillover effects or peer effects. We define the *average spillover effect* of having the neighborhood treatment set to level g versus g', when the unit is under the individual treatment z, by

$$\delta(g, g', z; V) = \mu(z, g; V) - \mu(z, g'; V) \tag{10}$$

Notice that V must be a subset of units belonging to both V_g and $V_{g'}$, i.e., $V \subseteq V_g \cap V_{g'}$. Finally, define the *average spillover effect* of intervention $\pi^\star$ vs π' by

$$\Delta(\pi^\star, \pi', z; V) = \int \mu(z, g; V) \pi^\star(g; \mathbf{X}_{\mathcal{N}_i}) dg - \int \mu(z, g; V) \pi'(g; \mathbf{X}_{\mathcal{N}_i}) dg \tag{11}$$

The hypothetical Intervention $\pi^\star(g; \mathbf{X}_{\mathcal{N}_i})$ vs $\pi'(g; \mathbf{X}_{\mathcal{N}_i})$ can be given by real experiments with assignment mechanism $p(\mathbf{Z} = \mathbf{z})$, which reflects into the probability distribution of the neighborhood treatment, or it can be directly defined as a probability distribution of G_i given covariates (see also Papadogeorgou et al., 2018).

The treatment effects $\tau(g; V)$ in (8) and spillover effects $\delta(g, g', z; V)$ in (10) are average comparisons of potential outcomes under fixed values of the individual and neighborhood treatment. Conversely, in the average treatment and spillover effects in (9) and (11), the individual treatment is kept fixed while the neighborhood treatment is drawn from the probability of hypothetical interventions.

3.4 Identifying Assumption: Unconfoundedness

Because the causal effects of interest depend on the comparison between two quantities $\mu(z, g; V)$ with different values of the individual and neighborhood treatments, identification results can focus on the identification of the ADRF $\mu(z, g; V)$.

Assumption 5 (Unconfoundedness of Individual and Neighborhood Treatment)

$$Y_i(z, g) \perp\!\!\!\perp Z_i, G_i \mid \boldsymbol{X}_i \qquad \forall z \in \{0, 1\}, g \in \mathcal{G}_i, \forall i \in V.$$

This assumption states that the individual and neighborhood treatments are independent of the potential outcomes of unit i, conditional on the vector of covariates $\boldsymbol{X}_i$.

Assumption 5 states that the vector $\mathbf{X}_i$ contains all the potential confounders of the relationship between the individual and the neighborhood treatment and the potential outcomes for each unit i. The plausibility of this assumption depends on how the vector $\mathbf{X}_i$ is defined in relation to the probability distribution of the treatment and to the network structure. Assumption 5 rules out the presence of latent variables (not included in $\mathbf{X}_i$) that affect both the probability of taking the treatment and/or the value of neighborhood treatment and the outcome. Neighborhood covariates, that is, the topology of the neighborhood or individual-level covariates among neighbors, are potential confounders only if they affect the outcome of the unit.

In addition, in principle the assumption does not rule out the presence of homophily, that is, the tendency of individuals who share similar characteristics to form ties. In fact, homophily does not violate the unconfoundedness assumption in the cases where characteristics driving the homophily mechanism i) are included in $\mathbf{X}_i$, ii) even if unobserved they do not affect the outcomes iii) they correspond to the treatment variable, that is people who share the same treatment/exposure variable tend to form ties. The only situation where homophily is a threat to identification is when variables underlying the network formation process are not included in $\mathbf{X}_i$ and affect the outcome.

Forastiere et al. (2020) show that under Assumption 5 the ADRF $\mu(z, g; V)$ is identified from the observed data, and estimation can be conducted by taking the average of the observed outcomes of units with $Z_i = z$ and $G_i = g$ within groups of units defined by covariates $\mathbf{X}_i$.

4. Bayesian Generalized Propensity Score Estimator for Causal Effects under Neighborhood Interference

Building on the generalized propensity score estimator proposed by Forastiere et al. (2020), here we develop a new Bayesian semi-parametric estimator for the ADRF $\mu(z, g; V)$, and in turn for the causal estimands in Section 3.3. The idea is to combine results on the generalized propensity score for multi-valued treatment proposed by Imbens (2000) and Hirano & Imbens (2004) (Section 2.9.1) and extended by Forastiere et al. (2020) to interference settings, with the two-step Bayesian propensity score estimator for a binary treatment without interference (see Section 2.10). The proposed estimator lies within the Bayesian imputation approach to causal inference reviewed in Section 2.6. In addition we will replace the parametric partial mean approach of Hirano & Imbens (2004) with a semi-parametric technique

based on penalized Bayesian multivariate splines. To take into account the dependence in the outcome, we also include random effects in the outcome model, with groups defined by a community detection algorithm. This Bayesian approach allows us to easily quantify the uncertainty due to the assignment of Z and G and to the inherent variability of the (missing) potential outcomes.

4.1 Individual and Neighborhood Propensity Scores

Under the unconfoundedness assumption (Assumption 5) the ADRF $\mu(z, g; V)$ could be estimated by taking the average of the observed outcomes within cells defined by covariates. Nevertheless, the presence of continuous covariates or a large number of covariates poses some challenges in the estimation. Under SUTVA, propensity score-based estimators are common solutions (see Section 2.8). Conversely, under the neighborhood interference assumption, Forastiere et al. (2020) propose a new propensity score-based estimator, based on the adjustment for the so-called individual and neighborhood propensity scores.

The *individual propensity score*, denoted by $\phi(z; \mathbf{X}_i^z)$, is the probability of having the individual treatment at level z conditional on covariates $\mathbf{X}_i^z$, i.e., $P(Z_i = z|\mathbf{X}_i^z = \mathbf{x}^z)$. Similarly, the *neighborhood propensity score*, denoted by $\lambda(g; z; \mathbf{X}_i^g)$, is the probability of having the neighborhood treatment at level g conditional on a specific value z of the individual treatment and on the vector of covariates $\mathbf{X}_i^g$, i.e., $P(G_i = g|Z_i = z, \mathbf{X}_i^g = \mathbf{x}^g)$. $\mathbf{X}_i^z \in \mathcal{X}^z \subset \mathcal{X}$ is the subset of covariates affecting the individual treatment, and $\mathbf{X}_i^g \in \mathcal{X}^g \subset \mathcal{X}$ is the subset of covariates affecting the neighborhood treatment. Typically, $\mathbf{X}_i^z$ should include individual characteristics and $\mathbf{X}_i^g$ is likely to include neighborhood characteristic. Nevertheless, $\mathbf{X}_i^z$ and $\mathbf{X}_i^g$ could also coincide and both include all kind of covariates.

Forastiere et al. (2020) show that the individual and neighborhood propensity scores satisfy the balancing and unconfoundedness properties. In particular, if Assumption 5 holds given $\mathbf{X}_i$, then the unconfoundedness assumption holds conditional on the two propensity scores separately, i.e., $Y_i(z, g) \perp\!\!\!\perp Z_i, G_i|\lambda(g; z; \mathbf{X}_i^g), \phi(z; \mathbf{X}_i^z), \forall z \in \{0, 1\}, g \in \mathcal{G}_i$. This property allows deriving a covariate-adjustment method that separately adjusts for the individual propensity score $\phi(z; X_i^z)$ and for the neighborhood propensity score $\lambda(g; z; X_i^g)$. Because $\phi(z; \mathbf{X}_i^z)$ is the propensity score of a binary treatment, we have that $\phi(0; \mathbf{X}_i^z) = 1 - \phi(1; \mathbf{X}_i^z)$ and thus, for every $z \in \{0, 1\}$, we can condition on the propensity score $\phi(1; \mathbf{X}_i^z)$ corresponding to the active treatment $z = 1$. Forastiere et al. (2020) propose the use of a sub-classification approach on the individual propensity score $\phi(1; x^z)$ and, within subclasses that are approximately homogenous in $\phi(1; x^z)$, a model-based approach for the neighborhood propensity score, similar to the one in Hirano & Imbens (2004). Here we replace the frequentist sub-classification and generalized propensity score-based estimator with a semi-parametric Bayesian approach.

4.2 Propensity Scores and Outcome Models

Below we detail alternative propensity scores and outcome models.

4.2.1 Individual and Neighborhood Propensity Scores Models

We first posit a model for the binary individual treatment Z_i

$$\begin{aligned} Z_i &\sim Ber(\phi(1;\mathbf{X}_i^z)) \\ \mathrm{h}^{(Z)}\big(\phi(1;\mathbf{X}_i^z)\big) &= \boldsymbol{\beta}^{(Z)T}\mathbf{X}_i^z \end{aligned} \tag{12}$$

where $h^{(G)}(\cdot)$ is the logit or probit link function, and a model for the neighborhood treatment G_i

$$\mathrm{h}^{(G)}\big(G_i\big) \sim f^{(G)}\Big(q^{(G)}\big(\mathbf{X}_i;\boldsymbol{\beta}^{(G)}\big),\nu^{(G)}\Big) \tag{13}$$

where again $h^{(G)}(\cdot)$ is a link function, $f^{(G)}$ is a probability density function (pdf), $q^{(G)}(\cdot)$ is a flexible function of the covariates depending on a vector of parameters $\boldsymbol{\beta}^{(G)}$, and $\nu^{(G)}$ is a scale parameter.

4.2.2 Outcome Model with Penalized Splines and Random Effects

We now postulate a model for the potential outcomes given $\phi(1;\mathbf{X}_i^z)$ and $\lambda(g;z;\mathbf{X}_i^g)$:

$$\mathrm{h}^{(Y)}\big(Y_i(z,g)\big) \sim f^{(Y)}\Big(q^{(Y)}\big(z,g,\phi(1;\mathbf{X}_i^z),\lambda(g;z;\mathbf{X}_i^g);\boldsymbol{\beta}^{(Y)}\big),\nu^{(Y)}\Big) \tag{14}$$

where as usual $h^{(Y)}(\cdot)$ is a link function, $f^{(Y)}$ is a probability density function (pdf), and $\nu^{(Y)}$ is a scale parameter. The key feature here is $q^{(Y)}(\cdot)$, which we model semi-parametrically using a set of penalized spline basis functions. Splines yield several advantages that include flexibility as well as interpretability via representations that use a compact set of basis functions and coefficients. In particular, the predictor $q^{(Y)}(z,\mathbf{V}(z,g)_i)$, where $\mathbf{V}(z,g)_i = [g,\phi(1;\mathbf{X}_i^z),\lambda(g;z;\mathbf{X}_i^g)]^T$, can be written in the mixed model representation (Ruppert et al., 2003):

$$\begin{aligned} &q^{(Y)}\big(z,\mathbf{V}_i(z,g)\big) = \boldsymbol{\beta}_V^{(Y)T}\mathbf{V}_i'(z,g) + \boldsymbol{\beta}_{Vz}^{(Y)T}\mathbf{V}_i'(z,g)z + \boldsymbol{b}_U^T\mathbf{U}_i(z,g) + \boldsymbol{b}_{Uz}^T\mathbf{U}_i(z,g)z + u_j \\ &u_j \sim \mathcal{N}(0,\Sigma_u) \qquad \boldsymbol{b}_U \sim \mathcal{N}(0,\sigma_{b_u}^2 I_K) \qquad \boldsymbol{b}_{Uz} \sim \mathcal{N}(0,\sigma_{b_{uz}}^2 I_K) \end{aligned} \tag{15}$$

where $\mathbf{V}_i'(z,g) = [1,g,\phi(1;\mathbf{X}_i^z),\lambda(g;z;\mathbf{X}_i^g),\lambda(g;z;\mathbf{X}_i^g)g]^T$, such that the first two terms of Equation (15) represent the linear predictor with interactions, and $\mathbf{U}_i(z,g)$ contains spline basis functions. In particular, we use multivariate smoothing splines with radial basis functions of the form

$$\mathbf{U}_{ik}(z,g) = C(||\mathbf{V}_i(z,g) - \mathbf{k}_k||)\Omega^{-1/2}; \qquad \Omega = \big[C(||\mathbf{k}_k - \mathbf{k}_{k'}||)\big]_{1\leq k,k',\leq K} \tag{16}$$

where $||\cdot||$ is the euclidean norm and $C(\cdot)$ is a basis function. Here our choice goes to thin plate splines of the form

$$C(||\mathbf{r}||) = \begin{cases} ||\mathbf{r}||^{2m-|\mathbf{r}|} \\ ||\mathbf{r}||^{2m-|\mathbf{r}|} log(\mathbf{r}) \end{cases} \tag{17}$$

where m is an integer satisfying $2m - |\mathbf{r}| > 0$, that controls the order of the spline and its smoothness (Ruppert et al., 2003; Wood, 2003). The default is to use the smallest

integer satisfying that condition. The advantage of radial basis functions in multivariate smoothing is that they are rotationally invariant. The distribution of the coefficients $\boldsymbol{b}_U$ and $\boldsymbol{b}_{Uz}$ is a mixed model representation of penalties. The variances $\sigma^2_{b_u}$ and $\sigma^2_{b_{uz}}$ are indeed the parameters controlling the degree of smoothness. A large value of these parameters , that is, a strong roughness penalty, leads to a smoother fit, while a small value (close to zero) leads to an irregular fit and essentially interpolation of the data. A key component in fitting splines is the choice of the number and the placement of knots (K). We address this issue by using penalized splines; wherein we choose a large enough number of knots that are sufficient to capture the local nonlinear features present in the data and control for overfitting by using a penalization on the basis coefficients. Knots are first placed on data locations. For large data sets we randomly subsample a maximum number of data locations (the default maximum number is 2000). Then a truncated eigen-decomposition is used to achieve a rank reduction (Wood, 2003, 2017).

In Equation 14, the outcome model depends on the individual treatment, the neighborhood treatment and both the individual and the neighborhood propensity scores. An alternative approach is to replace the model-based adjustment for the individual propensity score with a matching approach. The idea is to match units on the individual propensity score $\phi(1; \mathbf{X}_i^z)$ to create a matched sample where covariates $\mathbf{X}_i^z$ are balanced across the treated group ($Z_i = 1$) and the control group ($Z_i = 0$). Adjustment for the neighborhood propensity score is then handled, as previously, by a model-based generalized propensity score method applied to the matched samples. For matching with replacement or variable ratio matching weights need to be incorporated into the analysis (Dehejia & Wahba, 1999; Hill et al., 2004; Stuart, 2010). When matching with replacement, individuals receive a frequency weight that reflects the number of matched sets they belong to. When using variable ratio matching, units receive a weight that is proportional to the actual number of units matched to them. Therefore, we would assume a model for the outcome as in (14), with $V_i = [g, \lambda(g; z; \mathbf{X}_i^g)]$, $\mathbf{V}_i' = [1, g, \lambda(g; z; \mathbf{X}_i^g), \lambda(g; z; \mathbf{X}_i^g)g]^T$ and $\nu^{(Y)}$ scaled by the matching weights.

Finally, the last term in Equation (15), u_j, is the random effect for community j, with $j = 1, \ldots, J$. We include this term to take into account any dependence in the outcome data between a unit and his neighbors. We refer to this dependence as *ego-alter correlation*. It is worth noting that this type of dependence in the outcome data is fundamentally different from the mechanism of interference. Interference refers to the effect on a unit's outcome of the treatment received by their neighbors. On the contrary, the ego-alter correlation does not refer to a causal relationship but instead to a correlation among neighbors' outcomes, which can be due to unmeasured factors that are common among units sharing a link. In principle, each unit belongs to the neighborhoods $\mathcal{N}_k$ of all units $k \in \mathcal{N}_i$ in his own neighborhood. Such overlapping nature of neighborhoods complicates the estimation of the correlation structure. We propose an alternative dependence structure by identifying larger non-overlapping clusters incorporating the neighborhoods of multiple units. By defining random effects u_j at such cluster-level, we assume the presence of latent random variable which is shared by all units belonging to the same cluster. Clusters are defined using a community detection algorithm that identifies groups of nodes that are heavily connected among themselves, but sparsely connected to the rest of the network. This definition of

communities enables taking into account both the ego-alter correlation and a broader cluster correlation structure.

4.2.3 Community Detection

Unfolding the communities in real networks is widely used to determine the structural properties of these networks. Community detection or clustering algorithms aim at finding groups of related nodes that are densely interconnected and have fewer connections with the rest of the network. These groups of nodes are called communities or clusters. As communities are often associated with important structural characteristics of a complex system, community detection is a common first step in the understanding and analysis of networks. The search for communities that optimize a given quantitative performance criterion is typically an NP-hard problem, so in most cases one must rely on approximate algorithms to identify community structure. The problem of how to find communities in networks has been extensively studied and a substantial amount of work has been done on developing clustering algorithms (an overview can be found in (Schaeffer, 2007; Lancichinetti & Fortunato, 2009; Fortunato , 2010)). In the simulation section (Section 5), we will descrive the specific methods used for our simulations.

4.2.4 Priors

Within the Bayesian framework we should posit a prior distribution for the parameter vector

$$\boldsymbol{\theta} = \left[\boldsymbol{\theta^{(Z)}}, \boldsymbol{\theta^{(G)}}, \boldsymbol{\theta^{(Y)}}\right]$$

where $\boldsymbol{\theta^{(Z)}} = \boldsymbol{\beta^{(Z)}}$, $\boldsymbol{\theta^{(G)}} = [\boldsymbol{\beta^{(G)}}, \nu^{(G)}]$, $\boldsymbol{\theta^{(Y)}} = [\boldsymbol{\beta^{(Y)}}, \nu^{(Y)}, \Sigma_u, \sigma^2_{b_u}, \sigma^2_{b_{uz}}]$. In particular, we recommend the use of weakly informative priors to provide moderate regularization and help stabilize computation.

We assume a multivariate normal prior for all regression coefficients:

$$\boldsymbol{\beta^{(Z)}} \sim \mathcal{N}(\boldsymbol{\eta}^Z, K^Z) \qquad \boldsymbol{\beta^{(G)}} \sim \mathcal{N}(\boldsymbol{\eta}^G, K^G) \qquad \boldsymbol{\beta^{(Y)}} \sim \mathcal{N}(\boldsymbol{\eta}^Y, K^Y) \tag{18}$$

The priors on the scale parameters $\nu^{(G)}$ and $\nu^{(Y)}$ should depend on the neighborhood treatment distribution $f^{(G)}$ and on the outcome distribution $f^{(Y)}$, respectively. However, a general prior distribution could be

$$\nu^{(G)} \sim \text{Exp}(\gamma_{\nu^G}) \qquad \nu^{(Y)} \sim \text{Exp}(\gamma_{\nu^Y}) \tag{19}$$

The random effect covariance matrix Σ_u is decomposed into a diagonal matrix of standard deviations and the correlation matrix (McElreath, 2016):

$$\Sigma_u = \text{diag}(\boldsymbol{\sigma}_u)\Omega_u\text{diag}(\boldsymbol{\sigma}_u)$$

For the correlation matrix we use a prior distribution called LKJ, whose density is proportional to the determinant of the correlation matrix raised to the power of a positive regularization parameter minus one:

$$\Omega_u \sim LKJ(\zeta) \qquad p(\Omega_u) \propto \det(\Omega_u)^{\zeta-1}, \zeta > 0$$

so $\zeta = 1$ leads to a uniform distribution on correlation matrices, while the magnitude of correlations between components decreases as $\zeta \to \infty$. The standard deviations $\boldsymbol{\sigma}_u$ are in turn decomposed into the product of a simplex vector $\boldsymbol{\pi}_u$ and the trace of the covariance matrix:

$$\text{diag}(\boldsymbol{\sigma}_u) = \text{tr}(\Sigma_u)\boldsymbol{\pi}_u = Je^2\boldsymbol{\pi}_u$$

where the trace (total variance) is the product of the order of the matrix and the square of a scale parameter and the element π_j of the simplex vector is the proportion of the total variance attributable to the corresponding random effect u_j. For the scale parameter e we posit a Gamma prior, with shape and scale parameters both set to 1. For the simplex vector $\boldsymbol{\pi}_u$ we use a symmetric Dirichlet prior, which has a single concentration parameter $\chi > 0$.

For the smoothing parameters we posit the following prior distribution:

$$\sigma^2_{b_u} \sim \text{Exp}(\gamma_{b_u}) \qquad \sigma^2_{b_{uz}} \sim \text{Exp}(\gamma_{b_{uz}}) \tag{20}$$

4.3 Three-Step Estimation Procedure

Here we propose a three-step Bayesian estimator that extends the 'two-step Bayesian' estimator proposed by Hoshino (2008), McCandless et al. (2010) and Kaplan & Chen (2012) to the neighborhood interference setting, with the individual and the neighborhood propensity score.

The three steps refer to the posterior distributions of the parameters of the individual propensity score, the neighborhood propensity score and the outcome models. Since the outcome model involves the individual and neighborhood propensity scores, in principle the outcome model indirectly depends on all parameters, including the set of parameters of the two propensity score models, i.e., $\boldsymbol{\theta}^{(Z)}$, and $\boldsymbol{\theta}^{(G)}$. Therefore, the posterior distribution of these parameters should in part be informed by the outcome stage. To restrict the flow of information between models during MCMC computation and, hence, to avoid 'model feedback', we take a three-step approach which approximates the joint posterior distribution by drawing the parameters of the propensity score models from the approximate conditional distributions

$$p(\boldsymbol{\theta}^{(Z)}|\mathbf{Y}, \mathbf{G}, \mathbf{Z}, \mathbf{X}, \boldsymbol{\theta}^{(Y)}, \boldsymbol{\theta}^{(G)}) \propto \prod_i^N p(Z_i|\mathbf{X}_i, \boldsymbol{\theta}^{(Z)})p(\boldsymbol{\theta}^{(Z)})$$

and

$$p(\boldsymbol{\theta}^{(G)}|\mathbf{Y}, \mathbf{G}, \mathbf{Z}, \mathbf{X}, \boldsymbol{\theta}^{(Y)}, \boldsymbol{\theta}^{(Z)}) \propto \prod_i^N p(G_i|Z_i, \mathbf{X}_i, \boldsymbol{\theta}^{(G)})p(\boldsymbol{\theta}^{(G)})$$

which ignore the likelihood contribution from the outcome and, hence, depend neither on $\boldsymbol{\theta}^{(Y)}$ nor on the parameters of the neighborhood or individual propensity score model, respectively. The posterior distributions of the individual and the neighborhood propensity score models are then used as an input when deriving the posterior distribution of the parameters of the outcome model:

$$p(\boldsymbol{\theta}^{(Y)}|\mathbf{Y}, \mathbf{G}, \mathbf{Z}, \mathbf{X}, \boldsymbol{\theta}^{(Z)}, \boldsymbol{\theta}^{(G)}) \propto \prod_i^N p(Y_i, G_i, Z_i, \mathbf{X}_i|\boldsymbol{\theta}^{(Z)}, \boldsymbol{\theta}^{(G)}, \boldsymbol{\theta}^{(Y)})p(\boldsymbol{\theta}^{(Y)})$$

After the posterior distribution of all the parameters $\boldsymbol{\theta}$ is drawn, the posterior distribution of our finite-sample average dose-response function ADRF $\mu(z, g; V)$ is obtained by drawing from the posterior predictive distribution the potential outcomes $Y_i(z, g)$, for each value of z and g and for each unit $i \in V$. Then for each draw the ADRF $\mu(z, g; V)$ is computed by taking the average of the imputed potential outcomes $Y_i(z, g)$ over all units of the set V. Causal estimands are simply computed from the ADRF as comparisons of average potential outcomes at different levels.

We describe below all the steps of the algorithm.

Algorithm 4: Bayesian Three-Step Generalized Propensity Score

Input: Data set $\mathcal{D}$, Adjacency Matrix $\mathbf{A}$, Z model, G model, Y model, priors, Matching=FALSE
Output: Posterior distribution of ADRF $\mu(z, g)$

Community Detection Stage:

Run a *Community Detection algorithm* on $\mathbf{A} \Longrightarrow$ community indicators $C_i \in \{1, \ldots, J\}, \forall i \in \mathcal{N}$

Initialize parameters $\boldsymbol{\theta}^{(0)} = [\boldsymbol{\theta}^{(Z)(0)}, \boldsymbol{\theta}^{(G)(0)}, \boldsymbol{\theta}^{(Z)(0)}]$

for $m = 1$ *to* M **do**

PS Stage:

Define $\mathbf{X}_i^z \in \mathbf{X}_i$ */* including individual, network and neighborhood characteristics */*

Draw the parameters $\boldsymbol{\beta}^{(\boldsymbol{Z})(\boldsymbol{m})} \sim p(\boldsymbol{\beta}^{(\boldsymbol{Z})}|\mathbf{X}^z, \mathbf{Z})$

for $i = 1$ *to* N **do**

Predict $\widehat{\Phi}_i^{(m)} = \phi(1; \mathbf{X}_i^z; \boldsymbol{\beta}^{(\boldsymbol{Z})(\boldsymbol{m})})$

end

if *Matching=TRUE* **then**

Run a *Matching algorithm* with distance metric=$\widehat{\lambda}_i \Longrightarrow$ matched sets $\mathcal{S}_k$, $k = 1, \ldots, K$

Given $S_{ik} = I(i \in \mathcal{S}_k)$, define weights as $w_i = \frac{1}{\sum_{k=1}^K S_{ik}}$, set $\mathcal{M} = \{i : \sum_{k=1}^K S_{ik} > 0\}$, and define the matched sample as $\mathcal{D}^\star = \{(\mathbf{X}_i)_{i\in\mathcal{M}}, (Z_i)_{i\in\mathcal{M}}, (Y_i)_{i\in\mathcal{M}}\}$

end

else

Set $w_i = 1, \forall i \in \mathcal{N}$; $\mathcal{M} = \mathcal{N}$; $\mathcal{D}^\star = \mathcal{D}$

end

GPS Stage:

Compute $G_i = g_{\mathcal{N}}(\mathbf{Z}_{\mathcal{N}_i}, \mathbf{A}_i, \mathbf{X}_i, \mathbf{X}_{\mathcal{N}_i}), \forall i \in \mathcal{N}$

Define $\mathbf{X}_i^g \in \mathbf{X}_i$ */* including individual, network and neighborhood characteristics */*

Draw the parameters $\boldsymbol{\beta}^{(\boldsymbol{G})(\boldsymbol{m})} \sim p(\boldsymbol{\beta}^{(\boldsymbol{G})}|\mathbf{X}^g, \mathbf{Z}, \mathbf{G})$ and $\nu^{(G)(m)} \sim p(\nu^{(G)}|\mathbf{X}^g, \mathbf{Z}, \mathbf{G})$

for $i = 1$ *to* N **do**

Predict $\widehat{\Lambda}_i^{(m)} = \lambda(G_i; Z_i; \mathbf{X}_i^g; \boldsymbol{\beta}^{(\boldsymbol{G})(\boldsymbol{m})}, \nu^{(G)(m)})$

end

Outcome Stage:

if *Matching=TRUE* **then**

Define $\mathbf{V}^{(m)}(Z_i, G_i)_i = [G_i, \widehat{\Lambda}_i^{(m)}]^T$

end

else

Define $\mathbf{V}_i^{(m)}(Z_i, G_i) = [G_i, \widehat{\Phi}_i^{(m)}, \widehat{\Lambda}_i^{(m)}]^T$

end

Compute spline basis functions $\mathbf{U}_i^{(m)}(Z_i, G_i)$ as in (16) and (17)

Draw the parameters $\boldsymbol{\theta}^{(Y)(m)}$ of the outcome model in (14) and (15) using the Gibbs sampler algorithm 5

for $z = 0, 1$ **do**

for $g \in \mathcal{G}$ **do**

for $i = 1$ *to* N **do**

Impute potential outcomes $\widehat{Y}_i(z, g)$:

a. Predict the neighborhood GPS $\lambda(g; z; \mathbf{X}_i^g; \boldsymbol{\beta}^{(\boldsymbol{G})(\boldsymbol{m})}, \nu^{(G)(m)})$

b. Define $\mathbf{V}_i^{(m)}(z, g)$ and compute $\mathbf{U}_i^{(m)}(z, g)$

c. Predict $\widehat{Y}_i(z, g)$, given z,g, $\mathbf{V}_i^{(m)}(z, g)$, $\mathbf{U}_i^{(m)}(z, g)$, the random effects $\mathbf{u}^{(m)}$, and the parameters $\boldsymbol{\beta}^{(\boldsymbol{Y})(\boldsymbol{m})}$, $\boldsymbol{b}_U^{(m)}$, $\boldsymbol{b}_{Uz}^{(m)}$ and $\nu^{(Y)(m)}$

end

Average the potential outcomes over all units: $\widehat{\mu}(z, g) = \frac{1}{N}\sum_{i=1}^N \widehat{Y}_i(z, g)$

end

end

end

Algorithm 5: Gibbs Sampler for Parameters of the Outcome Model

Input: Data set $\mathcal{D}$, Adjacency Matrix $\mathbf{A}$, Y model, priors
Output: Posterior distribution of $\boldsymbol{\theta}^{(Y)(m)}$

1 Draw the parameters $\boldsymbol{\theta}^{(Y)(m)}$ of the outcome model in (14) and (15):

a. Draw the random effects $u_j^{(m)} \sim p(u_j|\Sigma_u^{m-1})$

b. Draw the coefficients $\boldsymbol{b}_U^{(m)} \sim p(\boldsymbol{b}_U|\sigma_{b_u}^{2(m-1)})$ and $\boldsymbol{b}_{Uz}^{(m)} \sim p(\boldsymbol{b}_{Uz}|\sigma_{b_{uz}}^{2(m-1)})$

c. Draw
$\Sigma_u^{(m)} \sim p(\Sigma_u|\mathbf{Y}, \mathbf{Z}, \mathbf{V}^{(m)}(\mathbf{Z}, \mathbf{G}), \mathbf{U}^{(m)}(\mathbf{Z}, \mathbf{G}), \mathbf{u}^{(m)}, \boldsymbol{\beta}^{(\boldsymbol{Y})(\boldsymbol{m-1})}, \boldsymbol{\nu}^{(\boldsymbol{Y})(\boldsymbol{m-1})}, \boldsymbol{b}_U^{(m-1)}, \boldsymbol{b}_{Uz}^{(m-1)})$

d. Draw the smoothing parameters:
$\sigma_{b_u}^{2(m)} \sim p(\sigma_{b_u}^{2(m)}|\mathbf{Y}, \mathbf{Z}, \mathbf{V}^{(m)}(\mathbf{Z}, \mathbf{G}), \mathbf{U}^{(m)}(\mathbf{Z}, \mathbf{G}), \mathbf{u}^{(m)}, \boldsymbol{b}_U^{(m)}, \boldsymbol{b}_{Uz}^{(m)}, \boldsymbol{\beta}^{(\boldsymbol{Y})(\boldsymbol{m-1})}, \boldsymbol{\nu}^{(\boldsymbol{Y})(\boldsymbol{m-1})})$
and $\sigma_{b_{uz}}^{2(m)} \sim p(\sigma_{b_{uz}}^{2(m)}|\mathbf{Y}, \mathbf{Z}, \mathbf{V}^{(m)}(\mathbf{Z}, \mathbf{G}), \mathbf{U}^{(m)}(\mathbf{Z}, \mathbf{G}), \mathbf{u}^{(m)}, \boldsymbol{b}_U^{(m)}, \boldsymbol{b}_{Uz}^{(m)}, \boldsymbol{\beta}^{(\boldsymbol{Y})(\boldsymbol{m-1})}, \boldsymbol{\nu}^{(\boldsymbol{Y})(\boldsymbol{m-1})})$

e. Draw $\boldsymbol{\beta}^{(\boldsymbol{Y})(\boldsymbol{m})} \sim p(\boldsymbol{\beta}^{(\boldsymbol{Y})}|\mathbf{Y}, \mathbf{Z}, \mathbf{V}^{(m)}(\mathbf{Z}, \mathbf{G}), \mathbf{U}^{(m)}(\mathbf{Z}, \mathbf{G}), \mathbf{u}^{(m)}, \boldsymbol{b}_U^{(m)}, \boldsymbol{b}_{Uz}^{(m)}, \boldsymbol{\nu}^{(\boldsymbol{Y})(\boldsymbol{m-1})})$

f. Draw $\nu^{(Y)(m)} \sim p(\nu^{(Y)}|\mathbf{Y}, \mathbf{Z}, \mathbf{V}^{(m)}(\mathbf{Z}, \mathbf{G}), \mathbf{U}^{(m)}(\mathbf{Z}, \mathbf{G}), \mathbf{u}^{(m)}, \boldsymbol{b}_U^{(m)}, \boldsymbol{b}_{Uz}^{(m)}, \boldsymbol{\beta}^{(\boldsymbol{Y})(\boldsymbol{m})})$

4.4 Posterior Predictive Checks

The unbiasedness of our proposed estimator relies on the structural Assumption 4 and 5 and on the correctness of the posited model for the individual and neighborhood propensity scores (Equations (12) and (13)) and for the outcome (Equation (14)). It is worth noting that Assumptions 4 can be seen as both a structural and a modeling assumption because, in addition to ruling out the presence of interference from units outside the neighborhood of unit i, it also assumes a specific function $g_{\mathcal{N}}(\cdot)$ through which interference takes place. As a method for model checking we propose the use of posterior predictive checks, where the posterior predictive distribution is compared to the observed outcome distribution (Gelman et al., 1996). This approach simply relies on the intuitive idea that if a model is a good fit to the data, then replicated data predicted from that model should look similar to the observed data. To simplify the comparison, a summary discrepancy measure $T(\cdot)$ is generally used.

In practice, a posterior predictive check can be incorporated in our estimator by sampling from the the posterior predictive distribution of the potential outcomes corresponding to the observed outcomes:

$$p(Y(Z_i, G_i)|\mathbf{Y}, \mathbf{G}, \mathbf{Z}, \mathbf{X}) = \int p(Y(Z_i, G_i)|\mathbf{Y}, \mathbf{G}, \mathbf{Z}, \mathbf{X}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{G}, \mathbf{Z}, \mathbf{X}) d\theta \tag{21}$$

After the models have been specified (Equations (12), (13) and (14)), together with the prior distribution $p(\boldsymbol{\theta})$, and the posterior distribution of the parameters has been sampled, we draw from the posterior predictive distribution many replicate data sets, that is, for each unit we draw outcomes corresponding to the observed individual and neighborhood treatments, i.e., $Y_i^{rep} = Y_i(Z_i, G_i)$. We then calculate a test statistic $T(\mathbf{Y}^{rep})$ for each replicate data set and we compare the distribution of $T(\mathbf{Y}^{rep})$ with the test statistic applied to the observed data $T(\mathbf{Y}^{obs})$. Parameter uncertainty is explicitly accounted for because the data realizations are generated from parameter values randomly drawn from the posterior distribution. Posterior predictive P-values are defined as posterior probability that the test statistic applied to replicated data, generated under the posited model with parameter sampled from the posterior distribution, exceeds the value of the test statistic applied to the observed data:

$$\text{p-value} = Pr(T(\mathbf{Y}^{rep}) \geq T(\mathbf{Y}^{obs})|\mathbf{Y}, \mathbf{G}, \mathbf{Z}, \mathbf{X}) \tag{22}$$

Posterior predictive P-values can easily be calculated as the proportion of data replications under the specified model resulting in a value of the test statistics that exceeds the observed value.

5. Simulation Study

To assess the performance of our proposed estimator we conducted a simulation study.

5.1 Data Generating Process and Estimators

The study consists of 12 (3 x 2 x 2) scenarios, characterized by a different data generating process given by the network type, the outcome functional form and the outcome correlation structure. Here we describe the general structure of the simulation scenarios. Details of the data generating models can be found in Appendix A.

With regard to the network type we generated the network from i) a stochastic block model, ii) a latent cluster position model, iii) the Add Health data set. Each network consists of 1000 nodes. For the first two cases, where the network was generated using a formation model, before generating the adjacency matrix two covariates —one continuous and one binary— were drawn for each unit. The two formation models differ by presence of homophily, that is the probability of forming links depending on covariates. In the stochastic block model (Holland et al., 1983) the probability of link between two units only depends on their community membership, whereas in the latent cluster position network (Handcock & Raftery, 2007) it also depends on the similarity in their characteristics. Finally, in the third case, we extracted 1000 students from the Add Health data set (Harris & Udry, 2017) and we used race, grade and sex as covariates.

Regarding the outcome functional form, for each network type we simulated the outcome from a normal distribution with the predictor $q^{(Y)}(\cdot)$ being a) a linear function of the individual treatment, the neighborhood treatment, the individual propensity score and neighborhood propensity score, b) a non-linear function of the two treatments and the two propensity scores, resulting in a sigmoid function of the neighborhood treatment.

Finally, for each scenario defined by the network type and the outcome functional form we generated the outcome with and without correlation between units belonging to the same community, that is with and without the inclusion of a community random intercept.

For each scenario, we ran 500 simulations where, given the sample of 1000 nodes and its adjacency matrix, for each unit i: 1) we generated the individual treatment Z_i from an individual propensity score model (see Appendix A); 2) we computed the neighborhood treatment G_i as the proportion of treated neighbors; 3) we generated the outcome Y_i from a model defined by the specific scenario (linear or non-linear and with or without a community intercept). The posterior variance of the estimator should capture the variability of the treatment and outcome stochastic models conditional on a given network. Therefore, in order for the variance across simulations to reproduce the posterior variance of the estimator in principle the treatment and the outcome should be repeatedly drawn on the same network of units. When the network was sampled from a formation model, we generated 5 different networks and we ran 100 simulations for each network, in order to assess the performance of the estimator beyond a specific sampled network.

In all 12 scenarios the average dose-response functions, and in turn the causal treatment and spillover effects, were estimated using the proposed estimator with different options. The first variant of the estimator was the use of a linear outcome model, i.e., $q^{(Y)}(\cdot)$ including only the first linear part, or the use of penalized splines, i.e., $q^{(Y)}(\cdot)$ including both the linear part and the non-linear part with spline basis functions. [2] Furthermore, we investigated the performance of the estimator with and without the inclusion of a community random intercept. Since the scope of the simulation study was to assess under different scenarios the frequentist performances of our proposed estimator with different modeling assumptions, in the case where community correlation structure was taken into account we defined random intercepts on the known communities as provided by the data generating process. In this way we could disentangle the performance of our Bayesian estimator based on the individual and neighborhood propensity score from the performance of community detection algorithms.

5.2 Results

We investigated the frequentist performance of the Bayesian generalized propensity score estimator in estimating the average dose-response function and the causal treatment and spillover effects. Given the posterior distribution of the ADRF and the causal effects for each replication, we then computed across the 500 replication the average bias and mean square error of the posterior mean and the coverage of the 95% credible intervals. In the tables below we report the bias, RMSE and coverage for the treatment and spillover effects. The performance of the estimator with respect to the ADRF can be seen in the figures. The figures represent posterior distribution of ADRF averaged across 500 simulations. The posterior mean and 95% credible intervals are plotted against the true average dose-response function.

In Table 1 we report the performance of our estimator for treatment effects $\tau(g, V_g)$, with $g \in \{0, 0.1, \ldots, 0.9, 1\}$, and spillover effects $\delta(g, g - 0.1, z, V_g \cup V_{g-0.1})$, with $g \in \{0.1, \ldots, 0.9, 1\}$ and $z \in \{0, 1\}$, under the scenario where the network is generated from a stochastic block model and the outcome is drawn from a linear model with random effects. The estimator that makes use of a linear model, reflecting the data generating model, performs well in terms of bias, mean square error and coverage (close to the nominal level) for all treatment and spillover effects. For treatment effects the bias and mean square error are constant across different levels of the neighborhood treatment g. On the contrary the mean square error and coverage of the estimator w.r.t spillover effects vary with g and get respectively higher and smaller as g departs from its median value. In these regions the overall performance of the estimator show severe deterioration when the linear model is replaced with penalized splines. In fact we can see in Figure 1 that splines result in large confidence intervals for the average dose-response function for values of the neighborhood treatment that are not found in less than 2% of the sample, that is, below the 1st percentile

2. In a simulation study, not reported here, we also assessed the use of a matching approach on the individual propensity score. The performance of this hybrid estimator, which uses matching on the individual propensity score and a model-based approach to adjust for the neighborhood propensity score, is similar to the fully model-based estimator, with a linear or non-linear model. However, we noticed a decrease in coverage which could be due to the fact that for each draw of the parameters of the individual treatment model we have a slightly different matching sample, which results in a slightly different condition set for our causal estimands.

and above the 99th percentile. This deterioration in areas where data points decrease is a typical drawback of the use of smoothing splines. Nevertheless, the splines-based estimator w.r.t. treatment effects still performs well in the entire range of g, with the bias and mean square error staying constant. When correlation is not taken into account, that is, the estimator does not include community random intercepts, the root MSE of both linear and splines-based estimators is doubled, with a slight increase in the bias and a higher increase in the variance. The coverage remains close to the nominal level, but we do see a small decrease for the linear estimator. This confirms the robustness of the Bayesian estimation approach, which generally has very good frequentist properties. Specifically, because the Bayesian approach does not rely on large sample approximations and specific estimation of the variance, it tends to be robust to (small) model misspecifications. In Table 2, where the network is again generated from a stochastic block model and the outcome is drawn from a linear model without random effects, we can see that the mean square error is insensitive to the inclusion of random intercepts in the estimator and is in general slightly lower but very close to the one that results from generating correlated outcomes and estimating the causal effects taking the correlation into account. Again we see that for the linear estimator the coverage is slightly higher when we try to estimate a correlation that is not present. This is not the case for the splines-based estimator. This means that in general including a random intercept when it is not needed is not harmful.

Figure 2 reports an example of posterior predictive checks, where draws from the posterior predictive distribution are plotted against the observed outcome distribution. We could also obtain the posterior predictive distribution of a test statistic (e.g. mean, median, standard deviation, quantiles) and compare it with its observed value. The figure shows that, as expected, the linear estimator gives rise to a posterior predictive distribution which does not match the observed one when the data generating model is non-linear. Moreover, the posterior predictive distribution resulting from a splines-based estimator has more variability than the one resulting from the linear estimator.

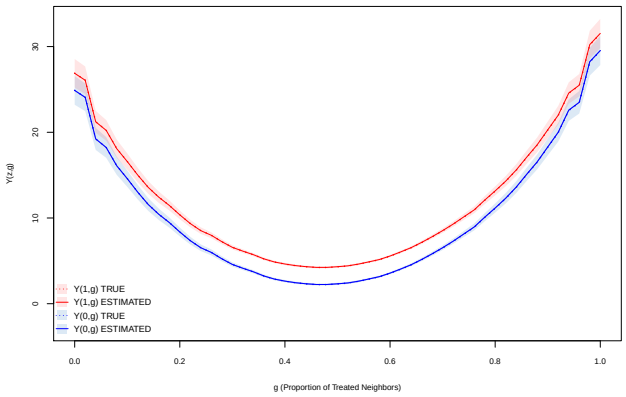
Tables 3 and 4 report the results under the scenarios where the network is still generated from a stochastic block model but the outcome is drawn from a non-linear model, respectively with and without random effects. In this case the use of a linear model fails to recover the average dose-response function. This results in low performances of the linear estimator w.r.t. the spillover effects. Nevertheless, the linear estimator is still able to yield unbiased estimates of the treatment effects with coverage close to the nominal rate. On the contrary, the splines-based estimator applied to the non-linear case is able to recover the average dose-response function and to unbiasedly estimate both treatment and spillover effects. In the critical regions of low and high values of the neighborhood treatment, the performance of the splines-based estimator seems to improve when such estimator is applied to non-linear data rather than linear data. In fact, here the bias and mean square error of this estimator w.r.t spillover effects at low or high values of g is reduced by a factor of 8-10 compared to the linear case.

Figure 3 represents a network graph generated from a stochastic block model with 100 communities of 10 units each. The nodes are colored according to community membership, with filled and un-filled circles corresponding to treatment status. The size of each node represents the neighborhood treatment.

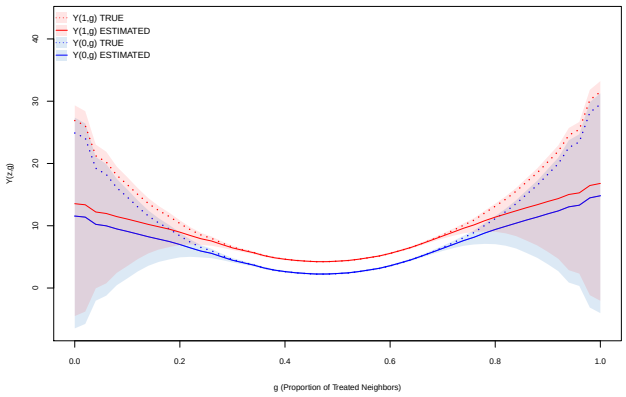
	Linear						Splines					
	RE			NO RE			RE			NO RE		
	Bias	RMSE	Coverage	Bias	RMSE	Coverage	Bias	RMSE	Coverage	Bias	RMSE	Coverage
$\tau(0, V_0)$	-0.004	0.083	0.960	0.001	0.173	0.948	-0.017	0.085	0.960	-0.028	0.177	0.964
$\tau(0.1, V_{0.1})$	-0.004	0.083	0.962	0.001	0.173	0.942	-0.017	0.085	0.958	-0.028	0.177	0.960
$\tau(0.2, V_{0.2})$	-0.004	0.083	0.966	0.001	0.173	0.948	-0.017	0.085	0.966	-0.028	0.177	0.962
$\tau(0.3, V_{0.3})$	-0.004	0.083	0.964	0.001	0.173	0.942	-0.017	0.085	0.958	-0.028	0.177	0.966
$\tau(0.4, V_{0.4})$	-0.004	0.083	0.968	0.001	0.173	0.942	-0.017	0.085	0.958	-0.028	0.177	0.962
$\tau(0.5, V_{0.5})$	-0.004	0.083	0.962	0.001	0.173	0.942	-0.017	0.085	0.956	-0.028	0.177	0.962
$\tau(0.6, V_{0.6})$	-0.004	0.083	0.962	0.001	0.173	0.944	-0.017	0.085	0.960	-0.028	0.177	0.964
$\tau(0.7, V_{0.7})$	-0.004	0.083	0.964	0.001	0.173	0.946	-0.017	0.085	0.954	-0.028	0.177	0.960
$\tau(0.8, V_{0.8})$	-0.004	0.083	0.962	0.001	0.173	0.942	-0.017	0.085	0.960	-0.028	0.177	0.966
$\tau(0.9, V_{0.9})$	-0.004	0.083	0.960	0.001	0.173	0.946	-0.017	0.085	0.958	-0.028	0.177	0.970
$\tau(1, V_1)$	-0.004	0.083	0.964	0.001	0.173	0.944	-0.017	0.085	0.960	-0.028	0.177	0.964
$\delta(0.1, 0, 0)$	0.006	0.379	0.944	0.019	0.801	0.944	7.826	8.778	0.806	8.363	9.173	0.898
$\delta(0.2, 0.1, 0)$	0.006	0.237	0.942	0.014	0.501	0.936	4.125	4.696	0.818	4.467	4.983	0.902
$\delta(0.3, 0.2, 0)$	0.005	0.153	0.942	0.012	0.322	0.936	1.272	1.575	0.840	1.440	1.789	0.904
$\delta(0.4, 0.3, 0)$	0.005	0.090	0.960	0.010	0.191	0.928	1.124	0.209	0.938	0.147	0.318	0.966
$\delta(0.5, 0.4, 0)$	0.005	0.042	0.998	0.008	0.089	0.934	0.013	0.054	1.000	0.019	0.111	0.998
$\delta(0.6, 0.5, 0)$	0.005	0.042	0.992	0.007	0.090	0.940	-0.005	0.056	0.994	0.006	0.110	0.994
$\delta(0.7, 0.6, 0)$	0.005	0.093	0.972	0.005	0.198	0.944	0.165	0.284	0.920	-0.240	0.418	0.954
$\delta(0.8, 0.7, 0)$	0.005	0.150	0.962	0.003	0.318	0.954	-1.597	1.900	0.856	-1.848	2.187	0.864
$\delta(0.9, 0.8, 0)$	0.005	0.236	0.952	0.001	0.500	0.956	-4.600	5.139	0.846	-4.993.	5.484	0.858
$\delta(1, 0.9, 0)$	0.005	0.380	0.950	-0.004	0.805	0.952	-8.350	9.274	0.856	-8.940	9.724	0.868
$\delta(0.1, 0, 1)$	0.006	0.379	0.944	0.019	0.801	0.944	7.826	8.778	0.806	8.363	9.173	0.898
$\delta(0.2, 0.1, 1)$	0.006	0.237	0.948	0.014	0.501	0.936	4.125	4.696	0.818	4.467	4.983	0.902
$\delta(0.3, 0.2, 1)$	0.005	0.153	0.946	0.012	0.322	0.936	1.272	1.575	0.840	1.440	1.789	0.904
$\delta(0.4, 0.3, 1)$	0.005	0.090	0.958	0.010	0.191	0.928	1.124	0.209	0.938	0.147	0.318	0.966
$\delta(0.5, 0.4, 1)$	0.005	0.042	0.994	0.008	0.089	0.934	0.013	0.054	1.000	0.019	0.111	0.998
$\delta(0.6, 0.5, 1)$	0.005	0.042	0.996	0.007	0.090	0.940	-0.005	0.056	0.994	0.006	0.110	0.994
$\delta(0.7, 0.6, 1)$	0.005	0.093	0.966	0.005	0.198	0.944	0.165	0.284	0.920	-0.240	0.418	0.954
$\delta(0.8, 0.7, 1)$	0.005	0.150	0.950	0.003	0.318	0.954	-1.597	1.900	0.856	-1.848	2.187	0.864
$\delta(0.9, 0.8, 1)$	0.005	0.236	0.946	0.001	0.500	0.956	-4.600	5.139	0.846	-4.993.	5.484	0.858
$\delta(1, 0.9, 1)$	0.005	0.380	0.946	-0.004	0.805	0.952	-8.350	9.274	0.856	-8.940	9.724	0.868

Table 1: Performance of Bayesian GPS estimator for treatment effects $\tau(g, V_g)$, with $g \in \{0, 0.1, \ldots, 0.9, 1\}$, and spillover effects $\delta(g, g-0.1, z, V_g \cup V_{g-0.1})$, with $g \in \{0.1, \ldots, 0.9, 1\}$ and $z \in \{0, 1\}$, under the scenario where the network is generated from a stochastic block model and the outcome is drawn from a linear model with random effects.

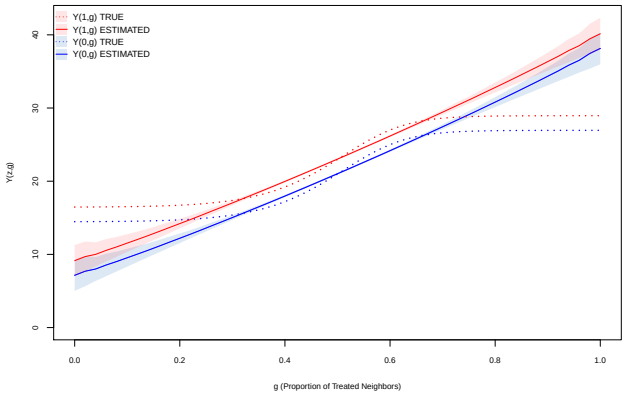
Tables 5, 6, 7 and 8 report the results of the scenarios where the network is generated from a latent cluster model, which is characterized by the presence of homophily along the observed covariates. We essentially see the same patterns shown in Tables 1, 2, 3 and 4. This means that the adjustment for the individual and neighborhood propensity score is able to disentangle the spillover effects from the mechanism of homophily when this is driven



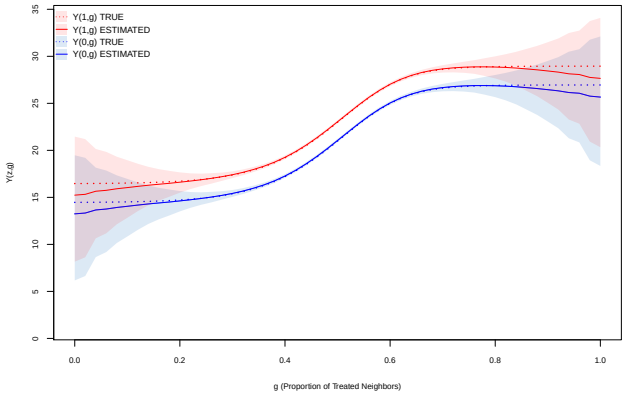
(a) Linear DGP - Linear Model



(b) Linear DGP - Splines



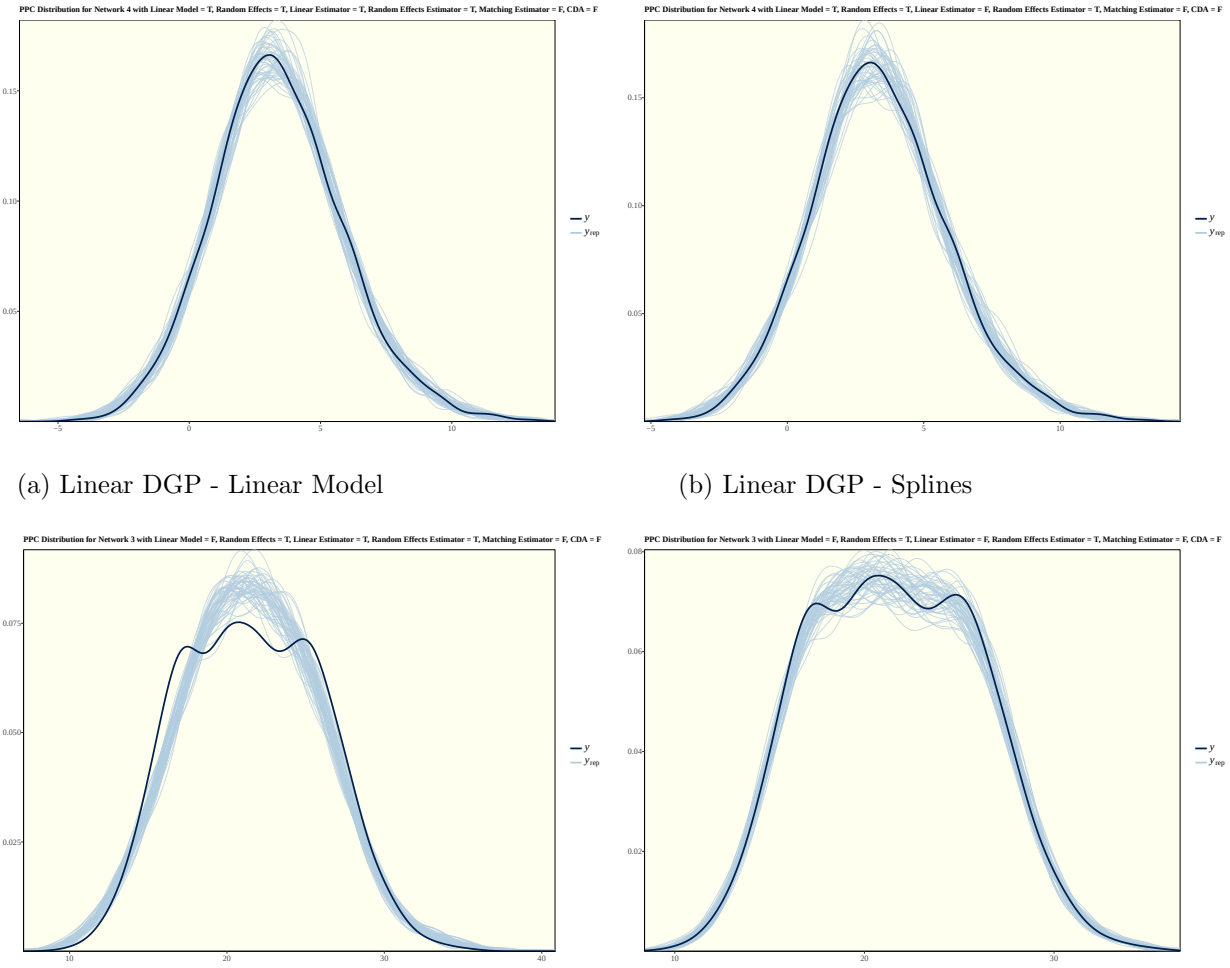
(c) Non-Linear DGP - Linear Model



(d) Non-Linear DGP - Splines

Figure 1: Estimated and true ()ADRF $\mu(z, g; V_g)$ under the scenario where the network is generated from a stochastic block model and the outcome is drawn from a linear model (top) or a non-linear model (bottom) with random effects. Average posterior mean (solid line) and 95 % credible intervals of $\mu(0, g; V_g)$ (blue) and $\mu(1, g; V_g)$ (red) of linear (left) and splines-based (right) estimator with random effects are represented with the true ADRFs (dotted line).

by observed characteristics. In addition to this result, which follows from the theory and is expressed by the uncounfoudedness assumption conditional on the two propensity scores (Forastiere et al., 2020), our empirical results show that the performance of our estimator is not affected by the presence of homophily in the network. The only difference is in a better performance of the estimators w.r.t the spillover effects at low and high values of g, that is, the decrease in performance happens at a lower rate than what we see with the stochastic block model network. This is presumably due to the distribution of the



(a) Linear DGP - Linear Model

(b) Linear DGP - Splines

(c) Non-Linear DGP - Linear Model

(d) Non-Linear DGP - Splines

Figure 2: Posterior predictive checks under the scenario where the network is generated from a stochastic block model and the outcome is drawn from a linear model (top) or a non-linear model (bottom) with random effects. Draws from the posterior predictive distribution (blue lines) are represented with the observed distribution (black line).

neighborhood treatment which is more spread for the latent cluster model with more data at the tails of the distribution.

Finally, Tables 9, 10, 11 and 12 report the results of the scenarios where the network is taken from the Add Health data, which is likely affected by homophily. Here the average degree is lower than the other two cases, resulting in a distribution of the neighborhood treatment which is sparse and concentrated around few values (see Appendix). Moreover, the size of the communities used to generate outcome correlation is bigger, since we define a community as the cluster of students in the same school and with the same grade.

	Linear						Splines					
	RE			NO RE			RE			NO RE		
	Bias	RMSE	Coverage	Bias	RMSE	Coverage	Bias	RMSE	Coverage	Bias	RMSE	Coverage
$\tau(0, V_0)$	-0.002	0.076	0.978	-0.002	0.076	0.950	-0.014	0.077	0.974	-0.013	0.077	0.974
$\tau(0.1, V_{0.1})$	-0.002	0.076	0.976	-0.002	0.076	0.950	-0.014	0.077	0.976	-0.013	0.077	0.974
$\tau(0.2, V_{0.2})$	-0.002	0.076	0.968	-0.002	0.076	0.950	-0.014	0.077	0.974	-0.013	0.077	0.976
$\tau(0.3, V_{0.3})$	-0.002	0.076	0.980	-0.002	0.076	0.950	-0.014	0.077	0.978	-0.013	0.077	0.982
$\tau(0.4, V_{0.4})$	-0.002	0.076	0.980	-0.002	0.076	0.950	-0.014	0.077	0.980	-0.013	0.077	0.976
$\tau(0.5, V_{0.5})$	-0.002	0.076	0.974	-0.002	0.076	0.950	-0.014	0.077	0.980	-0.013	0.077	0.978
$\tau(0.6, V_{0.6})$	-0.002	0.076	0.980	-0.002	0.076	0.950	-0.014	0.077	0.980	-0.013	0.077	0.976
$\tau(0.7, V_{0.7})$	-0.002	0.076	0.982	-0.002	0.076	0.950	-0.014	0.077	0.974	-0.013	0.077	0.976
$\tau(0.8, V_{0.8})$	-0.002	0.076	0.978	-0.002	0.076	0.950	-0.014	0.077	0.972	-0.013	0.077	0.976
$\tau(0.9, V_{0.9})$	-0.002	0.076	0.968	-0.002	0.076	0.950	-0.014	0.077	0.976	-0.013	0.077	0.976
$\tau(1, V_1)$	-0.002	0.076	0.980	-0.002	0.076	0.950	-0.014	0.077	0.976	-0.013	0.077	0.974
$\delta(0.1, 0, 0)$	0.009	0.365	0.944	0.008	0.364	0.952	7.386	8.434	0.800	7.375	8.451	0.794
$\delta(0.2, 0.1, 0)$	0.008	0.227	0.942	0.007	0.227	0.946	3.890	4.500	0.820	3.886	4.514	0.798
$\delta(0.3, 0.2, 0)$	0.008	0.146	0.942	0.007	0.146	0.948	1.212	1.527	0.830	1.215	1.537	0.840
$\delta(0.4, 0.3, 0)$	0.007	0.086	0.960	0.007	0.086	0.946	0.129	0.227	0.914	0.130	0.227	0.920
$\delta(0.5, 0.4, 0)$	0.007	0.040	0.998	0.007	0.040	0.938	0.011	0.054	0.994	0.012	0.054	0.998
$\delta(0.6, 0.5, 0)$	0.006	0.042	0.992	0.007	0.042	0.946	0.005	0.057	0.996	0.004	0.056	0.996
$\delta(0.7, 0.6, 0)$	0.007	0.092	0.972	0.007	0.091	0.944	-0.156	0.294	0.906	-0.157	0.296	0.896
$\delta(0.8, 0.7, 0)$	0.005	0.146	0.962	0.007	0.146	0.938	-1.498	1.851	0.750	-1.150	1.865	0.756
$\delta(0.9, 0.8, 0)$	0.005	0.229	0.952	0.006	0.229	0.942	-4.325	4.930	0.730	-4.322.	4.946	0.738
$\delta(1, 0.9, 0)$	0.004	0.369	0.950	0.005	0.368	0.942	-7.873	8.910	0.738	-7.861	8.926	0.740
$\delta(0.1, 0, 1)$	0.009	0.365	0.944	0.008	0.364	0.952	7.386	8.434	0.802	7.375	8.451	0.792
$\delta(0.2, 0.1, 1)$	0.008	0.228	0.948	0.007	0.227	0.946	3.890	4.500	0.818	3.886	4.514	0.806
$\delta(0.3, 0.2, 1)$	0.008	0.146	0.946	0.007	0.146	0.948	1.212	1.527	0.822	1.215	1.537	0.834
$\delta(0.4, 0.3, 1)$	0.007	0.086	0.958	0.007	0.086	0.946	0.129	0.227	0.922	0.130	0.227	0.918
$\delta(0.5, 0.4, 1)$	0.007	0.040	0.994	0.007	0.040	0.938	0.011	0.054	0.996	0.012	0.054	0.996
$\delta(0.6, 0.5, 1)$	0.007	0.042	0.996	0.007	0.042	0.946	0.005	0.057	0.996	0.004	0.056	0.998
$\delta(0.7, 0.6, 1)$	0.006	0.091	0.966	0.007	0.091	0.944	-0.156	0.294	0.910	-0.157	0.296	0.902
$\delta(0.8, 0.7, 1)$	0.005	0.146	0.950	0.007	0.146	0.938	-1.498	1.851	0.750	-1.150	1.865	0.762
$\delta(0.9, 0.8, 1)$	0.005	0.229	0.946	0.006	0.229	0.942	-4.325	4.930	0.726	-4.322.	4.946	0.732
$\delta(1, 0.9, 1)$	0.004	0.369	0.946	0.005	0.368	0.942	-7.873	8.910	0.736	-7.861	8.926	0.734

Table 2: Performance of Bayesian GPS estimator for treatment effects $\tau(g, V_g)$, with $g \in \{0, 0.1, \ldots, 0.9, 1\}$, and spillover effects $\delta(g, g-0.1, z, V_g \cup V_{g-0.1})$, with $g \in \{0.1, \ldots, 0.9, 1\}$ and $z \in \{0, 1\}$, under the scenario where the network is generated from a stochastic block model and the outcome is drawn from a linear model without random effects.

In Figure 5 we can see the Add Health social network, with colors representing the communities, the node symbol representing the individual treatment and the node size representing the neighborhood treatment.

In Tables 9 and 11 we can see that the mean square error of the estimators with random effects is slightly higher than in the previous scenarios, especially for treatment effects. In

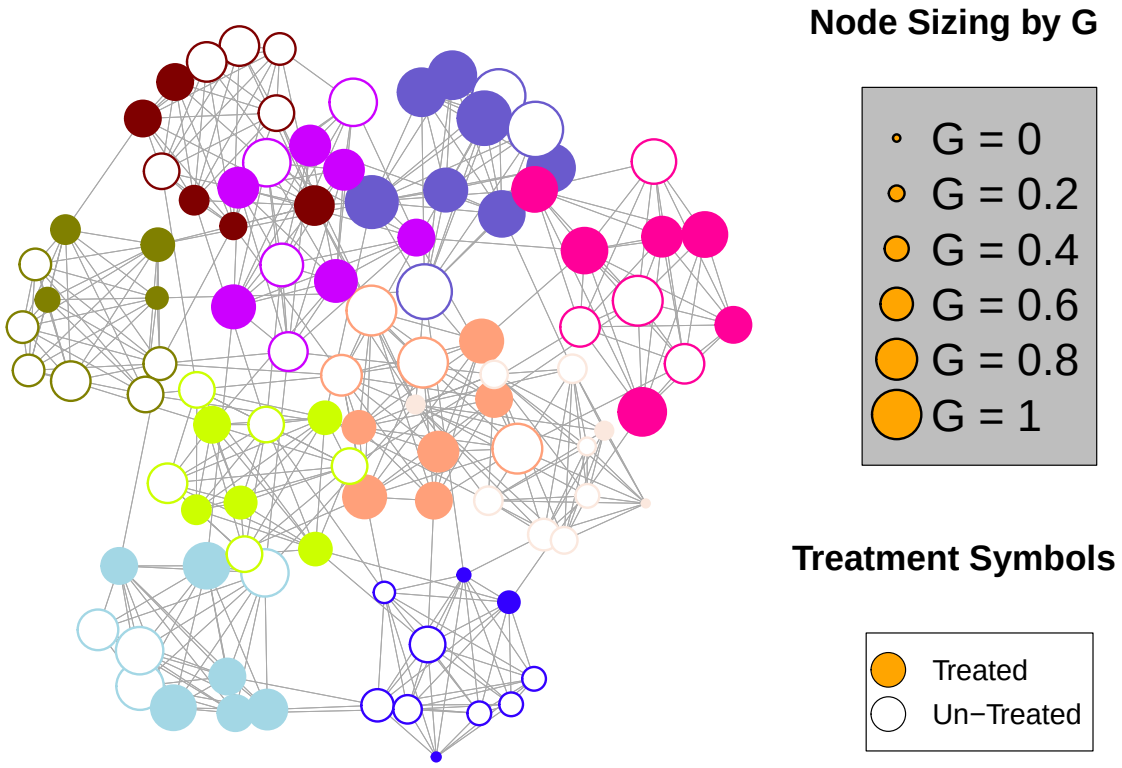


Figure 3: Stochastic Block Model Network

addition, when outcome correlation is not taken into account the MSE does not increase as much as in the previous cases. This is presumably due to the bigger size of the communities (the average size is 20.8 units). Nevertheless, for the linear estimator the coverage is still smaller when random intercepts are not included.

	Linear						Splines					
	RE			NO RE			RE			NO RE		
	Bias	RMSE	Coverage	Bias	RMSE	Coverage	Bias	RMSE	Coverage	Bias	RMSE	Coverage
$\tau(0, V_0)$	0.001	0.101	0.982	0.010	0.197	0.946	-0.010	0.082	0.976	-0.017	0.188	0.966
$\tau(0.1, V_{0.1})$	0.001	0.101	0.984	0.010	0.197	0.946	-0.010	0.082	0.974	-0.017	0.188	0.970
$\tau(0.2, V_{0.2})$	0.001	0.101	0.976	0.010	0.197	0.946	-0.010	0.082	0.976	-0.017	0.188	0.964
$\tau(0.3, V_{0.3})$	0.001	0.101	0.974	0.010	0.197	0.946	-0.010	0.082	0.976	-0.017	0.188	0.966
$\tau(0.4, V_{0.4})$	0.001	0.101	0.976	0.010	0.197	0.946	-0.010	0.082	0.972	-0.017	0.188	0.970
$\tau(0.5, V_{0.5})$	0.001	0.101	0.978	0.010	0.197	0.946	-0.010	0.082	0.972	-0.017	0.188	0.966
$\tau(0.6, V_{0.6})$	0.001	0.101	0.976	0.010	0.197	0.946	-0.010	0.082	0.974	-0.017	0.188	0.968
$\tau(0.7, V_{0.7})$	0.001	0.101	0.974	0.010	0.197	0.946	-0.010	0.082	0.974	-0.017	0.188	0.966
$\tau(0.8, V_{0.8})$	0.001	0.101	0.972	0.010	0.197	0.946	-0.010	0.082	0.972	-0.017	0.188	0.970
$\tau(0.9, V_{0.9})$	0.001	0.101	0.978	0.010	0.197	0.946	-0.010	0.082	0.976	-0.017	0.188	0.968
$\tau(1, V_1)$	0.001	0.101	0.978	0.010	0.197	0.946	-0.010	0.082	0.976	-0.017	0.188	0.974
$\delta(0.1, 0, 0)$	2.345	2.742	0.144	2.349	2.807	0.346	0.751	1.195	0.996	0.664	1.594	1.000
$\delta(0.2, 0.1, 0)$	2.464	2.611	0.008	2.466	2.638	0.054	0.373	0.731	0.994	0.368	1.051	0.998
$\delta(0.3, 0.2, 0)$	2.171	2.234	0.000	2.172	2.246	0.002	0.142	0.392	0.988	0.177	0.624	0.998
$\delta(0.4, 0.3, 0)$	1.111	1.149	0.000	1.112	1.155	0.004	0.017	0.164	0.984	0.042	0.294	0.984
$\delta(0.5, 0.4, 0)$	-0.731	0.739	0.000	-0.731	0.746	0.002	0.002	0.107	0.972	0.002	0.185	0.992
$\delta(0.6, 0.5, 0)$	-0.865	0.876	0.000	-0.866	0.883	0.000	-0.026	0.114	0.980	-0.075	0.206	0.984
$\delta(0.7, 0.6, 0)$	1.649	1.694	0.000	1.647	1.683	0.000	0.022	0.177	0.972	0.097	0.328	0.978
$\delta(0.8, 0.7, 0)$	3.084	3.146	0.000	3.082	3.147	0.000	-0.091	0.387	0.992	-0.019	0.324	0.992
$\delta(0.9, 0.8, 0)$	3.491	3.617	0.000	3.488	3.626	0.004	0.462	0.799	0.994	-0.398	1.060	1.000
$\delta(1, 0.9, 0)$	3.799	4.086	0.022	3.794	4.117	0.102	-0.792	1.236	0.996	-0.657	1.596	1.000
$\delta(0.1, 0, 0)$	2.345	2.742	0.146	2.349	2.807	0.346	0.751	1.195	0.996	0.664	1.594	1.000
$\delta(0.2, 0.1, 0)$	2.464	2.611	0.008	2.466	2.638	0.054	0.373	0.731	0.994	0.368	1.051	0.998
$\delta(0.3, 0.2, 0)$	2.171	2.234	0.000	2.172	2.246	0.002	0.142	0.392	0.988	0.177	0.624	0.998
$\delta(0.4, 0.3, 0)$	1.111	1.148	0.000	1.112	1.155	0.004	0.017	0.164	0.984	0.042	0.294	0.984
$\delta(0.5, 0.4, 0)$	-0.731	0.739	0.000	-0.731	0.746	0.002	0.002	0.107	0.972	0.002	0.185	0.992
$\delta(0.6, 0.5, 0)$	-0.865	0.876	0.000	-0.866	0.883	0.000	-0.026	0.114	0.980	-0.075	0.206	0.984
$\delta(0.7, 0.6, 0)$	1.649	1.694	0.000	1.647	1.683	0.000	0.022	0.177	0.972	0.097	0.328	0.978
$\delta(0.8, 0.7, 0)$	3.084	3.146	0.000	3.082	3.147	0.000	-0.091	0.387	0.992	-0.019	0.324	0.992
$\delta(0.9, 0.8, 0)$	3.491	3.617	0.000	3.488	3.626	0.004	0.462	0.799	0.994	-0.398	1.060	1.000
$\delta(1, 0.9, 0)$	3.799	4.086	0.022	3.794	4.117	0.102	-0.792	1.236	0.996	-0.657	1.596	1.000

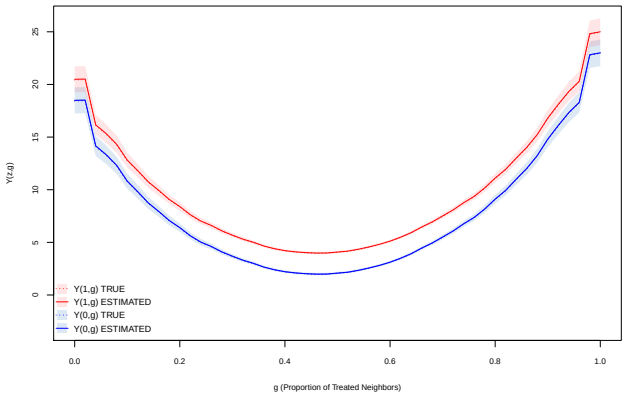
Table 3: Performance of Bayesian GPS estimator for treatment effects $\tau(g, V_g)$, with $g \in \{0, 0.1, \ldots, 0.9, 1\}$, and spillover effects $\delta(g, g-0.1, z, V_g \cup V_{g-0.1})$, with $g \in \{0.1, \ldots, 0.9, 1\}$ and $z \in \{0, 1\}$, under the scenario where the network is generated from a stochastic block model and the outcome is drawn from a non-linear model with random effects.

	Linear						Splines					
	RE			NO RE			RE			NO RE		
	Bias	RMSE	Coverage	Bias	RMSE	Coverage	Bias	RMSE	Coverage	Bias	RMSE	Coverage
$\tau(0, V_0)$	-0.001	0.105	0.960	-0.002	0.109	0.956	-0.009	0.080	0.978	-0.009	0.080	0.974
$\tau(0.1, V_{0.1})$	-0.001	0.105	0.958	-0.002	0.109	0.956	-0.009	0.080	0.974	-0.009	0.080	0.972
$\tau(0.2, V_{0.2})$	-0.001	0.105	0.956	-0.002	0.109	0.956	-0.009	0.080	0.976	-0.009	0.080	0.974
$\tau(0.3, V_{0.3})$	-0.001	0.105	0.980	-0.002	0.109	0.956	-0.009	0.080	0.976	-0.009	0.080	0.972
$\tau(0.4, V_{0.4})$	-0.001	0.105	0.960	-0.002	0.109	0.956	-0.009	0.080	0.972	-0.009	0.080	0.974
$\tau(0.5, V_{0.5})$	-0.001	0.105	0.958	-0.002	0.109	0.956	-0.009	0.080	0.972	-0.009	0.080	0.978
$\tau(0.6, V_{0.6})$	-0.001	0.105	0.960	-0.002	0.109	0.956	-0.009	0.080	0.974	-0.009	0.080	0.976
$\tau(0.7, V_{0.7})$	-0.001	0.105	0.964	-0.002	0.109	0.956	-0.009	0.080	0.972	-0.009	0.080	0.966
$\tau(0.8, V_{0.8})$	-0.001	0.105	0.958	-0.002	0.109	0.956	-0.009	0.080	0.972	-0.009	0.080	0.980
$\tau(0.9, V_{0.9})$	-0.001	0.105	0.956	-0.002	0.109	0.956	-0.009	0.080	0.976	-0.009	0.080	0.976
$\tau(1, V_1)$	-0.001	0.105	0.954	-0.002	0.109	0.956	-0.009	0.080	0.974	-0.009	0.080	0.972
$\delta(0.1, 0, 0)$	2.348	2.699	0.146	2.352	2.703	0.162	0.819	1.271	1.000	0.811	1.269	1.000
$\delta(0.2, 0.1, 0)$	2.465	2.595	0.002	2.467	2.598	0.010	0.416	0.761	0.998	0.412	0.761	0.998
$\delta(0.3, 0.2, 0)$	2.171	2.227	0.000	2.172	2.229	0.000	0.136	0.371	0.990	0.133	0.371	0.988
$\delta(0.4, 0.3, 0)$	1.110	1.144	0.000	1.111	1.145	0.000	0.021	0.170	0.970	0.020	0.169	0.972
$\delta(0.5, 0.4, 0)$	-0.731	0.739	0.000	-0.731	0.739	0.000	0.006	0.103	0.980	0.006	0.102	0.976
$\delta(0.6, 0.5, 0)$	-0.867	0.876	0.000	-0.867	0.876	0.000	-0.023	0.111	0.976	-0.023	0.111	0.974
$\delta(0.7, 0.6, 0)$	1.645	1.685	0.000	1.645	1.684	0.000	0.007	0.177	0.984	0.007	0.177	0.976
$\delta(0.8, 0.7, 0)$	3.080	3.134	0.000	3.079	3.133	0.000	-0.117	0.387	0.986	-0.113	0.386	0.988
$\delta(0.9, 0.8, 0)$	3.485	3.596	0.000	3.484	3.595	0.000	-0.500	0.832	0.994	-0.497	0.830	0.996
$\delta(1, 0.9, 0)$	3.792	4.044	0.012	3.790	4.043	0.022	-0.864	1.314	0.996	-0.857	1.312	1.000
$\delta(0.1, 0, 0)$	2.348	2.699	0.142	2.352	2.703	0.162	0.819	1.271	1.000	0.811	1.269	1.000
$\delta(0.2, 0.1, 0)$	2.465	2.595	0.008	2.467	2.598	0.010	0.416	0.761	0.998	0.412	0.761	0.998
$\delta(0.3, 0.2, 0)$	2.171	2.227	0.000	2.172	2.229	0.000	0.136	0.371	0.980	0.133	0.371	0.988
$\delta(0.4, 0.3, 0)$	1.110	1.144	0.000	1.111	1.145	0.000	0.021	0.170	0.970	0.020	0.169	0.972
$\delta(0.5, 0.4, 0)$	-0.731	0.739	0.000	-0.731	0.739	0.000	0.006	0.103	0.980	0.006	0.102	0.976
$\delta(0.6, 0.5, 0)$	-0.867	0.876	0.000	-0.867	0.876	0.000	-0.023	0.114	0.976	-0.023	0.111	0.974
$\delta(0.7, 0.6, 0)$	1.645	1.685	0.000	1.645	1.684	0.000	0.007	0.177	0.972	0.007	0.177	0.976
$\delta(0.8, 0.7, 0)$	3.080	3.134	0.000	3.079	3.133	0.000	-0.117	0.387	0.992	-0.113	0.386	0.988
$\delta(0.9, 0.8, 0)$	3.485	3.596	0.000	3.484	3.595	0.004	-0.500	0.832	0.994	-0.497	0.830	0.996
$\delta(1, 0.9, 0)$	3.792	4.044	0.014	3.790	4.043	0.022	-0.864	1.314	0.996	-0.857	1.312	1.000

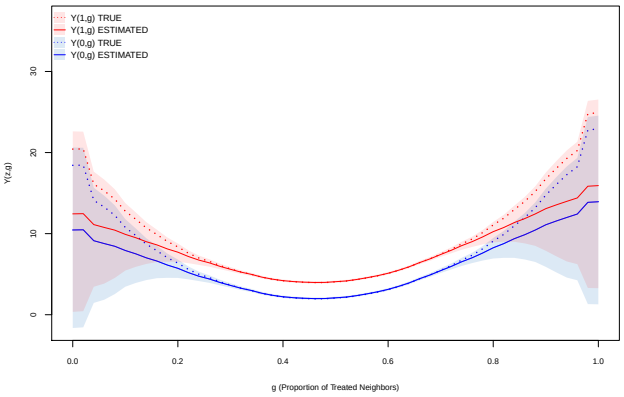
Table 4: Performance of Bayesian GPS estimator for treatment effects $\tau(g, V_g)$, with $g \in \{0, 0.1, \ldots, 0.9, 1\}$, and spillover effects $\delta(g, g-0.1, z, V_g \cup V_{g-0.1})$, with $g \in \{0.1, \ldots, 0.9, 1\}$ and $z \in \{0, 1\}$, under the scenario where the network is generated from a stochastic block model and the outcome is drawn from a non-linear model without random effects.

	Linear						Splines					
	RE			NO RE			RE			NO RE		
	Bias	RMSE	Coverage	Bias	RMSE	Coverage	Bias	RMSE	Coverage	Bias	RMSE	Coverage
$\tau(0, V_0)$	0.002	0.077	0.982	0.008	0.176	0.948	-0.011	0.078	0.980	-0.023	0.178	0.978
$\tau(0.1, V_{0.1})$	0.002	0.077	0.980	0.008	0.176	0.948	-0.011	0.078	0.976	-0.023	0.178	0.980
$\tau(0.2, V_{0.2})$	0.002	0.077	0.982	0.008	0.176	0.948	-0.011	0.078	0.976	-0.023	0.178	0.980
$\tau(0.3, V_{0.3})$	0.002	0.077	0.982	0.008	0.176	0.948	-0.011	0.078	0.976	-0.023	0.178	0.978
$\tau(0.4, V_{0.4})$	0.002	0.077	0.982	0.008	0.176	0.948	-0.011	0.078	0.980	-0.023	0.178	0.974
$\tau(0.5, V_{0.5})$	0.002	0.077	0.982	0.008	0.176	0.948	-0.011	0.078	0.978	-0.023	0.178	0.974
$\tau(0.6, V_{0.6})$	0.002	0.077	0.982	0.008	0.176	0.948	-0.011	0.078	0.978	-0.023	0.178	0.970
$\tau(0.7, V_{0.7})$	0.002	0.077	0.982	0.008	0.176	0.948	-0.011	0.078	0.974	-0.023	0.178	0.974
$\tau(0.8, V_{0.8})$	0.002	0.077	0.980	0.008	0.176	0.948	-0.011	0.078	0.970	-0.023	0.178	0.980
$\tau(0.9, V_{0.9})$	0.002	0.077	0.984	0.008	0.176	0.948	-0.011	0.078	0.978	-0.023	0.178	0.978
$\tau(1, V_1)$	0.002	0.077	0.980	0.008	0.176	0.948	-0.011	0.078	0.974	-0.023	0.178	0.976
$\delta(0.1, 0, 0)$	-0.018	0.273	0.956	-0.041	0.801	0.944	5.128	5.860	0.840	5.727	6.384	0.922
$\delta(0.2, 0.1, 0)$	-0.011	0.165	0.962	-0.025	0.501	0.936	2.207	2.591	0.850	2.525	2.926	0.930
$\delta(0.3, 0.2, 0)$	-0.007	0.109	0.974	-0.017	0.322	0.936	0.599	0.786	0.876	0.724	0.985	0.938
$\delta(0.4, 0.3, 0)$	-0.004	0.069	0.974	-0.010	0.191	0.928	0.064	0.133	0.958	0.085	0.227	0.980
$\delta(0.5, 0.4, 0)$	-0.002	0.032	1.000	-0.004	0.089	0.934	0.005	0.046	0.998	0.008	0.093	1.000
$\delta(0.6, 0.5, 0)$	0.001	0.033	0.998	0.002	0.090	0.940	0.004	0.049	1.000	-0.001	0.091	1.000
$\delta(0.7, 0.6, 0)$	0.004	0.071	0.982	0.009	0.198	0.944	-0.072	0.154	0.968	-0.112	0.251	0.964
$\delta(0.8, 0.7, 0)$	0.005	0.110	0.970	0.014	0.318	0.954	-0.777	0.972	0.808	-0.960	1.199	0.916
$\delta(0.9, 0.8, 0)$	0.011	0.179	0.968	0.025	0.500	0.956	-2.843	3.241	0.776	-3.253	3.650	0.896
$\delta(1, 0.9, 0)$	0.016	0.262	0.960	0.037	0.805	0.952	-5.348	6.010	0.786	-5.967	6.554	0.902
$\delta(0.1, 0, 1)$	-0.018	0.273	0.948	-0.041	0.801	0.944	5.128	5.860	0.840	5.727	6.384	0.922
$\delta(0.2, 0.1, 1)$	-0.011	0.165	0.942	-0.025	0.501	0.936	2.207	2.591	0.852	2.524	2.926	0.928
$\delta(0.3, 0.2, 1)$	-0.007	0.109	0.948	-0.017	0.322	0.936	0.599	0.786	0.866	0.724	0.985	0.934
$\delta(0.4, 0.3, 1)$	-0.004	0.069	0.952	-0.010	0.191	0.928	0.064	0.133	0.968	0.085	0.227	0.980
$\delta(0.5, 0.4, 1)$	-0.002	0.032	0.944	-0.004	0.089	0.934	0.005	0.046	0.998	0.008	0.093	1.000
$\delta(0.6, 0.5, 1)$	0.001	0.033	0.962	0.002	0.090	0.940	0.004	0.049	0.998	-0.001	0.091	0.998
$\delta(0.7, 0.6, 1)$	0.004	0.071	0.954	0.009	0.198	0.944	-0.072	0.154	0.972	-0.112	0.251	0.962
$\delta(0.8, 0.7, 1)$	0.005	0.110	0.960	0.014	0.318	0.954	-0.777	0.972	0.806	-0.960	1.199	0.922
$\delta(0.9, 0.8, 1)$	0.011	0.179	0.960	0.025	0.500	0.956	-2.843	3.241	0.780	-3.253	3.650	0.898
$\delta(1, 0.9, 1)$	0.016	0.262	0.960	0.037	0.805	0.952	-5.348	6.010	0.786	-5.967	6.554	0.902

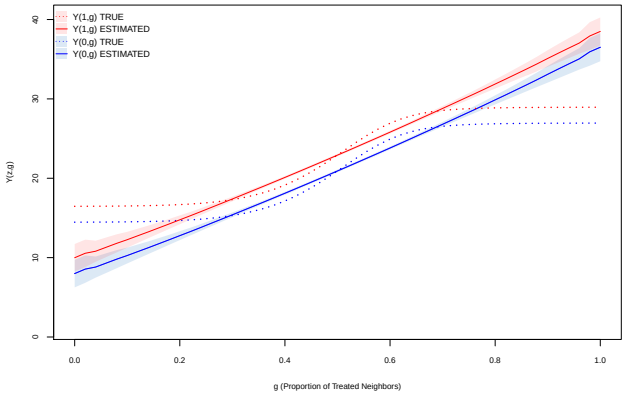
Table 5: Performance of Bayesian GPS estimator for treatment effects $\tau(g, V_g)$, with $g \in \{0, 0.1, \ldots, 0.9, 1\}$, and spillover effects $\delta(g, g-0.1, z, V_g \cup V_{g-0.1})$, with $g \in \{0.1, \ldots, 0.9, 1\}$ and $z \in \{0, 1\}$, under the scenario where the network is generated from a latent cluster model and the outcome is drawn from a linear model with random effects.



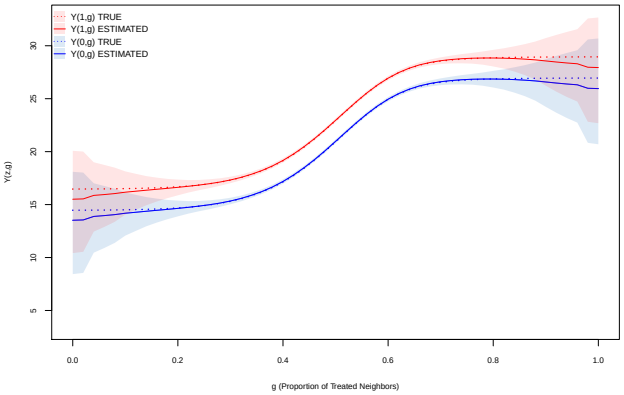
(a) Linear DGP - Linear Model



(b) Linear DGP - Splines



(c) Non-Linear DGP - Linear Model



(d) Non-Linear DGP - Splines

Figure 4: Estimated and true ()ADRF $\mu(z, g; V_g)$ under the scenario where the network is generated from a latent cluster model and the outcome is drawn from a linear model (top) or a non-linear model (bottom) with random effects. Average posterior mean (solid line) and 95 % credible intervals of $\mu(0, g; V_g)$ (blue) and $\mu(1, g; V_g)$ (red) of linear (left) and splines-based (right) estimator with random effects are represented with the true ADRFs (dotted line).

	Linear						Splines					
	RE			NO RE			RE			NO RE		
	Bias	RMSE	Coverage	Bias	RMSE	Coverage	Bias	RMSE	Coverage	Bias	RMSE	Coverage
$\tau(0, V_0)$	0.002	0.080	0.980	0.001	0.080	0.952	-0.010	0.082	0.974	-0.010	0.082	0.980
$\tau(0.1, V_{0.1})$	0.001	0.080	0.978	0.001	0.080	0.952	-0.010	0.082	0.980	-0.010	0.082	0.978
$\tau(0.2, V_{0.2})$	0.001	0.080	0.976	0.001	0.080	0.952	-0.010	0.082	0.970	-0.010	0.082	0.976
$\tau(0.3, V_{0.3})$	0.001	0.080	0.974	0.001	0.080	0.952	-0.010	0.082	0.976	-0.010	0.082	0.978
$\tau(0.4, V_{0.4})$	0.002	0.080	0.976	0.001	0.080	0.952	-0.010	0.082	0.976	-0.010	0.082	0.978
$\tau(0.5, V_{0.5})$	0.002	0.080	0.978	0.001	0.080	0.952	-0.010	0.082	0.972	-0.010	0.082	0.978
$\tau(0.6, V_{0.6})$	0.001	0.080	0.980	0.001	0.080	0.952	-0.010	0.082	0.976	-0.010	0.082	0.980
$\tau(0.7, V_{0.7})$	0.001	0.080	0.976	0.001	0.080	0.952	-0.010	0.082	0.980	-0.010	0.082	0.976
$\tau(0.8, V_{0.8})$	0.001	0.080	0.980	0.001	0.080	0.952	-0.010	0.082	0.982	-0.010	0.082	0.978
$\tau(0.9, V_{0.9})$	0.002	0.080	0.978	0.001	0.080	0.952	-0.010	0.082	0.972	-0.010	0.082	0.972
$\tau(1, V_1)$	0.001	0.080	0.972	0.001	0.080	0.952	-0.010	0.082	0.976	-0.010	0.082	0.974
$\delta(0.1, 0, 0)$	0.002	0.274	0.940	0.003	0.274	0.930	5.331	6.122	0.816	5.317	6.092	0.828
$\delta(0.2, 0.1, 0)$	0.002	0.165	0.942	0.003	0.265	0.928	2.303	2.739	0.818	2.301	2.731	0.822
$\delta(0.3, 0.2, 0)$	0.002	0.110	0.948	0.003	0.110	0.924	0.645	0.871	0.834	0.646	0.873	0.834
$\delta(0.4, 0.3, 0)$	0.002	0.069	0.970	0.003	0.069	0.928	0.076	0.153	0.926	0.077	0.154	0.938
$\delta(0.5, 0.4, 0)$	0.002	0.032	0.998	0.003	0.032	0.950	0.004	0.045	0.998	0.004	0.045	0.998
$\delta(0.6, 0.5, 0)$	0.002	0.031	1.000	0.003	0.031	0.944	0.004	0.045	1.000	0.004	0.046	0.998
$\delta(0.7, 0.6, 0)$	0.002	0.067	0.982	0.003	0.068	0.960	-0.066	0.153	0.946	-0.065	0.152	0.954
$\delta(0.8, 0.7, 0)$	0.002	0.108	0.972	0.003	0.108	0.962	-0.794	1.108	0.784	-0.789	1.100	0.794
$\delta(0.9, 0.8, 0)$	0.002	0.175	0.960	0.003	0.175	0.950	-2.918	3.362	0.732	-2.903	3.339	0.734
$\delta(1, 0.9, 0)$	0.002	0.260	0.960	0.003	0.261	0.946	-5.543	6.272	0.736	-5.196	6.232	0.744
$\delta(0.1, 0, 1)$	0.002	0.274	0.940	0.003	0.274	0.930	5.331	6.122	0.816	5.317	6.092	0.828
$\delta(0.2, 0.1, 1)$	0.002	0.165	0.942	0.003	0.265	0.928	2.303	2.739	0.818	2.301	2.731	0.822
$\delta(0.3, 0.2, 1)$	0.002	0.110	0.948	0.003	0.110	0.924	0.645	0.871	0.834	0.646	0.873	0.834
$\delta(0.4, 0.3, 1)$	0.002	0.069	0.970	0.003	0.069	0.928	0.076	0.153	0.926	0.077	0.154	0.938
$\delta(0.5, 0.4, 1)$	0.002	0.032	0.998	0.003	0.032	0.950	0.004	0.045	0.998	0.004	0.045	0.998
$\delta(0.6, 0.5, 1)$	0.002	0.031	1.000	0.003	0.031	0.944	0.004	0.045	1.000	0.004	0.046	0.998
$\delta(0.7, 0.6, 1)$	0.002	0.067	0.982	0.003	0.068	0.960	-0.066	0.153	0.946	-0.065	0.152	0.954
$\delta(0.8, 0.7, 1)$	0.002	0.108	0.972	0.003	0.108	0.962	-0.794	1.108	0.784	-0.789	1.100	0.794
$\delta(0.9, 0.8, 1)$	0.002	0.175	0.960	0.003	0.175	0.950	-2.918	3.362	0.732	-2.903	3.339	0.734
$\delta(1, 0.9, 1)$	0.002	0.260	0.960	0.003	0.261	0.946	-5.543	6.272	0.736	-5.196	6.232	0.744

Table 6: Performance of Bayesian GPS estimator for treatment effects $\tau(g, V_g)$, with $g \in \{0, 0.1, \ldots, 0.9, 1\}$, and spillover effects $\delta(g, g-0.1, z, V_g \cup V_{g-0.1})$, with $g \in \{0.1, \ldots, 0.9, 1\}$ and $z \in \{0, 1\}$, under the scenario where the network is generated from a latent cluster model and the outcome is drawn from a linear model without random effects.

	Linear						Splines					
	RE			NO RE			RE			NO RE		
	Bias	RMSE	Coverage	Bias	RMSE	Coverage	Bias	RMSE	Coverage	Bias	RMSE	Coverage
$\tau(0, V_0)$	-0.004	0.120	0.962	-0.011	0.201	0.938	-0.013	0.086	0.964	-0.034	0.184	0.960
$\tau(0.1, V_{0.1})$	-0.004	0.120	0.962	-0.011	0.201	0.938	-0.013	0.086	0.968	-0.034	0.184	0.962
$\tau(0.2, V_{0.2})$	-0.004	0.120	0.956	-0.011	0.201	0.938	-0.013	0.086	0.966	-0.034	0.184	0.964
$\tau(0.3, V_{0.3})$	-0.004	0.120	0.960	-0.011	0.201	0.938	-0.013	0.086	0.958	-0.034	0.184	0.966
$\tau(0.4, V_{0.4})$	-0.004	0.120	0.956	-0.011	0.201	0.938	-0.013	0.086	0.962	-0.034	0.184	0.968
$\tau(0.5, V_{0.5})$	-0.004	0.120	0.960	-0.011	0.201	0.938	-0.013	0.086	0.966	-0.034	0.184	0.964
$\tau(0.6, V_{0.6})$	-0.004	0.120	0.960	-0.011	0.201	0.938	-0.013	0.086	0.956	-0.034	0.184	0.954
$\tau(0.7, V_{0.7})$	-0.004	0.120	0.960	-0.011	0.201	0.938	-0.013	0.086	0.964	-0.034	0.184	0.956
$\tau(0.8, V_{0.8})$	-0.004	0.120	0.962	-0.011	0.201	0.938	-0.013	0.086	0.968	-0.034	0.184	0.964
$\tau(0.9, V_{0.9})$	-0.004	0.120	0.964	-0.011	0.201	0.938	-0.013	0.086	0.954	-0.034	0.184	0.958
$\tau(1, V_1)$	-0.004	0.120	0.962	-0.011	0.201	0.938	-0.013	0.086	0.962	-0.034	0.184	0.958
$\delta(0.1, 0, 0)$	2.231	2.450	0.064	2.255	2.522	0.198	0.621	1.052	0.996	0.535	1.323	1.000
$\delta(0.2, 0.1, 0)$	2.331	2.404	0.000	2.344	2.436	0.012	0.285	0.596	0.988	0.379	0.814	0.998
$\delta(0.3, 0.2, 0)$	1.999	2.033	0.000	2.007	2.051	0.000	0.069	0.269	0.986	0.081	0.445	0.990
$\delta(0.4, 0.3, 0)$	0.886	0.914	0.000	0.890	0.926	0.008	0.014	0.147	0.978	0.041	0.251	0.988
$\delta(0.5, 0.4, 0)$	-0.982	0.986	0.000	-0.982	0.988	0.000	-0.004	0.110	0.974	-0.011	0.195	0.994
$\delta(0.6, 0.5, 0)$	-1.110	1.115	0.000	-1.114	1.121	0.000	-0.020	0.108	0.978	-0.057	0.203	0.986
$\delta(0.7, 0.6, 0)$	1.371	1.400	0.000	1.362	1.397	0.000	0.031	0.155	0.970	0.046	0.380	0.986
$\delta(0.8, 0.7, 0)$	2.260	2.797	0.000	2.747	2.791	0.000	-0.060	0.282	0.990	0.006	0.456	0.990
$\delta(0.9, 0.8, 0)$	3.155	3.235	0.000	3.135	3.232	0.002	0.346	0.673	0.984	-0.255	0.888	0.996
$\delta(1, 0.9, 0)$	3.888	3.543	0.008	3.360	3.548	0.024	-0.641	1.053	0.994	-0.523	1.296	1.000
$\delta(0.1, 0, 1)$	2.231	2.450	0.064	2.255	2.522	0.198	0.621	1.052	0.996	0.535	1.323	1.000
$\delta(0.2, 0.1, 1)$	2.331	2.404	0.000	2.344	2.436	0.012	0.285	0.596	0.988	0.379	0.814	0.998
$\delta(0.3, 0.2, 1)$	1.999	2.033	0.000	2.007	2.051	0.000	0.069	0.269	0.986	0.081	0.445	0.990
$\delta(0.4, 0.3, 1)$	0.886	0.914	0.000	0.890	0.926	0.008	0.014	0.147	0.978	0.041	0.251	0.988
$\delta(0.5, 0.4, 1)$	-0.982	0.986	0.000	-0.982	0.988	0.000	-0.004	0.110	0.974	-0.011	0.195	0.994
$\delta(0.6, 0.5, 1)$	-1.110	1.115	0.000	-1.114	1.121	0.000	-0.020	0.108	0.978	-0.057	0.203	0.986
$\delta(0.7, 0.6, 1)$	1.371	1.400	0.000	1.362	1.397	0.000	0.031	0.155	0.970	0.046	0.380	0.986
$\delta(0.8, 0.7, 1)$	2.260	2.797	0.000	2.747	2.791	0.000	-0.060	0.282	0.990	0.006	0.456	0.990
$\delta(0.9, 0.8, 1)$	3.155	3.235	0.000	3.135	3.232	0.002	0.346	0.673	0.984	-0.255	0.888	0.996
$\delta(1, 0.9, 1)$	3.888	3.543	0.008	3.360	3.548	0.024	-0.641	1.053	0.994	-0.523	1.296	1.000

Table 7: Performance of Bayesian GPS estimator for treatment effects $\tau(g, V_g)$, with $g \in \{0, 0.1, \ldots, 0.9, 1\}$, and spillover effects $\delta(g, g-0.1, z, V_g \cup V_{g-0.1})$, with $g \in \{0.1, \ldots, 0.9, 1\}$ and $z \in \{0, 1\}$, under the scenario where the network is generated from a latent cluster model and the outcome is drawn from a non-linear model with random effects.

	Linear						Splines					
	RE			NO RE			RE			NO RE		
	Bias	RMSE	Coverage	Bias	RMSE	Coverage	Bias	RMSE	Coverage	Bias	RMSE	Coverage
$\tau(0, V_0)$	-0.005	0.105	0.982	-0.005	0.106	0.974	-0.008	0.077	0.972	-0.008	0.077	0.972
$\tau(0.1, V_{0.1})$	-0.005	0.105	0.976	-0.005	0.106	0.974	-0.008	0.077	0.970	-0.008	0.077	0.970
$\tau(0.2, V_{0.2})$	-0.005	0.105	0.978	-0.005	0.106	0.974	-0.008	0.077	0.974	-0.008	0.077	0.972
$\tau(0.3, V_{0.3})$	-0.005	0.105	0.982	-0.005	0.106	0.974	-0.008	0.077	0.972	-0.008	0.077	0.972
$\tau(0.4, V_{0.4})$	-0.005	0.105	0.978	-0.005	0.106	0.974	-0.008	0.077	0.972	-0.008	0.077	0.976
$\tau(0.5, V_{0.5})$	-0.005	0.105	0.978	-0.005	0.106	0.974	-0.008	0.077	0.972	-0.008	0.077	0.972
$\tau(0.6, V_{0.6})$	-0.005	0.105	0.984	-0.005	0.106	0.974	-0.008	0.077	0.972	-0.008	0.077	0.970
$\tau(0.7, V_{0.7})$	-0.005	0.105	0.976	-0.005	0.106	0.974	-0.008	0.077	0.974	-0.008	0.077	0.974
$\tau(0.8, V_{0.8})$	-0.005	0.105	0.982	-0.005	0.106	0.974	-0.008	0.077	0.974	-0.008	0.077	0.970
$\tau(0.9, V_{0.9})$	-0.005	0.105	0.974	-0.005	0.106	0.974	-0.008	0.077	0.976	-0.008	0.077	0.972
$\tau(1, V_1)$	-0.005	0.105	0.980	-0.005	0.106	0.974	-0.008	0.077	0.968	-0.008	0.077	0.974
$\delta(0.1, 0, 0)$	2.210	2.448	0.064	2.209	2.450	0.090	0.621	0.960	0.998	0.629	0.965	0.998
$\delta(0.2, 0.1, 0)$	2.320	2.401	0.000	2.320	2.401	0.000	0.274	0.524	0.998	0.278	0.523	0.996
$\delta(0.3, 0.2, 0)$	1.997	2.033	0.000	1.996	2.033	0.000	0.075	0.274	0.976	0.076	0.274	0.982
$\delta(0.4, 0.3, 0)$	0.889	0.918	0.000	0.888	0.918	0.000	0.008	0.135	0.982	0.007	0.135	0.976
$\delta(0.5, 0.4, 0)$	-0.975	0.979	0.000	-0.975	0.979	0.000	-0.004	0.105	0.980	-0.004	0.106	0.980
$\delta(0.6, 0.5, 0)$	-1.099	1.105	0.000	-1.100	1.105	0.000	-0.018	0.104	0.974	-0.017	0.104	0.968
$\delta(0.7, 0.6, 0)$	1.388	1.421	0.000	1.388	1.421	0.000	0.023	0.146	0.978	0.022	0.145	0.976
$\delta(0.8, 0.7, 0)$	2.782	2.823	0.000	2.782	2.823	0.000	-0.058	0.275	0.986	-0.058	0.276	0.988
$\delta(0.9, 0.8, 0)$	3.188	3.276	0.000	3.188	3.277	0.000	0.329	0.610	1.000	-0.331	0.613	0.996
$\delta(1, 0.9, 0)$	3.426	3.593	0.008	3.426	3.595	0.012	-0.639	0.967	1.000	-0.6483	0.971	1.000
$\delta(0.1, 0, 1)$	2.210	2.448	0.064	2.209	2.450	0.090	0.621	0.960	0.998	0.629	0.965	0.998
$\delta(0.2, 0.1, 1)$	2.320	2.401	0.000	2.320	2.401	0.000	0.274	0.524	0.998	0.278	0.523	0.996
$\delta(0.3, 0.2, 1)$	1.997	2.033	0.000	1.996	2.033	0.000	0.075	0.274	0.976	0.076	0.274	0.982
$\delta(0.4, 0.3, 1)$	0.889	0.918	0.000	0.888	0.918	0.000	0.008	0.135	0.982	0.007	0.135	0.976
$\delta(0.5, 0.4, 1)$	-0.975	0.979	0.000	-0.975	0.979	0.000	-0.004	0.105	0.980	-0.004	0.106	0.980
$\delta(0.6, 0.5, 1)$	-1.099	1.105	0.000	-1.100	1.105	0.000	-0.018	0.104	0.974	-0.017	0.104	0.968
$\delta(0.7, 0.6, 1)$	1.388	1.421	0.000	1.388	1.421	0.000	0.023	0.146	0.978	0.022	0.145	0.976
$\delta(0.8, 0.7, 1)$	2.782	2.823	0.000	2.782	2.823	0.000	-0.058	0.275	0.986	-0.058	0.276	0.988
$\delta(0.9, 0.8, 1)$	3.188	3.276	0.000	3.188	3.277	0.000	0.329	0.610	1.000	-0.331	0.613	0.996
$\delta(1, 0.9, 1)$	3.426	3.593	0.008	3.426	3.595	0.012	-0.639	0.967	1.000	-0.6483	0.971	1.000

Table 8: Performance of Bayesian GPS estimator for treatment effects $\tau(g, V_g)$, with $g \in \{0, 0.1, \ldots, 0.9, 1\}$, and spillover effects $\delta(g, g-0.1, z, V_g \cup V_{g-0.1})$, with $g \in \{0.1, \ldots, 0.9, 1\}$ and $z \in \{0, 1\}$, under the scenario where the network is generated from a latent cluster model and the outcome is drawn from a non-linear model without random effects.

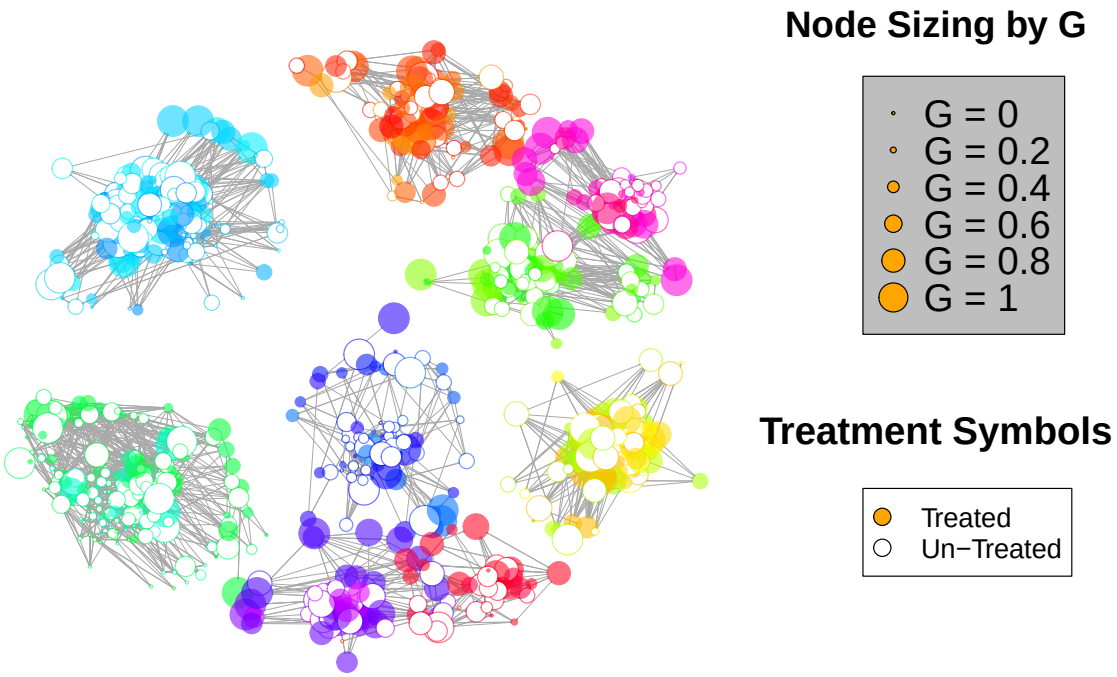


Figure 5: Add Health Network

	Linear						Splines					
	RE			NO RE			RE			NO RE		
	Bias	RMSE	Coverage	Bias	RMSE	Coverage	Bias	RMSE	Coverage	Bias	RMSE	Coverage
$\tau(0, V_0)$	0.007	0.132	0.988	0.006	0.141	0.958	-0.002	0.130	0.986	-0.002	0.140	0.992
$\tau(0.1, V_{0.1})$	0.007	0.132	0.986	0.006	0.141	0.958	-0.003	0.130	0.988	-0.001	0.140	0.990
$\tau(0.2, V_{0.2})$	0.007	0.132	0.986	0.006	0.141	0.958	-0.002	0.130	0.984	-0.001	0.140	0.990
$\tau(0.3, V_{0.3})$	0.007	0.132	0.990	0.006	0.141	0.958	-0.002	0.130	0.988	-0.002	0.140	0.992
$\tau(0.4, V_{0.4})$	0.007	0.132	0.988	0.006	0.141	0.958	-0.002	0.130	0.988	-0.002	0.140	0.984
$\tau(0.5, V_{0.5})$	0.007	0.132	0.988	0.006	0.141	0.958	-0.002	0.130	0.988	-0.002	0.140	0.992
$\tau(0.6, V_{0.6})$	0.007	0.132	0.988	0.006	0.141	0.958	-0.002	0.130	0.986	-0.002	0.140	0.996
$\tau(0.7, V_{0.7})$	0.007	0.132	0.988	0.006	0.141	0.958	-0.002	0.130	0.984	-0.002	0.140	0.992
$\tau(0.8, V_{0.8})$	0.007	0.132	0.988	0.006	0.141	0.958	-0.002	0.130	0.988	-0.002	0.140	0.990
$\tau(0.9, V_{0.9})$	0.007	0.132	0.986	0.006	0.141	0.958	-0.002	0.130	0.984	-0.001	0.140	0.988
$\tau(1, V_1)$	0.007	0.132	0.986	0.006	0.141	0.958	-0.002	0.130	0.986	-0.002	0.140	0.988
$\delta(0.1, 0, 0)$	-0.005	0.131	0.970	-0.009	0.150	0.888	0.845	1.019	0.896	0.841	1.008	0.916
$\delta(0.2, 0.1, 0)$	-0.004	0.106	0.982	-0.008	0.121	0.886	0.175	0.287	0.938	0.172	0.288	0.968
$\delta(0.3, 0.2, 0)$	-0.003	0.069	0.998	-0.005	0.081	0.884	0.020	0.099	1.000	0.019	0.110	0.996
$\delta(0.4, 0.3, 0)$	-0.002	0.045	1.000	-0.004	0.056	0.852	0.002	0.069	0.998	0.001	0.087	1.000
$\delta(0.5, 0.4, 0)$	-0.001	0.032	1.000	-0.003	0.042	0.792	0.001	0.055	1.000	0.001	0.077	1.000
$\delta(0.6, 0.5, 0)$	0.000	0.031	1.000	-0.002	0.042	0.794	0.002	0.053	1.000	0.000	0.069	0.998
$\delta(0.7, 0.6, 0)$	0.000	0.050	1.000	0.000	0.061	0.854	0.002	0.071	1.000	0.002	0.084	1.000
$\delta(0.8, 0.7, 0)$	0.001	0.058	0.996	0.001	0.070	0.866	-0.030	0.090	1.000	-0.027	0.098	0.996
$\delta(0.9, 0.8, 0)$	0.002	0.080	0.992	0.002	0.094	0.880	-0.134	0.191	0.966	-0.135	0.195	0.972
$\delta(1, 0.9, 0)$	0.004	0.151	0.966	0.006	0.174	0.888	-0.742	0.898	0.836	-0.748	0.901	0.860
$\delta(0.1, 0, 1)$	-0.005	0.131	0.970	-0.009	0.150	0.888	0.845	1.019	0.896	0.841	1.008	0.916
$\delta(0.2, 0.1, 1)$	-0.004	0.106	0.982	-0.008	0.121	0.886	0.175	0.287	0.938	0.172	0.288	0.968
$\delta(0.3, 0.2, 1)$	-0.003	0.069	0.998	-0.005	0.081	0.884	0.020	0.099	1.000	0.019	0.110	0.996
$\delta(0.4, 0.3, 1)$	-0.002	0.045	1.000	-0.004	0.056	0.852	0.002	0.069	0.998	0.001	0.087	1.000
$\delta(0.5, 0.4, 1)$	-0.001	0.032	1.000	-0.003	0.042	0.792	0.001	0.055	1.000	0.001	0.077	1.000
$\delta(0.6, 0.5, 1)$	0.000	0.031	1.000	-0.002	0.042	0.794	0.002	0.053	1.000	0.000	0.069	0.998
$\delta(0.7, 0.6, 1)$	0.000	0.050	1.000	0.000	0.061	0.854	0.002	0.071	1.000	0.002	0.084	1.000
$\delta(0.8, 0.7, 1)$	0.001	0.058	0.996	0.001	0.070	0.866	-0.030	0.090	1.000	-0.027	0.098	0.996
$\delta(0.9, 0.8, 1)$	0.002	0.080	0.992	0.002	0.094	0.880	-0.134	0.191	0.966	-0.135	0.195	0.972
$\delta(1, 0.9, 1)$	0.004	0.151	0.966	0.006	0.174	0.888	-0.742	0.898	0.836	-0.748	0.901	0.860

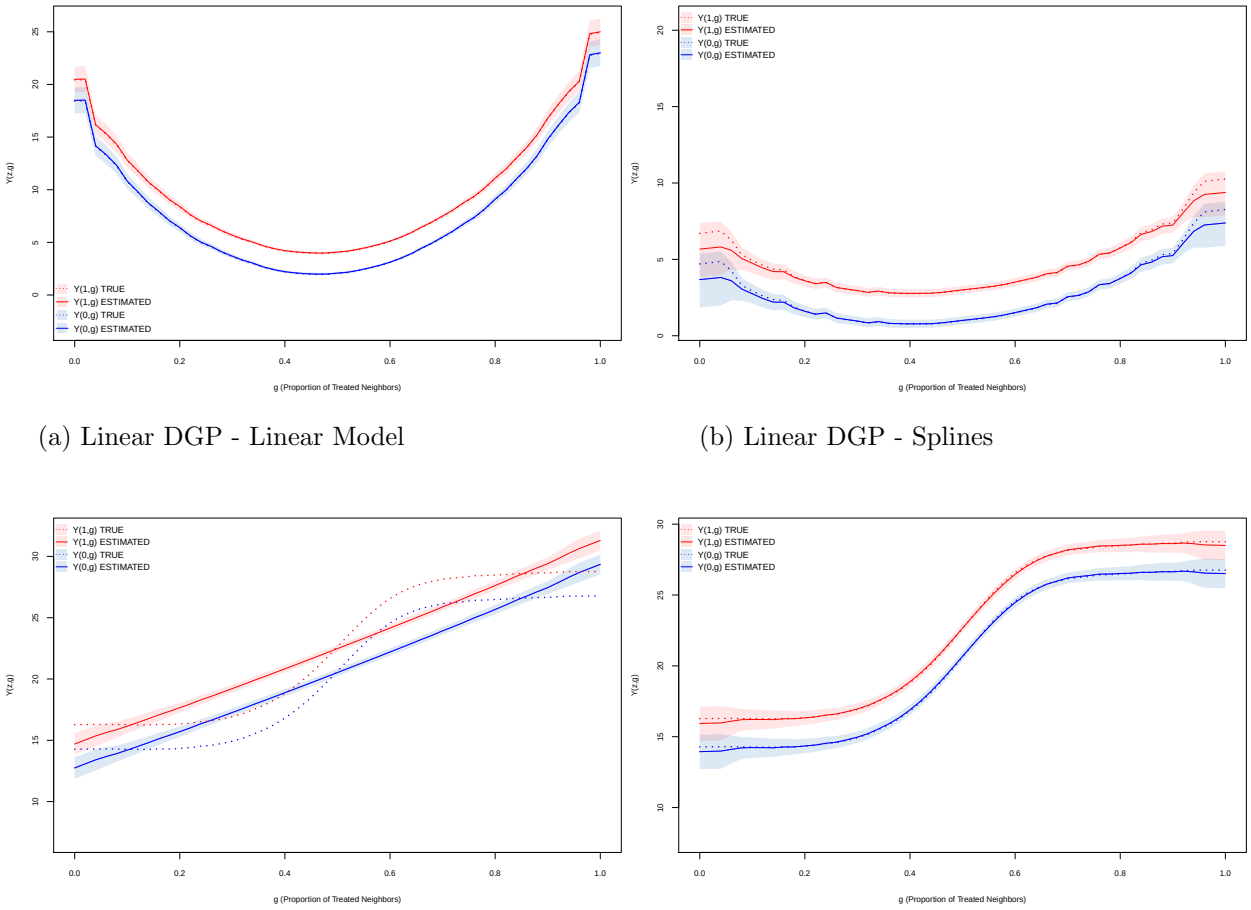
Table 9: Performance of Bayesian GPS estimator for treatment effects $\tau(g, V_g)$, with $g \in \{0, 0.1, \ldots, 0.9, 1\}$, and spillover effects $\delta(g, g-0.1, z, V_g \cup V_{g-0.1})$, with $g \in \{0.1, \ldots, 0.9, 1\}$ and $z \in \{0, 1\}$, under the scenario where the network is taken from the Add Health data and the outcome is drawn from a linear model with random effects.

	Linear						Splines					
	RE			NO RE			RE			NO RE		
	Bias	RMSE	Coverage	Bias	RMSE	Coverage	Bias	RMSE	Coverage	Bias	RMSE	Coverage
$\tau(0, V_0)$	0.000	0.066	0.984	0.000	0.066	0.952	-0.006	0.066	0.980	-0.005	0.066	0.980
$\tau(0.1, V_{0.1})$	0.000	0.066	0.984	0.000	0.066	0.952	-0.006	0.066	0.980	-0.006	0.066	0.978
$\tau(0.2, V_{0.2})$	0.000	0.066	0.984	0.000	0.066	0.952	-0.006	0.066	0.974	-0.005	0.066	0.978
$\tau(0.3, V_{0.3})$	0.000	0.066	0.984	0.000	0.066	0.952	-0.006	0.066	0.978	-0.006	0.066	0.976
$\tau(0.4, V_{0.4})$	0.000	0.066	0.984	0.000	0.066	0.952	-0.006	0.066	0.976	-0.005	0.066	0.978
$\tau(0.5, V_{0.5})$	0.000	0.066	0.984	0.000	0.066	0.952	-0.006	0.066	0.982	-0.006	0.066	0.976
$\tau(0.6, V_{0.6})$	0.000	0.066	0.984	0.000	0.066	0.952	-0.006	0.066	0.972	-0.006	0.066	0.978
$\tau(0.7, V_{0.7})$	0.000	0.066	0.984	0.000	0.066	0.952	-0.006	0.066	0.980	-0.006	0.066	0.978
$\tau(0.8, V_{0.8})$	0.000	0.066	0.984	0.000	0.066	0.952	-0.006	0.066	0.982	-0.006	0.066	0.978
$\tau(0.9, V_{0.9})$	0.000	0.066	0.984	0.000	0.066	0.952	-0.006	0.066	0.982	-0.006	0.066	0.976
$\tau(1, V_1)$	0.000	0.066	0.984	0.000	0.066	0.952	-0.006	0.066	0.982	-0.005	0.066	0.978
$\delta(0.1, 0, 0)$	0.000	0.057	0.986	0.001	0.057	0.940	0.555	0.737	0.894	0.553	0.740	0.882
$\delta(0.2, 0.1, 0)$	0.000	0.046	0.992	0.001	0.046	0.944	0.095	0.169	0.950	0.095	0.170	0.942
$\delta(0.3, 0.2, 0)$	0.000	0.030	1.000	0.000	0.030	0.946	0.009	0.045	0.998	0.009	0.044	0.994
$\delta(0.4, 0.3, 0)$	0.000	0.020	1.000	0.000	0.020	0.942	0.000	0.028	1.000	0.001	0.029	1.000
$\delta(0.5, 0.4, 0)$	0.000	0.014	1.000	0.000	0.013	0.944	0.000	0.024	1.000	0.000	0.024	1.000
$\delta(0.6, 0.5, 0)$	0.000	0.013	1.000	0.000	0.013	0.936	0.000	0.024	1.000	0.000	0.024	1.000
$\delta(0.7, 0.6, 0)$	0.000	0.021	1.000	0.000	0.021	0.946	0.002	0.031	1.000	0.002	0.031	1.000
$\delta(0.8, 0.7, 0)$	0.000	0.024	1.000	0.000	0.024	0.948	-0.016	0.047	0.996	-0.016	0.047	0.994
$\delta(0.9, 0.8, 0)$	0.000	0.034	1.000	0.000	0.034	0.948	-0.077	0.115	0.928	-0.079	0.115	0.926
$\delta(1, 0.9, 0)$	0.000	0.064	0.986	0.000	0.064	0.948	-0.444	0.578	0.852	-0.442	0.580	0.846
$\delta(0.1, 0, 1)$	0.000	0.057	0.986	0.001	0.057	0.940	0.555	0.737	0.894	0.553	0.740	0.882
$\delta(0.2, 0.1, 1)$	0.000	0.046	0.992	0.001	0.046	0.944	0.095	0.169	0.950	0.095	0.170	0.942
$\delta(0.3, 0.2, 1)$	0.000	0.030	1.000	0.000	0.030	0.946	0.009	0.045	0.998	0.009	0.044	0.994
$\delta(0.4, 0.3, 1)$	0.000	0.020	1.000	0.000	0.020	0.942	0.000	0.028	1.000	0.001	0.029	1.000
$\delta(0.5, 0.4, 1)$	0.000	0.014	1.000	0.000	0.013	0.944	0.000	0.024	1.000	0.000	0.024	1.000
$\delta(0.6, 0.5, 1)$	0.000	0.013	1.000	0.000	0.013	0.936	0.000	0.024	1.000	0.000	0.024	1.000
$\delta(0.7, 0.6, 1)$	0.000	0.021	1.000	0.000	0.021	0.946	0.002	0.031	1.000	0.002	0.031	1.000
$\delta(0.8, 0.7, 1)$	0.000	0.024	1.000	0.000	0.024	0.948	-0.016	0.047	0.996	-0.016	0.047	0.994
$\delta(0.9, 0.8, 1)$	0.000	0.034	1.000	0.000	0.034	0.948	-0.077	0.115	0.928	-0.079	0.115	0.926
$\delta(1, 0.9, 1)$	0.000	0.064	0.986	0.000	0.064	0.948	-0.444	0.578	0.852	-0.442	0.580	0.846

Table 10: Performance of Bayesian GPS estimator for treatment effects $\tau(g, V_g)$, with $g \in \{0, 0.1, \ldots, 0.9, 1\}$, and spillover effects $\delta(g, g-0.1, z, V_g \cup V_{g-0.1})$, with $g \in \{0.1, \ldots, 0.9, 1\}$ and $z \in \{0, 1\}$, under the scenario where the network is taken from the Add Health data and the outcome is drawn from a linear model without random effects.

	Linear						Splines					
	RE			NO RE			RE			NO RE		
	Bias	RMSE	Coverage	Bias	RMSE	Coverage	Bias	RMSE	Coverage	Bias	RMSE	Coverage
$\tau(0, V_0)$	-0.049	0.202	0.962	0.001	0.204	0.938	-0.012	0.138	0.980	-0.009	0.145	0.984
$\tau(0.1, V_{0.1})$	-0.048	0.202	0.954	0.001	0.204	0.938	-0.012	0.138	0.982	-0.009	0.145	0.982
$\tau(0.2, V_{0.2})$	-0.049	0.202	0.962	0.001	0.204	0.938	-0.012	0.138	0.980	-0.009	0.145	0.984
$\tau(0.3, V_{0.3})$	-0.049	0.202	0.964	0.001	0.204	0.938	-0.012	0.138	0.984	-0.009	0.145	0.986
$\tau(0.4, V_{0.4})$	-0.049	0.202	0.956	0.001	0.204	0.938	-0.012	0.138	0.982	-0.009	0.145	0.986
$\tau(0.5, V_{0.5})$	-0.049	0.202	0.952	0.001	0.204	0.938	-0.012	0.138	0.988	-0.009	0.145	0.982
$\tau(0.6, V_{0.6})$	-0.049	0.202	0.960	0.001	0.204	0.938	-0.012	0.138	0.982	-0.009	0.145	0.982
$\tau(0.7, V_{0.7})$	-0.049	0.202	0.962	0.001	0.204	0.938	-0.012	0.138	0.982	-0.009	0.145	0.980
$\tau(0.8, V_{0.8})$	-0.049	0.202	0.958	0.001	0.204	0.938	-0.012	0.138	0.984	-0.009	0.145	0.974
$\tau(0.9, V_{0.9})$	-0.049	0.202	0.960	0.001	0.204	0.938	-0.012	0.138	0.978	-0.009	0.145	0.980
$\tau(1, V_1)$	-0.048	0.202	0.960	0.001	0.204	0.938	-0.012	0.138	0.978	-0.009	0.145	0.984
$\delta(0.1, 0, 0)$	1.459	1.471	0.000	1.482	1.495	0.000	0.288	0.471	0.996	0.252	0.491	0.958
$\delta(0.2, 0.1, 0)$	1.443	1.451	0.000	1.466	1.474	0.000	0.055	0.276	0.982	0.050	0.282	0.984
$\delta(0.3, 0.2, 0)$	0.976	0.981	0.000	1.000	1.005	0.000	0.035	0.227	0.974	0.031	0.241	0.966
$\delta(0.4, 0.3, 0)$	-0.285	0.291	0.348	-0.260	0.269	0.030	0.026	0.197	0.976	0.034	0.206	0.980
$\delta(0.5, 0.4, 0)$	-2.173	2.173	0.000	-2.148	2.148	0.000	-0.039	0.178	0.964	-0.040	0.199	0.966
$\delta(0.6, 0.5, 0)$	-2.259	2.259	0.000	-2.233	2.234	0.000	-0.135	0.221	0.932	-0.140	0.237	0.934
$\delta(0.7, 0.6, 0)$	0.150	0.166	0.922	0.176	0.192	0.212	0.155	0.251	0.912	0.168	0.268	0.910
$\delta(0.8, 0.7, 0)$	1.409	1.412	0.000	1.435	1.438	0.000	-0.015	0.211	0.988	-0.006	0.232	0.966
$\delta(0.9, 0.8, 0)$	1.612	1.617	0.000	1.638	1.644	0.000	-0.049	0.222	0.992	-0.053	0.236	0.984
$\delta(1, 0.9, 0)$	1.788	1.806	0.000	1.817	1.834	0.000	-0.238	0.437	0.972	-0.236	0.464	0.958
$\delta(0.1, 0, 1)$	1.459	1.471	0.000	1.482	1.495	0.000	0.288	0.471	0.996	0.252	0.491	0.958
$\delta(0.2, 0.1, 1)$	1.443	1.451	0.000	1.466	1.474	0.000	0.055	0.276	0.982	0.050	0.282	0.984
$\delta(0.3, 0.2, 1)$	0.976	0.981	0.000	1.000	1.005	0.000	0.035	0.227	0.974	0.031	0.241	0.966
$\delta(0.4, 0.3, 1)$	-0.285	0.291	0.348	-0.260	0.269	0.030	0.026	0.197	0.976	0.034	0.206	0.980
$\delta(0.5, 0.4, 1)$	-2.173	2.173	0.000	-2.148	2.148	0.000	-0.039	0.178	0.964	-0.040	0.199	0.966
$\delta(0.6, 0.5, 1)$	-2.259	2.259	0.000	-2.233	2.234	0.000	-0.135	0.221	0.932	-0.140	0.237	0.934
$\delta(0.7, 0.6, 1)$	0.150	0.166	0.922	0.176	0.192	0.212	0.155	0.251	0.912	0.168	0.268	0.910
$\delta(0.8, 0.7, 1)$	1.409	1.412	0.000	1.435	1.438	0.000	-0.015	0.211	0.988	-0.006	0.232	0.966
$\delta(0.9, 0.8, 1)$	1.612	1.617	0.000	1.638	1.644	0.000	-0.049	0.222	0.992	-0.053	0.236	0.984
$\delta(1, 0.9, 1)$	1.788	1.806	0.000	1.817	1.834	0.000	-0.238	0.437	0.972	-0.236	0.464	0.958

Table 11: Performance of Bayesian GPS estimator for treatment effects $\tau(g, V_g)$, with $g \in \{0, 0.1, \ldots, 0.9, 1\}$, and spillover effects $\delta(g, g-0.1, z, V_g \cup V_{g-0.1})$, with $g \in \{0.1, \ldots, 0.9, 1\}$ and $z \in \{0, 1\}$, under the scenario where the network is taken from the Add Health data and the outcome is drawn from a non-linear model with random effects.



(a) Linear DGP - Linear Model

(b) Linear DGP - Splines

(c) Non-Linear DGP - Linear Model

(d) Non-Linear DGP - Splines

Figure 6: Estimated and true ()ADRF $\mu(z, g; V_g)$ under the scenario where the network is taken from the Add Health data and the outcome is drawn from a linear model (top) or a non-linear model (bottom) with random effects. Average posterior mean (solid line) and 95 % credible intervals of $\mu(0, g; V_g)$ (blue) and $\mu(1, g; V_g)$ (red) of linear (left) and splines-based (right) estimator with random effects are represented with the true ADRFs (dotted line).

	Linear						Splines					
	RE			NO RE			RE			NO RE		
	Bias	RMSE	Coverage	Bias	RMSE	Coverage	Bias	RMSE	Coverage	Bias	RMSE	Coverage
$\tau(0, V_0)$	-0.026	0.148	0.970	-0.001	0.148	0.958	-0.008	0.067	0.972	-0.008	0.067	0.974
$\tau(0.1, V_{0.1})$	-0.026	0.149	0.968	-0.001	0.148	0.958	-0.008	0.067	0.980	-0.008	0.067	0.974
$\tau(0.2, V_{0.2})$	-0.026	0.149	0.966	-0.001	0.148	0.958	-0.008	0.067	0.976	-0.008	0.067	0.972
$\tau(0.3, V_{0.3})$	-0.026	0.148	0.974	-0.001	0.148	0.958	-0.008	0.067	0.972	-0.008	0.067	0.980
$\tau(0.4, V_{0.4})$	-0.026	0.149	0.968	-0.001	0.148	0.958	-0.008	0.067	0.978	-0.008	0.067	0.978
$\tau(0.5, V_{0.5})$	-0.026	0.148	0.972	-0.001	0.148	0.958	-0.008	0.067	0.974	-0.008	0.067	0.972
$\tau(0.6, V_{0.6})$	-0.026	0.149	0.972	-0.001	0.148	0.958	-0.008	0.067	0.980	-0.008	0.067	0.980
$\tau(0.7, V_{0.7})$	-0.026	0.149	0.970	-0.001	0.148	0.958	-0.008	0.067	0.972	-0.008	0.067	0.974
$\tau(0.8, V_{0.8})$	-0.026	0.149	0.972	-0.001	0.148	0.958	-0.008	0.067	0.972	-0.008	0.067	0.98’
$\tau(0.9, V_{0.9})$	-0.026	0.149	0.968	-0.001	0.148	0.958	-0.008	0.067	0.972	-0.008	0.067	0.980
$\tau(1, V_1)$	-0.026	0.149	0.964	-0.001	0.148	0.958	-0.008	0.067	0.970	-0.008	0.067	0.978
$\delta(0.1, 0, 0)$	1.465	1.473	0.000	1.477	1.485	0.000	0.359	0.436	0.874	0.359	0.435	0.858
$\delta(0.2, 0.1, 0)$	1.450	1.455	0.000	1.462	1.467	0.000	0.050	0.165	0.978	0.049	0.164	0.982
$\delta(0.3, 0.2, 0)$	0.986	0.988	0.000	0.998	1.001	0.000	0.027	0.128	0.968	0.028	0.129	0.968
$\delta(0.4, 0.3, 0)$	-0.274	0.278	0.068	-0.261	0.265	0.030	-0.017	0.113	0.964	-0.017	0.113	0.970
$\delta(0.5, 0.4, 0)$	-2.161	2.161	0.000	-2.148	2.148	0.000	0.017	0.088	0.982	0.016	0.088	0.980
$\delta(0.6, 0.5, 0)$	-2.246	2.247	0.000	-2.233	2.233	0.000	-0.076	0.116	0.924	-0.075	0.116	0.920
$\delta(0.7, 0.6, 0)$	0.164	0.175	0.730	0.177	0.187	0.212	0.081	0.135	0.912	0.081	0.135	0.912
$\delta(0.8, 0.7, 0)$	1.423	1.426	0.000	1.438	1.440	0.000	-0.043	0.130	0.950	-0.042	0.130	0.964
$\delta(0.9, 0.8, 0)$	1.628	1.632	0.000	1.642	1.646	0.000	-0.014	0.145	0.976	-0.014	0.145	0.972
$\delta(1, 0.9, 0)$	1.809	1.822	0.000	1.824	1.837	0.000	-0.283	0.372	0.882	-0.283	0.373	0.868
$\delta(0.1, 0, 1)$	1.465	1.473	0.000	1.477	1.485	0.000	0.359	0.436	0.874	0.359	0.435	0.858
$\delta(0.2, 0.1, 1)$	1.450	1.455	0.000	1.462	1.467	0.000	0.050	0.165	0.978	0.049	0.164	0.980
$\delta(0.3, 0.2, 1)$	0.986	0.988	0.000	0.998	1.001	0.000	0.027	0.128	0.968	0.028	0.129	0.974
$\delta(0.4, 0.3, 1)$	-0.274	0.278	0.068	-0.261	0.265	0.030	-0.017	0.113	0.964	-0.017	0.113	0.968
$\delta(0.5, 0.4, 1)$	-2.161	2.161	0.000	-2.148	2.148	0.000	0.017	0.088	0.982	0.016	0.088	0.976
$\delta(0.6, 0.5, 1)$	-2.246	2.247	0.000	-2.233	2.233	0.000	-0.076	0.116	0.924	-0.075	0.116	0.920
$\delta(0.7, 0.6, 1)$	0.164	0.175	0.730	0.177	0.187	0.212	0.081	0.135	0.912	0.081	0.135	0.912
$\delta(0.8, 0.7, 1)$	1.423	1.426	0.000	1.438	1.440	0.000	-0.043	0.130	0.950	-0.042	0.130	0.958
$\delta(0.9, 0.8, 1)$	1.628	1.632	0.000	1.642	1.646	0.000	-0.014	0.145	0.976	-0.014	0.145	0.972
$\delta(1, 0.9, 1)$	1.809	1.822	0.000	1.824	1.837	0.000	-0.283	0.372	0.882	-0.283	0.373	0.872

Table 12: Performance of Bayesian GPS estimator for treatment effects $\tau(g, V_g)$, with $g \in \{0, 0.1, \ldots, 0.9, 1\}$, and spillover effects $\delta(g, g-0.1, z, V_g \cup V_{g-0.1})$, with $g \in \{0.1, \ldots, 0.9, 1\}$ and $z \in \{0, 1\}$, under the scenario where the network is taken from the Add Health data and the outcome is drawn from a non-linear model without random effects.

6. Concluding Remarks

After reviewing the literature on causal inference for observational studies, we have discussed definition, identification and estimation issues of causal effects of interventions in contexts with interconnected and interfering units. Under the neighborhood interference and unconfoundedness of the individual and neighborhood treatment assumptions, we have developed a new covariate-adjustment estimator for treatment and spillover effects in observational studies on networks. Estimation is based on a generalized propensity score that balances individual and neighborhood covariates across units under different levels of individual treatment and of exposure to neighbors' treatment. Adjustment for this generalized propensity score is performed using a penalized spline regression which allows a non-linear relationship of the outcome with the neighborhood treatment, the individual and neighborhood propensity scores. Inference capitalizes on a three-step Bayesian procedure which allows taking into account the uncertainty in the propensity score estimation and avoiding model feedback. Finally, correlation of interacting units is taken into account using a community detection algorithm and incorporating random effects in the outcome model. We conducted a simulation study to assess, under different data generating scenarios, the performance in terms of bias, MSE, and coverage of our proposed estimator and its variants —fully model-based versus matching adjustment for individual propensity score, linear versus non-linear outcome model, with versus without community random intercept. The simulation study has shown promising performance of the proposed methods. The splines-based estimator is superior to the linear estimator when the outcome does not linearly depend on the two treatments and the two propensity scores. The drawback of the use of splines is the large performance deterioration in regions where there are few observations. This, however, is not only specific to use of splines but it is a common problem due to lack of overlap. The ability to estimate the entire average dose-response function and the spillover effects at different levels of the neighborhood treatment clearly depends on both the sparsity of the graph and the distribution of the individual treatment and in turn of the neighborhood treatment. A sparse graph with most units with low degree would result in a distribution of the neighborhood treatment concentrated around few values and the dose-response function would be poorly estimated for the other values. On the other hand a dense graph with many units with a very large degree would result in a very large range of the neighborhood treatment, with few observations for every value. In this case a discretization of the neighborhood treatment along with an approximation of the dose-response function might be required, with obvious consequences in terms of causal effects' interpretation.

A possible correlation between neighbors has been a major concern in the field of causal inference on networks. The inclusion of a random intercept defined on non-overlapping communities, identified using a community detection algorithm, allows us to capture an outcome correlation due to latent characteristics. The ability to estimate and take into account the outcome correlation with a good approximation depends on the graph structure and on the performance of the community detection algorithm and its effectiveness in identifying communities such that connections between the nodes are denser than connections with the rest of the network. In the simulation study we have defined random intercepts on communities originating directly from the generating model or pre-specified in the Add

Health data. We have then used the same known communities in the estimation stage. In this way, we were able to assess the performances of our proposed estimator besides the community identification. The use of community detection algorithms would affect the performance of our estimator depending on their specific performance. Moreover, the inclusion of community random intercepts would succeed in taking into account the presence of outcome correlation to the extent to which the identified communities match the correlation structure with good approximation.

Finally, in the appendix we have included an additional simulation to show the bias resulting from assuming no-interference and the extent to which the bias depends on the level of interference and the correlation between neighbors' treatments. We have also investigated the bias reduction yielded by the inclusion of a community random effect, even when interference is not taken into account. This result could be useful when network connections are not measured, but there is still some information on the network structure, at least on the presence of dense communities of individuals. These communities, even if not be fully connected, could still be used to partly correct for the bias due to neglecting interference by capturing and controlling for some treatment correlation. Investigating the extent to which this bias reduction depends on the network, interference and correlation structure as well as its empirical application is left to future work.

Acknowledgments

We would like to thank Guido W. Imbens, Donald B. Rubin, the Associate Editor, and two anonymous reviewers, for detailed comments that helped improve this paper. This work was supported, in part, by the National Science Foundation (awards CAREER IIS-1149662 and IIS-1409177), by the Office of Naval Research (awards YIP N00014-14-1-0485 and N00014-17-1-2131), and by Italian ministerial funds (Dipartimenti Eccellenti 2018-2022).

Appendix A. Data Generating Models

Below we detail alternative data generating processes.

A.1 Network Model: Stochastic Block Model

For each unit $i = 1, \ldots, 1000$ we first generated a continuous covariate X_{i1} and a binary covariate X_{i2} from the following distributions:

$$X_{i1} \sim \text{Gamma}(0.5, 1) \quad X_{i2} \sim \text{Ber}(0.5)$$

We then generated the network from a stochastic block model, where the probability of link only depends on the community membership. Let $\mathbf{M}$ be the stochastic block matrix, where each element M_{lk} is the probability of link between a unit of community i and a unit of community k. In a stochastic block model the probability of link between two units i and j with community memberships C_i and C_j is

$$Pr(A_{ij} = 1 | C_i, C_j) = M_{C_i C_j}$$

In our simulation study we created 100 communities of 10 units each and we set the diagonal elements of $\mathbf{M}$ to 0.08 and the off-diagonal elements to 0.02.

A.2 Network Model: Latent Cluster Model

In scenarios with the network drawn from a latent cluster model, we first sampled covariates X_{i1} and X_{i2} as in the scenarios based on the stochastic block model. We created again 100 communities of 10 units each, with C_i being the community membership indicator. The adjacency matrix is then generated using the following model:

$$Pr(A_{ij} = 1|\mathbf{X}_i, \mathbf{X}_j, C_i, C_j) = \exp\big(\beta_0 + \boldsymbol{\beta}_X^T|\mathbf{X}_i - \mathbf{X}_j| + \beta_C|C_i - C_j|\big)$$

where $\beta_0 = -4.6$, $\boldsymbol{\beta}_X = [0.05, 0.005]^T$, and $\beta_C = 0.18$. The first term gives rise to homophily along the observed characteristics, that is, nodes with similar characteristics are more likely to be connected. The second term creates a cluster structure in the network, which can be interpreted as some homophily due to latent characteristics shared by group of units.

A.3 Network Model: Add Health Data

The Add Health Study is a large survey of adolescents attending schools in the United States who were listed on 7th through 12th grade enrollment rosters during the 1994 -1995 academic year. The Wave 1 survey collected data on socio-demographic characteristics of the respondents, education and occupation of parents, household structure, risk behaviors, health status, and friendships. As part of the in-home survey, students selected their five best male and five best female friends from a complete school roster. We used data on this friendship network together with 3 characteristics: sex, grade, race. We selected a sample of 1000 students in 8 schools. We used the directed friendship graph, meaning that only the students who were nominated by student i were considered his friends. Units with degree equal to 0 were not excluded a priori. They were in fact considered as alters of a student i who nominated them and their treatment and characteristics were used to compute the neighborhood treatment and neighborhood characteristics of student i.

In the simulation study communities are defined as clusters of students in the same school and same grade.

	Individual Characteristics		Neighborhood Characteristics	
	Mean	SD	Mean	SD
Sex	0.49	0.50	0.46	0.30
Race	0.71	0.46	0.65	0.42
Grade	9.46	1.68	8.60	3.13
Degree	5.07	2.92	·	·

Table 13: Summary Statistics of the Add Health data set.

A.4 Individual Treatment Model

In the scenarios based on the stochastic block model and the latent cluster model, the individual treatment Z_i was drawn for each unit using the following individual propensity score:

$$\text{logit}(Pr(Z_i = 1|\mathbf{X}_i)) = 2.6X_{1i} - 2.2X_{2i} \tag{23}$$

In the simulations based on the Add Health data we used all three characteristics and generated the individual treatment based on the following propensity score:

$$\text{logit}(Pr(Z_i = 1|\mathbf{X}_i)) = 0.7\text{sex}_i - 0.11\text{grade}_i + \text{race}_i \tag{24}$$

The neighborhood treatment G_i was then computed as the proportion of treated neighbors. The propensity scores in Equations (23) (24) applied on the 3 network structures, with degree distribution represented in Figure 7, has given rise to the distributions of the neighborhood treatment G_i shown in Figure 8.

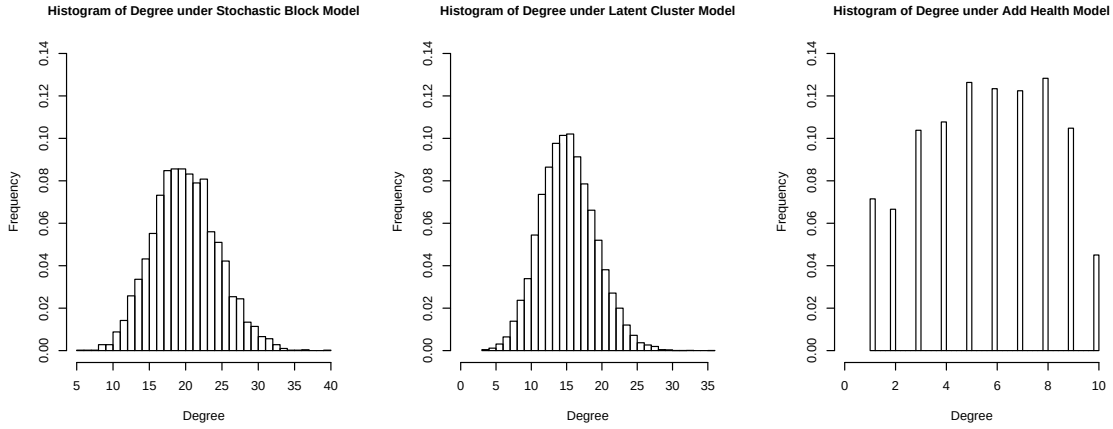


Figure 7: Histograms of the degree N_i in the 3 network types.

The neighborhood propensity score model was estimated using a binomial regression and the neighborhood propensity score $\Lambda_i = \lambda(G_i; Z_i; \mathbf{X}_i^g)$ was then predicted for each unit corresponding to the actual G_i affecting each unit, the individual treatment Z_i, the individual and neighborhood characteristics.

A.5 Potential Outcome Models

In scenarios where the outcome was drawn from a linear model, we used the following distribution:

$$Y_i \sim \mathcal{N}(-3 + 2Z_i + 4G_i - \Phi_i - 2\log(\Lambda_i) + u_j, 1) \tag{25}$$

with $u_j \sim \mathcal{N}(0, \Sigma_u)$. In scenarios with random effects $\Sigma_u = 2I$, whereas is scenarios without random effects $\Sigma_u = 0I$.

In scenarios where the outcome was drawn from a non-linear model, we instead used the following distribution:

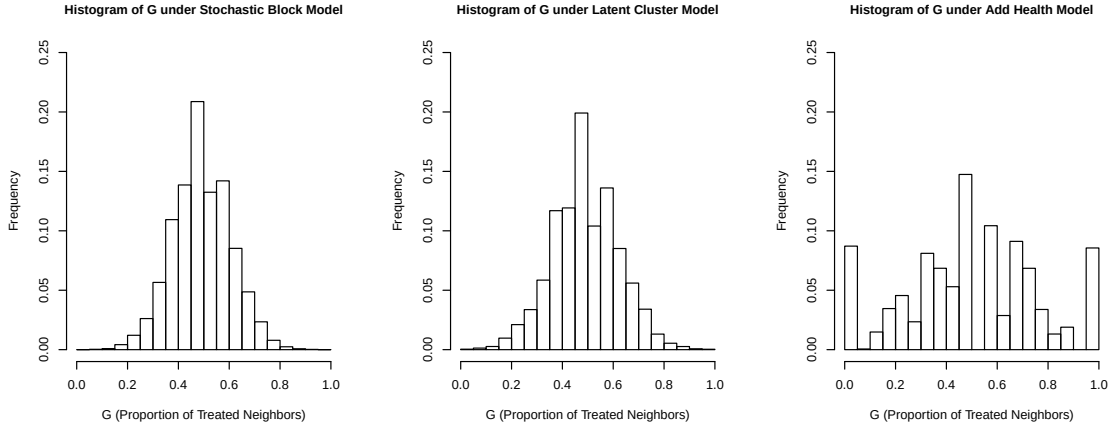


Figure 8: Histograms of the neighborhood treatment G_i in the 3 network types.

$$Y_i \sim \mathcal{N}(3 + 2Z_i + 25\,\text{sigmoid}\,(10\exp\,(-\frac{(G_i-1)^2}{0.12})) - 2\Phi_i - 2.5\Lambda_i + u_j, 1) \qquad (26)$$

where $\text{sigmoid}(x) = \frac{1}{1+\text{e}^{-x}}$ and again $u_j \sim \mathcal{N}(0, \Sigma_u)$ with $\Sigma_u = 2I$ or $\Sigma_u = 0I$.

Appendix B. Bias of the Treatment Effect When Neglecting Interference

We add here an additional simulation study to show the importance of taking interference into account even when the estimand of interest is the direct effect of the treatment. In Forastiere et al. (2020) we have derived bias formulas for the overall treatment effect, that is, the treatment effect averages over the observed distribution of the neighborhood treatment, when interference is neglected. Here we compare the average bias of a propensity score-based estimator for the treatment effects that neglects the presence of interference and with the bias of our Bayesian generalized propensity score-based estimator. An important result in Forastiere et al. (2020) is that the bias depends on the correlation between the individual treatment and the neighborhood treatment in addition to the level of interference. To show how these two factors play a role here in our empirical comparison, we simulated 9 scenarios with a different level of interference (low, medium, high) and an increasing correlation between Z and G (0.2, 0.5, 0.8). To generate this correlation we used an iterative approach where, after an initial draw of the individual treatments given the individual propensity scores, at each iteration for each unit the neighborhood treatment is computed as the proportion of treated neighbors and the individual treatment is drawn again from the initial individual propensity score with an additional dependence on the neighborhood treatment. This iterative approach replicates a realistic situation where the dependence between a unit's treatment and the treatment status of their neighbors is due to peer influence. We repeated this iterative generation of the individual and neighborhood treatment for each of the 500 simulations. The outcome was drawn from from a normal distribution with the predictor $q^{(Y)}(\cdot)$ being a linear function of the individual treatment, the neighborhood treatment,

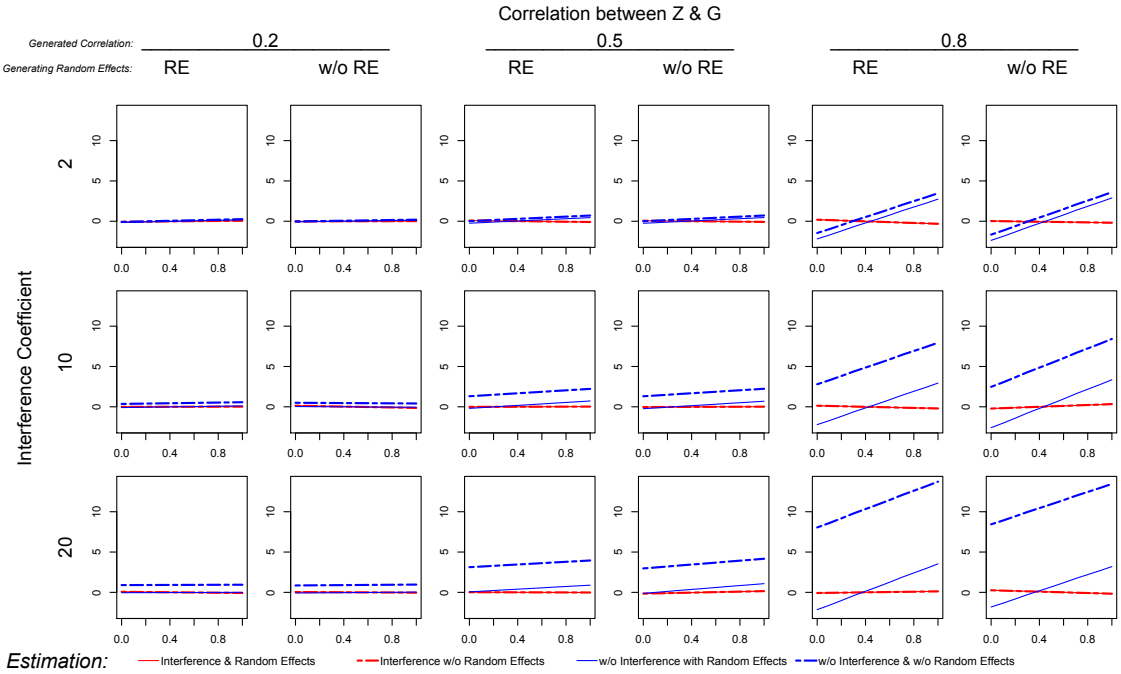


Figure 9: Average bias of the treatment effects $\tau(g, V_g)$ when including or neglecting interference for different levels of interference and correlation between Z and G.

the individual propensity score and neighborhood propensity score with and without the inclusion of a community random intercept:

$$Y_i \sim \mathcal{N}(-3 + 2Z_i + \delta G_i - \Phi_i - 2\log(\Lambda_i) + \gamma u_j, 1) \tag{27}$$

with $u_j \sim \mathcal{N}(0, \Sigma_u)$, and $\delta \in \{2, 10, 20\}$ and $\gamma \in \{0, 1\}$ defining the level of interference and the presence of the random intercept. All simulations were conducted using the friendship network of the 1000 students extracted from the Add Health data set (Harris & Udry, 2017) together with their covariates.

In all 9x2 scenarios the treatment effects $\tau(g, V_g)$ were estimated using i) the Bayesian generalized propensity score-based estimator proposed here for network interference, which included in the outcome model the individual and neighborhood treatments as well as the individual and neighborhood propensity scores, and ii) a Bayesian propensity score-based estimator neglecting interference, which included in the outcome model the individual treatment and the individual propensity score only. Furthermore, we investigated the inclusion of a community random effect in both estimators.

In Figure 9 we report the average bias for both estimators with and without random effects for the 9x2 scenarios defined by the level of correlation and interference and the presence of the random effect in the generation model. We can see that the bias of the estimator that neglects interference increases with both the level of correlation between the two treatments and with the level of interference. In addition, this estimator estimates a constant treatment effect $\tau(g, V_g)$ for all values of g since it does not include the neighbor-

hood treatment in the model. This results in a non-constant bias which, in this setting, increases with g.

Nevertheless, an important result to notice is the decrease of the bias when the community random intercept is included, regardless of the data generating process. Indeed, the neighborhood treatment of units belonging to a densely connected community tend to be correlated, and even more so when there is a correlation between a unit's individual treatment and that of their neighbors. Therefore, the inclusion of a community random effect is able to reduce the bias of the treatment effect due to the omission of the neighborhood treatment. Clearly, the bias reduction cannot retrieve the true treatment effects $\tau(g, V_g)$ because the estimated treatment effect will still be constant across all values of g, resulting in a non-constant bias. However, the overall bias of the treatment effect estimated neglecting interference with respect to the average treatment effect ($\int \tau(g, V_g)p(g)dg$) is smaller when a community random effect is included. This bias reduction is similar to the one yielded by a cluster random effect when cluster-level confounders are not measured or omitted. Indeed, neighborhood interference could be seen as creating a neighborhood-level confounder for the treatment effect. However, as opposed to the cluster random effects to account for cluster-level confounders, here communities included several overlapping neighborhoods, perhaps making the bias reduction less intuitive and more structure dependent. Further investigation is required to understand the extent to which this bias reduction depends on the community density and size and the average neighborhood size, the treatment correlation among neighbors, the direction of the bias, and the presence and level of an underlying outcome correlation.

References

Abadie, A. & Imbens, G.W.(2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, 74, 235–267.

Abadie, A. & Imbens, G.W.(2008). On the Failure of the Bootstrap for Matching Estimators. *Econometrica*, 76, 1537–1557.

Abadie, A. & Imbens, G.W.(2009). Matching on the Estimated Propensity Score. *NBER Working Papers 15301, National Bureau of Economic Research, Inc.*.

An, W. (2010). Bayesian propensity score estimators: incorporating uncertainties in propensity scores into causal inference. *Sociological Methodology*, 40, 151–189.

Angelucci, M. & De Giorgi, G.(2009). Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles' Consumption?. *American Economic Review*, 99, 486–508.

Aronow, P. M. (2012). A General Method for Detecting Interference Between Units in Randomized Experiments. *Sociological Methods & Research*, 41(3), 3–16.

Aronow, P. M. & Samii, C.(2017). Estimating Average Causal Effects Under General Interference. *Forthcoming. Annals of Applied Statistics. Preprint:arxiv:1305.6156.*

Athey, S., Eckles, D., & Imbens, G.W. (2015). Exact p-values for network interference. *NBER Working Paper 21313.*

Austin, P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research*, 46(3),

399–424.

Basse, G. W. & Airoldi, E.M.(2018a). Model-assisted design of experiments in the presence of network-correlated outcomes. *Biometrika*, 105(4), 849–858.

Basse, G. W. & Airoldi, E.M.(2018b). Limitations of design-based causal inference and A/B testing under arbitrary and network interference. *Sociological Methodology*, 48(1), 136–151.

Bia, M.& Mattei, A. (2008). A Stata package for the estimation of the dose response function through adjustment for the generalized propensity score. *Stata Journal, StataCorp LP*, 8(3), 354–373.

Bia, M., Flores, C.A., Flores-Lagunes, A., & Mattei, A. (2014). A Stata package for the application of semiparametric estimators of dose response functions. *Stata Journal, StataCorp LP*, 14(3), 580–604.

Bowers, J., Fredrickson, M.M., & Panagopoulos, C. (2013). Reasoning about interference between units: a general framework. *Political Analysis*, 21(1), 97–124.

Cai, C., Pouget-Abadie, J. & Airoldi, E.M. (2022). Optimizing randomized and deterministic saturation designs under interference. *arXiv* no. 2203.09682.

Cox, D. R. (1958). Planning of Experiments. *Wiley; New York.*

Dehejia, R.H., & Wahba, S. (1999). Causal effects in nonexperimental studies: Re-evaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94, 1053–62..

Doudchenko, N., Minzhengxiong, Z., Drynkin, D., Airoldi, E.M., Mirrokni, V. & Pouget-Abadie, J.(2020). Causal inference with bipartite designs. *arXiv* no. 2010.02108.

Eckles, D., Karrer, B., & Ugander, J. (2014). Design and analysis of experiments in networks: Reducing bias from interference. *arxiv:1404.7530.*

Flores, C. A., Flores-Lagunes A, Gonzales A., & Neuman T (2012). Estimating the Effects of Length of Exposure to Instruction in a Training Program: The Case of Job Corps. *The Review of Economics and Statistics.* 94. 153–171.

Forastiere, L., Mealli, F. & VanderWeele, T. (2017). Identification and Estimation of Causal Mechanisms in Clustered Encouragement Designs: Disentangling Bed Nets Using Bayesian Principal Stratification. *Journal of the American Statistical Association*, 111:514, 510–525.

Forastiere, L., Airoldi, E.M., & Mealli, F. (2020). Identification and estimation of treatment and interference effects in observational studies on networks. Forthcoming in the *Journal of the American Statistical Association.* Preprint at: *arxiv:1609.06245.*

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486, 75–174.

Gelfand, A.E., & Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398–409.

Gelman, A., Meng, X.,& Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6, 733–807.

Hahn, J. (1998). On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects. *Econometrica*, 66, 315–331.

Handcock, M. S. & Raftery, A. E.(2007). Model-based clustering for social networks. *Journal of the Royal Statistical Society: Series A*, 7,301–354.

Hansen, B. B. (2004). Full matching in an observational study of coaching for the SAT.. *Journal of the American Statistical Association*, 99,609–618.

Harris, K.M., & Udry, J. R. (2017). National Longitudinal Study of Adolescent to Adult Health (Add Health), 1994-2008 [Public Use]. Ann Arbor, MI: Carolina Population Center, University of North Carolina-Chapel Hill [distributor], Inter-university Consortium for Political and Social Research [distributor], 2017-10-24. https://doi.org/10.3886/ICPSR21600.v18.

Hernán, M. A., & Robins, J. M. (2018). Causal Inference. *Boca Raton: Chapman & Hall/CRC, forthcoming.*

Hill, J. & Reiter, J.(2006). Interval estimation for treatment effects using propensity score matching. *Statistics in Medicine*, 25, 2230–2256.

Hill, J., Reiter, J., &Zanutto, E. A. (2004). A comparison of experimental and observational data analyses. *T In: Gelman A, Meng X-L, editors. Applied Bayesian Modeling and Causal Inference from an Incomplete-Data Perspective. John Wiley & Sons, Ltd; 2004.*.

Hirano, K., & Imbens, G. W. (2004). The propensity score with continuous treatments. *In Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives, ed. A. Gelman and X.-L. Meng, 73–84. West Sussex, England: Wiley InterScience.*

Hirano, K., Imbens, G. W. & Ridder, G.(2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71, 1161–1189.

Holland, P. W. (1986). Statistics and causal inference (with discussion). *Journal of the American Statistical Association*, 81, 945–970.

Holland, P. W., Laskey, K. & Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks*, 5(2), 109–137.

Hong, G. & Raudenbush, S. W. (2006). Evaluating Kindergarten Retention Policy: A Case Study of Causal Inference for Multilevel Observational Data. *Journal of the American Statistical Association*, 101, 901–910.

Hoshino, T. (2008). A (b)ayesian propensity score adjustment for latent variable modeling and MCMC algorithm. *Computational Statistics & Data Analysis*, 52, 1413–1429.

Hudgens, M. G. & Halloran, M. E. (2008). Towards causal inference with interference. *Journal of the American Statistical Association*, 103, 832–842.

Imai, K., & Van Dyk, D.A. (2004). Causal treatment with general treatment regimes: Generalizing the Propensity Score. *Journal of the American Statistical Association.* 99: 854–866.

Imbens, G. W. (2000). The Role of the Propensity Score in Estimating Dose-Response Functions. *Biometrika*, 87, 706–710.

Imbens, G. W., & Rubin, D. B. (2015). Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. *Cambridge University Press New York, NY, USA, 2015.*

Kaplan, D. & Chen, J. (2012). A two-step bayesian approach for propensity score analysis: Simulations and case study. *Psychometrika*, 77, 581–609.

Karwa, V. & Airoldi, E.M.(2018). A systematic investigation of classical causal inference strategies under mis-specification due to network interference. *arXiv*, no. 1810.08259.

Kim, D.A., Hwong, A.R., Stafford, D., Hughes, D.A., O'Malley, A.J., Fowler, J.H., & Christakis, N.A. , & Zhao, Z.(2012). Social network targeting to maximise population behaviour change: a cluster randomised controlled trial. *Lancet*, 386(9989):145–153.

Kluve, J., Schneider, H., & Uhlendor, A. , & Zhao, Z.(2012). Evaluating continuous training programs using the Generalized Propensity Score. *BiomeJournal of the Royal Statistical Society, Series Atrika*, 175, 587–617.

Lancichinetti, A. & Fortunato, S. (2009). Community detection algorithms: A comparative analysis . *Physical Review E*, 80(5), 1–11.

Lechner, M. (2001). Identication and estimation of causal efects of multiple treatments under the conditional independence assumption. *Econometric Evaluation of Labour Market Policies, Heidelberg: Physica, M. Lechner, F.Pffeier (Eds.)*. 43–58.

Lee Yu, C., Airoldi, E.M., Borgs, C. & Chayes, J.T. (2022). Estimating Total Treatment Effect in randomized experiments with unknown network structure. *Proceedings of the National Academy of Sciences*, In press.

Liu, F., Bayarri, M.J., & Berger, J.O. (2009). Modularization in bayesian analysis, with emphasis on analysis of computer models. *Bayesian Analysis*, 4, 119–150.

Liu, L., Hudgens, M.G., & Becker-Dreps, S. (2016). On inverse probability-weighted estimators in the presence of interference. *Biometrika*, 103, 829–842.

Lunn, D., Best, N., Spiegelhalters, D., Graham, G., & Neuenschwander, B. (2009). Combining MCMC with 'sequential' PKPD modelling. *Journal of Pharmacokinetics and Pharmacodynamics*, 36, 19–38.

McCandless, L.C., Douglas, I.J., Evans, S.J. & Smeeth, L. (2010). Cutting feedback in bayesian regression adjustment for the propensity score. *The international journal of biostatistics*, 6, 122.

McCandless, L.C., Gustafson, P., & Austin, P.C. (2009). Bayesian propensity score analysis for observational data. *Statistics in Medicine*, 28, 94–112.

McCandless, L.C., Richardson, S., & Best, N. (2012). Adjustment for missing confounders using external validation data and propensity scores. *Journal of the American Statistical Association*, 107, 40–51.

McElreath, R. (2016). Statistical Rethinking: A Bayesian Course with Examples in R and Stan. *Chapman & Hall/CRC Press*.

Miguel, E. & Kremer, M. (2004). Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1), 159–217.

Neyman J. (1923). On the application of probability theory to agricultural experiments: Essay on Principles Section 9. Masters Thesis. *Portions translated into English by D. Dabrowska and T. Speed (1990)* Statistical Science, 5, 465–472.

Ogburn, E.L., Sofrygin, O., Diaz, I. & van der Laan, M.J. (2017). Causal inference for social network data. arXiv:1705.08527.

Ogburn, E.L., Shpitser, I., Lee, Y. & van der Laan, M.J. (2018). Causal inference, social networks, and chain graphs. arXiv:1812.04990.

Perez-Heydrich, C., Hudgens, M.G., Halloran, M.E., Clemens, J.D., Ali, M. & Emch, M.E. (2014). Assessing effects of cholera vaccination in the presence of interference. *Biometrics*, 70(3), 731–44.

Pouget-Abadie, J., Parkes, D.C., Mirrokni, V. & Airoldi, E.M. (2018). Optimizing cluster-based randomized experiments under a monotonicity assumption. *International Conference on Knowledge Discovery and Data Mining (KDD)*.

Rosenbaum, P. R. (1987). Model-Based Direct Adjustment. *Journal of the American Statistical Association*, 82, 387–394.

Rosenbaum, P. R. (2002). Observational studies. *New York, NY: Springer-Verlag*.

Rosenbaum, P. R. (2007). Interference Between Units in Randomized Experiments. *Journal of the American Statistical Association*, 102(477), 191–200.

Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41–55.

Rubin, B. D. (1974). Estimating causal effects of treatments in randomized and non randomized studies. *Journal of Educational Psychology* 66, 688–701.

Rubin, B. D. (1978a). Bayesian inference for causal effects. *Annals of Statistics*, 6, 34–58.

Rubin, D. B. (1980). Comment on "Randomization Analysis of Experimental Data in the Fisher Randomization Tes" by D. Basu. *Journal of the American Statistical Association*, 75, 591–593.

Rubin, D. B. (1985). The use of propensity scores in applied Bayesian inference. *In Bayesian Statistics 2, Bernardo JM, De Groot MH, Lindley DV, Smith AFM (eds). Valencia University Press, North-Holland: Amsterdam*, 63–72.

Rubin, D. B. (1986). Which Ifs have Causal Answers? Comment on "Statistics and Causal Inference" by P. Holland. *Journal of the American Statistical Association*, 81, 961–962.

Rubin, D. B. (2007). The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Statistics in Medicine*, 26, 20–36.

Rubin, D. B. (2008). For objective causal inference, design trumps analysis. *The Annals of Applied Statistics*, 2, 808–840.

Rubin, D. B. & Thomas, N. (1996). Matching using estimated propensity scores, relating theory to practice. *Biometrics*, 52, 249–264.

Ruppert, D., Wand, M.P. & Carrol, R.J. (2003). Semiparametric Regression. *Cambridge: Cambridge University Press*.

Schaeffer, S. E. (2007). Graph clustering. *Computer Science Review* 1, 27–64.

Sofrygin, O. & van der Laan, M. (2017). Semi-Parametric Estimation and Inference for the Mean Outcome of the Single Time-Point Intervention in a Causally Connected Population. *Journal of Causal Inference*, 5(1).

Stuart, E. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science* 25, 1–21.

Sussman, D.L. & Airoldi, E.M.(2018). Elements of estimation theory for causal effects in the presence of network interference. *arXiv*, no. 1702.03578.

Ugander, J., Karrer, B., Backstrom, L. & Kleinberg, J. (2013). Graph Cluster Randomization: Network Exposure to Multiple Universes. *Proc. 19th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining (KDD)*.

Tchetgen Tchetgen, E. J. & VanderWeele, T. J. (2012). On causal inference in the presence of interference. *Statistical Methods in Medical Research*, 21, 55–75.

Tchetgen Tchetgen, E. J., Fulcher, I & Shpitser, I. (2012). Auto-G-Computation of Causal Effects on a Network. arXiv:1709.01577.

Toulis, P. & Kao, E.K. (2013). Estimation of Causal Peer Influence Effects. *ICML*, (3) 2013: 1489–1497.

Toulis, P., Volfovsky, A. & Airoldi, E.M. (2018). Estimating causal effects when treatments are entangled by a network of units. *arXiv* no. 1801.07310.

Van der Laan, M.J. (2014). Causal Inference for a Population of Causally Connected Units. *Journal of Causal Inference*, 0, 2193–3677.

Zigler, C.M., &, Dominici, F. (2014). Uncertainty in propensity score estimation: Bayesian methods for variable selection and model averaged causal effects. *Journal of the American Statistical Association*, 109, 95–107.

Zigler, C.M., Watt, K., Yeh, R. W., , Wang, Y., , Coull, B. A., &, Dominici, F. (2013). Model feedback in {b}ayesian propensity score estimation. *Biometrics*, 69, 263–273.

Wood, S.N. (2003). Thin plate regression splines.. *Journal of the Royal Statistical Society: Series B - Statistical Methodology*, 65, 95.

Wood, S.N. (2017). Generalized Additive Models: An Introduction with R (2nd edition). *Chapman and Hall/CRC.*