IMT School for Advanced Studies, Lucca
Lucca, Italy

Essays on Causal Inference and Complex Networks

PhD Program in Institution, Markets and Technologies
Curriculum in Economics, Management and Data Science

XXXII Cycle

By

Costanza Tortú

2020

The dissertation of Costanza Tortú is approved.

PhD Program Coordinator: Massimo Riccaboni, IMT School for Advanced Studies Lucca

Advisor: Prof. Fabrizia Mealli, University of Florence

Co-Advisor: Prof. Irene Crimaldi, IMT School for Advanced Studies Lucca

Co-Advisor: Prof. Laura Forastiere, Yale University

The dissertation of Costanza Tortú has been reviewed by:

Prof. Tiziano Arduini, University of Bologna

Prof. Bruno Arpino, University of Florence

Prof. Fredrik Sävie, Yale University

IMT School for Advanced Studies Lucca
2020

To Clara

Contents

List of Figures

List of Tables

Vita

June 25, 1992	Born, Florence (FI), Italy
Sep 2011-Oct 2014	B.Sc. in Statistics Final mark: 110/110 cum laude University of Florence, Italy
Sep 2014-Oct 2016	M.Sc. in Statistics, Actuarial and Financial Sciences Final mark: 110/110 cum laude University of Florence, Italy
Oct 2016-Oct 2020	Ph.D in Economics, Management and Data Science IMT School for Advanced Studies Lucca, Italy
Feb 2019-May 2019	Visiting Postgraduate Student Department of Political Science, Yale Institute for Network Science Yale University, United States of America
Sep 2019-Dec 2019	Visiting Postgraduate Student Business and Economics Department University of Naples Parthenope, Italy

Publications

1. I. Crimaldi, L. Forastiere, F. Mealli, C. Tortú. "The Causal Effect of Immigration Policy on Income Inequality ," Forthcoming in *Proceedings of the conference of the Italian Statistical Society* (SIS 2020), ISBN=9788891910776.
2. C. Tortú., I. Crimaldi, F. Mealli, L. Forastiere. "Modelling Network Interference with Multi-valued Treatments: the Causal Effect of Immigration Policy on Crime Rates,". Under review to *Sociological Methods and Research*. arXiv, preprint arXiv: 2003.10525v3, .
3. F.J. Bargagli-Stoffi., C. Tortú., L. Forastiere. "Heterogeneous Treatment and Spillover Effects under Clustered Network Interference,". Under review to *Annals of Applied Statistics* arXiv, preprint arXiv:2008.00707v2.
4. C. Tortú., L. Forastiere., V. Leone Sciabolazza., D. Del Prete "Dyadic Treatment Effect on Network Formation using Multi-valued Propensity Score Matching: Lobbying Activities and Legislative Collaborations,". *Forthcoming* in arXiv, preprint arXiv:.
5. C. Tortú., I. Crimaldi, F. Mealli, L. Forastiere. "Causal Effects with Hidden Treatment Diffusion over Partially Unobserved Networks,". *Forthcoming* in arXiv, preprint arXiv: .

Presentations

1. C. Tortú., F.J Bargagli-Stoffi, "Heterogeneous Treatment and Spillover Effects under Clustered Network Interference,", web seminar at the reading group *Machine Learning and Networks in Economics(IMT School For Advanced Studies Lucca)*, Lucca, Italy, 2020.
2. C. Tortú., "Heterogeneous Treatment and Spillover Effects under Clustered Network Interference,", web seminar at *Missing Values and Causality Research Group (École Polytechnique)*, Paris, France, 2020.
3. C. Tortú., "Heterogeneous Treatment and Spillover Effects under Clustered Network Interference," poster presentation at *Causal Machine Learning Workshop*, St. Gallen, Switzerland, 2020.
4. C. Tortú., "Machine Learning for Heterogeneous Treatment and Spillover Effects under Network Interference", oral presentation at *IMT Research Symposium*, Lucca, Italy, 2020.
5. C. Tortú., "Modelling Network Interference with Multi-valued Treatments: the Causal Effect of Immigration Policy on Crime Rates," oral presentation at *Inclusive and Sustainable Development Workshop*, Naples, Italy, 2019.
6. C. Tortú., "Modelling Network Interference with Multi-valued Treatments: the Causal Effect of Immigration Policy on Crime Rates," oral presentation at *ARS '19 International Workshop on Social Network Analysis*, Vietri sul Mare, Italy, 2019.
7. C. Tortú., "Modelling Network Interference with Multi-valued Treatments: the Causal Effect of Immigration Policy on Crime Rates," oral presentation at *COSTNET19 (Statistics of Network Data) Conference*, Bilbao, Spain, 2019.
8. C. Tortú., "Modelling Network Interference with Multi-valued Treatments: the Causal Effect of Immigration Policy on Crime Rates", poster presentation at *IMT Research Symposium*, Lucca, Italy, 2019.

Abstract

This dissertation is a collection of articles that develop statistical methods for performing causal inference on network data. In bridging these two themes, causal inference and complex networks, the thesis develops four complementary methodological contributions in two main settings that often arise in network data: (i) both the treatment and the outcome are measured at the individual level but the treatment spills over through the network connections; (ii) both the treatment and outcomes are measured at dyadic level. In the first setting, it elaborates innovative techniques for assessing the direct and spillover effects of an intervention in a population of connected units, where the potential outcome of an agent is affected by the treatment status of other interfering agents. In particular, the articles featured in the dissertation expand the existing literature by developing methods that are useful for (i) estimating the effect of an observational multi-valued intervention in a sample of units connected through a weighted network; (ii) detecting and estimating heterogeneous treatment and spillover effects in presence of units who belong to exogenous clusters, and whose interactions are described by cluster-specific networks; (iii) accounting for hidden treatment diffusion processes in a partially unobserved network. In the second setting, the dissertation employs the potential outcomes framework to analyze causal relationships in network formation processes. Specifically, it develops an estimator for the causal effect that the existence of links in a "treatment network" has on the formation of links on an "outcome network," with both networks being directed.

Chapter 1

Introduction

1.1 Motivation

Causal inference methods are widely used in the statistical literature, as they provide frameworks and tools for assessing whether two variables are linked by a cause-and-effect relationship. In recent years, the increase in computing power has enabled researchers to manage big sources of data, which often provide information about complex patterns of interaction between agents. For this reason, a growing community of statisticians, economists and physicists have started to develop novel techniques for handling these types of complex network data. In particular, statisticians working on causal inference methods have been debating about how to extract causal information in the presence of connected units and about how to understand the causal factors driving the formation of interactions between units. The present dissertation intends to contribute to these growing streams of literature, by proposing novel methodologies for expanding the existing statistical tools towards a broader ensemble of empirical applications.

1.2 Introduction on Causal Inference and Complex Networks

The purpose of causal inference methods is to evaluate the effect that, in a given population, results from some intervention. An issue with these methods is that they are yet unsuited to handle populations whose units interact with one another. Relationships of this sort are usually depicted and analyzed through complex networks. In this section, I provide an introduction to the causal inference framework. In addition, I discuss some basic conceptual aspects about complex networks, as well as some motivating examples of real-world phenomena where units are likely connected to one another. With regard to both topics, this preliminary introduction focuses especially on those aspects that are expanded at length in the later chapters of this dissertation.

1.2.1 Causal Inference: Idea and Setting

In scientific research, a large proportion of analyses faces questions about causality (Dempster, 1990; G. W. Imbens & Rubin, 2015). The objective of these studies is to assess the existence of a cause-effect relationship between two phenomena which can be measured in terms of statistical variables. More technically, the general purpose of any causal inference examination is to prove the validity of a causal statement, which presumes the existence of a causal relationship between a dependent variable and an independent variable, which is suspected to explain a relevant amount of the dependent variable's variation (G. W. Imbens & Rubin, 2015; Pearl et al., 2009; Wiedermann & Von Eye, 2016). Causal statements are the basis of any scientific breakthrough in many empirical fields of science. For instance, epidemiologists may want to understand whether a new drug is effective for a given disease, or whether an experimental vaccine is safe and valid at protecting individuals from an infection (Halloran & Struchiner, 1995; Hirano et al., 2000; Kleinberg & Hripcsak, 2011; Muthén & Brown, 2009; VanderWeele & Tchetgen, 2011); economists may want to evaluate the effects of incentives on the performance of specific firms. (Arpino & Mattei, 2013; Bandiera et al., 2011; Buss, 2001; Hoover, 2006, 2012); finally, political scientists may want to assess whether a specific policy leads to the expected social benefits or whether a political phenomenon bears positive changes on societies (Bove & Böhmelt, 2016; Gilligan, Sergenti, et al., 2008; Keele et al., 2013; A. K. Mayer, 2011).

The existence of a cause-effect relationship between two events is usually assessed by studying their joint statistical variation (Wiedermann & Von Eye, 2016). However, one must be careful: this naive approach might lead to misleading conclusions (Goldthorpe, 2001; Holland, 1986). Suppose for instance that a policy maker is examining the effect of a given intervention Z upon an outcome variable Y in some reference population. Figure 1 represents the bi-dimensional scatter plot about these two variables. The graph shows a positive association between the two variables, in the sense that an increase in the "independent" variable Z

corresponds with an increase in the "dependent" variable Y, and vice versa.

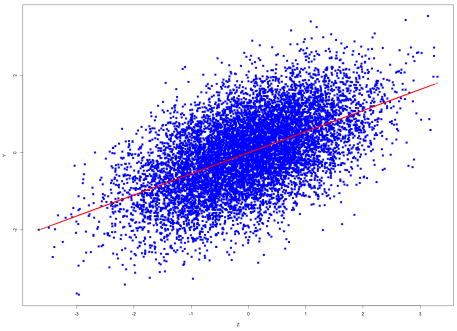


Figure 1: Example: positive association between Z and Y. The blue dots represent the reference population of units, while the red straight line depicts the linear fit between Y on Z obtained by the Ordinary Least Squares method.

However, the presence of an association of this sort does not necessarily imply that the two variables are actually linked by a cause-effect relationship (Grotzer, 2012). The positive association in question might have been generated by a variety of either causal or spurious mechanisms, which for the sake of illustration are graphically depicted in Figure 2. As one can easily observe, the two variables may show a significant association because either (i) Z causes Y (direct causation), (ii) Y causes Z (reverse causation), (iii) Z causes Y and Y causes Z (simultaneous causation), (iv) they are statistically associated due to coincidence (pure spurious association), (v) they are both caused by an additional factor, which is named *confounding variable*. Consequently, the empirical evidence displayed in Figure 1 is not sufficient to conclude that the intervention Z has for sure a positive causal effect on the outcome Y.

It is worth to expand upon some of the mechanisms that might induce to mistake a statistical association between two variables for a cause-effect relationship. Reverse causality issues commonly arise when the original intervention and its effect are confused: in this case, the two variables are still linked by a causal relationship, but one whose direc-

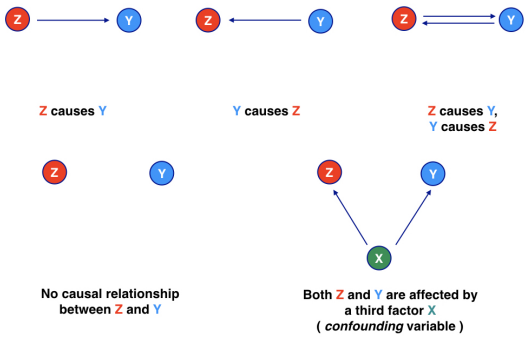


Figure 2: Causal or spurious relationships between Z and Y

tion is opposite to that which was originally hypothesized (Kramer et al., 2011; Sattar & Preiss, 2017). For instance, epidemiologists have discovered a negative causal relationship between breastfeeding and infant size, stating that newborns who have not received extended breastfeeding grow more rapidly during their first year of life than those who do. However, it has been recently verified that the true causal link among these two phenomena goes actually in the opposite direction: infants of slower growth are typically in need of prolonged breastfeeding period (Kramer et al., 2011). The economic relationship between health status and employment status is a solid example of reverse causality: while on the one hand an individual who suffers from diseases is less likely to take up a high-status occupation, on the other hand a poorly paid job may be at the onset of mental disorders or even health issues. Finally, a spurious relationship between two variables can occur because of sheer coincidence or even through some third factor which affects both variables. For instance, it has been shown that investment in science and the number of reported suicides are positively associated in the United States: these two mechanisms are clearly independent and the observation in question must be due to some pure coincidence. The observed positive association between average ice cream consumption and the incidence of house robberies is, instead, an example of spurious correlation

due to a confounding factor. This statistical correlation emerges without any causal link between the two variables: a third one – outdoor temperature – comes here into play, as it leads to an increase in both ice-cream consumption and in burglaries. This discussion clarifies how it is impossible to conclusively detect the existence of a cause-effect relationship without any further statistical investigation. In order to confidently conclude that a variable Z has a causal effect upon another variable Y, the policy maker must implement a suitable statistical approach, which effectively allows to disentangle the effect in question from other causal or spurious mechanisms linking Z with Y.

A statistical framework for performing causal inference has been introduced by Donald Rubin (G. W. Imbens & Rubin, 2015; Rubin, 1974, 1980); thus, it is commonly known as the Rubin Causal Model (RCM). In this framework, the individual characteristic of interest is named "outcome" variable and commonly denoted as Y, while the intervention that preasumibly affects the outcome is called the "treatment" variable and is often labeled as Z. The latter is typically a binary statistical variable; the presence of the intervention is expressed for an individual observation as $Z = 1$, and is to be compared against the complementary "control" treatment ($Z = 0$) which indicates the lack of treatment. Both treatment and outcome variables need to be measurable, fully observable. Thus, the RCM deals with a population of units $\mathcal{N} = \{1, \dots, i, \dots, N\}$, where the generic unit i is exposed to a given exposure condition Z_i, with $Z_i \in \mathcal{Z} = \{0, 1\}$. With $\boldsymbol{Z}$ and $\boldsymbol{Y}^{obs}$ I denote the corresponding vectors of the whole sample $\mathcal{N}$. Moreover, $\mathbf{X}_i$ denotes a vector of P covariates (or pre-treatment variables) that are not influenced by the treatment assignment. For each unit i, it is theoretically possible to observe the individual outcome Y_i^{obs}. The RCM is predicated on the thought experiment of comparing the two alternative individual *potential outcomes*, that is the outcome that would be observed if a given individual were exposed to the active treatment (the intervention) and the one that would occur if the same individual is exposed to the control treatment. The crucial issue is that in reality, these two outcomes cannot be jointly observed for the same unit of observation, because an individual can only be assigned

to either the treatment group or to the control group. The only outcome which is observable is the one corresponding to the type of treatment an individual is actually exposed to. Formally, I denote with $Y_i(z)$ the potential outcome of a given individual i under the treatment z, where $z \in \mathcal{Z} = \{0, 1\}$. The only outcome the policy maker is able to observe over i is the one that corresponds to the treatment he actually received Z_i, that is $Y_i^{obs} = Y_i(Z_i)$.

$$Y_i^{obs} = Y_i(Z_i) = \begin{cases} Y_i(0) & \text{if} \quad Z_i = 0, \\ Y_i(1) & \text{if} \quad Z_i = 1. \end{cases} \tag{1.1}$$

Suppose that a doctor is willing to assess whether a new drug ($Z_i = 1$) is effective against a eye disease by comparing it with the standard medical treatment ($Z_i = 0$). She tests the effect of the new drug over her five patients, who have all developed such a disease, and measures their diopters through an eye exam, after the first administration of the medical dose. She randomly assigns the experimental drug to two out of five patients, while the other three take the standard medicine. If she were an omniscient perfect doctor, she would be able to observe potential individual diopters both under the new drug and under the validated standard therapy for each of her patients. However, a real doctor lacks the possibility to inspect both outcomes for each personl: she is able to monitor only the outcome corresponding to the medical treatment that each patient has actually received. Table 1 displays some hypothetical information that the "real" doctor is able to observe against the more complete clinical picture that only a "perfect" omniscient doctor would be able to evaluate.

This example highlights how a causal inference cannot be performed only on the basis of a single individual i: instead, it must be implemented over a sample of units $\mathcal{N} = \{1, \ldots, i, \ldots, N\}$. In fact, a comparison based on single individual would be limited by the impossibility to observe the one of two potential outcomes which corresponds to the treatment level that was *not* experienced by the individual. In this sense, causal inference is actually about solving a *missing data problem*; consequently, the core idea behind every method of causal inference is about finding an

Real Doctor			
Unit	Z	Y(1)	Y(0)
1	1	7	?
2	0	?	2
3	0	?	3
4	1	9	?
5	0	?	0

Perfect Doctor			
Unit	Z	$Y_i(1)$	$Y_i(0)$
1	1	7	5
2	0	2	2
3	0	5	3
4	1	9	2
5	0	2	0

Table 1: Real Doctor vs Perfect Doctor: in both scenarios black numbers denote the realizations of the observed outcome. The real doctor is not able to observe the individual outcomes which correspond to the treatment that did not realize (whose value is replaced with the question mark "?"). Conversely, the perfect omniscient doctor has the power to monitor every aspects of reality, and the outcomes he only has the chance to observe are red-colored.

efficient strategy to impute "missing" potential outcomes. Every imputation strategy requires to deal with a consistent number of units which can potentially experience both treatment levels (Petersen et al., 2012; Rubin, 1974). Indeed, the individual missing potential outcome, that is the treatment level that a unit has *not* experienced, is imputed from the outcomes associated with the same treatment level but observed in other units from the same population of reference.

The RCM is capable to solve the missing data problem and achieve the desired policy evaluation objectives under a series of conditions that are formalized in terms of statistical assumptions. One such assumption, which is key, guarantees well-defined potential outcomes and hence that the treatment effect can be correctly outlined and estimated. This assumption, known as *Stable Units Treatment Assumption* (SUTVA) (Rubin, 1986), ensures that there are no multiple versions of the treatment and rules out the possibility of "interference" among units.

Assumption 1 (Stable Units Treatment Assumption (SUTVA))**.** *SUTVA is made up by two different components:*
(i) the Individualistic Treatment Response (ITR) (C. F. Manski, 2013) (or no interference*) states that there is no interference between units: every unit's*

potential outcomes are defined only by that unit's own treatment.
(ii) the consistency (or no multiple version of the treatment*) states that there are no different versions of the treatment levels. Formally, given two treatment assignment vectors,* $\boldsymbol{Z}$ *and* $\boldsymbol{Z'}$*, if* $Z_i = Z_i'$ *then* $Y_i(Z_i) = Y_i(Z_i')$

The assumption implies that the potential outcome of the unit i can be uniquely indexed with respect to its own treatment status, that is $Y_i(z)$: $z = \{0, 1\}$.

An important feature of proper causal inference studies is that the unit-level allocation to one of the two treatment levels cannot be analytically deducted from prior information. In other words, it should be impossible to identify a deterministic rule which allows the researcher to guess whether a given individual i is exposed to the "active" or, instead, to the control level of the treatment, at least not before that the actual treatment assignment (intervention) takes place. The individual allocation to a given treatment level should thus not be the result of a deterministic principle; it rather follows a stochastic rule and relates to the so-called *treatment assignment mechanism* (G. W. Imbens & Rubin, 2015; Rubin, 1974, 1980). The treatment assignment mechanism is the stochastic process governing which units receive the active level of treatment and which ones receive the control level:

$$P\left(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{Y(0)}, \boldsymbol{Y(1)}\right).$$

Each unit can receive the treatment or the control level depending on which assignment vector $\boldsymbol{Z}$ will show up. Note that, in order to fairly define and estimate the treatment effect, one needs to assume that, for each unit i, the assignment probability $\pi_i(z) = P(\boldsymbol{Z}_i|\boldsymbol{X}_i, \boldsymbol{Y_i(0)}, \boldsymbol{Y_i(1)}$ is strictly bounded between 0 and 1. This means that each unit has a nonzero probability of being exposed to either treatment status. This assumption is known as the *Positivity Assumption*, and can be formally expressed as

Assumption 2 (Positivity)**.** $0 < \pi_i(z) < 1$ *for all* $z \in \mathcal{Z}$*.*

Under the SUTVA and the Positivity Assumption, it is possible to define the *treatment effect*, which represents the element of major interest in

policy evaluation settings. Here, we express the *Average Treatment Effect* (ATE) τ while adopting a "superpopulation" point of view, i.e. by assuming that the sample $\mathcal{N}$ is a random draw from an infinite population. Formally, the ATE is defined as follows:

$$\tau = \mathbb{E}\left[Y_i(Z_i = 1) - Y_i(Z_i = 0)\right].$$

Randomized settings vs Observational Studies

The specific imputation strategy to be implemented in order to solve the missing data issue is typically chosen after taking into account various elements of the empirical setting at hand. The main aspect that a researcher must consider while choosing such a strategy is the shape of the treatment assignment mechanism which, as discussed above, is key in the determination of individual treatment statuses. This mechanism can be either known *a priori* by the researcher (this is the case if, for example, the treatment is randomly assigned to units, according to some randomization criteria and parameters) or it can be unknown (that is, the treatment status depends in a non-deterministic way upon some observable *and* unobservables characteristics of the units; in this sense, the units can possibly affect their very treatment status) (Benson & Hartz, 2000; Concato et al., 2000; Kovesdy & Kalantar-Zadeh, 2012). The first scenario refers is that of a *randomized study* (Leuven et al., 2010; Manning et al., 1987), while the second scenario is that of an *observational study* (Agarwal et al., 2008; Segal et al., 2004; Victora et al., 2004). An example of a randomized trial is the clinical evaluation of the effects of a new drug, where the researcher has the opportunity to manipulate the intervention by randomly assigning the experimental treatment to a specific sub-sample of the initial population (Gueyffier et al., 1997; Park et al., 2012). Conversely, analyses about the effect of smoking on individual health are prime examples of observational studies, (C. F. Manski, 2013; Sandler et al., 1985). Since policy makers cannot force people to smoke – for both ethical and practical reasons: they can only passively observe which of the men and women included in the analysis usually smoke cigarettes – a randomized control over the treatment assignment

mechanism is impossible. In political science, assessing the social effect of a local policy necessarily implies an observational approach, since administrative entities independently choose whether to implement such a policy or not (Johansson & Palme, 2002; Moore & Rhodes, 1973; Perotti, 2005). In summary, observational studies refer to all those empirical scenarios where the treatment assignment mechanism cannot be manipulated by the researcher (Rosenbaum & Rubin, 1983b).

From a statistical perspective a pure randomized setting is preferable, as it ensures that the treatment and the control group are sufficiently similar with respect of baseline covariates. Indeed the randomization design, especially if it follows a stratified approach, usually guarantees a perfect balance among treated and untreated units with respect to individual baseline characteristics (whether these are observed or not). Otherwise, a significant unbalance in the two treatment arms could cause a dependence between potential outcomes and treatment variable, thereby introducing a bias in the estimate of the treatment effect (Hansen & Bowers, 2008; Morgan, Rubin, et al., 2012). However, the "specter of covariate unbalance" can loom in the horizon in randomized settings too: if the randomization plan has been compromised by an imperfect compliance to the treatment assignment or, for whatever reason, has not succeeded in generating sufficient covariate balance, the analysis requires a strategy to addressing the unbalance (Hennessy et al., 2016; Senn, 1989). This issue is even more crucial in observational studies: there, a randomization plan does not even exist; and all the imputation criteria that are adopted in those settings, are conceived for the sake of guaranteeing an adequate balance between the baseline characteristics of the actively treated units and the control group. In other words, their objective is to make the observational study as close as possible to a randomized study (Rosenbaum, 2002; Rosenbaum & Rubin, 1983b). In an observational setting, conditioning for baseline covariates is essential to avoid predictable dependencies between treatment and potential outcomes. This assumption is known as *Unconfoundedness Assumption* (Rosenbaum & Rubin, 1983b, 1984) and it can be formally expressed as

Assumption 3 (Unconfoundedness)**.**

$$P(Z_i = 1, |Y_i(z), \boldsymbol{X}_i) = P(Z_i = 1, |\boldsymbol{X}_i).$$

This assumption enables researchers to meaningfully and safely infer causal effects even in observational settings. However, it can neither be tested nor it has testable implications. Furthermore, it requires conditioning upon a (potentially) large set of unit-level covariates (Guo & Fraser, 2014; Rosenbaum & Rubin, 1983b). To avoid an high-dimensionality problem, the general approach – which is well grounded in the statistical literature for causal evaluation studies – consists in conditioning upon a *scalar synthesis* of covariates, which is called the *propensity score* (Imai & Van Dyk, 2004; Rosenbaum & Rubin, 1983b). This quantity represents the individual probability of being exposed to the active treatment, conditioning on baseline covariates. Formally, the propensity score is defined as follows:

Definition 1 (Propensity Score)**.**

$$\phi(z, \boldsymbol{x}) = P(Z_i = z|\boldsymbol{X}_i = \boldsymbol{x}).$$

The propensity score can be formally shown to be a "balancing score:" this means that conditioning upon the entire set of covariates is statistically equivalent to conditioning only upon the propensity score (Rosenbaum & Rubin, 1984).

Nature of the treatments

So far we have discussed the causal inference framework assuming that the treatment variable has a binary characterization. However, this needs not be always the case. A relevant portion of policy evaluation studies involve more complex treatments, such as *multi-valued* treatments or *continuous* treatments. Multi-valued treatments are interventions which vary over more than two categories; while highly diffused in nature, they are also widely employed in all those empirical studies whose aim is to compare different characterization of the treatment variable (Linden et al., 2016; Lopez, Gutman, et al., 2017; S. Yang et al., 2016). For instance,

epidemiologists may need to compare different types of drug (Linden et al., 2016). Other analysts may want instead to evaluate different sorts of incentives for visiting museums that aimed at students (Forastiere et al., 2019a). This types of treatment are also exploited to analyze more complex types of treatment (such as, for example, institutional political attitudes towards convoluted issues)(Tortù et al., 2020).

Continuous treatments are similarly widespread (Croxatto et al., 1993; Del Prete et al., 2019; N. Wilson et al., 1995). They are related to all types of interventions whose intensity varies over a continuous domain. For instance, epidemiologists might need to test a drug by experimenting over a (quasi) continuous ensemble of doses (Kondo & Togari, 2011). In both cases of a multi-level and a continuous treatment, coercing the intervention to a binary characterization implies a relevant informative loss: in the multi-valued setting it averts from evaluating differential treatment effects (Cattaneo, 2010), while in the continuous scenario it prevents to meaningfully appreciate the consequences of variation to treatment exposure. Figure 3 provides a graphical representation of the various kinds of treatment discussed thus far.

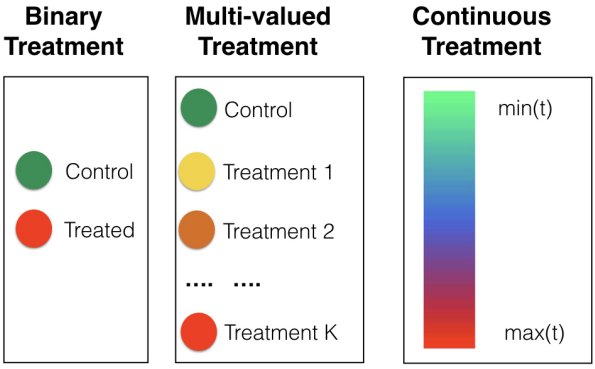


Figure 3: Nature of the treatments: (Left) binary treatments varying over a binary domain; (Center) multi-valued treatments characterized by more than two categories; (Right) continuous treatments moving over a continuous (or quasi-continuous) domain

Heterogeneous Causal Effects

In recent years, the statistical community dedicated to causal inference methods and their applications has matured a growing interest towards heterogeneous treatment effects, which concerns how the treatment effect varies across different sub-populations of units (Bargagli-Stoffi et al., 2019; Cockx et al., 2019; K. Lee et al., 2018; Zhao et al., 2017). This follows from the recognition that treatment effects likely vary along the features of the units themselves. Assessing the possible heterogeneity in the treatment effect is especially relevant from the perspective of policy makers, as it allows to understand on which sub-populations of units the treatment is especially effective and what are the major individual characteristics driving heterogeneity. For instance, one can imagine the extreme case where the average effect of a treatment (i.e. a drug) is positive on the population (the drug is generally effective against a specific disease), but for a sub-population of units with certain characteristics the treatment is either ineffective or it even has a negative impact. Evaluating heterogeneous causal effects means, in a broader sense, conditioning on a given partition of the covariate space. Formally, by denoting some P-variate partition of the feature space (where P denotes the number of observed characteristics) as $\boldsymbol{x}$, we can define the *Conditional Average Treatment Effect* which measured on those units i with $\boldsymbol{X}_i = \boldsymbol{x}$) as:

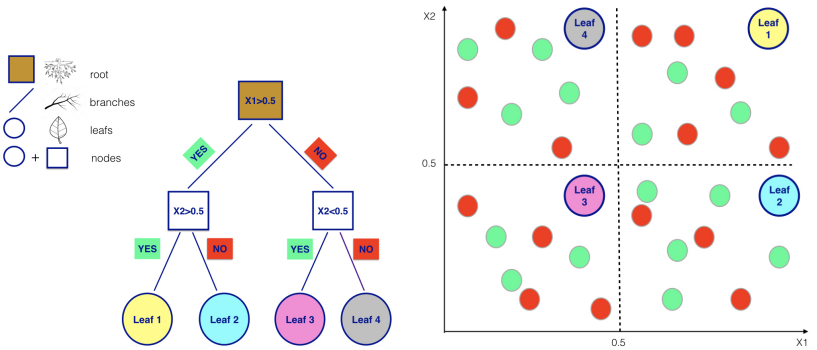
$$\tau(\boldsymbol{x}) = \mathbb{E}\left[Y_i(Z_i = 1|\boldsymbol{X}_i = \boldsymbol{x}) - Y_i(Z_i = 0|\boldsymbol{X}_i = \boldsymbol{x})\right].$$

Although the sub-populations with heterogeneous causal effects can be externally specified using prior knowledge, these entail the risk that are inherent in erroneous suppositions; in addition defining sub-populations by relying on prior information implies that heterogeneity is evaluated only on those specific sub-populations, preventing researchers from assessing which characteristics drive heterogeneity in the treatment effect. Thus, it can be preferable to use data-driven approaches instead. In this regard, the general approach is to evaluate heterogeneous effects by running a data-driven machine learning algorithm tailored for causal inference, which identifies the most significant drivers of heterogeneity

(Athey & Imbens, 2016; Athey & Imbens, 2015; Athey et al., 2019; Foster et al., 2011; Hahn et al., 2020; Hill, 2011; Lechner, 2019; K. Lee et al., 2020; Starling et al., 2019; Su et al., 2012; Wager & Athey, 2018). The intuition behind these machine learning algorithms is that sub-populations are partitioned by iteratively separating those groups whose estimated conditional average treatment effect deviates the most from the average treatment effect estimated in the population as a whole.

These algorithms are usually tree-based algorithms (Athey & Imbens, 2016; Athey & Imbens, 2015; Wager & Athey, 2018) and they all have their roots in the the Classification And Regression Trees (CART) algorithm (Friedman et al., 1984). CART is a widely used algorithm for the construction of trees, and are typically such that each node is split into only two branches (i.e. binary trees). Binary trees are sprouted by recursively partitioning the observations from the *root* (the set containing all the observations) into two child *nodes*. This procedure is iterated until the tree reaches the final nodes, called *leaves*. The CART algorithm requires two key elements: *i.* a criterion function, which determines the splitting process of units and *ii.* a stopping rule, which establishes the conditions under which the algorithm stops splitting. The standard CART identifies heterogeneity in the relationship between the observed outcome and the baseline characteristics so to accurately predicting the outcome variable. Consequently the criterion function is defined such that at every step of the splitting process the prediction error is minimized. Once that the algorithm has met one of the stopping conditions it returns a tree Π , which is a partition of the covariate space $\mathcal{X}$ into M non-overlapping regions: $\Pi = \{\ell_1, \ldots, \ell_M\}$, where $\bigcup_{m=1}^{M} \ell_m = \mathcal{X}$, and with $\ell(\mathbf{x}, \Pi) : \mathcal{X} \to \Pi$ a function that maps each vector $\mathbf{x}$ of the covariate space into a region. Figure 4 shows how recursive partitioning works in a binary tree: in this simple tree, there are only two continuous predictors that are bounded between 0 and 1, that is $X_1 \in [0, 1]$ and $X_2 \in [0, 1]$. Units are mapped into the final leafs according to their observed values of X_{i1} and X_{i2}.

The counterpart of CARTs in the causal environment are the Causal Trees (Athey & Imbens, 2016; Athey & Imbens, 2015). The intuition behind causal trees is akin to that of CARTs, and is based on recursive parti-



(a) Recursive partitioning

(b) Mapping leafs according to the individual vector of characteristics

Figure 4: (Left) An example of a binary tree. The internal nodes are labelled by their splitting rules and the terminal nodes. (Right) The corresponding partition of the sample space.

tioning. The difference is that in the causal framework, the criterion function is defined to detect, at every step of the splitting process, the subpopulations which exhibit the maximal heterogeneity in the treatment effect, while also penalizing those subgroups displaying higher variance in the estimated causal effects (Athey & Imbens, 2016).

1.2.2 Complex networks: idea and notation

Many real-world phenomena can be described in terms of interactions among agents. Think at the huge amount of interactions which occur everyday in the modern social networks, such as Facebook: users easily become virtual friends, they share virtual content and react to the one shared by their friends (Traud et al., 2012). In a similar vein, consider the interactions among neurons that are found in our brain (Chua & Yang, 1988): they are numerous and dynamic. These simple examples points out that in nature it is possible to observe very complex interactions between objects and that their complexity is due to both their shape and amount. In recent years, also thanks to the growing computing power

that has enabled researchers to manage complex and big data sources, the interest about complex networks has been growing rapidly (Albert & Barabási, 2002; Boccaletti et al., 2006; Caldarelli, 2007; Cimini et al., 2019; Strogatz, 2001). Complex networks are used to represent a large variety of natural phenomena of various sorts, while modeling the nature of the agents they are representing. When connections concern human beings and describe individuals interacting with one another, complex networks are denoted with the term *social networks* (Catanzaro et al., 2004; Degenne & Forsé, 1999; Mislove et al., 2007). Social networks help describe multiple aspects of human behavior: from friendship or parental ties (Cai et al., 2015; Hendrickson et al., 2011; South & Haynie, 2004) to professional collaborations (Ahuja, 2000; Becatti, Crimaldi, et al., 2019; Fowler, 2006a; Newman, 2001a, 2001b). Interactions between agents are extremely relevant: for example, several recent studies have demonstrated that social networks are vehicles for the spread of information which is possibly inaccurate or even malicious (Del Vicario et al., 2016; Quattrociocchi et al., 2016). Figure 5 shows an example of a collaboration network. Specifically, it depicts the "co-sponsorships" networks between legislators in the U.S. House of Representatives (HoR) during the 111^{th} legislature. In this setting, the agents are the politicians elected at the U.S. HoR in that Cycle, while ties between them signal the presence of a co-sponsorship tie (i.e. that one of the two connected legislators has supported a bill which is also sponsored by the other). This particular empirical setting is expanded at length in a later chapter of this dissertation.

Figure 6 provides an example of a friendship network: it describes friendship ties between pupils belonging to the same classroom.

Economic agents interact too. Firms are linked by economic or legal relationships (M. O. Jackson, 2010; Schweitzer et al., 2009) such as input/output ties (Blöchl et al., 2011; Contreras & Fagiolo, 2014), ownership and control relations (Conyon, Muldoon, et al., 2008; Rungi et al., 2017) and more. Likewise, banks are connected through financial ties of heterogeneous intensity; it has been observed that this plays a kay role in the propagation of financial shocks (Battiston & Caldarelli, 2013; Elliott et al., 2014; Gai & Kapadia, 2010). Finally, large institutions like as

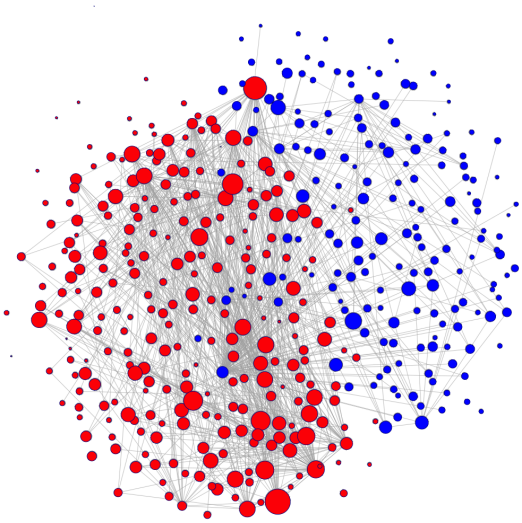


Figure 5: Collaboration network: co-sponsorship ties between politicians in the US House of Representatives. Nodes are colored according to their party membership (*red* nodes represents democrats, while *blue* vertexes are republicans

national or local governments interact intensively. They may be linked via explicit agreements and political alliances (H. Chen & Chen, 2002; C. K. Jackson et al., 2015), though trade links (Furusawa & Konishi, 2007; Squartini et al., 2011), or by their geographic position, defined in a broad sense (Barthélemy, 2011; Crucitti et al., 2006; Del Prete et al., 2019). Figure 7 provides a simple example of a boundaries-based spatial network. Nodes represent those OECD countries located in Europe, while links between them highlight that the two countries share a boundary (this particular network too is expanded in a later chapter).

Some Basic Notation

In the scientific literature the term "network" denotes a natural phenomenon, while the word "graph" is used to discuss the mathematical object that represents the given phenomenon. Hence, the relationships

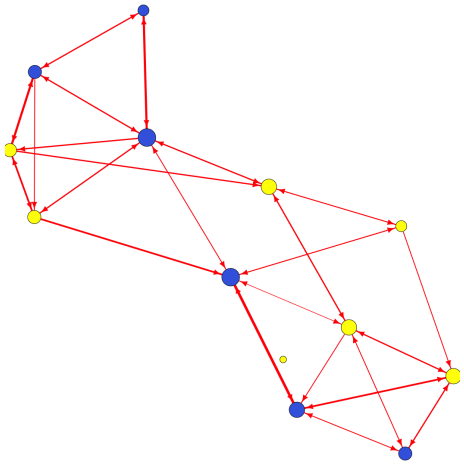


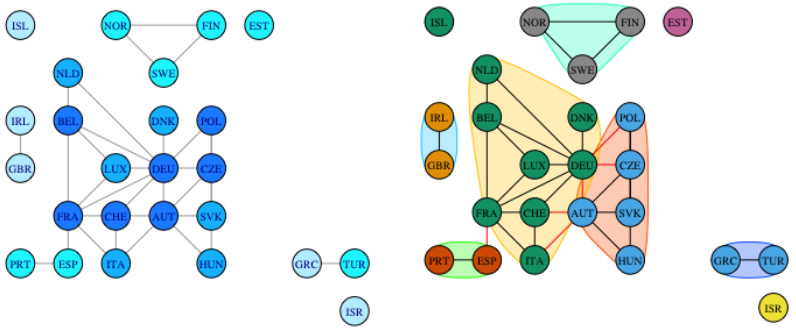
Figure 6: Friendship network: friendship ties linking schoolchildren of a given class. Edges' width is proportional to the strength of the tie, while nodes' color signals whether scholars are more inclined to establish relationship within their classroom (blue nodes) or outside their classroom (yellow nodes)

that govern the dynamics of complex systems are graphically depicted by a *network* where the observed nodes are the elements of the population of interest, while the observed links represent the relationships that relate nodes to one another. The mathematical tool used to describe such relationships is a graph $\mathrm{G} = (\mathcal{N}, \mathcal{E})$.

- $\mathcal{N}$ *Nodes*: this set represents the statistical units that appear to be the nodes in the graph.

- $\mathcal{E}$ *Edges*: this set contains links between units.

The joint characterization of nodes and edges determines the type of network. If edges are not characterized by a "structural direction" the network is said to be an *undirected* one. This kind of network describes "reciprocated" ties, which cannot be active only in a given direction, but not in the other. Given two units i and j belonging to the population $\mathcal{N}$, a reciprocated tie between node i and node j is interpreted as a mutual



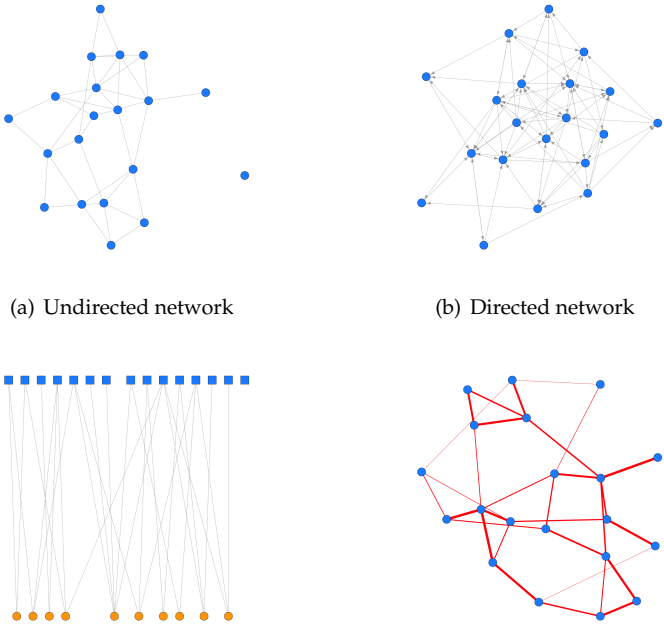
(a) Geographical centrality (b) Geographical Communities

Figure 7: Spatial network: boundary-based network of OECD countries located in Europe. (Left) Countries are colored according to their spatial centrality: the darker is the blue shadow the higher is the number of countries with which they share a boundary. (Right) Countries are grouped according to spatial communities, which have been identified through a data-driven for community detection (Clauset et al., 2004).

relationship that affects both nodes in a symmetric way and is necessarily reciprocal; it is usually expressed as $i \leftrightarrow j$. To make an example, Facebook ties are undirected, as (virtual) friendships must to be accepted and reciprocated in order to exist. The boundary-based network connecting European OECD countries is also an undirected network, as sharing a boundary is necessarily a symmetric relationship involving a pair of countries. Conversely, all those networks where links are possibly characterized by an asymmetric orientation are said to be *directed*. Such networks might include both mutual ties, denoting that two units i and j belonging to $\mathcal{N}$ are linked through a reciprocated relationship $i \leftrightarrow j$, and unilateral ties, such that *either* the tie goes from i to j (i.e $i \rightarrow j$), or the tie goes from j to i (i.e that is $j \rightarrow i$). The links in the social networks examples from the previous subsection are characterized by an explicit direction: cosponsorship ties among legislators are directed since a politician can possibly support another but not vice versa; similarly, friendship relationships between classmates as in Figure 6 are interpreted under a di-

rectional perspective, since a scholar might regard another classmate as a friend but not vice versa. Finally, networks describing the ownership ties within a sample of firms are structurally directed: if a firm owns another, one cannot possibly observe the same relationship going in the opposite direction at the same time. The three types of networks I have just presented differ in the characterization of their edges. However, nodes can also have different profiles: in particular,nodes can be divided into disjoint sets and the edges ($\mathcal{E}$) connect nodes that belong to different sets.. A network of this sort is commonly said to be *bipartite*. (Barber, 2007; Saracco et al., 2015; Shang et al., 2010; T. Zhou et al., 2007). In a bipartite network $\mathrm{G}^{Bip} = \{\mathcal{N}_1, \mathcal{N}_2, \mathcal{E}_{\mathcal{N}_1,\mathcal{N}_2}\}$, the entire set of nodes $\mathcal{N}$ is partitioned into two separated subsets, respectively labeled as $\mathcal{N}_1$ and $\mathcal{N}_2$, and edges connect nodes belonging to different sets. Bipartite networks are widely diffused in nature. Consider for instance a variation of the previously illustrated network about congressmen co-sponsorship: the network of contributions that are directed from firms to politicians to support their electoral campaigns. Such a network is evidently not only directed, but also bipartite: politicians and firms belong to two separated layers and network ties (representing the monetary transactions) can only connect nodes from different layers.

Regardless of the specific joint characterization of nodes and edges, complex networks can also be possibly characterized by information about the "intensity" of individual edges. Networks of this sort are said to be *weighted* (Newman, 2004; Squartini et al., 2013). Consider again the networks from the previous examples: their edges all bear a specific "weight." To recall one such example, in a network of supply relationships between firms the weight can represent the quantity of exchanged products, or the value of the corresponding monetary transactions. While the latter is an example of a network which both directed and weighted; weighted networks can also be undirected or bipartite. For instance, rating networks where users evaluate products by assigning them a vote (Becatti, Caldarelli, et al., 2019) are an example of a bipartite weighted network. Figure 8 provides graphical illustration about the four types of networks we have just discussed. Finally, there is a particular type of net-



(a) Undirected network

(b) Directed network

(c) Bipartite network

(d) Weighted network

Figure 8: Types of Networks

work, known as *multiplex* network, which will be extensively employed in the last Chapter of the dissertation. A multiplex is characterized by the presence of a unique set of nodes, whose elements interact according to two distinct networks. It may be regarded as a particular type of multilayer network. Figure 9 shows graphical rappresentation of a multiplex.

Any graph based on a monopartite network $\mathrm{G}(\mathcal{N}, \mathcal{E})$ admits a unique representation in terms of its adjacency matrix $\boldsymbol{A} = \{a_{ij} : i, j \in \mathcal{N}\}$. The element a_{ij} characterizes the relationship between i and j. If ties are not characterized by a specific weight, the adjacency matrix $\boldsymbol{A}$ is a binary matrix, where the generic element a_{ij} equals 1 if units i and j are connected,

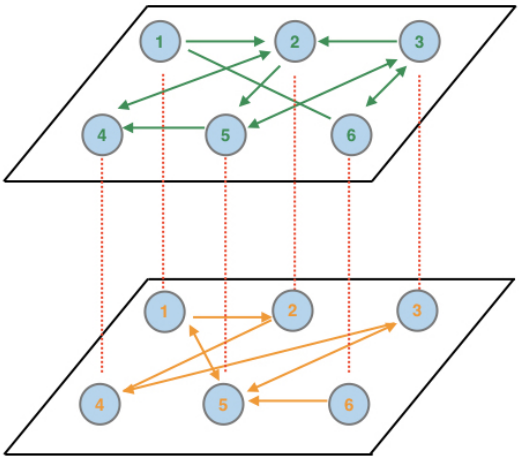
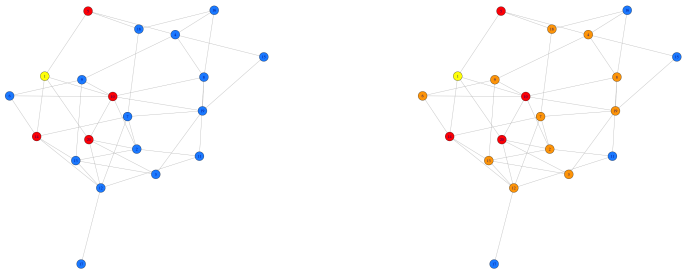


Figure 9: Multiplex network: nodes are connected according to separated networks; links of the first network are *orange* colored, while ties of the second network are *green* colored

and 0 otherwise. If the corresponding network is undirected, the adjacency matrix is also symmetric. In the case of a directed network, a_{ij} and a_{ij} might differ for some $(i, j) \in \mathcal{N}^2$: there could be pairs of units characterized by tie going from one node to the other, say i to j ($a_{ij} > 0$), but the relation is not reciprocated ($a_{ji} = 0$). For each node $i \in \mathcal{N}$ it is possible to characterize the set of its immediate neighbors through G. This set, that we denote by $\mathcal{N}_i$, includes all the other nodes having an immediate connection with i. If G is undirected, this set can be formally written as $\mathcal{N}_i = \{j : (i, j) = (j, i) \in \mathcal{E}\}$. If otherwise ties feature an explicit direction, we distinguish between the set of nodes that "deliver" an in-going link to i $\mathcal{N}_i^{in} = \{j : (j, i) \in \mathcal{E}\}$ and the set of nodes which instead "receive" an out-going tie originating from i, $\mathcal{N}_i^{out} = \{j : (i, j) \in \mathcal{E}\}$. These sets identify the first order neighborhood of unit i (i.e the nodes who have an *immediate* connection with i). One can also easily identify, for each node i, its k^{th} order neighborhood, that is the set of nodes that are indirectly connected to i by at least k subsequent links: for instance, the second order neighborhood of node i includes its immediate neighbors as well as

the neighbors of its immediate neighbors. Figure 10 represents first and second order neighborhoods for a hypothetical node i.



(a) First Order Neighborhood (b) Second Order Neighborhood

Figure 10: k^{th} order neighborhood of a given node. The node i (the *yellow*-colored node) is the vertex of interest: *red* nodes identify vertexes in the immediate neighborhood of i; *orange* nodes represent nodes in a second order neighborhood of i; *blue* nodes represent nodes who do not have at least a second order connection with i

The cardinalities of these node-specific sets are variously named as different types of *degree* of a node. In undirected networks, there is only one degree measure for every node i, which counts the number of their active connections: $N_i = |\mathcal{N}_i| = \sum_{i<j} a_{ij}$. In directed networks one must specify two different degree measures: the *out-degree* counts the number of a node's out-going connections, $N_i^{out} = |\mathcal{N}_i^{out}| = \sum_{i:1}^{N} a_{ij}$; while the *in-degree* measures the number of in-going ties, $N_i^{in} = |\mathcal{N}_i^{in}| = \sum_{i:1}^{N} a_{ji}$.

1.3 Bridging the two themes

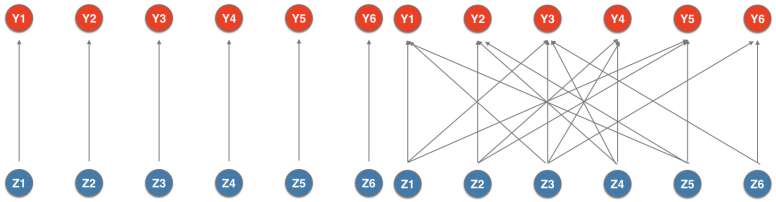
After having briefly and separately introduced causal inference and complex networks, in what follows I discuss some statistical mechanisms that connect both topics at hand. In fact, the interplay between the two is the thread that unites all the contributions elaborated in the upcoming chapters. This dissertation bridges causal inference and complex networks by

developing two methodological innovations: (i) on the one hand, we address issues that arise while evaluating the effect of a given intervention on a population of interconnected units, by proposing new methods that expand the growing statistical literature about causal inference under interference; (ii) on the other hand, we present novel tools to employ the potential outcomes framework in order to analyze the dynamics of network formation in a multiplex. By introducing some conceptual points that are extensively discussed in the following chapters, the upcoming subsections provide some general intuition that help appreciate both the methodological contributions of this thesis.

1.3.1 Causal inference under interference

Policy evaluation studies intend to estimate the effect of an intervention. However, the causal evaluation might be complicated by the presence of interactions between units. From the causal inference perspective, understanding these interactions and accounting for them in the analysis is crucial. Interactions between units may generate an *interference* mechanism, which formally occurs when the potential outcome of a given unit is affected by the treatment assignment of other units (Cox, 1958) (the interference mechanism is also known as *spillover*). Interference may arise due to social, physical or virtual interactions among social and economic agents (Crimaldi et al., 2020; Tortù et al., 2020). In epidemics, the introduction of a new vaccine benefit also unprotected individuals, as their probability to be infected decreases in the wake of an overall reduction in the reservoir of infection (Bridges et al., 2000; Nichol et al., 1995). In education, students that are assigned to a learning program may interfere with their untreated peers through knowledge transmission paths (Chin et al., 2013; de Heer et al., 2011). In political science, policies implemented by administrative entities may impact also neighboring or allied territories (Fang et al., 2019; Naranjo, 2010). In economics, incentives targeted to firms affect also those firms which do not directly benefit of the incentive but do have an economic or juridical relationship with favored companies (Chuang & Lin, 1999; Cohen et al., 2002). In finance,

a monetary shock smashing into some financial institutions may propagate over those entities involved in their transactions (Squartini et al., 2011; G.-J. Wang et al., 2017; J. Yang & Zhou, 2013). In labor market, a job placement assistance directly helps job seekers who decide to rely on this service, but indirectly impacts the other job seekers who compete on the same job market and are penalized by a competing disadvantage. In marketing, individuals who are exposed to an advertisement may adjust their consuming behavior and influence their friends.
In all these scenarios, the potential outcomes of units $\boldsymbol{Y}(z)$ are affected by the treatment assignment of agents who interact with them. Figure 11 provides a graphical intuition of what interference is (Tortù et al., 2020). When interference arises, the potential outcome framework, in its



(a) No interference (b) With interference

Figure 11: No interference scenario vs interference scenario: under no-interference (left side figure) units' potential outcomes (red dots) are affected only by their own treatment level Z_i (blue dots); in the presence of interference (right side figure), the potential outcome of a given unit depends by the treatment status of other units

standard formulation, is not valid anymore. The standard Rubin Causal Model rules out the possibility of interference, by relying on SUTVA (Assumption 1): hence, the spillover mechanism constitutes a violation of this key assumption, which sustains the entire potential outcomes framework. This violation introduces a bias in the estimates and may lead to inaccurate conclusions about the real effect of an intervention. As a consequence, the RCM requires to be rearranged so to account for spillover effects, while estimating the treatment effect. Furthermore, spillover effects may represent themselves the object of interest, as they may help in

fully understanding the real impact of interventions. Moreover, understanding spillovers may be crucial in designing experiments (Angelucci & Di Maro, 2015; Baird et al., 2018; Eckles et al., 2017; Kang & Imbens, 2016; Sinclair et al., 2012). The design phase of a given randomized intervention would gain a significant advantage from knowing which individuals guarantee a beneficial spillover of the treatment. This information would also allow the policy-maker to decrease the percentage of agents assigned to the active treatment, relying on the fact that untreated individuals will be exposed to the intervention anyway, due to their interactions with their treated and influential peers.

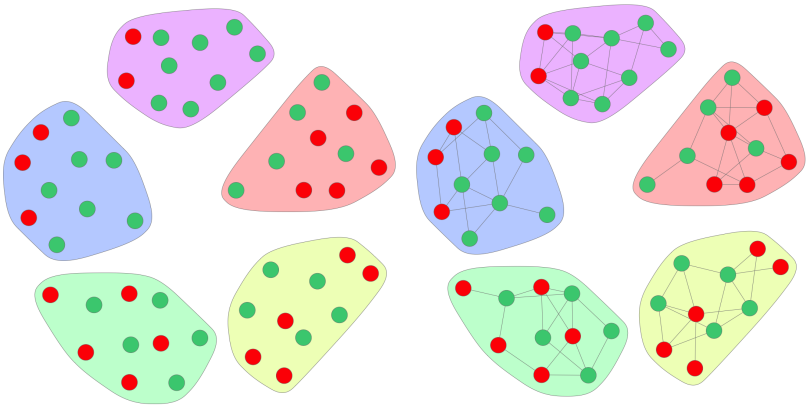
To provide a solution for this methodological issue, researchers have recently started to work on interference, by extending the standard potential outcome framework so to account for the spillover mechanism. Most of the existing works focus on the role of interference in *randomized trials* (Aronow, 2012; Aronow & Samii, 2017; Athey et al., 2018b; Bowers et al., 2013; Forastiere et al., 2019a; Hudgens & Halloran, 2008; Imai et al., 2020; Kang & Imbens, 2016; Liu & Hudgens, 2014; Rosenbaum, 2007; VanderWeele et al., 2014). Just a few contributions explicitly deal with observational settings (Forastiere et al., 2020; Forastiere et al., 2018; Hong & Raudenbush, 2006; Sofrygin & van der Laan, 2017; Tchetgen & VanderWeele, 2012; van der Laan, 2014) [1].

Types of interference

The interference mechanism can be classified in various types, according to the shape of the interactions which drive spillovers. In a broader sense, it is possible to distinguish among three types of interference: i) agents belong to exogenous clusters and the spillover mechanisms occur only within clusters, that is, units are equally exposed to the treatment assignments of individuals referring to the same cluster (Basse & Feller,

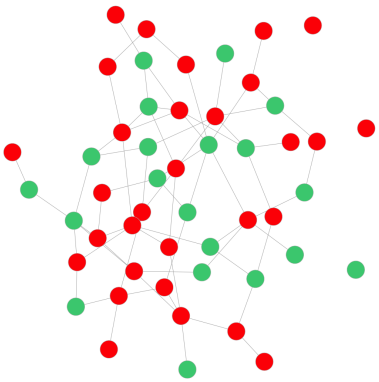
[1] This is just preliminary literature review of the existing works on interference. Each of the three upcoming chapters that focuses on spillovers will start with a detailed literature about causal inference and interference presenting existing works which are particularly related to the present contribution

2018; Hudgens & Halloran, 2008; Sobel, 2006b) (*clustered interference*); ii) agents belong to exogenous clusters and the spillover mechanisms occur only within clusters, according to the links of a cluster-specific network (*clustered-network interference*); iii) interactions are described by a complex network (*general (network) interference*). Figure 12 provides a graphical intuition about the different types of interference.



(a) Clustered interference

(b) Clustered network interference



(c) Network interference

Figure 12: Types of interference: (Top Left) interactions homogeneously happens within exogenous clusters; (Top Right) interactions occur within exogenous clusters, according to cluster-specific networks; (Bottom) interactions are described by a complex network

Mechanisms of interference

So far I have discussed about interference, in its broader terms, and we have classified the interference mechanism according to the shape of in-

teractions between units. However, this phenomenon may arise as a result of various mechanisms. Indeed, there is a large variety of ways in which the treatment of one unit can affect the outcomes of others: i) the intervention affects the outcome of a given unit and the outcome spreads to other individuals (*outcome spreading*); ii) the treatment received by one individual has an indirect effect in determining the outcome of interfering units (*indirect effect*); iii) treated units tangibly spread the active treatment to interfering units (*treatment diffusion*). I will focus on the latter two mechanisms throughout this introduction: those exploited in the contributions collected in this dissertation. Both mechanisms will be presented under the assumption of neighborhood interference: this assumption limits the interference mechanism in the immediate neighborhood of a given vertex and it remains valid in all articles of the present work [2].

In the presence of neighborhood interference the treatment assignment of a given unit indirectly impacts the potential outcome of their interfering units (Forastiere et al., 2020). The neighborhood of unit i is labeled as $\mathcal{N}_i$, while $\mathcal{N}_{-i}$ denotes the complementary of this set, collecting all those units who do not have a connection with node i. Under first order interference, the no-interference component of the SUTVA is replaced by the Assumption 4.

Assumption 4 (Neighborhood Interference)**.** *There exists a function* $g_i : \{0,1\}^{N_i} \rightarrow \mathbb{G}$*, with* $\mathbb{G} \subseteq \mathbb{R}$ *such that, for all* $\boldsymbol{Z}_{\mathcal{N}_{-i}}, \boldsymbol{Z}^1_{\mathcal{N}_{-i}}$ *and* $\boldsymbol{Z}_{\mathcal{N}_i}, \boldsymbol{Z}^1_{\mathcal{N}_i}$ *with* $g_i(\boldsymbol{Z}_{\mathcal{N}_i}) = g_i(\boldsymbol{Z}^1_{\mathcal{N}_i})$*, we have*

$$Y_i(Z_i, \boldsymbol{Z}_{\mathcal{N}_i}, \boldsymbol{Z}_{\mathcal{N}_{-i}}) = Y_i(Z_i, \boldsymbol{Z}^1_{\mathcal{N}_i}, \boldsymbol{Z}^1_{\mathcal{N}_{-i}}).$$

This assumption admits first-order spillover effects, and they are modelled through the function g_i. This function operates on the treatment assignment vector of units who are in the neighborhood of unit i: it may count the number of treated neighbors, the proportion of treated ties etc. In broader terms, it uniquely summarizes the information coming from the treatment assignment vector of peers. The variable resulting from applying the $g()$ function on the i's neighbors' treatment vec-

[2] The potential outcomes framework under neighborhood interference is presented here and in Chapter 2 while particularly referring to Forastiere et al., 2020

tor $G_i = g_i(Z_{N_i})$ is interpreted as the *neighborhood treatment* and measures the extent of the indirect exposure. The neighborhood treatment is treated as an additional treatment that characterizes the overall exposure of a given unit i. Hence, each individual i is exposed to a *joint treatment* (Z_i, G_i), where the first component Z_i represents the treatment assignment she/he has actually received (individual treatment), and G_i the extent of their indirect exposure to the treatment (neighborhood treatment). Under Neighborhood Interference and Consistency, potential outcomes can be indexed only with respect to the joint treatment, that is $Y_i(Z_i = z, G_i = g) = Y_i(z, g)$. By comparing individual potential outcomes, it is possible to define individual causal effects. In this setting, the treatment effect is defined while controlling for the neighborhood exposure. Moreover, it is feasible to outline individual spillover effects, measuring the extent of the individual susceptibility with respect to the indirect exposure. Formally, the direct effects result from the comparison of the two individual treatment status, while maintaining the neighborhood exposure fixed at a given level g, that is

$$\tau(g) = E[Y_i(Z_i = 1, G_i = g) - Y_i(Z_i = 0, G_i = g)|\ i \in \mathcal{N}_g],$$

where $\mathcal{N}_g = \{i \in N : g \in \mathrm{G}_i\}$ collects units who have a nonzero probability of being characterized by the level g of the individual exposure [3]. On the contrary, spillover effects compare different values of the indirect exposure, while keeping the individual treatment fixed at a certain status z.

$$\delta(z, g) = E[Y_i(Z_i = z, G_i = g) - Y_i(Z_i = z, G_i = 0)|i \in \mathcal{N}_g]$$

Spillovers may also manifest in the tangible spreading of the intervention (An, 2018). Such a phenomenon, known as *treatment diffusion*, may occur when the intervention might be tangibly spread and it is considerable if the outcome becomes observable with a certain lag with respect to the initial exposure. For instance, consider the case of a firm ownership network, in which researches aim to estimate the effect of extra

[3] For instance, suppose the $g_i()$ function counts the number of treated neighbors: an agent who has a number of neighbors that is lower than g has no possibility to be exposed to a given value g of the neighborhood exposure (thus having g treated friends in their neighborhood)

fundings given to some big firms on their future revenues: beneficiaries may tangibly transfer a portion of extra money to the smaller firms they own. A second example could be the study of the effects on smoking of a prevention campaign using informative videos assigned to students.; in this case it might be useful to consider students interactions (for instance their friendship network) and the possibility that students share the link of the video with their friends. If treatment diffusion arises, some individuals who have been originally assigned to the control group, and were not provided with treatment by design, might have actually received the intervention because of a link with treated users. The process generates a relevant statistical issue, concerning the missclassification in the treatment variable (Braun et al., 2014; Grandjean et al., 2004; Lewbel, 2007; McCaffrey et al., 2013; Vanderweele, 2012; Yanagi, 2018). The treatment diffusion process is an important issue in causal inference but, in general, exact information about its evolution are not easy to retrieve. Sometimes the missing information also concerns the underlying diffusion network on which the treatment propagates.

Formally, in the presence of a treatment diffusion process, the initial treatment assignment vector $\boldsymbol{Z}$ does not truly represent the unit-level allocation in the two treatment groups. The real treatment assignment status is correctly represented by $\boldsymbol{Z}'$, which is usually unobservable. The missclassification issue introduces a bias in the estimation of the treatment effect: if diffusion plays a role, the treatment effect that is meaningful to be estimated is τ^*, where

$$\tau^* = \mathbb{E}\left[Y_i(Z_i' = 1) - Y_i(Z_i' = 0)\right],$$

If estimates rule out the possibility of a treatment spreading, then the effect that is evaluated is $\tau = \mathbb{E}\left[Y_i(Z_i = 1) - Y_i(Z_i = 0)\right]$. The estimation bias, given by the difference between τ^* and τ constitutes a relevant issue to be concerned with. Dependencies among the treatment diffusion process and the potential outcomes or heterogeneous treatment effects in the population might further complicate the analysis.

Although the diffusion process is not observed, it is possible to advance simplifying assumptions to characterize the process, by account-

ing for the network information of data. A plausible assumption that can be considered is that the diffusion spreading happens through edges that are characterized by the presence of a treated node and an untreated node. We call these edges *treatment diffusion paths*. Note that, in a directed network the active diffusion paths correspond to the edges that treated units deliver to their neighbors. Figure 13 signals the active diffusion paths in an undirected network. Nodes are colored with respect to their treatment status: *red* nodes identify treated individuals, the *green* color characterize vertexes belonging to the control group (this rule will be followed throughout this work, while dealing with binary interventions).

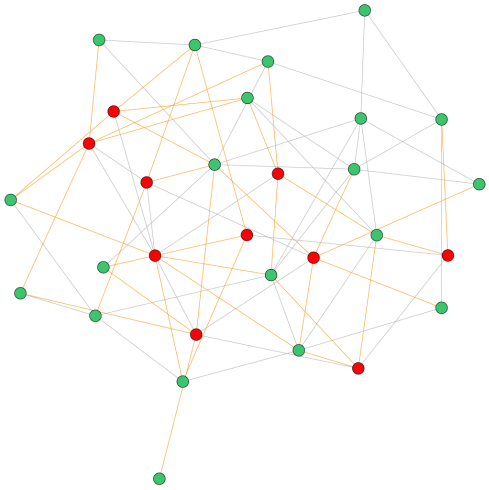
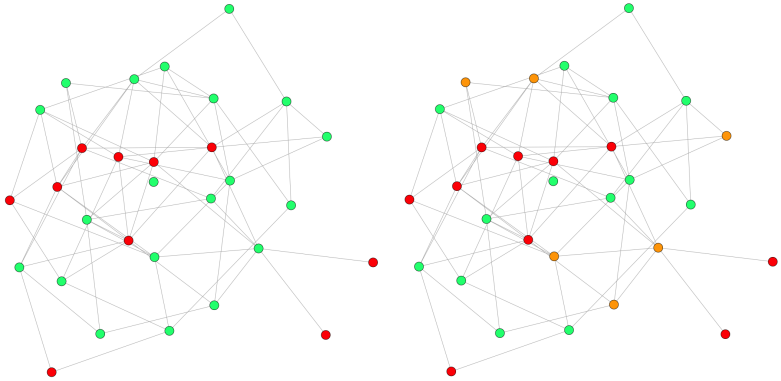


Figure 13: Active treatment diffusion paths in an undirected network

An example of what the diffusion process may generate can be found in Figure 14. Consider a network of units, who have been randomly assigned to either the active treatment or the control treatment. The situation that the policy maker observes right after the treatment assignment (time t) is depicted in the leftsubfigure. At a subsequent point in time, that I denote with t', the situation appears as different (right subfigure):

some of the initially untreated nodes (the *orange* nodes) have actually received the active treatment through diffusion process.



(a) Time t (before the diffusion process) (b) Time t' (after the diffusion process)

Figure 14: Treatment diffusion, before and after the occurrence of a random diffusion process: *red* nodes represent individuals while *green* nodes label untreated units; nodes who were initially untreated but have actually received the treatment due to the diffusion process are colored in *orange*

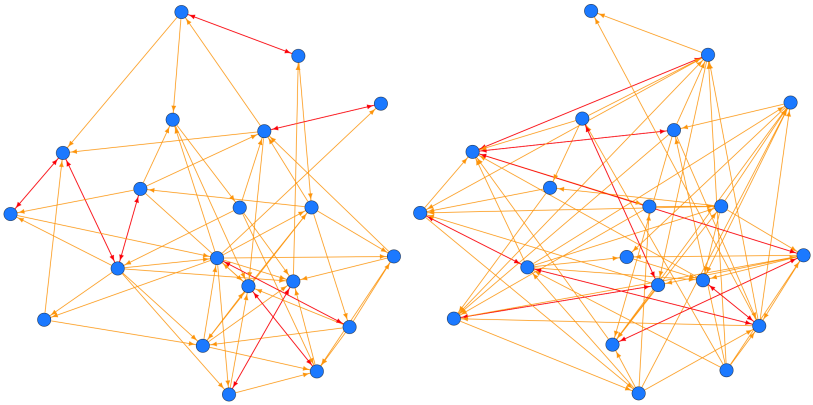
1.3.2 Causal inference for network formation

Given the increasing variability of network data, the interest about interactions among different objects has rapidly grown, and researchers have started to analyze complex networks, by observing their patterns and studying their topological characteristics (Falkowski et al., 2006; Routray et al., 2015). However, the statistical approach towards network does not limit to a passive observation of interactions, but it intends to understand the mechanisms generate them. The analysis of the causes that play a role in the formation of a network structure is called *network formation analysis*. The standard statistical literature about network formation is based on Exponential Random Graph Models (ERGM) (De Stefano & Zaccarin, 2012; P. Wang et al., 2016; Zaccarin & Rivellini, 2010), Stochastic Block

Models (Y. J. Wang & Wong, 1987; Xu & Hero, 2014), Latent Space Models (Sewell & Chen, 2015; Zhao et al., 2017), Additive and Multiplicative Effects models (Hoff, 2015; Minhas et al., 2019)(AMEN) (see Falk and Kosfeld, 2012 and Chandrasekhar, 2016 for a review of network formation models). For instance, ERGMs model the probability that two nodes are linked, with respect to their individual characteristics and to features of the network (i.e attitude towards reciprocity, presence of triangles). However, these approaches are usually seen as descriptive and cannot be used to draw causal information.

Employing the potential outcome framework

To draw such a causal information, it is necessary to implement a proper causal inference analysis, by reworking the potential outcomes framework so to suitably model network data. In this setting, the network itself represents the main outcome of interest. As a consequence the sample over which the analysis is performed consists in dyads and the outcome variable may be the presence, the strength or the direction of a social tie. The treatment variable may be of various kinds, depending on the empirical phenomenon of interest: it may be intrinsically dyadic or it may result from the joint observation of individual treatments (Arpino et al., 2017). Suppose a researcher intends to assess whether the individual participation of a school-organized extra learning course has encouraged friendship relationships among participants: in that case, students are individually assigned to participate to the course, and the dyadic treatment may be signal the simultaneous participation of the students identifying a given dyad. Conversely, while assessing the role of being assigned as deskmates in prompting social ties, one has to deal with a pure dyadic treatment. In the last Chapter of this dissertation we will analyze a dyadic treatment which is actually a mixture of these two sorts. Figure 15 suggests a type of network formation analysis that can be performed by adapting the potential outcomes framework to networks. Here, the researcher intends to assess whether the type of a social tie in a treatment network (it can be either *absent*, *asymmetric* or *symmetric*) affects the likelihood of observing another type of social tie in an outcome network



(a) Treatment Network (b) Outcome Network

Figure 15: Network formation analysis, example: edges are colored according to their type: *red* denotes symmetric edges, while *orange* denote asymmetric ties

Regardless on the nature of the dyadic treatment, the potential outcomes framework for network formation is exploited by considering a dyadic population $\mathcal{D}$ where the single pair (i, j) is characterized by a pair specific treatment Z_{ij}, an observed outcome with Y_{ij}^{obs} and by a vector of covariates $\boldsymbol{X}_{ij}$, which may include both individual and dyadic covariates. All the assumptions and tools of the standard RCM are rephrased in the context of dyads and require additional discussions.

1.4 Outline of the thesis

The next chapters elaborate on the connection between causal inference methods and complex networks from different statistical perspectives. The first three chapters develop a discussion about methodological issues that may arise while estimating the effect of an intervention in the presence of connected units, while the last chapter makes use of the potential outcome framework for estimating both unconditional and conditional dyadic treatment effects in a multi-layer network. In this Section

I summarize the contents discussed in each of these four chapters.

In recent years, researchers have proposed novel tools to account for interference in observational studies. However, this existing works allow for binary treatments only. Chapter 2 addresses this shortcoming and develops a novel methodology for estimating treatment effects in observational studies under network interference, allowing for a multi-valued treatment and an interference structure which is shaped by a weighted network. The estimation strategy is based on a joint multiple generalized propensity score and allows the researcher to estimate direct effects, controlling for both individual and network covariates. The proposed methodology is employed in order to analyze the impact of the national immigration policy on the crime rate. The contribution proposes a multi-valued characterization of political attitudes towards migrants and assumes that the extent to which each country can be influenced by another country is modeled by an appropriate indicator, summarizing their cultural and geographical proximity.

Chapter 3 represents the first attempt to unite these two of the most popular streams of literature on causal inference: the literature about causal effects in the presence of interference and the literature about machine learning methods for detecting and estimating heterogeneous causal effects. Recent studies have pointed out that, in those settings where the effects of a treatment can spill from one unit to its neighbours, assuming no interference can introduce large biases in the estimation of causal effects while also neglecting spillover effects. On the other hand, several recent contributions emphasize the importance of understanding heterogeneity in the treatment effect. When interference takes place, a policy maker may desire to account for the spillover mechanism and to understand the heterogeneity both in the treatment effect and in the spillover effect. Indeed, in some empirical studies it could be useful to determine not only the sub-populations on which the treatment is more effective, but also those who are characterized by highly influential and susceptible individuals. This Chapter proposes a novel machine learning algorithm - named *Network Causal Tree*(NCT) - to assess the heterogeneity of treatment and spillover effects under clustered-network interference.

The performance of the NCT is evaluated in series of Monte Carlo simulations. Additionally, the Chapter provides an application on real-world data from a randomized experiment aimed at assessing the effects of a new weather insurance policy in rural China.

Chapter 4 focuses on the impact of a treatment diffusion process. This issue may arise when the intervention can be tangibly diffused among interacting units and when the outcome variable is observed with a certain lag with respect to the initial exposure. In most scenarios, this process is hidden. For instance, if the intervention of interest is an information campaign realized through a video or a flyer, some treated units might share the treatment with their friends though emails or social networks. If this happens, a specter looms in the horizon: that of missclassifying the treatment variable. In fact, some of the controlled units might have actually received the treatment through diffusion from their treated neighbors. Inspired by a recent experiment studying the effect of various types of school-incentives for prompting students to attend museums, the Chapter proposes a novel approach for dealing with the hidden diffusion process, while also accounting for an incomplete network structure describing links among units. The proposed method addresses the missing links issue by implementing a machine learning algorithm based on random forests to multiply predict whether a missing tie is present or not. Subsequently, the chapter develops a sensitivity analysis for assessing the extent to which the estimates vary depending on the unknown diffusion process, while also accounting for uncertainty in the network structure. This procedure simulates various diffusion scenarios within a plausible range of sensitivity parameters and compares the treatment effect which is estimated in each scenario against the one obtained while ignoring the diffusion process.

Chapter 5 develops an estimator for causal effects of the existence of links in a "treatment" network on the formation of links in an "outcome network," with both networks being directed. The approach is based on the definition of "conditional causal effects" concerning the effect on the presence, symmetry and direction of links; the estimator is based on an extension of the propensity score matching approach to simultaneously

handle multi-valued treatments, network data and conditional effects. The chapter showcases the methodological framework while assessing the effect of lobbying pressure on legislators on their legislative collaborations in the US House of Representatives. Firms and corporate companies support the electoral campaign of political candidates running for a seat, trying to influence their future political agenda. Politicians willing to achieve their political goals are encouraged to collaborate with those colleagues who are pushed by common objectives. If these two mechanisms hold, we expect that two politicians who are supported by the same lobbies are likely to collaborate. The work measures the extent of this effect, while also defining conditional dyadic effects for investigating the specific causal mechanisms playing a role in this setting.

Finally, Chapter 6 concludes with a discussion of results and with potential lines for further research in the field of causal inference and complex networks.

Chapter 2

Modeling Network Interference with Multi-valued Treatments: the Causal Effect of Immigration Policies on Crime Rates

This Chapter is a joint work with my supervisors Prof. Irene Crimaldi and Prof. Fabrizia Mealli, and Prof. Laura Forastiere. The full text of the article is also available from the arXiv repository, preprint number arXiv:2003.10525v3.
The methodology presented in this Chapter is also employed by Irene Crimaldi, Laura Forastiere, Fabrizia Mealli and Costanza Tortú in the article named "The Causal Effect of Immigration Policy on Income Inequality ," forthcoming in *Proceedings of the conference of the Italian Statistical Society* (SIS 2020).

2.1 Introduction

2.1.1 Motivation

Policy evaluation studies aim to assess the effect of an intervention. Social sciences such as economics or political science often evaluate complex interventions, which have not been randomly assigned in the population. In some real-world settings, the analysis can be further complicated by the presence of interference between units. This phenomenon occurs because both economic and social agents are interconnected. Firms are connected by a wide mixture of juridical or commercial relationships including trading links, ownership or control ties and strategic alliances (Reinert et al., 2009). On the other side, individuals also interact through various mechanisms involving friendship or parental links, working collaborations or informative communications. In addition, even political entities are linked by way of explicit or velled agreements, or according to their specific geographical and cultural collocation with respect to a reference environment. These relations are depicted by a *network*: the observed nodes are the elements of the population of interest, while network links represent the relations between them.

Causal inference on a population of agents who are connected through a network faces some statistical challenges, including how to take into account the spillover mechanism that may arise. The typical causal inference framework (Rubin, 1980) relies on a key assumption, called *Stable Unit Treatment Value Assumption (SUTVA)*, which rules out the presence of interference among units. However, if agents are linked, experiments as well as observational studies may be affected by the presence of *interference*, which formally occurs when the treatment of one unit has an effect on the response of other units (in addition to the unit's own outcome) (Cox, 1958). In the presence of interference, the causal effect of a treatment on one unit may be altered by the treatment received by other interfering units. For example, incentives targeted to some firms or companies may also benefit all those firms that are linked to them, according to juridical or economic relationships. In addition, policies implemented by single administrative entities also affect the outcomes

of interfering territories. Dealing with interference is of paramount importance: wrongly assuming SUTVA can introduce a significant bias in the estimates and, consequently, lead to deceptive conclusions about the real effect of an intervention.

2.1.2 Related Works

For this reason, in recent years, a growing community of statisticians has started reasoning about interconnected units, developing novel methods and techniques which allow to account for interference in causal inference studies. The existing works extend the standard framework to include the network information in the definition of individual potential outcomes. Most of these works examine interference in *randomized trials*. The limitations of the present statistical tools in dealing with possible dependencies among units have been first pointed out by Rosenbaum, 2007, who has also developed non-parametric tests to evaluate treatment and spillover effects in the presence of interference. This last study was extended, after a few years, by Aronow, 2012, who presented a novel method to detect interference, and by Bowers et al., 2013, who proposed tools to model various dependency scenarios, also showing how to test hypotheses about causal effects according to the specific model that is supposed to depict interference. Recently, Aronow and Samii, 2017 rearranged the Horwitz-Thompson estimator allowing for the presence of interference to obtain unbiased estimators for all the effects of interest, main and spillovers. Athey et al., 2018b computed exact p-values for a variety of sharp null hypotheses about treatment effect in an experimental design where units are connected in an observed network. Interference may even play a role in the *design of experiments*. Having proved that wrongly assuming SUTVA leads to biased results, Eckles et al., 2017 formalized a model of experiments in networks, proposing novel techniques for reducing this bias directly through the experimental design itself. Some other works focus on a particular type of interference known as *partial (clustered) interference*, where units belong to exogenous groups and the spillover mechanism can occur only within clusters. The term

"partial" is used here to counterpoise this scheme of clustered dependencies with the "general" interference scenario, where units interact according to a network. The partial interference assumption was formally introduced by Sobel, 2006b and it was further advanced also by Hudgens and Halloran, 2008, who investigated the role of interference in the spreading of infectious diseases, where the probability that a person becomes infected is lower if the proportion of vaccinated individuals in his group is high (Basse & Feller, 2018). Moreover, Barkley et al., 2017 addressed the issue of a possible treatment selection among connected individuals and proposed causal estimands allowing for clustered dependence in the treatment selection (Papadogeorgou et al., 2019). There are just a few articles that explicitly deal with general interference in *observational studies*. For instance, Hong and Raudenbush, 2006 evaluated the policy of retaining low-achieving children in kindergarten rather than promoting them to first grade, using a multilevel propensity score model. van der Laan, 2014 and Sofrygin and van der Laan, 2017 proposed a targeted minimum loss-based estimation (TMLE) estimator. A propensity score approach under spillover effects was first introduced by Forastiere et al., 2020, who presented a reworked formalization of the standard propensity score, named joint propensity score (JPS), with the aim of estimating the dose-response function in presence of interference. This work analyzes a binary treatment and models interference through an observed binary network. This last framework was employed by Del Prete et al., 2019 to explore trade distortions in agricultural markets: here, the authors rearranged the JPS formulation in order to model a continuous individual treatment which in turn leads to a continuous characterization of the indirect exposure to the treatment.

2.1.3 Contribution

The existing statistical literature tackling interference in observational studies deals with binary or continuous treatments only. However, many policy evaluation studies involve more complex treatments, as, for example, treatments which are defined over more than two categories, known

as multi-valued treatments. *Multi-valued treatments* are highly diffused in nature. They are commonly used when the empirical aim consists in comparing various characterizations of an intervention and, above all, they are particularly employed in studies yearning to get the picture of complex and many-faceted phenomena, which may vary across multiple dimensions. For instance, evaluating the impact of different political attitudes towards puzzling macro-themes (immigration, national healthcare, economy) often calls for a multi-valued approach and requires also to account for interference, since the treatment may spill over to different political entities. Since our empirical attempt is to evaluate the impact of immigration policy, we expand the theoretical framework proposed by Forastiere et al., 2020 to the case of an individual multi-valued treatment, in observational studies.

Generalization of the standard techniques (such as subclassification and propensity score methods) for binary treatments to the multi-valued scenario is not straight-forward and requires additional assumptions (Lopez, Gutman, et al., 2017; S. Yang et al., 2016; Linden et al., 2016). The methodological approach becomes even more complicated if we decide to allow for the presence of interference, relaxing SUTVA and allowing for first-order spillover effects. The key idea is that under a multi-valued treatment, in the presence of interference, each unit is individually assigned to a treatment level and, simultaneously, they can be exposed to all the treatment levels, due to the interaction with their neighbors. Therefore, units experiment a *multiple neighborhood exposure*, where each of their neighbors contributes in increasing the exposure to her own individual treatment level. In addition, the multiple neighborhood exposure mapping accounts for weights, which quantify the extent of dependencies, if they are observable. Weighted networks are widely spread in real-world data. For instance, networks of transactions between entities are usually enriched by the information about transactions' amount, social networks sometimes are coupled with the strength of friendship between units, scientific collaborations networks often provide the number of collaborations, political networks frequently measure the strength of connections between administrative and political entities by specific indicators. In

settings with multi-valued treatments and weighted network, each unit is exposed to an individual treatment, which is categorical with a given number of categories, and to a neighborhood treatment, which is a multivariate continuous variable that measures the unit's exposure to all treatment levels, resulting from the interaction of their neighbors and given the strength of these interactions. Since we move in an observational study setting, where neither the individual treatment nor the neighborhood treatment are randomly assigned in the population, we propose an estimation strategy based on the usage of an extended version of the joint propensity score proposed by Forastiere et al., 2020. We employ the JPS approach for exploiting the direct effect of a multi-valued intervention, while accounting for a multivariate continuous indirect exposure to the treatment [1]. Our definition of propensity score, that we call *Joint Multiple Generalized Propensity Score (JMGPS)*, allows to handle a multi-valued treatment and a multiple neighborhood exposure. The JMGPS is a type of generalized propensity score (Hirano & Imbens, 2004), where the estimation strategy relies on a three-stage approach: i) we first assume a parametric distribution for both the individual and the neighborhood treatment and for the outcome variable; ii) for all the possible values that the joint treatment can assume, we use these models to predict missing potential outcomes; iii) we estimate the effects of interest by comparing potential outcomes, and we use bootstrap to compute the estimated standard errors.

We make use of this methodology for the analysis of the causal effect of *immigration policies* on *crime rates*. In the last decades, the relevance of the immigration process has rapidly grown and shew the way to the spreading of a wide and open debate about the effects of migration. Some political parties, single politicians and citizens all around the world do believe that immigration represents a risk for national identity and, moreover, that it leads to a lower public safety. Consequently, they support governments that implement restrictive immigration poli-

[1] The proposed estimation strategy finds its roots in Del Prete et al., 2019, but intends to model the effect of a multi-valued intervention (instead of handling a continuous individual treatment), in the presence of a multivariate continuous exposure (instead of accounting for a univariate continuous exposure

cies. However, the causal effect of immigration policies on crime or social conditions in general has not been tested yet. In particular, there are not quantitative studies that involve and compare many countries, over a wide time frame. We analyze policies using the *IMPIC* (Immigration Policies in Comparison) dataset that numerically measures all the immigration policies that have been implemented in the OECD countries from 1980 and 2010 in terms of restrictiveness. We include in the analysis 22 OECD countries that are located in Europe over the whole time frame covered by the IMPIC dataset. Our purpose is to investigate the impact of a national towards migrants on the *crime rate*. In this application, the treatment of interest represents the restrictiveness of immigration policies, which is measured in the IMPIC Dataset through the evaluation of a series of single policies. Each policy refers to *regulations* or *control* protocols. The former are all the binding legal provisions that create or constrain rights (Helbling et al., 2017), while the latter refer to the directives that have been adopted with the aim of monitoring whether the regulations are observed. Therefore, by aggregating items referring to these two political dimensions, we obtain two indicators measuring the country-year restrictiveness towards migrants, with respect to regulations and control mechanisms separately. Using this information, we define a multi-valued treatment by looking at the joint value of the two indicators, for each country-year profile. In this empirical setting, SUTVA is unlikely to hold. The political strategy towards migrants that a single country decides to implement may also affect crime rate of other countries. The possible spillover effect of the adopted political attitude towards immigration arises because migrants try to avoid countries with highly restrictive laws, and tend to move to states that appear to be more welcoming. Since migrants tend to move to countries with specific characteristics of their choice, the extent to which each country is affected by other countries' policies depends on their level of similarity. Following this intuition, we derive an indicator summarizing the main factors which may prompt the spillover mechanism. These factors refer to various measures of similarity, which we reasonably believe to be the primary mechanisms driving interference. Specifically, this index, that we

call *Influence Index* (II), gives a measure of potential interference between each pair of countries at a given year and combines information about *geographical proximity* and *cultural similarity*, which in turn are summarized by specific indicators.

This work is organized as follows. In Section 2.2 we focus on methodology. We first summarize the existing causal inference framework under interference in observational studies and then we present our methodological novelties: we introduce a multi-valued treatment and we propose a novel tool that allows one to model the neighborhood exposure in the presence of multi-valued treatment and weighted interference. We introduce the joint generalized multiple propensity score and we illustrate the estimation strategy. In Section 2.3 we motivate the importance of the empirical application, giving a broad overview of the existing literature and briefly describing data. Moreover, we give a more detailed characterization of the Influence Index and we provide a deeper explanation on the definition of treatment nominal categories. In Section 2.4 we present the main empirical results. Then, in the appendix, we collect the proofs of the theoretical propositions we present in Section 2.2, we give the precise definition of II also adding further details about the definition of the neighborhood treatment variable, and we present the detailed results of all the models we implement, reporting some descriptives and checking the robustness of the main findings with respect to alternative definitions of the treatment variable.

2.2 Methodology

In this section we explain the main methodological developments. We start from the existing causal inference framework under interference and then we present the novel approach for multi-valued treatments (Subsection 2.2.1). Finally, we define the Joint Multiple Generalized Propensity Score (Subsection 2.2.2) together with its properties and we propose the estimation strategy (Subsection 2.2.3).

2.2.1 Causal Inference Under Network Interference

The main scope of causal inference is estimating the effect of a treatment on some outcome variable in a population of units. Let us consider a sample $\mathcal{N}$ composed of N units. Denote as K the number of treatment levels and let $Z_i \in \{1, \dots, K\}$ be a categorical variable representing the treatment assigned to unit i and Y_i^{obs} the observed outcome for the same unit. By $\mathbf{Z}$ and $\mathbf{Y}^{obs}$ we denote the corresponding vectors of the whole sample $\mathcal{N}$. Moreover, $\mathbf{X}_i$ denotes a vector of P covariates (or pre-treatment variables) that are not influenced by the treatment assignment. Following Rubin, 1974, 1980, we postulate, for each unit, the existence of K potential outcomes, one for each treatment vector, $Y_i(\mathbf{Z})$. Most causal inference relies on the Stable Units Treatment Assumption (SUTVA) (Rubin, 1986). SUTVA consists of two different components: (i) the Individualistic Treatment Response (ITR) (C. F. Manski, 2013) (or *no interference*) assumption, which states that there is no interference between units, each unit's potential outcomes are defined only by the unit's own treatment; (ii) the *consistency* (or *no multiple version of the treatment*) assumption, which states that there are no different versions of the treatment levels. As a consequence, under SUTVA, potential outcomes can be indexed only by Z_i -i.e $Y_i(Z_i)$- and the observed outcome is the one corresponding to the treatment that each unit i has actually received: $Y_i^{obs} = Y_i(Z_i)$.

SUTVA completely rules out the presence of interference among units. However, in many real situations, this no-interference assumption is violated. This phenomenon can occur in various and heterogeneous frameworks. For instance, in economics, firms assigned to a program of incentives can be affected by incentives received by other firms. In epidemics, vaccines are known to benefit the whole community, including unprotected individuals, because they reduce the reservoir of infection and the infectiousness. Finally, in political sciences, policies implemented in some administrative regions may have an effect also on neighboring territories. All these examples refer to empirical situations in which one unit's outcome may be influenced by other units' treatment level.

When the spillover mechanism comes into play, wrongly assuming SUTVA leads to biased results and, consequently, to inaccurate or even misleading conclusions about the effects of interest. In order to model interference, we must look at the relationships between units. We consider an observed undirected network $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ is the set of nodes (the population of interest) and $\mathcal{E}$ represents the set of edges indicating links between nodes. For each node i, we identify a partition of $\mathcal{N}$ into two subsets: i) the neighborhood of node i, $\mathcal{N}_i$, that includes all the nodes j with a link with node i, $i \leftrightarrow j$, and we denote by N_i the cardinality of $\mathcal{N}_i$; ii) the No-Neighborhood of node i, $\mathcal{N}_{-i}$, including all the nodes j without a link with node i, $i \nleftrightarrow j$. According to these partitions, we define, for each node i, the following partitions of the treatment vector and of the outcome vector, $(Z_i, \mathbf{Z}_{\mathcal{N}_i}, \mathbf{Z}_{\mathcal{N}_{-i}})$, $(Y_i, \mathbf{Y}_{\mathcal{N}_i}, \mathbf{Y}_{\mathcal{N}_{-i}})$. Figure 16 shows the neighborhood of a given node.

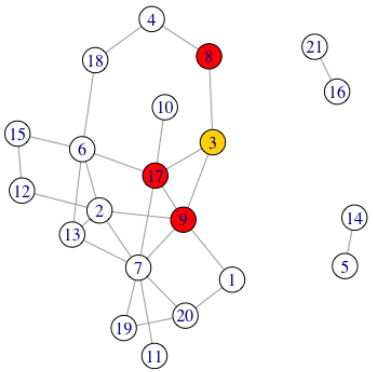


Figure 16: Neighborhood of a given node: the figure shows a given unit (yellow-colored unit) and highlights his own neighbors (red-colored units), in a population of connected agents.

Admitting network interference in the analysis implies the replacement of SUTVA by an assumption on the interference structure. Forastiere et al., 2020 make a neighborhood interference assumption, which allows for the existence of first-order spillover effects between neighbors, in the

context of binary treatments. More precisely, using the notation $Y_i(\boldsymbol{Z})$ for potential outcomes of unit i, we have:

Assumption 5 (Stable Unit Treatment on Neighborhood Value Assumption (SUTNVA))**.** *SUTNVA is constituted by two components:*

1. *No Multiple Versions of Treatment (*Consistency*):* $Y_i(\boldsymbol{Z}) = Y_i(\boldsymbol{Z}^1) \ \ \forall \boldsymbol{Z}, \boldsymbol{Z}^1$ *such that* $\boldsymbol{Z} = \boldsymbol{Z}^1$*, that is, the mechanism used to assign the treatments does not matter.*

2. *Neighborhood Interference: There exists a function* $g_i : \{0,1\}^{N_i} \to \mathbb{G}$*, with* $\mathbb{G} \subseteq \mathbb{R}$ *such that, for all* $\boldsymbol{Z}_{\mathcal{N}_{-i}}, \boldsymbol{Z}^1_{\mathcal{N}_{-i}}$ *and* $\boldsymbol{Z}_{\mathcal{N}_i}, \boldsymbol{Z}^1_{\mathcal{N}_i}$ *with* $g_i(\boldsymbol{Z}_{\mathcal{N}_i}) = g_i(\boldsymbol{Z}^1_{\mathcal{N}_i})$*, we have*

$$Y_i(Z_i, \boldsymbol{Z}_{\mathcal{N}_i}, \boldsymbol{Z}_{\mathcal{N}_{-i}}) = Y_i(Z_i, \boldsymbol{Z}^1_{\mathcal{N}_i}, \boldsymbol{Z}^1_{\mathcal{N}_{-i}}).$$

   This assumption basically states that there is interference and it is modelled by the function g_i*.*

The variable $G_i = g_i(\boldsymbol{Z}_{\mathcal{N}_i})$, called *neighborhood treatment*, represents the unit's exposure to the treatment, due to the influence of his neighbors. The function g_i can be defined in many different ways, according to the interference mechanism that is assumed to take place. For instance, it can simply count the number of treated neighbors or it can measure the proportion of treated neighbors. Note that under SUTNVA interference is assumed to arise only from neighborhood of each unit and that any higher order interference is completely ruled out. This means that unit i is not influenced by units other than their neighbors. This restriction over the interference structure may appear to be strong in some scenarios, but it seems to be plausible in many empirical applications.

In many real-world applications treatments are implicitly or explicitly multi-valued. In epidemics, researchers are interested in comparing between drugs (Linden et al., 2016). In economics, firms are exposed to different types of incentives. In training programs, participants receive different types of coaching (Cattaneo, 2010). Finally, political scientists evaluate political strategies towards highly complex and multi-faceted issues which involve different sub-fields. In such scenarios, a common practice is to collapse the multi-valued treatment into a binary variable,

but this approach implies a relevant loss in terms of information and it prevents the possibility of capturing differential effects across treatment levels (Cattaneo, 2010). For this reason, researchers have started to study how to extract causal information under multi-valued treatments, developing novel assumptions and techniques that extend to the multi-valued scenario standard causal inference methods as matching, subclassification, inverse probability weighting on the propensity score (Lopez, Gutman, et al., 2017). However, no existing work suggests how to deal with interference in the multi-valued scenario. In this work we fill in this gap.

Under first-order spillover effects, each unit is exposed to a neighborhood treatment, defined as a numerical synthesis neighbors' treatment status. In the binary setting, this synthesis is usually expressed by a single value, $G_i \in \mathbb{G} \subseteq \mathbb{R}$, while when the individual treatment is defined by multiple categories this definition of G_i is too simplistic. Categorical treatments imply a more complex definition of the neighborhood treatment exposure, as the neighborhood treatment must summarize the individual network exposure to *each* treatment level. The mathematical tool that we introduce to model the neighborhood treatment under multi-valued individual treatment is the *Neighborhood Treatment Exposure Matrix* $\boldsymbol{G}$:

Definition 2 (Neighborhood Treatment Exposure Matrix (NTEM), $\boldsymbol{G}$)**.** *The NTEM is an $N \times K$ matrix $\boldsymbol{G}$ that collects the unit neighborhood exposure to all the treatment levels:*

$$\boldsymbol{G} = \begin{pmatrix} G_{1,1} & \dots & G_{1,z} & \dots & G_{1,K} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ G_{i,1} & \dots & G_{i,z} & \dots & G_{i,K} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ G_{N,1} & \dots & G_{N,z} & \dots & G_{N,K} \end{pmatrix}.$$

Each element $G_{i,z} \in \mathbb{G} \subseteq \mathbb{R}$ indicates the exposure of unit i to the treatment level $z \in \{1, \dots, K\}$. Each row is the neighborhood treatment vector for the unit i, $\boldsymbol{G}_i \in \mathbb{G}^K \subseteq \mathbb{R}^K$. Therefore, the neighborhood treatment is not a scalar measure, as in the binary treatment setting. It is instead a K-dimensional vector whose components describe the unit's neighborhood exposure to each treatment level.

The Neighborhood Treatment Exposure Matrix is essential for modelling a multi-valued indirect exposure to the treatment. A discrete scalar characterization of that exposure is not feasible in the multi-valued scenario, as it does not allow to capture the categorical distribution of the individual treatment over the neighborhood of each unit.

In the multi-valued scenario, the first component of the recalled SUTNVA for binary treatments, i.e. the no multiple versions of treatment assumption, is confirmed as stated above; while the second component, i.e. the neighborhood interference assumption, is here replaced by a more general assumption, which handles the spillover mechanism generated by a multi-valued individual treatment:

Assumption 6 (Multiple Neighborhood Interference)**.** *There exists a function* $g_i : \{1, \dots, K\}^{N_i} \to \mathbb{G}^K$*, with* $\mathbb{G}^K \subseteq \mathbb{R}^K$ *, such that, for all* $\boldsymbol{Z}_{\mathcal{N}_{-i}}, \boldsymbol{Z}^1_{\mathcal{N}_{-i}}$ *and* $\boldsymbol{Z}_{\mathcal{N}_i}, \boldsymbol{Z}^1_{\mathcal{N}_i}$ *with* $g_i(\boldsymbol{Z}_{\mathcal{N}_i}) = g_i(\boldsymbol{Z}^1_{\mathcal{N}_i})$*, we have*

$$Y_i(Z_i, \boldsymbol{Z}_{\mathcal{N}_i}, \boldsymbol{Z}_{\mathcal{N}_{-i}}) = Y_i(Z_i, \boldsymbol{Z}^1_{\mathcal{N}_i}, \boldsymbol{Z}^1_{\mathcal{N}_{-i}}).$$

This assumption states that interference is modelled by the function g_i *(with components* $g_{i,z}$*,* $z \in \{1, \dots, K\}$*) which maps the neighborhood exposure of unit* i *over a* K*-variate domain, that is* $\boldsymbol{G}_i = g_i(\boldsymbol{Z}_{\mathcal{N}_i})$*.*

If interference is modeled through a weighted network, the function g_i must take into account the weights, I_{ij}, measuring the strength of the link between i and the neighbor j. For instance, given an individual treatment with K categories, we can set $G_{i,z} = \sum_{j \in \mathcal{N}_i} I_{ij}\delta_{zj}$, where δ_{zj} is a dummy variable that equals 1 if $Z_j = z$ and 0 otherwise. Each unit i is exposed to a *joint treatment* $(Z_i, \boldsymbol{G}_i)$: the *individual treatment* Z_i, which is a categorical variable with K levels, and the *neighborhood multi-treatment* $\boldsymbol{G}_i$, which is a K-variate variable. Hence, potential outcomes, for each unit i, are indexed by the joint treatment: $Y_i(Z_i, \boldsymbol{G}_i) = Y_i(Z_i = z, \boldsymbol{G}_i = \boldsymbol{g})$. The observed outcome is the one corresponding to the actual joint treatment each unit is exposed to: $Y_i^{obs} = Y_i(Z_i, \boldsymbol{G}_i)$.

Regarding the effects of interest, the number of the possible comparisons is $\binom{K}{2} = \frac{K!}{(K-2)!2!}$. Under the multi-valued individual treatment, the direct effect of a given treatment z' with respect to the treatment z,

keeping the neighborhood treatment as fixed, can be expressed as

$$\tau_{z'z}(\boldsymbol{g}) = E\big[Y_i(z', \mathbf{g}) - Y_i(z, \mathbf{g})\big]. \tag{2.1}$$

This quantity represents the individual causal effect of a direct exposure, when the neighborhood treatment is set to $\boldsymbol{g}$. The overall main effect can be define averaging the individual treatment effect over the multivariate probability distribution of the neighborhood treatment, that is

$$\tau_{z'z} = \sum_{\boldsymbol{g} \in \mathbb{G}^K} \tau_{z'z}(\boldsymbol{g}) P(\boldsymbol{G}_i = \boldsymbol{g}). \tag{2.2}$$

2.2.2 Joint Multiple Generalized Propensity Score (JMGPS)

In this work, we focus on observational studies, where neither the individual nor the neighborhood treatment are randomly assigned in the population. The general strategy in observational studies is to control for baseline covariates such that, conditioning on them, the treatment assignment becomes as good as random. In other words, we can exclude any dependence between treatment variable and potential outcomes. This assumption is known as *unxonfoundedness* (Rosenbaum & Rubin, 1983b). In some empirical applications, the number and nature of covariates makes it hard to control for all of them without relying on strong parametric assumptions and, extrapolating in these settings, researchers, instead of conditioning on the set of covariates, prefer to work with a scalar synthesis of them, called *propensity score* (Rosenbaum & Rubin, 1983b). In the binary treatment setting with no interference, propensity score is defined as the conditional probability of receiving the treatment, given the values of the covariates. If the unxonfoundedness assumption is valid when conditioning on individual covariates, it remains valid when conditioning on the propensity score. Using this approach, researchers benefit from a relevant dimensionality reduction in the analysis. This general approach, which is well grounded in the standard causal inference literature, can be extended to the setting with multi-valued treatment and interference. Here the unxonfoundedness assumption must be related to the joint treatment and the joint potential

outcomes. Following the motivations proposed by S. Yang et al., 2016, we rely on the weaker version of unxonfoundedness with respect to the individual multi-valued treatment. Hence, instead of considering the actual multi-valued treatment variable Z_i, we refer to K *treatment indicator variables* representing the presence (or absence) of a given treatment level z, $D_i(z)$. Thus, we advance the following assumption:

Assumption 7 (Weak Unconfoundedness of the Joint Treatment)**.**

$$P(D_i(z) = 1, \boldsymbol{G}_i = \boldsymbol{g} \,|\, Y_i(z, \boldsymbol{g}), \boldsymbol{X}_i) = P(D_i(z) = 1, \boldsymbol{G}_i = \boldsymbol{g} \,|\, \boldsymbol{X}_i)$$
$$\forall z \in \{1, \ldots, K\} \;\; \forall \boldsymbol{g} \in \mathbb{G}^K.$$

Note that, in presence of interference, $\boldsymbol{X}_i$ can include purely individual covariates as well as neighborhood covariates. From now on, we denote as $\mathbf{X}_i^{ind}$ the individual covariates and as $\mathbf{X}_i^{neigh}$ the neighborhood covariates.

In the presence of interference, the propensity score is the joint probability of receiving a value z of the individual treatment and, simultaneously, being exposed to a value $\boldsymbol{g}$ of the neighborhood treatment, given the unit's baseline covariates. Forastiere et al., 2020 formally introduced propensity score under network interference in the case of a binary treatment. We expand their definition allowing for a multi-valued individual treatment and a multivariate neighborhood treatment. Therefore, we introduce the *Joint Multiple Generalized Propensity score (JMGPS)* as follows:

Definition 3 (Joint Multiple Generalized Propensity Score (JMGPS))**.** *The Joint Multiple Generalized Propensity Score (JMGPS), labelled as $\psi(z, \boldsymbol{g}, \boldsymbol{x})$, is the probability of being jointly exposed to a K-variate individual treatment equal to z and to a K-dimensional neighborhood treatment equal to $\boldsymbol{g}$, conditioning on baseline covariates.*

$$\psi(z, \boldsymbol{g}; \boldsymbol{x}) = P(Z_i = z, \boldsymbol{G}_i = \boldsymbol{g} | \boldsymbol{X}_i = \boldsymbol{x}) \tag{2.3}$$

As the standard propensity score, the Joint Multiple Generalized Propensity Score is a balancing score, that is, it guarantees balance with respect to neighborhood and individual covariates. JMGPS has the following properties.

Proposition 1 (Balancing Property of JMGPS). *The joint propensity score is a balancing score, that is*

$$P(Z_i = z, \boldsymbol{G}_i = \boldsymbol{g}|\boldsymbol{X}_i) = P(Z_i = z, \boldsymbol{G}_i = \boldsymbol{g}|\psi(z, \boldsymbol{g}; \boldsymbol{X}_i)),$$
$$\forall\, z \in \{1, \dots, K\} \text{ and } \forall\, \boldsymbol{g} \in \mathbb{G}^K$$

Proof in Appendix A.1.1

Furthermore, conditioning on JMGPS, we can exclude any dependency between the treatment variable and potential outcomes.

Proposition 2 (Conditional Unconfoundedness of $D_i(z)$ and $\boldsymbol{G}_i$ given JMGPS). *Under Assumption 7, for all $z \in \{1, \dots, K\}$ and $\boldsymbol{g} \in \mathbb{G}^K$*

$$P(D_i(z) = 1,\, \boldsymbol{G}_i = \boldsymbol{g} \,|\, Y_i(z, \boldsymbol{g}), \psi(z, \boldsymbol{g}; \boldsymbol{X}_i)) = P(D_i(z) = 1,\, \boldsymbol{G}_i = \boldsymbol{g} \,|\, \psi(z, \boldsymbol{g}; \boldsymbol{X}_i)).$$

Proof in Appendix A.1.2

Following Forastiere et al., 2020, we rely on the factorization of the joint propensity score in neighborhood propensity score and individual propensity score.

Definition 4 (Factorization of the Joint Multiple Generalized Propensity score (JMGPS)). *JMGPS can be factorized as follows*

$$\begin{aligned}\psi(z, \boldsymbol{g}; \boldsymbol{x}) &= P(Z_i = z, \boldsymbol{G}_i = \boldsymbol{g}|\boldsymbol{X}_i = \boldsymbol{x}) \\ &= P(\boldsymbol{G}_i = \boldsymbol{g}|Z_i = z, \boldsymbol{X}_i^g = \boldsymbol{x}^g) P(Z_i = z|\boldsymbol{X}_i^z = \boldsymbol{x}^z) \\ &= \lambda(\boldsymbol{g}; z, \boldsymbol{x}^g)\phi(z; \boldsymbol{x}^z),\end{aligned}$$

where $\lambda(\boldsymbol{g}; z, \boldsymbol{x}^g)$ is the neighborhood propensity score and $\phi(z; \boldsymbol{x}^z)$ is the individual propensity score. $\boldsymbol{X}_i^z$ and $\boldsymbol{X}_i^g$ are vectors collecting covariates that affect the individual and the neighborhood treatment, respectively. Note that the two sets corresponding to the covariates included in $\boldsymbol{X}_i^z$ and $\boldsymbol{X}_i^g$ may differ. In particular, $\boldsymbol{X}_i^g$, can collect individual covariates as well as neighborhood covariates, while $\boldsymbol{X}_i^z$ includes individual variables only.

Using the factorization that we have just presented, we illustrate another key property of JMGPS.

Proposition 3 (Conditional Unconfoundedness of $D_i(z)$ and $\boldsymbol{G}_i$ given individual and neighborhood propensity scores)**.** *Under Assumption 7, for all* $z \in \{1, \ldots, K\}$ *and* $\boldsymbol{g} \in \mathbb{G}^K$*, we have*

$$P(D_i(z) = 1, \boldsymbol{G}_i = \boldsymbol{g} \,|\, Y_i(z, \boldsymbol{g}), \phi(z; \boldsymbol{X}_i^z), \lambda(\boldsymbol{g}; z, \boldsymbol{X}_i)) =$$
$$P(D_i(z) = 1, \boldsymbol{G}_i = \boldsymbol{g} \,|\, \phi(z; \boldsymbol{X}_i^z), \lambda(\boldsymbol{g}; z, \boldsymbol{X}_i)).$$

Proof in Appendix A.1.3

This property indicates that conditioning on the two components separately still guarantees the validity of the conditional unxonfoundedness property.

2.2.3 Estimation Procedure

The JMGPS is the fundamental element of the estimation procedure that we propose here in this section. Both its components can be seen as peculiar characterizations of the generalized propensity score proposed by Hirano and Imbens, 2004. This procedure follows a parametric approach and imputes missing potential outcomes for all configurations of the joint treatment and then compares them to estimate the direct effects of interest [2]. Standard errors and confidence intervals are computed using bootstrap methods. The proposed estimation strategy can be summarized in three main steps.

1) Model treatment and outcome variables

(1.a) Assume a distribution for Z_i, $\boldsymbol{G}_i$ and Y_i.

Formally:

$$Z_i \sim f^z(\boldsymbol{X}_i^z; \theta^z),$$
$$\boldsymbol{G}_i \sim f^g(Z_i, \boldsymbol{X}_i^g; \theta^g),$$
$$Y_i(z, \boldsymbol{g}) \sim f^y(z, \boldsymbol{g}, \phi(z; \boldsymbol{X}_i), \lambda(\boldsymbol{g}; z, \boldsymbol{X}_i); \theta^y).$$

Of course, the multi-valued characterization of Z_i demands for the definition of a statistical model for categorical responses, with respect to the

[2] The proposed estimation strategy find its roots in Del Prete et al., 2019, whose approach has been rearranged to model a multi-valued treatment and a multiple neighborhood exposure (Del Prete et al., 2019 deal with a continuous individual treatment and with a continuous and univariate neighborhood exposure.)

individual propensity score model. Furthermore, $\boldsymbol{G}_i$ requires the definition of a multivariate model Note that this stage requires to rely on some parametric assumptions about the distribution of Z_i and $\boldsymbol{G}_i$. The more credible are these assumptions, the more accurate are the estimates of the effects of interest. In this sense, a prior knowledge about the empirical phenomenon that is object of study is of paramount importance. If this knowledge lacks or it is imprecise, the researcher should perform the whole procedure under various parametric assumptions, then checking the robustness of results with respect to these assumptions.

(1.b) Predict actual individual and neighborhood propensity score

Estimate the parameters θ^z and θ^g of the models for Z_i and $\boldsymbol{G}_i$; Use the estimated parameters in Step 1, $\widehat{\theta^z}$ and $\widehat{\theta^g}$, to predict for each unit $i \in \mathcal{N}$ the actual individual propensity score and the actual neighborhood propensity score, that is, the probabilities of being exposed to the individual treatment and the multivariate neighborhood treatment they have actually being exposed to:

$$\widehat{\Phi_i} = \phi(Z_i; \boldsymbol{X}_i^z; \widehat{\theta^z}),$$

$$\widehat{\Lambda_i} = \lambda(\boldsymbol{G}_i; Z_i, \boldsymbol{X}_i^g; \widehat{\theta^g}).$$

(1.c) Estimate parameters of the outcome model

Use the predicted propensity scores $\widehat{\Phi}_i$ and $\widehat{\Lambda}_i$, in order to estimate the parameters θ^y of the outcome model $Y_i(z, \boldsymbol{g})$:

$$Y_i^y(Z_i, \boldsymbol{G}_i, \widehat{\Phi}_i, \widehat{\Lambda}_i; \theta^y).$$

2) Impute Missing Potential Outcomes

Consider the domain of the joint treatment $(Z_i = z, \boldsymbol{G}_i = \boldsymbol{g})$. In particular, $\boldsymbol{G}_i$ is a K-dimensional continuous variable. For each possible value of the joint treatment, that is, for each combination $(Z_i = z, \boldsymbol{G}_i = \boldsymbol{g})$ s.t $z \in \{1, \ldots, K\},\ \boldsymbol{g} \in \Gamma$ [3], with $\Gamma \subset \mathbb{G}^K$, and for each unit $i \in \mathcal{N}$,

(2.a) Predict the individual propensity score corresponding to that level of z, $\widehat{\phi}(z; \boldsymbol{X}_i^z)$.

[3] In order to explore a multivariate domain, one common practice is constructing a K-dimensional discrete grid that scours the possible values of $\boldsymbol{g}$, over its K components' respective domain. Let us denote this grid as Γ, $\Gamma \subset \mathbb{G}^K$.

(2.b) Predict the neighborhood propensity score corresponding to that level of $\boldsymbol{g}$, $\widehat{\lambda}(\boldsymbol{g}; z, \boldsymbol{X}_i^g)$.
(2.c) Use that estimated parameters to impute the potential outcome $\widehat{Y}_i(z, \boldsymbol{g})$, that is,

$$\widehat{Y}_i(z, \boldsymbol{g}) \sim f^y(z, \boldsymbol{g}, \widehat{\phi}(z; \boldsymbol{X}_i^z), \widehat{\lambda}(\boldsymbol{g}; z, \boldsymbol{X}_i^g); \widehat{\theta}^y).$$

3) Estimate the effects of interest and their corresponding variance
(3.a) Estimate the final direct effects of interest, averaging potential outcomes over $\lambda(\boldsymbol{g}; z, \boldsymbol{X}_i^g)$:

$$\widehat{\tau}_{z'z} = \frac{1}{N} \sum_{i=1}^{N} \Big[\sum_{\boldsymbol{g} \in \Gamma} \big(\widehat{Y}_i(z', \boldsymbol{g}) - \widehat{Y}_i(z, \boldsymbol{g}) \big) \Big] \frac{\widehat{\lambda}(\boldsymbol{g}; z, \boldsymbol{X}_i^g)}{\sum_{\boldsymbol{g'} \in \Gamma} \widehat{\lambda}(\boldsymbol{g'}; z, \boldsymbol{X}_i^g)}.$$

(3.b) Compute variance through bootstrap (Efron, 1992; Forastiere et al., 2020). This procedure works as follows. Choose a large value R. For R times, $r : \{1, \ldots, R\}$, draw a random sample $\mathcal{N}_r$ with replacement from $\mathcal{N}$. Estimate the effects of interest over the subsample $\mathcal{N}_r$, so getting $\widehat{\tau}_{z'z}(r)$. Consider the distribution of $\widehat{\tau}_{z'z}(r)$ over the R repetitions $\widehat{T}^r_{z'_z}$. The estimated standard error Std.Er$(\widehat{\tau}_{z'z})$ is the standard error of $\widehat{T}^r_{z'_z}$. [4]

2.3 Empirical Application

In this section, we focus on the empirical application. We first explain the relevance of our empirical research question with respect to the existing literature about immigration (Subsection 2.3.1). Second, we describe the different data sources that we have merged (Subsection 2.3.2). Finally, we formalize and discuss the influence index (Subsection 2.3.3) and we explain how we derive the treatment categories (Subsection 2.3.4) also

[4] The methodological setting makes it difficult to derive a closed-form estimator for variances. We are aware that the bootstrap methodology holds under the assumption of no correlation among neighbors' outcomes. We could experiment alternative sampling techniques to explicitly account for the network information into data. However, graph partitioning methods are not easy to be applied on our data, due to the complexity of the network structure. The network information has been directly included into the definition of the neighborhood treatment, which we treat as an additional intervention. Then, we perform bootstrap by employing the standard unit-level sampling with replacement, similarly to Del Prete et al., 2019 and Forastiere et al., 2020

showing how to implement the estimation strategy we have presented in the previous section (Subsection 2.3.5).

2.3.1 Empirical Research Question

In the last decades, interest about immigration has rapidly grown, so that it has become a major topic both in academic and real life debates (Helbling et al., 2017). Immigration flows significantly increased, since many people attempted to move away from countries which have been suffering long periods of wars and bad economic conditions. The consequence of this process is that the world has become multicultural: migrants have started to be socially included into the hosting countries, searching for a new job and establishing social relationships. Moreover, migrants have diffused their own social and religious beliefs. However, immigration has entailed not only positive outcomes. Some countries that have embraced a relevant number of migrants have experienced the rising of social tensions (Rudolph, 2003). Over the last decades, economic conditions gradually get worse: unemployment rates rose up, and real wages went down. In addition, people have noticed a relevant worsening in the perception of individual security. The conviction that immigration may have aggravated these negative processes has slowly taken root in the public opinion. Politicians and common citizens have started to evaluate problematic consequences about immigration and globalization. Concerns about migration spread up in three main directions. First, native people perceive immigration as a risk for the preservation of national identity. Integration results to be not always easy and multiculturalism tends to be perceived more as a threat than as an opportunity. Second, as migrants move looking for better living conditions, they represent, in the common belief, competitive profiles for job. Finally, people tend to blame migrants for raising crime (Bigo, 2002).

In recent years, many researchers have started studying the effects of the increasing migration flows. For instance, Bove and Böhmelt, 2016 has assessed the effect of migration on the diffusion of terrorism, while Rudolph, 2006 has evaluated the effects on national security. Many epi-

demiological studies as Polissar, 1980, Stillman et al., 2007 and Hildebrandt and McKenzie, 2005 have analyzed the consequences on the spreading of some diseases and on public health, in general. Furthermore, Coleman, 2008 and Keely, 2000 have studied the causal effect of migration on some demographic outcomes. Bianchi et al., 2008 Bianchi et al., 2012 and Stansfield, 2016 have studied the impact of flows on crime. The existing works assessing the causal link between migration flows and crime present findings which are conceptually in contrast to common perception, suggesting that increasing immigration flows does not lead to higher crime rates. Some of them also state that there is actually a negative effect of immigration on crime.

The public discussion about migration has also involved the immigration policies that national governments implement with the aim of controlling and ruling the immigration process. Brochmann and Hammar, 1999 defined immigration policies as the *"government's statements of what it intends to do or not do (including laws, regulations, decisions or orders) in regards to the selection, admission, settlement and deportation of foreign citizens residing in the country"*. These policies can be more or less restrictive and, therefore, can discourage or encourage migrants, respectively. Some political parties of various countries all over the world support the idea that implementing restrictive immigration policies limits the negative effects of migrations and, consequently, leads to better living and economic conditions for the natives. On the other side, many politicians and intellectuals argue that the legislative system of a country should encourage immigrants and facilitate their settling.

In this work, we investigate the causal effect of immigration policies on crime rates. Specifically, we study the effect of the restrictiveness of the implemented immigration policy on one year lagged national *crime rate*, expressed in terms of homicides every 10.000 inhabitants. We approach this research question from a country level perspective: in particular, we focus on the subset of OECD countries that are located in the continental Europe and we inspect their policies towards migrants from 1980 and 2010. These policies have been measured in terms of restrictiveness in the *IMPIC (Immigration Policies in Comparison)* Dataset (Helbling

et al., 2017, Schmid and Helbling, 2016), that properly conceptualizes and quantitatively compares many national policies that affect migrants (Munck & Verkuilen, 2002). Starting from the restrictiveness measures supplied by the IMPIC Dataset and taking into account the conceptualization of the observed policies that the same dataset proposes, we evaluate the national immigration policy over two political dimensions: restrictiveness of regulations and restrictiveness of control strategies. From now on, we denote by *Reg* and *Cont* the two variables representing those two dimensions. We present a treatment variable that qualitatively distinguishes country-year profiles with respect to these two measures and we pairwise compare different political strategies.

Our empirical analysis covers 22 OECD countries that are situated in the continental Europe [5]. These countries are characterized by very different immigration experiences: there are countries that have experienced increasing immigration since one or two centuries (Great Britain, Germany, France), countries that recently turned from emigration to immigration (Italy, Spain) and countries that have experienced very limited immigration (Finland) (Helbling et al., 2017). However, they are still highly comparable from an institutional point of view, as they are all fully developed democracies and are all located in Europe. In Figure 17 we show a map of the 22 countries included. We believe that this is an innovative contribution to the existing literature about migration. Indeed, there are just some recent studies (Geddes and Scholten, 2016, Messina, 2007) about immigration policies but they look at particular behaviors of individual countries or describe a small number of countries (Helbling et al., 2017). Even if there are existing works that assess the effects of migration flows on crime rates, they focus on single countries comparing subnational administrative entities and they do not take into account the national political strategy towards migrants. Moreover, they all rule out spillover effects.

In this empirical scenario, interference may play a relevant role. Migrants may choose to avoid highly restrictive countries and to settle in

[5] 24 OECD countries are located in the continental Europe but we remove from the analysis Hungary and Estonia as they present extreme values of the crime rate.

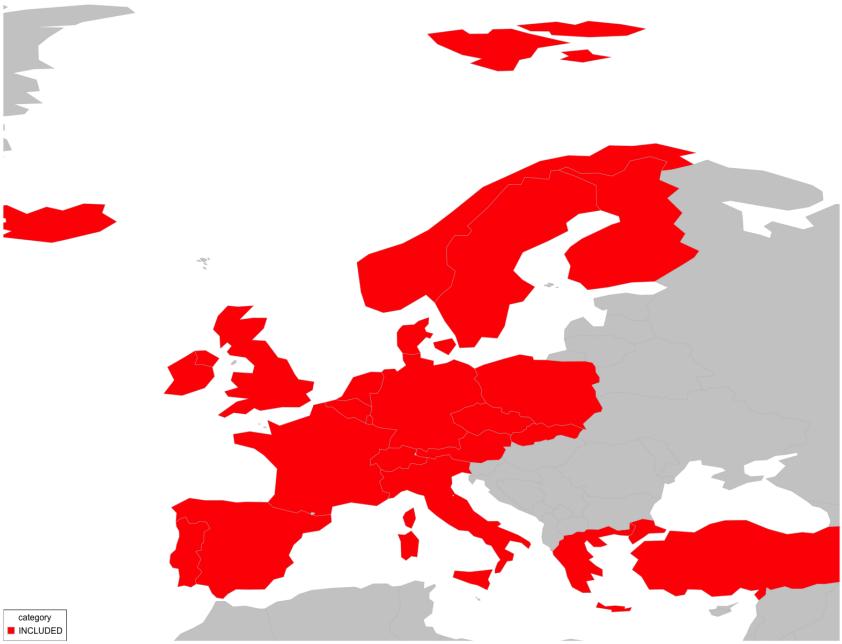


Figure 17: Included countries: red colored countries are the ones included in our analysis. They are the 22 ORCD countries located in continental Europe

places where laws appear to be more welcoming. But we expect that they try to preserve some characteristics of their settling choice. Thus, the general idea is that dependence between two countries is related to their level of similarity. We assume that two mechanisms may drive interference: *geographical proximity* and *cultural proximity*. Thus, we build a continuous indicator that analytically captures these driving mechanisms of interference. Each component contributes according to a given weight. We test various configurations of the influence weights in order to check the robustness of our results with respect to different restrictions about dependencies.

2.3.2 Data

This work merges different data sources. First, we use the *IMPIC (Immigration Policies in Comparison)* Dataset (Helbling et al., 2017) that provides information about national immigration policies. In particular, this dataset includes data on migration policies for all the OECD countries over thirty years (from 1980 to 2010). Policies are measured with respect to their restrictiveness, from 0 (less restrictive) to 1 (highly restrictive). Data include more than 50 policies for each country-year profile and items are aggregated in different indicators of the general restrictiveness towards migrants with respect to the *regulation* and *control* protocols. The former aspect is related to all the laws that discipline immigrants and their life in the hosting country, while the latter is referred to the mechanisms that help in monitoring whether the regulations are abided by (Schmid & Helbling, 2016).

Second, we handle different datasets to assemble the *Influence Index* (II), which measures the extent of dependency between each pair of countries at a given year, that is, the extent to which a country's immigration policy influences the crime rate of another country. As we will fully discuss in the forthcoming section, this index is a convex combination of two complex indicators quantifying *geographical proximity* and *cultural similarity* between two countries at a given year. We build up the *geographical proximity indicator* starting from the CEPII Dist Dataset (T. Mayer and Zignago, 2011) which includes different measures of bilateral distances (in kilometers) and a dummy variable denoting pairwise contiguity. Furthermore, we explore *cultural similarity* between each pair of countries at a given year looking at the *linguistic similarity* through the CEPII Language Dataset (Melitz & Toubal, 2014) and at the *religious similarity* through CEPII Gravity Dataset (Fouquin, Hugot, et al., 2016).

Third, we make use of some datasets that provide country-year features. Specifically, we collect information about crime rates relying on the *World Countries Homicide rate dataset* which comprises information about the country-year specific number of homicides per 10.000 inhabitants. In addition, we manage the *World Development Indicators* dataset, provided

by the World Bank (Coppedge et al., 2018a, Lindberg et al., 2014 and Coppedge et al., 2018b) which contains highly detailed country-year indicators referring to various aspects of society: they quantitatively mark out the economic situation, the demographic features, the state of the social welfare and democracy and even the level of equality, freedom and justice.

The observed population is characterized by country-year observations: we deal with $C = 22$ countries observed over $T = 30$ years where the initial time $t = 1$ is year 1980 while the ending time $t = T$ is year 2010). [6]. Therefore, the generic unit i is a pair (c, t) and the total number of units is $N = C \times T$. We indicate as $\mathbf{Y}^{obs} = \{Y_{ct}^{obs}\}$ the $(N \times 1)$ observed (country,year) outcome vector. Furthermore, we take into account of the pre-treatment covariates matrix $\mathbf{X}$ with dimension $N \times P$: each row of this matrix represents a country-year observation, while each column refers to a specific baseline factor. The included covariates can be grouped in four sets, according to the main issue they refer to: i) *Economy*: GDP per capita, equal distribution of resources index, state ownership of economy index; ii) *Inequality*: educational inequality index, income inequality index, health equality index, power distributed to gender index, equal access; iii) *Freedom and Participation*: civil participation index, freedom of expression index, freedom of religion; iv) *Demography*: life expectancy, fertility rate. We denote as $\mathcal{X}$ the set collecting these variables.

We point out that we assume that there is a *one-year lag effect* of baseline covariates on treatment and of treatment on outcome variables. We state that the covariates of one country c at time t affect its individual as well as neighborhood treatment at time $t + 1$ and that the joint treatment in turn affects the outcomes at time $t + 2$. Figure 18 provides an intuition of this conceptual idea. There is no way to test this assumption, but a lagged process of causation seems plausible in the considered empirical scenario. In addition, in order to avoid reverse causality issues, in the

[6] Estonia starts to be included in the analysis from 1991, after its Independence. Czech Republic and Slovak are instead considered only from 1993: they both became independent countries after the Dissolution of Czechoslovakia which took effect on Jan 1, 1993.

Figure 18: Temporal structure of effects: the covariates of one country c at time t affect his individual as well as neighborhood treatment at time $t+1$ and that the joint treatment in turn affects the outcomes at time $t+2$

propensity score estimation we control for no-lagged outcome variables. For instance, we consider the baseline covariates of one country at time t to model his joint treatment at time $t+1$ and consequently his outcome at time $t+2$ and in the set of the pre-treatment variables at time t we include the outcome at time t as well.

2.3.3 Modelling Interference: Influence Index (I)

Here, we must define the interference structure taking into account the possible mechanisms that could make immigration policies in one country affecting the crime levels of other countries. The idea is that immigrants avoid highly restrictive countries and settle to areas that are similar to the first choice with respect to some characteristics, but more politically welcoming. Thus, the relevance of spillover between each pair of countries depends on their pairwise similarity. We assume that the kind of similarity that plays a role in this mechanism is the geographic proximity (meaning, the geographic distance between countries) and the cultural similarity. In some sense, we state that a migrant, who is willing to move, chooses the most welcoming alternative among the countries that are relatively near and culturally similar to the first choice option, that though implements highly restrictive laws. Therefore, we build up a composite indicator which numerically summarizes these two mechanisms which we reasonably believe are the key prompters of dependency. The two components contribute to the determination of the global

index according to some weights, α and β. This index, that we call *Influence Index (I)*, gives a unique information about how much one country c interfere with a country c' at time t. Formally: [Influence Index (I)]

$$I_{cc',t} = \alpha \times IG_{cc'} + \beta \times IC_{cc',t}$$

where $IG_{cc'} \in [0,1]$ is the *geographic proximity indicator* which measures the geographical proximity between country c and country c' and $IC_{cc',t} \in [0,1]$ is the *cultural similarity indicator* that measures the cultural similarity between two countries time t. The constants α and β, with $\alpha + \beta = 1$, are the Influence Inputs Weights (IIW) that determine the extent to which each component contributes to the global index. Note that, since the Influence Index is a convex combination of two indicators bounded between 0 and 1, it is in turn bounded between 0 and 1, that is, $I_{cc',t} \in [0,1]$. More details about the construction of the Influence index can be found in the Appendix A.2. We test various allocations of II to check the robustness of our results with respect to different assumptions over the interference structure. Following the same approach of many existing works in economics and social sciences (Del Prete et al., 2019), we have ruled out the presence of intertemporal links, that is we set $I_{(ct),(c't')} = 0 \ \ \forall c, c', t, t'$ with $t \neq t'$. This assumption rules out any form of intertemporal dependencies between country-year profiles. This means that the potential outcome of a given country c observed at time t does not depend on the treatment of any other country (included itself) observed at a different point in time t'. The structure of between-countries dependencies is uniquely ruled by the year specific influence index $I_{cc't}$. Conversely, within-country intertemporal dependencies are uniquely modelled according to the temporal structure of effects that we have introduced in the Subsection 2.3.2.

2.3.4 Treatment Categories

IMPIC dataset provides indicators which measure the country-year restrictiveness towards migrants with respect to *regulations* and *control* mechanisms. Let us denote as reg_i the reported value of the restrictiveness in

terms of regulations of the generic country-year profile $i = (c, t)$ and with $cont_i$ the corresponding value in terms of control protocols. We define the nominal treatment categories looking at the joint distribution of the two indicators. In particular, denoting as med_{Reg} and med_{Cont} the median of the distribution of the regulations indicator and of the control one, respectively, we define the treatment categories as follows

Definition 5 (Nominal Treatment Categories)**.** *Individual treatment is obtained by applying the following categorization criterion.*

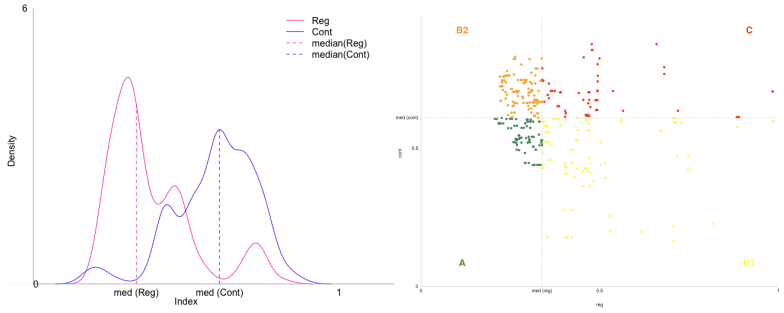
- *Z_i=LL if $reg_i \leq med_{Reg}$ and $cont_i \leq med_{Cont}$: this category identifies profiles that are barely restrictive with respect to the two mechanisms.*
- *Z_i=HL if $reg_i > med_{Reg}$ and $cont_i \leq med_{Cont}$: this category detects profiles which implement restrictive regulations but weak control strategies.*
- *Z_i=LH if $reg_i \leq med_{Reg}$ and $cont_i > med_{Cont}$: this category indicates a welcoming attitude in terms of regulations but intense control protocols.*
- *Z_i=HH if $reg_i \geq med_{Reg}$ and $cont_i \geq med_{Cont}$: this category denotes an highly restrictive policy towards migrants with respect to both regulations and control.*

These categories intend to summarize the country-year political strategy towards migrants. They have been defined by following an approach which ensures to have an accurate but still interpretable differentiation of the examined political profiles. The two measures on which this categorization depends, regulations and control mechanisms, are considered in the existing literature about immigration policies the two main determinants of the (latent) overall political approach towards immigrants (Helbling et al., 2017). Therefore, we can reasonably believe that the proposed categorization plausibly represents the country-year's attitude towards migrants, while guaranteeing both accuracy and interpretability [7]

Figure 19 provides a graphical idea of the previously described definition procedure. The left subfigure shows the density distributions of the *regulation* and *control* indexes, across the country-year profiles: their

[7] Since this categorization is arbitrary, we will check the robustness of results under various plausible definitions of the treatment variable.

corresponding median values are identified by dotted lines (Helbling et al., 2017). The right subfigure shows the individual treatment collocation based on their own values of the *regulation* and *control* indexes.



(a) Density distributions of *regulation* (vi-(b) Individual collocation in the treatolet line) and *control* (blue line) indexes,ment categories according to the regulaand their respective medians (dottedtion and control indexes lines)

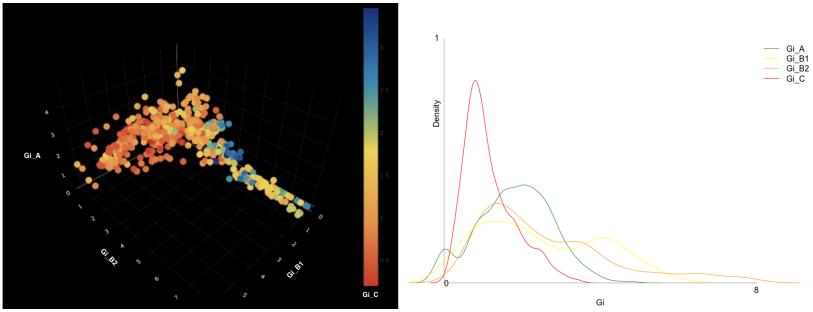
Figure 19: Treatment Categories: definition

Hence, we deal with a K-valued individual treatment, where $K = 4$. Let us denote as $\mathbf{Z} = \{Z_{ct}\}$, the $(N \times 1)$ multi-valued treatment vector where $Z_{ct} \in \{LL, HL, LH, HH\}$. Following Definition 5 and assuming the Influence Index as the ruling mechanism of dependencies, we explicit the neighborhood treatment $\boldsymbol{G}_{ct}$ as

$$\boldsymbol{G}_{ct} = \begin{pmatrix} G_{ctLL} \\ G_{ctHL} \\ G_{ctLH} \\ G_{ctHH} \end{pmatrix} = \begin{pmatrix} \sum_{c' \in \mathcal{N}_{ct}} I_{cc',t} \delta_{Ac't} \\ \sum_{c' \in \mathcal{N}_{ct}} I_{cc',t} \delta_{HLc't} \\ \sum_{c' \in \mathcal{N}_{ct}} I_{cc',t} \delta_{LHc't} \\ \sum_{c' \in \mathcal{N}_{ct}} I_{cc',t} \delta_{HHc't} \end{pmatrix},$$

where $\delta_{LLc't}, \delta_{HLc't}, \delta_{LHc't}, , \delta_{HHc't}$ are dummy variables such that $\delta_{LLc't} = 1$ if $Z_{c',t} = LL$ and 0 otherwise; $\delta_{HLc't} = 1$ if $Z_{c',t} = HL$ and 0 otherwise; $\delta_{LHc't} = 1$ if $Z_{c',t} = LH$ and 0 otherwise; $\delta_{HHc't} = 1$ if $Z_{c't} = HH$ and 0 otherwise. Consequently, the potential outcomes are defined as $Y_{ct}(Z_{ct}, \boldsymbol{G}_{ct})$. Figure 20 displays the distribution of the neighborhood treatment variable under the hypothesis of equal contribution to the Influence index of the cultural and geographical subcomponents, ($\alpha = \beta =$

$\frac{1}{2}$). The pre-treatment variables that we employ in this analysis are the



(a) Tridimensional scatterplot of the neighborhood treatment variable

(b) Density distribution of the singular components that constitutes $\boldsymbol{G}_{ct}$

Figure 20: Neighborhood treatment, under $\alpha = \beta = \frac{1}{2}$: (left) tridimensional scatterplot of the neighborhood treatment variable: the first three components correspond to the three axes that are present in the figure, while the value of the latter one is analyzed by looking at the color intensity; (right) density distribution of the singular components that constitutes the multivariate neighborhood treatment variable

ones described and motivated in Subsection 2.3.2. They effectively summarize four aspects that may impact both the country-year's attitude towards migrants and their expected crime rate: economy, inequality, freedom and participation and demography. Since covariates cover heterogeneous characteristics of countries and they refer to the year previous to the realization of the treatment (as extensively motivated in Subsection 2.3.2), we can reasonably assume that, conditioning on these characteristics, the weak unconfoundedness assumption holds and so that the country-year's attitude towards migrants and its potential crime rate are independent.

In order to estimate the causal effects of interest, we follow the estimation procedure described in Section 2.2.

2.3.5 Joint Multiple Generalized Propensity Score (JMGPS) Estimation

We estimate the two components of JMGPS (see Definition 3), that is the individual propensity score and the neighborhood propensity score.

Individual Propensity Score

The individual propensity score $\phi(z; \boldsymbol{x}^z)$ is the individual probability of receiving an individual treatment z conditioning on unit-level baseline covariates. If the individual treatment is a categorical variable with K nominal categories the estimation strategy consists in fitting a model for categorical responses. Here we use the *Multinomial Logit Model* (Agresti, 2018; Long et al., 2006 and Menard, 2002), where the reference category is set to "LL", that is,

$$P(Z_i = LL) = \frac{1}{1 + \sum_{z \neq LL} \exp \boldsymbol{\beta}_z \boldsymbol{X}_i^z},$$

$$P(Z_i = HL) = \frac{\exp \boldsymbol{\beta}_{HL} \boldsymbol{X}_i^z}{1 + \sum_{z \neq LL} \exp \boldsymbol{\beta}_z \boldsymbol{X}_i^z},$$

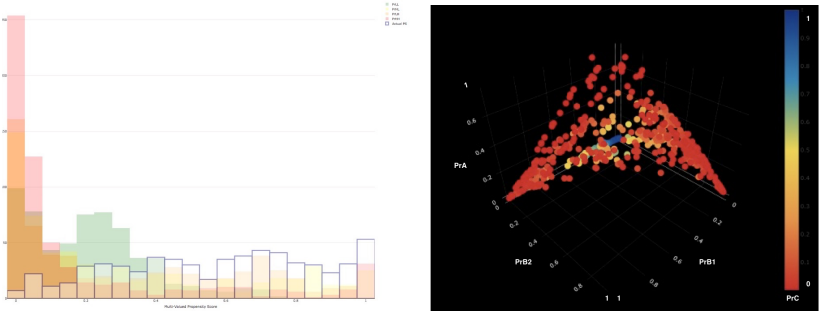
$$P(Z_i = LH) = \frac{\exp \boldsymbol{\beta}_{LH} \boldsymbol{X}_i^z}{1 + \sum_{z \neq LL} \exp \boldsymbol{\beta}_z \boldsymbol{X}_i^z},$$

$$P(Z_i = LH) = \frac{\exp \boldsymbol{\beta}_{HH} \boldsymbol{X}_i^z}{1 + \sum_{z \neq LL} \exp \boldsymbol{\beta}_z \boldsymbol{X}_i^z}.$$

Given the vector of estimated parameters $\widehat{\theta}^z = \{\widehat{\boldsymbol{\beta}}_{HL} \cup \widehat{\boldsymbol{\beta}}_{LH} \cup \widehat{\boldsymbol{\beta}}_{HH}\}$, we denote the estimated *individual propensity score* corresponding to the actual treatment Z_i as $\widehat{\Phi_i} = \phi(Z_i; \boldsymbol{X}_i^z; \widehat{\theta}^z)$. We include in $\boldsymbol{X}_i^z$ the whole set of covariates $\mathcal{X}$ we have described in Section 2.3.2. Figures 21 provides a graphical intuition of the marginal and joint distribution of predicted propensity scores.

Neighborhood Propensity Score

In the considered empirical scenario, the neighborhood treatment is a quadrivariate continuous variable, $\boldsymbol{G}_{ct}$.



(a) Histograms of propensity scores

(b) Tridimensional scatterplot of propensity scores: colors refer to $\phi(HH; \boldsymbol{X}_i^z; \widehat{\theta}^{HH})$

Figure 21: Individual Propensity Score

We first apply a transformation on each component of the neighborhood multi treatment (more details can be found in the Appendix A.3) so that, after the transformation, we can state that the obtained variables $G^*_{i,z}$ follow a normal distribution. This transformation allows us to fairly assume a multivariate normal distribution over G^*_i and, for this reason, we do not take into account alternative distributions of G^*_i.

Specifically, the four transformed components jointly follow a quadrivariate-normal distribution:

$$\boldsymbol{G}^*_i \sim \mathcal{MN}(\boldsymbol{\mu}_{\boldsymbol{G}^*_i}, \boldsymbol{\Sigma}_{\boldsymbol{G}^*}),$$

where the vector of the means $\boldsymbol{\mu}_{\boldsymbol{G}^*_i}$ depends on the individual treatment and on units' covariates through some parameters,

$$\begin{aligned}\boldsymbol{\mu}_{\boldsymbol{G}^*_i} =& \left[\mu_{G^*_{i,LL}}, \mu_{G^*_{i,B1}}, \mu_{G^*_{i,B2}}, \mu_{G^*_{i,C}}\right] \\ & \left[\alpha_{G^*_{LL}} + \boldsymbol{\beta}^T_{G^*_{LL}} \mathbf{X}^g_i + \boldsymbol{\beta}^T_{G^*_{LL}} Z_i, \alpha_{G^*_{HL}} + \boldsymbol{\beta}^T_{G^*_{HL}} \mathbf{X}^g_i + \boldsymbol{\beta}^T_{G^*_{HL}} Z_i,\right. \\ & \left.\alpha_{G^*_{LH}} + \boldsymbol{\beta}^T_{G^*_{LH}} \mathbf{X}^g_i + \boldsymbol{\beta}^T_{G^*_{LH}} Z_i, \alpha_{G^*_{HH}} + \boldsymbol{\beta}^T_{G^*_{HH}} \mathbf{X}^g_i + \boldsymbol{\beta}^T_{G^*_{HH}} Z_i\right]\end{aligned}$$

and the variance-covariates matrix $\boldsymbol{\Sigma}_{\boldsymbol{G}^*}$ is a (4×4) matrix such that the diagonal embraces the elements $\sigma^2_{G^*_z}$, which represent the variances of each singular component, while the elements that are out of the diagonal

represent the covariances between each pair of G_z^* and $G_{z'}^*$, with $z, z' \in \mathcal{Z}$, that is, $\rho_{(G_z^*, G_{z'}^*)}\sigma_{G_z^*}\sigma_{G_{z'}^*}$

To estimate these quantities, we fit *Multivariate Multiple Linear Regression Model*, (Davis, 1982,Duchesne and De Micheaux, 2010), regressing the (transformed) unit neighborhood treatment $\boldsymbol{G}_i^*$ on the individual treatment Z_i and on the predictors that are candidate to influence the neighborhood treatment, $\boldsymbol{X}_i^g$. Here we include as explanatory variables $\boldsymbol{X}_i^g$ the whole set of characteristics $\mathcal{X}$, the individual treatment Z_i and a measure of vertex centrality. This procedure determines $\widehat{\boldsymbol{\mu}}_{G_i^*}$.
The variance-covariance matrix is estimated looking at the residuals of the model. In particular, we first compute residuals of the model and, then, we estimate the variance and covariance matrix of the residuals $\widehat{\boldsymbol{\Sigma}}_{G^*}$, that results to be an unbiased estimator of $\boldsymbol{\Sigma}_{\boldsymbol{G}^*}$. Therefore, the neighborhood propensity score corresponds to the quantity

$$\widehat{\Lambda}(\boldsymbol{g}; z, \boldsymbol{X}_i^g) = \frac{1}{(2\pi)^{\frac{3}{2}}\left|\widehat{\boldsymbol{\Sigma}}_{G^*}\right|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\left(\boldsymbol{g} - \widehat{\boldsymbol{\mu}}_{G^*}\right)^T \widehat{\boldsymbol{\Sigma}}_{G^*}^{-1}\left(\boldsymbol{g} - \widehat{\boldsymbol{\mu}}_{G^*}\right)^T \right].$$

2.4 Empirical Results

In this section, we illustrate the main empirical findings of this work [8]. We evaluate the impact of immigration policies on crime rates evaluating pairwise comparisons between the four treatment levels. To assess the robustness of results with respect to different assumptions on the influence structure, we check the following configurations of the Influence Inputs Weights (IIW): i) $\alpha = \beta = \frac{1}{2}$, *(gc)*: both geographical proximity and cultural similarity shape dependencies between units, and contribute in determining the influence index with equal weight; ii) $\alpha = 1, \beta = 0$, *(g)*: only *geographical proximity* drives interference; iii) $\alpha = 0, \beta = 1$, *(c)*: the influence structure depends on cultural similarity only and iv) $\alpha = 0, \beta = 0$, *(noint)*: *no interference* mechanism comes into play.

[8] Here, we just present conclusions about the causal effects of interest, more detailed results about the models we implemented in the whole analysis can be found in the Appendix A.6. Descriptives about included covariates are provided by Appendix A.4.

Figures 22 graphically shows the main empirical results, which are numerically reported in Table 2. The general conclusion is that severe approaches towards immigration imply higher crime rates, compared with a welcoming political receipt. This finding holds when the comparison is with strategies with restrictive regulations only (*HL-LL*), systems where only control protocols are particularly strict (*LH-LL*) and profiles adopting a restrictive legislative plan in terms of both regulations and control mechanisms (*HH-LL*). If we look at how results change with different definitions of the influence weights, we can state that ignoring the possible spillover mechanism (*noint*) leads to a downward bias in the estimates. This conclusion is stable in all the contrasts of interest. On the contrary, allowing for the presence of interference increases the size of the effects. In particular, introducing the cultural similarity in the mechanism of dependencies enhances the effects' intensities (*c*). Geographical proximity mitigates the impact of interference on results, but also assuming that geography is the only prompter of the spillover mechanism steers to stronger conclusions, compared to the no-interference scenario (*g* and *gc*). These considerations hold in all the considered comparisons.

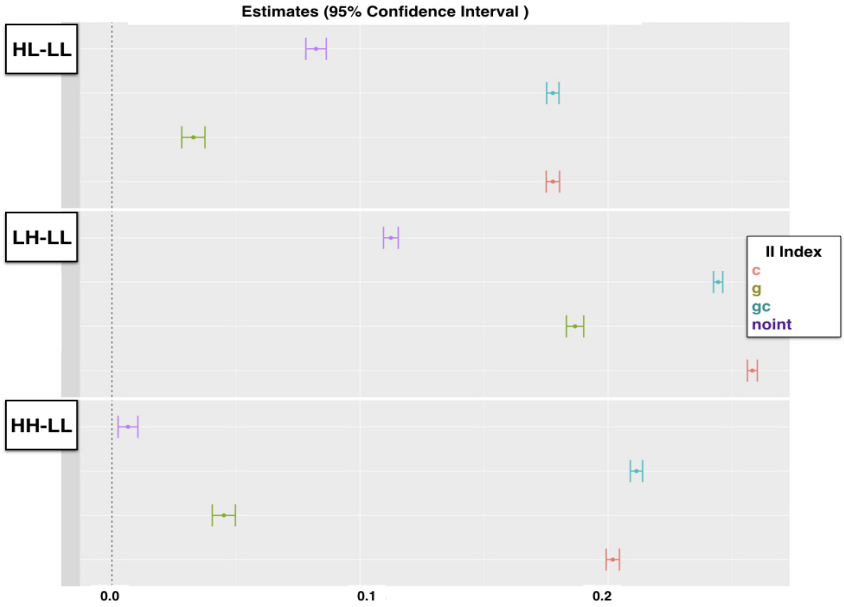


Figure 22: Direct treatment effects: point estimates and 95% Confidence intervals. Colors signal the different assumption about interference: *gc*(lightblue), *g*(green), *c*(red), *noint*(purple)

Table 2: Direct Treatment Effects for the contrasts of interest: point estimates and 95% Confidence intervals

	Effects of Interest		
IIW (α, β)	HL-LL	LH-LL	HH-LL
$(\frac{1}{2}, \frac{1}{2})$	0.17774 ***	0.24439 ***	0.21145 ***
	(0.17501;0.18008)	(0.24246;0.24618)	(0.20907;0.21409)
$(1, 0)$	0.03281 ***	0.1867 ***	0.0451 ***
	(0.02768;0.03721)	(0.18308;0.19007)	(0.04062;0.05006)
$(0, 1)$	0.17778 ***	0.25819 ***	0.20191 ***
	(0.17483;0.1803)	(0.2561;0.26012)	(0.19934;0.20476)
$(0, 0)$	0.08228 ***	0.11245 ***	0.00647 ***
	(0.07842;0.08657)	(0.10927;0.11517)	(0.00213;0.01038)

As we fully discuss in Appendix A.5, these results are robust to different specifications of the multi-valued treatment (we introduce an alternative definition of the multi-valued treatment collapsing the LH and HL categories into one M category).

2.5 Concluding Remarks and Discussion

This work extends the existing framework of causal inference under interference allowing for a multi-valued treatment and for an interference structure shaped through a weighted network. This is a very common setting that can be found in a wide wide ensemble of applications. For example, political science often deals with policy evaluation settings with a multi-valued strategy, as treatments vary across multiple dimensions, so calling for an high level of complexity. Here we evaluate the effect of the national immigration policy on the crime rate. Given the multi-valued nature of the individual treatment, the neighborhood exposure cannot be summarized by a single measure, as in the binary setting. Our idea is to introduce a multi-valued network exposure, where each unit is exposed to their neighbors' treatment, weighted by the strength of their interaction. Information about the whole exposure mapping is depicted by the Neighborhood Treatments Exposure Matrix (NTEM). This framework implies an extended definition of the joint propensity score, called Joint Multiple Generalized Propensity Score (JMGPS), which models a multi-valued individual treatment and a multivariate neighborhood treatment. Direct effects of interest are pairwise comparisons of all treatment levels and they are computed comparing imputed potential outcomes controlling for the multi-valued network exposure. Our empirical results show that implementing a welcoming immigration policy causes a reduction in the crime rate. These findings suggest that welcoming immigration policies may contribute in reducing the social unrest between immigrants and natives. One possible explanation is that adopting a legislative system, which allows migrants to be actively involved in the hosting community, conceding them civil and social rights, encourages the integration process and reduces frictions. Results also show that ig-

noring multi-valued interference leads to weaker estimates.

Chapter 3

Heterogeneous Treatment and Spillover Effects under Clustered Network Interference

This Chapter is a joint work with my supervisor Prof. Laura Forastiere and with my colleague Dr. Falco J. Bargagli-Stoffi. The full text of the article will be soon available from the arXiv repository.

3.1 Introduction

3.1.1 Motivation

According to Cox, 1958, there is *interference* between different units if the outcome of one unit is affected by the treatment assignment of other units. In the case of policy interventions or socio-economic programs, interference may arise due to social, physical or virtual interactions. For instance, in the case of infectious diseases, unprotected individuals can still benefit from public health measures taken in the rest of the population, such as vaccinations or preventive behaviors, because these reduce the reservoir of infection (Bridges et al., 2000; Nichol et al., 1995), the vector of transmission (Binka et al., 1998; Howard et al., 2000) and the number of susceptible individuals (Broderick et al., 2008; Kelso et al., 2009; Kissler et al., 2020). In labor market, job placement assistance can affect job seekers that use this service, but it can also have an effect on other job seekers that are competing on the same job market. In education, learning programs may spill over to untreated peers through knowledge transmission paths. In marketing, the exposure to an advertisement might directly affect the consuming behaviour of the exposed individuals, and indirectly affect other individuals that are influenced by the consuming choices of people in their social network. If the exposure to the advertisement takes place in social media, the spillover effect of being targeted by a marketing campaign might also be explained by the propagation of the advertisement to non-exposed users that are virtually connected to the targeted ones.

In economics and social sciences there has been increasing interest in estimating spillover effects on networks in many different contexts: Duflo and Saez, 2003 study the role of information and social interactions in retirement plan decisions in the academic community; Cai et al., 2015 assess the spillover effect of training sessions on uptake of weather insurance in rural areas of China; Muralidharan and Sundararaman, 2015 study the aggregate effects of school choices while Imai et al., 2018 evaluate the effects of the Indian National health insurance. In the presence of interference, the effect of the treatment status of other units on one's

outcome is usually referred to a *spillover effect*. Spillovers are a crucial component in understanding the full impact of an intervention at the population-level. In fact, the scale-up phase of an intervention requires knowledge about the mechanism of spillover and how this would take place in the population where the intervention will be rolled out. Information about spillovers of public interventions can also support decisions about how best to deliver interventions and could be used to guide public funds allocation. Indeed, the presence of beneficial spillover effects allows treating a lower percentage of the population, because the untreated individuals would still benefit from the treatment provided to the targeted sample. The use of spillover effects to save resources could be further improved if we were able to target specific individuals who would increase the overall impact of the intervention. A targeting (or seeding) strategy aims at delivering the intervention to certain individuals such that the impact on the overall population is maximized (e.g., Kim et al., 2015; Montes et al., 2020; Valente, 2012; VanderWeele & Christakis, 2019). Typically, seeding strategies are designed in settings where either an element of the intervention (e.g., information, flyers or coupons provided during the intervention) or the outcome (e.g., the adoption of a behavior or a product) diffuse through the network. In these settings, the goal is the identification of the set nodes in network that, if targeted, would maximize contagion or diffusion cascades. To do so, seeding strategies are designed using information on the network structure and the dynamics of contagion or diffusion. This 'influence maximization' problem is NP-hard and computer scientist have developed approximate algorithms that usually rely on simplified contagion processes (Kempe et al., 2003a). Indeed, it is typically assumed that susceptibility to direct treatment and to others as well as the influential power are homogeneous across individuals.

Here we take a different perspective. First, we investigate spillover effects of a unit's treatment on other units' outcome without specifying the mechanism through which this might take place, Second, we focus on the assessment of the heterogeneity of susceptibility to direct and indirect treatment. In the field of personalized medicine it is well known that in-

dividuals with different characteristics might respond differently to the treatment (e.g., Chakraborty & Murphy, 2014; Kosorok & Laber, 2019; Murphy, 2003). In the presence of interference, we also have that different people might be more or less susceptible to the treatment received by other units. This means that not only the treatment effect but also spillover effects are heterogeneous. Understanding these heterogeneities can help policy-makers in the scale-up phase of the intervention, it can guide the design of targeting strategies with the ultimate goal of making the interventions more cost-effective, and it might even allow generalizing the level of treatment spillover effects in other populations.

3.1.2 Related Work

In the past decade causal inference literature has experienced a growing interest in the mechanism of interference, leading to the assessment of bias for causal effects estimated under the no-interference assumption (Forastiere et al., 2020; Sobel, 2006a), the design of experiments to either avoid or assess interference (Angelucci & Di Maro, 2015; Baird et al., 2018; Kang & Imbens, 2016), and the estimation of casual effects under interference. Estimators for treatment and spillover effects have first been developed under the assumption of partial interference, allowing interference within groups but not across different groups (Basse & Feller, 2018; Forastiere et al., 2019a; Forastiere et al., 2016; Hudgens & Halloran, 2008; Liu & Hudgens, 2014; Liu et al., 2016; Tchetgen & VanderWeele, 2012). However, the assumption of group-level interference is often invalid. Hence, several works focus on the estimation of causal effects in the context of units interconnected on networks, both in randomized experiments (Aronow & Samii, 2017; Athey et al., 2018a; Leung, 2020) and in observational studies (Forastiere et al., 2020; Forastiere et al., 2018; Ogburn et al., 2017; Sofrygin & van der Laan, 2017). In the context of social networks, even in randomized experiments where the treatment is randomized at the unit-level, exposure to other units' treatment is not. Therefore the propensity of being exposed to different levels of treatments among network contacts will depend on the network struc-

ture. Aronow and Samii, 2017 developed an Horvitz-Thomson estimator to adjust for the imbalance in this propensity across units under different individual and contacts' treatment status.

In parallel to this field of research on interference, in recent years, thanks to the availability of increased computing power and large data sets, researchers have started to think about advanced data-driven methods to assess the heterogeneity of treatment effects with respect to large numbers of features. In fact, the standard methods for subgroup analysis to investigate heterogeneous effect has several drawbacks: (1) they strongly rely on the subjective decisions on the specific variables defining the heterogeneous sub-populations; and (2) they fail to discover heterogeneities other than the ones that are *a priori* defined by the researchers. In addition, a data-driven method avoids potential problems related to cherry-picking the subgroups with extremely high/low treatment effects (Assmann et al., 2000; Cook et al., 2004). Hence, many data-driven algorithms for the estimation of heterogeneous causal effects have been proposed in recent years (Athey & Imbens, 2016; Athey et al., 2019; Foster et al., 2011; Hahn et al., 2020; Hill, 2011; Lechner, 2019; Starling et al., 2019; Su et al., 2012; Wager & Athey, 2018).[1] The aim of these methods is to detect 'causal' rules defining subsets of the study population where the treatment effect for that subgroup deviates from the average treatment effect. This is done by selecting the most important features and their values that define a partition of the covariate space (the tree) where the treatment effect is 'significantly' heterogeneous. Among these algorithms, many rely on extensions of the standard Classification and Regression Trees (CART) (Breiman, 2001; Friedman et al., 1984), and are adapted to different settings (Athey & Imbens, 2016; Athey & Imbens, 2015; Bargagli Stoffi & Gnecco, 2018, 2019; Guber, 2018; Johnson et al., 2019; K. Lee et al., 2018; G. Wang et al., 2018; Zhang et al., 2017).

Causal trees have already been applied to various scenarios for the discovery of heterogeneous effects of air pollution (K. Lee et al., 2018), employment incentives (Bargagli Stoffi & Gnecco, 2019), job training programs (Cockx et al., 2019), development finance projects (Zhao et al.,

[1] For a recent review the reader can refer to Athey and Imbens, 2019.

2017), cardiovascular surgeries (G. Wang et al., 2017), cancer treatments (Zhang et al., 2017), and health insurance (Johnson et al., 2019). The wide usage of tree-based algorithms is due, in particular, to their ability to deal with non-parametric settings in an efficient and interpretable way. Indeed, these algorithms do not assume any specific shape of the treatment effect function. Causal trees and similar tree-based methodologies built on the Classification and Regression Trees (CART) algorithm (Friedman et al., 1984) have been widely employed because of various attractive features: i.e., they can deal with a large number of variables that are potentially responsible for the heterogeneity, they are simple to understand and visualize, easy to interpret, computationally scalable, and they can deal with non-linear relationships in the covariates without the need of data pre-processing. Nevertheless, tree-based methods for the estimation of heterogeneous causal effect have been developed ruling out the presence of spillover effects by assuming no-interference between the units. On the other hand, the growing literature on spillover effects has focused on the estimation of population average spillover effects, neglecting potential heterogeneous spillover effects. There have been few articles dealing with different types of heterogeneity in spillover effects. Forastiere et al., 2019a; Forastiere et al., 2016 estimated the heterogeneity of spillover effects with respect to principal strata defined by the compliance behaviors in response to a cluster randomized treatment. However, the latent nature of these strata makes it difficult to effectively use the detected heterogeneity to design targeting strategies or to generalize the conclusions to a different population. Observed heterogeneity is instead studied in Arduini et al., 2014 and Arduini et al., 2019 where the focus is on peer effects from other units' outcomes and their heterogeneity across two groups, and the estimation relies on linear-in-means models and two-stage least squares. To the best of our knowledge there are no studies dealing with the heterogeneity of spillover effects on networks.

3.1.3 Contributions

In this paper, we bridge the gap between these two bodies of causal inference literature by proposing a new algorithm for the discovery and estimation of heterogeneous treatment and spillover effects with respect to a large number of characteristics, including individual and network features. Our method is designed for randomized experiments affected by the presence of clustered network interference, that is, units are organized in a clustered structure, with no interactions between clusters and a network of connections within clusters (e.g., friendship networks within schools). Randomization is assumed at the individual-level, resulting in treated and untreated units in the same cluster. Under this setting, spillover effects are confined within clusters and are assumed to take place on network interactions.

Our proposed method, *network causal tree* (NCT), builds upon the *causal trees* proposed by Athey and Imbens, 2016, by modifying the splitting criterion to target treatment and spillover effects under interference. Splits are made so as to maximize the heterogeneity of the targeted causal effect(s), treatment and/or spillover effects, across the population. This criterion relies on the unbiasedness of the estimator of the effect(s) within each subset of the population. Therefore, we also contribute to the existing literature on interference by developing an unbiased estimator for conditional treatment and spillover effects. We extend the Horvitz-Thomson estimator in Aronow and Samii, 2017 to conditional causal effects under clustered network interference and we prove its consistency under the clustered network setting. This estimator is then used in our network causal trees to decide the binary splits that maximize the heterogeneity and to finally estimate the heterogeneous causal effects within the selected sub-populations.

In order to use the selected partition of the covariate space and the estimated treatment and spillover effects to guide policies, the heterogeneous sub-populations should be identified based on the causal effects that will be part of the decision rule. For instance, a policy that assigns the treatment to those who benefit the most from it requires the discov-

ery of the subsets of the populations with heterogeneous treatment effect. Alternatively, a policy that is designed to target those who will respond to both their own treatment and the neighbors' treatment will need to identify sub-populations with high treatment and spillover effects. Hence, the proposed *network causal tree* (NCT) is optimized to detect the heterogeneity in treatment and spillover effects (i) either *simultaneously*, or (ii) *separately*. By reworking the criterion function of the seminal causal trees algorithm to account for interference, we allow the algorithm to spot heterogeneity in treatment and/or spillover effects. The discovery of the causal rules (the variables and their values defining heterogenous sub-populations) representing the heterogeneity of one causal effect (treatment or spillover) can be achieved by using a splitting criterion minimizing the MSE of that causal effect. On the contrary, if our goal is to identify a partition of the covariate space that can explain the heterogeneity of multiple causal effects, we propose the use of a composite splitting function that is designed to minimize the MSE of all the effects. This flexibility allows scholars and policy-makers to customize their investigations depending on their targeting goals. For instance, if a policy-maker wants to target individuals with the highest treatment effect and the lowest spillover effect (with the motivation that those with higher spillover effects can benefit from the treatment received by others) the NCT algorithm would be implemented with the use of a composite splitting function, to detect specific subsets of the population where both treatment and spillover effects are heterogenous and the decision criterion can be applied. Conversely, if a targeting strategy is designed to target just individuals who would benefit the most from receiving the treatment, regardless of other people's assignment, a tree would be build using a single splitting criterion targeted to minimize the MSE of the treatment effect. Similarly, a single criterion targeted to a spillover effect would be used in the case of targeting strategies only involving that spillover effect.

It is important to note that the use of our algorithm to design implementation strategies is possible thanks to its high level of interpretability. Our network causal trees provide interpretable inference on heteroge-

neous treatment and spillover effects by discovering a set of causal rules that can be represented through a binary tree. As argued by K. Lee et al., 2018 and K. Lee et al., 2020 it is important to provide interpretable information on simple causal rules that can be targeted to improve policy effectiveness and to ensure that stakeholders and policy-makers understand (and, in turn, trust) the functionality of these models. Valdes et al., 2016 argue that a learning algorithm is interpretable if one can explain its classification by a conjunction of conditional statements (i.e., if-then rules). In this regard, tree-based algorithms based on if-then rules, such as the proposed NCT, are optimal for interpretability.

To assess the performance of the proposed NCT algorithm, we run a series of Monte Carlo simulation. In particular, we investigate the performance of the proposed algorithm with respect to two dimensions: its ability (i) to correctly identify the actual heterogeneous sub-populations and, (ii) to precisely estimate the conditional treatment and spillover effects. While the latter performance assessment is quite standard in the literature, the former is critical for interpretable algorithms for heterogeneous causal effects (Bargagli-Stoffi et al., 2019).

Finally, we apply are NCT algorithm to a randomized experiment conducted in China to assess the impact of information sessions on the purchase of a new weather insurance policy (Cai et al., 2015). Besides estimating the population average treatment and spillover effects (as already investigated in Cai et al., 2015), our aim is to detect the strata of the population where one or both effects are heterogeneous and estimate these effects within these strata.

The remainder of the paper is organized as follows. In Section 3.2 we introduce the notation, setting, and assumptions that we employ throughout the paper. In Section 3.3 we define the conditional causal effects in a general partition of the covariate space and develop a Horvitz-Thomson estimator. Section 3.4 presents the proposed network causal tree algorithm, which is based on effect-specific or composite splitting functions for causal effects under interference. We then conduct a simulation study to assess the performance of the algorithm and estimator under different scenarios in Section 3.5 and we illustrate the application

of the network causal tree on a randomized experiment in Section 3.6. Section 3.7 concludes the paper with a discussion of the proposed algorithm and directions for further research.

3.2 Clustered Network Interference and Unit-Level Randomization

3.2.1 Notation and Setting

Let us consider a sample $\mathcal{N}$ of N units organized in K separate clusters $\mathcal{K}$. Let k ($k = 1, \ldots, K$) be the cluster indicator and let $i = 1, \ldots, n_k$ be the unit indicator in each cluster k. Let now consider a connection structure such that units belonging to the same cluster might share a link whereas units belonging to different clusters are not connected. This network structure is represented by the graph $G = (N, \mathcal{E})$, where $\mathcal{N}$ defines the set of nodes and $\mathcal{E}$ defined the set of edges, that is, the collection of links between each connected pair of nodes. In a clustered network G is in turn an ensemble of K disjoint sub-graphs: $G_k = (\mathcal{N}_k, \mathcal{E}_k), \quad k = 1 \ldots K$. The adjacency matrix $\boldsymbol{A}$ corresponding to the graph G, is a block-diagonal matrix with K blocks, $\boldsymbol{A}_k, \quad k : 1 \ldots K$, where each element $a_{ij,k}$ is equal to 1 if there is a link between unit i and unit j in cluster k, that is if the edge $\epsilon_{ij,k} \in \mathcal{E}_k$.

Let now $Z_{ik} \in \{0, 1\}$ be a binary variable representing the treatment assigned to unit i in cluster k and let Y_{ik} be the observed outcome. We denote by $\mathbf{Z}_k$ and $\mathbf{Y}_k$ the treatment and outcome vectors in each cluster k. Similarly, $\mathbf{Z}$ and $\mathbf{Y}$ denote the treatment and outcome vectors in the whole sample. It worth noting that treatment status is not the same for all units belonging to the same cluster. This treatment allocation corresponds to a unit-level randomization where treatment is assigned to each unit of a cluster independently (as in a Bernoulli trial) or with some level of dependency (as in a completely randomized trial) based on a probability distribution $P(\mathbf{Z}_k)$ (See Section 3.2.3 for further details. Moreover, for each unit ik we observe a vector $\mathbf{X}_{ik}$ of P covariates (or pre-treatment variables) that are not influenced by the treatment assignment. The vec-

tor of covariates might include individual characteristics (e.g., age, sex, socio-economic status, ...), cluster-level characteristics (e.g., cluster size, location, ...), as well as network characteristics representing aggregated individual characteristics (e.g., average age or proportion of males and females, ...) or the network topology (e.g., degree, centrality, transitivity, ...). Figure 23 provides a graphical intuition on the clustered network structure and treatment assignments at the unit-level. Edges indicate links between units, within each cluster. Colors refer to the individual treatment assignment: red colored nodes represent treated units, while green colored vertices signal individuals who have been assigned to the control group.

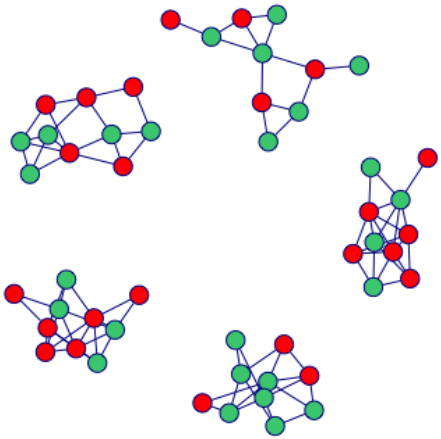


Figure 23: Clustered Network Structure

3.2.2 Clustered Network Interference

Following the potential outcome framework (Holland, 1986; Rubin, 1974), we denote by $Y_{ik}(\mathbf{z})$, with $z \in \{0, 1\}$ the potential outcome that unit i in cluster k would experience if the treatment vector in the whole sample were $\mathbf{z}$. Under the assumption of no-interference, the potential outcome could be indexed only by the individual treatment assignment Z_{ik},

that is, $Y_{ik}(Z_{ik} = z)$. In combination with the assumption of *consistency*, this assumption is known as Stable Units Treatment Assumption (SUTVA)(Rubin, 1986). The no-interference assumption is clearly violated in many real-world scenarios. For instance, the evaluation of the effect of introducing vaccines against some disease can be affected by the presence of unprotected individuals that remain healthy because they have less probability to be in contact with other infected people.

Here, we focus on a particular type of interference: *clustered network interference*. As we will show later, focusing on this type of interference is critical to ensure asymptotic properties of the estimator for conditional causal effects as well as to allow the network causal tree to divide the sample into a training set and an estimation set (and a testing set, if applicable). [2] The assumption of clustered network interference implies that: i) interference is restricted to nodes of the same cluster and interference between clusters is ruled out, that is, one's outcome is only affected by the treatment received by units belonging to the same cluster; ii) interference occurs through a function of the cluster treatment $\mathbf{Z}_k$, that is, one's outcome will depend on a summarizing function of $\mathbf{Z}_k$, not on the whole vector.

Let $Z_{k/i}$ be the vector collecting the treatment status of all units in cluster k except unit i. Let $g(\cdot) : \{0,1\}^{n_k-1} \longrightarrow \Delta_{ik}$ be a function that maps a cluster assignment vector $\mathbf{Z}_{k/i}$ to an exposure value. We define it as a function of the dot product between the cluster assignment vector and a vector of weights $\boldsymbol{\delta}_i(A_k, \mathbf{X}_k)$, which in turn depends on the adjacency matrix A_k and the covariate matrix $\mathbf{X}_k$, i.e., $g(\mathbf{Z}_{k/i}, \boldsymbol{\delta}_i(A_k, \mathbf{X}_k)) = f(\mathbf{Z}_{k/i} \cdot \boldsymbol{\delta}_i(A_k, \mathbf{X}_k))$. For instance, the function $g(\cdot)$ could result in the number or proportion of treated units in a cluster. In this case the weight vector would be equal to $\boldsymbol{\delta}_i(A_k, \mathbf{X}_k)) = \mathbf{1}_{n_k-1}$ or $\boldsymbol{\delta}_i(A_k, \mathbf{X}_k)) = (\frac{\mathbf{1}}{n_k-1})_{n_k-1}$, respectively. Alternatively we could use the adjacency matrix to compute the geodesic distance $d(i,j)$ between each pair of nodes in cluster k

[2] Alternatively, network causal trees could also be extended to the case of one single network as long as the amount of dependency is limited (to ensure the consistency of the estimator), and the network has a high level of clustering and could be approximated by separate communities that could be identified using a community detection algorithm (to divide the sample into the different sets needed for the causal tree algorithm)

and let $g(\mathbf{Z}_{k/i}, \boldsymbol{\delta}_i(A_k, \mathbf{X}_k)) = \sum_{j=1}^{n_k-1} \frac{Z_{jk}}{d(i,j)}$. The function $(\cdot)$ is similar to the 'effective treatments' function in C. Manski, 2013 and the 'exposure mapping' function in Aronow and Samii, 2017, although it applies to the cluster treatment vector only. To ease notation, throughout we will omit the weight vector $\boldsymbol{\delta}_i(A_k, \mathbf{X}_k))$ in the function $g(\cdot)$. We can now formalize the clustered network interference assumption as follows.

Assumption 8 (Clustered Network Interference)**.** *Given a function* $g(\cdot) : \{0,1\}^{n_k-1} \longrightarrow \Delta_{ik}$, $\forall k \in \mathcal{K}$, $\forall i \in \mathcal{N}_k$, *and* $\forall\ \mathbf{Z}, \mathbf{Z}' \in \{0,1\}^N$ *such that* $Z_{ik} = Z'_{ik}$, $g(\mathbf{Z}_{k/i}) = g(\mathbf{Z}'_{k/i})$, *the following equality holds:* $Y_{ik}(\mathbf{Z}) = Y_{ik}(\mathbf{Z}')$.

Again, Assumption 8 states that the outcome of a unit i in cluster k depends on the *individual treatment* Z_{ik} and a function of the treatment status of the other members of cluster k, i.e., $g(\mathbf{Z}_{k/i})$, regardless of the specific treatment status of each member. This assumption can be viewed as an intermediate assumption between (i) assuming no interference and (ii) making no assumptions about the nature of interference. In a way, it is similar to the *partial interference* or the *stratified interference* in Hudgens and Halloran, 2008, which are special cases of the clustered network interference assumption, with $g(\mathbf{Z}_{k/i}) = \mathbf{Z}_{k/i}$ and $g(\mathbf{Z}_{k/i}) = \sum_{j=1}^{n_k} Z_{ij}$.

Let $G_{ik} = g(\mathbf{Z}_{k/i})$, referred to throughout as *network exposure*. Under Assumption 8, each unit has $|\Delta_{ik}| \times 2$ potential outcomes, which we can write in terms of the individual treatment and the network exposure exposure as $Y_{ik}(z, g)$, representing the potential outcome of unit ik under $Z_{ik} = z$ and $Gik = g(\mathbf{Z}_{k/i}) = g$.

We also assume the following consistency assumption:

Assumption 9 (Consistency)**.**

$$Y_{ik} = Y_{ik}(Z_{ik}, G_{ik})$$

This assumption rules out different versions of the treatment and different ways in which a value of the network exposure can affect the outcome of a particular unit. Under a 'finite sample perspective', we assume the potential outcomes of each unit to be fixed but unknown, except the observed $Y_{ik}(Z_{ik}, G_{ik})$. Therefore, the only source of randomness in the potential outcomes is given by the random assignment to the treatment

and the random network exposure induced by the random cluster assignment.

Assumptions 8 and 9 together are alternative to SUTVA when interference is present and is limited to within clusters. When the weight function $\boldsymbol{\delta}_i(A_k, \mathbf{X}_k))$ is such that elements $\delta_{ij}(A_k, \mathbf{X}_k)) = 0$ if $j \in \mathcal{N}_k : a_{ij,k} = 0$, that is, units that are not directly connected to unit i receive a weight equal to zero, then interference is limited to the neighborhood $\mathcal{N}_{ik}$ of each unit, with $\mathcal{N}_{ik} = \{j \in \mathcal{N}_k : a_{ij,k} = 1\}$. In this case, Assumptions 8 and 9 correspond to the SUTNVA Assumption in Forastiere et al., 2020. We denote by $\mathcal{N}_{ik}^g$ the set of units defining the network exposure, that is, $\mathcal{N}_{ik}^g = \{j \in \mathcal{N}_k : \text{if } Z'_{jk} \neq Z_{jk} \text{ then } g(\mathbf{Z}'_{k/i}) \neq g(\mathbf{Z}_{k/i}),\ \forall Z'_{hk}, Z_{hk}, h \neq j\} = \{j \in \mathcal{N}_k : \delta_{ij}(A_k, \mathbf{X}_k)) \neq 0\}$. In most of the literature on spillover effects this is set is either a cluster k (Hudgens & Halloran, 2008) or the neighborhood of unit i (Forastiere et al., 2020). Alternative specifications are also possible and might involve higher-order neighbors. Here, we consider an 'exposure mapping' function $g(\cdot)$ such that $g(\cdot) : \{0,1\}^{n_k - 1} \longrightarrow \Delta_{ik} \subset \mathbb{Z}$, where $\mathbb{Z}$ is the set of integers. In particular, we rely on the following assumption:

Assumption 10 (Discrete Network Exposure)**.** *We assume that $G_{ik} \in \Delta_i \subset \mathbb{Z}$, that is, the network exposure is a discrete variable*

For instance, we can define a binary network exposure based on a threshold function applied to the number of treated neighbors:

$$G_{ik} = \mathbb{1}\big(\big(\sum_{j \in \mathcal{N}_{ik}} Z_{jk}\big) \geq h\big), \tag{3.1}$$

where h is a threshold. Hence, the $g(\cdot)$ exposure mapping function behaves like a threshold function which sums the elements of its argument (i.e the treatment assignment vector that characterizes the neighborhood of each unit) taking the value of 1 if the resulting value exceeds a certain threshold (e.g., at least one treated neighbor is treated, the majority of the neighbors are treated, ...). In our simulation study as well as in the application we have chosen the following definition: $G_{ik} = \mathbb{1}\big(\big(\sum_{j \in \mathcal{N}_{ik}} Z_{jk}\big) \geq 1\big)$, that is, the network exposure is 1 if at least

one network neighbor is treated. As a consequence, both the individual treatment and the network exposure are defined as binary variables, $Z_{ik} \in \{0,1\}$ and $G_{ik} \in \{0,1\}$. It follows that the support of the joint treatment variable (Z_{ik}, G_{ik}) is finite and comprises four possible realizations, given by the combination of the two marginal domains. Hence, $(Z_{ik}, G_{ik}) \in \{(z,g) = (0,0), (1,0), (0,1), (1,1)\}$. A discrete network exposure is crucial for our causal tree algorithm, at least in the version proposed in this paper. Indeed, the algorithm relies on the presence of enough observations for each treatment and exposure value to allow the estimation of the causal effects. Depending on the stopping rule that might rely on the accuracy of the estimation of conditional effects or on the number of observations (see Section 3.4), if the sample size is not large enough with respect to the number of categories of the network exposure and/or its distribution is non-uniform and highly skewed, the network causal tree algorithm might result in a tree with low depth and low granularity, that is, with highly heterogeneous causal effects even within the terminal leaves. Therefore, the maximum number of categories for the network exposure depend on the sample size, the number of covariates and their nature, as well as on the extent of the heterogeneity in the causal effects. It follows, that an eventual researcher, who intends to apply this methodology under more complex definitions of the joint treatment, should put his effort in credibly reducing the dimensionality of the problem. For instance, one could define a set of plausible exposure thresholds and implement the proposed algorithm, after having defined the binary network exposure variable, according to various values of these thresholds. By following this strategy, the researcher is also able to evaluate the robustness of results with respect to different characterizations of the treatment variable.

3.2.3 Unit-Level Randomization and Induced Joint Distributions

In this work, we consider an experimental design with a unit-level randomization of the treatment, which is independent between clusters but

might dependent within them. Therefore, the treatment vector $\mathbf{Z}$ is a random vector with probability distribution $P(\mathbf{Z} = \mathbf{z})$ and the following assumption holds.

Assumption 11 (Independent treatment allocation between clusters)**.**

$$P(\mathbf{Z} = \mathbf{z}) = \prod_{k=1}^{K} P(\mathbf{Z}_k = \mathbf{z}_k)$$

where $\mathbf{Z}_k$ is the treatment vector in each cluster k.

We denote by π_{ik}^{Z} the unit-level probability that Z_{ik} is equal to 1, under the experimental design in place. In a randomized experiment π_{ik}^{Z} is known. In the case of a Bernoulli trial, where each unit is independently assigned to the individual treatment, π_{ik}^{Z} is constant and equal to α. [3] An example of a design with randomization independent between clusters but dependent within clusters is that of a completely randomized experiment taking place in each cluster. In this case, $\pi_{ik}^{Z} =$ would be equal to m/n_k, where m is fixed number of treated units, and the treatment assignment for each unit does depend on the treatment status of other units. Since the network exposure is a deterministic function $g(\cdot)$ of the cluster assignment vector $\mathbf{Z}_{k/i}$, then the randomization distribution $P(\mathbf{Z} = \mathbf{z})$ induces, together with the definition of the function $g(\cdot)$, a probability distribution of the vector of network exposures $\mathbf{G}$ in the whole sample. Hence, the probability for a unit of being exposed to a specif value of the network exposure $G_{ik} = g$ given the individual treatment z, denoted by $\pi_{ik}^{G|Z}(g|z)$, is known and can in principle be computed from the probability distribution $P(\mathbf{Z})$. Note that, when the randomization is independent between units, we can drop the dependency from the individual treatment and write $\pi_{ik}^{G}(g)$. Let $\Delta = \{0, 1\} \times \bigcup_{ik \in \mathcal{N}} \Delta_{ik}$ be the domain of the joint individual and network treatment status, that is, $(z, g) \in \Delta$. Let $\pi_{ik}(z, g)$ denote the *marginal probability* for unit ik of being assigned to individual treatment z and being exposed to the network status g. This is equal to the expected proportion of assignment vectors inducing an individual treatment z and a

[3] The unit-level assignment probability could also vary across clusters as in a two-stage randomization

network exposure g:

$$\begin{aligned}\pi_{ik}(z,g) = &\sum_{\mathbf{z}\in\{0,1\}^N} \mathbb{1}(Z_{ik}=z, G_{ik}=g)P(\mathbf{Z}=\mathbf{z}) \\ =&(\pi_{ik}^z)^z(1-\pi_{ik}^z)^{1-z} \times \pi_{ik}^{G|z}(g|z)\end{aligned}$$

This marginal probability is a crucial component of the Horvitz-Thomson estimator for causal effects under network interference. If, for instance, the experimental design is a Bernoulli trial with unit-level probability α and the network exposure is defined by a threshold function on the neighborhood as in Equation 3.1, then the joint probability could be computed as follows:

$$\begin{aligned}\pi_{ik}(z,g) =&\alpha^z(1-\alpha)^{1-z} \times \left[1-\sum_{l=0}^{h-1}\binom{N_{ik}}{l}p^l(1-p)^{N_{ik}-l}\right]^g \\ &\times \left[\sum_{l=0}^{h-1}\binom{N_{ik}}{l}p^l(1-p)^{N_{ik}-l}\right]^{1-g}\end{aligned}$$

where N_{ik} is the number of neighbors ('degree') of unit ik. To deal with well-defined potential outcomes, we must assume that each unit has a nonzero probability of being exposed to each (z,g):

Assumption 12 (Positivity). $\pi_{ik}(z,g) > 0 \quad \forall i \in \mathcal{N},\ k \in \mathcal{K}$ *and* $\forall (z,g) \in \Delta$.

When $\pi_{ik}(z,g) = 0$ for some units, then the average potential outcomes and causal effects involving these values z and g must be restricted to the subset of units for which $\pi_{ik}(z,g) > 0$. For instance, if the network exposure is defined as in Equation 3.1, then the positivity assumption is violated for units who cannot be exposed to a value g, that is, those with a degree N_{ik} lower than the threshold h. Consequently, the analysis must be restricted only to the subset of the population satisfying the positivity criterion. The estimator that we propose below also requires the so-called *pairwise exposure probabilities*, which describe the joint probability of pairs of units being exposed to a given individual treatment and network status. Hence, given specific exposure conditions (z,g) and (z',g') , a

pairwise exposure probability, denote by $\pi_{ikjh}(z, g; z', g')$, quantifies the probability that the two events $(Z_{ik} = z, G_{ik} = g)$ and $(Z_{jh} = z', G_{jh} = g')$ occur, i.e., $\pi_{ikjh}(z, g; z', g') = P(Z_{ik} = z, G_{ik} = g, Z_{jh} = z', G_{jh} = g')$. In general, this can be written as:

$$\pi_{ikjk'}(z, g, z', g') = \sum_{\mathbf{z} \in \{0,1\}^N} \mathbb{1}(Z_{ik} = z, G_{ik} = g, Z_{jk'} = z', G_{jk'} = g') P(\mathbf{Z} = \mathbf{z})$$

Under the event of both units being exposed to the same condition (z, g) we denote the pairwise exposure probability by $\pi_{ikjh}(z, g)$.

In the case of an experimental design assigning treatment independently between clusters, under the clustered network interference the two events $(Z_{ik} = z, G_{ik} = g)$ and $(Z_{jh} = z', G_{jh} = g')$, with $k \neq h$, are independent and the pairwise exposure probability equals the product of the two joint probabilities: $\pi_{ikjh}(z, g; z', g') = \pi_{ik}(z, g) \times \pi_{jh}(z', g')$. In the of a Bernoulli trial and the network exposure defined on the neighborhood only, this is also true for units belonging to the same cluster, i.e., $k = h$, but with neighborhoods $\mathcal{N}_{ik}$ and $\mathcal{N}_{jk}$ not overlapping, that is when i and j are not connected and do not share any neighbors.

Note that if $\pi_{ik}(z, g)$ or $\pi_{jh}(z', g') = 0 \Rightarrow \pi_{ikjh}(z, g; z', g') = 0$, but not the reverse. Indeed, the joint probability of the two events $(Z_{ik} = z, G_{ik} = g)$ and $(Z_{jh} = z', G_{jh} = g')$ might be zero if the network exposures G_{ik} and G_{jh} are defined on two subsets of units that coincide or include jh and ik, respectively. For example, if the network exposure is defined as in Equation 3.1 with threshold equal to 1 (i.e., having at least one treated neighbor) then, if unit ik is treated, with Z_{ik} and belongs to the neighborhood $\mathcal{N}_{jh}$ of unit jh, the network exposure G_{jh} cannot be 0.

3.3 Conditional Treatment and Spillover effects and Horvitz-Thomson Estimator

3.3.1 Conditional Treatment and Spillover Effects

In this section we will define our causal effects of interests. Our ultimate goal is to *detect* the regions of the covariate space exhibiting a high

level of heterogeneity in the causal effects and *estimate* the causal effects of interest in these heterogenous regions. In this section we will focus on the definition and estimation of conditional treatment and spillover effects and we will assume that the heterogenous regions that we want to investigate have already been identified, either a priori according to subject-matter knowledge or thanks to data-driven methods.

Let us denote with Π a partition of the covariate space $\mathcal{X}$ into M non-overlapping regions : $\Pi = \{\ell_1, \ldots, \ell_M\}$, where $\bigcup_{m=1}^{M} \ell_m = \mathcal{X}$, and with $\ell(\mathbf{x}, \Pi) : \mathcal{X} \rightarrow \Pi$ a function that maps each vector $\mathbf{x}$ of the covariate space into a region. Let $N(\ell_m)$ be the size of each region ℓ_m, with $m = 1, \ldots, M$, and let $\mathcal{N}_k(\ell_m)$ be the subset of units belonging to region ℓ_m in cluster k, with $k = 1, \ldots, K$. In the machine learning literature on CART, these non-overlapping regions are referred to as *leaves*. For consistency, throughout we will use this terminology, regardless of whether the partition Π has been a priori defined or is the result of a tree-based algorithm. In addition, to ease notation, we will drop the reference to the partition Π from the mapping function $\ell(\cdot)$.

When units are organized in a network, it is worth noting that a partition Π of the covariate space partitions the sample units into subpopulations according to similarities in their characteristics, regardless of their network distance. Hence, two units connected units might belong to different regions of the partition. However, in an homophilous network, where the probability of forming a link depends on the similarity in certain features and, hence, connected units are likely to share similar characteristics,a partition of the covariate space is also likely to cluster together connected units.

Given a partition Π, we now define conditional average potential outcomes under each individual treatment and network exposure condition $(z, g) \in \Delta$. For the subset of units $\mathcal{S}_m$ with covariate vectors $\mathbf{x} \in \mathcal{X}$ that are mapped to the same region by the function $\ell(\mathbf{x})$, i.e., $\mathcal{S}_m = \{ik \in N : \ell(\mathbf{X}_{ik}) = \ell_m\}$, we define the leaf-specific average potential outcome under treatment and exposure condition $(z, g) \in \Delta$ as

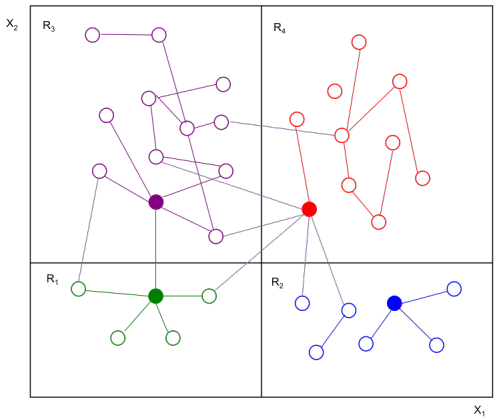


Figure 24: Partition of the covariate space with connected units.

follows:

$$\mu_{(z,g)}(\ell(\mathbf{x})) = \frac{1}{N(\ell(\mathbf{x}))} \sum_{k=1}^{K} \sum_{i=1}^{n_k} Y_{ik}(z,g) \mathbb{1}(\mathbf{X}_{ik} \in \ell(\mathbf{x})) \tag{3.2}$$

Note that $\mu_{(z,g)}(\ell(\mathbf{x}))$ is a sample average, that is, it is the average potential outcomes for all units in the sample N with a covariate vector mapped to the same region $\ell(\mathbf{x})$.

Leaf-specific conditional average average causal effect (CACE) can be defined by comparing average potential outcomes under two different conditions:

$$\begin{aligned}\tau_{(z,g;z',g')}(\ell(\mathbf{x})) =& \frac{1}{N(\ell(\mathbf{x}))} \sum_{k=1}^{K} \sum_{i=1}^{n_k} Y_{ik}(z,g) \mathbb{1}(\mathbf{X}_{ik} \in \ell(\mathbf{x})) - \\ & \frac{1}{N(\ell(\mathbf{x}))} \sum_{k=1}^{K} \sum_{i=1}^{n_k} Y_{ik}(z',g') \mathbb{1}(\mathbf{X}_{ik} \in \ell(\mathbf{x})) \\ =& \mu_{(z,g)}(\ell(\mathbf{x})) - \mu_{(z',g')}(\ell(\mathbf{x})). \end{aligned} \tag{3.3}$$

We denote by $\mathcal{T}$ the set of possible contrast we are interested in. For instance if both the individual treatment and the network exposure are

binary, then $\mathcal{T} \subseteq \{(1,0,0,0),(1,1,0,1),(0,1,0,0),(1,1,1,0),(1,1,0,0)\}$. We define as *leaf-specific treatment effects* causal contrasts $\tau_{(z,g,z',g')}(\ell(\mathbf{x}))$ that keep the network exposure fixed at a level g while changing the individual treatment from z' to z, that is, when $g = g'$. These represent causal effects of receiving the treatment while the treatment status of all other units is kept fixed or is mapped to the same network exposure g. On the contrary, we define as *leaf-specific spillover effects* causal contrasts $\tau_{(z,g,z',g')}(\ell(\mathbf{x}))$ that keep the individual treatment fixed at a level z while changing the network exposure from g' to g, that is, when $z = z'$. These spillover effects can be seen as causal effects of a change in the treatment status of other units such that the network exposure also changes, while the individual treatment status is kept fixed. It should be emphasized that $\mu_{(z,g)}(\ell(\mathbf{x}))$ corresponds to a unit-level intervention setting the treatment and network exposure of each unit to specific values. The focus on these type of average potential outcomes, as opposed to the ones based on population-level hypothetical interventions as in Hudgens and Halloran, 2008, is due to our purpose of investigating heterogeneous responses to the individual treatment and network status across units with different characteristics. If interested in assessing the heterogeneity of the average response to the network exposure resulting from a hypothetical treatment allocation, our approach could be extended to marginalized causal effects as the ones in Forastiere et al., 2020.

3.3.2 Estimator for leaf-specific CACE

Here we develop an Horvitz-Thomson estimator for leaf-specific conditional average causal effects. The derivation of the proposed estimator builds upon the estimator for average causal effects under network interference proposed by Aronow and Samii, 2017. Following Horvitz and Thompson, 1952 and Aronow and Samii, 2017, a design-based estimator for the leaf-specific average potential outcome under individual treatment z and network exposure g, $\mu_{(z,g)}(\ell(\mathbf{x}))$, can be expressed as:

$$\widehat{\mu}_{(z,g)}(\ell(\mathbf{x})) = \frac{1}{N(\ell(\mathbf{x}))} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \frac{Y_{ik}}{\pi_{ik}(z,g)} \mathbb{1}(Z_{ik} = z, G_{ik} = g, \mathbf{X}_{ik} \in \ell(\mathbf{x}))$$

where $\pi_{ik}(z,g)$ denotes the probability of a given unit ik, that belongs to the leaf $\ell(\mathbf{x})$ (in the partition Π), to be exposed to the treatment condition (z,g).

The variance estimator of $\widehat{\mu}_{(z,g)}(\ell(\mathbf{x})$ can be expressed as:

$$
\begin{aligned}
\widehat{\mathbb{V}}\Big(\widehat{\mu}_{(z,g)}(\ell(\mathbf{x}))\Big) = & \frac{1}{N(\ell(\mathbf{x}))^2} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \mathbb{1}(Z_{ik} = z, G_{ik} = g, \mathbf{X}_{ik} \in \ell(\mathbf{x})) \\
& [1 - \pi_{ik}(z,g)] \left[\frac{Y_{ik}}{\pi_{ik}(z,g)} \right]^2 + \frac{1}{N(\ell(\mathbf{x}))^2} \times \\
& \left[\sum_{k=1}^{K} \sum_{i=1}^{n_k} \sum_{j \neq i} \mathbb{1}(Z_{ik} = z, G_{ik} = g, \mathbf{X}_{ik} \in \ell(\mathbf{x})) \right. \\
& \times \mathbb{1}(Z_{jk} = z, G_{jk} = g, \mathbf{X}_{jk} \in \ell(\mathbf{x})) \\
& \left. \times \frac{\pi_{ikjk}(z,g) - \pi_{ik}(z,g)\pi_{jk}(z,g)}{\pi_{ikjk}(z,g)} \frac{Y_{ik}}{\pi_{ik}(z,g)} \frac{Y_{jk}}{\pi_{jk}(z,g)} \right]
\end{aligned}
$$

This expression extends the variance estimator derived in Aronow and Samii, 2017 (Equation 7) to the case of conditional average potential outcomes and clustered interference. In fact, the second term in the previous equation includes the covariance between the individual treatment and network exposure of two units belonging to the same leaf $\ell(\mathbf{x})$. Under an experimental design with independent treatment allocation between clusters and under clustered interference such covariance between two units belonging to different clusters is zero and the second term should be restricted to units j in the same cluster as i. In addition, the covariance between the joint treatment of two units is non-zero if the set of units defining the network exposure, i.e., $\mathcal{N}_{ik}^{g}$, (e.g., the whole cluster or the unit's neighborhood) is shared between them or includes them. Formally, let $\mathcal{N}_{ik}^{wg} = \mathcal{N}_{ik}^{g} \cup ik$. Even under independent treatment assignment, if $\mathcal{N}_{ik}^{wg} \cap \mathcal{N}_{jk'}^{wg} \neq 0$ the joint treatment of the units ik and jk' will be dependent, that is, $\pi_{ikjk'}(z,g) - \pi_{ik}(z,g)\pi_{jk'}(z,g) \neq 0$. Hence, two units belonging to the same leaf are more likely to have intersecting sets $\mathcal{N}_{ik}^{wg}$ and $\mathcal{N}_{jk'}^{wg}$ (e.g., shared neighbors) if the sets are homogeneous,

that is, units belonging to these sets share similar characteristics. Settings with homophilous networks are investigated in the appendix. An estimator for the leaf-specific conditional average causal effect of the exposure condition (z, g) compared with the configuration (z', g') can be written as:

$$\widehat{\tau}_{(z,g;z',g')}(\ell(\mathbf{x})) = \widehat{\mu}_{(z,g)}(\ell(\mathbf{x})) - \widehat{\mu}_{(z',g')}(\ell(\mathbf{x}), \Pi). \tag{3.4}$$

The estimated variance of the estimator $\widehat{\tau}_{(z,g;z',g')}(\ell(\mathbf{x}))$ can be decomposed as follows:

$$\begin{aligned}\widehat{\mathbb{V}}\Big(\widehat{\tau}_{(z,g;z',g')}(\ell(\mathbf{x}))\Big) =& \widehat{\mathbb{V}}\Big(\widehat{\mu}_{(z,g)}(\ell(\mathbf{x}))\Big) + \widehat{\mathbb{V}}\Big(\widehat{\mu}_{(z',g')}(\ell(\mathbf{x}))\Big) \\ & - 2\Big[\widehat{\mathbb{C}}\Big(\widehat{\mu}_{(z,g)}(\ell(\mathbf{x})), \widehat{\mu}_{(z',g')}(\ell(\mathbf{x}))\Big)\Big].\end{aligned} \tag{3.5}$$

with the covariance estimator taking the following expression for the case when $\pi_{ikjk}(z, g, z', g') > 0 \; \forall i, j, k$:

$$\begin{aligned}\widehat{\mathbb{C}}\Big(\widehat{\mu}_{z,g}(\ell(\mathbf{x})), \widehat{\mu}_{z',g'}(\ell(\mathbf{x}))\Big) \quad = \quad & \frac{1}{N(\ell(\mathbf{x}))^2} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \sum_{j \neq i} \frac{1}{\pi_{ikjk}(z, g; z', g')} \\ \times \quad & \mathbb{1}(Z_{ik} = z, G_{ik} = g, \mathbf{X}_{ik} \in \ell(\mathbf{x})) \\ \times \quad & \mathbb{1}(Z_{jk} = z', G_{jk} = g', \mathbf{X}_{jk} \in \ell(\mathbf{x})) \\ \times \quad & [\pi_{ikjk}(z, g; z', g') - \pi_{ik}(z, g)\pi_{jk}(z', g')] \\ \times \quad & \frac{Y_{ik}}{\pi_{ik}(z, g)} \frac{Y_{jk}}{\pi_{jk}(z', g')} \\ - \quad & \frac{1}{N(\ell(\mathbf{x}))^2} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \sum_{j \neq i} \Big[\\ & \frac{\mathbb{1}(Z_{ik} = z, G_{ik} = g, \mathbf{X}_{ik} \in \ell(\mathbf{x})) Y_{ik}^2}{2\pi_{ik}(z, g)} \\ + \quad & \frac{\mathbb{1}(Z_{ik} = z', G_{ik} = g', \mathbf{X}_{ik} \in \ell(\mathbf{x})) Y_{ik}^2}{2\pi_{ik}(z', g')} \Big].\end{aligned} \tag{3.6}$$

Further details about the variance estimator of leaf-specific CACE can be found in Appendix B.2.

Properties of the Horvitz-Thomson Estimator

Here we will describe the properties of the Horvitz-Thomson estimator of leaf-specific causal effects. Asymptotic results will rely on a growth process that is commonly assumed with cluster data. In particular, we consider a sequence of nested samples $\mathcal{N}$ of size N, where $\mathcal{N}$ consists of K separate clusters $\mathcal{N}_k$ of size n_k, $k = 1, \ldots, K$. We let the sample size $N \longrightarrow \infty$ by letting the number of clusters go to infinity, i.e., $K \longrightarrow \infty$, while the cluster size n_k, $k = 1, \ldots, K$ remains fixed.

Proposition 4 (Unbiaseness)**.**

$$E\Big[\widehat{\mu}_{(z,g)}(\ell(\mathbf{x}))\Big] = \mu_{z,g}(\ell(\mathbf{x}))$$

and

$$E\Big[\widehat{\tau}_{(z,g,z'g')}(\ell(\mathbf{x}))\Big] = c(\ell(\mathbf{x}))$$

Proof. Proof in Appendix B.1. □

The unbiaseness of the estimator of leaf-specific CACE is conditional on the partition Π and the function $\ell(\cdot)$. When building causal trees to assess the heterogeneity of causal effects, we will rely on this property to derive the splitting criterion and to estimate leaf-specific causal effects. However, the unbiaseness of the estimator $\widehat{\tau}_{z,g,z'g'}(\ell(\mathbf{x})$ does not ensure the identification of subsets with the highest heterogeneity. The performance of the causal tree in identifying heterogeneous regions depends on the splitting criterion, the algorithm and the sample.

Proposition 5 (The variance estimator of $\widehat{\mu}_{(z,g)}$ is unbiased)**.** *If* $\pi_{ikjk}(z,g) > 0 \ \forall i, j, k$ *then*

$$\mathbb{E}\Big[\widehat{\mathbb{V}}\Big(\widehat{\mu}_{(z,g)}(\ell(\mathbf{x}))\Big)\Big] = \mathbb{V}\Big(\widehat{\mu}_{(z,g)}(\ell(\mathbf{x})\Big).$$

The proof follows directly from the unbiaseness of the Horvitz-Thomson estimator. A conservative estimator for the case when $\pi_{ikjk}(z,g) = 0$ for some units can be found in Appendix B.2.

Proposition 6 (The variance estimator of $\widehat{\tau}_{(z,g;z',g')}$ is conservative)**.**

$$\mathbb{E}\Big[\widehat{\mathbb{V}}\Big(\widehat{\tau}_{(z,g;z',g')}(\ell(\mathbf{x}))\Big)\Big] \geq \mathbb{V}\Big(\widehat{\tau}_{(z,g;z',g')}(\ell(\mathbf{x}))\Big).$$

A proof follows from Aronow and Samii, 2017. The restriction on the covariances does not change the proof.

Proposition 7 (Consistency)**.** *Consider the asymptotic regime where the number of clusters K go to infinity, i.e., $K \longrightarrow \infty$, while the cluster size remains bounded, i.e., $n_k \leq B(\ell(\mathbf{x})) \leq B$ for some constant B. In addition, assume that $|Y_{ik}(z,g)|/\pi_{ik}(z,g) \leq C < 1, \forall i,k,z,g$. Then as $K \longrightarrow \infty$*

$$\widehat{\tau}_{(z,g;z',g')}(\ell(\mathbf{x})) \xrightarrow{p} \tau_{(z,g;z',g')}(\ell(\mathbf{x})).$$

Proof. See Appendix B.1. □

Note that cluster network interference and independent treatment allocation between clusters ensure that the amount of dependence across units is limited. This limited independence is the condition required to ensure consistency (Aronow & Samii, 2017). [4]

Proposition 8 (Asymptotic Normality)**.** *Given an cluster independent design and the clustered network interference assumption, then:*

$$\sqrt{N(\ell(\mathbf{x}))}\Big(\widehat{\tau}_{(z,g;z',g')}(\ell(\mathbf{x}))-\tau_{(z,g;z',g')}(\ell(\mathbf{x}))\Big) \xrightarrow{d} N(0,\mathbb{V}\Big(\widehat{\tau}_{(z,g;z',g')}(\ell(\mathbf{x}))\Big)).$$

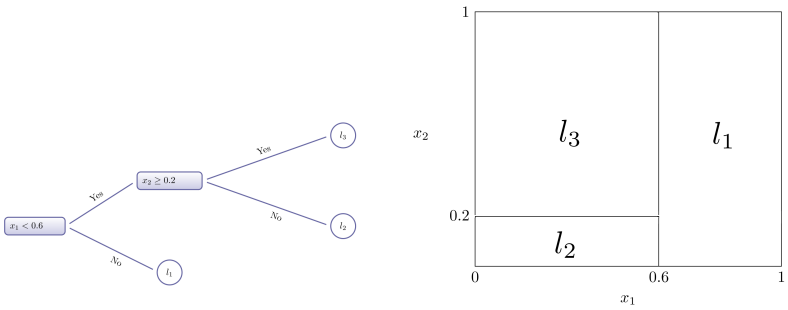
An independent treatment allocation between clusters and the clustered network interference ensure the limited dependence condition required in Aronow and Samii, 2017. This condition allows us to rely on a central limit theorem derived via Stein's method (L. H. Y. Chen & Shao, 2004) to achieve the asymptotic normality of the estimator. The variance estimators will depend on the size of the sample belonging to leaf $\ell(\mathbf{x})$ in each cluster, i.e., $n_k(\ell(\mathbf{x})) \leq B(\ell(\mathbf{x})) \leq B, k = 1 = \ldots, K$, and the maximum conditional degree $D(\ell) = \max_{ik \in \mathcal{N}:\mathbf{X}_{ik} \in \ell(\mathbf{x})} N^g_{ik}(\ell(\mathbf{x}))$, where $N^g_{ik}(\ell) = N^g_{ik} \cap \mathcal{N}_k(\ell(\mathbf{x}))$. Given that these quantities are bounded, we can show that $\mathbb{V}\Big(\mathbb{V}\Big(\widehat{\tau}_{(z,g;z',g')}(\ell(\mathbf{x}))\Big)\Big) = O(1/N(\ell(\mathbf{x})))$, that is, the rate of convergence will be $1/\sqrt{N(\ell(\mathbf{x}))}$, with $N(\ell(\mathbf{x})) \leq KB(\ell(\mathbf{x}))$ (the proof follows the one in Aronow and Samii, 2017).

[4] Note that for the variance of the estimator to go to zero as $N \longrightarrow \infty$ we must have $\sum_{ik}\sum_{jk'} \mathbb{1}(\mathbf{X}_{ik} \in \ell(\mathbf{x}), \mathbf{X}_{jk} \in \ell(\mathbf{x}))\mathbb{1}\big(\pi_{ikjk'}(z,g) - \pi_{ik}(z,g)\pi_{jk'}(z,g) \neq 0\big) = o(N(\ell(\mathbf{x}))^2)$. This is guaranteed given that the joint treatment is independent between units in different clusters, that is, $\pi_{ikjk'}(z,g) = \pi_{ik}(z,g)\pi_{jk'}(z,g), \forall k'$, and given that the cluster size is bounded.

3.4 Network Causal Trees for Heterogeneous Causal Effects under Clustered Network Interference

In the previous section we have introduced and developed an estimator for causal effects conditional on sub-populations of units defined by a partition Π of the covariate space $\mathcal{X}$. Here we develop a a data-driven machine learning algorithm to identify the partition Π at the aim of investigating the heterogeneity in the effects of interest. Our proposed algorithm, named *Network Causal Tree* (NCT), builds upon the Causal Tree (CT) algorithm introduced by Athey and Imbens, 2016, which in turn finds its roots in the Classification and Regression Tree (CART) algorithm (Friedman et al., 1984). CART is a widely used nonparametric method to partition the feature space. It relies on a tree-based algorithm which recursively splits the sample. In particular, trees are constructed by recursively partitioning the observations from the *root* (that contains all the observations in the learning samples) into two child *nodes*. This procedure is repeated until the tree reaches the final nodes which are called *leaves*. Because each node is always split into two sub-nodes, these trees are called *binary trees*. Binary trees are called *regression trees* when the outcome is a continuous variable, while they are called *classification trees* when the outcome is either a discrete or a binary variable. The aim of the tree construction is to identify heterogeneities in the relationship between the observed outcome and the features to best predict the outcome variable. Therefore, splits are made with the aim of minimizing the prediction error. With this aim different splitting criteria could be specified For additional details on CART, we refer to the seminal paper by Friedman et al., 1984. Figure 25 illustrates an example of binary partitioning in a simple case with just two predictors $x_1 \in [0, 1]$ and $x_2 \in [0, 1]$.

Building on CART, Athey and Imbens, 2016 developed a causal decision tree algorithm with the aim of detecting the of causal effects. In particular, they modified the splitting function to minimize the estimation error of conditional effects. Moreover, Athey and Imbens, 2016 introduced honest inference by using a sub-sample to build the tree (discov-



(a) Undirected Network (b) Directed Network

Figure 25: (Left) An example of a binary tree. The internal nodes are labelled by their splitting rules and the terminal nodes are labelled with the corresponding parameters l_i.
(Right) The corresponding partition of the sample space.

ery sample) and a separate sub-sample to perform inference (estimation sample). This sample-splitting approach is transparent and furthermore even efficient in high-dimensional settings (K. Lee et al., 2020).

Our proposed NCT differs form the standard causal tree algorithm in two critical aspects: (i) it estimates heterogeneous causal effects – both treatment and spillover effects – in the presence of clustered network interference, and (ii) it possibly models heterogeneity with respect to more than one effect at the same time through a composite splitting criterion. In this Section, we describe and motivate the splitting criteria for our NCT algorithm (Subsection 3.4.1) and its detailed structure (Subsection 3.4.2).

3.4.1 Splitting Criteria

The NCT algorithm is built to detect and estimate heterogeneous treatment and spillover effects, in the presence of clustered network interference. Moreover, NCT is able to discover heterogeneity driven by more than one estimand at the same time. Here, we present the three criteria that rule the splitting procedure of NCT: the first two of them lead

to trees targeted to single effects, while the last criterion sprouts a tree which is targeted to multiple effects.

Let $\mathbb{P}$ be the space of partitions. Given a causal effect $\tau_{(z,g,z',g')}$, we can use recursive splitting to look for the best partition $\Pi \in \mathbb{P}$ with respect to a splitting criterion $Q_{(z,g,z',g')}(\Pi)$. Formally:

$$\Pi = \text{argmax}_{\Pi \in \mathbb{P}} \, Q_{(z,g,z',g')}(\Pi) \tag{3.7}$$

Given our goal of describing the relationship between the causal effect and the covariate space and detecting subsets that exhibit a high level of heterogeneity, we can define a splitting criterion that maximizes accuracy in the prediction of conditional effects $\tau_{(z,g,z',g')}(\mathbf{X}_{ik})$ in the whole sample N. This translates into the minimization of expected value of the mean square error (MSE):

$$\begin{aligned} Q_{(z,g,z',g')}(\Pi) &= -EMSE\Big(\widehat{\tau}_{(z,g,z',g')}(\ell(\mathbf{x}), \Pi)\Big) \\ &= -\mathbb{E}\Big[\Big(\tau_{(z,g,z',g')}(\mathbf{x}) - \widehat{\tau}_{(z,g,z',g')}(\ell(\mathbf{x}, \Pi)\Big)^2\Big] \end{aligned} \tag{3.8}$$

the expected value is taken over the sampling distribution. When this splitting criterion is used to select the partition Π, we maximize the function in (3.8) evaluated in the sample used to build the tree, i.e., the *training set*. In the machine learning literature in this case the objective function is referred to as the *in-sample splitting function* and we denote this by $Q^{in}_{(z,g,z',g')}(\Pi)$. As opposed to the EMSE of the observed outcome prediction, the true causal effect $\tau_{z,g,z',g'}(\mathbf{x})$ is unknown. However, we can use the training data to estimate the EMSE for the in-sample spitting rule. Thanks to the unbiaseness of the estimator $\widehat{\tau}_{(z,g,z',g')}(\ell(\mathbf{x}), \Pi)$ with respect to the population causal effect $E[\tau_{(z,g,z',g')}(\mathbf{X}_{ik})|\mathbf{X}_{ik} \in \ell(\mathbf{x})]$ (Proposition 9 in the Appendix), following Athey and Imbens, 2016 we can estimate the EMSE as follows:

$$\begin{aligned} Q^{in}_{(z,g,z',g')}(\Pi) &= -\widehat{EMSE}\Big(\widehat{\tau}_{(z,g,z',g')}(\ell(\mathbf{x}), \Pi)\Big) \\ &= \frac{1}{N^{tr}} \sum_{k \in \mathcal{K}^{tr}} \sum_{i=1}^{n_k} \big(\widehat{\tau}_{(z,g,z',g')}(\ell(\mathbf{X}_{ik}, \Pi))\big)^2 \end{aligned} \tag{3.9}$$

where $\mathcal{K}^{tr}$ is the subset of clusters belonging to the training set and N^{tr} is the sample size [5]. Therefore, the maximization of this splitting function results in the maximization of the heterogeneity across leaves. In fact, if two sub-populations ℓ_1 and ℓ_2 have a different causal effect $\tau_{(z,g,z',g')}$, i.e., $\tau_{(z,g,z',g')}(\ell_1) \neq \tau_{(z,g,z',g')}(\ell_2)$, a partition Π that splits them would yield a higher $Q_{(z,g,z',g')}(\Pi)$ than the partition Π^c that combines the two sub-populations into one leaf ℓ_{1+2} (a simple proof can be found in the appendix). To avoid using the same information for selecting the partition and for the estimation, Athey and Imbens, 2016 propose to estimate the effects in a separate sample than the one used to build the tree. They call this an 'honest' causal tree. We denote by $\mathcal{N}^{est}$ the estimation set and by $\mathcal{N}^{tr}$ the training set (or discovery set) that is used to build the tree (by evaluating the splitting function). The training set and the estimation set are here obtained by taking two random subsets $\mathcal{K}^{tr}$ and $\mathcal{K}^{est}$ of the clusters $\mathcal{K}$. Hence, $N^{est} = \bigcup_{k\in\mathcal{K}^{est}} \mathcal{N}_k$ and $N^{tr} = \bigcup_{k\in\mathcal{K}^{tr}} \mathcal{N}_k = \mathcal{N}/\mathcal{N}^{est}$. This random split of the sample avoids any dependencies between training and estimation sub-samples. In this 'honest' version, the splitting function can be estimated as follows:

$$
\begin{aligned}
Q^{in,H}_{(z,g,z',g')}(\Pi) =& \frac{1}{N^{tr}} \sum_{k\in\mathcal{K}^{tr}} \sum_{i=1}^{n_k} \left(\hat{\tau}_{(z,g,z',g')}(\ell(\mathbf{X}_{ik},\Pi))\right)^2 \\
&- \left(\frac{1}{N^{tr}} + \frac{1}{N^{est}}\right) \sum_{\ell\in\Pi} \widehat{\mathbb{V}}\left(\widehat{\tau}_{(z,g,z',g')}(\ell;\Pi)\right)
\end{aligned}
\tag{3.10}
$$

where $N^{est} = |\mathcal{N}^{est}|$. The proof follows from Athey and Imbens, 2016. In (3.10) we can see that the splitting function is such that splits will be chosen so to maximize the heterogeneity across leaves as well as to minimize the average variance in the estimated effect. The idea is to identify the most heterogeneous partitions while introducing a penalization term which corrects the objective function to minimize the leaf-specific variation in the estimated effect. This penalization term has also

[5] In standard CART the training set is a subset of the whole sample together with the testing set, which is used to evaluate the objective function in order to choose the best partition selected in the training set that maximizes out-of-sample prediction accuracy.

the effect of reducing the depth of the tree depending on the sample size, because leaves with a small number of observations $N(\ell)$ will exhibit a higher variance. In addition to this penalization we will also add a stopping rule which is required to avoid having leaves where the effect $\tau_{(z,g,z',g')}$ cannot be estimated because there are no observations with observed treatment (Z_{ik}, G_{ik}) equal to the values (z, g) or (z', g') that we are comparing.

We now introduce a composite splitting rule targeted to multiple causal estimands. When interference is in place targeting strategies might involve both treatment and spillover effects. For example, in settings with limited resources the treatment should be provided to those who would benefit from it, i.e., with a non-zero treatment effect, whereas we could save resources by not giving the treatment to those who would benefit from other people being treated, i.e., with high spillover effects. This is the case in marketing interventions where we can provide advertisements only to those who would be affected and who are less likely to get the information from someone else. Another interesting example can be found in the potential challenges of the COVID-19 vaccine distribution, which represents an illustrative scenario where seeding strategies play a role. Those at high risk of getting infected even if those in close contact were immune.i.e., those with high treatment effect under treated neighbors, should be targeted. On the contrary, those who are in contact with a low number of people and can greatly gain from having one of these contacts vaccinated could be left without the vaccine, at least in the early stages of the distribution.

For these kind of targeting strategies involving more than one causal effect we must partition the population into sub-groups that show high level of heterogeneity in all estimands of interest. Building a separate tree for each causal effect would provide us with different partitions that cannot be used for the design of multi-effect strategies. Therefore, we propose a composite splitting function that would result in a tree that maximises heterogeneity in all the causal estimands of interest. This composite objective function is a weighted average of the effect-specific

splitting functions:

$$Q_{\mathcal{T}}(\Pi) = \sum_{(z,g,z',g') \in \mathcal{T}} \gamma_{(z,g,z',g')} Q_{(z,g,z',g')}(\Pi) \quad \text{with} \quad \gamma_{(z,g,z',g')} = \frac{\omega_{(z,g,z',g')}}{\left(\hat{\tau}^{(z,g,z',g')}\right)^2} \tag{3.11}$$

where $\omega_{(z,g,z',g')} \in [0,1]$ is a customized weight for each estimand and $\hat{\tau}_{(z,g,z',g')}$ is the estimated effect in the whole sample. Each effect $\tau_{(z,g;z',g')}$, where $(z,g;z',g') \in \mathcal{T}$, contributes to the global objective function according to a specific weight $\gamma_{(z,g;z',g')}$. The element $\gamma_{(z,g;z',g')}$ is proportional to a customized weight $\omega_{(z,g,z',g')}$, which is set by the researcher according to the extent to which the estimand $\tau_{(z,g,z',g')}$ is of interest, and is normalized by the the estimated effect in the whole sample to rule out any dependence on the magnitude of the effect. The composite criterion requires that at least two of the four weights are strictly greater than zero. A similar composite objective function can be derived from the splitting functions for the 'honest' causal trees:

$$Q^H_{\mathcal{T}}(\Pi) = \sum_{(z,g,z',g') \in \mathcal{T}} \gamma_{(z,g,z',g')} Q^H_{(z,g,z',g')}(\Pi) \quad \text{with} \quad \gamma_{(z,g,z',g')} = \frac{\omega_{(z,g,z',g')}}{\left(\hat{\tau}^{(z,g,z',g')}\right)^2} \tag{3.12}$$

3.4.2 Network Causal Tree Algorithm

Compared with the standard HCT algorithm, its main novelties are the introduction of interference and the possibility of including more than one effect. Specifically, the extent to which each effect $\tau_{(z,g;z',g')}$, with $(z,g;z',g') \in \mathcal{T}$, contributes to the determination of the tree is specified by the weight $w(z,g;z',g')$. Here we describe the key steps of the NCT algorithm, including the recursive partitioning based on the splitting functions and the stopping rules.

Key steps of the NCT algorithm

The proposed algorithm takes mainly six elements as inputs.

- The sample $\mathcal{N}$, which collects for each unit ik the individual treatment assignment status Z_{ik}, the observed outcome Y_{ik} and a vector of characteristics $\mathbf{X}_{ik}$;

- The network information, which is fully described by the global adjacency matrix $\boldsymbol{A}$, including the cluster-specific blocks $\boldsymbol{A}_k$;

- The specification of the exposure mapping function $g()$ which together with the adjacency matrix and possibly covariate matrix will be translated in the computation of the observed network exposure G_{ik} for each unit;

- The experimental design which will determine the computation of the probabilities $\pi_{ik}(z, g)$ and $\pi_{ik}(z, g, z', g')$.

- The fifth aspect pertains to the weight $\omega_{z,g,z'g'}$ for each causal effect;

- The specification of the two parameters: *maximum depth*, that is the maximum depth of the tree, and the *minimum size*, that is the minimum number of units falling in each exposure condition (z, g) in each leaf.

After some preliminary steps, the algorithm consists of two main phases. The first phase is focused on the selection of the partition, i.e. the tree, while the second phase is concerned with the estimation of causal effects and returns point estimates and standard errors of the conditional average causal effects, for all the comparisons of interest and within each leaf of the detected partition. We report below the key steps of the NCT algorithm:

0. **Phase 0** (Preliminaries): In a preliminary stage the algorithm computes some quantities and tools that will be used in the subsequent steps.

   (a) Given the adjacency blocks $\boldsymbol{A}_k$ and potentially the covariate matrix $\boldsymbol{X}$, for each unit the *network exposure* variable G_{ik} is computed according to the rule expressed in Assumption 10.

(b) The joint exposure probabilities $\pi_{ik}(z, g)$ and $\pi_{ikjk'}(z, g, z', g')$ (as in Subsection 3.2) or estimated (as in Aronow and Samii, 2017).

(c) Finally, the algorithm randomly splits the clusters between the training set $\mathcal{N}^{tr}$ and the estimation set N^{est}. [6]

1. **Phase 1** (Tree Discovery): the first step of the algorithm sprouts the Network Causal Tree, that is, it detects the relevant heterogeneous partitions. Note that this step is performed over the discovery set only. In particular, the NCT algorithm works with the clusters belonging to the set $\mathcal{K}^{tr}$ and builds the tree using a binary recursive partitioning.

   (a) Recursive Partitioning. The algorithm *grows a tree* by maximizing the in-sample splitting criterion at each binary split. At iteration $r-1$The partition can be written as follows:

$$\Pi^{r-1} = \{\mathbf{x} \in \mathcal{X} : \bigcap_{m=1}^{r-1} x^m \in \mathcal{A}_m^{h_m}\}_{\mathbf{h} \in \{L,R\}^{r-1}}$$

   where $x^m \in \{x_p\}_{p=1,\dots,P}$ is the feature that was split at iteration m and $\mathcal{A}_m^L = \{x^m \leq c_m\}$, $\mathcal{A}_m^L = \{x^m \; c_m\}$ for some cutoff point c_m. The variable x^m split at iteration m together with the cutoff point c_m compose a *node* of the tree. At iteration r, the partition will be complemented with a split of a variable $x^r \in \{x_p\}_{p=1,\dots,P} / \{x^m\}_{m=1\dots,r-1}$ at some cutoff point c_r:

$$\Pi^r = \{\mathbf{x} \in \mathcal{X} : \bigcap_{m=1}^{r-1} x^m \in \mathcal{A}_m^{h_m} \bigcap x^r \in \mathcal{A}_r^{h_r}\}_{\mathbf{h} \in \{0,1\}^r}$$

   Among all the candidate splits x^r and c_r, the algorithm will choose the one that maximizes the *in-sample* splitting function in (3.9) or (3.10).

[6] Following Athey and Imbens, 2016 we suggest to assign half of the clusters to the discovery sample and another half to the estimation sample.

(b) Stopping Rule. The recursive partitioning stops when at least one *stopping condition* is met (i) the NCT has reached the specified *maximum depth*; (ii) the current split r generate at least one leaf ℓ where the set of units $matchalN(\ell)\_(z,g) = \{ik \in N^{tr} : \mathbf{X}_{ik} \in \ell, Z_{ik} = z, G_{ik} = g\}$ with a number of observations $|N(\ell)_{(z,g)}|$ lower than the specified *minimum size*, for at least one exposure condition (z, g).

This step generates a network causal tree which corresponds to a partition Π of the feature space $\mathcal{X}$ into M leaves: $\Pi = \{\ell_1, \ldots, \ell_M\}$, with $\bigcup_{m:1}^{M} \ell_m = \mathcal{X}$ and $l(\mathbf{x}, \Pi) : \mathcal{X} \to \Pi$.

2. **Phase 2** (Estimation): the second phase of the algorithm takes as input the Network Causal Tree Π built in Phase 1 and computes all the point estimates, the standard errors and the confidence intervals of the leaf-specific causal effects of interest in all its nodes $\ell(\mathbf{x}, \Pi)$. This is done using the Horvitz-Thomson estimator in Section 3.3. In the 'honest' version, at this stage the NCT algorithm works with the clusters belonging to the set $\mathcal{K}^{est}$.

3.5 Simulation Study

Our algorithm provides an interpretable method to detect and estimate heterogeneous effects in the presence of clustered network interference. In this section we evaluate through a set of simulations the performance of the proposed algorithm with respect to both discovery and estimation. In particular, we investigate its ability to correctly identify the actual heterogeneous sub-populations, comparing the use of single or composite splitting functions, and we assess the performance of the Horvitz-Thomson estimator for leaf-specific treatment and spillover effects. While the latter performance assessment is quite standard in the literature, the former is critical for the development of interpretable algorithms for heterogeneous causal effects (Bargagli Stoffi & Gnecco, 2019; K. Lee et al., 2020). We evaluate the performance of the algorithm and the estimator

Algorithm 1 Overview of the NCT algorithm

- **Inputs**: i) Observed data $\{Z_{ik}, Y_{ik}, \mathbf{X}_{ik}\}_{ik \in \mathcal{N}}$; ii) Global adjacency matrix $\boldsymbol{A}$, which comprises the cluster-specific blocks $\boldsymbol{A}_k$; iii) Experimental Design; iv) vector of weights $\omega(z, g; z', g')$, where $(z, g; z', g') \in \mathcal{T}$; v) Tree parameters: *maximum depth* and *minimum size*.
- **Outputs**: (1) a partition Π if the covariate space, and (2) point estimates, standard errors and confidence intervals of the conditional average causal effects:
    1. Phase 0 (Preliminaries): compute G_{ik} and both the marginal and joint exposure probabilities $\pi_{ik}(z, g)$ and $\pi_{ikjk'}(z, g, z', g')$. Then, randomly assign clusters to discovery and estimation samples.
    2. Phase 1 (Tree Discovery): build a tree according to the in-sample splitting criterion and stop when either the tree has reached its maximum depth or any additional split would generate leaves, which are not sufficiently representative of the four exposure conditions.
    3. Phase 2 (Estimation) on the estimation sample): Use the Horvitz-Thomson estimator to estimate the leaf-specific CACE and their standard errors in each leaf.

in settings that differ with respect to three main factors: (1) the structure of the heterogeneity, (2) the level of the effect heterogeneity, and (3) the number of clusters. In Appendix B.3, we also consider two additional factors: (4) the correlation structure in the covariate matrix and (5) the presence of homophily in the network structure. Regarding the structure of the heterogeneity, we are particularly interested in settings where the structure of the causal tree representing heterogeneity is different for each causal effect. In particular, causal trees differ if they have different nodes corresponding to the split of a feature, that is, if covariates driving the heterogeneity are different, or if they have different terminal leaves where the causal effect is heterogenous, i.e., non-zero. We call *causal rule*

these heterogenous terminal leaves.

For each simulation scenario we simulated $M = 500$ samples and applied our NCT algorithm to detect sub-populations with heterogeneous causal effects (or causal rules) and to estimate our causal effects of interest. To evaluate the performance of our composite splitting function under different settings, splits rely on either effect-specific splitting criteria or on the composite function.

All simulations are performed under Bernoulli trials, that is treatment is randomly assigned independently to each unit with a fixed probability $\pi_{ik}^{W} = \alpha$. In our simulation study we also assume that interference only takes place at the neighborhood level and we choose the following definition of the network exposure:

$$G_{ik} = \mathbb{1}\big(\big(\sum_{j \in \mathcal{N}_{ik}} Z_{jk}\big) \geq 1\big), \tag{3.13}$$

that is, the network exposure of unit ik is 1 if at least one neighbor is treated. A binary network exposure together with a binary individual treatment results in a joint treatment with four categories, i.e., $(Z_{ik}, G_{ik}) \in \{(z, g) = (0,0), (1,0), (0,1), (1,1)\}$.

The binary definition of the network exposure is chosen to allow the growth of deeper trees. Given that the minimum size requirement stops the algorithm when the number of units in a child leaf is not enough to estimate a conditional causal effect, a joint treatment with four categories ensures that this stopping condition is unlikely to be met during the first few splits. In addition, the assumption of neighborhood interference allows the computation of the marginal and joint probabilities without the need for intensive estimation procedures. In fact, the approximate algorithm for estimating the marginal and joint probabilities proposed by Aronow and Samii, 2017 is computationally demanding and could not be incorporated in our simulation study. However, in the case of Bernoulli trial and network exposure defined as in (3.13), the probability $\pi_{ik}(z, g)$ can be computed using the formula in (3.2.3) while the joint probability $\pi_{ikjk'}(z, g, z', g')$ is simply the product of $\pi_{ik}(z, g)$ and $\pi_{jk'}(z', g')$ if the two units ik and jk' are independent. On the contrary, if $\mathcal{N}_{ik}^{wg}$ and

$\mathcal{N}_{jk'}^{wg} \neq 0$ overlap the joint treatment of the units ik and jk' will be dependent, that is, $\pi_{ikjk'}(z, g, z', g') \neq \pi_{ik}(z, g)\pi_{jk'}(z', g')$. In this case the joint probability can still be readily computed using combinatorics formulas on two overlapping sets.

3.5.1 Data generating process

For each simulation $m = 1 \ldots, M$ we generated a K clusters and within each cluster we simulate Erdos-Renyi Random Graph with $n_k = 100$ nodes and a fixed probability (0.01) to observe a link. Given the definition of the network exposure we remove isolated nodes from the analysis to make the Assumption 12 hold. Z_{ik} and any covariate X_{ip} are sampled from independent Bernoulli distributions with probability 0.5

$$W_i \sim Ber(0.5) \quad \text{and} \quad X_{ip} \sim Ber(0.5)$$

. In the simulation study we focus on two main effects: the pure treatment effect $\tau_{(1,0,0,0)}$ and the pure spillover effect $\tau_{(0,1,0,0)}$. To ease notation, We denote by τ the treatment effect and by δ the spillover effect. Once these two effects are set depending on the simulation scenarios, we generate the four different potential outcomes as follows:

$$\begin{aligned}
Y_{ik}(0,0) &\sim \mathcal{N}(0,1); \\
Y_{ik}(1,1) &\sim \mathcal{N}(0,1); \\
Y_{ik}(1,0) &= Y_{ik}(0,0) + \tau(\mathbf{X}_{ik}); \\
Y_{ik}(0,1) &= Y_{ik}(0,0) + \delta(\mathbf{X}_{ik});
\end{aligned}$$

Finally, the observed outcome is generated as:

$$Y_{ik} = \sum_{w=0}^{1} \sum_{g=0}^{1} \mathbb{1}(Z_{ik} = z, G_{ik} = g) Y_{ik}(z, g).$$

We now detail how we varied the three factors (1), (2), and (3). We simulated two different scenarios with respect to the heterogeneity struc-

ture (1). In the first scenario we have:

$$\tau(\mathbf{X}_{ik}) = \begin{cases} h & \text{if} \quad \mathbf{X}_{ik} \in \ell_1 = \{X_{ik1} = 0, X_{ik2} = 0\} \\ -h & \text{if} \quad \mathbf{X}_{ik} \in \ell_2 = \{X_{ik1} = 1, X_{ik2} = 1\} \\ 0 & \text{otherwise} \end{cases}$$

$$\delta(\mathbf{X}_{ik}) = \begin{cases} h & \text{if} \quad \mathbf{X}_{ik} \in \ell_1 = \{X_{ik1} = 0, X_{ik2} = 0\} \\ -h & \text{if} \quad \mathbf{X}_{ik} \in \ell_2 = \{X_{ik1} = 1, X_{ik2} = 1\} \\ 0 & \text{otherwise} \end{cases}$$

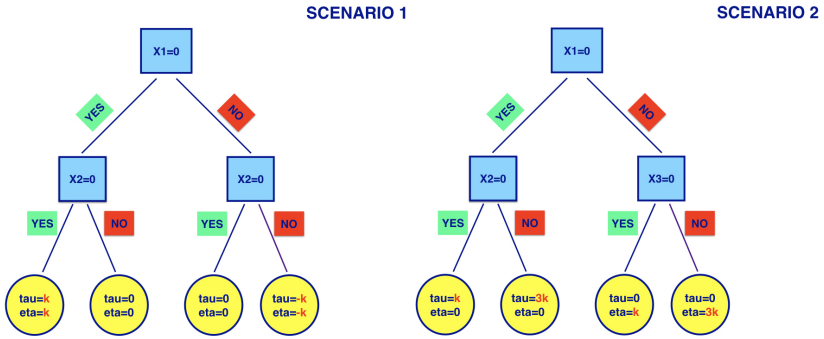
Hence, in this scenario the heterogeneity driving variables (HDV), i.e., X_{i1} and X_{i2}, are the same for both the treatment effect τ and the spillover effect δ and the two causal rules overlap. In the second scenario, we introduce a change in the drivers of the heterogeneity in the following way:

$$\tau(\mathbf{X}_{ik}) = \begin{cases} h & \text{if} \quad \mathbf{X}_{ik} \in \ell_{\tau 1} = \{X_{ik1} = 0, X_{ik2} = 0\} \\ 3h & \text{if} \quad \mathbf{X}_{ik} \in \ell_{\tau 2} = \{X_{ik1} = 0, X_{ik2} = 1\} \\ 0 & \text{otherwise} \end{cases}$$

$$\delta(\mathbf{X}_{ik}) = \begin{cases} h & \text{if} \quad \mathbf{X}_{ik} \in \ell_{\delta 1} = \{X_{ik1} = 1, X_{ik2} = 0\} \\ 3h & \text{if} \quad \mathbf{X}_{ik} \in \ell_{\delta 2} = \{X_{ik1} = 1, X_{ik2} = 1\} \\ 0 & \text{otherwise} \end{cases}$$

Hence, in the second scenario the heterogeneity drivers are different for the two causal effects. Specifically, we have: X_{i1} and X_{i2} for the treatment effects, and X_{i1} and X_{i3} for the spillover effects. In addition, we have two causal rules for the treatment effect, namely $\{X_{ik1} = 0, X_{ik2} = 0\}$ and $\{X_{ik1} = 0, X_{ik2} = 1\}$ and two different causal rules for the spillover effect, namely $\{X_{ik1} = 1, X_{ik2} = 0\}$ and $\{X_{ik1} = 1, X_{ik2} = 1\}$. For each structural scenario we varied the effect size: $(h \cdot 0.1)_{k=1}^{10}$ with $h \in \mathbb{N}$ Figure 26 graphically represents the two simulations' scenarios. Moreover, we changed the number of clusters keeping their size fixed to $K = (10, 20, 30)$. [7] For each scenario we construct three NCTs: one tree implementing the composite splitting rule for the treatment and spillover

[7]Note that K/2 clusters will be assigned to the discovery sample and the remaining clusters will be in the estimation set as in Athey and Imbens, 2016.



(a) First scenario (b) Second scenario

Figure 26: Simulations' scenarios

effects as in (3.12), one tree implementing the singular splitting rule for the treatment effect $Q^{in,H}_{(1,0,0,0)}(\Pi)$ as in 3.10, and one tree implementing the singular splitting rule for the spillover effect $Q^{in,H}_{(0,1,0,0)}(\Pi)$ as in (3.10).

We believe that the two examined simulated scenarios, together with the additional two that will be presented in Appendix B.3, sufficiently represent the ensemble of plausible scenarios that the researcher may actually face in real world data. Indeed, the illustrated scenarios not only differ for the structure and the magnitude of heterogeneity but also account for the number of clusters, for the eventual statistical correlation among predictors and for the network structure.

3.5.2 Performance measures

In all the scenarios the performance of the NCT algorithm is evaluated using the following measured averaged over the M generated datasets:

- average number of correctly discovered heterogeneous causal rules corresponding to the leaves of the generated NCTs (reported with respect to the effect sizes in Figures 27 and 28) in the discovery sample;
- average conditional treatment and spillover effects estimated for

each correctly detected heterogenous terminal leaf (reported as $\hat{\tau}$ and $\hat{\delta}$ in Tables 3, 4, 5 and 6);

- average standard errors estimated for each correctly detected heterogenous terminal leaf (reported as $\hat{se}(\hat{\tau})$ and $\hat{se}(\hat{\delta})$ in Tables 3, 4, 5 and 6);

- Monte-Carlo Bias in the estimation sample:

$$\begin{aligned}\text{Bias}_m(\mathcal{N}^{est}) =& \frac{1}{N^{\text{est}}} \sum_{k \in \mathcal{K}^{est}} \sum_{i=1}^{n_k} \Big(\tau_{(z,g,z'g')}(\mathbf{X}_{ik}) - \\ & \widehat{\tau}_{(z,g,z'g')}(\ell(\mathbf{X}_{ik}, , \Pi_m), \mathcal{N}^{est}) \Big),\end{aligned}$$

$$\text{Bias}(\mathcal{N}^{est}) = \frac{1}{M} \sum_{m=1}^{M} \text{Bias}_m(\mathcal{N}^{est});$$

  where Π_m is the partition selected in simulation m.

- Monte-Carlo MSE in the estimation sample:

$$\begin{aligned}\text{MSE}_m(\mathcal{N}^{est}) =& \frac{1}{N^{\text{est}}} \sum_{k \in \mathcal{K}^{est}} \sum_{i=1}^{n_k} \Big(\tau_{(z,g,z'g')}(\mathbf{X}_{ik}) - \\ & \widehat{\tau}_{(z,g,z'g')}(\ell(\mathbf{X}_{ik}), \Pi_m, \mathcal{N}^{est}) \Big)^2,\end{aligned}$$

$$\text{MSE}(\mathcal{N}^{est}) = \frac{1}{M} \sum_{m=1}^{M} \text{MSE}_m(\mathcal{N}^{est});$$

- Coverage, computed as the average proportion of units for whom where the estimated 95% confidence interval of the causal effect in the assigned leaf includes the true value:

$$\begin{aligned}\text{C}_m(\mathcal{N}^{est}) =& \frac{1}{N^{\text{est}}} \sum_{k \in \mathcal{K}^{est}} \sum_{i=1}^{n_k} \mathbb{1} \Big(\tau_{(z,g,z'g')}(\mathbf{X}_{ik}) \\ & \in \widehat{\text{CI}}_{95} \Big(\widehat{\tau}_{(z,g,z'g')}(\ell(\mathbf{X}_{ik}, \Pi_m), \mathcal{N}^{est}) \Big) \Big),\end{aligned}$$

$$\text{C}(\mathcal{N}^{est}) = \frac{1}{M} \sum_{m=1}^{M} \text{C}_m(\mathcal{N}^{est});$$

3.5.3 Results

We will first analyze the ability of the algorithm to correctly detect the heterogeneous subgroups in the first simulation scenario, that is, when the heterogeneity is the same for the two causal effects of interest. Figure 27 reports the average number of correctly discovered heterogeneous causal rules with composite splitting rule or effect-specific splitting rules targeted to the treatment effect or the spillover effect, in the case of 10, 20 and 30 clusters. As you can see from the three plots in the Figure the algorithm is always able to detect all the 4 correct leaves for both effects with all the splitting rules. As the number of cluster grows the minimum effect size allowing the algorithm to optimally discover all the heterogeneous sub-populations gets lower.

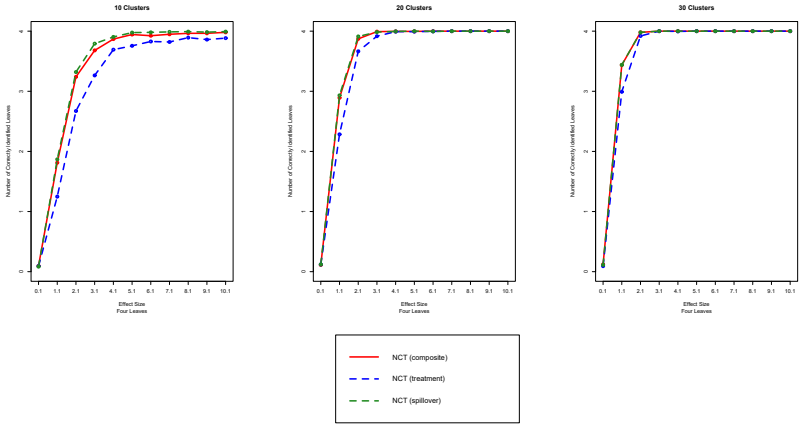


Figure 27: Simulations' results for correctly discovered leaves in the first scenario with 10, 20 and 30 clusters, respectively.

Tables 3, 4, 5 report the results for the first scenario 6 for the performance of the estimator in the correctly detected leaves. We only report the results of the estimation procedure on the tree built with the composite splitting rule, as the spitting rule would only affect the identification of the heterogenous sub-populations but not the estimation of

the causal effects once these sub-populations are correctly detected. The estimator is able to estimate the heterogeneous treatment and spillover effects without bias and its precision grows as the number of clusters increases. Interestingly, NCT provides more accurate estimates of the heterogeneous spillover effects than the treatment effects. This is partly due to the larger number of units with $(Z_{ik} = 0, G_{ik} = 1)$ than those with $(Z_{ik} = 1, G_{ik} = 0)$, by definition of the network exposure. Hence, this reduces the standard error of the estimator for the spillover effects. The higher precision in the estimation of the spillover effects is also reflected in the identification of the heterogenous subgroups, which is more accurate when splits are targeted to the minimization of the MSE of the spillover effect (Figure 27).

Table 3: Simulations' results for the first scenario (10 clusters)

	Treatment Effects						
Effect Size	$\hat{\tau}_{\ell_1}$	$\hat{se}(\hat{\tau}_{\ell_1})$	$\hat{\tau}_{\ell_2}$	$\hat{se}(\hat{\tau}_{\ell_2})$	MSE	Bias	Coverage
0.1	0.353	0.416	-0.139	0.350	0.106	0.107	1.000
1.1	1.108	0.514	-1.132	0.515	0.261	-0.012	0.936
2.1	2.124	0.740	-2.156	0.741	0.570	-0.016	0.941
3.1	3.004	1.003	-3.050	0.997	0.934	-0.023	0.945
4.1	4.061	1.293	-4.179	1.287	1.843	-0.059	0.928
5.1	5.201	1.565	-5.197	1.585	2.297	0.002	0.946
6.1	6.179	1.889	-6.086	1.824	3.282	0.046	0.931
7.1	7.125	2.140	-7.110	2.128	4.276	0.007	0.921
8.1	8.011	2.378	-8.149	2.425	5.087	-0.069	0.931
9.1	9.153	2.729	-9.128	2.713	6.803	0.012	0.927
10.1	9.991	2.947	-9.916	2.974	8.007	0.038	0.934
	Spillover Effects						
	δ_{ℓ_1}	$\hat{se}(\delta_{\ell_1})$	δ_{ℓ_2}	$\hat{se}(\delta_{\ell_2})$	MSE	Bias	Coverage
0.1	0.225	0.380	-0.249	0.322	0.082	-0.012	0.958
1.1	1.114	0.415	-1.103	0.402	0.155	0.005	0.960
2.1	2.085	0.542	-2.138	0.545	0.230	-0.026	0.967
3.1	3.047	0.716	-3.094	0.716	0.408	-0.023	0.972
4.1	4.098	0.892	-4.079	0.882	0.646	0.009	0.966
5.1	5.119	1.063	-5.054	1.057	0.862	0.032	0.958
6.1	6.090	1.250	-6.045	1.242	1.132	0.022	0.958
7.1	7.086	1.437	-7.099	1.442	1.542	-0.006	0.973
8.1	8.075	1.618	-8.019	1.618	1.986	0.028	0.961
9.1	8.936	1.811	-9.197	1.828	2.586	-0.130	0.957
10.1	10.078	2.023	-10.055	2.026	2.730	0.011	0.971

Table 4: Simulations' results for the first scenario (20 clusters)

	Treatment Effects						
Effect Size	$\hat{\tau}_{\ell_1}$	$\hat{se}(\hat{\tau}_{\ell_1})$	$\hat{\tau}_{\ell_2}$	$\hat{se}(\hat{\tau}_{\ell_2})$	MSE	Bias	Coverage
0.1	0.053	0.286	-0.072	0.284	0.061	-0.010	0.929
1.1	1.120	0.382	-1.102	0.377	0.145	0.009	0.954
2.1	2.102	0.538	-2.114	0.538	0.270	-0.006	0.949
3.1	3.105	0.710	-3.140	0.721	0.468	-0.018	0.946
4.1	4.055	0.905	-4.198	0.947	0.848	-0.072	0.948
5.1	5.058	1.100	-5.034	1.108	1.132	0.012	0.940
6.1	6.155	1.332	-6.034	1.306	1.609	0.061	0.933
7.1	7.058	1.519	-7.036	1.521	2.000	0.011	0.954
8.1	8.141	1.737	-8.118	1.755	2.919	0.011	0.961
9.1	9.130	1.952	-9.160	1.949	3.313	-0.015	0.959
10.1	10.088	2.150	-10.071	2.142	4.021	0.008	0.947
	Spillover Effects						
	δ_{ℓ_1}	$\hat{se}(\delta_{\ell_1})$	δ_{ℓ_2}	$\hat{se}(\delta_{\ell_2})$	MSE	Bias	Coverage
0.1	0.085	0.234	-0.139	0.255	0.064	-0.027	0.929
1.1	1.092	0.300	-1.106	0.297	0.075	-0.007	0.967
2.1	2.072	0.393	-2.083	0.396	0.132	-0.006	0.965
3.1	3.067	0.509	-3.084	0.511	0.208	-0.009	0.973
4.1	4.124	0.641	-4.088	0.639	0.296	0.018	0.971
5.1	5.118	0.772	-5.052	0.772	0.436	0.033	0.974
6.1	6.082	0.900	-6.154	0.915	0.611	-0.036	0.972
7.1	7.114	1.042	-7.101	1.040	0.802	0.007	0.968
8.1	8.023	1.183	-8.156	1.185	1.027	-0.067	0.975
9.1	9.106	1.312	-9.139	1.323	1.262	-0.017	0.979
10.1	10.210	1.465	-10.151	1.460	1.472	0.029	0.976

Table 5: Simulations' results for the first scenario (30 clusters)

	Treatment Effects						
Effect Size	$\hat{\tau}_{\ell_1}$	$\hat{se}(\hat{\tau}_{\ell_1})$	$\hat{\tau}_{\ell_2}$	$\hat{se}(\hat{\tau}_{\ell_2})$	MSE	Bias	Coverage
0.1	0.092	0.238	-0.068	0.230	0.016	0.012	1.000
1.1	1.083	0.311	-1.115	0.306	0.091	-0.016	0.950
2.1	2.101	0.435	-2.107	0.436	0.170	-0.003	0.949
3.1	3.104	0.584	-3.113	0.588	0.305	-0.005	0.953
4.1	4.086	0.754	-4.114	0.746	0.546	-0.014	0.946
5.1	5.168	0.921	-5.170	0.931	0.794	-0.001	0.956
6.1	6.110	1.091	-6.059	1.074	1.041	0.025	0.956
7.1	7.133	1.259	-7.135	1.251	1.422	-0.001	0.960
8.1	8.078	1.420	-7.952	1.409	1.729	0.063	0.946
9.1	9.199	1.618	-9.047	1.580	2.264	0.076	0.961
10.1	10.148	1.763	-10.171	1.765	2.862	-0.012	0.958
	Spillover Effects						
	δ_{ℓ_1}	$\hat{se}(\delta_{\ell_1})$	δ_{ℓ_2}	$\hat{se}(\delta_{\ell_2})$	MSE	Bias	Coverage
0.1	0.022	0.208	-0.086	0.213	0.019	-0.032	0.967
1.1	1.082	0.249	-1.108	0.244	0.056	-0.013	0.964
2.1	2.080	0.322	-2.115	0.326	0.083	-0.017	0.974
3.1	3.094	0.423	-3.065	0.418	0.132	0.015	0.969
4.1	4.119	0.528	-4.083	0.525	0.209	0.018	0.973
5.1	5.092	0.634	-5.110	0.637	0.263	-0.009	0.983
6.1	6.085	0.745	-6.126	0.746	0.384	-0.020	0.975
7.1	7.061	0.857	-7.160	0.860	0.517	-0.050	0.977
8.1	8.101	0.974	-8.051	0.973	0.695	0.025	0.979
9.1	9.111	1.088	-9.103	1.091	0.827	0.004	0.984
10.1	10.184	1.215	-10.123	1.201	0.993	0.030	0.981

For the second scenario with different causal rules for each causal effect, we only report the results for simulations with 30 clusters. Figure 28 depicts the average number of correctly discovered heterogeneous causal rules with composite splitting rule or effect-specific splitting rules

targeted to the treatment effect or the spillover effect. When we are interested in building a tree that can represent the heterogeneity of all causal effects simultaneously (left panel), the composite splitting rule NCT is able to correctly identify all the heterogeneous causal rules (four in this example), while the other two NCT, targeted to either the treatment effect or the spillover effect, only detect the two leaves where the corresponding causal effect is heterogeneous. This is even more clear from the other two plots where we depict the ability to detect just the treatment effect rules (central panel) and the spillover effect rules (right panel). Indeed, when we are interested in subgroups that are heterogenous with respect to only one causal effect, both the effect-specif spitting rule targeted to that effect or the composite spitting function can be used and perform similarly, while the use of the effect-specif spitting rule targeted to the other effect results in a poor detection of the correct causal rules. The results from this second scenario show the clear added value of the composite splitting rule. Indeed, when the HDV are different for treatment and spillover effects implementing this splitting rule enables the researcher to correctly spot all the true causal rules *simultaneously*. Finally, Table 6 shows how the Horvitz-Thomson estimator performs well in estimating the conditional treatment and spillover effects in the two corresponding heterogenous leaves.

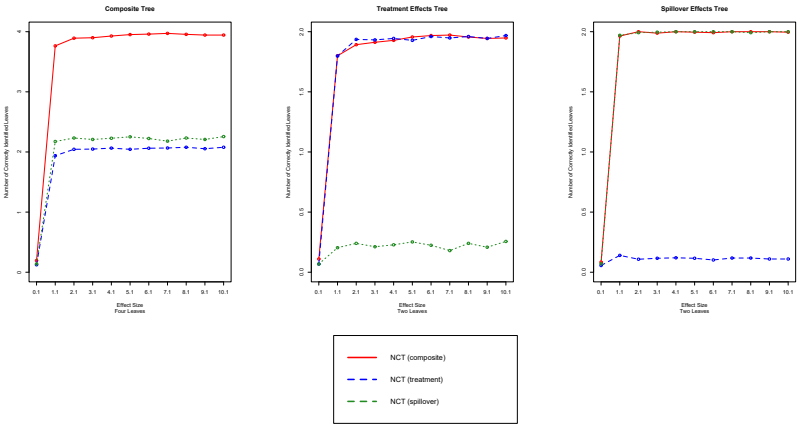


Figure 28: Simulations' results for correctly discovered leaves in the second scenario with 30 clusters.

Table 6: Simulations' results for second scenario (30 clusters)

	Treatment Effects						
Effect Size	$\hat{\tau}_{\ell_{\tau 1}}$	$\widehat{se}(\hat{\tau}_{\ell_{\tau 1}})$	$\hat{\tau}_{\ell_{\tau 2}}$	$\widehat{se}(\hat{\tau}_{\ell_{\tau 2}})$	MSE	Bias	Coverage
0.1	0.194	0.254	0.351	0.241	0.059	0.072	0.964
1.1	1.103	0.310	3.239	0.607	0.223	-0.029	0.942
2.1	2.088	0.438	6.379	1.121	0.650	0.034	0.952
3.1	3.100	0.589	9.317	1.634	1.226	0.008	0.970
4.1	4.113	0.756	12.285	2.137	2.353	-0.001	0.945
5.1	5.120	0.917	15.418	2.708	4.106	0.069	0.944
6.1	6.106	1.074	18.263	3.159	5.469	-0.015	0.947
7.1	7.086	1.241	21.312	3.712	6.246	-0.001	0.955
8.1	8.167	1.433	24.257	4.209	8.790	0.012	0.949
9.1	9.028	1.569	27.099	4.671	9.975	-0.136	0.954
10.1	10.024	1.733	30.617	5.288	14.454	0.120	0.951
	Spillover Effects						
	$\delta_{\ell_{\delta 1}}$	$\widehat{se}(\delta_{\ell_{\delta 1}})$	$\delta_{\ell_{\delta 2}}$	$\widehat{se}(\delta_{\ell_{\delta 2}})$	MSE	Bias	Coverage
0.1	0.086	0.187	0.281	0.209	0.041	-0.017	1.000
1.1	1.089	0.248	3.321	0.444	0.105	0.005	0.968
2.1	2.116	0.326	6.325	0.773	0.237	0.021	0.977
3.1	3.120	0.421	9.313	1.114	0.491	0.016	0.980
4.1	4.109	0.527	12.277	1.455	0.832	-0.007	0.972
5.1	5.123	0.637	15.407	1.816	1.441	0.065	0.976
6.1	6.119	0.746	18.341	2.159	1.990	0.030	0.972
7.1	7.074	0.858	21.393	2.516	2.177	0.033	0.979
8.1	8.095	0.976	24.312	2.854	3.168	0.003	0.974
9.1	9.048	1.090	27.264	3.210	3.953	-0.044	0.975
10.1	10.053	1.201	30.416	3.565	4.610	0.034	0.980

3.6 Empirical Application

In this section we provide an empirical application of the proposed methodology. We use data from a randomized experiment designed to assess

the effectiveness of intensive information sessions to promote the uptake of a new weather insurance policy among farmers in rural China. The promoted policy is addressed to rice farmers, and it is aimed at protecting their products from adverse weather shocks (Cai et al., 2015). The sample consists of 4,569 households belonging to 47 different villages. Households were randomly assigned to the intensive training session [8]. Households are linked according to an observed village-specific friendship network, represented in Figure 29. Relationships between households belonging to different villages are negligible.

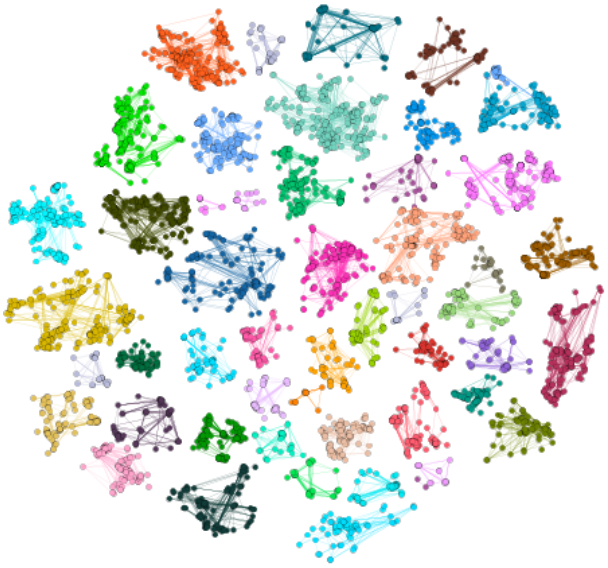
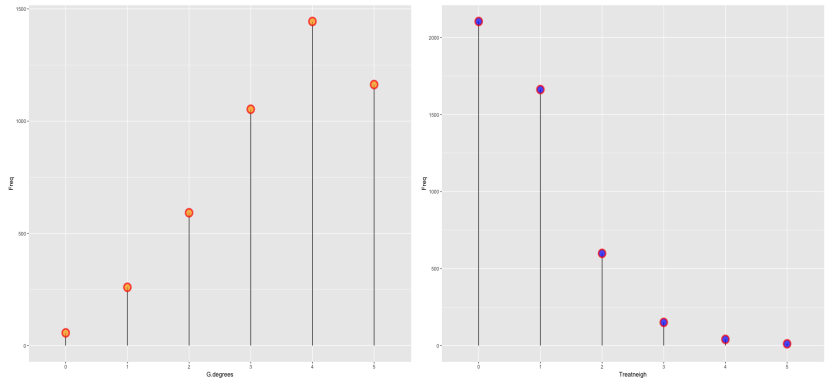


Figure 29: Friendship network between households living in rural villages in China. Colors refer to different villages.

Here the individual treatment variable $Z_{ik} \in \{0, 1\}$ for each household i living in village k represents the assignment to the intensive ($Z_{ik} =$

[8]The original experiment also includes a village-level randomization on price variation and a second round of sessions Cai et al., 2015. Here we only consider the household-level randomization to the first round of sessions.

1) or simple ($Z_{ik} = 0$) information session. Friendship relationships between households in a given rural village k are fully described by the adjacency matrix $\boldsymbol{A}_k$, where the generic element $A_k(i, j)$ equals 1 if the household i in village k has nominated the household j in the same village k. Therefore, the adjacency matrix is not symmetric. We denote with $\mathcal{N}_{ik}$ the set of nominated friends of unit ik, and with N_{ik} the out-degree. Figure 3.30(a) shows the overall degree distribution, while Figure 3.30(b) represents the distribution of the number of treated neighbors. In this setting interference could take place because the information received in the intensive information session could be transferred to network contacts. We assume here an exposure to the network treatments only at the neighborhood level and, in particular, we define the network exposure variable as in (3.13). Hence, G_{ik} is equal to 1 if a unit has at least one treated friend, and 0 otherwise. This definition of the network exposure is justifiable by the fact that the decision of a farmer to buy the weather insurance might depend only on the information either received directly in the intensive session or indirectly by one of his friends. This is also what has been found in Cai et al., 2015 where the only significant spillover effect is from first neighbors. We drop from the analysis



(a) Number of neighbors (b) Number of treated neighbors

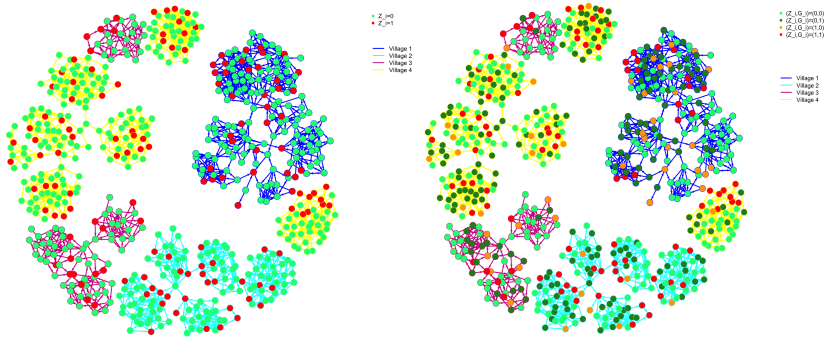
Figure 30: Degree and treated neighbors: distribution

the 57 households without any friends. The total number of remaining households is 4512, residing in 47 villages. Among them, 977 families were assigned to the intensive training sessions, while 3535 belong to the control group and, thus, they have undergone a less intensive session. 2075 households don't have any treated friends, while 2437 of them undertake friendship relationships with at least one treated household. The joint distribution of the individual and neighborhood treatments is summarized in the following Table 7.

Table 7: Distribution of the Joint Treatment

	$G_i = 0$	$G_i = 1$
$Z_i = 0$	1621	1914
$Z_i = 1$	454	523

Figure 31 represent the treatment distribution in four villages. In the left subfigure nodes are colored according to their individual treatment assignment, while in right subfigure node colors refer to the joint treatment status. The outcome variable $Y_{ik} \in \{0, 1\}$ is a binary variable which



(a) Individual treatment in four villages (b) Joint treatment in four villages

Figure 31: Zoom on four villages

equals 1 if the household i residing in village k purchased the insurance after the information session. At the end of the experiment, 2495 fami-

lies chose not to accept the proposed insurance policy, while 2017 were positively persuaded by the session and accepted the weather insurance.

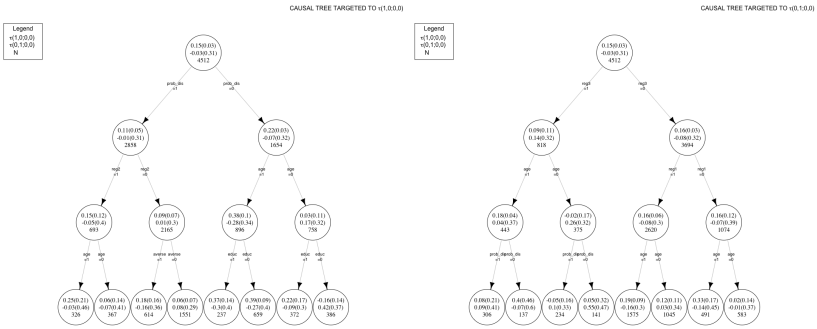
In order to evaluate the heterogeneity of treatment and spillover effects, we include in the analysis all the observed characteristics that reasonably prompt heterogeneity: three dummies representing the household's location in the production areas 1, 2 and 3 (`reg1`, `reg2` and `reg3`, respectively); a dummy variable (`age`) which equals 1 if the household's head is at least 50 years old; a dummy variable which distinguishes highly educated households (`educ`) being 1 if the household's head has successfully completed high school; a dummy variable representing families who are strongly worried about weather phenomena (`prob_dis`), which is marked as 1 if the household's perceived probability of a relevant weather disaster happening in the coming year exceeds 0.30; a dummy variable which identifies families who declare to be risk averse (`averse`).

In the remaining part of this section, we present the most relevant empirical findings. The following Figures report the Network Causal Trees targeted to single effects (Figure 32) and to multiple effects (Figure 33). We also report the estimates of the principal treatment and spillover effects and the overall size of the selected sub-populations. Each tree has been obtained according to the following rules: we have randomly assigned 20 clusters to the training set, while the remaining villages have been allocated to the estimation set.[9] Furthermore, we have set the maximum depth at 3 to maintain a high level of interpretability, while we have set at 20 the minimum number of units that must be present in the child leaves for each of the four exposure conditions.

We can see that, in the whole population, the treatment has a positive effect on insurance take up, while the spillover effect is negligible. The most relevant heterogeneity drivers result to be the perceived probability of disaster and the production area (specifically, living in the third area). However, the estimated tree slightly changes under different specifications of the spitting rule. Figures 3.32(a) and 3.32(b) sep-

[9]The random assignment of the villages to the training and estimation samples has been kept fixed in all the trees.

arately represent the trees targeted to $\tau(1,0,0,0)$ and $\tau(0,1,0,0)$. The most important variable driving the heterogeneity of the treatment effect is the perceived probability of disaster. The treatment appears to be particularly effective in the sub-population which identify less concerned and older households, with both a high or low level of education. We can speculate that younger households who are more worried about possible disasters benefit less from an intensive information session because even less information would prompt them to purchase a weather insurance. On the contrary, living in the third region is the characteristic which plays a prominent role in determining the spillover effect's variation across sub-populations. In this region younger households with a lower concern about a possible disaster are those who benefit the most from receiving the information about the weather insurance from some of their friends. Figure 33 depicts the partition selected using a compos-

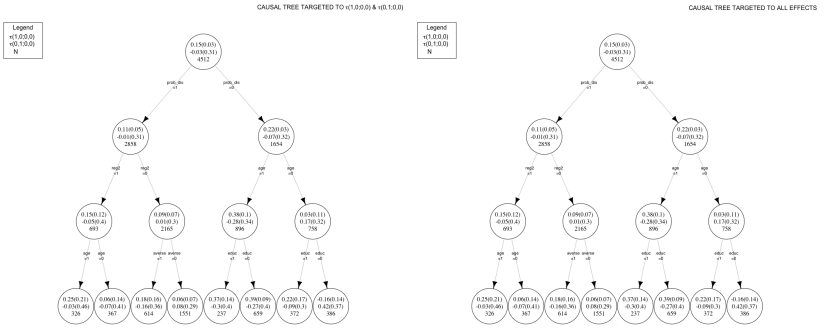


(a) NCT targeted to $\tau(1,0,0,0)$ (b) NCT targeted to $\tau(0,1,0,0)$

Figure 32: Network causal trees Targeted to single effects

ite splitting function targeted to both causal effects. Specifically, Figure 3.33(a) refers to the network causal tree targeted to both $\tau(1,0,0,0)$ and $\tau(0,1,0,0)$, such that each component contributes in determining the objective function with equal weight (0.5). Figure 3.33(b) is related to the tree which has been built assigning an equal weight to all the four effects in $\mathcal{T}$. In this application, the composite tree which considers both $\tau(1,0,0,0)$ and $\tau(0,1,0,0)$ coincides with the tree based on $\tau(1,0,0,0)$

only. Therefore, the latter leads the composite variability across partitions. Finally, the network causal tree that incorporates all the four effects shows slightly different results: here the sub-population where the treatment results to be more effective is the one including older households who are currently settled in the second production area.



(a) NCT targeted to $\tau(1,0,0,0)$ and $\tau(0,1,0,0)$

(b) NCT targeted to all effects

Figure 33: Network causal trees Targeted to multiple effects

We can conclude that intensive training sessions encouraged Chinese rural households to take up the insurance policy: the characteristics which emerge as the main determinants of the heterogeneity are the production area and age. Indirect effects do not have a significant impact in this study, and, hence, composite criteria are mostly ruled by treatment effects.

3.7 Conclusions

Depending on our characteristics we might respond differently to a treatment or intervention. Similarly, we might be more or less susceptible to the influence of other people who have experienced the treatment. Understanding the heterogeneity of the effect of a treatment with the aim of targeting people who would benefit from it has been the focus of a recent

field of research, especially applied to medicine. Investigating how different people respond differently to the treatment received by others can be crucial, particularly in settings with limited resources where spillover effects could be leveraged. In this paper, we have introduced a new algorithm to estimate heterogeneous causal treatment and spillover effects in the presence of clustered network interference. The proposed network causal tree model bridges the gap between two streams of causal inference literature: estimators for causal effects under interference and tree-based methods for the discovery of heterogeneous sub-populations. We build upon the seminal algorithm proposed by Athey and Imbens, 2016 to account for clustered network interference through a rework of the criterion function. Leaf-specific causal effects are then estimated using the Horvitz-Thomson estimator proposed by Aronow and Samii, 2017.

The proposed NCT algorithm has enhanced interpretability and shows an excellent performance in a set of Monte Carlo simulations. In particular, the algorithm is able both to spot the relevant sources of heterogeneity in the data and to consistently estimate the conditional treatment and spillover effects.

Moreover, we introduce a composite splitting function that allows the researchers to simultaneously detect the sub-populations where both treatment and spillover effects are heterogenous. The identification of these multi-effect heterogenous subgroups is crucial for the design of targeting strategies that involve multiple effects. For instance, in marketing campaigns a person might not be affected by an advertisement received directly by a company but might be susceptible to the advertisement received by her friends. In this case, resources could be saved by promoting the product among people who are could be directly susceptible and letting those who could be more influenced by their friends receive the advertisement indirectly. Our simulation study shows that the use of such composite splitting rule is able to correctly detect all the heterogenous sub-populations defined by both treatment and spillover effects, as well the ones defined by one effect only. Therefore, the selected partition of the population can be used to design strategies whose objective function incorporates multiple effects, but can also be used a posteriori

to target subgroups maximizing either a treatment or a spillover effect.

When applied to real-world data, the NCT algorithm provided useful insights on the effectiveness of intensive training sessions among Chinese rural households on the uptake of a weather insurance policy. We found that the main characteristics responsible for the heterogeneity of the effects are perception of a possible disaster, the production area, and the age of the farmers. However, these heterogeneity drivers play a different role with respect to the two main treatment and spillover effects and when the tree is built using composite splitting function. Nevertheless, the proposed algorithm may suffer from the limitations common to tree-based methods: instability to the random allocation of units in the training sample and the potential impact of outlier observations in the node-specific estimations. Here, we propose an algorithm that selects a single tree because of its high interpretability that plays a fundamental role in policy relevant scenario. Indeed, the single tree algorithms are suitable for the discovery of heterogenous sub-populations, which can give useful insights into the main variables that drive the heterogeneity and can be used to design targeting strategies. Moreover, we did not find instability in neither the detection nor the estimation of heterogeneous effects in the Monte Carlo simulations. Nonetheless, the proposed algorithm could be extended to tree-based ensemble methods, following Wager and Athey, 2018 and Athey et al., 2019. By averaging the estimates from many single trees, this extension could enhance estimation precision at the cost of reduced interpretability (see K. Lee et al., 2020 for a discussion on the trade-off between accuracy and interpretability).

In addition, our approach might be rearranged to deal with settings where the network structure is known cannot be partitioned into well-defined and pre-specified clusters. However, in some network structures, clusters could be detected by implementing a network-based community detection algorithm (Fortunato, 2010) and the NCT algorithm could be applied on the detected communities, while the estimator should take into account the uncertainty in group membership .

Furthermore, here we assume that the network exposure variable is discrete, with the performance of the algorithm being affected by the

number of categories resulting from the exposure mapping function. In our simulation study and application we have used a binary neighborhood exposure, which allowed us to grow deeper trees, reduce the number of possible causal effects, and have enough observations for each exposure condition to maintain the variance in a reasonable range. However, alternative specifications could be use. For instance, the network exposure could be defined as the proportion of treated neighbors, perhaps categorized into few bins. A more complex definition of the network exposure, possibly resulting in a continuous variable, would require some methodological adjustments in the estimation strategy, but it would allow to model a wider ensemble of real-world interference mechanisms. We leave this extension to future work.Finally, further research is needed to use the selected partition to actually design targeting strategies involving both treatment and spillover effects. Furthermore, these strategies should also rely on the average susceptibility of network contacts as well as on heterogenous influential power.

Chapter 4

Causal Effects with Hidden Treatment Diffusion over Partially Unobserved Networks

This Chapter is a joint work with my supervisors Prof. Irene Crimaldi and Prof. Fabrizia Mealli, and Prof. Laura Forastiere. The full text of the article will be soon available from the arXiv repository.

4.1 Introduction and motivation

Policy evaluation studies intend to estimate the effect of an intervention. However, in a wide variety of real world scenarios, the treatment of interest can be diffused among units (An, 2018; An & VanderWeele, 2019). This phenomenon occurs when agents interact with one another and when units have the possibility to actually spread the intervention with their interfering neighbors, by means of real or virtual social ties. For instance, promotional videos or advertising links can be shared on social media, advertising flyers can be distributed by hand to interacting individuals, even monetary incentives can be transferred among economic or social agents. It follows that in all those studies where the treatment of interest is transferable by nature and it is not reasonable to completely rule out the presence of interactions among agents, the researcher should account for the possibility that a treatment diffusion process takes place. If such a phenomenon happens, some individuals who have been originally assigned to the control group, and were not provided with treatment by design, might have actually received the intervention because of a link with treated users.

Treatment diffusion may be regarded as a specific mechanism of interference (also known as *spillover*). Indeed, a generic interference mechanism takes place when the potential outcome of a given unit is affected by the treatment assignment of other units (Cox, 1958). In the treatment diffusion setting, the potential outcomes of one unit are affected by the treatment assignment vector of her neighbors, as treated neighbors have the chance to directly spread the intervention with her. As it is possible to grasp, treatment diffusion is not the only existing mechanism of interference. Specifically, spillover effects mainly involve three different types of mechanisms that occur in different stages of the causal process: i) the direct effect of one's treatment on their own outcome coupled with the diffusion of the outcome to other individuals; ii) the indirect effect of one's treatment on the outcome of connecting units; iii) the diffusion of the treatment to interacting individuals coupled with the effect of indirectly receiving the treatment on one's own outcome. Generic spillovers

may arise in a wide ensemble of real applications. In epidemics, the introduction of a new vaccine benefit also unprotected individuals, as their probability to be infected decreases in the wake of an overall reduction in the reservoir of infection (Bridges et al., 2000; Nichol et al., 1995). In education, students that are assigned to a learning program may interfere with their untreated peers through knowledge transmission paths (Chin et al., 2013; de Heer et al., 2011). In economics, incentives targeted to firms affect also those firms which do not directly benefit of the incentive but do have an economic or juridical relationship with favored companies (Chuang & Lin, 1999; Cohen et al., 2002). In finance, a monetary shock smashing into some financial institutions may propagate over those entities involved in their transactions (Squartini et al., 2011; G.-J. Wang et al., 2017; J. Yang & Zhou, 2013). In marketing, individuals who are exposed to an advertisement may adjust their consuming behavior and influence their friends.

The causal inference literature has developed a number of statistical methods to estimate spillover effects as a whole (Aronow & Samii, 2017; Aronow et al., 2019; Forastiere et al., 2020; Loh et al., 2020; Miles et al., 2019; Papadogeorgou et al., 2019; Tortù et al., 2020), without disentangling among the specific interference mechanisms. However, investigating the treatment diffusion process differs from simply exploring spillover mechanisms as a whole. Indeed, the diffusion analysis focuses on exploiting the real treatment spreading among units, while examining spillovers as a whole requires to account for the overall indirect exposure to the intervention. When we are primarily concerned with assessing the effect of actually receiving the treatment, directly or indirectly, the main mechanism of interest is the treatment diffusion process and the estimation strategy should precisely account for treatment diffusion, and not for spillovers as a whole.

Treatment diffusion produces an alteration in the original treatment assignment mechanism of a given experiment, where units have been initially randomly assigned to treatment and control groups. When the diffusion process arises, some of the units, who have been assigned to the control group, may have yet received the active treatment, by means of

their interaction with treated neighbors. In some sense, the treatment diffusion process might give rise to a measurement error in the treatment variable. Measurement error issues have been widely discussed in the recent statistical literature (Bound et al., 2001; R. J. Carroll et al., 2006; Fuller, 2009; Grace, 2017) and some studies have also investigated the role of measurement error in the causal inference framework, by exploring the possible misclassification of the outcome variable (Grace, 2017; Shu & Yi, 2019a, 2019b) and the treatment variable (Babanezhad et al., 2010; Braun et al., 2014; Braun et al., 2016; Grandjean et al., 2004; Imai & Yamamoto, 2010; Lewbel, 2007; McCaffrey et al., 2013; Vanderweele, 2012; Yanagi, 2018). These works have all pointed out that a misclassification of the exposure might induce a bias in the estimate of the average treatment effect both when the measurement error is independent on potential outcomes, conditioning on the true value of the treatment and on observed covariates (*nondifferential* scenario), and when the misspecification of the treatment depends on potential outcomes and on individual characteristics (*differential* measurement error). The direction of this bias is difficult to be determined, as it results from various interacting mechanisms. However, Lewbel, 2007 states that if (i) misclassification is mean independent on potential outcomes, conditioning on the true treatment and the observed covariates and (ii) the sum of mis-classification probabilities is less than one (that is, the individual probability of presenting a misclassified treatment is not huge and the probability that the misclassified treatment coincides with the actual exposure is higher than the probability that a pure random guess reflects the actual exposure), the misspecified treatment effect underestimates the real treatment effect. This statement means that, when the misclassification is independent on potential outcomes and it is mostly determined by a random component, the observed effect actually under-estimates the real effect of the intervention. Vice versa, in the presence of a measurement error which depends on potential outcomes, the missclassifcation process may lead to both an overestimation and underestimation of the real treatment effect, depending on the mechanism which drives the contingency among the outcome variable and the missclassifcation process and on the

expected potential outcomes characterizing those sub-populations who have mostly experienced the diffusion process. This aspect is particularly significant in the presence of a relevant heterogeneity in the treatment effect: if the diffusion process mostly impacts those sub-populations who exhibit a particular responding behavior to the intervention, then the effects of the misclassification will reflect their specific attitude towards the treatment.

As we have already hinted, the treatment diffusion setting represents a framework, where a misclassification of the treatment variable might potentially arise, especially if the outcome is observed with a certain lag with respect to the treatment exposure. This kind of mis-classification of the treatment variable introduces a bias in the estimates. In most cases a correction for this type of bias cannot be performed because the diffusion process is usually unobserved. Note that the treatment diffusion process is sometimes the main object of interest: indeed, from the policy maker perspective, it may be useful to investigate the real diffusion spreading, using the acquired information to maximize the effectiveness of an intervention. An additional issue that may arise in this setting is that the network ties connecting treated and untreated individuals may not be completely known. In this contribution, we point out that ignoring the diffusion process, if present, can lead to overestimate or underestimate the real effect of an intervention and we propose a novel approach to deal with unknown treatment diffusion in a partially unknown network structure. Specifically, we propose a sensitivity analysis to assess the robustness of the estimates with respect to different diffusion scenarios: the analysis accounts for both the uncertainty in the network structure and the uncertainty in the real individual treatment status.

This work has been inspired by a recent experiment, which was designed to assess the effect of different kinds of school-incentives aimed at promoting museums attendance among students (Forastiere et al., 2019b; Lattarulo et al., 2017). In particular, this experiment was targeted to students living in Florence, for encouraging them to attend the museum of Palazzo Vecchio, which represents the administrative center of the city, and for pushing long-term museums attendance. In this cluster random-

ized experiment each class was randomly assigned to treatment or control. All students receive a flyer containing basic information about the museum. Students belonging to classes assigned to treatment were provided with a video presentation, where an art expert from the museum directly promotes the art exhibit. The outcome of interest is the number of self-reported museum visits in the eight months between the baseline and the follow-up survey. In this setting, the treatment diffusion might potentially be plausible because students receiving the video could have shared it with their friends. The presence of the treatment diffusion could lead to a bias in the estimation of the actual effect of the treatment. Furthermore, only intra-class ties are observable, while network links between students of different classes are not known. Here, we develop a sensitivity analysis (Rosenbaum & Rubin, 1983a) to assess the robustness of the estimated treatment effect against different diffusion scenarios. The partially unknown network is reconstructed through multiple imputation (Rubin, 1996, 2004): specifically, starting from the observation of known ties, we multiply predict the presence (or the absence) of inter-class friendship ties using random forests with chained equations (Buuren & Groothuis-Oudshoorn, 2010; Doove et al., 2014; Shah et al., 2014). The key inputs of the algorithm are a batch of dyadic similarity indicators, which measure the baseline degree of affinity between a given pair of units, with respect to hobbies, school attitudes, cultural interests and personal background. As a consequence, the sensitivity analysis procedure accounts for i) the missing inter-class network ties and for ii) the hidden diffusion process, which plausibly can cause a switch in the treatment status of the initially untreated students.

For ease of interpretation, we assume to deal with a three time step process: at the initial time, the treatment is assigned in the population; at an intermediate point in time, the treatment might diffuse over the network and, finally, at the end of the process, the individual outcome becomes observed. The three-step temporal characterization of the process implies that the treatment spreading is assumed to happen at a single point in time. Here, we also make the assumption that one's outcome is not affected by other units' treatment if not through the contagion of

the treatment itself, that is, interference is manifest only in the form of the treatment diffusion mechanism. Furthermore, we also advance the hypothesis that the treatment might spread from a treated unit to an untreated neighbor according to a "Bernoulli mechanism" with a diffusion parameter, which does not vary across dyads. This strategy allows us to detect the existence of an empirical threshold of the diffusion parameter, above which estimates significantly vary: if the threshold represents a plausible diffusion parameter, then the estimates obtained while ignoring the diffusion process are not reliable; if instead the threshold corresponds to an unrealistic empirical diffusion parameter, then ignoring the diffusion process might not constitute a relevant issue for the study and its major findings.
The paper is organized as follows. Section 4.2 presents the main aspects of the used methodology: it characterizes the theoretical framework, formalizing the treatment diffusion process, discussing the direction of the treatment diffusion bias and introducing the key steps of the applied sensitivity analysis. Section 4.3 illustrates how the sensitivity analysis can effectively reduce the treatment diffusion bias in some exemplifying simulations' scenarios. In Section 4.4, we focus on the application: we motivate its empirical relevance, we describe the dataset and we provide the main findings.

4.2 Sensitivity Analysis for Treatment Diffusion: Methodology

4.2.1 Setup and Notation

We here give a formal characterization of the treatment diffusion mechanism and its resulting contamination of the treatment arms. Let us consider a randomized experiment aimed of evaluating the effect of a given intervention on an outcome in a given population. We denote by $\mathcal{N}$ the population of interest, where the generic individual i, with $i = 1, \ldots, card(\mathcal{N}) = N$, can be randomly assigned to the active treatment or to the control group. Let $Z_{it} \in \{0, 1\}$ be the binary variable

representing the treatment assigned to unit i at time t and $Y_{it''}$ the individual outcome, observed at the lagged time t'', with $t'' > t$. The symbols $\boldsymbol{Z}_t$ and $\boldsymbol{Y}_{t''}$ denote the corresponding vectors associated to the entire population $\mathcal{N}$, while $\boldsymbol{Z}_{-it}$ denotes the vector of the treatments to all the units different from i. We assume that each unit has a non-zero probability of being initially allocated in each of the two treatment arms, that is $0 < P(Z_{it} = 1) = \pi_{it}(1) < 1$. According to the initial randomization of treated and untreated individuals, $\boldsymbol{Z}_t$, we define two sets of nodes: i) the set of treated units at time t, $\mathcal{T}_t = \mathcal{T}_t(\boldsymbol{Z}_t) = \{i \in \mathcal{N} : Z_{it} = 1\}$, with cardinality $T_t = T_t(\boldsymbol{Z}_t) = card(\mathcal{T}_t) = \sum_{i=1}^{N} Z_{it}$) and ii) the set of untreated units at time t, $\mathcal{C}_t = C_t(\boldsymbol{Z}_t) = \{i \in \mathcal{N} : Z_{it} = 0\}$, with cardinality $C_t = C_t(\boldsymbol{Z}_t) = card(\mathcal{C}_t) = N - T_t$). We possibly observe also a $N \times P$ baseline covariate matrix, $\boldsymbol{X}$, that are assumed not to change over time.

We denote by $\mathrm{G} = (\mathcal{N}, \mathcal{E})$ the graph describing the relations among units. We denote by $\boldsymbol{A} = \{a_{ij} : i, j \in \mathcal{N}\}$ the adjacency matrix associated to G, where the generic element a_{ij} signals the presence of an edge between unit i and unit j. Note that this matrix is not necessarily symmetric: that is, the relations between units could be not reciprocated, and so there could have $a_{ij} \neq a_{ji}$. Therefore, we distinguish between the set of nodes having an in-going link with i, $\mathcal{N}_i^{in} = \{j : (j, i) \in \mathcal{E}\}$ with cardinality N_i^{in}, and the set of nodes with an out-going tie to i, $\mathcal{N}_i^{out} = \{j : (i, j) \in \mathcal{E}\}$ with cardinality N_i^{out}. When the matrix $\boldsymbol{A}$ is symmetric, we have $\mathcal{N}_i^{in} = \mathcal{N}_i^{out}$ and we will simply write $\mathcal{N}_i$. Network relations are assumed fixed. Here we work in a setting where the connections in $\boldsymbol{A}$ are either fully or partially observed. For instance, in our motivating application, we have just a partial information about the network structure: we observe friendship ties between students who are enrolled in the same class, but we do not measure inter-class links. We will show one possible approach to predict the full network structure.

If the treatment spreads among individuals, then the initial treatment assignment vector $\boldsymbol{Z}_t$ does not truly represent the real allocation of units among the treatment arms. The actual treatment status of units is represented by an unknown treatment vector $\boldsymbol{Z}_{t'}$, which collects the individual treatment indicators after the diffusion process. For ease of inter-

pretation, we assume to deal with a simplified process, characterized by three time points only:

1. At the initial time t, the treatment is randomly assigned over the population, defining the treatment assignment vector $\boldsymbol{Z}_t$.

2. At time t', treatment may spread in the network: treated nodes may contaminate untreated neighbors, by sharing the intervention with them. The new treatment status is represented by the treatment variable $Z_{it'}$.

3. At time t'', the outcome $\boldsymbol{Y}_{t''}$ becomes observable.

This process is graphically represented in Figure 34. Figure 35 provides

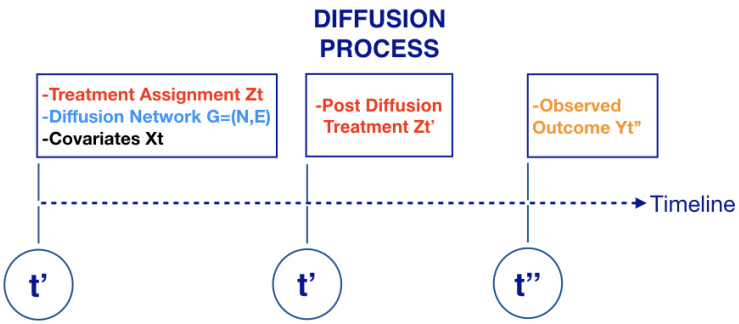


Figure 34: Three-steps Diffusion Process: timeline

a graphical example of how the diffusion process looks like. As you can see, in the example the population of interest includes ten units, linked by some edges which remain unchanged over time. At time t, a half of the population is assigned to the active treatment, while the remaining half falls into the control group. At time t', some of the initially controlled individuals receive the active treatment by diffusion, due to their interaction with treated neighbors. It is worthwhile to note that we have assumed to have a complete knowledge about the graph G. However, in many empirical scenarios this could not be the case. For instance, in our

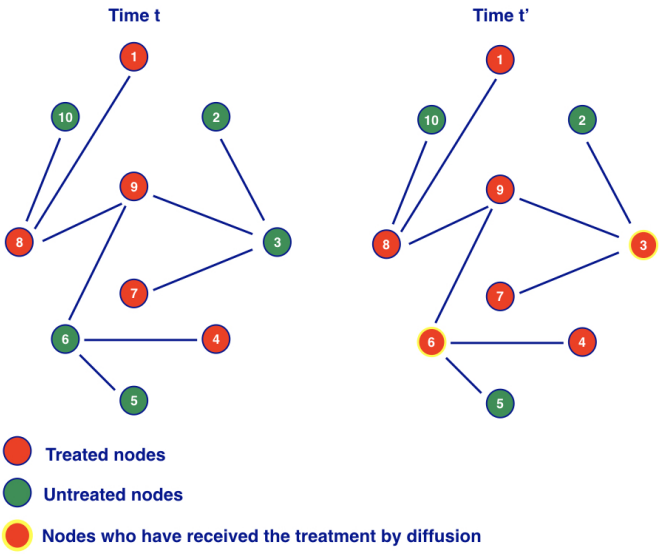


Figure 35: Treatment diffusion process: units numbered as 3 and 6 were not initially assigned to the active treatment, but they have received the treatment by diffusion .

motivating application, we have just a partial information about the network structure: we perfectly observe friendship ties between students who are enrolled in the same class, but we ignore inter-class links. We will show one possible approach to predict the full network structure and to use the reconstructed information in accounting for the diffusion process.

4.2.2 The Diffusion Process: Assumptions and Preliminaries

This study represents the first attempt to account for unobserved diffusion in causal inference. It is worthwhile to note that we firstly consider the entire network G known. Then, in Subsection 4.2.5 we will discuss the case of a partially unknown network G.

We model the diffusion process under the following simplifying as-

sumptions:

Assumption 13 (Single diffusion)**.** *For all the untreated units, diffusion can occur only at the same single time point between t and t'.*

This assumption permits not to consider multiple diffusion steps and not to account for (eventual) spreadings of the intervention occurred after the fixed time step t'. Under this assumption, it is possible to uniquely index the individual potential outcomes at time t'' in terms of the real treatment assignment vector (observed after the treatment diffusion process)

Assumption 14 (Diffusion process with a fixed probability)**.** *Given the graph G and the pre-treatment covariates $\boldsymbol{X}$, an untreated node i may receive the treatment only by a unit k in $\mathcal{N}_i^{in} = \{k \in \mathcal{N} : a_{ki} = 1\}$ with $Z_{kt} = 1$ and, each of such units, say k, can diffuse the treatment to i with probability $\overline{p}_i = \overline{p}_i(\mathbf{X}_i)$ (diffusion parameter associated to i), independently of the other units $k' \in \mathcal{N}_i^{in}$ with $Z_{k't} = 1$.*

Note that a particular case is when $\overline{p}_i = \overline{p}$ for all i and in this case we refer to $\overline{p}$ as the diffusion parameter of the model.

By introducing this assumption, we simplify the modelling of the actual treatment diffusion, relying on the Bernoulli distribution, which has been proved to suitably model a wide ensemble of real world scenarios. In this setting, is essential to advance some parametric assumptions on the treatment process. The proposed framework could effectively handle with alternative and more flexible specifications, that might be employed if data provide additional information or the researcher has a deep a priori knowledge on the diffusion phenomenon she is investigating. This assumption also states that the event that any untreated nodes switches its status due to the diffusion process occurs independently on the same event affecting any other individual in the network (included neighboring individuals). By assuming independence, we are able to model the treatment diffusion process as a result of independent events. In particular, we model the individual probability of gaining a tangible exposure to the active treatment by adopting a simplified rearrangement of the *Independent Cascade Model* (Kempe et al., 2003b; Saito et al., 2008; C. Wang et al., 2012), where contagion may occur at only one time frame by means

of a fixed probability of infection, treated units are all active spreaders of the treatment and untreated nodes who have treated neighbors are all susceptible of being infected. The possibility of a node to effectively gain the treatment by second-order neighbors is ruled out. In this setting, treatment propagates from treated to untreated friends. Nodes who have initially been assigned to the active treatment cannot receive the treatment by diffusion. In addition, an untreated node who does not have treated friends in their neighborhood cannot get the treatment. Figure 36 provides a graphical example of this diffusion process: the color of the nodes is related to their treatment status: *red* nodes are treated nodes, while *green* nodes represents units, who have been initially assigned to the control group. As you can see, units can receive the treatment by diffusion if both the following conditions are met: i) they have been originally allocated to the control group and ii) they have at least one treated friend among their direct neighbors.

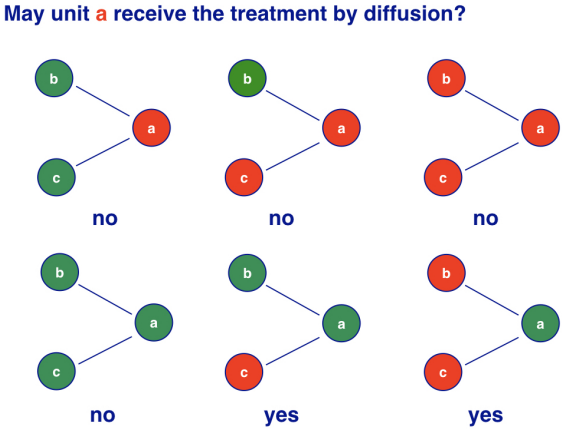


Figure 36: Diffusion Process: who can propagate the treatment to whom? The color of the nodes is related to their treatment status: *red* nodes are treated nodes, while *green* nodes represent units, who have been initially assigned to the control group. As you can see, units can receive the treatment by diffusion if both the following conditions are met: i) they have been originally allocated to the control group and ii) they have at least one treated neighbor.

Note that the simplified scenario considered here, with only three time frames and time-invariant network structure, is reasonable in most settings when the time interval between treatment assignment and follow-up is small. All the assumptions that have been introduced in this section contribute in characterizing the hidden treatment diffusion process, by simplifying the structure of the process and by imposing a plausible statistical parameterization on the actual treatment spreading. The proposed batch of assumptions simplifies the methodological scenario, and allows to partially reconstruct the hidden treatment diffusion process, about which we do not have prior information. Note that these assumptions, even if they are strong, result to be plausible in a wide ensemble of real-world phenomena and they have been conceived so to be both informative and generalizable. When prior information about a real-world diffusion process is available, these assumptions may be revisited so to be more precise and targeted to the specific empirical phenomenon: the more the researcher has valid prior information about the particular diffusion process she is analyzing the better she is able to effectively characterize this process.

Formally, given the graph G, we set $\mathcal{T}_{it} = \mathcal{T}_{it}(\boldsymbol{Z}_{-it}) = \{j : j \in \mathcal{N}_i^{in} \text{ and } Z_{jt} = 1\}$ the set of treated in-neighbors of node i at time t and we denote by $T_{it} = T_{it}(\boldsymbol{Z}_{-it})$ its cardinality, i.e. the number of treated in-neighbors of i at time t. Moreover, according to the treatment status at time t and the number of treated in-neighbors, we identify three different sets of units, that inform about the nodes' possible treatment status after the diffusion process, with respect to their eventual eligibility to experience it:

- The set of **surely treated** nodes at time t', that exactly coincides with the set $\mathcal{T}_t = \mathcal{T}_t(\boldsymbol{Z}_t)$ of units that have been randomly assigned to treatment at time t (one node treated at t cannot subsequently pass to a control status).

- The set of **surely untreated** nodes at time t'. These nodes have been assigned to control at time t and it is impossible for them to receive

treatment through diffusion at t' because of two reasons

1. node i has not in-neighbors, that is $\mathcal{N}_i^{in} = \emptyset$
2. node i has not treated nodes in his in-neighborhood, that is $\mathcal{T}_{it} = \emptyset$

We denote this set by $\mathcal{I} = \mathcal{I}(\boldsymbol{Z}_t) = \{i : Z_{it} = 0 \wedge (\mathcal{N}_i^{in} = \emptyset \vee \mathcal{T}_{it} = \emptyset)\}$.

- The set of nodes whose treatment variable at time t' is **uncertain** is given by $\mathcal{D} = \mathcal{D}(\boldsymbol{Z}_t) = \{i : Z_{it} = 0, \mathcal{T}_{it} \neq \emptyset\}$. This set of $D = D(\boldsymbol{Z}_t) = card(\mathcal{D})$ nodes identifies the units that at time t have been assigned to control but can receive the treatment at time t' through diffusion with a non-zero probability. We denote by $\mathcal{DT} = \mathcal{DT}(\boldsymbol{Z}_t, \boldsymbol{Z}_{t'})$ the subset of units belonging to $\mathcal{D}$ that will receive treatment through diffusion, i.e., $\mathcal{DT} = \mathcal{DT}(\boldsymbol{Z}_t, \boldsymbol{Z}_{t'}) = \{i \in \mathcal{D} : Z_{it'} = 1\}$, and by $\mathcal{DC} = \mathcal{DC}(\boldsymbol{Z}_t, \boldsymbol{Z}_{t'})$ the subset of units belonging to $\mathcal{D}$ that will not receive treatment, i.e., $\mathcal{DC} = \mathcal{DC}(\boldsymbol{Z}_t, \boldsymbol{Z}_{t'}) = \{i \in \mathcal{D} : Z_{it'} = 0\}$.

As we can see from this characterization, the vector of treatment assignments at time t', $\mathbf{Z}_{t'}$, is partially unknown, and we only have partial information about the units' allocation in the two treatment arms. Specifically, we are informed about the treatment status of units belonging to $\mathcal{T}_t$ and $\mathcal{I}$ (who are surely treated and surely untreated, respectively) but we do not know whether those individuals, who are eligible to gain the treatment by diffusion, have effectively received it. Nevertheless, we can compute units' overall probability of being treated at time t'. Indeed, each individual can be treated at time t' because either he was initially assigned to the active treatment or he has received the treatment by diffusion. Therefore, given the graph G, we can express the conditional probability of being treated at time t', given the treatment vector at time t in the rest of the network, i.e. $\mathbf{Z}_{-it}$, and the individual characteristics,

i.e. $\mathbf{X}_i$, by means of the law of total probabilities, that is

$$\begin{aligned}\pi_{it'}(1;\mathbf{Z}_{-it},\mathbf{X}_i) &= P(Z_{it'}=1|\mathbf{Z}_{-it},\mathbf{X}_i)\\ &= P(Z_{it'}=1|Z_{it}=1,\mathbf{Z}_{-it},\mathbf{X}_i)P(Z_{it}=1|\mathbf{Z}_{-it},\mathbf{X}_i)+\\ &\quad P(Z_{it'}=1|Z_{it}=0,\mathbf{Z}_{-it},\mathbf{X}_i)P(Z_{it}=0|\mathbf{Z}_{-it},\mathbf{X}_i)\\ &= \pi_{it}(1;\mathbf{Z}_{-it},\mathbf{X}_i)+\rho_i(1-\pi_{it}(1;\mathbf{Z}_{-it},\mathbf{X}_i)),\end{aligned}$$

where $\pi_{it}(1;\mathbf{Z}_{-it},\mathbf{X}_i)$ is the conditional probability that the unit i is initially assigned to the treatment, given the treatment vector $\mathbf{Z}_{-it}$ of the other units and the characteristics $\mathbf{X}_i$ of the unit, and $\rho_i = \rho_i(\mathbf{Z}_{-it},\mathbf{X}_i) = P(Z_{it'}=1|Z_{it}=0,\boldsymbol{Z}_{-it},\boldsymbol{X}_i)$ is the conditional probability for the unit i of receiving the treatment by diffusion given that unit i has been not assigned to the active treatment at time t and given the initial treatment assignment vector $\mathbf{Z}_{-it}$ referred to all units but i and on the vector of characteristics $\boldsymbol{X}_i$. Moreover, we can express the conditional probability of being treated at time t', given the covariate $\mathbf{X}_i$, as

$$\begin{aligned}\pi_{it'}(1;\mathbf{X}_i) = P(Z_{it'}=1|\mathbf{X}_i) &= P(Z_{it'}=1|Z_{it}=1,\mathbf{X}_i)P(Z_{it}=1|\mathbf{X}_i)+\\ &\quad P(Z_{it'}=1|Z_{it}=0,\mathbf{X}_i)P(Z_{it}=0|\mathbf{X}_i)\\ &= \pi_{it}(1;\mathbf{X}_i)+P(Z_{it'}=1|Z_{it}=0,\mathbf{X}_i)(1-\pi_{it}(1;\mathbf{X}_i))\\ &= \pi_{it}(1;\mathbf{X}_i)+\mathbb{E}[\rho_i|Z_{it}=0,\mathbf{X}_i](1-\pi_{it}(1;\mathbf{X}_i)),\end{aligned}$$

where $\pi(1;\mathbf{X}_i) = P(Z_{it}=1|\mathbf{X}_i)$ and $\mathbb{E}[\rho_i|Z_{it}=0,\mathbf{X}_i] = \sum_{\mathbf{z}_{-i}}\rho_i(\mathbf{z}_{-i},\mathbf{X}_i)P(\mathbf{Z}_{-it}=\mathbf{z}_{-i}|Z_{it}=0,\mathbf{X}_i)$. Finally, the overall individual probability of being treated at time t' is

$$\begin{aligned}\pi_{it'}(1) = P(Z_{it'}=1) &= P(Z_{it'}=1|Z_{it}=1)P(Z_{it}=1)+\\ &\quad P(Z_{it'}=1|Z_{it}=0)P(Z_{it}=0)\\ &= \pi_{it}(1)+P(Z_{it'}=1|Z_{it}=0)(1-\pi_{it}(1))\\ &= \pi_{it}(1)+\mathbb{E}[\rho_i|Z_{it}=0](1-\pi_{it}(1)),\end{aligned}$$

where $\pi_{it}(1) = P(Z_{it}=1)$ and $\mathbb{E}[\rho_i|Z_{it}=0]$ is the mean value of ρ_i given $Z_{it}=0$ and, when $\mathbf{X}_i$ is a discrete random variable, we have $\mathbb{E}[\rho_i|Z_{it}=0] = \sum_{\mathbf{x}}\sum_{\mathbf{z}_{-i}}\rho_i(\mathbf{z}_{-i},\mathbf{x})P(\mathbf{Z}_{-it}=\mathbf{z}_{-i},\mathbf{X}_i=\mathbf{x}|Z_{it}=0)$.

In our setting, by Assumption 14, we have

$$\rho_i = \rho_i(\mathbf{Z}_{-it}, \mathbf{X}_i) = P(Z_{it'} = 1 | Z_{it} = 0, \boldsymbol{Z}_{-it}, \mathbf{X}_i) = P(Z_{it'} = 1 | Z_{it} = 0, T_{it}, \mathbf{X}_i)$$
$$= 1 - (1 - \overline{p}_i)^{T_{it}} = 1 - (1 - \overline{p}_i(\mathbf{X}_i))^{T_{it}(\mathbf{Z}_{-it})}.$$

Remark:
Note that, if $\{Z_{it} : i \in \mathcal{N}\}$ and $\{\mathbf{X}_i : i \in \mathcal{N}\}$ are two independent collections of independent random variables, we have

$$\pi_{it}(1; \mathbf{Z}_{-it}, \mathbf{X}_i) = \pi_{it}(1; \mathbf{X}_i) = \pi_{it}(1)$$

and $\mathbb{E}[\rho_i | Z_{it} = 0] = \mathbb{E}[\rho_i]$.

So far we have defined the probability of receiving the treatment by diffusion, and the overall probability of being treated at time t', but we have not characterized the exact subset of units who actually experiment the diffusion process. If we denote by S_i the random variable which equals 1 if unit i switches his status by means of diffusion and 0 otherwise, we have

$$S_i = D_i I_{\{Z_{it}=0\}} \qquad \text{with } D_i \sim Bernoulli(\rho_i)\,,$$

where the random variables D_i, $i = 1, \ldots, N$, are conditional independent, given $\mathbf{Z}_t$ and $\mathbf{X}$. Note that we have $\{S_i = 1\} = \{Z_{it} = 0, D_i = 1\} = \{Z_{it} = 0, Z_{it'} = 1\}$, while $\{S_i = 0\}$ is the union of $\{Z_{it} = 1\} = \{Z_{it} = 1, Z_{it'} = 1\}$ and $\{Z_{it} = 0, D_i = 0\} = \{Z_{it} = 0, Z_{it'} = 0\}$. Summing up, after the diffusion process at time t' the total number of treated units will be the number of initially treated nodes plus the (random) number of units that received the treatment by diffusion. Therefore, if we denote by $\mathbf{S}$ the vector of the random variables S_i and by $\mathcal{T}_{t'} = \mathcal{T}_{t'}(\mathbf{Z}_t, \mathbf{Z}_{t'}) = \mathcal{T}_{t'}(\mathbf{Z}_t, \mathbf{S})$ the set representing the total random number of treated units at time t' and by $T_{t'} = T_{t'}(\mathbf{Z}_t, \mathbf{Z}_{t'}) = T_{t'}(\mathbf{Z}_t, \mathbf{S})$ its cardinality, we can state that $\mathcal{T}_{t'} = \mathcal{T}_t \cup \mathcal{DT}$ and $T_{t'} = T_t + \sum_{i=1}^{N} S_i$.

4.2.3 Potential Outcomes and Causal Effects

The treatment diffusion process might heavily compromise the causal inference setting. This unknown mechanism can introduce a relevant bias

in the estimates, leading to inaccurate conclusions about the real effect of an intervention. We here define the causal effects, under the potential outcomes framework. Following the Rubin Causal Model (RCM) (Rubin, 1974), we denote by $Y_{it''}(\mathbf{Z}_t)$ the potential outcome of unit i at time t'' under the initial treatment vector assigned at time t.

By definition, the treatment status of unit i at time t' is a function of the entire treatment vector at time t: $Z_{it'} = Z_{it'}(\mathbf{Z}_t)$. This function is stochastic and is determined by the diffusion process. In principle, we should define the potential outcome as a function of the whole treatment vector at time t and the whole potential treatment vector at time t', i.e., $Y_{it''}(\mathbf{Z}_t, \mathbf{Z}_{t'}(\mathbf{Z}_t))$. We make three important assumptions: i) the treatment status of unit i at time t has no effect on their outcome at time t'' if not through their treatment status at time t' ii) the treatment status of other units at time t has no effect on the outcome of unit i at time t'' if not through the treatment status of unit i at time t'; iii) the outcome of unit i at time t'' is not affected by the treatment status of other units at time t'. This means that there is no interference between units if not trough the diffusion process. These assumptions can be formally expressed as follows:

Assumption 15 (Exclusion Restriction and No-interference of other units' treatment at time t')**.** *Given two different assignment treatment vector* $\mathbf{Z}_t$ *and* $\mathbf{Z}'_t$*, resulting in the same treatment status at time* t' *for unit* i*, i.e.,* $Z_{it'}(\mathbf{Z}_t) = Z_{it'}(\mathbf{Z}'_t)$*, but different treatment status for some of the other units, i.e.,* $Z_{kt'}(\mathbf{Z}_t) \neq Z_{kt'}(\mathbf{Z}'_t)$ *for some* $k \neq i$*, then*

$$Y_{it''}(\mathbf{Z}_t, \mathbf{Z}_{t'}(\mathbf{Z}_t)) = Y_{it''}(\mathbf{Z}'_t, \mathbf{Z}_{t'}(\mathbf{Z}'_t)).$$

Under Assumption 15, we can index the potential outcomes by the treatment status at time t', that is $Y_{it''}(\mathbf{Z}_t, \mathbf{Z}_{t'}(\mathbf{Z}_t)) = Y_{it''}(Z_{it'})$. Hence, we postulate the existence of two potential outcomes for each unit, $Y_{it''}(Z_{it'} = 0)$ and $Y_{it''}(Z_{it'} = 1)$, representing the potential outcome that would be observed for unit i at time t'' under control and under (directly or indirectly received) active treatment, respectively. As a consequence, the

observed outcome can be expressed by

$$Y_{it''} = Y_{it''}(Z_{it'} = z) = \begin{cases} Y_{it''}(0) & \text{if} \quad z = 0, \\ Y_{it''}(1) & \text{if} \quad z = 1. \end{cases} \tag{4.1}$$

Thanks to the randomization of the initial treatment and the assumed diffusion process (Assumption 14), we can forward the following set of assumptions:

Assumption 16 (Unconfoundedness)**.**

$$Y_{it''}(Z_{it'} = z) \perp\!\!\!\perp \mathbf{Z}_t, \mathbf{X}_i \qquad \forall z \in \{0, 1\}$$

$$Y_{it''}(Z_{it'} = z) \perp\!\!\!\perp Z_{it'} | \mathbf{Z}_{-it}, Z_{it} = 0, \mathbf{X}_i \qquad \forall z \in \{0, 1\}.$$

The first sub-assumption simply reflects the randomization of the initial treatment assignment, while the second sub-assumption states that the treatment at time t' is unconfounded, given that unit i is not treated at time t, the treatment vector at time t of the other units and the vector of characteristics $\mathbf{X}_i$. The last two components drive the diffusion process: in the hypothesized process described in Section 4.2.2 the probability of receiving the treatment when initially assigned to control only depends on the covariates of the unit and the number of its treated neighbors in the network. We formally define the average treatment effect τ^* as the average comparison between the two potential outcomes:

$$\tau^* = \mathbb{E}\left[Y_{it''}(Z_{it'} = 1)\right] - \mathbb{E}\left[Y_{it''}(Z_{it'} = 0)\right] = \mathbb{E}[Y_{it''}(1)] - \mathbb{E}[Y_{it''}(0)].$$

The effect τ^* represents the causal effect of receiving the treatment. In the presence of diffusion, the actual treatment received $Z_{it'}$ does not coincide with the assigned treatment Z_{it}. In this case, τ^* is a comparison of the potential outcomes defined under the actual (but hidden) treatment status $\boldsymbol{Z}_{t'}$ and it represents the causal effect of receiving the treatment directly as assigned or indirectly through diffusion.

Under the unconfoundedness assumption 16, we have that

$$\begin{aligned}\tau^* = \mathbb{E}\left[Y_{it''}(1)|Z_{it}=1\right]P(Z_{it}=1)+ \\ \sum_{\mathbf{z}_{-it}}\sum_{\mathbf{x}}\Big(\mathbb{E}\left[Y_{it''}(1)|Z_{it'}=1, Z_{it}=0, \mathbf{Z}_{-it}=\mathbf{z}_{-it}, \mathbf{X}_i=\mathbf{x}\right]- \\ \mathbb{E}\left[Y_{it''}(0)|Z_{it'}=0, Z_{it}=0, \mathbf{Z}_{-it}=\mathbf{z}_{-it}, \mathbf{X}_i=\mathbf{x}\right]\Big) \\ P(Z_{it}=0, \mathbf{Z}_{-it}=\mathbf{z}_{-it}, \mathbf{X}_i=\mathbf{x}).\end{aligned}$$

Proof can be found in Appendix C.3. Thus, if $Z_{it'}$ were observed for all units, the causal effect τ^* would be identified from the observed data. For example, when $0 < \pi_{it'}(1; \mathbf{Z}_{-it}, \mathbf{X}_i) < 1$ (this happens, for instance, when all the random variables Z_{it}, $\mathbf{X}_i$, with $i \in \mathcal{N}$, are independent so that the term $\pi_{it}(1; \mathbf{Z}_{-it}, \mathbf{X}_i)$ in the expression for $\pi_{it'}(1; \mathbf{Z}_{-it}, \mathbf{X}_i)$ coincides with $\pi_{it}(1)$) we could use the following Horvitz-Thomson estimator:

$$\widehat{\tau}^\star = \frac{1}{N}\Big[\sum_{i=1}^{N} Z_{it'}\frac{Y_{it''}}{\pi_{it'}(1; \mathbf{Z}_{-it}, \mathbf{X}_i)} - \sum_{i=1}^{N}(1-Z_{it'})\frac{Y_{it''}}{1-\pi_{it'}(1; \mathbf{Z}_{-it}, \mathbf{X}_i)}\Big].$$

where $\pi_{it'}(1; \mathbf{Z}_{-it}, \mathbf{X}_i) = P(Z_{it'}=1|\mathbf{Z}_{-it}, \mathbf{X}_i)$ is the conditional probability for i of receiving the treatment at time t' conditional on the initial treatment vector $\mathbf{Z}_{-it}$ of the other units and the characteristics of unit i. [1] In Appendix C.3 we prove that the above estimator is unbiased under Assumption 16.

The actual diffusion is usually unobserved. Therefore, one would be tempted to neglect any diffusion mechanism, even if plausible, and estimate the treatment effect relying on the initial treatment assignment We denote by $\widehat{\tau}^b_{obs}$ a naïve estimator under the assumption of no-diffusion. This can be written as

$$\tau^b_{obs} = \mathbb{E}\left[Y_{it''}|Z_{it}=1\right] - \mathbb{E}\left[Y_{it''}|Z_{it}=0\right]. \tag{4.2}$$

As its estimator, we can use the Horvitz-Thomson estimator (Aronow & Middleton, 2013; Horvitz & Thompson, 1952) based on the initial treatment variables Z_{it} (see Subsection 4.2.6 below).

[1] In Section 4.2.6 we make use of this estimator after simulating the diffusion process.

However, if a treatment diffusion arises, τ^b_{obs} does not represent the real effect of the intervention, as the initial treatment allocation may have been altered by the treatment spreading. Hence, $\widehat{\tau}^b_{obs}$ will be a biased estimate of the treatment effect τ^*.

Furthermore, because the presence of diffusion is one of the mechanisms in which interference manifests, that is the outcome of unit i will be affected by the treatment assigned to other units, then $\widehat{\tau}^b_{obs}$ will not even estimate an intent-to-treat effect, that is, the effect of the assignment as defined in (4.2). Instead, in order to estimate the causal effect of the assignment to treatment, one would need to take into account interference and consider the treatment assigned to other individuals. In a general network setting, even if connections were measured, this could be hard, because treatment diffusion does not restrict interference to the neighboring units. The estimation of the intent-to-treat effect under treatment diffusion is beyond the scope of this paper and we will leave this issue to further works.

Our interest here is to deal with the bias with respect to the causal effect of the treatment receipt τ^*. In most scenarios, the treatment spreading is completely unobserved. However, sometimes we do have some information on the relationships between units. In fact, some studies might collect social interactions among participants, yielding a partial or a full knowledge of the social network. Alternatively, oftentimes we have geographic information on participants or a partial information on the social structure (e.g. schools and classes, social groups, ...). Full or partial knowledge of the social network G could be used to perform a sensitivity analysis for treatment diffusion. The sensitivity analysis developed here relies on full or partial information of the network to predict treatment diffusion scenarios that might have plausibly occurred. For this purpose, we must make a set of assumptions on the diffusion process.

4.2.4 Bias Analysis when Diffusion is Neglected

So far we have stated that ignoring the diffusion process, when present, introduces a bias in the estimate of the treatment effect. In this subsec-

tion, we illustrate the formula of the bias (the computations are collected in Appendix C.4) and we discuss the direction of the bias by investigating the possible settings which can generate either an underestimation or an overestimation of the causal treatment effect. If the policy maker neglected the possibility of any diffusion process playing a role in the analysis, she would estimate the quantity $\tau^b_{obs} = \mathbb{E}\left[Y_{it''}|Z_{it} = 1\right] - \mathbb{E}\left[Y_{it''}|Z_{it} = 0\right]$. Conversely, the average effect of the actual treatment is defined as $\tau^* = \mathbb{E}\left[Y_{it''}(1)\right] - \mathbb{E}\left[Y_{it''}(0)\right]$. We define the bias b as the difference between these two quantities, that is $b = \tau^b_{obs} - \tau^*$. This is the difference between the two quantities targeted under either the assumption of no-diffusion and, in turn, no-interference, or taking into account the diffused treatment. Therefore, this bias is not affected by the potential bias of the estimators used to estimate these quantities. As proven in Appendix C.4, under the first part of Assumption 16, the bias can be expressed as:

$$\begin{aligned} b &= \tau^b_{obs} - \tau^* \\ &= \mathbb{E}\left[Y_{it''}(0)\right] - \mathbb{E}\left[Y_{it''}(0)|Z_{it'} = 0, Z_{it} = 0\right](1 - \mathbb{E}[\rho_i|Z_{it} = 0]) \\ &\quad - \mathbb{E}\left[Y_{it''}(1)|Z_{it'} = 1, Z_{it} = 0\right]\mathbb{E}[\rho_i|Z_{it} = 0], \end{aligned}$$

where $\mathbb{E}[\rho_i|Z_{it} = 0] = P(Z_{it'} = 1|Z_{it} = 0)$ is the average probability of receiving the treatment by diffusion. If $b > 0$, neglecting the diffusion process implies to over-estimate the real effect of the intervention and the intervention appears to be more effective than it really is; while if $b < 0$ our analysis is affected by under-estimation and the intervention appears less effective. Clearly, in the absence of the diffusion process ($\rho_i = 0$), since we have $\mathbb{E}[\rho_i] = 0$ and $\mathbb{E}[Y_{it''}(0)|Z_{it'} = 0, Z_{it} = 0] = \mathbb{E}[Y_{it''}(0)|Z_{it} = 0] = \mathbb{E}[Y_{it''}(0)]$ (by the first part of Assumption 16), we trivially have $b = 0$. Moreover, we note that, if $\mathbb{E}\left[Y_{it''}(0)|Z_{it} = 0, Z_{it'} = 0\right] = 0$ and $\mathbb{E}\left[Y_{it''}(0)\right] = 0$, then:

- $\mathbb{E}\left[Y_{it''}(1)|Z_{it'} = 1, Z_{it} = 0\right] < 0$ leads to $b > 0$, that is an overestimation of the real effect;

- $\mathbb{E}\left[Y_{it''}(1)|Z_{it'}=1, Z_{it}=0\right] > 0$ leads to $b < 0$, that is an under-estimation of the real effect.

These are the cases that have inspired the simulations in Section 4.3 (although in the simulations Assumption 16 is not verified).

When we have a constant potential outcome under control, we have $\mathbb{E}\left[Y_{it''}(0)|Z_{it}=0, Z_{it'}=0\right] = \mathbb{E}[Y_{it''}(0)]$ and so

$$b = \tau^b_{obs} - \tau^* = \left(\mathbb{E}\left[Y_{it''}(0)\right] - \mathbb{E}\left[Y_{it''}(1)|Z_{it'}=1, Z_{it}=0\right]\right)\mathbb{E}[\rho_i|Z_{it}=0].$$

When the potential outcome under treatment is also constant, we have $\mathbb{E}\left[Y_{it''}(1)|Z_{it'}=1, Z_{it}=0\right] = \mathbb{E}[Y_{it''}(1)]$ and so the bias becomes

$$b = \left(\mathbb{E}[Y_{it''}(0)] - \mathbb{E}[Y_{it''}(1)]\right)\mathbb{E}[\rho_i|Z_{it}=0]\,.$$

Therefore the over-estimation and the under-estimation depend on the sign of the difference $(\mathbb{E}[Y_{it''}(0)] - \mathbb{E}[Y_{it''}(1)])$.

4.2.5 Dealing with a partially unknown network structure

In some situations, a full information about the network $\mathrm{G} = (\mathcal{N}, \mathcal{E})$ could not be possible. In particular, in our motivating example, we know links between students belonging to the same class, while we ignore the relationships among students in different classes. Generally, a wide variety of empirical contexts provide an incomplete network information. Link-prediction is a growing research topic within the network theory and, in a broader sense, within the statistical literature of network data. The key idea of link-prediction models is to use a snapshot of a network to predict missing links. Figure 37 provides a graphical intuition concerning the missing-links issue: given an observed network (left side network), link-prediction models use various and heterogeneous statistical techniques to impute missing links, generating a complete network (right side network). Links are predicted (purple-dotted arcs) according to a prediction model, which is determined by the specific setting.

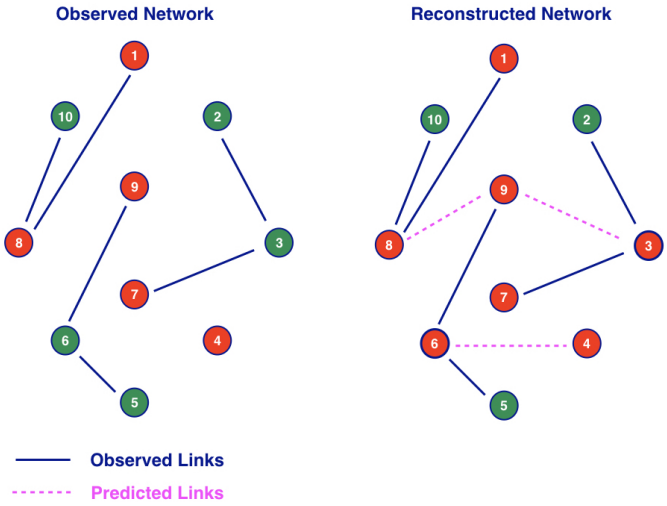


Figure 37: Observed Network vs Reconstructed Network

There are a variety of prediction models for link imputation. Prediction models can be classified into the following broad classes: i) purely statistical models, which use a parametric strategy for imputation, first estimating each link probability, then imputing missing links according to the estimated probability (Cranmer & Desmarais, 2011; Fellows & Handcock, 2012); ii) network reconstruction models, mostly referring to growing literature of statistical physics for complex networks: these approaches are highly flexible and impute missing links considering either similarity-based methods, likelihood-based criteria or entropy-based strategies (Liben-Nowell & Kleinberg, 2007; Lü & Zhou, 2011; T. Zhou et al., 2009). These latter approaches do not focus on single dyads only, but they impute missing links by taking into account the characteristics and the link patterns observed in the entire network.

We decide to perform a slightly alternative approach, which could be considered as a mixture of the two outlined strategies: we implement a multiple imputation of missing links relying on a machine learning algorithm, which predicts missing values using a more flexible approach,

based on recursive partitioning (Buuren & Groothuis-Oudshoorn, 2010). The idea is to consider the missing information about the presence (or absence) of a given tie as a missing value, and to multiple impute all that lacking values M times, so to generate M different reconstructed network. Specifically, we use *random forests* (Breiman et al., 1984) to impute missing values. Random forests derive from a theoretical extension of the classification and regression trees (CART), which are predictive models that recursively split data according to the values of the predictors. Multiple CARTs sprout a random forest: in fact, a random forest consists in an ensemble of trees (CARTs), each generated over a distinct sample of units and predictors (both selected randomly, using bootstrap). Random forests are particularly flexible, perform well in managing with possible nonlinearities or interactions, and do not require specific assumptions (Doove et al., 2014; Shah et al., 2014). Multiple imputation techniques (Rubin, 1996, 2004) are not specifically targeted for network data and link prediction issues: they are usually implemented in all those circumstances, where data include variables, which for some reasons have not been observed in some of the entries. However, they provide versatile statistical tools, which perform well in dealing with heterogeneous missing data issues.

In our empirical scenario, an highly flexible approach like the one that we have just described appears to be particularly appropriate: using a machine learning algorithm saves us from advancing strong assumptions on the imputation model, while multiple imputation allows us to efficiently handle the uncertainty over link predictions and to work with an ensemble of reconstructed scenarios. From now on, we will denote by $\mathrm{G} = (\mathcal{N}, \mathcal{E})$ the initially observed network and by $\mathrm{G}^m = (\mathcal{N}, \mathcal{E}^m)$, where $m = \{1, \ldots, M\}$, the m reconstructed network. Reconstructed networks are collected in the ensemble $\mathcal{G}$.

4.2.6 Sensitivity analysis for estimating causal effects in the presence of an unknown diffusion process: Procedure

We propose a sensitivity analysis for the unobserved treatment diffusion process, with the aim of assessing the degree of sensitivity of the naïve estimates $\widehat{\tau}^b$ of the treatment effect. The key idea is to simulate a set of diffusion scenarios and compare estimates of the treatment effect accounting for diffusion with the naïve estimates under the assumption of no-diffusion. This sensitivity analysis will allow us to assess whether ignoring treatment diffusion would lead to an over-estimate or under-estimate of the treatment effect or whether any hidden diffusion process would not have a significant impact on results. In our specific application, we must not only reconstruct the diffusion scenario, but also deal with a partially unknown network structure. It follows, that the strategy that we are going to describe accounts both for the variation in the diffusion parameters and for the managing of the distinct M reconstructed complete networks. Since we are handling a randomized setting, we make use of the Horvitz-Thompson estimator (Horvitz & Thompson, 1952) to compute the estimate of the treatment effect, and its corresponding standard error. Note that the sensitivity analysis that we propose here implicitly assumes that (i) the model for imputing missing links accurately predict hidden relationships among scholars of different classes and (ii) the hidden diffusion process satisfies the assumptions that we have advanced in Subsection 4.2.2. Relying on specific parametric assumptions is unusual in a pure sensitivity analysis. However, we call this approach a *sensitivity analysis* as it still allows to assess the robustness of results, with respect to plausible realizations of the treatment diffusion process (as pointed out by recent reviews of sensitivity analysis methodologies Hamby, 1994; Iooss and Lemaıtre, 2015, there exist sensitivity approaches which implicitly rely on specific models).

The strategy that we propose accounts for the hidden diffusion process, in the presence of a partially unknown network and consists in the following steps (note that we assume $\overline{p}_i = \overline{p}$ for all units):

1. *Naïve estimates under the assumption of no-diffusion*. Estimate the treatment effect under the assumption of no-diffusion $\widehat{\tau}^b_{obs}$ by means of the Horvitz-Thompson estimator, that is

$$\widehat{\tau}^b_{obs} = \widehat{\mu}^b_{(1)} - \widehat{\mu}^b_{(0)} = \frac{1}{N}\left[\sum_{i=1}^{N} Z_{it}\frac{Y_{it''}}{\pi_{it}(1)} - \sum_{i=1}^{N}(1-Z_{it})\frac{Y_{it''}}{1-\pi_{it}(1)}\right],$$

where $\pi_{it}(1)$ denotes the probability of unit i to be initially assigned to the treatment group (determined directly by the initial randomization plan) and compute also the corresponding estimated standard error $\widehat{\sigma}(\widehat{\tau}^b_{obs})$.

2. *Dealing with the unknown diffusion process*. Set the fixed diffusion probability $\overline{p}$, representing the probability that a treated node passes the treatment to an untreated out-neighbor. If no a-priori knowledge about this parameter is available, we will let $\overline{p}$ vary over a grid of P values, thus $\overline{p} \in \overline{\mathcal{P}} = \{\overline{p}_1, \ldots, \overline{p}_P\}$.
For each $\overline{p} \in \overline{\mathcal{P}}$, consider the M reconstructed networks and:

   (a) For each of the M reconstructed networks $\mathrm{G}^m = (\mathcal{N}, \mathcal{E}^m)$ with $m \in \{1, \ldots M\}$, and $\mathrm{G}^m \in \mathcal{G}$

   - Compute for each node the number of treated neighbors T^m_{it}.
   - Compute the elements of the N-dimensional vector $\boldsymbol{\rho}^{\overline{p},m} = (\rho^{\overline{p},m}_1, \ldots, \rho^{\overline{p},m}_N)$, where each element represents the unit-level conditional probability to switch status due to the diffusion process, according to the fixed probability $\overline{p}$, in the network m.
   - Using the vector $\boldsymbol{\rho}^{\overline{p},m}$, compute the unit level probability of being exposed to the active treatment at time t':

     $$\pi^{\overline{p},m}_{it'}(1; \mathbf{Z}_{-it}, \mathbf{X}_i) = \pi_{it}(1; \mathbf{Z}_{-it}, \mathbf{X}_i) + \rho^{\overline{p},m}_i(1 - \pi_{it}(1; \mathbf{Z}_{-it}, \mathbf{X}_i)).$$

   - The individual treatment status at time t', for those units who have been initially assigned to the control group, say

$Z^{\overline{p},m}_{1t'|Z_{it}=0}$, is obtained by sampling from a Bernoulli distribution with parameter $\rho_i^{\overline{p},m}$, that is

$$Z^{\overline{p},m}_{it'|Z_{it}=0} \sim Bernoulli(\rho_i^{\overline{p},m}).$$

Initially treated units will remain treated, that is $Z^{\overline{p},m}_{it'|Z_{it}=1} = 1$.

- Sample from $\mathbf{Z}_{t'}^{\overline{p},m}$ several times, say R, where $r = 1, \ldots, R$, and obtain a certain sample space $\mathcal{Z}^{\overline{p},m} = \{\mathbf{Z}_{t'}^{\overline{p},m,1}, \ldots, \mathbf{Z}_{t'}^{\overline{p},m,R}\}$. In other words, $\mathcal{Z}^{\overline{p},m}$ collects R different configurations of treatment assignment vectors at time t', given the fixed diffusion probability $\overline{p}$ and the network structure m. For each configuration $\mathbf{Z}_{t'}^{\overline{p},m,r} \in \mathcal{Z}^{\overline{p},m}$:
  - Compute an estimate of the treatment effect $\widehat{\tau}^{\overline{p},m,r}$ using the Horvitz-Thompson estimator, that is

$$\begin{aligned}\widehat{\tau}^{\overline{p},m,r} = \widehat{\mu}_{(1)}^{\overline{p},m,r} - \widehat{\mu}_{(0)}^{\overline{p},m,r} &= \frac{1}{N}\left[\widehat{y}_{HT,(1)}^{\overline{p},m,r} - \widehat{y}_{HT,(0)}^{\overline{p},m,r}\right] \\ &= \frac{1}{N}\Big[\sum_{i=1}^{N} Z_{it'}^{\overline{p},m,r} \frac{Y_{it''}}{\pi_{it'}^{\overline{p},m}(1;,\mathbf{Z}_{-it},\mathbf{X}_i)} - \\ &\quad \sum_{i=1}^{N}(1 - Z_{it'}^{\overline{p},m,r}) \frac{Y_{it''}}{1 - \pi_{it'}^{\overline{p},m}(1;,\mathbf{Z}_{-it},\mathbf{X}_i)}\Big].\end{aligned}$$

  - Compute the standard error of the estimate $\widehat{\sigma}(\widehat{\tau}^{\overline{p},m,r})$.
- Obtain an entire set of estimates of the overall effect of the treatment, under a specific configuration of the network m and given a fixed diffusion probability $\overline{p}$: $\Psi^{\overline{p},m} = \{\widehat{\tau}^{\overline{p},m,r} : r = 1, \ldots, R \ ; m = 1, \ldots, M\}$.

(b) The set $\Psi^{\overline{p}} = \bigcup_{m:1}^{M} \Psi^{\overline{p},m}$ contains all the treatment effect estimates, computed under a fixed probability $\overline{p}$. These estimates evaluate different topologies of the network and distinct random realizations of the unknown after-diffusion treatment assignment vector. In other terms, the values in $\Psi^{\overline{p}}$ represent an empirical distribution of the estimated effects, under diffusion

probability $\overline{p}$. Hence, we can compute the key quantities that allow us to characterize the distribution of the effects, under the diffusion parameter $\overline{p}$:

- The average value of estimated effects, under $\overline{p}$, $\overline{\widehat{\boldsymbol{\tau}}^{\overline{p}}}$, where

$$\overline{\widehat{\boldsymbol{\tau}}^{\overline{p}}} = \overline{\Psi^{\overline{p}}} = \frac{1}{R \times M} \sum_{m=1}^{M} \sum_{r=1}^{R} \widehat{\tau}^{\overline{p},m,r}.$$

- The total variance in the estimated effects, that results from the sum of two components, a within variance $s^{2^{(W)}}_{\widehat{\boldsymbol{\tau}}^{\overline{p}}}$ and a between variance $s^{2^{(B)}}_{\widehat{\boldsymbol{\tau}}^{\overline{p}}}$, that is

$$s^2_{\widehat{\boldsymbol{\tau}}^{\overline{p}}} = s^{2^{(T)}}_{\widehat{\boldsymbol{\tau}}^{\overline{p}}} = s^{2^{(B)}}_{\widehat{\boldsymbol{\tau}}^{\overline{p}}} + s^{2^{(W)}}_{\widehat{\boldsymbol{\tau}}^{\overline{p}}},$$

where

$$s^{2^{(B)}}_{\widehat{\boldsymbol{\tau}}^{\overline{p}}} = \frac{1}{R \times M - 1} \sum_{m=1}^{M} \sum_{r=1}^{R} (\widehat{\tau}^{\overline{p},m,r} - \overline{\widehat{\boldsymbol{\tau}}^{\overline{p}}})^2 \quad \text{and}$$

$$s^{2^{(W)}}_{\widehat{\boldsymbol{\tau}}^{\overline{p}}} = \frac{1}{R \times M - 1} \sum_{m=1}^{M} \sum_{r=1}^{R} (\widehat{\sigma}(\widehat{\tau}^{\overline{p},m,r}))^2.$$

The between variance $s^{2^{(B)}}_{\widehat{\boldsymbol{\tau}}^{\overline{p}}}$ captures the estimates' variance in the $M \times R$ joint realization of different (imputed) networks and different realization of the treatment assignment vector at time t'. Conversely, the within variance component $s^{2^{(W)}}_{\widehat{\boldsymbol{\tau}}^{\overline{p}}}$ averages on the estimated standard errors which have been computed within each joint realization m, r. The composite the variance is introduced to account for both the variability that comes from the sensitivity analysis and the intrinsic variability of the effect.

3. The key point of the sensitivity analysis stands in the comparison between the average value of the estimated effects under $\overline{p}$, i.e. $\overline{\widehat{\boldsymbol{\tau}}^{\overline{p}}}$, together with its corresponding total standard error, i.e. $s_{\widehat{\boldsymbol{\tau}}^{\overline{p}}}$, and the estimated treatment effect obtained while ignoring the

possibility of treatment diffusion, i.e. $\widehat{\tau}^b_{obs}$, together with its estimated standard error. For assessing the significance of the estimation bias, due to the diffusion process, one strategy could be to assume the estimated effects to be normally-distributed and to compare the confidence intervals $[\widehat{\tau}^b_{obs} - z_{\alpha/2}\widehat{\sigma}(\widehat{\tau}^b_{obs}), \widehat{\tau}^b_{obs} + z_{\alpha/2}\widehat{\sigma}(\widehat{\tau}^b_{obs})]$ and $[\overline{\widehat{\boldsymbol{\tau}}^{\overline{p}}} - z_{\alpha/2}s_{\widehat{\boldsymbol{\tau}}^{\overline{p}}}, \overline{\widehat{\boldsymbol{\tau}}^{\overline{p}}} + z_{\alpha/2}s_{\widehat{\boldsymbol{\tau}}^{\overline{p}}}]$, where $z_{\alpha/2}$ represents the critical value of the Normal distribution, associated to the significance level $(1-\alpha)$.

This procedure has to be repeated within the grid of possible values for the diffusion parameter $\overline{p}$. It is worthwhile to note that this approach can be used also to identify a "critical" threshold for the diffusion parameter: that is, testing various values of $\overline{p}$, it is possible to detect the value above which diffusion has a relevant impact on the results. Therefore, if this threshold is plausible in the specific empirical scenario, then researchers cannot really trust in the no-diffusion estimates and must also keep into account the simulated results in summarizing their findings. If instead the threshold appears to be greater than the reasonable diffusion parameters in the considered empirical framework, then no-diffusion hypothesis can be fairly assumed to be valid.

4.3 Sensitivity Analysis for Treatment Diffusion: Illustrative Simulations

In this section, we illustrate how the proposed procedure performs in some simulated scenarios. In particular, we show how the sensitivity analysis effectively succeeds in reducing the estimation bias due to the hidden diffusion process. As extensively motivated in the last section, the sensitivity analysis reconstructs an ensemble of plausible diffusion scenarios, by relying on the assumptions that characterize the diffusion process. Then, it inspects the robustness of results with respect to these reasonable realizations of the process. If the treatment diffusion process actually occurs, the researcher must front a miss-classification in the treatment variable, so that the treatment assignment vector that she ac-

tually observes does not truly represent the individual allocation in the two treatment arms. In these illustrative scenarios, we simulate a real diffusion process which introduces a bias in the estimates and leads to an over-estimation or to an under-estimation of the treatment effect. Once introduced the estimation bias, we observe how the sensitivity analysis performs in shrinking this bias and in towing the estimates towards the right treatment effect.
Here, we showcase the methodology under a simplified scenario, where the network is supposed to be entirely known. This simplification implies that the sensitivity analysis accounts for the hidden diffusion process only, and not for the network structure.

4.3.1 Data Generating Process (DGP)

In this subsection we detail the data generating process. We consider a sample made up by $N = 1000$ observations. We generate the structure of interactions by simulating an Erdős-Rényi random graph (Erdős & Rényi, 1959) $\mathcal{G}$ with N nodes and a fixed probability (0.01) to observe a link. The initial individual treatment assignment at time t, Z_{it}, and a pre-treatment characteristic X_i are sampled from independent Bernoulli distributions with probability 0.5. Hence,

$$Z_{it} \sim Ber(0.5) \quad \text{and} \quad X_i \sim Ber(0.5).$$

Thus, the individual probability to be treated at time t is $\pi_{it} = \pi_{it}(1) = 0.5$. According to the network $\mathcal{G}$ and to the initial treatment assignment vector $\boldsymbol{Z}_t$, we compute the number of treated neighbors T_{it}. Then, we randomly generate a treatment diffusion process, which we assume to be the one who truly realizes. The diffusion parameter p^* is a function of a fixed contagion parameter $\overline{p}^*$ and of a parameter δ, which introduces a statistical dependence between the diffusion parameter p^* and the variable X_i, that is

$$p_i^* = p^*(X_i) = \begin{cases} \overline{p}^* + \delta & \text{if} \quad X_i = 1; \\ \overline{p}^* & \text{if} \quad X_i = 0\,. \end{cases}$$

Once computed the real diffusion probability, we similarly compute the real conditional probability of receiving the treatment by diffusion, $\rho_i^* = \rho_i^{p_i^*} = \rho_i^{p_i^*}(\mathbf{Z}_{-it}, X_i)$, and the real conditional probability of being treated at time t', $\pi_{it'}^* = \pi_{it''}^*(1; \mathbf{Z}_{-it}, X_i)$. Formally,

$$\rho_i^{p_i^*} = \rho_i^{p_i^*}(\mathbf{Z}_{-it}, X_i) = 1-(1-p_i^*)^{T_{it}} \qquad \pi_{it'}^* = \pi_{it'}^*(1; \mathbf{Z}_{-it}, X_i) = \pi_{it}+\pi_{it}*\rho_i^*.$$

(Note that $\pi_{it}(1; \mathbf{Z}_{-it}, X_i) = \pi_{it}$ by the initial independence of all the random variables Z_{it} X_i, with $i \in \mathcal{N}$.)

The real individual treatment status at time t', for those units who have been initially assigned to the control group, say $Z_{1t'|Z_{it}=0}^{p_i^*}$, is obtained by uniquely sampling from a Bernoulli distribution with parameter ρ_i^*, that is

$$Z_{it'|Z_{it}=0}^{p_i^*} \sim Bernoulli(\rho_i^*).$$

As more extensively motivated in the last subsection, for those untreated units who have not treated individuals in their neighborhood and those units who have been treated at time t the treatment status remains unchanged (to the control and to the active status, respectively). The vector $\boldsymbol{Z}_{it'}^{p_i^*}$, that, for ease of notation, we also denote by $\boldsymbol{Z}_{it'}^*$, represents the real individual treatment status after the real treatment diffusion process. From $\boldsymbol{Z}_{it'}^*$ we can derive the dummy variable S_i^*, which signals whether the given unit i has really received the unit by diffusion ($S_i^* = 1$), or not ($S_i^* = 0$).

Once that we have generated the real treatment diffusion process, we introduce the estimation bias. In the Subsection 4.2.4, we have discussed the causal mechanisms, which may lead to overestimate or underestimate the real causal effect of the intervention. Following those considerations, we have generated the real individual effect. The variable $Y_{it''}(\mathbf{Z}_{t'}^*)$ represents the potential outcome of unit i at time t'' under the real treatment status defined at time t'. In particular, the potential outcome of the unit i under no real exposure to the intervention is sampled from a Standard Normal Distribution, $Y_{it''}(Z_{it'}^* = 0) = Y_{it''}(0) \sim \mathcal{N}(0, 1)$, while the potential outcome under an active exposure to the intervention after the diffusion process includes the individual response to the intervention,

say τ_i^*, that is, $Y_{it''}(Z_{it'}^* = 1) = Y_{it''}(1) = Y_{it''}(0) + \tau_i^*$. Summing up, the observed outcome is

$$Y_{it''} = Y_{it''}(0)(1 - Z_{it'}^*) + Y_{it''}(1)Z_{it'}^*.$$

The quantity $\tau_i^* = \tau_i^*(S_i^*, X_i)$ depends on a fixed parameter k and is a function of the switching indicator S_i^* and of the variable X_i. More precisely, given that $q_{s,x}$ denotes the proportion of individuals such that $S_i^* = s$ and $X_i = x$, we introduce an *over-estimation bias* by setting the real individual treatment effect τ_i^* as

$$\tau_i^* = \tau_i^*(s,x) = \begin{cases} -2k; & \text{if} \quad S_i^* = s = 1 \quad \& \quad X_i = x = 1 \\ -k; & \text{if} \quad S_i^* = s = 1 \quad \& \quad X_i = x = 0 \\ \frac{+2k*q_{1,1}}{q_{0,1}}; & \text{if} \quad S_i^* = s = 0 \quad \& \quad X_i = x = 1 \\ \frac{+k*q_{1,0}}{q_{0,0}}; & \text{if} \quad S_i^* = s = 0 \quad \& \quad X_i = x = 0\,. \end{cases}$$

Conversely, we introduce an *under-estimation bias* setting

$$\tau_i^* = \tau_i^*(s,x) = \begin{cases} +2k; & \text{if} \quad S_i^* = s = 1 \quad \& \quad X_i = x = 1 \\ +k; & \text{if} \quad S_i^* = s = 1 \quad \& \quad X_i = x = 0 \\ \frac{-2k*q_{1,1}}{q_{0,1}}; & \text{if} \quad S_i^* = s = 0 \quad \& \quad X_i = x = 1 \\ \frac{-k*q_{1,0}}{q_{0,0}}; & \text{if} \quad S_i^* = s = 0 \quad \& \quad X_i = x = 0\,. \end{cases}$$

Note that in both scenarios the effect is normalized, so that the effect in the whole population is equal to 0. Moreover, note that, since we have $\mathbb{E}\left[Y_{it''}(0)|Z_{it} = 0, Z_{it'} = 0, X_i = x\right] = \mathbb{E}[Y_{it''}(0)] = 0$ and $\mathbb{E}[Y_{it''}(1)] = 0$, by what said in Appendix C.4, we have:

- $b > 0$ when $\mathbb{E}\left[Y_{it''}(1)|Z_{it} = 1, X_i = x\right] = \tau_i^*(0,x) > 0$ and $\mathbb{E}\left[Y_{it''}(1)|Z_{it'} = 1, Z_{it} = 0, X_i = x\right] = \tau_i^*(1,x) < 0$, that is in the "over-estimation scenario";
- while $b < 0$ when $\mathbb{E}\left[Y_{it''}(1)|Z_{it} = 1, X_i = x\right] = \tau_i^*(0,x) < 0$ and $\mathbb{E}\left[Y_{it''}(1)|Z_{it'} = 1, Z_{it} = 0, X_i = x\right] = \tau_i^*(1,x) > 0$, that is in the "under-estimation scenario".

At this stage, we can estimate the real treatment effect in the whole population $\widehat{\tau^*}$ (note that this measure is designed to be 0 from the DGP). This estimate is primarily compared with the estimated treatment effect under the assumption of no treatment diffusion, $\widehat{\tau^b}$. The difference between these two estimates is the estimation bias b. The goal of the simulation design is to see whether the sensitivity analysis is able to reduce this bias and to produce estimates which are closer to the real value (i.e. 0).

4.3.2 Results

Here we present the main simulations' results. The simulated scenarios differ in terms of i) the size of the effect k, ii) the heterogeneity parameter δ, iii) the real fixed diffusion probability $\overline{p}^*$ and the iv) direction of the bias (overestimation or underestimation).
We start from the underestimation scenario. Here those units who have received the treatment by diffusion exhibit a positive response to the intervention. Figure 38 represents the simulations' results for the underestimation setting, where $k = 1$. The matrix of plot needs to be read as follows: from the top to the bottom the plots show the results under increasing values of the real fixed diffusion probability $\overline{p}^*$; from the left to the right plots represents increasing values of the heterogeneity parameter δ. In all the plots are depicted: i) the line corresponding to the estimated treatment effect under no diffusion (orange dotted line); ii) the line corresponding to the estimated treatment effect under the real diffusion process (blue dotted line); iii) the zero line, which signals the real treatment effect generated by the DGP (black line); iv) the box-plots representing the distribution of the estimated treatment effect obtained under the R simulated diffusion processes, which are all ruled by the fixed contagion probability $\overline{p}$ represented on the x-axis (colored box plots); v) the mean estimated treatment effect, under a fixed contagion probability $\overline{p}$ (red dots); vi) the corresponding confidence intervals of these estimates (colored triangles). As we may observe in the figure, ignoring the treatment diffusion process leads to an underestimation of the real treat-

ment effect. When the analysis accounts for the possibility of a treatment spreading, the estimation bias get reduced and the sensitivity analysis leads to estimates, which are closer to the true value 0. The performance of the proposed procedure increases as the heterogeneity shift increases, as the algorithm predicts with more accuracy the real treatment allocation of units. Note that when the real treatment diffusion probability is high and the heterogeneity parameter is small, the algorithm could even lead to a switch in the sign of the estimation bias (so, it leads to overestimate the effect of the intervention). This phenomenon occurs because the algorithm is less effective in differentiating between high values of diffusion probabilities and in predicting which units actually receive the treatment by diffusion. However, even in this scenario, the sensitivity analysis leads to estimates that are nearer to the real value of the treatment effect than the estimated effect under no diffusion. Globally, we can state that in all scenarios the sensitivity analysis, when the treatment diffusion process actually happens, performs well in reducing the estimation bias and moves the estimates closer to the real effect.

Figure 39 shows the same underestimation scenario, but with the effect k set to 1.5. The general conclusions are similar to the ones that we have advanced to the previous scenario. So, we can similarly observe that i) the estimated effect under no diffusion really underestimates the effect of the intervention and that ii) the sensitivity analysis contributes in reducing the estimation bias, moving the estimates towards 0. Moreover, in the presence of an higher overall effect, the estimation bias increases and the sensitivity analysis becomes even more accurate in catching the real treatment effect. Now, the sensitivity procedure allows to get accurate findings even under relatively small values of the fixed diffusion probability $\overline{p}$. We pass now to the overestimation scenario. In this setting, those units who have actually received the treatment through the diffusion process, have a negative treatment effect. Figure 40 shows the main results of the overestimation setting, under $k = 1$. As we notice from the figure, the estimated treatment effect under no diffusion overestimates the real effect of the intervention. The sensitivity analysis causes a downward shifting in the estimates, by moving them towards 0. As

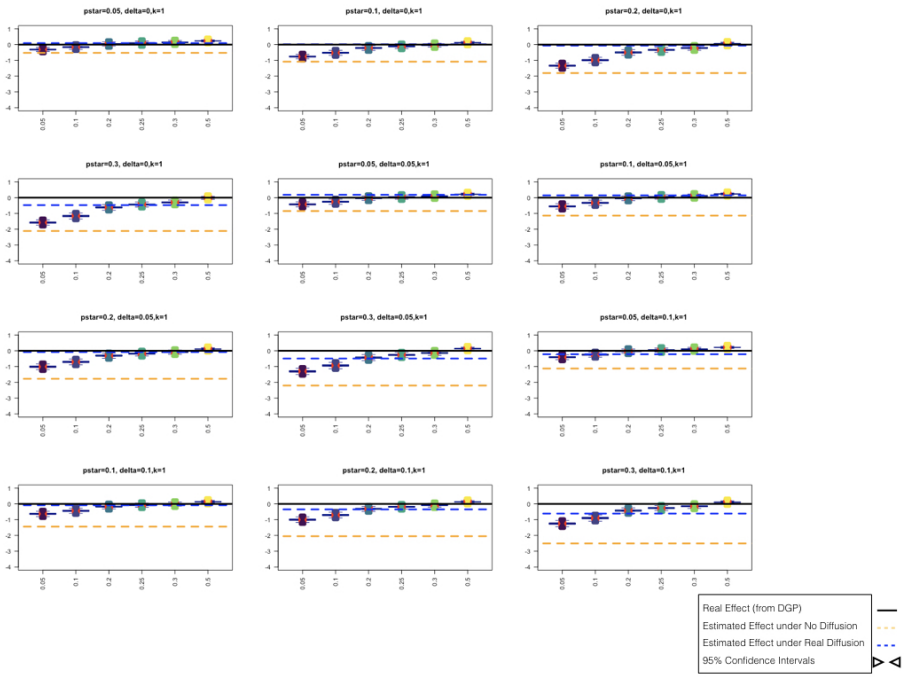


Figure 38: Underestimation, $k = 1$

before, the performance of the procedure increases as the heterogeneity parameter δ increases. The algorithm is less performing when the real diffusion probability is substantially high and the heterogeneity parameter is small. However, we can globally state that the procedure shows a good capability of towing the estimates towards the true value, by reducing the estimation bias due to having wrongly ignored the diffusion process. Finally, Figure 41 depicts the simulations' results for the overestimation scenario, where $k = 1.5$. As in the underestimation setting, an increasing in the size of the overall effect leads to an higher initial estimation bias. However, this estimation bias is effectively reduced by the sensitivity analysis, which rapidly tows the estimates towards the true value. We can definitely state that the sensitivity analysis, when

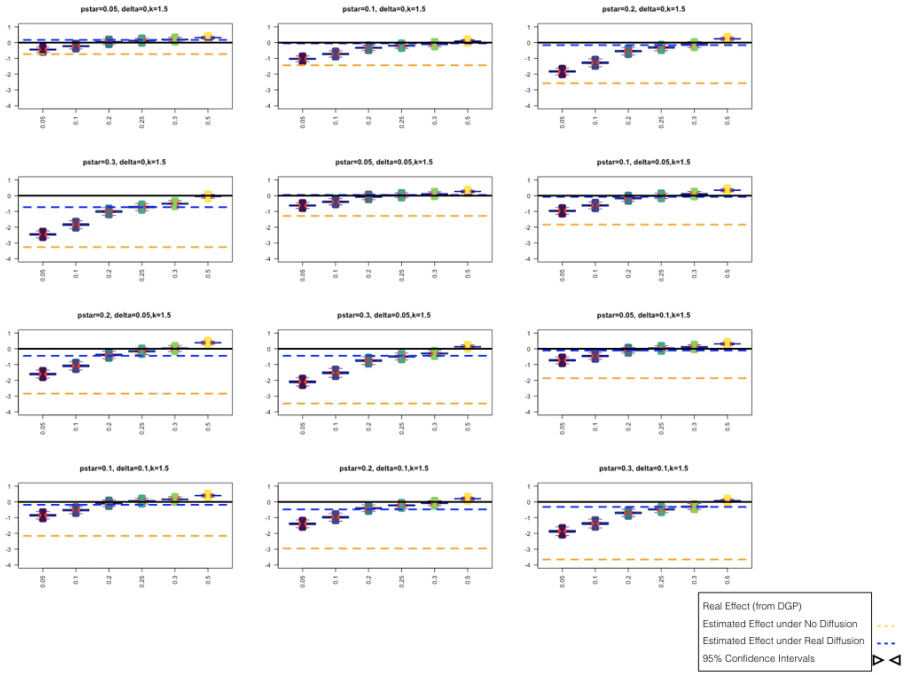


Figure 39: Underestimation, $k = 1.5$

the treatment diffusion process occurs, helps the researcher in reducing the estimation bias, by towing the estimates towards the real value. As expected, the proposed procedure performs better when some a priori information about the real process is actually known (i.e when the researcher is aware that there is some heterogeneity in the diffusion probabilities and/or in the treatment effect).

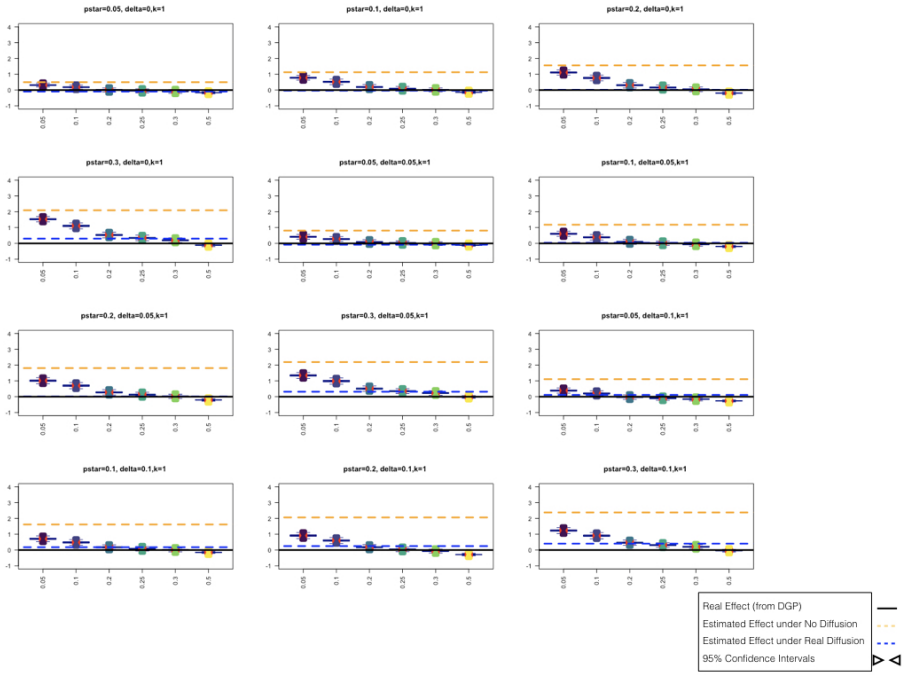


Figure 40: Overestimation, $k = 1$

4.4 Encouraging students to visit museums: issue and data

4.4.1 Empirical Motivation

The empirical application is taken from the field of Education Economics. The school system plays a relevant role in the process of the cultural capital acquisition as it can help scholars in increasing their cultural capital, encouraging them to approach the cultural and artistic heritage that surrounds them. This can happen through the active involvement and participation to theatrical performances, museum visits and art exhibitions. Although several studies (Bourdieu, 2011) have pointed out that the family represents the primary focal entity in transmitting the cultural capital

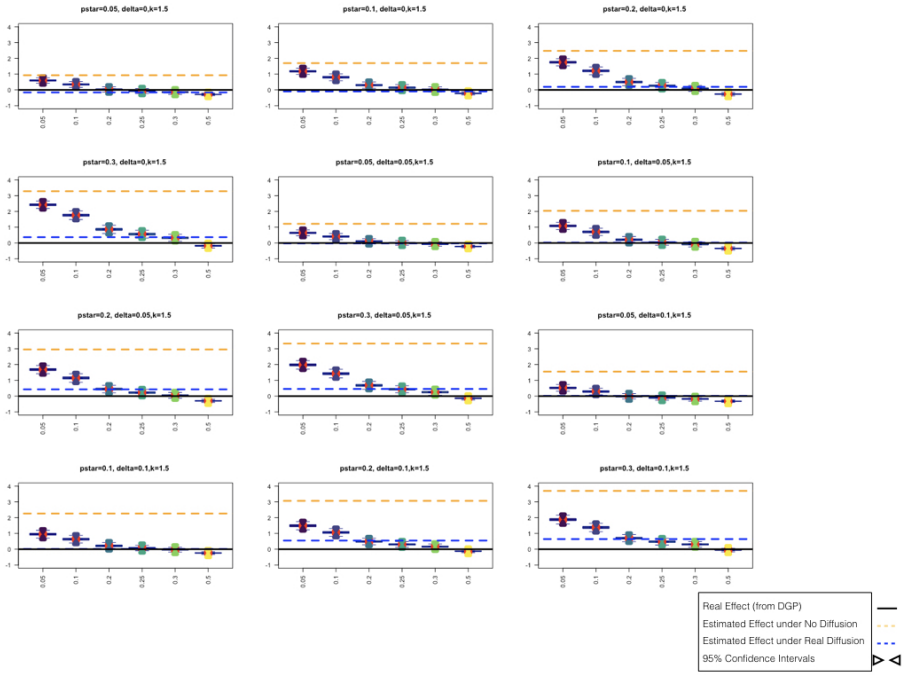


Figure 41: Overestimation, $k = 1.5$

to children, other contributions (DiMaggio, 1982; DiMaggio & Useem, 1978) have highlighted the centrality of school as an institution, which may heavily contribute to provide a cultural exposure to those children who have not adequately benefited from it in their domestic environment. Recently, Kisida et al., 2014 has underlined that cultural exposure of scholars sparks a real virtuous circle, according to which students become active cultural consumers, who are always more motivated to acquire extra cultural capital.

Although this issue is definitely relevant for the social development of the societies, there is not a so wide literature about the effects of school-promoted incentives to students (Forastiere et al., 2019b; Lattarulo et al., 2017). The randomized experiment motivating this work contributes to

filling this gap. The field experiment was a *Cluster Randomized Encouragement Designs (CEDs)* implemented in Florence, Italy in 2014, with the aim of assessing the effect of different kinds of school-promoted incentives on encouraging students to visit art museums [2]. The general goal of the experiment was to detect the most effective strategy to increase the teens' museum attendance and change their attitude towards art. In the study, classes of a school in Florence were randomly assigned to experience two different incentives: some of the classes received a flier about the importance of museum attendance, while the remaining classes received, in addition to the flier, a video presentation about an art exhibition [3]. The experiment might be potentially altered by the diffusion of the video link among students.

4.4.2 Data

Data involve $N = 176$ students, enrolled in $C = 10$ different classes, with $c = \{1, \dots, C\}$, which are in turn afferent to high schools in the city of Florence. A time t (Spring 2014) a set $\mathcal{C}_T$ of $C_T = 5$ classes were randomly assigned to the two types of cultural encouragement. Given the cluster randomized design, all the students who are enrolled in a class were then exposed to the treatment assigned to their own class. Therefore, denotiong by $C(i)$ the class membership of student i, the probability for each student i of being assigned to the video presentation, i.e., $Z_{it} = 1$, is $\pi_{it}(1) = P(Z_{it} = 1) = P(C(i) \in \mathcal{C}_T) = C_T/C = 0.5$. At time t'', 8 months after the initial assignment, students were asked to report the number of museum visits they had attended during those 8 months: this variable represents our outcome variable $Y_{it''}$.

At baseline, students were also asked to report their friendship ties.

[2] Data regarding this experiment have been collected and organized by Patrizia Lattarulo (IRPET – Tuscany's Regional Institute for Economic Planning,), Marco Mariani (IRPET – Tuscany's Regional Institute for Economic Planning) and Laura Razzolini (University of Alabama). An extensive discussion about data can be found in Lattarulo et al., 2017

[3] The original experiment by Lattarulo et al., 2017 includes a third type of encouragement: extra-credit points towards their final school grade. To simplify the interpretation of the impact of the plausible diffusion process, we decide to omit this third arm from the analysis.

Specifically, they were asked to declare who among their classmates they consider as friends, also ranking the existing ties with respect to their strength (however, here we do not consider the friendship rank, we just gauge the presence or the absence of the friendship tie). The whole network structure is described by the graph $G = (\mathcal{N}, \mathcal{E})$, which consists of C disjointed subgraphs, $\mathrm{G}_c = (\mathcal{N}_c, \mathcal{E}_c), \quad c = 1, \ldots, C$. The adjacency matrix $\boldsymbol{A}$ corresponding to the graph G, is a block-diagonal matrix with C blocks, $\boldsymbol{A}_c, \quad c : 1 \ldots C$. We have that there are no links between units belonging to different clusters, that is $A_{ij} = 0 \; \forall i, j \;$ s.t $\; C(i) \neq C(j)$. Figure 42 provides a graphical representation of the overall network structure.

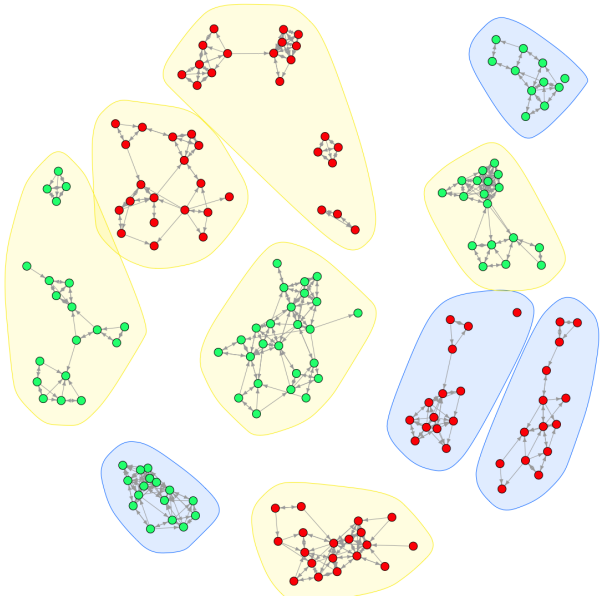
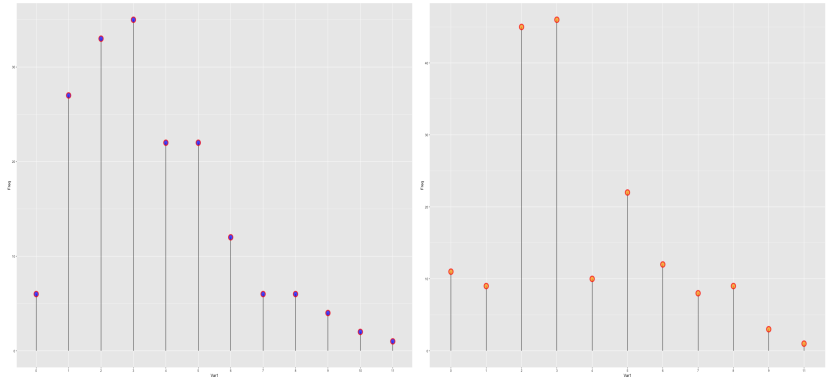


Figure 42: Observed network: nodes are colored with respect to their treatment status (red identifies treated nodes, while green untreated units; polygons encircle classes (note that students belonging to the same classes are characterized by an identical treatment status), while polygons' color represents the school membership.

In our setting, relations between students in a given class c are fully described by the adjacency matrix $\boldsymbol{A}_c$, where the generic element $A_c(i, j)$

equals 1 if the student i, enrolled in the class c, has included the student j, still enrolled in c, among his friends. Note that friendship may not be symmetric: i may regard j as a friend, but not vice versa. We denote by $\mathcal{N}_i^{out}$ the set of students that the unit i has nominated as friends, and with N_i^{out} its corresponding cardinality (the *out-degree*). Instead, we identify as $\mathcal{N}_i^{in}$ the set of students that the have included i among their friends, and with N_i^{in} its corresponding cardinality (the $in-degree$). Figure 4.43(a) shows the in-degree distribution, in the entire population, while Figure 4.43(b) displays the out-degree distribution .



(a) In-Degree distribution (b) Out-Degree distribution

Figure 43: Degree Distribution

4.4.3 Reconstructing the Unknown Diffusion Network

The experiment does not consider the possibility of inter-class links. Although they have not been explicitly reported in the survey, they are likely to be present. Inter-class links might also have been vectors for the spreading of the treatment. In fact, students belonging to different classes, but enrolled in the same school for the same year, are likely to know each other: they maybe are connected through social networks and they share similar hobbies and activities. For this reason, we intend

to reconstruct inter-class missing links through multiple imputation, as exhaustively discussed in Subsection 4.2.5. Specifically, we use random forests to multiply impute missing links, via chained equation. The imputation process takes as inputs a batch of dyadic covariates, which the algorithm employs to recursively split data, until predicting the presence (or absence) of the dyadic tie. These dyadic covariates represent four measures of similarities that have been defined starting from unit-level characteristics (the entire set of unit-level characteristics can be found in Lattarulo et al., 2017). Specifically, we include in the analysis four variables, which assess the baseline degree of similarity between two given students with respect to hobbies, school attitudes, cultural interests and personal background (details about these similarity indicators can be found in the Appendix C.1).

These similarities measures capture the key mechanisms that might prompt a friendship tie between students belonging to different classes. They are not affected by the treatment variable and they represent the key inputs of multiple imputation algorithm. In addition to the similarity measures, we have included in the multiple imputation algorithm two individual-specific indicators: they measure their number of in-class friends and the specific school environment where they are particularly inclined to establish friendship ties (mostly within their class or mostly outside their class). We generated $M = 500$ distinct imputed datasets, which correspond to 500 (complete) reconstructed networks. Note that these networks are identical with respect to the known links, but differ in the imputed ties. Figure 44 shows the densities of the tie indicators in the original (blue line) and imputed (red lines) datasets. As expected, the percentage of present links is less in the imputed datasets. This finding is in line with the general intuitive idea that it is easier to become friends for students who belong to the same class and it demonstrates that the algorithm has performed fairly in predicting the links. In fact, the similarity measures that we have given to the algorithm as inputs are intrinsically higher for pairs of students enrolled in the same class and therefore, it makes sense that imputed inter-class links are (in percentage) fewer with respect to the intra-class observed ties. Therefore, even

if observed ties and missing ties are intrinsically different, the dyadic covariates we account for allow to catch the diverse nature of links and guarantee the empirical validity of the imputation algorithm. Figures 45

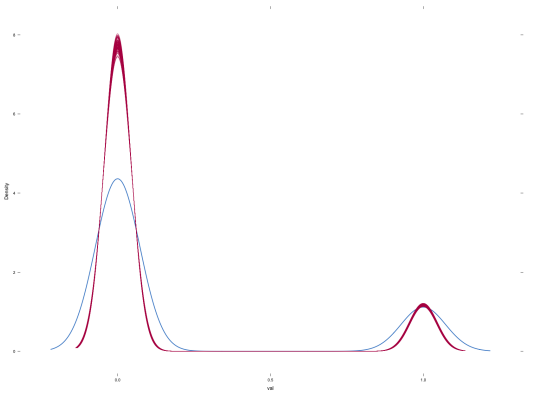


Figure 44: Densities of observed vs imputed links indicators. 1 denotes the presence of the link, 0 means the absence of a relevant tie. Blue line shows the density in the original dataset, while red lines depicts densities in the imputed links

provides a graphical example of how a (complete) reconstructed network looks like, in our setting. The plot refers to the first of the 500 generated networks. Nodes are colored according to their initial assignment status (red nodes are treated units, while green characterizes untreated units). The figure displays two kinds of links: blue links denote observed intra-class links, while violet edges depict inter-class links.

4.4.4 Sensitivity Analysis: Empirical Results

We here discuss the key empirical findings of our sensitivity analysis. The multiple imputation procedure has generated M separated reconstructed networks, which embrace the observed intra-class links as well as the predicted inter-class ties. The ensemble of the M generated networks encompasses the variation boundary of the entire network structure and it represents one of the inputs of the sensitivity analysis algo-

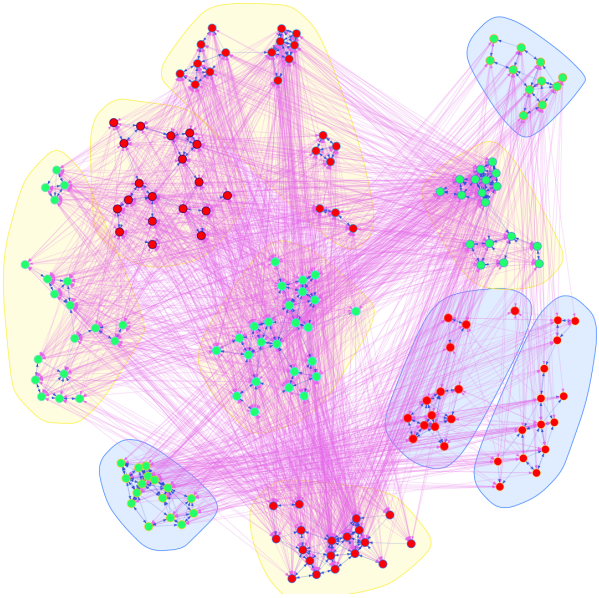


Figure 45: Reconstructed network: an example of the M generated networks

rithm we have introduced in Subsection 4.2.6. Here we set the grid of the eligible values for the diffusion parameter $\overline{p}$ to $\overline{\mathcal{P}} = \{0.01, 0.05, 0.10, 0.20, 0.25\}$. Finally, we fix the number of sampled configurations of the (unknown) treatment vector at time t' to $R = 200$.

Figure 46 gives an idea about the switching status process that happens in the presence of a plausible treatment spreading. In particular, it depicts the distributions of the probability of receiving the treatment by diffusion, for various configurations of the fixed diffusion parameter. Under small values of $\overline{p}$, very few units are eligible to gain the treatment by diffusion. As the fixed diffusion parameter increases, the number of initially untreated units who receive the treatment gets higher.

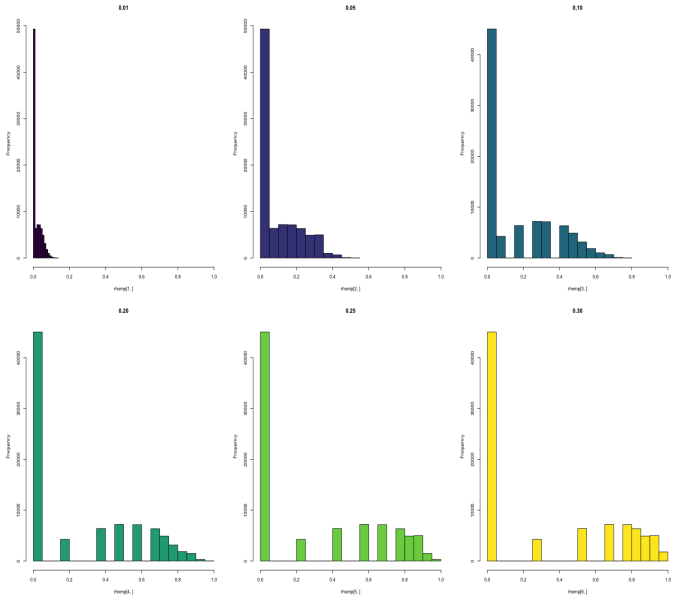


Figure 46: Probability of receiving the treatment by means of the diffusion process: histograms of the probabilities of receiving the treatment by diffusion.

Figure 47 graphically summarizes the key finding of the sensitivity analysis. The procedure accounts both for the uncertain network structure (multiply imputing the M reconstructed networks) and for unknown after-diffusion treatment vector (generating R different treatment assignment vectors at time t'). In particular, Figure 47 shows the box-plot of the treatment effect estimates, obtained under the various configurations of m and r. In addition, it shows the extremes of the 95% confidence interval, that, in case of a positive treatment diffusion probability, have been constructed so to incorporate both the between variation and the within variation (as shown in Subsection 4.2.6). Distributions refer to various possible characterizations of the diffusion parameter $\overline{p}$. Under the no-diffusion assumption, the intervention has a positive and significant impact on students' museum visits ($\widehat{\tau}^b = 2.8295$ with a 95% confidence interval which equals $\left[1.4863, 4.1727\right]$) . The graphs suggest that

ignoring the treatment diffusion process leads to an underestimation of the treatment effect. In fact, even small values of the treatment diffusion probability generate a significant upward shift in the distribution of estimated treatment effects. On the other hand, the presence of a strong treatment diffusion process heavily increases the variability of the estimates, so that even a very tiny shift in the specification of the fixed diffusion parameter causes a relevant increase in the estimated total standard errors.

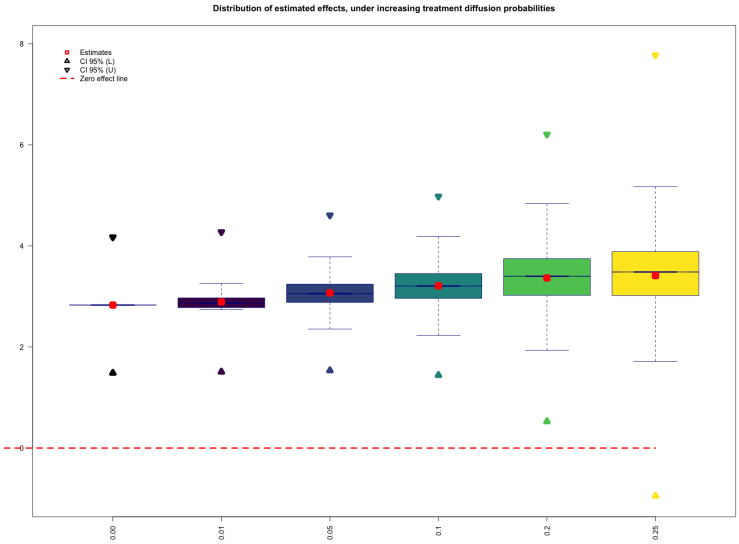


Figure 47: Box-plots of estimated treatment effects and their corresponding 95% Confidence Intervals, under increasing treatment diffusion probabilities.

According to the discussion we have in Section 4.2.4, the over-estimation issue that we face in this empirical setting may be due to a positive effect of the presentation on those students who have received it by diffusion, so that their expected outcome under the active exposure to the presentation is greater than the same expectation referred to initially treated students.

To summarize, we can state that treatment diffusion could have plausibly affected the results of the experiment that we have revisited in this work. However, as the direction of the possible bias caused by having ignored treatment diffusion is positive, the sensitivity analysis is in support of the major finding of the experiment (the video-presentation has a positive impact on students and encourages them to attend museums visits).

4.5 Concluding Remarks and Future Developments

This contribution represents the first methodological attempt of handling an unknown treatment diffusion process, in the presence of a partially unknown network structure. The proposed approach, which is based on a sensitivity analysis on the unknown after-diffusion treatment assignment vector and on the multiple predictions of missing links, is highly motivated by the specific case study we have analyzed, but it can be easily rearranged so to be suitable in a wide variety of empirical scenarios. The assumed mechanism of the treatment diffusion process appears to be reasonable for describing several real diffusion processes. Furthermore, the multiple-prediction of unobserved ties using random forests and chained equations, represents an original strategy for addressing missing links issues and it is particularly helpful when it is required to account for the uncertainty of predictions.

Our findings suggest that ignoring the treatment diffusion process, when it plausibly arises, paves the way to an inaccurate evaluation of the causal effect of interest. Specifically, in our empirical application, the effect of the intervention is larger when accounting for the treatment diffusion mechanism. Under treatment spreading, the intervention is even more effective that what it seemed from the initial estimates (obtained under the no-diffusion assumption). We actually do not have enough elements to state if diffusion really happens or not, but the threshold of the fixed contagion probability, which leads to a significant bias in the estimates, is quite small. This means that even an apparently minor dif-

fusion probability heavily alters the final results. Consequently, we are pretty confident that the real effect of the intervention is under-estimated.

Finally, the illustrated methodology might provide a lot of possible theoretical extensions, as the issues related to the hidden treatment diffusion process in program evaluation studies have not been explicitly addressed in causal inference literature yet. Future developments might involve a different temporal characterization of the treatment diffusion process, different definitions of the diffusion probabilities (for instance, they could be a function of the proportion of initially treated neighbors or be dependent on specific dyadic covariates, which in turn could involve individual and network characteristics) or an extended theoretical identification of the general estimating framework, which could be designed for accounting for spillover mechanisms. It would be also interesting to study how to account for treatment diffusion in designing experiments. The acquired information about treatment spreading in a network may be used in the experimental design, and the randomization strategy may be planned with the aim of maximizing the total number of individuals who can benefit from the intervention, either by the initial design or by diffusion.

Chapter 5

Dyadic Treatment Effect on Network Formation using Multi-valued Propensity Score Matching: Lobbying Activities and Legislative Collaborations

This Chapter is a joint work with my supervisor Prof.Laura Forastiere and with Prof. Davide Del Prete and Prof. Valerio Leone Sciabolazza, my supervisors during the three months I spent as a vising student at the University of Naples Parthenope. The full text of the article will be soon available from the arXiv repository.

5.1 Introduction

5.1.1 Motivation

Companies and firms usually strive to strengthen their political connections, at the aim of increasing their bargaining power with respect to relevant political decisions, even conditioning the approval or the reject of particular bills (Baron, 2006; Dekel et al., 2008; Denzau & Munger, 1986; Diermeier & Myerson, 1999; Groseclose & Snyder Jr, 1996; Helpman & Persson, 2001; Persson, 1998; Snyder Jr, 1991). The lobbying activity of firms towards politicians occurs through a variety of mechanisms. One possible way for firms to exert power over politicians is related to *campaign contributions*: companies manifest their political connections by choosing to officially support given candidates and to contribute in financing their electoral campaign (Bertrand et al., 2018; Powell & Grimmer, 2016; Richter & Werner, 2017; Romer & Snyder Jr, 1994; Teseo, 2020). Despite the share of lobbying companies is relatively small (95% of U.S. public companies do not even participate in campaign finance (Fouirnaies & Hall, 2018)), donations from corporations may have a relevant political impact, as they are only motivated by the desire of firms to lobby over politicians (Teseo, 2020). Indeed, by means of campaign contributions, companies intend to directly determine the political agenda of supported politicians (Besley & Coate, 2001; Dixit, 1996; Dixit et al., 1997; Grossman & Helpman, 1994; Stigler, 1971).

To more effectively fulfill their political agenda and advance their primary issues in the legislative body where they have been elected, politicians are used to establish political collaborations, with those colleagues who share similar objectives in their political agenda (Battaglini & Patacchini, 2018; Battaglini et al., 2019; Volden et al., 2020). As a consequence, if two politicians share a similar political agenda, because they are pushed by common lobbying groups, we expect them to collaborate. In other words, if campaign contributions truly play a role in shaping the legislative agenda of politicians, then we should observe that contributions have an impact also on their collaborations, by encouraging those legislators, whose political conduct is pushed by similar lobbying

interests, to team up.

This project focuses on the causal link between lobbying activities and political behaviors of legislators: in particular, it examines the effect of a pair of politicians sharing common supporting firms on their legislative collaborations. Data refer to the US House of Representatives, specifically to the 111^{th}, 112^{th} and 113^{th} Congress, and include information about Political Action Committees (PACs) and details about lawmaking collaborations, measured in terms of the number of bills that each pair of politicians has reciprocally cosponsored. The idea is that if campaign contributions can be used as a form of lobbying activity to orientate the political agenda of a legislator, and legislators use their connections to advance their political agenda, then we expect that legislators are more likely to cosponsor one another if they have been funded by the same PACs. Furthermore, we expect that, after controlling for other variables that have been identified as pushing factors for political collaborations, the lobbying activity emerges to be determining in the decision to collaborate. If these hypotheses hold, we should observe that the direction of the cosponsorship tie follows the direction of the tie of common PACs. Finally, as a matter of investigation, we examine the heterogeneity of the effects across different sub-populations of pairs of politicians, defined by specific traits: factors that reasonably prompt heterogenity of results the state of election of the two legislators, their political positioning in the legislative chamber and the pair-specific collocation in terms of party membership and ideological radicalism.

Results show that two legislators are more prone to collaborate with each other if they share a significant amount of financing supporters. Moreover, we observe that a strong link of common lobbies encourages mutual cosponsorships, rather than unilateral cosponsorships and that the cosponsorship tie follows the direction of the strong tie. These findings hold also after controlling for other factors, which are known to drive the cosponsorship activity: the common party membership, the common geographical area and the joint connection to ethnic, religious or gender minorities (Battaglini & Patacchini, 2018; Cranmer & Desmarais, 2011). Moreover, our findings suggest that the characteristics that mostly

drive heterogeneity in the estimated effects are the party membership and the state of election of the two legislators.

Lobbying activities are evaluated by monitoring PACs, following a recent stream of literature on this field (Bertrand et al., 2018; Teseo, 2020). Committees originally include not only firms and corporate groups, but also political parties, associations and organizations: given that we intend to focus on the political pressure exerted by companies, we isolate them from the rest of the PACs through a novel string-matching algorithm. Then, we remove from the analysis all those entities, which have been not successfully detected by the algorithm. Furthermore, since companies and corporate firms usually support not a unique candidate but a batch of politicians running for a seat, we account only for those links that we identify as statistically significant: for each politician, we look at the distribution of the number of lobbies that she shares with each of her colleagues and we detect those ties, which are statistically relevant from her own point of view. This approach is consistent with the idea of the *strong ties* in politics (Battaglini et al., 2019; Granoveter, 1973; Kirkland, 2011): in our setting, by applying this procedure, we are detecting strong ties of common supporting companies. Note that, according to the definition of the strong tie relationship, this relation is not necessarily symmetric: a given politician may have a relevant tie of shared financiers with one of her colleagues, but not vice versa.
It is important to point out that regarding PACs as a form of lobbying activity is still debated in the literature (Teseo, 2020). The most arguable aspect related to the usage of PACs records as measures for lobbying activities concerns the relative relevance of PACs on the entire amount of campaign contributions. Indeed, most of the donations in the US political campaign come from private citizens, parties or associations who are commonly deemed as ideological donors (Ansolabehere et al., 2003). However, this statement could be criticized according two main motivations: i) as Teseo, 2020 points out, there are many individuals (- the so-called *corporate elites* -) who contribute to the political campaign on behalf of their company, while not being moved by an ideological intent; ii) our contribution shows that the relative weight of corporate donations

is anything but negligible (about one third of the total amount of funds comes from real US companies).

Legislative collaborations are instead measured by looking at the *cosponsorship networks* (Baller, 2017; Bratton & Rouse, 2011; Fowler, 2006a, 2006b). There is an open debate in the applied literature about how to exhaustively measure the extent to which two politicians are connected each to the other. Some studies have measured the intensity of a connection among two politicians by looking at the frequency on agreements on roll call votes (Rice, 1927; Truman, 1959). However, this kind of approach allows to detect purely ideological affinities, without bringing out the presence of social relationships between legislators. Moreover, some researchers have also advanced some critiques about the usage of cosponsorship networks in explaining social and political relationships between legislators. Indeed, each politician supports a large number of bills and the decision whether to cosponsor a proposal presented by others does not imply monetary costs (Kessler & Krehbiel, 1996). However, despite the number of cosponsored bills is high, it still represents a very tiny fraction of the total amount of proposals which are presented in the US HoR during a given legislature (Fowler, 2006a). As a consequence, legislators must face a significant search cost while deciding which acts are worth to be supported. Therefore, it is reasonable to expect that the decision of a legislator to cosponsor a bill is also motivated by the fact that she shares the same political agenda of the bill's sponsor. This lends further support to the idea that cosponsorship networks are appropriate for describing relationships between politicians in our empirical setting. Note that even the cosponsorship activity is not characterized by an intrinsic reciprocity: one politician may have cosponsored some bills promoted by another but not vice versa [1].

In this setting, the outcome of interest is related to the existence of a cosponsorship tie between two members of the US House of Representatives in a given legislature. The intervention is instead related to the presence of strong ties of common supporting companies between two

[1] In the US parliamentary system, a bill is presented by one primary sponsor and it can be successively cosponsored by other officials

politicians. We intend to assess whether sharing a significant amount of common PACs impacts the likelihood of two politicians to collaborate. To face this empirical research question, we extend the standard Rubin Causal Model (RCM)(Angrist et al., 1996; Rubin, 1974, 1980) for policy evaluation studies, to deal with dyadic treatments and outcomes on network data.

5.1.2 Methodological Issues, Related Works and Contributions

The traditional Rubin Causal Model was not designed to deal with network data in that both the treatment and the outcome variable are defined at an individual level. However, in many real world scenarios units can be linked through a wide variety of relationships: for instance, individuals may be connected by means of social relationships such as friendships (Haynie, 2002; Hendrickson et al., 2011), collaborations (Newman, 2001a, 2001b) or parental ties; firms are linked by economic or legal relationships (M. O. Jackson, 2010; Schweitzer et al., 2009) such as input/output ties (Blöchl et al., 2011; Contreras & Fagiolo, 2014), ownership and control relations (Conyon, Muldoon, et al., 2008; Rungi et al., 2017). The natural tendency of describing real-world phenomena in terms of interacting units, coupled with the increasing data availability, has sparked a scientific debate about how to make the causal inference framework suitable for connected units. In recent years the causal inference literature on network data has dealt with two main issues: i) interference, where the network is the mean through which an individual treatment has an effect on the outcome of other; ii) network formation, where the network is the outcome of interest and we wish to make inference on the causal determinants of the formation of links.

The recent literature of causal inference with interconnected units has focused primarily on interference between units e.g (Aronow & Samii, 2017; Forastiere et al., 2020; Hudgens & Halloran, 2008), where the outcome of one unit is also affected by the treatment assigned to the units in their neighborhood (Cox, 1958). It has been pointed out that, neglecting

interference, when it plausibly arises, introduces a significant bias in the estimates (Forastiere et al., 2020; Sobel, 2006a) and may lead to inaccurate conclusions about a real effect of an intervention. For this reason, researchers have developed a novel framework and estimators allowing to estimate treatment and spillover effects under interference, both when the spillover mechanism occurs within groups (Barkley et al., 2017; Basse & Feller, 2018; Forastiere et al., 2019b; Forastiere et al., 2016; Hudgens & Halloran, 2008; Liu & Hudgens, 2014; Liu et al., 2016; Papadogeorgou et al., 2019; Tchetgen & VanderWeele, 2012) or when interfering interactions are described by an observed network (Aronow & Samii, 2017; Arpino & Mattei, 2013; Forastiere et al., 2020; Forastiere et al., 2018; Ogburn et al., 2017; Sofrygin & van der Laan, 2017).

On the other hand, there has been little work on causal inference for network formation. Indeed, the issue of investigating determinants of the formation of links has been one of the main research areas in the social network literature. The social network literature on network formation deals with assessing the individual and dyadic factors that might have shaped the network we observe (Battaglini et al., 2019; De Stefano & Zaccarin, 2012; Shumate & Palazzolo, 2010; P. Wang et al., 2016; Zaccarin & Rivellini, 2010). In recent years, many researchers have started exploring the causation of links by employing various econometric tools, such as dyadic regression (see (Graham, 2019) for a detailed review. However, these models are usually seen as descriptive and cannot be used to draw causal conclusions. Arpino et al., 2017 made one of the first attempts to extend the potential outcome framework to causal inference to answer questions related to network formation. They used the propensity score matching method to estimate the causal effect of the exposure of a dyad to a binary treatment on the formation of a link.

In our setting, we have dyadic data, collecting information about the presence of a strong tie of common financing companies and the presence of cosponsorships between two legislators of the US HoR. We may regard the strong tie network as a *treatment* network and the cosponsorship network as an *outcome* network. These two networks are both directed. The resulting setting consists in a *multiplex*, where the elements of the sam-

ple are connected according to two distinct network and the aim of the research is to assess whether the presence, symmetry, and direction of a tie in a treatment network has an impact in determining the likelihood of the presence, symmetry, and direction of a tie in an outcome network. There are a variety of real-world scenarios where this framework may be employed. For instance, economists may be willing to assess whether ties in a input-output network of firms are motivated by existing agreements among them; sociologists may be interested to investigate whether social ties among scientists have prompted their scientific collaborations; finally, political scientists may be yearning to assess whether the network of political alliances among countries has shaped the network of bilateral migration flows. In our setting, the treatment network is the network describing strong relationships of common supporting companies among politicians, while the outcome network is the network signaling their legislative cosponsorships. Here, we disentangle the specific mechanisms driving the effect of the treatment network on outcome network by defining conditional causal effects representing the effect on the presence, symmetry and direction of links. In particular, causal effects on the symmetry and direction of links are defined conditioning on the potential presence of a link under each level of the dyadic treatment. Building upon the literature on survival average causal effects (Comment et al., 2019; Hernán & Robins, 2010; Tchetgen Tchetgen, 2014), we advance further assumptions allowing to identify these effects. In addition, since the treatment of interest has not been randomly assigned, the relationship between the dyadic treatment and dyadic outcome might be confounded by individual or dyadic covariates. To adjust for potential confounders, we implement an estimation strategy based on *propensity score* (Hirano & Imbens, 2004; Imai & Ratkovic, 2014; Rosenbaum & Rubin, 1983b, 1984). In particular, we develop an estimator which is based on an extension of the propensity score matching approach to handle multi-valued treatments (S. Yang et al., 2016), in the presence of network data and conditional effects.

This work is organized as follows. Section 5.2 exploits the main aspects of methodology, outlining the framework of the Rubin Causal Model

in a dyadic population and discussing the structure and the properties of the multi-level matching estimator for network data. The entire framework is formalized so to be suitable for estimating both unconditional and conditional average causal effects. Section 5.3 motivates the relevance of the empirical research question, briefly describing the sources of data. Section 5.4 presents the empirical strategy to face the empirical research question. Section 5.5 presents the main results. Section 5.6 concludes the paper with a discussion of results and potential lines for future research.

5.2 Methodological Framework

5.2.1 Dyadic Set Up

In this section, we reformulate the causal inference framework, in particular the Rubin Causal Model also known as the potential outcome framework (Rubin, 1974, 1980), to handle dyadic data. In Arpino et al., 2017, who were the first to accomodate the potential outcomes framework to dyadic data, the focus was on one undirected networks, the outcome variable was defined by the presence of links and the treatment was an attribute defined at the dyadic level. On the contrary, here we focus on two distinct directed networks defining the treatment and the outcome variables, and we assess the effect of a function of treatment ties on a function of outcome ties. These functions may, for instance, define the presence, symmetry or direction of a tie. The entire data structure may be regarded as a two-layer *multiplex*, such that the examined nodes are linked by two disjoint networks. Formally, we consider a given population of units $\mathcal{N}$, where $i : 1, \ldots, N$, over which we observe a directed treatment-network $\mathrm{G}^z = (\mathcal{N}, \mathcal{E}^z)$ and a directed outcome-network $\mathrm{G}^y = (\mathcal{N}, \mathcal{E}^y)$. The two sets of edges, $\mathcal{E}^z$ and $\mathcal{E}^y$, collect links among units in $\mathcal{N}$. The data structure is represented in the multiplex $\mathrm{G} = (\mathcal{N}, \mathcal{E}^z, \mathcal{E}^y)$. In this setting, both the treatment and the outcome of interest are defined at a dyadic level. Hence, the population of interest is the dyadic population $\mathcal{D} : \{(i, j) : 1, \ldots, D\}$, which comprise all the pairs of units belonging

to $\mathcal{N}$ [2]. Let us denote with (e^z_{ij}, e^z_{ji}) the pair of edges that describe the interactions in the dyad (i, j) in the treatment network G^z: in particular, e^z_{ij} characterizes the relationship from i to j and it is strictly greater than 0 if i is connected to j in the treatment network, and it equals 0 otherwise; the edge e^z_{ji} describes instead the relationship from j to i and it is greater than 0 if the node j is linked to i in the treatment network [3]. Similarly, we label as (e^y_{ij}, e^y_{ji}) the pair of edges that give information about the interactions of the dyad (i, j) in the outcome network G^y: specifically, e^y_{ij} explains the relationship from i to j and it is strictly greater than 0 if i is linked to j in the outcome network, and it equals 0 otherwise; the edge e^y_{ji} describes instead the relationship from j to i and it is greater than 0 if the node j is linked to i in the outcome network. For each pair (i, j), it is possible to define a dyadic treatment variable and a dyadic outcome variable, which result from applying a function over their ties in the treatment network, (e^z_{ij}, e^z_{ji}) and in the outcome network (e^y_{ij}, e^y_{ji}), with $e^z_{ij}, e^z_{ji} \in \mathcal{E}^z$ and $e^y_{ij}, e^y_{ji} \in \mathcal{E}^y$. In particular, the dyadic treatment Z_{ij} results from applying $f(\cdot)$ function on the pair-specific ties, observed in the treatment network, that is, $Z_{ij} = f(e^z_{ij}, e^z_{ji})$. Similarly, the dyadic outcome results from applying $g(\cdot)$ function on the pair-specific ties observed in the outcome network $Y^{obs}_{ij} = g(e^y_{ij}, e^y_{ji})$. The functions $f(\cdot)$ and $g(\cdot)$ take as input the edges (and their attributes, if present) observed in the treatment network and in the outcome network, respectively, and can be defined in various ways, according to the causal mechanism one plans to assess. For instance, $f(\cdot)$ and $g(\cdot)$ may signal whether the relationship between i and j is symmetric ($e_{ij} > 0,\ e_{ji} > 0$), asymmetric ($e_{ij} + e_{ji} > 0\ \ , e_{ij} \times e_{ji} = 0$), or absent ($e_{ij} = e_{ji} = 0$); they may generate a discrete variable summing the weights of the two edges ($e_{ij} + e_{ji}$) or may simply discern present links (of whatever kind) ($e_{ij} + e_{ji} > 0$) from

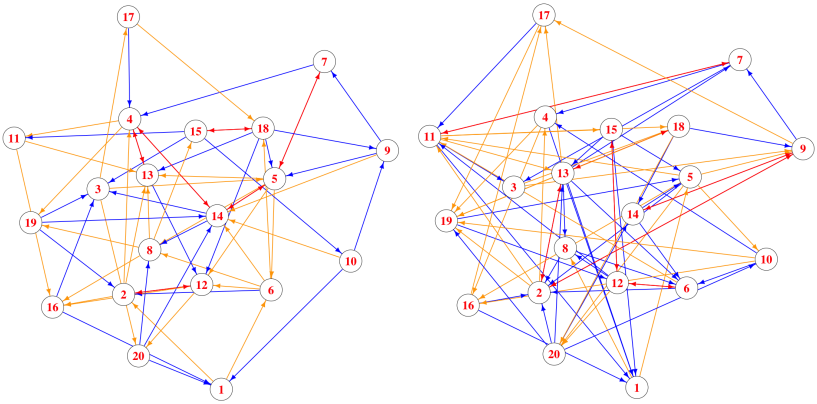
[2] Here we build up the dyadic population so that it includes unique pairs of units. Consequently, the pair is defined so that it has not an explicit direction: for instance, the dyad composed by units i and j, (i, j), with both i and j belonging to $\mathcal{N}$, is conceptually identical to the dyad (j, i). However, different empirical settings may require a direct characterization of the dyadic population, that would not imply any relevant theoretical modification in the methodological framework that we are going to propose.

[3] In case of binary networks the edges e^z_{ij} and e^z_{ji} equal 1 if the relationship is present, 0 otherwise.

absent relationships ($e_{ij} = e_{ji} = 0$). Independently on the formulation of these functions, we can state that they map ties in the treatment network and in the outcome network so to uniquely define a pair-specific treatment variable and outcome variable. These functions take values over the sets $\mathcal{Z}$ and $\mathcal{Y}$, which represent the domain of the treatment variable and the outcome variable, respectively. Here, we allow $\mathcal{Z}$ for the possibility to be not necessarily binary, and to embrace K distinct elements. For each pair (i, j), we also observe a P-dimensional vector of baseline dyadic covariates, $\boldsymbol{X}_{ij}$, such that $\boldsymbol{X}$ denotes the corresponding $(D \times P)$ covariate matrix. Note that the $\boldsymbol{X}$ matrix can include both purely dyadic covariates and features resulting from the joint distribution of individual characteristics, which separately describe the two individuals belonging to the pair.

Figure 48 suggests the idea of the setting we are handling in this work, with a toy example. We deal with a given sample composed by 20 elements, who are connected according to two disjoint networks: the treatment network and the outcome network. The elements of the sample are the nodes in the two graphs and they are labelled according to a numeric ID going from 1 to 20. Edges in the two networks can be either symmetric, asymmetric or absent. Note that the direction of asymmetric interactions matters here. In this example, we interpret the edges' direction according to the nodes' IDs: "ordinary" links are connections where the vertex ID of the node who delivers the tie is lower than the vertex ID of the node who receives the tie; "opposite" ties are the ones where the vertex ID of the node who delivers the tie is greater than the vertex ID of the node who receives the tie. The dyadic treatment is defined according to ties in the treatment network, while the dyadic outcomes is defined by considering connections in the outcome network.

Following the Rubin Causal Model (RCM) (Rubin, 1974), we postulate the existence of *dyadic potential outcomes*. For each dyad (i, j), with $(i, j) \in \mathcal{D}$, we theorize the existence of Z potential outcomes, where K denotes the number of the treatment levels. Given a dyadic multi-valued treatment $Z_i j \in \mathcal{Z}$, with $|\mathcal{Z}| = K$, the potential outcome of the pair (i, j) under a given treatment level z, with $z : 1, \ldots, K$, $Y_{ij}(z)$, represents the



(a) Treatment-network (b) Outcome-network

Figure 48: Causal inference for network formation over a directed multiplex: example. Nodes represent units in the population and are characterized by a numeric ID. Edges are colored according to the type of interaction they describe: *orange* edges represent asymmetric ties going in the "ordinary" direction; *blue* edges represent asymmetric ties going in the "opposite" direction;*red* denotes mutual ties; missing ties are not represented in the figure.

outcome that would have been observed, if the dyad (i, j) had been exposed to the treatment level z. The vector $\boldsymbol{Y}(z)$ collects the dyadic potential outcomes, under a dyadic treatment set to z.

The standard causal inference framework is based on a key assumption, known as *Stable Unit Treatment Value Assumption (SUTVA)*, which allows to deal with well-defined potential outcomes and, consequently, to properly estimate the effect of an intervention. SUTVA is composed of two components that rule out the possibility of having multiple versions of the treatment and the presence of interference among units. Here, the SUTVA requires to be discussed with respect to the dyadic population, leading to the *Dyadic Stable Unit Treatment Value Assumption*. This assumption has still two components (i) Individualistic Treatment Response (ITR) (*No interference*): there is no interference among dyads, that is, each pair's dyadic potential outcomes are not affected by the treat-

ments received by the other dyads; (ii) *No hidden versions of treatment*: there are no different versions of the treatment levels assigned to each dyad, which may lead to different potential outcomes. In a network setting, our proposed dyadic formulation of SUTVA implies that the dyadic treatment is well-specified and that there are no hidden versions of the intervention. Furthermore, dyadic SUTVA entails that the treatment, assigned to a given pair, does not affect the outcome of other dyads.

5.2.2 Average Dyadic Treatment Effects

We can define the *dyadic average treatment effects* as comparisons of dyadic average potential outcomes, that outline the mean of the outcomes for each given configuration of the dyadic treatment. Here, we formalize the dyadic average treatment effects so that the proposed mathematical definition may be suitable to model both conditional and unconditional causal effects. Conditional causal effects are employed when one intends to compare potential outcomes, while conditioning upon particular realizations of the treatment and the outcome: in such a setting, one must consider that for some units a given potential outcome may be undefined. This idea reflects the approach that is common in the literature on survival average causal effects (Comment et al., 2019; Hernán & Robins, 2010; Tchetgen Tchetgen, 2014): in examining the effect of an intervention in longitudinal studies, the researcher must keep into account that individual potential outcomes at follow-up is undefined for those people who die before attending the follow-up visit. This phenomenon is known as *truncation by death*. The treatment effects estimated while ignoring the truncation by death issue might be biased (Tchetgen Tchetgen, 2014). A similar concern arises in cross-sectional studies, when the causal effects are estimated while conditioning on specific values of the actual treatment and the observed outcome. To account for that issue in the analysis, we define the causal effects of interest only for those dyads, over which both potential outcomes that are compared for outlining the given effect exist and are well-defined. Formally, we introduce two conditioning mapping functions, to identify those dyads with respect to that

a given potential outcome exists. Let the variable $Z^*_{ij} = f^*(e^z_{ij}, e^z_{ji})$ result from applying a conditioning mapping function $f^*()$ on the treatment values, such that for those dyads which satisfy a given conditioning on their treatment values, (e^z_{ij}, e^z_{ji}), we have that $Z^*_{ij} = 1$. The function $g^*()$ maps instead a certain conditioning on the outcome values (e^y_{ij}, e^y_{ji}), that is $Y^*_{ij} = g^*(e^y_{ij}, e^y_{ji})$: the $g^*()$ function is defined such that if the resulting variable equals 0, then the outcome variable $Y_{ij} = g(e^y_{ij}, e^y_{ji})$ is undefined. We can outline the *average dyadic treatment effect* of the treatment z with respect to the treatment z', while focusing only on those dyads whose potential outcomes under both treatment levels, z and z', exist: this stratum of units is called *always-defined principal stratum*. The effect of interest, that we call $\tau_{(z,z')}$, is formally defined as

$$\begin{aligned}\tau_{(z,z')} =& E\left[Y_{ij}(z)|Z^*_{ij} = 1, Y^*_{ij}(z) = Y^*_{ij}(z') = 1\right] - \\ & E\left[Y_{ij}(z')|Z^*_{ij} = 1, Y^*_{ij}(z) = Y^*_{ij}(z') = 1\right]\end{aligned}$$

Potential outcomes cannot be jointly observed, because each dyad can be exposed to one treatment level only. It follows, that the only outcome that we can observe for the pair (i, j) is the one which corresponds to the treatment level it is actually exposed to: $Y^{obs}_{ij} = Y_{ij}(Z_{ij})$. The remaining outcomes, which are related to a treatment level that the given pair has not experienced, are not observable. Therefore, the statistical challenge here consists in detecting the best strategy to impute dyadic missing potential outcomes across treatment arms, then comparing them to extract causal information, about the effect of the dyadic treatment on network formation. Since we move in an observational setting, we propose to estimate the dyadic average treatment effects by implementing a propensity score matching across dyads, accounting for dyadic baseline characteristics.

5.2.3 Multi-level Propensity Score Matching with Network Data

Propensity score matching is a well-grounded statistical approach in policy evaluation studies. However, it is usually implemented with in-

dividual data and with binary treatments, while we intend to let the dyadic treatment the possibility to vary over a multi-valued domain. A novel strategy for dealing with multi-valued treatments in propensity score matching has been recently proposed by S. Yang et al., 2016. We in turn expand their innovative framework so to handle network data. As we move in an observational setting, we must account for baseline (dyadic) covariates. Instead of conditioning on the entire batch of available covariates, researchers sometimes prefer to work with the so called propensity score, which represents a numerical synthesis of pretreatment covariates (Rosenbaum & Rubin, 1983b). In its standard formulation, propensity score measures the individual probability of being exposed to the active treatment, given baseline characteristics. This quantity embodies all the relevant information expressed by covariates and it can be employed for covariate adjustment using different appraoches, such as sub-classification and matching (Howard et al., 2000). In particular, propensity score matching represents one of the most widespread statistical techniques for extracting causal information in observational studies (Abadie & Imbens, 2016; Dehejia & Wahba, 2002; Rosenbaum & Rubin, 1983b). The idea of any matching algorithm is to compare units who exhibit similar values of propensity score, but differ in the actual treatment status. Here we build upon the stardard approach to develop a propensity score-based estimator for dyadic average causal effects, possibly conditional on the always-defined principal stratum. Given the possibly multi-valued nature of the dyadic treatments we rely on the generalized propensity scores (G. W. Imbens, 2000).

Definition 6 (Generalized Propensity Score (GPS) for network data)**.** *Generalized propensity score is defined as the dyadic conditional probability of receiving each treatment level, conditioning on the pre-treatment dyadic covariates matrix and on not to have unspecified potential outcomes for that given treatment level*

$$\pi_{ij}(z|\boldsymbol{x}, z^*, y^*) = P(Z_{ij} = z|\boldsymbol{X}_{ij} = \boldsymbol{x}, Z^*_{ij} = 1, Y^{obs,*}_{ij} = 1)$$

In order to properly define and estimate the effects of interest, we must assume that each dyad has a nonzero probability of being exposed

to each of the treatment categories. Therefore, we advance the following Assumption,

Assumption 17 (Positivity). $\pi_{ij}(z) > 0 \quad \forall (ij) \in \mathcal{D},$ *and* $\forall z \in \mathcal{Z}$

In addition, the necessary condition for GPS to be used for the estimation of dyadic treatment effects is that the conditional probability for a dyad of being exposed to each of the treatment level is strictly positive, for all the pair-specific realization of dyadic covariates. This means, that Assumption 18 must hold.

Assumption 18 (Overlap)**.** *For all values of* $\boldsymbol{x}$*, the conditional probability of receiving any level of the treatment is strictly positive*

$$\pi_{ij}(z|\boldsymbol{x}, z^*, y^*) > 0 \;\; \forall z, \;\; \forall \boldsymbol{x} \;\; and \;\; \forall (i,j) \in \mathcal{D}$$

The unconfoundedness assumption, readjusted to the network scenario, would imply that the dyadic treatment and potential outcomes are independent, conditioning on (dyadic) baseline characteristics. However, this assumption presents some peculiar implications in handling multi-valued treatments: in fact, in the presence of a treatment defined over multiple categories, the standard unconfoundedness assumption would require the conditioning over a $Z-1$-dimensioned batch of propensity scores. This aspect highly limits the benefits of the propensity score approach in terms of dimensionality reduction, which is really the primary aspect that makes propensity score methods preferable, compared to the strategy of managing the entire set of covariates. To achieve a dimensionality reduction of the analysis, S. Yang et al., 2016 suggests to rely on a weaker version of unconfoundedness, which is formally presented in the following assumption. Here we extend the weak unconfoundess assumption to handle conditional causal effects as the ones we have presented in the previous subsection.

Assumption 19 (Weak Unconfoundedness for dyadic treatments and conditional causal effects)**.** *The assignment mechanism is weakly unconfounded if for all* $z \in \mathcal{Z}$ *and for all* $(i,j) \in \mathcal{D}$ *we have that*

$$\left[I_{ij}(z) \perp\!\!\!\perp Y_{ij}(z) | \boldsymbol{X}_{ij}, Z^* = 1, Y^*_{ij}(z) = Y^*_{ij}(z') = 1\right]$$

where

$$I_{ij}(z)\begin{cases}1 \;\; \textit{if} \;\; Z_{ij}=z,\\ 0 \;\; \textit{if} \;\; Z_{ij}\neq z,\\ \textit{unspecified} \;\; \textit{otherwise}\end{cases}$$

This assumption works with dyadic treatment indicators $I_{ij}(z)$, which are dummy variables signaling the dyadic exposure to a given treatment status z. If the assignment mechanism is weakly unconfounded conditioning on baseline covariates, then it remains unconfounded if we condition on the generalized propensity score evaluate at z. This means that under Assumption 19 the following relation holds, for all $z \in \mathcal{Z}$ and for all dyads $(i,j) \in \mathcal{D}$ for which potential outcomes are always defined.

$$I_{ij}(z) \perp\!\!\!\perp Y_{ij}(Z_{ij})|\pi_{ij}(z|\boldsymbol{x},z^*,y^*), Z^*_{ij}=1, Y^*_{ij}(z)=Y^*_{ij}(z')=1$$

Because our average dyadic treatment effects are defined conditioning on the principal stratum where the outcome is always defined, we make two additional assumptions allowing to identify the causal effects from the observed data.

Assumption 20 (Further assumptions for Conditional Average Causal Effects)**.** *For each pair $i,j \in \mathcal{D}$, the two following assumptions hold,*
Conditional independence of $Y_{ij}(z)$ and $Y^*_{ij}(z')$, given $Y^*_{ij}(z)$ and the propensity score.

$$Y_{ij}(z) \perp\!\!\!\perp Y^*_{ij}(z')|Y^*_{ij}(z), \pi_{ij}(z|\boldsymbol{x},z^*,y^*); \textit{with } z'\neq z$$

ii) Conditional independence of $Y^*_{ij}(z)$ and $Y^*_{ij}(z')$, given propensity score.

$$Y^*_{ij}(z) \perp\!\!\!\perp Y^*_{ij}(z')|\pi_{ij}(z|\boldsymbol{x},z^*,y^*), \textit{with } z'\neq z$$

Under Assumptions 17,18,19 and 20, the Dyadic Average treatment Effect $\tau_{(z,z')}$ is identified as:

$$\begin{aligned}&\mathbb{E}\big[Y_{ij}(z')-Y_{ij}(z)|Z^*_{ij}=1, Y^*_{ij}(z)=Y^*_{ij}(z')=1\big]=\\ &\mathbb{E}\big[\mathbb{E}[Y^{obs}_{ij}|Z_{ij}=z', \pi_{ij}(z'|\boldsymbol{x},z^*,y^*), Z^*_{ij}=1, Y^{obs,*}_{ij}=1)]\big]-\\ &\mathbb{E}\big[\mathbb{E}[Y^{obs}_{ij}|Z_{ij}=z, \pi_{ij}(z|\boldsymbol{x},z^*,y^*), Z^*_{ij}=1, Y^{obs,*}_{ij}=1)]\big]\end{aligned}$$

Under Assumption 19 we can match dyads with similar generalized propensity score and estimate the average causal effects comparing the

observed outcomes of the matched pair of dyads. In particular, here we perform a nearest-neighbor matching algorithm such that the matched sets are obtained by solving the following minimization problem

$$m_{gps}(z, \pi_{ij}(z|\boldsymbol{x}, z^*_{ij} = 1, y^*_{ij}), Z^*_{ij} = 1, Y^{obs,*}_{ij} = 1) = \underset{(kl)\in\mathcal{D}: Z_{kl}=z, Z^*_{kl}=1, Y^{obs,*}_{kl}=1}{\arg\min} \parallel \pi_{kl}(z|\boldsymbol{X}_{kl} = \boldsymbol{x}, Z^*_{kl} = 1, Y^{obs,*}_{kl} = 1) - \pi_{ij}(z|\boldsymbol{X}_{ij} = \boldsymbol{x}, Z^*_{ij} = 1, Y^{obs,*}_{ij} = 1) \parallel$$

The output of this function is the pair (k, l) (or multiple pairs in case of multiple matches), who best satisfies the minimization criterion. It follows, that the imputed potential outcome for the dyad (i, j) with respect to the dyadic treatment status z corresponds to the observed outcome of their nearest match, conditioning on the existence of the potential outcome under the treatment level z, for both (i, j) and its match [4]. This means that an estimator of the average treatment effect of the dyadic treatment z with respect to the treatment z' can be expressed as follows:

$$\widehat{\tau}^{gps}(z, z') = \frac{1}{D} \sum_{(i,j): Z^*_{ij}=1, Y^{obs,*}_{ij}=1} \Big[Y^{obs}_{m_{gps}(z, \pi_{ij}(z|\boldsymbol{x}, z^*_{ij}=1, y^*_{ij}))} - Y^{obs}_{m_{gps}(m_{gps}(z', \pi_{ij}(z|\boldsymbol{x}, z^*_{ij}=1, y^*_{ij})))} \Big]$$

The estimated variance of $\widehat{\tau}^{gps}(z, z')$ is derived in S. Yang et al., 2016 and finds its roots in Abadie and Imbens, 2006.

[4] Note that the imputed potential outcome which corresponds to the treatment status the dyad is actually exposed to coincides with its own observed potential outcome. Note also that the number of selected matches M can be greater than 1: this means that for each dyad we can identify more than one matching pair, and, as a consequence, in the case of $M > 1$, the imputed potential outcome for a given dyad is obtained by computing the mean of the outcomes observed among detected matches. For estimating the effects of interest, we need to compare dyadic potential outcomes.

5.3 Political Lobbying and Collaborations: Issue and Data

5.3.1 Background

Companies and firms intend to influence the political agenda of elected politicians. From a broader point of view, they intend to monitor, stimulate, amend or prevent pubic policies, that are pertinent to their economic exercise. The process of organizations and enterprises exerting power on public servants is known as *lobbying*. The National Institute for Lobbying and Ethics defined lobbying as follows (for Lobbying & Ethics, n.d.): *"Lobbying is a legitimate and necessary part of our democratic political process. Government decisions affect both people and organizations, and information must be provided in order to produce informed decisions. Public officials cannot make fair and informed decisions without considering information from a broad range of interested parties"*. Therefore, according to the definition of the NILE, lobbying is a constituent mechanism of a democratic political competition. However, the presence of interactions between business interests and public welfare introduces clear or masked distortions in the legislative behavior of legislators: the lobbying pressure induces politicians to work on issues which are believed to be considerable and urgent by the lobbies, while overshadowing and postponing those acts that are inspired by ideological and political priorities. In some cases, the lobbying conduct may be also not easy to be detected: for instance, if lobbies intend to block a particular proposal referred to a given issue, an external observer is not able to perceive any political pressure, since all that she/he is able to judge is the maintaining of a legislative status quo (Drutman, 2011); moreover, sometimes business groups demand for clauses pertained to very specific matters: therefore, the outcomes of the lobbying behavior can be concealed into massive and complex legislative acts (Bertrand et al., 2018).

There are relevant works examining how the interaction between government and business groups affects the national political attitude towards key issues (Marsh & Lewis, 2014; Paster, 2018; Vogel, 2003), while

other project focus on looking at the causal link between this interaction and politicians' individual behavior (see Battaglini and Patacchini, 2018 for a recent review). Influence occurs by means of heterogeneous mechanisms, such as informative agreements, employment opportunities or campaign contributions (Bombardini & Trebbi, 2019). Campaign contributions have recently been identified as powerful channels for firms to apply a political pressures on candidates, who compete for a seat in a legislative body (Powell & Grimmer, 2016; Richter & Werner, 2017; Teseo, 2020). All politicians running for a national electoral competition look for external financiers willing to support their campaign and wealthy organizations are likewise disposed to strengthen their political connections for future favors. As a matter of transparency, campaign contributions have to be declared to national guaranteeing institutions, which supervise the fairness of any electoral competition. In the United States of America, political candidates must notify the *Federal Election Commision* (FEC) about their electoral funds and, for that reason, FEC now represents the primary source of data for contributions campaign. In particular, FEC collects any monetary transaction, which is made by single individuals or *Political Action Committees* (PACs) for financing candidates. Given data availability, an intense scientific debate on the effects of lobbying activity has developed over the last decades. The recent economic literature is actually discussing about whether PACs are valid instruments for measuring the lobbying pressure over politicians (Teseo, 2020) and whether lobbies effectively prevail in shaping the political agenda of elected politicians.

It is reasonable to believe that politicians are inclined to collaborate with colleagues that share a similar political agenda. A possible approach for measuring the intensity of legislators' political connections is analysing the cosponsorship network, in a given Congress (Battaglini et al., 2019; Fowler, 2006a, 2006b; R. K. Wilson & Young, 1997). The US Congress system states that each legislative proposal must be promoted by one leading proponent and that it can be additionally explicitly supported by other members of the Congress. Legislator may choose to support a legislative act promoted by a colleague either because they

have participated in writing the proposal or because they are willing to demonstrate their ideological approval to the act, by following either their personal opinion or a political strategy. Those politicians who sign a submitted bill, are called *cosponsors*, while the first signatory is the *sponsor*. Despite cosponsorship ties are likely to be observed, as each legislator supports a large number of bills, they are nevertheless informative in signaling the presence of a real collaborative tie between legislators. Indeed, each politician faces a significant search cost in deciding which acts might be worth supporting, among the huge amount of bills presented at the HoR (Fowler, 2006a). Therefore, cosponsorship networks have become widely popular in the political research. The cosponsorship network is a direct network, where the set of nodes represents elected politicians, and the set of edges describes their cosponsorship links. In particular, a present link starting from a node i and running out into the node j signals that i has cosponsored at least one act, which has been presented by his colleague j. In most empirical studies, this network is also weighted, where the weight of the link measures the number of cosponsored bills.

If companies and corporations succeed in influencing the political agenda of elected politicians, by supporting their electoral campaign through PACs, and if politicians are more likely to work together with colleagues sharing similar political priorities, then it is reasonable to believe that two politicians sponsored by common lobbies are more likely to collaborate. In particular, we assess the effect of two politician having common financing sponsors on their joint legislative behavior, that is, their legislative collaborations. The idea is that sharing funding supporters leads politicians to strengthen their working relation, inducing them to work on similar issues and to co-operate about specific proposals, while being implicitly or explicitly stimulated by lobbies. This mechanism may partially tweak the role of the standard incentives of pairs of elected officials to co-operate: the components that are commonly regarded as pushing factors for collaborations are the common party membership, the common geographical area, where the two politicians have settled their political roots, and the joint connection to ethnic, religious or gender minori-

ties (Battaglini & Patacchini, 2018; Cranmer & Desmarais, 2011; Fowler, 2006a). Additionally, since politicians who receive funds from firms are likely to share at least one common financing firm, we identify strong ties of common supporters, by detecting and isolating those links that are statistically relevant. We assess whether the kind of the strong tie relationship affects the legislative collaborations among legislator and whether it encourages mutual cosponsorships rather then unilateral ties. Furthermore, as both the common supporting companies networks and the cosponsorship networks provide oriented ties, we test whether the direction of the collaboration tie follows the direction of the strong tie, hence directly identifying the role of sharing common lobbies in determining cosponsorships.

5.3.2 Data Description

Data refer to three legislative periods, namely the $111^{th.}$, $112^{th.}$ and $113^{th.}$ Congress. Consequently, the time-indicator t varies over a three-dimensional domain, $t = \{1, 2, 3\}$, where each element corresponds to the first, second and third period, respectively. The whole data structure can be represented as a temporal bipartite network $\mathrm{G}^{t,BIP} = (\mathcal{F}^t, \mathcal{P}^t, \mathcal{E}^t_{FP}, \mathcal{E}^{t,COSP}_{PP})$, where the first set $\mathcal{F}^t = \{1, \ldots, N^t_F\}$ includes those firms, who have financially supported any legislators at time t, while the second set $\mathcal{P}^t = \{1, \ldots, N^t_P\}$ embraces the members of the Congress at time t. Politicians are connected by cosponsorship ties, which are collected in $\mathcal{E}^{t,COSP}_{PP}$. We denote the cosponsorship network among elected members of the HoR at time t by $\mathrm{G}^{t,COSP}_P = (\mathcal{P}^t, \mathcal{E}^{t,COSP}_{PP})$. Links between the two layers $\mathcal{E}^t_{FP}$ represent financial contributions of firms to the benefit of politicians, during their electoral campaign: thus links between the two networks are weighted [5]. These monetary transactions are characterized by specific attributes, such as the date and the exact amount. Both firms and politicians are characterized by attributes. The idea of the data structure is depicted in Figure 49. Some descptive statistics can be found in Appendix

[5]Data provide also connections among firms: they are connected by ownership ties. However, this aspect is not addressed in the present work.

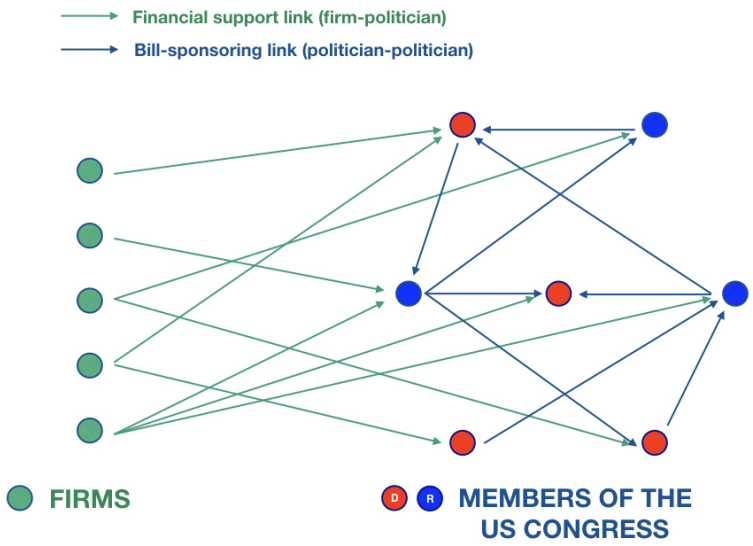


Figure 49: Data Structure: bipartite network. First layer: firms (green nodes). Second Layer: members of the Congress (red and blue nodes). Colors refer to party-membership: red denotes democrats, while blue labels republicans

D.2. In the remaining part of this section, we describe the sources of data and we show how we have merged the different data repositories.

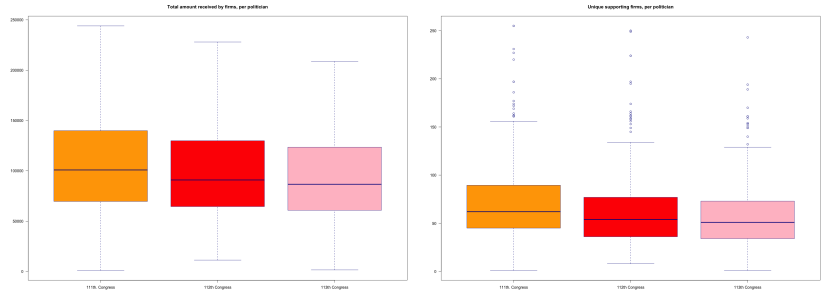
Committee Data

Campaign contributions data from the Federal Election Commission (FEC) files are collected by the Center for Responsive Politics (CRP) (Stewart & Woon, 2007). This data repository is extremely rich, as it collects the complete lists of monetary transactions at the benefit of elected politicians. In particular, data provide details of each single monetary transaction, which has been made from a given Political Action Committee (PAC) to the benefit of candidates, at the aim of supporting them during their electoral campaign. In particular, each payment is characterized by the date when it has been executed, by the type and industry with which the cor-

responding PAC is associated, by the exact amount of the transaction and by the final beneficiary. PACs include associations, organizations, single individuals, political parties and firms. Since the focus of this work is on the lobbying activities of companies, we remove from the analysis all those transactions involving PACs, which are not related to real US firms. We detected companies-related PACs by performing a string-matching algorithm, whose detailed steps can be found in the Appendix D.1. In the 111^{th}, 112^{th} and 113^{th} Congress, we have found $N_F^1 = 1007$, $N_F^2 = 970$ and $N_F^3 = 897$ unique lobbying firms, respectively. Not all elected politicians receive funds from companies: in the 111^{th} Congress politicians who were sponsored by firms were $N_P^1 = 379$ of the 439 members of the HoR, in 112^{th} Congress they were $N_P^2 = 374$ of 434 and, finally, in the 113^{th} Congress, $N_P^3 = 328$ of 416. These contributions represent the links between the two layers of the network, $\mathcal{F}^t$ and $\mathcal{P}^t$, $\mathcal{E}_{FP}^t$. In all legislative cycles, the proportion of transactions involving firms is about one third. This means that contributions coming from firms represent an extremely relevant quote of the total amount contributions registered in an electoral campaign. Figure 50 shows the distribution of the total amount of money (expressed in US dollars), which the HoR members have received from firms and, furthermore, it depicts the distribution of the number of unique supporting companies, per politician. As you may observe in the figure, the amount of money received from firms is large, and there is a large array of firms participating to campaign contributions.

Cosponsorship Network Data

In the US House of Representatives a bill is proposed by at most one politician, and other members of the Congress may decide to support it, officially cosponsoring it. Data about cosponsorships are collected by the US Government and each potential user can access them for free, by consulting the website GovTrack.us. In the recent decades, cosponsorships data have been organized and widely employed in various empirical studies (Battaglini et al., 2019; Fowler, 2006a; Kessler & Krehbiel, 1996; R. K. Wilson & Young, 1997). In our setting, cosponsorhip network data provide information about the number of cosponsored bills between



(a) Total amount (b) Unique firms

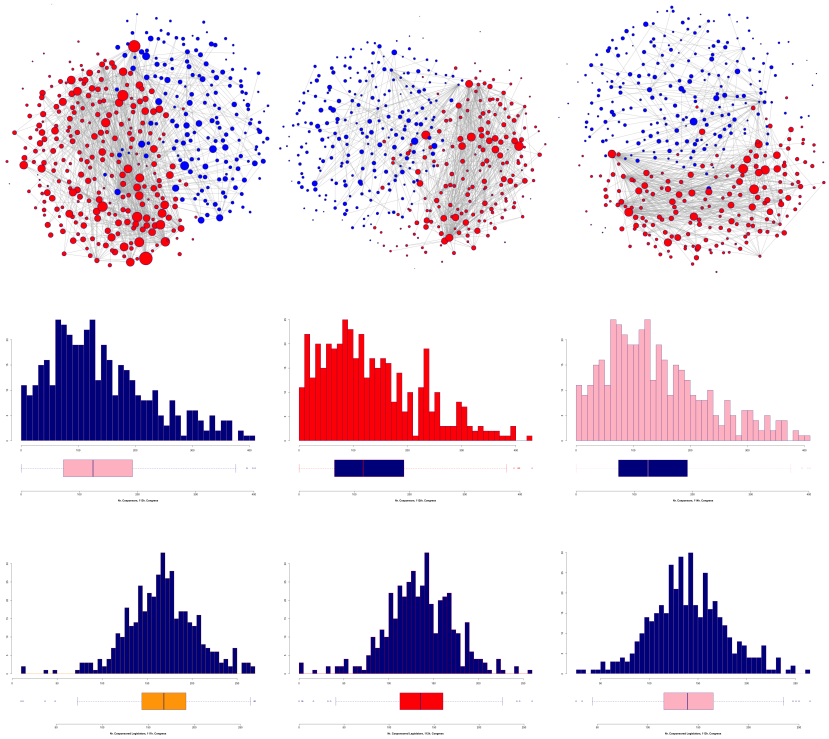
Figure 50: Box-plot: distribution of the total amount received by firms and of the number of unique supporting companies, per politician, over the three legislatures. Only those politicians who have received funds by firms are shown

HoR members, in the 111^{th}-113^{th} Congress. The network $\mathrm{G}_P^{t,COSP}$ describes cosponsorships among elected officials during the period t. It admits an equivalent representation in terms of its adjacency matrix $\boldsymbol{C}^t = \{c_{ij}^t : i, j \in \mathcal{P}^t\}$. The generic element c_{ij}^t counts how many times the HoR member i has cosponsored a bill, which was promoted by his colleague j, during the period t. Consequently, the matrix $\boldsymbol{C}^t$ is not necessarily symmetric. Figure 51 illustrates cosponsorship networks of the 111^{th}, 112^{th} and 113^{th} U.S Congress, where each node corresponds to each legislator, the color of each node is related to his/her own party membership (democrats are *red* colored, while republicans are *blue* colored) and the size is proportional to his/her in-degree in the cosponsorship network (i.e the number of legislators who have cospoonsored his/her bills at least once) [6].Note that cosponsorship networks are high density [7] networks (density is 0.3808469 in the 111^{th} Congress, 0.3115282 in the 112^{th} Congress and 0.3392203 in the 113^{th} Congress). It also depicts the histograms and the corresponding box-plots related to distributions of the

[6]For ease of observation, we are plotting only those links c_{ij}^t that are in the top 1% in terms of weight

[7]density in a network measures the ratio between the number of present ties and the number of feasible connections

total number of politicians each member of the HoR has cosponsored at least once during the legislative period and of the total number of unique cosponsors, over the three legislative cycles we take into account (i.e his/her out-deree and in-degree in the cosponsorship network, respectively). This figure points out that collaborations in the U.S HoR are intense, since each politician establish many collaborative ties with his colleagues, by cosponsoring bills that they promote.



(g) 111^{th} Congress (h) 112^{th} Congress (i) 113^{th} Congress

Figure 51: Cosponsorships, in the three legislative Cycles ($111^{th} - 113^{th}$): cosponsorship network (top), distribution of the total number of cosponsors, per politician (middle) and distribution of the total number of cosponsored legislators, per politician (bottom).

Politicians Data

Politicians data aggregate political, demographic and ideological characteristics concerning those politicians, who have been elected at the US House of Representatives (C & Wiseman, 2017; Stewart & Woon, 2007). Political attributes include party membership, legislative seniority, ideological collocation (in terms of distance with respect to the ideological "center"(R. Carroll et al., 2011)). Politicians' ideological collocation is a continuous variable bounded between 0 (highly moderate) and 1 (highly extreme): we employ this measure to classify legislators in *moderate* (ideological distance lower or equal than 0.5) or *extreme*(ideological distance greater than 0.5) . General and demographic features report the US state they politically refer to and signal whether a given politician belongs to ethnic or gender minorities. These factors are employed in our analysis to control for alternative drivers of cosponsorships: indeed, some studies have recently pointed out that being of the same party, coming from the same US State and being of the same gender represent significant elements that prompt cosponsorships between legislators (Battaglini & Patacchini, 2018; Cranmer & Desmarais, 2011). Moreover, Baller, 2017 has verified that the common party membership and the common state origin drive collaborations in the US Parliament as well. Hence, as we reasonably expect these factors to play a role in our setting, we control for this information in order to identify the effect of lobbying activities on political collaborations.

Firms Data

In our analysis we make use of the Orbis database (Van Dijk, 2013), a large multi-country firm-level dataset that contains cross border ownership information. The Orbis database, compiled by the Bureau van Dijk, a Moody's Analytics Company, is acknowledged as a reliable source of firm-level data (Del Vicario et al., 2016; Kalemli-Ozcan et al., 2015), albeit not yet exploited in studies related to the economics of political connections and cosponsorships. In particular, the Orbis database collects general characteristics and financial information, about companies and

firms that are officially recognized by United States of America. This is a huge database, which is widely employed in the economic research, because of its extensive information of all-size firms located in different countries all over the world. Each company is uniquely identified by a BVD ID number and the ORBIS Database provides information about its exact location and the economic sector to which its business activity is related. Moreover, the firm is monitored over time in terms of activities, acquisitions and economic performances. We focus on data from 2010 to 2016, so to completely cover the whole time frame that has been taken into account in the analysis. By handling this database, we can discern which organizations among the ones included in the committee data are real US companies.

5.3.3 Merging Data Sources

This work merges the four sources of data we have previously described: (i) *committee data*, which provide information about the campaigns contributions to the HoR candidates; (ii) *cosponsorship network data*, which monitor the legislative collaborations among elected politicians; (iii) *firms data*, which cover the most relevant US firms and yield general attributes of the companies, also providing some financial indicators, and their variation over time; (iv) *politicians data*, which give a unique portrait of the members of the Congress, examining their physical, cultural, ideological and political characteristics. These four data repositories are merged according to the map that can be observed in Figure 52. Cosponsorship network data and politicians data are immediately merged through an ID, which is assigned by the FEC for uniquely identifying the members of the Congress. The same ID is also used to match politicians' data together with committee data (which originally involve also candidates who have not been elected). Committee data and firms data are instead matched using a pure string-matching algorithm, that is fully described and discussed in Appendix D.1. This algorithm has been implemented in order to isolate firms or corporate companies among all the entities, whose supporting contributions are included in the committee

data. We assume that all the organizations, which have financially supported elected candidates but have not been matched by the algorithm, are not registered as US companies and, therefore, can be removed from our analysis.

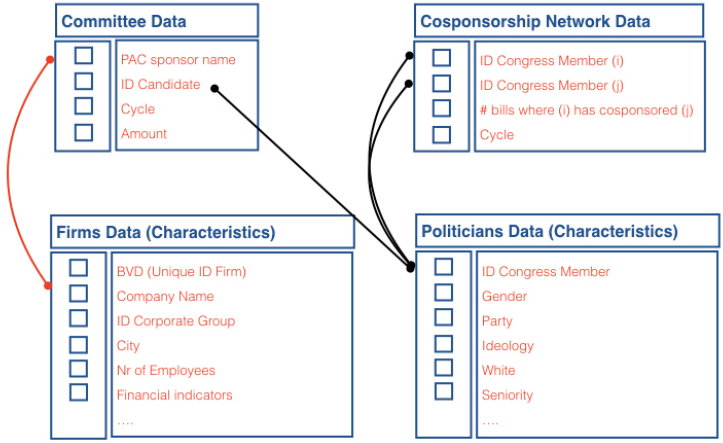


Figure 52: Data Map: four data repositories, perfectly merged through unique IDs (black arch) or non-perfectly matched using a string-matching algorithm (red arch)

5.4 Empirical Strategy

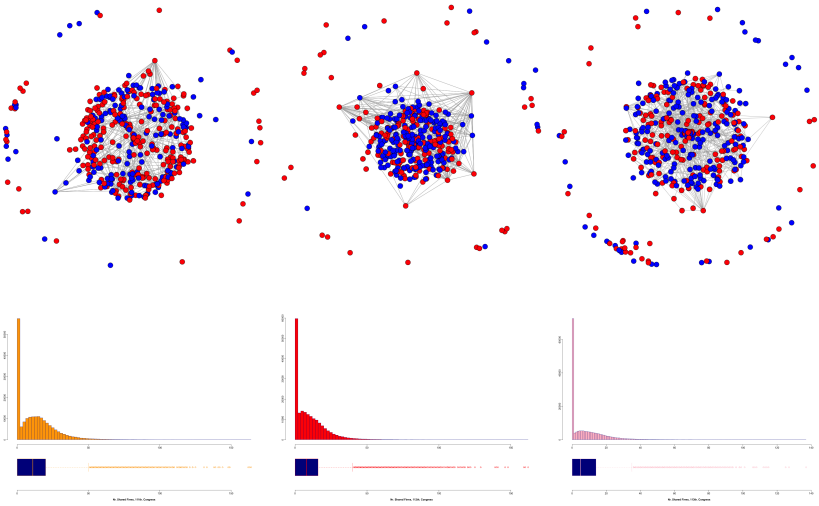
This work plans to assess the causal effect of the lobbying pressure exerted by firms on politicians on their legislative cosponsorships. Indeed, supposing that companies support politicians during their electoral campaign in an attempt to sway their political agenda and that politicians establish legislative collaborations to persuade colleagues for the approval of acts representing their political priorities, we expect that legislators pushed by the same lobbying interests to collaborate. Specifically, this project intends to appraise the exact mechanisms which may rule the causal relationship between lobbying activities and cosponsorships ties, by testing some empirical hypotheses. These hypotheses investigate whether

the type of the strong tie among two politicians (absent, asymmetric or mutual) encourages cosponsorships (unilateral or mutual) and whether the direction of the cosponsorship tie follows the direction of the strong tie. The methodological approach that we propose has been extensively discussed in Section 5.2. The whole methodological architecture is focused on singular dyads (i, j), that is, unique pairs of politicians are the units of interest. The treatment variable is defined starting from the network describing strong ties of common lobbies between politicians (which represent the treatment-network), while the dyadic outcome variable is defined starting from ties observed in the cosponsorhip networks (the outcome-network). We will provide different definitions of the treatment and outcome variables according to different specifications of the functions applied to the treatment and outcome network. In the upcoming section, we present the empirical strategy that we have implemented to define the treatment network, representing strong links of common supporting firms among politicians.

5.4.1 Strong Links of Common Supporting Lobbies

Our treatment of interest concerns the presence of common financing companies between pairs of politicians. To define the specific treatment network we employ in this work, we first create a *Shared Supporting Firms (SF)* network, $\mathrm{G}_P^{t,SF} = (\mathcal{P}^t, \mathcal{E}_{PP}^{t,SF})$, which is obtained by collapsing the bipartite network $\mathrm{G}^{t,BIP}$ over $\mathcal{P}^t$. Its edges $\mathcal{E}_{PP}^{t,SF}$ signal whether two given politicians share at least one common supporting companies and they are characterized by a specific edge-attribute that provides information about the exact number of unique firms that two given members of HoR both have among their financial supporters. However, since each firm allocates money so to benefit a high number of elected politicians, the number of zero-links in this network is negligible. Dyads which involve politicians who have both received funds from firms, are likely to present at least one shared sponsor. Figure 53 exactly pictures this phenomenon: shared supporting firms networks are highly dense networks (densities are 0.7327883, 0.7322932 and 0.6055723, respectively), where

the isolated nodes gravitating around the gathered vertexes mostly represent units who have not been supported by companies. The distribution of the recorded values of shared supporting firms, over the network dyads, has a relevant peak at 0, but, this is due to the presence of no-sponsored politician: the conditional dyadic probability of having a common supporter, given the presence of both sponsored units, is about 0.98, in all the three legislative cycles. It is relevant to point out that this particular aspect does not allow us to discern those ties that portrait a real relevant connection among politicians, in terms of shared supporting companies.



(d) 111^{th} Congress (e) 112^{th} Congress (f) 113^{th} Congress

Figure 53: Networks of shared supporting firms (top), in the three legislative Cycles (111^{th} - 113^{th}; edges signal that two nodes share at least one supporting firms; colors refer to party membership: democrats (*red*) and republicans (*blue*). Distribution of shared supporters (bottom) , per politician

As we have previously hinted, this phenomenon is induced by the common practice of companies to allocate money among many political candidates. Figure 54 shows the distribution of the number of sponsored

politicians, per firm. The median number of sponsored members of the Congress per company is about 9, in the three cycles. There are even companies, which decide to distribute their lobbying budget over hundreds of elected politician. This aspect is particularly relevant as it shows the firms' attitude to diversify their lobbying risk, by widening the circle of supported politicians so to minimize the possibility to support a politician who will not be elected. For this reason, we decide to fur-

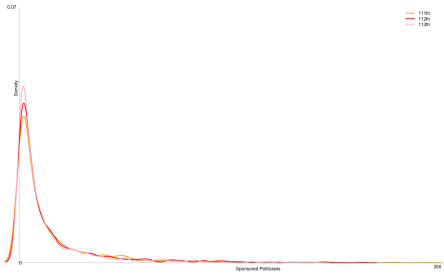


Figure 54: Density distribution of the number of supported politicians, per firm, in the three legislative Cycles (111^{th} - 113^{th})

ther examine the network $G^{t,SF}$, at the aim of detecting those links that are statistically relevant, for each unit. Identified connections are considered to be *strong links* (Battaglini et al., 2019; Granoveter, 1973; Kirkland, 2011), in terms of common financiers, for each politician. Formally, we consider the set of politicians $\mathcal{P}^t$ elected in the legislative period t, with given units i and j both belonging to $\mathcal{P}^t$. We denote with $\epsilon_{ij}^{t,SF}$ the number of common financing firms between i and j (note that, by construction, $\epsilon_{ij}^{t,SF} = \epsilon_{ji}^{t,SF}$). Moreover, let $\boldsymbol{\epsilon}_{i\cdot}^{t,SF}$ be the vector collecting the number of financing firms that i shares with her colleagues. We state that i has a *strong link* of sharing supporters with j in period t, $S_{ij}^t = 1$, if the observed value of shared supporters stays at the extreme of the distribution, that is,

$$S_{ij}^t = 1 \quad \text{if} \quad \epsilon_{ij}^{t,SF} \geq \left(\overline{\boldsymbol{\epsilon}_{i\cdot}^{t,SF}} + sd(\boldsymbol{\epsilon}_{i\cdot}^{t,SF})\right),$$

where $\overline{\boldsymbol{\epsilon}_{i\cdot}^{t,SF}}$ denotes the sampling mean of $\left(\boldsymbol{\epsilon}_{i\cdot}^{t,SF}\right)$, while $sd(\boldsymbol{\epsilon}_{i\cdot}^{t,SF}$ is the corresponding standard deviation. Similarly, j has a *strong link* of sharing

supporters with i if

$$\mathrm{S}_{ji}^{t} = 1 \quad \text{if} \quad \epsilon_{ij}^{t,SF} \geq \left(\overline{\boldsymbol{\epsilon}_{j\cdot}^{t,SF}} + sd(\boldsymbol{\epsilon}_{j\cdot}^{t,SF})\right).$$

where $\overline{\boldsymbol{\epsilon}_{j\cdot}^{t,SF}}$ denotes the sampling mean of $\boldsymbol{\epsilon}_{j\cdot}^{t,SF}$, while $sd(\boldsymbol{\epsilon}_{j\cdot}^{t,SF})$ is the corresponding standard deviation. This procedure allows to detect relevant links, from the legislator's perspective. Strong links are collected in the edge set $\mathcal{E}_{PP}^{t,SL}$, where $\mathcal{E}_{PP}^{t,SL} \subset \mathcal{E}_{PP}^{t,SF}$. Hence, the set $\mathcal{E}^{t,SL}$ embraces only those edges in $\mathcal{E}^{t,SF}$, that have been detected as strong, from the politicians point of view. It follows that the network $\mathrm{G}_{P}^{t,SL} = (\mathcal{P}^{t}, \mathcal{E}_{PP}^{t,SL})$ is a sub-graph of the original shared supporting firm network $\mathrm{G}_{P}^{t,SF}$. Finally we indicate as $\boldsymbol{SL}^{t}$ the $N_P^t \times N_P^t$ binary matrix representing strong links of sharing supporting firms. Specifically, the element $\mathrm{sl}_{i,j}^{t}$ of this matrix equals 1 if $S_{ij}^{t} = 1$. It is a direct matrix, as i may have a strong relation with j but not vice versa. This is due to our proposed characterization of a strong link relationship: the statistical relevance of each link is measured with respect to each politician, by gauging his own distribution of shared supporting firms. This means that the number of supporters that two legislators i and j - both elected at time t -, s_{ij}^{t}, may be statistically relevant i) from the i's perspective only ($\mathrm{S}_{ij}^{t} = 1$ and $\mathrm{S}_{ji}^{t} = 0$), ii) from the j's perspective only ($\mathrm{S}_{ij}^{t} = 0$ and $\mathrm{S}_{ji}^{t} = 1$), iii) from both perspectives ($\mathrm{S}_{ij}^{t} = 1$ and $\mathrm{S}_{ji}^{t} = 1$) or iv) for neither i nor j ($\mathrm{S}_{ij}^{t} = 0$ and $\mathrm{S}_{ji}^{t} = 0$) .

This procedure allows to identify, for each member of the HoR, those colleagues with whom he shares a relevant number of financing firms.

5.4.2 Assessing the Effect of Lobbying Activities on Cosponsorships

The network that we have defined in the previous subsection, $\mathrm{G}_{P}^{t,SL}$, represents the treatment network of our methodological design and depicts strong links of common supporting firms among politicians. while, the outcome network is the cosponsorships network, $\mathrm{G}_{P}^{t,COSP}$. In Section 5.2, we have clarified that the dyadic treatment variable and the dyadic outcome variable are defined by applying a function over edges

collected in the treatment network and in the outcome network, respectively. However, our motivating application provides networks, which are characterized by a temporal variation. Before going deeper in describing the empirical strategy for assessing the effect of lobbying activities on legislative collaborations, we point out here how we deal with the temporal dimension. We assume that observations come from one cross-sectional study, developed in the three legislative periods that we include in the analysis. We consider the whole set of politicians $\mathcal{P}$, which results by the union of the legislative specific sets: $\mathcal{P} = \bigcup_{t:1}^{T} \mathcal{P}^t$, with $\parallel \mathcal{P} \parallel = P$. Hence, the entire dyadic population, that we denote by $\mathcal{D}$, will comprise all the pairs of politicians over the three periods, $\mathcal{D} = \bigcup_{t:1}^{T} \mathcal{D}^t$, where the element $\mathcal{D}^t$ refers to the legislature-specific dyadic population, $\mathcal{D}^t : \{(i,j)^t : 1, \ldots, D^t\}$, with both i and j belonging to $\mathcal{P}^t$. Some politicians have run in more than one legislative cycle and will appear in the data multiple times [8]. Consequently, the dyadic population will also contain duplicates.. We are assuming that a given politician i who is in charge at time t is a different statistical unit with respect to the same politician i who is elected at time t', with $t' \neq t$. The same assumption holds for dyads: a pair of officials (i,j) can be defined over many legislative periods, but identical pairs, which are considered in distinct time frames, are assumed to be statistically different. This assumption seems reasonable in our empirical setting. Indeed, the individual legislative behavior of a legislator is analyzed with respect to a specific composition of the legislative arena, characterized by specific political equilibria. As a result, we look at the same individual in two different congresses as two different individuals to take into account the change of political equilibria occurring when a different Congress is formed. This hypothesis has been similarly advanced in many empirical studies (Battaglini et al., 2019), and it is motivated exactly by the volatility of the political scenario. Any adjustment in the political equilibrium leads to a change in the legislators' behavior. If we relaxed this assumption and admit-

[8] There are 244 politicians who are elected over the three legislative cycles. 334 politicians among the ones who were in charge during the 111^{th} Congress have been elected in the 112^{th} congress as well, while 324 among the ones who were in charge during the 113^{th} Congress have been elected in the 113th congress as well.

ted time dependencies, we would still imposing a strong assumption on data, namely we would implicitly assume that the political equilibrium remains unaltered during the three congresses (which cover more 12 years). Since this assumption is difficult to be justified and it is not considered to be plausible by the existing literature on cosponsorships, we prefer to rely on the independence assumption.

Under this assumption, it is possible to consider the network $\mathrm{G}_P^{SL} = (\mathcal{P}, \mathcal{E}_{PP}^{SL})$, whose edges describe strong ties of common supporting companies among politicians, over the three congresses, $\mathcal{E}_{PP}^{SL} = \bigcup_{t:1}^{T} \mathcal{E}_{PP}^{t,SL}$, and whose nodes represent the entire set of politicians $\mathcal{P}$. It represents the union of the three temporal sub-graphs $\mathrm{G}_P^{t,SL}$, with $t = 1, 2, 3$, and its corresponding adjacency matrix $\boldsymbol{SL}$ is a $(P \times P)$ three block-diagonal adjacency matrix with blocks $\boldsymbol{SL}^t$. The network G_P^{SL} will be our treatment network G^z, with $\mathcal{E}_{PP}^{SL} = \mathcal{E}^z$. Similarly, it is possible to define the network $\mathrm{G}_P^{COSP} = (\mathcal{P}, \mathcal{E}_{PP}^{COSP})$, whose edges signal cosponsorships among politicians, over the three congresses $\mathcal{E}_{PP}^{COSP} = \bigcup_{t:1}^{T} \mathcal{E}_{PP}^{t,COSP}$, and whose nodes still represent the entire set of politicians $\mathcal{P}$. It represents the union of the three temporal sub-graphs $\mathrm{G}_P^{t,COSP}$, with $t = 1, 2, 3$ and its corresponding adjacency matrix $\boldsymbol{C}$ is a $(P \times P)$ three block-diagonal adjacency matrix with blocks $\boldsymbol{C}^t$. The network G_P^{COSP} represents our outcome network G^y, with $\mathcal{E}_{PP}^{COSP} = \mathcal{E}^y$.

As specified in Section 5.2, for each pair politicians (i, j), the dyadic treatment variable and the outcome treatment variable will be defined, applying a mapping function over their observed ties in the treatment network and in the outcome network, respectively. In particular, the dyadic treatment Z_{ij} results from applying the $f(\cdot)$ function on the pair specific ties observed in the treatment network, $Z_{ij} = f(e_{ij}^z, e_{ji}^z)$. The dyadic outcome Y_{ij} results from applying the $g(\cdot)$ function on the pair specific ties observed in the treatment network, $Y_{ij} = g(e_{ij}^y, e_{ji}^y)$.

Therefore, given G^z and G^y, we can define dyadic causal effects to investigate the mechanisms which rule the causal link between lobbying activities and cosponsorships. The first aspect that it is relevant to assess is whether the presence of strong ties between two legislators affects their likelihood to collaborate. The two variables S_{ij} and S_{ji} signal the

presence and symmetry of a strong relationship of common supporting lobbies between i and j. This strong tie may not be symmetric, as, for instance, i may have a statistically relevant tie of common financiers with j ($S_{ij} = 1$) but not vice versa ($S_{ji} = 0$). Consequently, we can state that the strong relationship of common supporters between two politicians may be of three types: i) *symmetric (mutual)*, (ii) *present and asymmetric* or iii) *absent*. The first dyadic causal effect we are interested in is the causal effect of the type of strong tie of common supporters on the presence of any co-sponsorded bills. Therefore, for the definition of this causal effect, the treatment variable Z_{ij}^{symm} has a multi-valued characterization, and it varies over three nominal categories. Formally,

$$Z_{ij}^{symm} = \begin{cases} = 0 & \text{if } S_{ij} = S_{ji} = 0 \quad \text{(no strong tie)} \\ = 1 & \text{if } S_{ij} + S_{ji} = 1 \quad \text{(asymmetric strong tie)} \\ = 2 & \text{if } S_{ij} = S_{ji} = 1 \quad \text{(mutual strong tie)} \end{cases}$$

To evaluate the effect of strong ties on cosponsorships, one needs to assess the effect of Z_{ij}^{symm} on an outcome variable, which signals the presence of cosponsored bills in the dyad. Specifically, the outcome Y_{ij}^{any} is a dummy, which signals if *either* i has cosponsored j at least one bill during the legislature *or* j has cosponsored i. Denoting with C_{ij} the number of bills which have been promoted by j and successively supported by i and with C_{ji} the number of proposals cosponsored by j, after having been fostered by i, we formally define the outcome variable as

$$Y_{ij}^{any} \begin{cases} = 0 & \text{if } C_{ij} = C_{ji} = 0 \quad \text{(no cosponsorships)} \\ = 1 & \text{if } C_{ij} + C_{ji} > 0 \quad \text{(at least one cosponsorship)} \end{cases}$$

Since the treatment variable Z_{ij}^{symm} varies over $K = 3$ nominal categories, assessing the effect of Z_{ij}^{symm} on Y_{ij}^{any} amounts to perform binary comparisons between z and z', with z and$z' \in \{0, 1, 2\}$. These effects can be formally written as,

$$\tau_{(z,z')}^{M1} = E\left[Y_{ij}^{any}(Z_{ij}^{symm} = z)\right] - E\left[Y_{ij}^{any}(Z_{ij}^{symm} = z')\right]$$

The effect of Z_{ij}^{symm} on Y_{ij}^{any}, $\tau_{(z,z')}^{M1}$, represents the key causal investigation of our work, as it enables us to state whether strong ties do

play a role in determining collaborations among legislators. The subsequent hypothesis that requires to be tested regards the type of cosponsorships that are mostly encouraged by strong ties: in other words, we ask whether strong ties encourage asymmetric cosponsorships rather than mutual cosponsorships. Hence, conditioning on observing at least one cosponsorship in the dyad, we test the effect of the strong common supporters tie on the presence of mutual cosponsorhip (*M2*). We state that two politicians i and j mutually cosponsor one each other if both the numbers of legislative proposals that the two officials have reciprocally supported is strictly greater than 0, that is, $C_{ij} > 0$ and $C_{ji} > 0$. Instead, two members of the HoR are connected through a one-sided cosponsorship tie if one politician belonging to the observed pair has officially sustained at least one bill which has been submitted by the other, but not vice versa. This analysis is restricted only on those dyads who have at least a one-sided cosponsorship tie: consequently, pairs of politicians who do not reciprocally support any bill during the given legislature are excluded from this specific examination. Hence, this analysis is addressed to a subset of the entire population of dyads $\mathcal{D}$: labelling as $D^{(M2)}$, with $D^{(M2)} \subset D.$, the restricted sample of dyads we take into account. Therefore, here the treatment variable is identical to the one that we have introduced in the previous subsection (Z_{ij}^{symm} for all the pairs $(i,j) \in \mathcal{D}^{(M2)}$), while the outcome is defined as a dummy which equals 1 in case of mutual cosponsorship and 0 under one-sided cosponsorship and it is still defined only for those dyads who exhibit at least a one-sided cosponsorship tie.

$$\left[Y_{ij}^{symm} \mid Y_{ij}^{any} = 1\right] = \begin{cases} = 0 & \text{if } C_{ij} + C_{ji} > 0,\ C_{ij} = 0 \text{ or } C_{ji} = 0 \\ & \text{(one sided cosponsorships)} \\ = 1 & \text{if } C_{ij} > 0 \text{ and } C_{ji} > 0 \\ & \text{(mutual cosponsorship)} \end{cases}$$

The effects of interest are conditional average causal effects as they are defined while conditioning on those dyads that are characterized by at

least a one-sided cosponsorship tie. They can be formalized as

$$\begin{aligned}\tau^{M2}_{(z,z')} =& E\left[Y^{symm}_{ij}(Z^{symm}_{ij} = z)|Y^{any}_{ij} = 1\right]- \\ & E\left[Y^{symm}_{ij}(Z^{symm}_{ij} = z')|Y^{any}_{ij} = 1\right]\end{aligned}$$

After investigating the existence of a causal relationship between strong ties and cosponsorships, we focus on the directions of those links: specifically, we wish to assess whether the direction of the strong ties follows the direction of the cosponsorship tie. To achieve this goal, we analyze the direction of ties, under two different perspectives: i) focusing on those dyads who exhibit asymmetric relationships both in the treatment and in the outcome network, we test whether the presence of a strong tie in a given direction increases the probability of observing the cosponsorship tie on the same direction (rather than in the opposite direction) (*MOD3*); ii) focusing on those dyads who exhibit asymmetric behaviors in the treatment network only, we test whether the presence of a strong tie in a given direction determines the predominant direction of the total cosponsorship activity (i.e the direction of the most relevant number of cosponsored bills, in the pair)(*MOD4*).

In *MOD3*, we assess whether the direction of the present cosponsorship tie reflects the direction of the strong link of common sponsors while focusing on probabilities and conditioning on those dyads who exhibit asymmetric relationships both in the strong common lobbies tie and in the cosponsorship behavior. Hence, the empirical investigation is confined on the subset of dyads, that show asymmetries both in the strong tie relationship and in the cosponsorship activity. That is, we focus on a subset, $\mathcal{D}^{(M3)}$, with $\mathcal{D}^{(M3)} \subset \mathcal{D}$, that can be formally described as $\mathcal{D}^{(M3)} = \{(i,j) \in \mathcal{D} : \ \ S_{ij} + S_{ji} = 1 \text{ and } C_{ij} + C_{ji} > 0 \ \ C_{ij} \times C_{ji} = 0\}$. Here, the treatment variable $Z^{(M3)}_{ij}$ simply represents the direction of the

asymmetric strong link, that is

$$\left[Z_{ij}^{dir} \mid Z_{ij}^{symm} = 1, Y_{ij}^{symm} = 0\right] = \begin{cases} = 0 & \text{if } S_{ij} = 1 \text{ and } S_{ji} = 0 \\ & (i \text{ has a strong tie with } j \\ & \text{but not vice versa}) \\ = 1 & \text{if } S_{ij} = 0 \text{ and } S_{ji} = 1 \\ & (j \text{ has a strong tie with } i \\ & \text{but not vice versa}) \end{cases}$$

Similarly, the outcome variable $Y_{ij}^{(M3)}$ is defined as a binary indicator expressing the direction of the cosponsorship behavior. Formally,

$$\left[Y_{ij}^{dir} \mid Z_{ij}^{symm} = 1, Y_{ij}^{symm} = 0\right] = \begin{cases} = 0 & \text{if } C_{ij} > 0 \text{ and } C_{ji} = 0 \\ & (i \text{ has cosponsored } j \\ & \text{at least once, but not vice versa}) \\ = 1 & \text{if } C_{ij} = 0 \text{ and } C_{ji} > 0 \\ & (j \text{ has cosponsored } i \\ & \text{at least once, but not vice versa}) \end{cases}$$

Here, the effect of interest is a conditional average causal effect as it us defined while conditioning on those dyads who are characterized by an asymmetric behavior both in the strong tie relationship and in the cosponsorship tie. It can be expressed as

$$\begin{aligned} \tau_{(z,z')}^{M3} = & E\left[Y_{ij}^{dir}(Z_{ij}^{dir} = z) | Z_{ij}^{symm} = 1, Y_{ij}^{symm} = 0\right] - \\ & E\left[Y_{ij}^{dir}(Z_{ij}^{dir} = z') | Z_{ij}^{symm} = 1, Y_{ij}^{symm} = 0\right] \end{aligned}$$

In *MOD4*, we estimate the effect of the direction of the strong tie on the predominant direction of the total cosponsorship activity, conditioning on observing asymmetric relationships in the strong link of common supporting firms. Hence, we aim to assess whether the direction of the strong link of common sponsors has an impact on whom of the two politicians mostly supports the other, in his legislative attempts. This issue is evaluated by focusing on those pairs of politicians, who exhibit an asymmetric tendency with respect to the presence of a strong link of shared financiers. This subset can be formally described as $\mathcal{D}^{(M4)} =$

$\{(i,j) \in \mathcal{D} : \quad S_{ij} + S_{ji} = 1\}$ [9]. The treatment variable coincides, in its mathematical formulation, with the one that we have proposed in the previous examination, Z_{ij}^{dir}. However, it is defined over a different subsample of units (those showing asymmetries in the outcome network only - and not in both the treatment network and in the outcome network, as we have previously investigated -). The outcome, instead, measures the difference between the reciprocal cosponsored bills between i and j, $Y_{ij}^{diff} = C_{ji} - C_{ij}$. If this difference results to be greater then 0, then we deduce that the politician who has mostly supported the other was j. If, vice versa, the outcome is negative, then we can recognize i as the most collaborative element in the pair. The effect is outlined in the following expression,

$$\begin{aligned}\tau_{(z,z')}^{M4} =& E\left[Y_{ij}^{diff}(Z_{ij}^{dir} = z)|Z_{ij}^{symm} = 1\right] - \\ & E\left[Y_{ij}^{diff}(Z_{ij}^{dir} = z')|Z_{ij}^{symm} = 1\right]\end{aligned}$$

5.4.3 Assessing Heterogeneities in the Effects of Interest

The effects we have described and motivated in the previous subsection, all outlined with the aim of investigating the causal relationship between lobbying activities and legislative collaborations, are estimated both in the entire dyadic sub-samples, in which they are respectively defined, and in specific dyadic sub-populations, characterized by particular traits. This approach allows us to assess the heterogeneity of results and to investigate i) the dyadic sub-populations whose legislative behavior is particularly "responsive" to the presence of strong ties and ii) the dyadic characteristics that mainly drive heterogeity. We define the dyadic sub-populations of interest by using prior information about the phenomenon of interest and not relying on a data-driven approach. In particular, we suppose that factors that may prompt heterogeneities the state of election of the two legislators (`state`), their political collocation in the electoral chamber (`majority`), and their posi-

[9] Note that the set of dyads, which is employed in the previous examination (*M3*), $\mathcal{D}^{(M3)}$, is in turn a subset of $\mathcal{D}^{(M4)}$, that is, the following relation holds: $\mathcal{D}^{(M3)} \subset \mathcal{D}^{(M4)} \subset \mathcal{D}$

tioning in terms of party membership and ideological extremism (`party` and `ideology`). The latter aspect leads to the definition of the following eight sub-populations of dyads: democrats, moderate-democrats, extreme-democrats, mixed ideology-democrats, republicans, moderate-republicans, extreme-republicans, mixed ideology-republicans. Specifically, moderate pairs are those whose ideological distance to the center is lower than 0.5, extreme pairs are those whose ideological distance to the center is greater to 0.5, mixed ideology profiles denote dyads who are composed by one extreme legislator and one moderate legislator. Dealing with heterogeneity of the effects yields to a meaningful interpretation about peculiar characteristics of those pairs of politicians whose cooperation appears to be particularly motivated by the presence of common pushing companies.

5.5 Results

The effects of interest have been estimated by implementing a propensity score matching for network data we have described in Section 5.2. In all the four examinations, we include in the batch of pre-treatment covariates some baseline dyadic features, which cannot be affected by the treatment: in particular, we include a dummy variable which signals whether the two elements of the pair are of the same gender (`gender`), a dummy variable which signals whether the two elements of the pair are both referred to an ethnic minority (`nowhite`), a dummy variable which equals 1 if the two politicians belong to the same party (`party`), a dummy variable which identifies those pairs where the two politicians have been elected in the same US State (`state`) and, finally, a quantitative variable measuring the absolute value of the difference between the two officials' legislative seniority (expressed in terms of years spent in the US HoR before the current legislature)(`seniority`) [10]. These elements have been included as they represent factors that the existing literature about

[10]Note that small values of this last variable denote pair of politicians, where both elements are either similarly politically young officials or similarly politically experienced officials

cosponsorships have identified as variables that independently motivate legislative collaborations (as motivated in Subsection 5.3.1. Moreover, by conditioning on these variables, we may reasonably state that the dyadic intervention is independent on the dyadic potential outcomes (thus, we can presume the validity of the unconfoundedness assumption). Indeed, the dyadic pre-treatment characteristics that we consider in the analysis allow to control for those mechanisms which may have plausibly affect both the likelihood of legislators to share an high number of financial supporters and their legislative collaborations

The main examination of our work regards the effect of the type of the strong tie on the presence of cosponsorships (*MOD 1*). As we have previously outlined, the strong tie of common supporting firms is defined from the point of view of the single politician and, consequently, it is not necessarily symmetric. The same entry concerning the number of shared lobbies among two politicians may result to be statistically significant for both of them, only one of them or nobody of them. As a consequence, the strong tie can be mutual ($Z_{ij}^{symm} = 2$), asymmetric ($Z_{ij}^{symm} = 1$) or absent ($Z_{ij}^{symm} = 0$). We test whether the type of the strong tie among politicians affects the presence of cosponsored bills between them ($Y_{ij}^{any} = 1$).

Our findings show that the probability of forming a cosponsorship tie (either asymmetric or mutual) is higher with respect to the probability of not forming a cosponsorship tie if Congress members are linked by a strong tie (either asymmetric or mutual). In particular, this effect is stronger for mutual strong ties: a mutual strong tie leads to an increasing of 0.13164 in the probability of establishing a cosponsorship tie, with respect to an absent strong tie relationship. Instead, an asymmetric strong tie increases the probability of establishing a cosponsorship tie by 0.04063, with respect to an absent strong tie. Finally, a mutual strong tie raises the probability of collaborations by 0.09101, compared with an asymmetric strong tie. Details can be found in Table 8.

Table 8: Effect of the type of the strong tie on the presence of a cosponsorship tie (**M1**). Total Number of Dyads: $D = 276422$; Number of dyads per treatment status: $Dt = 0\ (222330),\ \ 1\ (46378),\ \ 2\ (7714)$

$\widehat{\tau}^{M1}_{(z,z')}$	ATE	Y(1)-Y(0)	Y(2)-Y(0)	Y(2)-Y(1)
M1	Estimate	0.04063 ***	0.13164 ***	0.09101***
	Std.Er	(0.00293)	(0.00671)	(0.00713)

We now examine the heterogeneity of this effect. We perform a stratified analysis building up sub-populations of dyads that are characterized by given dydadic features. Results show that the effect is stronger on those pairs of politicians where the two legislator have not been elected in the same US State (*different origin*), compared with those where the two legislators refer to the same state (*common origin*). Being both majoritarian legislators (*majoritarian dyad*) leads to a decreasing in the effect, with respect to the overall population. Party membership plays a relevant role: specifically, the impact of the presence of strong ties of common lobbies is particularly stronger on those dyads composed by republicans only (*republicans*), especially if the two legislators are characterized by a radical ideology (*extreme-republicans*). Democrats' legislative behavior is less affected by the presence of strong ties, and the effect is relevant only on those pairs, whose ideological collocation is moderate (*moderate-democrats*). Detailed results are shown in Table 9

Table 9: Effect of the type of the strong tie on the presence of a cosponsorship tie: Heterogeneity of the effect (**M1**)

$\widehat{\tau}^{M1}_{(z,z')}$	ATE	Y(1)-Y(0)	Y(2)-Y(0)	Y(2)-Y(1)
State of Election	Common origin (US state)	0.02832 ***	0.06384 ***	0.03553 ***
		(0.00952)	(0.01732)	(0.01873)
	Different origin (US state)	0.05417 ***	0.14750 ***	0.09333 ***
		(0.00305)	(0.00852)	(0.00892)
Majority	Majoritarian dyad	0.03248 ***	0.09847 ***	0.06599 ***
		(0.00310)	(0.00618)	(0.00655)
	Non-Majoritarian dyad	0.07673 ***	0.19022 ***	0.11349 ***
		(0.00510)	(0.01566)	(0.01635)
Party and ideology	Democrats	0.02473***	0.07222***	0.04749***
		(0.00422)	(0.00894)	(0.00937)
	Moderate - democrats	0.03610***	0.09683***	0.06073***
		(0.00491)	(0.00919)	(0.00972)
	Extreme - democrats	0.00928	-0.03944	-0.04872
		(0.01947)	(0.08512)	(0.08624)
	Mixed ideology - democrats	-0.00431	0.04718***	0.05149***
		(0.00874)	(0.02291)	(0.02374)
	Republicans	0.04144***	0.12398***	0.08253***
		(0.00464)	(0.01076)	(0.01126)
	Moderate - republicans	0.02816	-0.06285	-0.09102
		(0.02523)	(0.06002)	(0.06351)
	Extreme - republicans	0.03507***	0.12922***	0.09415***
		(0.00574)	(0.01058)	(0.01133)
	Mixed ideology - republicans	0.03459 ***	0.14172***	0.10713***
		(0.00842)	(0.01620)	(0.01740)

The second empirical examination we have implemented in this work regards the effect of the type of the strong tie on mutual cosponsorships. By means of the previous analysis, we have proved that having a strong tie relationship (either asymmetric or multual) motivates politicians to work together, establishing a cosponsorship tie. What we are testing now is whether a strong tie mainly encourages pairs of officials to form a mutual cosponsorship link ($Y_{ij}^{symm} = 1$), rather than of an asymmetric one ($Y_{ij}^{symm} = 0$). A mutual cosponsorship tie is present in all those dyads who have reciprocally cosponsored at least one bill in the legislative cycle. Instead, this relationship is only asymmetric if only one of the two politicians involved in the dyad has cosponsored a bill which was initially promoted by the other, but not vice versa. Note that the treatment variable we employ in this analysis exactly coincides with the one we

have defined for the previous examination, but this specific evaluation is enacted only on those pairs of units who are characterized by the presence of at least a one-sided cosponsorship tie.

Our findings suggest that the probability of forming a mutual cosponsorship tie is higher with respect to the probability of forming an asymmetric cosponsorship tie if Congress members are linked by an asymmetric strong tie or a mutual strong tie. Specifically, the magnitude of the estimated effect is higher for mutual strong ties: a mutual strong tie leads to an increasing of 0.08725 in the probability of establishing a mutual cosponsorship tie (instead of an asymmetric strong tie), with respect to an absent strong tie. Similarly, an asymmetric strong tie increases the probability of establishing a mutual cosponsorship tie by 0.01346, compared with an absent strong tie. Finally, a mutual strong tie raises the probability of collaborations by 0.07379, in comparison with an asymmetric strong tie. Complete results are collected in Table 10.

Table 10: Effect of the type of the strong common supporters tie on the presence of mutual cosponsorhips (**M2**): Total number of Dyads :$D = 144006$; Dyads per treatment status $Dt = 0(108464), 1(29732), 2(5810)$

$\widehat{\tau}^{M2}_{(z,z')}$	ATE	EY(1)-EY(0)	EY(2)-EY(0)	EY(2)-EY(1)
MF2	Estimate	0.01346 ***	0.08725 ***	0.07379***
	Std.Er	(0.00341)	(0.00726)	(0.00768)

We now examine heterogeneity of this effect. The effect is stronger on those pairs where the two politicians have been elected in different US states (*different origin*) than on those where the two legislators share a common political origin. The impact of strong ties is particularly relevant on non-majoritarian dyads (*non-majoritarian dyads*). As in the previous examination, the effect is particularly stronger for dyads whose members are republicans (*republicans*): however, in this examination, the ideological collocation plays a less relevant role, conditioning on party membership. Detailed results can be found in Table 11.

Table 11: Effect of the type of the strong tie on the presence of a mutual cosponsorship tie: Heterogeneity of the effect (**M2**)

$\widehat{\tau}^{M2}_{(z,z')}$	ATE	Y(1)-Y(0)	Y(2)-Y(0)	Y(2)-Y(1)
State of Election	Common origin (US State)	0.03722 ***	0.05675 ***	0.01953
		(0.01269)	(0.01872)	(0.02026)
	Different origin (US State) 6	0.01517 ***	0.08261 ***	0.06744 ***
		(0.00343)	(0.00829)	(0.00871)
Majority	Majoritarian dyad	0.00585	0.10178 ***	0.09593 ***
		(0.00396)	(0.00824)	(0.00869)
	Non-Majoritarian dyad	0.03261 ***	0.03250 ***	-0.00011
		(0.00640)	(0.01622)	(0.01723)
Party and ideology	Democrats	-0.00432	0.06203***	0.06635***
		(0.00549)	(0.01177)	(0.01233)
	Moderate - democrats	0.01091	0.07686 ***	0.06595 ***
		(0.00634)	(0.01272)	(0.01339)
	Extreme - democrats	0.00748	0.04485	0.03738
		(0.03722)	(0.08397)	(0.08902)
	Mixed ideology - democrats	-0.01501	0.11471 ***	0.12972 ***
		(0.01175)	(0.03283)	(0.03390)
	Republicans	0.01693 ***	0.13715 ***	0.12021 ***
		(0.00579)	(0.01237)	(0.01305)
	Moderate - republicans	-0.02959	0.11928	0.14887*
		(0.02730)	(0.07823)	(0.08069)
	Extreme - republicans	0.01618 ***	0.12932***	0.11314***
		(0.00732)	(0.01655)	(0.01735)
	Mixed ideology - republicans	0.02457 ***	0.16661***	0.14204***
		(0.01080)	(0.02652)	(0.02780)

The last two examinations intends to focus on the direction of links, by assessing whether the direction of the strong ties follows the direction of the cosponsorship tie. The analysis is performed over under two different perspectives. Specifically, the third causal evaluation (*MOD3*) focuses on those dyads who exhibit asymmetric relationships both with respect to the strong tie of common supporters and with respect to the cosponsorship activity and tests whether the cosponsoring tie follows the direction of the strong tie in terms of probabilities. Both the tretament and the outcome variable are dummy variables signaling the direction of the examined tie. The fourth causal examination (*MOD4*) intends to assess whether the direction of the strong tie of shared lobbies determines the most productive cosponsorship direction. The treatment of interest still represents the direction of the strong tie relationship: the treatment

variable equals 0 if i has a strong tie with j but not vice versa ($Z_{ij} = 0$), and equals 1 if j has a relevant relationship with i, but not vice versa ($Z_{ij} = 1$). We consider as outcome variable the difference in the number of cosponsored bills between i and j: this difference results to be greater than zero if j has cosponsored a larger number of bills promoted by i than vice versa.

Our findings suggest that the cosponsorship tie reflects the direction of the strong tie relationship, both in terms of probabilities and in terms of the number of bills. This means that legislators are encouraged to support those politicians who they regard as particularly relevant in terms of shared financiers. Formally, according to *MOD3*, the probability of observing j cosponsoring i ($Y_{ij}^{dir} = 1$) is 20% higher when j has a strong tie with i ($Z_{ij}^{dir} = 1$), with respect to the situation where i has a strong tie with j ($Z_{ij}^{dir} = 0$).. Results of *MOD4* confirm the finding: the statement that "the direction of the cosponsorship tie follows the direction of the strong tie relationship" holds even if we look at the difference in the number of cosponsored bills between i and j, rather than using probabilities. This means that j cosponsors a larger number of bills to i than vice versa, if j has a strong tie with i, and i has not a strong tie to j. Specifically, dyads where j has a strong tie with i and not vice versa, experiment an average increasing by 0.83250 in the difference between the number of bills cosponsored by j to i and the number of proposals cosponsored by i to j. Results are shown in Table 12.

Table 12: Effect of the direction of the strong common supporters tie on the direction of cosponsorhip activity: i) (**M3**); Total number of dyads: $D = 18518$; Dyads per treatment status $Nt = 0(7681), 1(10837)$, ii) (**MF4**); Total number of dyads: $N = 46378$; Dyads per treatment status: $Nt = 0(18897), 1(27481)$;

$\widehat{\tau}^{M3}_{(z,z')}$ $\widehat{\tau}^{M4}_{(z,z')}$	ATE	Y(1)-Y(0)
MF3	Estimate	0.195728 ***
	Std.Er	(0.01603)
MF4	Estimate	0.83250 ***
	Std.Er	(0.03268)

Estimated effects in the heterogeneous sub-populations confirm the general finding, that is, the direction of cosponsorships goes after the direction of the strong tie. According to both definitions of direction, the effect is higher in sub-populations of dyads whose elements are both referred to the parliamentary majority (*majoritarian dyads*) and republicans (*republicans*) (especially if they both have a radical ideological opinion). As in all the previous examinations, the legislative behavior of pairs whose elements do not have a common political origin is particularly affected by the presence of strong ties (*different state origin*) . Detailed results can be observed in Table 13. These findings are statistically significant.

Table 13: Effect of the type of the strong tie on the presence of a mutual cosponsorship tie: Heterogeneity of the effect (**M3**) and (**M4**)

$\widehat{\tau}^{M3}_{(z,z')}$ $\widehat{\tau}^{M4}_{(z,z')}$	ATE	Y(1)-Y(0) **MOD 3**	Y(1)-Y(0) **MOD 4**
State of election	Common origin (US state)	0.07948***	0.73941
		(0.03511)	(0.19003)
	Different origin (UD state)	0.19612***	0.85996***
		(0.00822)	(0.03265)
Majority	Majoritarian Dyad	0.22653***	0.95490***
		(0.00893)	(0.03758)
	Non-Majoritarian Dyad	0.11233***	0.45490***
		(0.01765)	(0.06319)
Party and ideology	Democrats	0.16760 ***	0.37488 ***
		(0.01301)	(0.03334)
	Moderate - democrats	0.16491 ***	0.39943 ***
		(0.01462)	(0.03321)
	Extreme - democrats	-0.06275	0.62572
		(0.14439)	(0.60898)
	Mixed - ideology	0.16070 ***	0.40789 ***
		(0.03090)	0.09134
	Republicans	0.28427 ***	0.69919 ***
		(0.01283)	(0.03100)
	Moderate - republicans	0.02070	0.27951 *
		(0.06755)	(0.16867)
	Extreme - republicans	0.29973 ***	0.78091 ***
		(0.01662)	(0.04076)
	Mixed ideology - republicans	0.29320 ***	0.52139 ***
		(0.02271)	(0.05124)

Summing up, we can state that the presence of strong ties of common

supporting companies motivate legislators to collaborate and, in particular, it encourages mutual cosponsorship ties, rather than asymmetric ties. Moreover, the direction of the cosponsorship tie follows the direction of the strong tie, both in terms of probabilities and in terms of cosponsored bills. Dyads whose legislative behavior is particularly influenced by the presence of strong ties are those where the two legislators have not been elected in the same US state and are both republicans.

5.6 Concluding Remarks and Discussion

In this paper, we have explored the causal relationship between the lobbying pressure exerted by companies over politicians and their legislative behavior. In particular, we have investigated whether the presence of a strong tie in terms of common supporting firms encourages legislative collaborations between two politicians. The political connections of firms have been detected by observing the committee data, which report the campaign contributions in favor of politicians. Instead, the legislative collaborations among two politicians have been measured in terms of the number of reciprocally cosponsored bills. The methodological framework reworks the standard potential outcomes framework by Rubin, 1974 so to model the dyanmics of network formation. In particular, the methodological contribution is the development of an estimator for causal effects of the formation of links on a 'treatment' network on the formation of links on an 'outcome network'. Note that the methodological approach that we propose here is suitable for scenarios where treatments are intrinsically dyadic as well as for empirical studies, where the dyadic treatment variable results from joint observation of interventions that are originally defined at individual level. We have estimated the effects of interest by implementing a propensity score matching algorithm for network data: in particular, in the specific empirical evaluations, which have required a multi-valued characterization of the treatment variable, we have implemented the innovative algorithm for propensity score matching under multi-valued treatments (recently proposed by S. Yang et al., 2016) on network data. Results point out that the

presence of a strong tie relationship, either asymmetric or symmetric, encourages both asymmetric and mutual cosponsorships between the two politicians. Moreover, we observe that the direction of the cosponsorship tie reflects the direction of the strong tie, expressed in terms both of probabilities and of the number of cosponsored bills.

This work adopts an innovative approach, which refers to the potential outcome framework, for evaluating how the pressure exerted by lobbies modifies the officials' legislative behavior, by encouraging cosponsorships among two politicians who are pushed by similar interests. From the empirical point of view, our project contributes to the growing literature about the effects of lobbies interests on politicians (Battaglini & Patacchini, 2018). However, the methodological approach that it proposes partially deviates from the standard econometric techniques that are usually implemented in previous studies. It is important to highlight that this study is the first work which effectively detects the presence of almost 1000 lobbying firms in the US electoral campaign. Moreover, it points out that firms allocate money over various political candidates. We have not identified which economic mechanisms drive the money allocation of firms over candidates. Moreover, we have not proven the existence of any type of coordination among firms, which are pushed by similar business interests. Finally, we could expect that the findings that we have obtained in this work would be even stronger, if we took into account the matter of cosponsored bills. We leave this for future research.

Chapter 6

Conclusions

In this concluding Chapter I summarize the main objectives of this dissertation alongside its main findings. In addition, I illustrate potential directions for future research.

6.1 Discussion

The present dissertation proposes novel statistical methodologies and tools for performing causal inference on network data. This work is motivated by two major issues that arise while dealing with units that are interconnected in a network: (i) social or economic connections induce mechanisms of *interference*, meaning that the outcome of a unit is affected by the treatment status of its interfering units; (ii) interactions themselves are of intrinsic interest for researchers, who are often keen on investigating the causal mechanisms driving the formation of links; yet there is scant availability of methodologies suited for this purpose. The first three chapters of this dissertation address the first of such issues, thus contributing to expand the extant statistical literature about causal inference under interference. In particular, they propose innovative strategies to: (i) estimate the effect of an observational multi-valued intervention in a sample of units connected through a weighted network; (ii) detect and estimate heterogeneous treatment and spillover effects in the presence of clustered-network interference; (iii) account for an hidden treatment diffusion process, where the intervention may spill over according to partially unobserved links. The fourth chapter instead addresses the second aforementioned issue, and it employs the potential outcomes framework to assess the causal effect that the presence links in a "treatment network" has on the formation of links on an "outcome network," with both networks being directed. In what follows I summarize the main purposes and findings of each chapter at once.

The first Chapter proposes a strategy to assess the direct effect of an observational multi-valued intervention in a sample of units connected through a weighted network. The proposed methodology, which is based on a joint multiple generalized propensity score, allows the researcher to estimate direct effects following a parametric approach, while con-

ditioning on both individual and network features. This method is employed to investigate the causal effect of national immigration policies on crime rates. In this framework, national political attitudes towards migrants are expressed with a multi-valued characterization. The interference structure is designed under the assumption that countries influence one another as a function of their cultural and geographical proximity, which is summarized through a composite indicator. The empirical findings suggest that implementing highly restrictive immigration policies leads to an increase in the national crime rate, and that ignoring multi-valued interference leads to the estimation of an effect which, while still positive, is weaker, thereby not allowing to appreciate the magnitude of the effect in its full extent.

The second Chapter develops a novel framework to discover and estimate heterogeneous treatment and spillover effects under a particular type of interference, known as "clustered network interference," which features units belonging to exogenously given clusters and interacting in networks that are cluster-specific. The proposed algorithm, named Network Causal Tree (NCT), allows to estimate treatment and spillover effects that are both heterogeneous, as well as to identify the most heterogeneous sub-populations while considering multiple effects at the same time. The NCT displays excellent performance in several simulated scenarios. Furthermore, the Chapter showcases the potential of the NCT algorithm in a real-world application: the analysis of intensive training sessions for Chinese rural households, where interest falls on the effect of the training on the uptake of a weather insurance policy.

The third Chapter proposes a novel strategy to account for a hidden treatment diffusion process, where the intervention may spill over in the population according to linkages that are partially unknown. I address the issue of missing links by performing a machine learning algorithm that multiply imputes the presence of links starting from observed dyadic and individual characteristics. The characterization of the treatment spreading process is based on a set of simplifying assumptions which restrain the temporal dimension of the process to a three time step process, while limiting the treatment spreading to immediate neighbors

and assuming a fixed contagion probability among dyads. I formalize the bias that can occur while neglecting the treatment diffusion process and discuss the causal mechanisms that drive the bias to either direction. Finally, I perform a sensitivity analysis to assess whether the conclusions about the effect of an intervention are robust with respect to the (unknown) treatment diffusion process.

All three chapters show that ignoring the mechanism of interference in those scenarios when it can plausibly arise introduces a significant bias in the estimates, thus leading to inaccurate conclusions about the effect of an intervention. For instance, the first Chapter illustrates that allowing for multi-valued interference leads to weaker estimates about the effect of national immigration policies on crime rates. The second Chapter highlights the importance of assessing heterogeneity in the treatment and the spillover effects, so as to obtain powerful insights about the total effect of an intervention and to identify those sub-populations of units that are particularly responsive to the intervention as well as those that are able to spark a positive spillover mechanism. The third Chapter points out that the bias caused by neglecting a hidden treatment diffusion process can pave the way to either overestimation or underestimation of an intervention's effect, depending on the causal mechanisms that drive the spreading process and on the characteristics of the units that experienced it. Taken together, all these contributions underline the importance of taking into account spillover mechanisms in policy evaluation settings, for the sake of avoiding possible miss-interpretations about the real effect of an intervention and of exploring the role of interactions among agents in multiplying the consequences of a policy. When interference is likely to arise, the policy maker should make an effort into precisely characterizing interactions between units and conjecturing the possible mechanisms that may drive spillovers. Indeed, the methodologies that are proposed in this dissertation (and all the other existing strategies for estimating causal effects in presence of interconnected units), can be applied to a huge variety of empirical scenarios and can be exploited to further confirm (or to review) the validity of relevant causal statements put forward by prior research that ignored issues of in-

teraction. In addition, these methods allow to assess other causal effects that have never been studied yet.

The fourth Chapter stands out, as it addresses instead the issue of network formation. In so doing, it explores causal mechanisms in a setting where the structure of interactions among units in a population can be represented as a multiplex, where agents are connected by both a treatment network and an outcome network. This contribution is centered on an estimator for the causal effects that the existence of links in a treatment network has on the formation of links in an outcome network. I define conditional effects to investigate the symmetry and the direction of ties; the proposed estimator extends the established propensity score matching approach so as to handle multi-valued treatments, network data and conditional effects. I apply this framework to evaluate the causal effect of firms' lobbying activities on legislative collaborations among politicians: in this setting, the treatment network is defined by the existence of strong ties between legislators that are due to presence of shared supporting companies, while the outcome network consists in their co-sponsorship ties. The results show that the "treatment" strong ties of common pushing encourage politicians to collaborate and that co-sponsorship ties follow the direction of the strong ties. In this setting, and unlike those about interference, the network is the main object of interest, and interactions are more merely features to control for or to take into account while explaining the overall effect of an intervention. The contribution highlights the importance of drawing causal conclusions on network formation processes and showcases the empirical potential of the proposed methodology through a relevant application to the fields of economics and political science.

To conclude, the four contributions all underline the importance of dealing with interactions among units in policy evaluation settings. Even if they approach issues about units connected in a network from different perspectives, with distinct statistical objectives and by adopting different methodologies, they all show how policy evaluation can gain meaningful insights from handling issues of complexity in networked interactions, whether these are (i) something to account for while assessing the

effect of an intervention or (ii) the actual objective of the analysis.

6.2 Future Research

The interplay of causal inference and complex networks is a frontier topic in the statistical literature. Thus, there is plenty of related outstanding issues, both methodological and empirical, that are ripe for exploration by the scientific community in the next future.

There is ample space for mode methodological contributions, with respect to both the directions explored in this dissertation: the estimation of treatment and spillover effects in the presence of connected units and that of the causal effects on network formation. The extant statistical literature about spillovers has focused primarily on randomized settings (Aronow, 2012; Aronow & Samii, 2017; Athey et al., 2018b; Bowers et al., 2013; Forastiere et al., 2019a; Hudgens & Halloran, 2008; Imai et al., 2020; Kang & Imbens, 2016; Liu & Hudgens, 2014; Rosenbaum, 2007; VanderWeele et al., 2014), while there are way less contributions exploring spillover mechanisms in observational studies (Forastiere et al., 2020; Forastiere et al., 2018; Sofrygin & van der Laan, 2017; Tchetgen & VanderWeele, 2012; van der Laan, 2014). In both cases, the literature has largely focused on the analysis of first order spillover effects in clustered dependencies: only few recent contributions have exploited networks to represent the interference structure, and for the most part they only allow for first-order interactions in binary network settings. The investigation may be expanded to admit higher order spillover effects and to model interactions in more complex network structures, such as hierarchical networks or multi-layer networks. In particular, analyzing interference in multi-layer networks would contribute to expand the very recent stream of literature that models interference in a bipartite network (Zigler & Papadogeorgou, 2018), where units belong to two different sets: one containing units to which the treatment is assigned and one featuring those units displaying the outcome variable of interest.

Likewise, there is a similarly ample room for additional work about heterogeneity under interference. The second Chapter of this thesis rep-

resents the very first attempt to estimate heterogeneous treatment and spillover effects in presence of interference and to simultaneously deal with multiple effects for detecting the most heterogeneous sub-populations. This is the first step towards addressing a methodological issue which, in the upcoming years, is likely to raise an increasing interest in the statistical community. As it is common when taking the initial steps along a dark alley, the contribution featured in this dissertation leverages much upon simplifying assumptions that,in future research need to be overcome, so as to expand the ensemble of applications to which this methodology is applicable. I believe that issues to be taken into account, expanded and elaborated in future research are the following. First, the present work focuses on a particular type of interference, clustered network interference, which is aimed here at preserving the *honesty* of the framework (Athey & Imbens, 2016; Athey & Imbens, 2015); in future research however, one may want to extend the concept of interference to more general scenarios. Second, these contributions are based upon a binary categorization of both the individual treatment and the neighborhood treatment: in future work one could expand the feasible characterizations of the joint intervention that can be exploited by the algorithm. Third, this work models a randomized setting, while numerous applied contributions may be interested into exploring heterogeneity under spillovers in an observational setting. Fourth, the presented contributions assume perfect knowledge about clusters, which is unrealistic for many real-world applications; this framework can be meaningfully expanded by allowing for individual group memberships to be predicted via a data-driven approach (say, a community detection algorithm), while also accounting for the uncertainty in the prediction of groups. Fifth, and last, one contribution builds upon the causal tree and is based on the construction of a single tree: addressing heterogeneity and spillovers with more complex machine learning tools (such as random forests (Wager & Athey, 2018), Bayesian random forests (Hill, 2011) or Super Learner (Kreif & DiazOrdaz, 2019)) can be a worthwhile undertaking of future work. Finally, another methodological issue that is worth of future elaboration concerns imperfect compliance in randomized tri-

als: in presence of interactions between units, the imperfect compliance of agents might also be caused by the behavior and influence of those other units that interact with them.

With regard to the network formation framework, future research may expand upon the fourth Chapter of this dissertation by also examining (i) longitudinal, (ii) bipartite and (iii) multi-layer networks. The first case concerns networks observed at different points in time, such that, in each time interval, not only one can witness the establishment, persistence or breakup of ties between agents, but nodes themselves can also appear, disappear or reappear into and from the analysis. The second case is about two distinct set of nodes, which are mutually connected through inter-layers links, and that are also characterized by within-layer interactions; one can envision modeling the formation of within-set ties that are observed in the second set of nodes as a function of the inter-layers links and of the within-layer ties that are observed in the first set of nodes. The third case is one where agents are connected according to multiple networks (note that in the contribution from this dissertation there are just two such networks) and researchers aim to evaluate the causal effect of links in one layer on the formation of links in all the other layers..

On the empirical side, it is clear that the presented methodologies (regarding both the interference setting and the network formation setting) may be applied to a wide variety of empirical scenarios. As it is discussed at length in the introduction of the present dissertation, many real-world phenomena can be described in terms of interactions among different agents or objects. Consequently, applied researchers can utilize each of the methodologies that are exposed throughout this dissertation in order to address relevant research questions in various empirical fields, chiefly among them epidemiology, economics and political science.

Appendix A

Appendix

A.1 Proofs

A.1.1 Balancing property of JMGPS

We have to prove that

$$P(Z_i = z, \boldsymbol{G}_i = \boldsymbol{g}|\boldsymbol{X}_i) = P(Z_i = z, \boldsymbol{G}_i = \boldsymbol{g}|\psi(z, \boldsymbol{g}; \boldsymbol{X}_i)).$$

The expression on the leften side exactly equals JMGPS, by definition, that is

$$P(Z_i = z, \boldsymbol{G}_i = \boldsymbol{g}|\boldsymbol{X}_i) = \psi(z, \boldsymbol{g}; \boldsymbol{X}_i).$$

We focus now on the righten side. By iterated equation we have that

$$\begin{aligned} P(Z_i = z, \boldsymbol{G}_i = \boldsymbol{g}|\psi(z, \boldsymbol{g}; \boldsymbol{X}_i)) &= \mathbb{E}_{\boldsymbol{X}}\left[P(Z_i = z, \boldsymbol{G}_i = \boldsymbol{g}|\boldsymbol{X}_i, \psi(z, \boldsymbol{g}; \boldsymbol{X}_i))|\psi(z, \boldsymbol{g}; \boldsymbol{X}_i)\right] \\ &= \mathbb{E}_{\boldsymbol{X}}\left[P(Z_i = z, \boldsymbol{G}_i = \boldsymbol{g}|\boldsymbol{X}_i)|\psi(z, \boldsymbol{g}; \boldsymbol{X}_i)\right] \\ &= \mathbb{E}_{\boldsymbol{X}}\left[\psi(z, \boldsymbol{g}; \boldsymbol{X}_i)|\psi(z, \boldsymbol{g}; \boldsymbol{X}_i)\right] = \psi(z, \boldsymbol{g}; \boldsymbol{X}_i) \end{aligned}$$

The second equality holds as the joint multiple generalized propensity score, by definition, is functionally related to the characteristics $\boldsymbol{X}_i$. The third equality follows from the definition of JMGPS.

Both above expressions are equal to the joint multiple propensity score itself and, hence, they are also equal to each other.

A.1.2 Conditional unconfoundedness of $D_i(z)$ and G_i given JMGPS

We have to show that

$$P(D_i(z)=1,\boldsymbol{G}_i=\boldsymbol{g}|Y_i(z,\boldsymbol{g}),\psi(z,\boldsymbol{g};\boldsymbol{X}_i))=P(D_i(z)=1,\boldsymbol{G}_i=\boldsymbol{g}|\psi(z,\boldsymbol{g};\boldsymbol{X}_i)).$$

Righten side. We first focus on the expression that lies at the righten side. By the fact that $(D_i(z)=1)=(Z_i=z)$ and Proposition 1, we have

$$P(D_i(z)=1,\boldsymbol{G}_i=\boldsymbol{g}|\psi(z,\boldsymbol{g};\boldsymbol{X}_i))=P(Z_i=z,\boldsymbol{G}_i=\boldsymbol{g}|\psi(z,\boldsymbol{g};\boldsymbol{X}_i))=\psi(z,\boldsymbol{g};\boldsymbol{X}_i).$$

Leften side. By iterated equations, we have

$$\begin{aligned}
&P(D_i(z)=1,\boldsymbol{G}_i=\boldsymbol{g}|Y_i(z,\boldsymbol{g}),\psi(z,\boldsymbol{g};\boldsymbol{X}_i))\\
&=\mathbb{E}_{\boldsymbol{X}}\big[P\big(D_i(z)=1,\boldsymbol{G}_i=\boldsymbol{g}|\boldsymbol{X}_i,\psi(z,\boldsymbol{g};\boldsymbol{X}_i),Y_i(z,\boldsymbol{g})\big)|Y_i(z,\boldsymbol{g}),\psi(z,\boldsymbol{g};\boldsymbol{X}_i)\big]\\
&=\mathbb{E}_{\boldsymbol{X}}\big[P\big(D_i(z)=1,\boldsymbol{G}_i=\boldsymbol{g}|Y_i(z,\boldsymbol{g}),\boldsymbol{X}_i\big)|Y_i(z,\boldsymbol{g}),\psi(z,\boldsymbol{g};\boldsymbol{X}_i)\big]\\
&=\mathbb{E}_{\boldsymbol{X}}\big[P\big(D_i(z)=1,\boldsymbol{G}_i=\boldsymbol{g}|\boldsymbol{X}_i\big)|Y_i(z,\boldsymbol{g}),\psi(z,\boldsymbol{g};\boldsymbol{X}_i)\big]\\
&=\mathbb{E}_{\boldsymbol{X}}\big[\psi(z,\boldsymbol{g};\boldsymbol{X}_i)|Y_i(z,\boldsymbol{g}),\psi(z,\boldsymbol{g};\boldsymbol{X}_i)\big]=\psi(z,\boldsymbol{g};\boldsymbol{X}_i),
\end{aligned}$$

where the second equality is obtained taking into account that the joint multiple generalized propensity score is a function of covariates, the third equality results from applying the Assumption 7, while the forth equality holds recalling that $(D_i(z)=1)=(Z_i=z)$ and Definition 3.

A.1.3 Conditional unconfoundedness of $D_i(z)$ and G_i given individual and neighborhood propensity scores

We have to show that

$$P(D_i(z)=1,\boldsymbol{G}_i=\boldsymbol{g}|Y_i(z,\boldsymbol{g}),\phi(z;\boldsymbol{X}_i^z),\lambda(\boldsymbol{g};z,\boldsymbol{X}_i^g))=P(D_i(z)=1,\boldsymbol{G}_i=\boldsymbol{g}|\phi(z;\boldsymbol{X}_i^z)$$

where $\phi(z;\boldsymbol{X}_i^z)=P(D_i(z)=1|\boldsymbol{X}_i^z)$ and $\lambda(\boldsymbol{g};z,\boldsymbol{X}_i^g)=P(\boldsymbol{G}_i=\boldsymbol{g}|Z_i=z,\boldsymbol{X}_i^g)$. We proceed showing that both sides of the equation are equal to the joint multiple generalized propensity score.

Righten side. By iterated equations, we have

$$
\begin{aligned}
&P(D_i(z)=1, \boldsymbol{G}_i=\boldsymbol{g}|\phi(z;\boldsymbol{X}_i^z), \lambda(\boldsymbol{g};z,\boldsymbol{X}_i^g)) \\
&= \mathbb{E}_{\boldsymbol{X}}\left[P(D_i(z)=1, \boldsymbol{G}_i=\boldsymbol{g}|\boldsymbol{X}_i, \phi(z;\boldsymbol{X}_i^z), \lambda(\boldsymbol{g};z,\boldsymbol{X}_i^g))|\phi(z;\boldsymbol{X}_i^z), \lambda(\boldsymbol{g};z,\boldsymbol{X}_i^g)\right] \\
&= \mathbb{E}_{\boldsymbol{X}}\left[P(D_i(z)=1, \boldsymbol{G}_i=\boldsymbol{g}|\boldsymbol{X}_i)|\phi(z;\boldsymbol{X}_i^z), \lambda(\boldsymbol{g};z,\boldsymbol{X}_i^g)\right] \\
&= \mathbb{E}_{\boldsymbol{X}}\left[\psi(z,\boldsymbol{g};\boldsymbol{X}_i)|\phi(z;\boldsymbol{X}_i^z), \lambda(\boldsymbol{g};z,\boldsymbol{X}_i^g)\right] = \psi(z,\boldsymbol{g};\boldsymbol{X}_i).
\end{aligned}
$$

The above equalities result from the fact that both $\phi(z;\boldsymbol{X}_i^z)$ and $\lambda(\boldsymbol{g};z,\boldsymbol{X}_i^g)$ are function of $\boldsymbol{X}_i$ (second equality) and from the factorization $\psi(z,\boldsymbol{g};\boldsymbol{X}_i) = \phi(z;\boldsymbol{X}_i^z)\lambda(\boldsymbol{g};z,\boldsymbol{X}_i^g)$ (third equality).

Leften side. By iterated equations, we have

$$
\begin{aligned}
&P(D_i(z)=1, \boldsymbol{G}_i=\boldsymbol{g}|Y_i(z,\boldsymbol{g}), \phi(z;\boldsymbol{X}_i^z), \lambda(\boldsymbol{g};z,\boldsymbol{X}_i^g)) \\
&= \mathbb{E}_{\boldsymbol{X}}\left[P(D_i(z)=1, \boldsymbol{G}_i=\boldsymbol{g}|\boldsymbol{X}_i, Y_i(z,\boldsymbol{g}), \phi(z;\boldsymbol{X}_i^z), \lambda(\boldsymbol{g};z,\boldsymbol{X}_i^g))|Y_i(z,\boldsymbol{g}), \phi(z;\boldsymbol{X}_i^z), \lambda(\boldsymbol{g}\right. \\
&= \mathbb{E}_{\boldsymbol{X}}\left[P(D_i(z)=1, \boldsymbol{G}_i=\boldsymbol{g}|\boldsymbol{X}_i, Y_i(z,\boldsymbol{g})|Y_i(z,\boldsymbol{g}), \phi(z;\boldsymbol{X}_i^z), \lambda(\boldsymbol{g};z,\boldsymbol{X}_i^g)\right] \\
&= \mathbb{E}_{\boldsymbol{X}}\left[P(D_i(z)=1, \boldsymbol{G}_i=\boldsymbol{g}|\boldsymbol{X}_i)|Y_i(z,\boldsymbol{g}), \phi(z;\boldsymbol{X}_i^z), \lambda(\boldsymbol{g};z,\boldsymbol{X}_i^g)\right] \\
&= \mathbb{E}_{\boldsymbol{X}}\left[\psi(z,\boldsymbol{g};\boldsymbol{X}_i)|Y_i(z,\boldsymbol{g}), \phi(z;\boldsymbol{X}_i^z), \lambda(\boldsymbol{g};z,\boldsymbol{X}_i^g)\right] \\
&= \psi(z,\boldsymbol{g};\boldsymbol{X}_i),
\end{aligned}
$$

where the second equality results from the fact that the two propensity scores are function of the covariates, the third equality comes from Assumption 7, the fourth equality is obtained recalling Definition 3 and that $(D_i(z)=1) = (Z_i = z)$ and, finally, the last equality follows from the factorization of the JMGPS.

A.2 Influence Index detailed construction

The Influence index (I) is formally defined as

$$I_{cc',t} = \alpha \times IG_{cc'} + \beta \times IC_{cc',t}$$

where IG denotes the geographical proximity indicator while IC states for the cultural similarity indicator. The former is time invariant, while

the latter provides a temporal variation. Here, we discuss the detailed construction of these two indexes, which determine the interference structure.

We build up the geographical proximity index taking into account of two variables: a boundaries-related variable and a geographical distance-related variable. The former, that we denote by Sp counts the minimum number of states one needs to cross by, at the aim of reaching country c' starting from country c. Thus, if we consider a graph collecting all the national states, this variable represents the length of the shortest path between c and c' [1]. The latter, that we denote as $Dist^{std}$ is a standardized measure of geographic distance between the most populated cities belonging to the two countries. Formally, the geographical proximity indicator is computed as follows

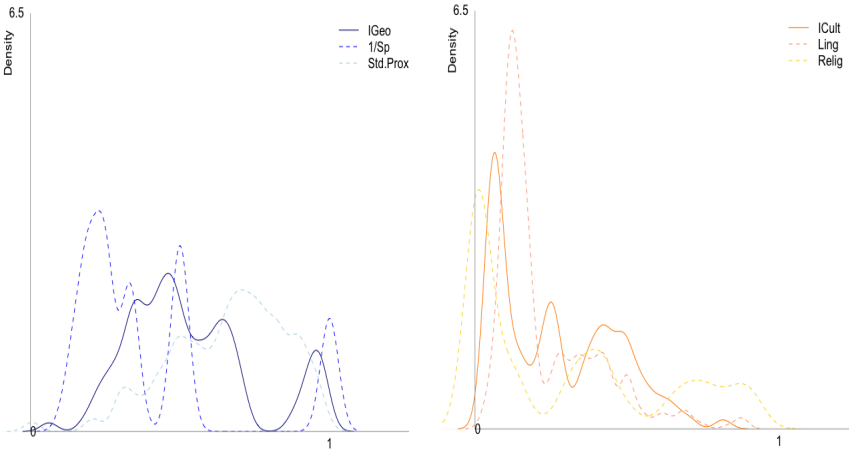
$$IG_{cc'} = 0.5 \times \frac{1}{\text{Sp}_{cc'}} + 0.5 \times (1 - \text{Dist}^{std}_{cc'}) = 0.5 \times \frac{1}{\text{Sp}_{cc'}} + 0.5 \times \text{Prox}^{std}_{cc'}$$

On the other side, the cultural similarity indicator measures the level of cultural similarity between two countries c and c' at a given time t. We summarize this aspect evaluating the linguistic similarity and the religious similarity, through the variables $Ling$ and $Relig$. These measures have been defined by the CEPII Linguistic Dataset Melitz and Toubal, 2014 and CEPII Gravidata Dataset (Fouquin, Hugot, et al., 2016), respectively. The *linguistic proximity indicator* gives a unique measure of how much the whole linguistic systems differ in the two countries, both in terms of the distribution of spoken languages over the population and in terms of the linguistic roots. The *religious similarity indicator* takes into account of the distribution of practised religions: an high value of this variable signals an high similarity in terms of prevalence of the various religious communities at time t.

$$IC_{cc',t} = 0.5 \times \text{Ling}_{cc',t} + 0.5 \times \text{Relig}_{cc',t}$$

[1] We assume that the pairs of countries France and Great Britain, Ireland and Great Britain share a common boundary, as, even if they're formally separated by the English Channel and the Irish Sea, respectively, they are very near and connections are extremely simple

Figure 55 shows the density distributions of the two indicators that contribute in determining the Influence Index, as well as of their respective sub-components.



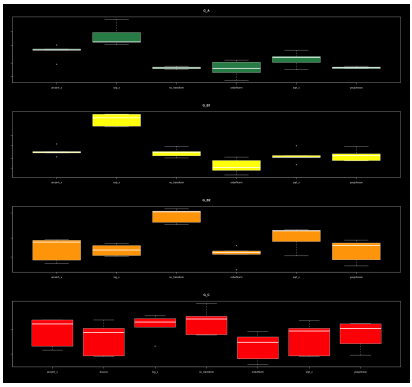
(a) Geographical proximity Indicator (b) Cultural similarity Indicator

Figure 55: Influence index components density distributions

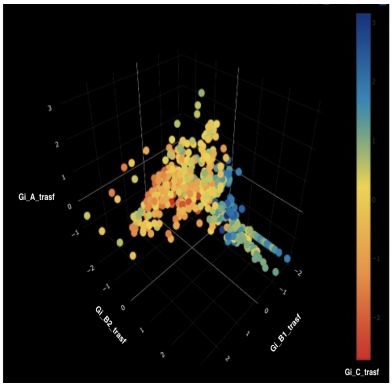
A.3 Transformation of the NTEM components

We run some checks about the normality of the components $G_{i,z}$. The Shapiro Tests for Normality (Shapiro and Wilk, 1965), separately conducted on the four components, $G_{i,LL}$, $G_{i,HL}$, $G_{i,LH}$, $G_{i,HH}$, rejects the Normality null-Hypothesis.
Hence, we decide to apply a transformation to each of the $G_{i,z}$. We conduct some tests experimenting various transformation methods and we compare them selecting the best approach according to the Pearson P statistic for Normality (divided by its degrees of freedom). We use repeated cross validation to estimate the out-of-sample performance of all these methods. Figure A.56(a) shows the box plots of the out of sample estimated normality statistics for all the techniques that we experiment,

over the four variables of interest (under $\alpha = \frac{1}{2}$ and $\beta = \frac{1}{2}$). We find out that the method that performed better in handling the $G_{i,z}$ variables is the *Ordered Quantile (ORQ) transformation*, for all the various configurations of the II input weights. Figure A.56(b) represents the tridimensional scatterplot of transformed variables.



(a) Best Normalizing methods: comparison among different methods, through box plots



(b) Tridimensional plot of transformed variables

Figure 56: Best Normalizing method

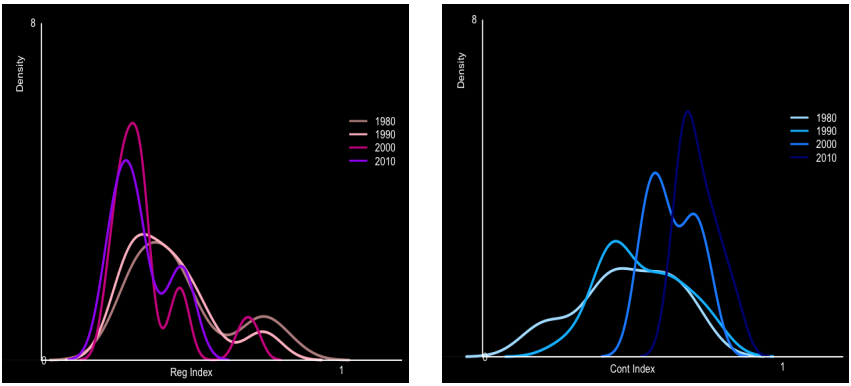
Ordered Quantile trasformation (Bartlett, 1947,Van der Waerden, 1952)is based on ranks. Essentially, the values of a variable, judged as a vector, are mapped to their percentile, and then to the same percentile of the Standard Normal Distribution. As long as the number of ties is negligible, this method guarantees that the transformed variable follows a Normal Distribution. Formally, each variable $G_{i,z}$ is transformed according to the following formula:

$$G^*_{i,z} = \Phi^{-1}\left(\frac{\text{rank}(G_{iz})}{N+1}\right),$$

where Φ is the cumulative density function of a Standard Normal distribution, N is the number of observations. We denote as $G^*_{i,z}$ the variable resulting from the Ordered Quantile transformation.

A.4 Descriptives

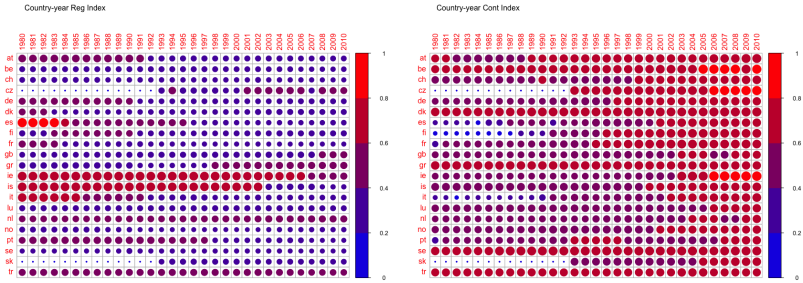
This paragraph provides some descriptives. Figure 57 shows the density distributions of the indicators measuring the restrictiveness of *regulations* (*Reg*) and *control strategies* (*Cont*), over years. Regulations have become more welcoming over time while control strategies have turned to a more severe attitude.



(a) Regulations (b) Control

Figure 57: Indicators measuring the restrictiveness of immigration policies over years: the countries' attitude towards migrants with respect to *regulations* and *control* mechanisms observed at four points in time (over the 30 years time frame covered by data - 1980-2010 -)

Figure 58 consents to inspect the strictness of regulations and control implemented policies in each country-year profile.



(a) Regulations (b) Control

Figure 58: Indicators measuring the restrictiveness of immigration policies over country year proriles: *regulations* and *control* protocols

Figure 59 represents the variation of the distributions of the *Reg* (violet), *Control* (blue) and *ImPol* (yellow) indicators in the 22 countries that we have included in the analysis.

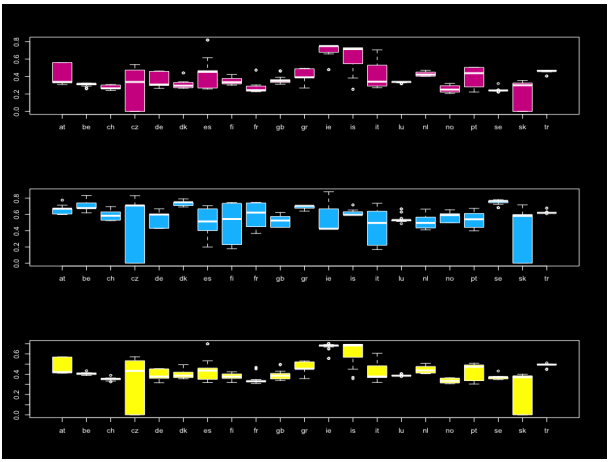


Figure 59: Countries' attitude towards migrants in terms of i) *regulations* (violet), ii) *control*(blue) protocols and iii) both components (yellow).

Figure 60 shows the tridimensional plot of the two indicators measur-

ing the restrictiveness towards migrants, and the corresponding crime rate.

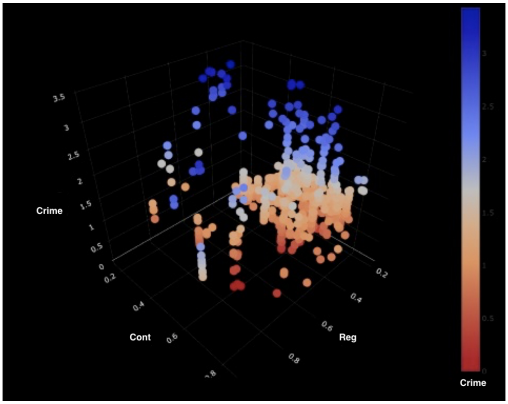


Figure 60: Tridimensional plot of the two indicators *Reg* and *Cont*, with respect to the *crime rate*.

Table 14 shows the basic descriptives of all the variables we have included in the analysis.

Table 14: Descriptive statistics

Variable	Mean	St. dev	Min	Pctl(25)	Pctl(75)	Max
Crime rate (every 10.000 inhab.)	1.326	0.636	0.000	0.910	1.560	3.430
Fertility rate	1.675	0.498	0.000	1.440	1.840	4.360
Power distributed to gender Index	1.876	0.973	−0.854	1.408	2.394	3.714
Health equality Index	2.418	0.588	0.612	1.972	2.775	3.792
Educational inequality Gini index (/60)	0.269	0.163	0.000	0.149	0.361	0.893
Income inequality Gini index (/60)	0.485	0.139	0.000	0.432	0.564	0.867
Equal access index	0.871	0.137	0.290	0.856	0.945	0.986
Equal distribution of resources index	0.924	0.070	0.588	0.908	0.964	0.986
Civil partecipation index	0.641	0.111	0.161	0.616	0.690	0.885
Access to justice index	0.940	0.121	0.165	0.944	0.989	0.995
State ownership of economy index	1.219	0.616	−0.536	0.890	1.636	2.731
Freedom of expression index	0.934	0.128	0.128	0.955	0.979	0.991
Freedom of religion index	1.958	0.751	−1.003	1.749	2.519	2.766
Life expectancy (/ 100)	0.746	0.126	0.000	0.749	0.785	0.824
GDP per capita (/ 10.000)	2.666	1.154	0.610	1.954	3.373	8.192

A.5 Results under different configurations of the treatment

This section shows results under alternative specifications of the treatment variable of interest. In particular, as Definition 7 clarifies, we test two secondary ways of detecting the treatment categories.

Definition 7. *Alternative specifications of the treatment variable Let us indicate Z_i^K a generic treatment variable defined over K categories. We consider the following treatment classifications*

1. *Multi-valued treatment with three categories, $Z_i^{(3)}$, which have been defined collapsing the categories HL and LH of the original individual treatment variable.*

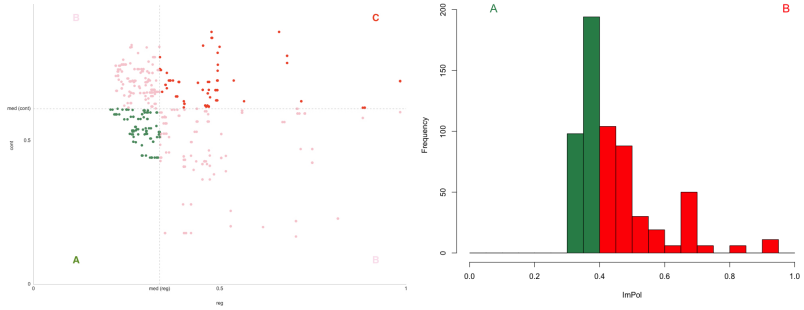
   - *Z_i^3=L if $reg_i \leq med_{reg}$ and $cont_i \leq med_{cont}$: this category identifies profiles that are barely restrictive with respect to the two mechanisms.*
   - *Z_i^3=H if $reg_i \geq med_{reg}$ and $cont_i \geq med_{cont}$: this category denotes an highly restrictive policy towards migrants with respect to both regulations and control.*
   - *Z_i^3=M otherwise* [2]

2. *Binary treatment with two categories, $Z_i^{(2)}$ defined as follows*

   - *Z_i^2=L if $impol_i \leq med_{ImPol}$*
   - *Z_i^2=H if $impol_i > med_{ImPol}$*

Figure 61 graphically represents these two alternative treatment characterizations. Table 15 shows results under these two definitions of the treatment variable. As it is immediate to observe, these results are robust with the main findings of this paper.

[2] note that the A and C categories exactly coincide with the A and C categories of the four-valued treatment



(a) Multi-valued treatment with three categories, $Z_i^{(3)}$

(b) Binary treatment, $Z_i^{(2)}$

Figure 61: Alternative definitions of the treatment variable

Table 15: Results under alternative definitions of the treatment variable

	Treatment categories		
	3 (L,M,H)		2 (L,H)
	Effects of Interest		Effects of Interest
IIW	M-L	H-L	H-L
(α, β)			
$(\frac{1}{2}, \frac{1}{2})$	0.06648 ***	0.04986 ***	0.03875***
	(0.06485;0.06815)	(0.04774;0.05196)	(0.01424;0.06133)
$(1, 0)$	0.04363 ***	0.01781 ***	0.04126 ***
	(0.04203;0.04527)	(0.01573;0.01987)	(0.01686;0.06374)
$(0, 1)$	0.09282 ***	0.03523 ***	0.03587 ***
	(0.09228;0.09338)	(0.03452;0.03592)	(0.01288;0.05705)
$(0, 0)$	0.0727 ***	'0.0008	0.03506***
	(0.07027;0.07524)	(-0.00386;0.00303)	(0.01443;0.05789)

Taking into consideration of a binary treatment (which we obtain simply differentiating the country-year profiles whose observed value of the general immigration policies indicator is above its reference median) still leads to positive results, regardless of the assumption about interference. But in the last two scenarios, effects are significantly weaker.

A.6 Models Results

A.6.1 Model for Z

Table 16: Model for the individual treatment Z_i: multinomial logit

	Dependent variable:		
	Z_i		
	HL	LH	HH
(Intercept)	32.79243***(12.78648)	-10.11091(15.25728)	80.91698(15.50048)
rate	0.16626(0.26154)	0.05287(0.2613)	0.69135(0.43628)
ferrate	2.58746***(0.96028)	-1.67848(1.06485)	4.32561***(1.12471)
powgend	-3.09196***(0.49015)	0.714(0.41834)	-0.54734(0.51338)
eq_health	0.10112(0.69408)	2.75906***(0.78272)	4.62427***(1.1305)
ineq_educ	-2.09696(2.33622)	5.29634***(2.39962)	-10.3432***(2.70765)
ineq_inc	5.32876***(2.59477)	-9.08242***(2.35931)	10.89922***(3.40414)
eq_access	28.56299***(5.5844)	-5.09848(6.35618)	31.06497***(7.08197)
eq_redist	-40.91293***(12.89674)	-38.18629***(16.62724)	-114.90727***(19.31919)
civilpart	-9.59618***(2.78319)	5.08124***(2.47201)	-6.36655(4.05762)
accjust	-1.89435(9.22897)	-5.28742(10.20244)	-75.66229***(12.17123)
ecocont	0.42103(0.29668)	-0.02385(0.3009)	-0.00473(0.41129)
freexp	-15.37981**(7.11022)	45.12384***(11.12141)	46.82842***(10.41181)
freerelig	1.26604***(0.57401)	0.68092(0.62653)	-0.2062(0.74503)
lifexp	2.02242(2.7095)	6.54206***(2.64635)	12.97523***(5.51876)
gdppc	-0.8286***(0.21716)	-0.14708(0.17659)	-0.15251(0.2456)
Note:			*p<0.1; **p<0.05; ***p<0.01

A.6.2 Model for G

Table 17: Models for the neighborhood treatment $\boldsymbol{G}_i$: multivariate linear model. Results are obtained while choosing different configurations of the interference weights

	Dependent variable:		
	$\boldsymbol{G}^*$		
	Statistic	Statistic	Statistic
Omnibus Effect	35.31***	38.27***	41.42***
(Intercept)	21.63 ***	28.17***	9.79***
Z_i	12.09 ***	14.62***	9.81***
rate	2.71 **	4.80 **	0.91
ferrate	18.01 ***	20.04 ***	14.95 ***
powgend	3.76 ***	6.22 ***	3.44 ***
eq_health	6.61 ***	12.35***	2.34 ***
ineq_educ	0.38	0.07	5.43***
ineq_inc	11.17 ***	11.28 ***	8.85***
eq_access	7.81***	12.64***	3.32 ***
eq_resdist	15.75 ***	24.51***	4.97***
civilpart	16.13 ***	18.06***	14.58 ***
accjust	20.43 ***	21.78 ***	12.37
econcont	16.66 ***	16.39 ***	20.85 ***
freeexp	15.85 ***	16.85 ***	9.86 ***
freerelig	14.43 ***	13.58 ***	18.86 ***
lifeexp	1.01	0.64	4.51 ***
gdppc	191.40 ***	196.56***	180.59 ***
Vertex_centr	72.79 ***	81.82 ***	124.79 ***
IIW			
α	1/2	1	0
β	1/2	0	1
Note:	*p<0.1; **p<0.05; ***p<0.01		

A.6.3 Models for Y

Table 18: Models for Y: linear model with time fixed effects. Results are obtained while choosing different configurations of the interference weights

	Dependent variable:		
	Y		
$Z_{i,HL}$	0.18591**(0.08543)	0.04817(0.09693)	0.1866**(0.07333)
$Z_{i,LH}$	0.25183**(0.08896)	0.20069**(0.08139)	0.26622***(0.08028)
$Z_{i,HH}$	0.20918*(0.11826)	0.04082(0.08269)	0.19946*(0.11675)
$G^*_{i,LL}$	-0.44931(0.32531)	0.46268(0.32837)	-0.92524***(0.28568)
$G^*_{i,HL}$	-0.26138(0.35755)	-1.4981**(0.58176)	-0.00591(0.22521)
$G^*_{i,LH}$	0.9401**(0.47619)	1.80662***(0.53027)	0.33123(0.3205)
$G^*_{i,HH}$	1.46947***(0.36545)	0.55599(0.36047)	1.05639***(0.3828)
$\phi(z_i; \boldsymbol{X}_i^z)$	-0.07777(0.06924)	-0.14616**(0.07232)	-0.08396(0.0706)
$\lambda(\boldsymbol{g_i}; z_i, \boldsymbol{X}_i^g)$	0.33052(**0.16081)	0.36958**(0.15836)	0.05239(0.14535)
II w			
α	1/2	1	0
β	1/2	0	1
Note:			*p<0.1; **p<0.05; ***p<0.01

Appendix B

Appendix

B.1 Proofs

Proposition 4 states that

$$E[\widehat{\mu}_{(z,g)}(\ell(\mathbf{x}))] = \mu_{(z,g)}(\ell(\mathbf{x}))$$

Proof.

$$\begin{aligned}\mathbb{E}[\widehat{\mu}_{(z,g)}] =& \mathbb{E}\left[\frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\mathbb{1}(Z_{ik}=z, G_{ik}=g,)\frac{Y_{ik}}{\pi_{ik}((z,g))}\right] \\ =& \frac{1}{N(\ell(\mathbf{x}))}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\mathbb{E}\big[\mathbb{1}(Z_{ik}=z, G_{ik}=g)\big]\frac{Y_{ik}(z,g)}{\pi_{ik}(z,g)} \\ =& \mu_{(z,g)}\end{aligned} \tag{B.1}$$

where the expectation is over the randomization distribution of Z_{ik} and the induced distribution on G_{ik} and the first equality holds by consistency. □

Proposition 9 (Population Unbiaseness)**.** *The estimator is unbiased with respect to the population mean of the potential outcomes:*

$$E[\widehat{\mu}_{(z,g)}(\ell(\mathbf{x}))] = E[Y_{ik}(z,g)|\mathbf{X}_{ik} \in \ell(\mathbf{x}))] = \mu^{P}_{(z,g)}(\ell(\mathbf{x}))$$

and

$$E[\widehat{\tau}_{(z,g,z',g')}(\ell(\mathbf{x}))] = E[Y_{ik}(z,g) - Y_{ik}(z',g')|\mathbf{X}_{ik} \in \ell(\mathbf{x}))] \\ = E[\tau_{(z,g,z',g')}(\mathbf{X}_{ik})|\mathbf{X}_{ik} \in \ell(\mathbf{x}))] = \mu^P_{(z,g)}(\ell(\mathbf{x}))$$

where the expected value is taken over the sampling distribution.

Proof.

$$\begin{aligned}
\mathbb{E}[\widehat{\mu}_{(z,g)}] =& \mathbb{E}\left[\frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\mathbb{1}(Z_{ik}=z, G_{ik}=g,)\frac{Y_{ik}}{\pi_{ik}((z,g))}\right] \\
=& \frac{1}{N(\ell(\mathbf{x}))}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\frac{\mathbb{E}\big[\mathbb{1}(Z_{ik}=z, G_{ik}=g)Y_{ik}\big]}{\pi_{ik}(z,g)} \\
=& \frac{1}{N(\ell(\mathbf{x}))}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\frac{\mathbb{E}\big[Y_{ik}|Z_{ik}=z, G_{ik}=g\big]\pi_{ik}(z,g)}{\pi_{ik}(z,g)} \qquad \text{(B.2)} \\
=& \frac{1}{N(\ell(\mathbf{x}))}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\mathbb{E}\big[Y_{ik}(z,g)\big] = \mathbb{E}\big[Y_{ik}(z,g)\big] \\
=& \mu^P_{(z,g)}
\end{aligned}$$

The proof of the unbiaseness of $\widehat{\tau}_{(z,g,z',g')}$ follows directly. □

Recall Proposition 7
Consider the asymptotic regime where the number of clusters K go to infinity, i.e., $K \longrightarrow \infty$, while the cluster size remains bounded, i.e., $n_k \leq B$ for some constant B. In addition, assume that $|Y_{ik}(z,g)|/\pi_{ik}(z,g) \leq C < 1$, $\forall i,k,z,g$. Then as $K \longrightarrow \infty$

$$\widehat{\tau}_{(z,g;z',g')}(\ell(\mathbf{x})) \xrightarrow{p} \tau_{(z,g;z',g')}(\ell(\mathbf{x})).$$

Proof. As in Proposition 4 $\widehat{\mu}_{(z,g))}(\ell(\mathbf{x}))$ is unbiased. Hence, for consistency to hold we need to prove that the variance goes to 0 as N goes to infinity. Following Aronow and Samii, 2017, it is easy to show that the

variance of $\widehat{\mu}_{(z,g))}(\ell(\mathbf{x}))$ is given by:

$$\begin{aligned}\mathbb{V}\Big(\widehat{\mu}_{(z,g)}(\ell(\mathbf{x}))\Big) =& \frac{1}{N(\ell(\mathbf{x}))^2}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\mathbb{1}(\mathbf{X}_{ik}\in\ell(\mathbf{x}))\pi_{ik}(z,g)[1-\pi_{ik}(z,g)]\left[\frac{Y_{ik}(z,g)}{\pi_{ik}(z,g)}\right]^2 \\ &+ \frac{1}{N(\ell(\mathbf{x}))^2}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\sum_{j\neq i}\mathbb{1}(\mathbf{X}_{ik}\in\ell(\mathbf{x}),\mathbf{X}_{jk}\in\ell(\mathbf{x})) \\ &\times[\pi_{ikjk}(z,g)-\pi_{ik}(z,g)\pi_{jk}(w,z)])\frac{Y_{ik}(z,g)}{\pi_{ik}(z,g)}\frac{Y_{jk}(z,g)}{\pi_{jk}(z,g)} \quad \text{(B.3)}\end{aligned}$$

Since $|Y_{ik}(z,g)|/\pi_{ik}(z,g) \leq C < 1$ and given that in each cluster the sample size belonging to leaf $\ell(\mathbf{x})$ is bounded, i.e., $n_k(\ell(\mathbf{x})) \leq B(\ell(\mathbf{x})) \leq B$, we have

$$(K\times B)^2\mathbb{V}\Big(\widehat{\mu}_{(z,g)}(\ell(\mathbf{x}))\Big) \leq\ C^2\times K\times B + C^2\times K\times B^2$$

Consistency of $\widehat{\mu}_{(z,g)}(\ell(\mathbf{x}))$ is therefore ensured since $\mathbb{V}\Big(\widehat{\mu}_{(z,g)}(\ell(\mathbf{x}))\Big) \longrightarrow 0$ as $K \longrightarrow \infty$. Consistency of $\widehat{\tau}_{(z,g;z',g')}(\ell(\mathbf{x}))$ follows by Slutsky's Theorem. □

Proposition 10. *The partition* $\Pi^\star$ *such that*

$$\Pi^\star = \textit{argmax}_{\Pi\in\mathbb{P}}Q_{(z,g,z',g')}(\Pi) = \frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\left(\widehat{\tau}_{(z,g,z',g')}(\ell(\mathbf{X}_{ik},\Pi))\right)^2$$

maximizes the heterogeneity across leaves.

Proof. Let ℓ_1 and ℓ be two sub-populations with a different causal effect $\tau_{(z,g,z',g')}$, i.e., $\tau_{(z,g,z',g')}(\ell_1) \neq \tau_{(z,g,z',g')}(\ell_2)$. Let Π be a partition that splits ℓ_1 and ℓ_2 into two leaves and let Π^c be the partition that combines the two sub-populations into one leaf ℓ_{1+2}. Then we have that $Q_{(z,g,z',g')}(\Pi) > Q_{(z,g,z',g')}(\Pi^c)$. The proofs follows from Jensen's inequality. In fact, for partition Π the splitting function can be written as follows:

$$\begin{aligned}Q_{(z,g,z',g')}(\Pi) &= \frac{1}{|\ell_1|+|\ell_2|}\sum_{ik\in\ell_1\cup\ell_2}\left(\widehat{\tau}_{(z,g,z',g')}(\ell(\mathbf{X}_{ik},\Pi))\right)^2 \\ &= \frac{1}{|\ell_1|+|\ell_2|}\Big(\sum_{ik\in\ell_1}\left(\widehat{\tau}_{(z,g,z',g')}(\ell_1)\right)^2 + \sum_{ik\in\ell_2}\left(\widehat{\tau}_{(z,g,z',g')}(\ell_2)\right)^2\Big)\end{aligned}$$

For partition Π^c we have:

$$
\begin{aligned}
Q_{(z,g,z',g')}(\Pi^c) &= \frac{1}{|\ell_1|+|\ell_2|}\sum_{ik\in\ell_1\cup\ell_2}\big(\widehat{\tau}_{(z,g,z',g')}(\ell(\mathbf{X}_{ik},\Pi^c))\big)^2 \qquad \text{(B.4)}\\
&= \frac{1}{|\ell_1|+|\ell_2|}\sum_{ik\in\ell_1\cup\ell_2}\big(\widehat{\tau}_{(z,g,z',g')}(\ell_{1+2})\big)^2\\
&= \frac{1}{|\ell_1|+|\ell_2|}\sum_{ik\in\ell_1\cup\ell_2}\Big[\frac{1}{|\ell_1|+|\ell_2|}\Big(\sum_{ik\in\ell_1}\widehat{\tau}_{(z,g,z',g')}(\ell_1)+\sum_{ik\in\ell_2}\widehat{\tau}_{(z,g,z',g')}(\ell_2)\Big)\Big]^2\\
&= \Big[\frac{1}{|\ell_1|+|\ell_2|}\Big(\sum_{ik\in\ell_1}\widehat{\tau}_{(z,g,z',g')}(\ell_1)+\sum_{ik\in\ell_2}\widehat{\tau}_{(z,g,z',g')}(\ell_2)\Big)\Big]^2
\end{aligned}
$$

where the second-last equality holds because of the properties of the Horvitz-Thomson estimator. Thanks to Jensen's inequality

$$
\begin{aligned}
&\frac{1}{|\ell_1|+|\ell_2|}\Big(\sum_{ik\in\ell_1}\big(\widehat{\tau}_{(z,g,z',g')}(\ell_1)\big)^2+\sum_{ik\in\ell_2}\big(\widehat{\tau}_{(z,g,z',g')}(\ell_2)\big)^2\Big)\geq\\
&\qquad\Big[\frac{1}{|\ell_1|+|\ell_2|}\Big(\sum_{ik\in\ell_1}\widehat{\tau}_{(z,g,z',g')}(\ell_1)+\sum_{ik\in\ell_2}\widehat{\tau}_{(z,g,z',g')}(\ell_2)\Big)\Big]^2
\end{aligned}
$$

Hence, $Q_{(z,g,z',g')}(\Pi)\geq Q_{(z,g,z',g')}(\Pi^c)$. □

B.2 Further Details of the Variance Estimator of Leaf-Specific CACE

If in the examined leaf $\ell(\mathbf{x})$ there are some pairs of units (i,j) whose joint probability of the exposure condition (z,g) is zero, that is, $\pi_{ikjk}(z,g)=0$, the variance of $\widehat{\mu_{(z,g)}(\ell(\mathbf{x}))}$ can be estimated following a result from Aronow and Samii, 2017. Such estimator, denoted by $\widehat{\mathbb{V}}^c\Big(\widehat{\mu}_{(z,g)}(\ell(\mathbf{x}))\Big)$, is the sum of two components: (i) the estimated variance of leaf-specific potential outcomes, $\widehat{\mathbb{V}}\Big(\widehat{\mu}_{(z,g)}(\ell(\mathbf{x})\Big)$ in (3.3.2) for the case when $\pi_{ikjk}(z,g)>0\ \forall i,j,k$, and (ii) a correction term $\widehat{A}_{(z,g)}(\ell(\mathbf{x}))$:

$$
\widehat{\mathbb{V}}^c\Big(\widehat{\mu}_{(z,g)}(\ell(\mathbf{x}))\Big)=\widehat{\mathbb{V}}\Big(\widehat{\mu}_{(z,g)}(\ell(\mathbf{x}))\Big)+\widehat{A}_{(z,g)}(\ell(\mathbf{x})). \qquad \text{(B.5)}
$$

where

$$\widehat{A}_{(z,g)}(\ell(\mathbf{x}))\frac{1}{N(\ell(\mathbf{x}))^2}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\sum_{j\neq i:\pi_{ikjk}(z,g)=0}\left[\frac{\mathbb{1}(W_{ik}=z,G_{ik}=g,\mathbf{X}_{ik}\in\ell(\mathbf{x}))Y_{ik}^2}{2\pi_{ik}(z,g)}\right.$$
$$\left.\frac{\mathbb{1}(W_{jk}=z,G_{jk}=g,\mathbf{X}_{ik}\in\ell(\mathbf{x}))Y_{jk}^2}{2\pi_{jk}(z,g)}\right].$$

Note that the correction term $\widehat{A}_{(z,g)}(\ell(\mathbf{x}))$ is zero if the leaf does not have pairs of units such that $\pi_{ikjk}(z,g,z,g)=0$. Furthermore, as in Aronow and Samii, 2017, $\widehat{\mathbb{V}}^c\Big(\widehat{\mu}_{(z,g)}(\ell(\mathbf{x}))\Big)$ is a conservative estimator of the leaf-specific variance, as the following holds:

$$\mathbb{E}\left[\widehat{\mathbb{V}}^c\Big(\widehat{\mu}_{(z,g)}(\ell(\mathbf{x}))\Big)\right]=\mathbb{E}\left[\widehat{\mathbb{V}}\Big(\widehat{\mu}_{(z,g)}(\ell(\mathbf{x}))\Big)+\widehat{A}_{(z,g)}(\ell(\mathbf{x}))\right]\geq\mathbb{V}\Big(\widehat{\mu}_{(z,g)}(\ell(\mathbf{x}))\Big).$$

We now explicit the covariance $\widehat{\mathbb{C}}^c\Big(\widehat{\mu}_{(z,g)}(\ell(\mathbf{x})),\widehat{\mu}_{(z',g')}(\ell(\mathbf{x}))\Big)$ in the case we have pairs of units (i,j), whose joint probability of experiencing the conditions (z,g) and (z',g'), respectively, is zero, that is, $\pi_{ikjk}(z,g,z',g')=0$:

$$
\begin{aligned}
\widehat{\mathbb{C}}\Big(\widehat{\mu}_{z,g}(\ell(\mathbf{x})),\widehat{\mu}_{z',g'}(\ell(\mathbf{x}))\Big) &= \frac{1}{N(\ell(\mathbf{x}))^2}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\sum_{j\neq i:\pi_{ikjk}(z,g,z',g')>0} \\
&\quad \frac{\mathbb{1}(Z_{ik}=z,G_{ik}=g,\mathbf{X}_{ik}\in\ell(\mathbf{x}))}{\pi_{ikjk}(z,g,z',g')} \\
&\quad \frac{\mathbb{1}(Z_{jk}=z',G_{jk}=g',\mathbf{X}_{jk}\in\ell(\mathbf{x}))}{\pi_{ikjk}(z,g,z',g')} \\
&\times \quad [\pi_{ikjk}(z,g,z',g')-\pi_{ik}(z,g)\pi_{jk}(z',g')] \\
&\times \quad \frac{Y_{ik}}{\pi_{ik}(z,g)}\frac{Y_{jk}}{\pi_{jk}(z',g')} \\
&- \quad \frac{1}{N(\ell(\mathbf{x}))^2}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\sum_{j\neq i:\pi_{ikjk}(z,g,z',g')=0} \\
&\quad \Big[\frac{\mathbb{1}(Z_{ik}=z,G_{ik}=g,\mathbf{X}_{ik}\in\ell(\mathbf{x}))Y_{ik}^2}{2\pi_{ik}(z,g)} \\
&+ \quad \frac{\mathbb{1}(Z_{ik}=z',G_{ik}=g',\mathbf{X}_{ik}\in\ell(\mathbf{x}))Y_{ik}^2}{2\pi_{ik}(z',g')}\Big] \qquad \text{(B.6)}
\end{aligned}
$$

B.3 Additional Monte Carlo Simulations

We included in the simulation study two additional sets of simulations where i) we introduce correlation between the covariates, and ii) we replace the Erdos-Renyi model for network formation with an exponential random graph (ERGM) model introducing homophily within the clusters. These two additional simulations are conducted with 30 clusters and under the first scenario introduced in Section 3.5, where the heterogeneity is the same for the two causal effects of intetest.

B.3.1 Correlated covariates

In Figure 62 we report the number of correctly detected leaves under low and high correlation (0.25 and 0.5), while in Tables 19 and 20 we report the estimated treatment and spillover effects with their standard error in the two heterogenous leaves, together with the MSE, bias and coverage

of the average treatment and spillover effects in the sample. In Figure 62 we can see that the correlation between covariates compromises the ability of the algorithm to correctly identify the heterogeneous subgroups. This is due to the fact that, as the covariates become more *similar* to each other, it becomes harder for the algorithm to detect the true HDVs. Such a problem is common to all tree-based algorithms. Hence, we argue that one should carefully check the correlation patterns between the variables to get a sense of the reliability of the discovered subgroups.

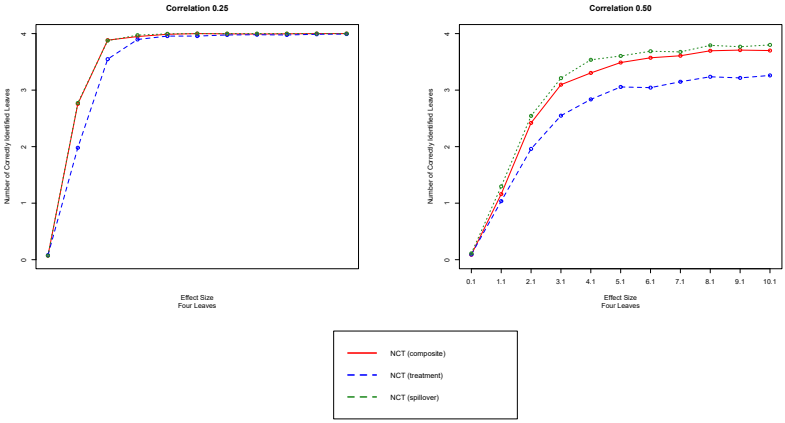


Figure 62: Simulations' results for correctly discovered leaves in the first scenario with correlated covariates

Table 19: Simulations' results for the first scenario with correlated covariates (0.25)

	Treatment Effects						
Effect Size	$\hat{\tau}_{l_1}$	$\hat{se}(\hat{\tau}_{l_1})$	$\hat{\tau}_{l_2}$	$\hat{se}(\hat{\tau}_{l_2})$	MSE	Bias	Coverage
0.1	0.116	0.288	-0.230	0.225	0.071	-0.057	1.000
1.1	1.093	0.280	-1.078	0.288	0.080	0.008	0.946
2.1	2.084	0.403	-2.067	0.401	0.147	0.009	0.939
3.1	3.112	0.553	-3.134	0.555	0.273	-0.011	0.958
4.1	4.082	0.692	-4.140	0.699	0.400	-0.029	0.961
5.1	5.126	0.843	-5.162	0.852	0.673	-0.018	0.950
6.1	6.121	1.007	-6.157	1.008	1.000	-0.018	0.948
7.1	7.019	1.161	-7.067	1.165	1.214	-0.024	0.958
8.1	8.018	1.309	-8.085	1.309	1.557	-0.033	0.959
9.1	9.221	1.509	-9.088	1.491	2.014	0.066	0.953
10.1	10.214	1.655	-10.195	1.648	2.441	0.009	0.957
	Spillover Effects						
	$\hat{\delta}_{l_1}$	$\hat{se}(\hat{\delta}_{l_1})$	$\hat{\delta}_{l_2}$	$\hat{se}(\hat{\delta}_{l_1})$	MSE	Bias	Coverage
0.1	0.069	0.185	-0.200	0.194	0.025	-0.065	0.944
1.1	1.102	0.227	-1.090	0.227	0.046	0.006	0.970
2.1	2.111	0.302	-2.117	0.302	0.078	-0.003	0.972
3.1	3.093	0.392	-3.116	0.393	0.112	-0.011	0.983
4.1	4.086	0.488	-4.093	0.490	0.169	-0.004	0.979
5.1	5.131	0.589	-5.078	0.587	0.246	0.026	0.971
6.1	6.071	0.692	-6.051	0.693	0.331	0.010	0.979
7.1	7.109	0.801	-7.087	0.802	0.458	0.011	0.980
8.1	8.092	0.906	-8.063	0.902	0.577	0.014	0.976
9.1	9.110	1.016	-9.060	1.010	0.659	0.025	0.990
10.1	10.122	1.123	-10.101	1.124	0.861	0.010	0.980

Table 20: Simulations' results for the first scenario with correlated covariates (0.50)

	Treatment Effects						
Effect Size	$\hat{\tau}_{l_1}$	$\hat{se}(\hat{\tau}_{l_1})$	$\hat{\tau}_{l_2}$	$\hat{se}(\hat{\tau}_{l_2})$	MSE	Bias	Coverage
0.1	0.129	0.223	-0.101	0.217	0.050	0.014	0.909
1.1	1.093	0.268	-1.120	0.272	0.066	-0.014	0.972
2.1	2.140	0.392	-2.114	0.378	0.158	0.013	0.949
3.1	3.104	0.503	-3.065	0.505	0.241	0.019	0.946
4.1	4.123	0.652	-4.188	0.664	0.399	-0.033	0.960
5.1	5.081	0.795	-5.108	0.801	0.618	-0.013	0.953
6.1	6.138	0.960	-6.146	0.943	0.825	-0.004	0.959
7.1	7.154	1.110	-7.063	1.096	1.088	0.045	0.956
8.1	8.093	1.227	-8.174	1.257	1.388	-0.041	0.949
9.1	9.080	1.380	-9.116	1.389	1.637	-0.018	0.959
10.1	10.086	1.529	-10.122	1.520	1.952	-0.018	0.952
	Spillover Effects						
	$\hat{\delta}_{l_1}$	$\hat{se}(\hat{\delta}_{l_1})$	$\hat{\delta}_{l_2}$	$\hat{se}(\hat{\delta}_{l_1})$	MSE	Bias	Coverage
0.1	0.144	0.190	-0.202	0.175	0.030	-0.029	0.864
1.1	1.091	0.214	-1.087	0.212	0.047	0.002	0.952
2.1	2.103	0.282	-2.137	0.285	0.064	-0.017	0.975
3.1	3.119	0.364	-3.072	0.365	0.097	0.024	0.977
4.1	4.098	0.458	-4.090	0.458	0.156	0.004	0.962
5.1	5.056	0.551	-5.101	0.553	0.218	-0.022	0.976
6.1	6.127	0.654	-6.093	0.649	0.302	0.017	0.980
7.1	7.093	0.750	-7.130	0.752	0.370	-0.018	0.988
8.1	8.162	0.852	-8.038	0.844	0.476	0.062	0.984
9.1	9.122	0.948	-9.026	0.943	0.571	0.048	0.982
10.1	10.043	1.046	-10.177	1.052	0.698	-0.067	0.982

Nevertheless, for both correlation levels (0.25 and 0.50) the estimator seems to perform well within correctly detected leaves (see Tables 19 and 20).

B.3.2 Network homophily within the clusters

Table 21 shows the results in the case of network homophily within the clusters. In this case, We find larger standard errors than the original scenario reported in 5 without homophily. As a consequence, the Monte-Carlo MSE is also slightly larger.

Table 21: Simulations' results for the first scenario with network homophily within the clusters (30 clusters)

	Treatment Effects						
Effect Size	$\hat{\tau}_{l_1}$	$\widehat{se}(\hat{\tau}_{l_1})$	$\hat{\tau}_{l_2}$	$\widehat{se}(\hat{\tau}_{l_2})$	MSE	Bias	Coverage
0.1	0.152	0.241	-0.108	0.259	0.045	0.022	1.000
1.1	1.069	0.312	-1.094	0.315	0.100	-0.013	0.953
2.1	2.130	0.454	-2.119	0.448	0.188	0.006	0.959
3.1	3.149	0.608	-3.083	0.606	0.335	0.033	0.955
4.1	4.099	0.775	-4.084	0.763	0.477	0.007	0.959
5.1	5.115	0.946	-5.116	0.943	0.736	0.000	0.961
6.1	6.162	1.127	-6.150	1.127	1.043	0.006	0.970
7.1	7.076	1.285	-7.175	1.289	1.377	-0.049	0.955
8.1	8.129	1.465	-8.000	1.446	1.752	0.064	0.957
9.1	9.029	1.621	-9.136	1.629	2.249	-0.053	0.953
10.1	10.163	1.814	-10.133	1.815	2.977	0.015	0.957
	Spillover Effects						
	δ_{l_1}	$\widehat{se}(\delta_{l_1})$	δ_{l_2}	$\widehat{se}(\delta_{l_1})$	MSE	Bias	Coverage
0.1	0.084	0.221	-0.142	0.222	0.036	-0.029	0.971
1.1	1.082	0.268	-1.100	0.274	0.069	-0.009	0.957
2.1	2.102	0.370	-2.084	0.366	0.120	0.009	0.964
3.1	3.116	0.483	-3.112	0.486	0.172	0.002	0.974
4.1	4.065	0.603	-4.114	0.610	0.260	-0.025	0.976
5.1	5.101	0.740	-5.077	0.737	0.389	0.012	0.980
6.1	6.076	0.871	-6.045	0.869	0.538	0.015	0.979
7.1	7.088	1.004	-7.054	0.997	0.738	0.017	0.972
8.1	8.041	1.132	-8.150	1.142	0.936	-0.055	0.971
9.1	9.157	1.275	-9.030	1.263	1.070	0.064	0.976
10.1	10.066	1.408	-10.093	1.407	1.391	-0.014	0.982

Appendix C

Appendix

C.1 Similarity Measures

In this section, we detail the four dyadic similarity measures that we employed for imputing missing links among dyads. They catch the baseline degree of similarity between each pair of students with respect to hobbies, school attitudes, cultural interests and personal background. They all have been derived starting from individual-level covariates and they have been computed as follows:

1. Hobbies Similarity $x1_{ij}$: Jaccard similarity between hobbies of units i and j. : in particular, students were asked whether they are interested in politics, whether they practice sports, gymnastics or volunteering and whether they love painting, listening music, chatting on social networks and watching TV.

2. School Attitudes Similarity $x2_{ij}$: this variable represents a measure of similarity with respect to the school attitudes. This quantity involves both an evaluation of the individual academic performance, expressed in terms of grade point average (gpa), and the attendance of specific extracurricular activities offered by the school (music lessons, language lessons, humanities lessons). The dyadic similarity in school performance between units i and j, that

we call $x2_{ij}^a$, is measured according to the formula $x2_{ij}^a = 1 - \frac{|gpa_i - gpa_j|}{\max(gpa) - \min(gpa)}$. While, the variable $x2_{ij}^b$ quantifies the extent of similarity between i and j in terms of school activities through a Jaccard measure. The final $x2_{ij}$ value results from the mean of these two measures, that is $x2_{ij} = \frac{x2_{ij}^a + x2_{ij}^b}{2}$.

3. Cultural Interests Similarity $x3_{ij}$: this variable indicates the level of similarity with respect to the individual baseline attitude towards culture. Students have been asked to grade the frequency of how they practice the following interests: book reading, symphony listening, theatrical shows watching, cinema going. Higher values correspond to a more frequent accomplishment of that specific activity. The $x3_{ij}$ variable is obtained by computing the Euclidean distance among these measures (then subtracting the resulting value from 1, so to get a similarity measure, instead that a measure of distance). This value has been in turn standardized, so to get a measure which varies between 0 and 1.

4. Personal Background Similarity $x4_{ij}$: this variable measures the level of similarity among individual personal characteristics. Formally, given some students i and j, both belonging to $\mathcal{N}$, $x4_{ij}$ is defined as a Jaccard similarity of their respective personal features: in particular, the personal characteristics that are included in this evaluation are related to the gender, to the seniority, to the geographical origin of the individual (the survey asks the student to declare if she/he is born abroad or not) and to the current living area (the survey asks the student to declare if she/he is living in suburban areas or not).

C.2 Multiple Imputation Algorithm: Stability

Figure 63 shows the trace plot concerning the mean and the standard deviation of the link indicator variable. The algorithm multiply imputes missing links after a given number of iterations (we have set this number at 5, which is the default value) for making the prediction more stable. In

each iteration, the multiple imputation algorithm is based on a random forest composed by five trees and recursively splits data to predict the dyadic outcome (i.e the presence / absence of the link). The graph suggests that the prediction is highly stable also after a very few number of iterations.

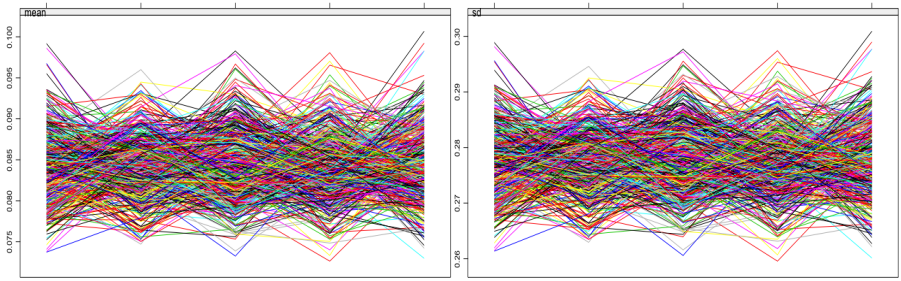


Figure 63: Mean and standard deviation of the link indicator variable: trace plot monitoring the trend of this values over the 5 iterations performed by the algorithm, in the $M = 500$ distinct data-imputations.

C.3 Proofs for τ^* and its estimator

The causal effect of actually receiving the treatment, denoted with tau^*, can be expressed as the average comparison of the two potential outcomes,

$$\tau^* = \mathbb{E}\left[Y_{it''}(1)\right] - \mathbb{E}\left[Y_{it''}(0)\right]$$

First, we want to prove that, under Assumption 16, it can be formally written as

$$\begin{aligned}\tau^* = \mathbb{E}\left[Y_{it''}(1)|Z_{it}=1\right]P(Z_{it}=1)+ \\ \sum_{\mathbf{z}_{-it}}\sum_{\mathbf{x}}\Big(\mathbb{E}\left[Y_{it''}(1)|Z_{it'}=1, Z_{it}=0, \mathbf{Z}_{-it}=\mathbf{z}_{-it}, \mathbf{X}_i=\mathbf{x}\right]- \\ \mathbb{E}\left[Y_{it''}(0)|Z_{it'}=0, Z_{it}=0, \mathbf{Z}_{-it}=\mathbf{z}_{-it}, \mathbf{X}_i=\mathbf{x}\right]\Big) \\ P(Z_{it}=0, \mathbf{Z}_{-it}=\mathbf{z}_{-it}, \mathbf{X}_i=\mathbf{x}).\end{aligned}$$

Proof.

$$
\begin{aligned}
\tau^* &= \sum_{\mathbf{z}} \sum_{\mathbf{x}} \Big(\mathbb{E}\left[Y_{it''}(1) | \mathbf{Z}_t = \mathbf{z}, \mathbf{X}_i = \mathbf{x} \right] - \mathbb{E}\left[Y_{it''}(0) | \mathbf{Z}_t = \mathbf{z}, \mathbf{X}_i = \mathbf{x} \right] \Big) P(\mathbf{Z}_t = \mathbf{z}, \mathbf{X}_i = \mathbf{x}) \\
&= \sum_{\mathbf{z}_{-it}} \sum_{\mathbf{x}} \mathbb{E}\left[Y_{it''}(1) | Z_{it} = 1, \mathbf{Z}_{-it} = \mathbf{z}_{-it}, \mathbf{X}_i = \mathbf{x} \right] P(Z_{it} = 1, \mathbf{Z}_{-it} = \mathbf{z}_{-it}, \mathbf{X}_i = \mathbf{x}) + \\
&\quad \sum_{\mathbf{z}_{-it}} \sum_{\mathbf{x}} \mathbb{E}\left[Y_{it''}(1) | Z_{it} = 0, \mathbf{Z}_{-it} = \mathbf{z}_{-it}, \mathbf{X}_i = \mathbf{x} \right] P(Z_{it} = 0, \mathbf{Z}_{-it} = \mathbf{z}_{-it}, \mathbf{X}_i = \mathbf{x}) - \\
&\quad \sum_{\mathbf{z}_{-it}} \sum_{\mathbf{x}} \mathbb{E}\left[Y_{it''}(0) | Z_{it} = 1, \mathbf{Z}_{-it} = \mathbf{z}_{-it}, \mathbf{X}_i = \mathbf{x} \right] P(Z_{it} = 1, \mathbf{Z}_{-it} = \mathbf{z}_{-it}, \mathbf{X}_i = \mathbf{x}) - \\
&\quad \sum_{\mathbf{z}_{-it}} \sum_{\mathbf{x}} \mathbb{E}\left[Y_{it''}(0) | Z_{it} = 0, \mathbf{Z}_{-it} = \mathbf{z}_{-it}, \mathbf{X}_i = \mathbf{x} \right] P(Z_{it} = 0, \mathbf{Z}_{-it} = \mathbf{z}_{-it}, \mathbf{X}_i = \mathbf{x}) \\
&= \sum_{\mathbf{z}_{-it}} \sum_{\mathbf{x}} \mathbb{E}\left[Y_{it''}(1) | Z_{it} = 1, \mathbf{Z}_{-it} = \mathbf{z}_{-it}, \mathbf{X}_i = \mathbf{x} \right] P(\mathbf{Z}_{-it} = \mathbf{z}_{-it}, \mathbf{X}_i = \mathbf{x} | Z_{it} = 1) P(Z_{it} = 1) + \\
&\quad \sum_{\mathbf{z}_{-it}} \sum_{\mathbf{x}} \mathbb{E}\left[Y_{it''}(1) | Z_{it'} = 1, Z_{it} = 0, \mathbf{Z}_{-it} = \mathbf{z}_{-it}, \mathbf{X}_i = \mathbf{x} \right] P(Z_{it} = 0, \mathbf{Z}_{-it} = \mathbf{z}_{-it}, \mathbf{X}_i = \mathbf{x}) - \\
&\quad \sum_{\mathbf{z}_{-it}} \sum_{\mathbf{x}} \mathbb{E}\left[Y_{it''}(0) | Z_{it'} = 0, Z_{it} = 0, \mathbf{Z}_{-it} = \mathbf{z}_{-it}, \mathbf{X}_i = \mathbf{x} \right] P(Z_{it} = 0, \mathbf{Z}_{-it} = \mathbf{z}_{-it}, \mathbf{X}_i = \mathbf{x}) \\
&= \mathbb{E}\left[Y_{it''}(1) | Z_{it} = 1 \right] P(Z_{it} = 1) + \sum_{\mathbf{z}_{-it}} \sum_{\mathbf{x}} \Big(\mathbb{E}\left[Y_{it''}(1) | Z_{it'} = 1, Z_{it} = 0, \mathbf{Z}_{-it} = \mathbf{z}_{-it}, \mathbf{X}_i = \mathbf{x} \right] \\
&\quad - \mathbb{E}\left[Y_{it''}(0) | Z_{it'} = 0, Z_{it} = 0, \mathbf{Z}_{-it} = \mathbf{z}_{-it}, \mathbf{X}_i = \mathbf{x} \right] \Big) P(Z_{it} = 0, \mathbf{Z}_{-it} = \mathbf{z}_{-it}, \mathbf{X}_i = \mathbf{x}).
\end{aligned}
$$

□

Second, we intend to prove that, under Assumption 16, if the real treatment assignment vector were fully observed over the population, we would be able to identify it by means of the following Horvitz-Thompson estimator

$$
\widehat{\tau}^{\star} = \frac{1}{N} \Big[\sum_{i=1}^{N} Z_{it'} \frac{Y_{it''}}{\pi_{it'}(1; \mathbf{Z}_{-it}, \mathbf{X}_i)} - \sum_{i=1}^{N} (1 - Z_{it'}) \frac{Y_{it''}}{1 - \pi_{it'}(1; \mathbf{Z}_{-it}, \mathbf{X}_i)} \Big].
$$

Proof.

$$\mathbb{E}\left[\frac{Y_{it''}Z_{it'}}{\pi_{it'}(1;\mathbf{Z}_{-it},\mathbf{X}_i)}\right]=\mathbb{E}\left[\frac{Y_{it''}I(Z_{it'}=1)}{\pi_{it'}(1;\mathbf{Z}_{-it},\mathbf{X}_i)}\right]=\mathbb{E}\left[\frac{Y_{it''}(1)I(Z_{it'}=1)}{\pi_{it'}(1;\mathbf{Z}_{-it},\mathbf{X}_i)}\right]$$
$$=\mathbb{E}\left[\mathbb{E}\left[\frac{Y_{it''}(1)I(Z_{it'}=1)}{\pi_{it'}(1;\mathbf{Z}_{-it},\mathbf{X}_i)}|\mathbf{Z}_{-it},\mathbf{X}_i\right]\right]$$
$$=\mathbb{E}\left[\frac{1}{\pi_{it'}(1;\mathbf{Z}_{-it},\mathbf{X}_i)}\mathbb{E}\left[Y_{it''}(1)I(Z_{it'}=1)|\mathbf{Z}_{-it},\mathbf{X}_i\right]\right].$$

Now, it is enough to observe that

$$\mathbb{E}\left[Y_{it''}(1)I(Z_{it'}=1)|\mathbf{Z}_{-it},\mathbf{X}_i\right]=$$
$$\mathbb{E}\left[Y_{it''}(1)I(Z_{it'}=1)|Z_{it}=1,\mathbf{Z}_{-it},\mathbf{X}_i\right]P(Z_{it}=1|\mathbf{Z}_{-it},\mathbf{X}_i)+$$
$$\mathbb{E}\left[Y_{it''}(1)I(Z_{it'}=1)|Z_{it}=0,\mathbf{Z}_{-it},\mathbf{X}_i\right]P(Z_{it}=0|\mathbf{Z}_{-it},\mathbf{X}_i)=$$
$$\mathbb{E}\left[Y_{it''}(1)|Z_{it}=1,\mathbf{Z}_{-it},\mathbf{X}_i\right]\pi_{it}(1;\mathbf{Z}_{-it},\mathbf{X}_i)+$$
$$\mathbb{E}\left[Y_{it''}(1)|Z_{it}=0,\mathbf{Z}_{-it},\mathbf{X}_i\right]P(Z_{it'}=1|Z_{it}=0,\mathbf{Z}_{-it},\mathbf{X}_i)(1-\pi_{it}(1;\mathbf{Z}_{-it},\mathbf{X}_i))=$$
$$\mathbb{E}[Y_{it''}(1)]\pi_{it}(1;\mathbf{Z}_{-it},\mathbf{X}_i)+\mathbb{E}[Y_{it''}(1)]\rho_i(1-\pi_{it}(1;\mathbf{Z}_{-it},\mathbf{X}_i))=$$
$$\mathbb{E}[Y_{it''}(1)]\pi_{it'}(1;\mathbf{Z}_{-it},\mathbf{X}_i).$$

(For the second equality we have used that $Z_{it}=1$ implies $Z_{it'}=1$, the second par of Assumption 16 and the definition of $\pi_{it}(1;\mathbf{Z}_{-it},\mathbf{X}_i)$. For the third equality we have used the first part of Assumption (16) and the definition of ρ_i. Finally, for the last equality, we have used the definition of $\pi_{it'}(1;\mathbf{Z}_{-it},\mathbf{X}_i)$.)

Similarly, we have

$$\mathbb{E}\left[\frac{Y_{it''}(1-Z_{it'})}{(1-\pi_{it'}(1;\mathbf{Z}_{-it},\mathbf{X}_i))}\right]=\mathbb{E}\left[\frac{Y_{it''}I(Z_{it'}=0)}{(1-\pi_{it'}(1;\mathbf{Z}_{-it},\mathbf{X}_i))}\right]=$$
$$\mathbb{E}\left[\frac{Y_{it''}(0)I(Z_{it'}=0)}{(1-\pi_{it'}(1;\mathbf{Z}_{-it},\mathbf{X}_i))}\right]=\mathbb{E}\left[\mathbb{E}\left[\frac{Y_{it''}(0)I(Z_{it'}=0)}{(1-\pi_{it'}(1;\mathbf{Z}_{-it},\mathbf{X}_i))}|\mathbf{Z}_{-it},\mathbf{X}_i\right]\right]=$$
$$\mathbb{E}\left[\frac{1}{(1-\pi_{it'}(1;\mathbf{Z}_{-it},\mathbf{X}_i))}\mathbb{E}\left[Y_{it''}(0)I(Z_{it'}=0)|\mathbf{Z}_{-it},\mathbf{X}_i\right]\right],$$

where

$$\mathbb{E}\left[Y_{it''}(0)I(Z_{it'}=0)|\mathbf{Z}_{-it},\mathbf{X}_i\right]=$$
$$\mathbb{E}\left[Y_{it''}(0)I(Z_{it'}=0)|Z_{it}=1,\mathbf{Z}_{-it},\mathbf{X}_i\right]P(Z_{it}=1|\mathbf{Z}_{-it},\mathbf{X}_i)+$$
$$\mathbb{E}\left[Y_{it''}(0)I(Z_{it'}=0)|Z_{it}=0,\mathbf{Z}_{-it},\mathbf{X}_i\right]P(Z_{it}=0|\mathbf{Z}_{-it},\mathbf{X}_i)=$$
$$\mathbb{E}\left[Y_{it''}(0)|Z_{it}=0,\mathbf{Z}_{-it},\mathbf{X}_i\right]P(Z_{it'}=0|Z_{it}=0,\mathbf{Z}_{-it},\mathbf{X}_i)(1-\pi_{it}(1;\mathbf{Z}_{-it},\mathbf{X}_i))=$$
$$\mathbb{E}[Y_{it''}(0)](1-\rho_i)(1-\pi_{it}(1;\mathbf{Z}_{-it},\mathbf{X}_i))=\mathbb{E}[Y_{it''}(0)](1-\pi_{it'}(1;\mathbf{Z}_{-it},\mathbf{X}_i)).$$

Remark:
The estimator

$$\widehat{\tau}^{\star\star}=\frac{1}{N}\left[\sum_{i=1}^{N}Z_{it'}\frac{Y_{it''}}{\pi_{it'}(1)}-\sum_{i=1}^{N}(1-Z_{it'})\frac{Y_{it''}}{1-\pi_{it'}(1)}\right]$$

is trivially unbiased if we have

$$Y_{it''}(Z_{it'}=z)\perp\!\!\!\perp\mathbf{Z}_{t'}\qquad\forall z\in\{0,1\}.$$

Note that we always have $0<\pi_{it'}(1)<1$ (since the term $\pi_{it}(1)$ in its expression belongs to $(0,1)$ by assumption) and, moreover, it coincides with the standard Horvitz-Thompson estimator when there is no possibility of diffusion. □

C.4 Diffusion Bias

Here, we derive the bias due to the hidden treatment diffusion process. If the policy maker neglected the possibility of any diffusion process playing a role in the analysis, she would estimate the quantity $\tau^b_{obs} = \mathbb{E}\left[Y_{it''}|Z_{it}=1\right] - \mathbb{E}\left[Y_{it''}|Z_{it}=0\right]$, that is

$$
\begin{aligned}
\tau^b_{obs} &= \mathbb{E}\left[Y_{it''}|Z_{it}=1\right] - \mathbb{E}\left[Y_{it''}|Z_{it}=0\right] \\
&= \Bigg[\mathbb{E}\left[Y_{it''}|Z_{it'}=1, Z_{it}=1\right] P(Z_{it'}=1|Z_{it}=1) + \\
&\quad \mathbb{E}\left[Y_{it''}|Z_{it'}=0, Z_{it}=1\right] P(Z_{it'}=0|Z_{it}=1) \Bigg] - \\
&\quad \Bigg[\mathbb{E}\left[Y_{it''}|Z_{it'}=1, Z_{it}=0\right] P(Z_{it'}=1|Z_{it}=0) + \\
&\quad \mathbb{E}\left[Y_{it''}|Z_{it'}=0, Z_{it}=0\right] P(Z_{it'}=0|Z_{it}=0) \Bigg] \\
&= \mathbb{E}\left[Y_{it''}(1)|Z_{it'}=1, Z_{it}=1\right] - \mathbb{E}\left[Y_{it''}(1)|Z_{it'}=1, Z_{it}=0\right] \mathbb{E}[\rho_i|Z_{it}=0] - \\
&\quad \mathbb{E}\left[Y_{it''}(0)|Z_{it'}=0, Z_{it}=0\right] (1 - \mathbb{E}[\rho_i|Z_{it}=0]) \\
&= \mathbb{E}\left[Y_{it''}(1)|Z_{it}=1\right] - \mathbb{E}\left[Y_{it''}(1)|Z_{it'}=1, Z_{it}=0\right] \mathbb{E}[\rho_i|Z_{it}=0] - \\
&\quad \mathbb{E}\left[Y_{it''}(0)|Z_{it'}=0, Z_{it}=0\right] (1 - \mathbb{E}[\rho_i|Z_{it}=0]) \,,
\end{aligned}
$$

where we have used that

$$
\begin{aligned}
P(Z_{it'}=1|Z_{it}=1) &= 1, \\
P(Z_{it'}=0|Z_{it}=1) &= 0 \,.
\end{aligned}
$$

and $\rho_i = P(Z_{it'}=1|Z_{it}=0, \boldsymbol{Z}_{-it}, \boldsymbol{X}_i)$ and, in the case of a discrete $\mathbf{X}_i$, we have

$$
\mathbb{E}[\rho_i|Z_{it}=0] = \sum_{\mathbf{x}} \sum_{\mathbf{z}_{-i}} \rho_i(\mathbf{z}_{-i}, \mathbf{x}) P(\mathbf{Z}_{-it} = \mathbf{z}_{-i}, \mathbf{X}_i = \mathbf{x}|Z_{it}=0).
$$

Note that, when the random variables Z_{it} and $\mathbf{X}_i$, with $i \in \mathcal{N}$, are all independent, we simply have $\mathbb{E}[\rho_i|Z_{it}=0] = \mathbb{E}[\rho_i]$. Moreover, under the first assumption included in Assumption 16, we have $\mathbb{E}\left[Y_{it''}(1)|Z_{it} = 1\right] = \mathbb{E}[Y_{it''}(1)]$.

The treatment diffusion bias can be obtained by computing the difference between the above quantity and $\tau^* = \mathbb{E}[Y_{it''}(1)] - \mathbb{E}[Y_{it''}(0)]$ and, under the first assumption included in Assumption 16, we have:

$$\begin{aligned} b = \tau^b_{obs} - \tau^* &= \mathbb{E}\left[Y_{it''}(1)\right] - \mathbb{E}\left[Y_{it''}(1)|Z_{it'}=1, Z_{it}=0\right]\mathbb{E}[\rho_i|Z_{it}=0] - \\ &\quad \mathbb{E}\left[Y_{it''}(0)|Z_{it'}=0, Z_{it}=0\right](1-\mathbb{E}[\rho_i|Z_{it}=0]) - \\ &\quad \mathbb{E}[Y_{it''}(1)] + \mathbb{E}[Y_{it''}(0)] \\ &= \mathbb{E}\left[Y_{it''}(0)\right] - \mathbb{E}\left[Y_{it''}(0)|Z_{it'}=0, Z_{it}=0\right](1-\mathbb{E}[\rho_i|Z_{it}=0]) - \\ &\quad \mathbb{E}\left[Y_{it''}(1)|Z_{it'}=1, Z_{it}=0\right]\mathbb{E}[\rho_i|Z_{it}=0]. \end{aligned}$$

If the Assumption 16 is not satisfied (as in the above simulations), the bias is given by the above quantity plus $\mathbb{E}[Y_{it''}(1)|Z_{it}=1] - \mathbb{E}[Y_{it''}(1)]$.

Appendix D

Appendix

D.1 String-Matching Algorithm

We perform a string-matching algorithm in order to merge committee data and firms data. In other words, the goal of our procedure is to detect which of the supporting entities are lobbying firms. Committee data involve firms, parties, associations and private citizens and, since our project focuses on the role of lobbying firms, we must detect them and isolate them from the rest of the entities. In view of the fact that firms collected in committee data are not characterized by a unique firm identifier (such as the BVD ID number) which can attest that they are official companies, we must rely on a string-matching algorithm that catches similarities between their reported names and the names of all certified US companies. We assume that all the entities which are not identified as firms by our string-matching algorithm are not real US firms. The string-matching algorithm is organized in few steps and it takes as input the list of organizations, which have financially supported at least one elected politician in the three legislative cycles, and the list of the official US companies, which are regularly registered in the US lists [1]. The

[1] We do not consider all the existing US firms and companies, we take into account of a sub-sample of official companies (about 2 millions) with respect to that we have sufficient information in terms of location, operating sectors and financial reports. They plausibly represent the most relevant US companies.

structure of the string-matching algorithm is organized according to the following steps.

1. Remove foreign entities: we remove from the list of political supporters in committee data all the organizations which are not located in the US.

2. Remove non-firms according to their code: remove from the committees all the entities whose code explicitly identifies them as parties, associations or political organizations. Committees are characterized by a five character code identifying their industry or ideology. The fist digit of this code signals their operating sectors: we delete from the analysis entities referring to sectors Z,X,J or Y, which pertain to political parties and political organizations.

3. Remove non-firms according to their name: remove from the committee dataset all the entries whose name contains the words "association", "federation" or "assn". The presence of one of these words clearly signal that the given organization is not a pure company.

4. Clean up the strings (preliminary cleaning): remove from the strings, located both in committee data and in firms data, all the spurious elements (erroneous punctuation, special symbols, additional final signs - such as, for instance, "inc." and "co."-). Then, implement all the string-adjustments for making the committee names' nomenclature compatible with the one which characterizes company names in the ORBIS Database (for instance, the sign "u.s" is unrecognized in the ORBIS database and needs to be transformed to "us").

5. Look for perfect matches: for each unique entry of the committee database (i.e for each name of a political supporting entity) skim the firms database and grab whether the string spots perfect matches (i.e this happens if in the firms database there are one or more firms which are named *exactly* like the given committee).

Collect the perfectly merged committees of this step in A (together with their corresponding match in the ORBIS Database).

6. Clean up the strings from recurrent terms (additional cleaning). Consider all the entities which have not been successfully matched at the previous step. Remove from their names all the recurrent terms (such as "corporation", "group", "company" etc.) that may taint the matching-search procedure (for instance the entities "Vitamingas company" and "Vitamingas" clearly refer to the same company, but they would have not been matched in the previous step because their names do not perfectly coincide).

7. Look for perfect matches: repeat the step 5, after the further cleaning in the strings (performed at the previous step). Collect the perfectly merged committees of this step in $A1$ (together with their corresponding match in the ORBIS Database).

8. Look for fuzzy matches: consider all the entities (in both datasets) which have not successfully matched at steps 5 and 7. Since any attempt of perfect matching is failed for them, we perform a fuzzy matching algorithm, where we pairwise compare strings and we search for similar patterns. The fuzzy merged committees of this step are collected in $A2$ (together with their corresponding match in the ORBIS Database).

9. Manage multi-matches: the steps 5,7 and 8 may have generated multiple matches, in the sense that a given entity in the committee data may have been matched to more than one firm in the ORBIS database (since at each step we remove matched firms, multiple matches have occurred at the same step). Since we do not have elements to discern the right match among the possible ones, we decide to pair the multiply matched entity in the committee dataset, with the most relevant of its matches (i.e the firm with the highest number of employees at 2010 [2]). This step generates a complete list

[2] if we are not able to identify a unique most relevant firm among the candidate matches we do not match that firm anyway

of pairwise matches A, which do not contain multiple matches.

10. Collect all the matched entities: the matched organizations are collected in A, so that now each committee, which has been identified as a real US company by the algorithm, is characterized by its own BVD firm identifier (that corresponds to the BVD of its correspondent match).

These steps exhaustively describe the procedure that we have followed to match committees with US official companies. It follows that all those committees, which have not been identified as firms by the algorithm, are removed from the analysis.

D.2 Descriptive Analysis

This Section provides some descriptive statistics which refer to the data, we have analyzed in Chapter 5. All the aspects are investigated with respect to the three legislatures we have included in our analysis, namely the 111^{th}, 112^{th} and 113^{th} Congresses in the US House of Representatives.
The first Table presents some main statistics. The key aspects that this table points out are the following: i) the number of companies involved in lobbying activities is extremely relevant in all the three Congresses and transactions from firms are about one third of the total number of transactions (with respect to both their number and their amount); ii) an high quote of legislators has received funds from at least one firm, during the electoral campaign and conditioning on having received funds from at least one PAC, the probability of having been financed from at least one firm is about 0.99.

Table 22: Main statistics

	111^{th}	112^{th}	113^{th}
Number of transactions	76434	68546	58137
Number of transactions involving firms	26715	23236	19285
Unique entities involved in transactions	3702	3668	3387
Unique firms involved in transactions	1007	970	897
Unique groups involved in transactions	798	768	718
Total amount of transactions	143761574	138910053	120496348
Total amount of transactions involving firms	44344588	43963589	33223769
Politicians who received funds	383	376	330
Politicians who received funds from firms	379	374	328
Politicians who didn't receive funds	56	58	86
Total elected politicians	439	434	416

The following Table shows the distributions concerning the number of transactions at the benefit of each politician, over the three legislatures. It confirms the general intuition: the financing support coming from firms is extremely relevant for legislators.

Table 23: Number of transactions, per politician

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	N
	Number of transactions, per politician [3]						
111^{th}	1	125.00	176	199.56658	265.0	627	383
112^{th}	1	121.00	160	182.30319	223.5	623	378
113^{th}	1	119.25	155	176.17273	218.0	607	330
	Number of transactions involving firms, per politician [4]						
111^{th}	1	45.50	62	70.48813	90.0	258	379
112^{th}	8	36.00	54	62.12834	78.0	252	374
113^{th}	1	35.00	51	58.79573	74.0	243	328

The following Table investigates the number of unique financing supporters, per politician. As it is possible to notice, a single politician is supported by a huge amount of entities, during her electoral campaign. About one third of these supporters are real US companies

Table 24: Unique supporters

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	N
	Unique supporting entities, per politician [5]						
111^{th}	1	124.5	175	197.62924	262.50	625	383
112^{th}	1	120.0	160	180.86702	223.00	621	378
113^{th}	1	118.0	153	174.64848	217.75	604	330
	Unique supporting firms, per politician [6]						
111^{th}	1	45.0	62	70.01583	89.50	255	379
112^{th}	8	36.0	54	61.86096	77.00	250	374
113^{th}	1	34.0	51	58.47256	73.00	243	328

The following Table shows the distribution of supported politicians, for firm, in the three legislature. It points out that companies are used not to concentrate their lobbying budget over a single politician only: they prefer to differentiate their risk and to support a relevant number of legislators.

Table 25: How many politicians each company supports

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	N
111^{th}	1	3	10	26.52929	30.00	340	1007
112^{th}	1	3	9	23.95464	26.75	287	970
113^{th}	1	3	9	21.49944	24.00	249	897

The last Table focuses on the amount of transactions. It highlights that PACs invest a huge amount of money in the US electoral campaigns and companies save a relevant amount of funds for lobbying activities.

Table 26: Amount of Transactions

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	N
	Mean amount received from transactions, per politician [7]						
111^{th}	865	1475.403	1644.699	1806.176	1870.791	9081.00	383
112^{th}	1000	1500.686	1698.479	2032.584	1996.976	20216.77	378
113^{th}	1000	1510.387	1724.321	2102.864	2007.122	20216.77	330
	Mean amount received from transactions involving firms, per politician [8]						
111^{th}	1000.000	1432.671	1591.857	1665.778	1782.681	3533.20	379
112^{th}	1052.833	1479.880	1682.338	2059.556	1956.589	83033.62	374
113^{th}	1088.519	1481.391	1695.131	1821.320	1965.387	11176.19	328
	Total amount received from transactions, per politician [9]						
111^{th}	1000	195722.0	291139.0	375356.6	480150.5	1593409	383
112^{th}	1000	205891.8	281096.5	369441.6	433709.2	3497501	378
113^{th}	1000	207673.2	279846.5	365140.4	437193.2	3497501	330
	Total amount received from transactions involving firms, per politician [10]						
111^{th}	1000	69600.00	100850.0	117004.2	139878.0	664860	379
112^{th}	11284	64468.50	90905.5	117549.7	129533.5	3072244	374
113^{th}	1500	60801.75	86643.0	101292.0	123515.5	445176	328

Bibliography

Abadie, A., & Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *econometrica*, *74*(1), 235–267.

Abadie, A., & Imbens, G. W. (2016). Matching on the estimated propensity score. *Econometrica*, *84*(2), 781–807.

Agarwal, A., Deepinder, F., Sharma, R. K., Ranga, G., & Li, J. (2008). Effect of cell phone usage on semen analysis in men attending infertility clinic: An observational study. *Fertility and sterility*, *89*(1), 124–128.

Agresti, A. (2018). *An introduction to categorical data analysis*. Wiley.

Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative science quarterly*, *45*(3), 425–455.

Albert, R., & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, *74*(1), 47.

An, W. (2018). Causal inference with networked treatment diffusion. *Sociological Methodology*, *48*(1), 152–181.

An, W., & VanderWeele, T. J. (2019). Opening the blackbox of treatment interference: Tracing treatment diffusion through network analysis. *Sociological Methods & Research*, 0049124119852384.

Angelucci, M., & Di Maro, V. (2015). *Program evaluation and spillover effects*. The World Bank.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, *91*(434), 444–455.

Ansolabehere, S., De Figueiredo, J. M., & Snyder Jr, J. M. (2003). Why is there so little money in us politics? *Journal of Economic perspectives*, *17*(1), 105–130.

Arduini, T., Patacchini, E., & Rainone, E. (2014). Identification and estimation of outcome response with heterogeneous treatment externalities. *Bank of Italy Temi di Discussione (Working Paper) No*, *974*.

Arduini, T., Patacchini, E., & Rainone, E. (2019). Treatment effects with heterogeneous externalities. *Journal of Business & Economic Statistics*, 1–13.

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

Aronow, P. M., & Middleton, J. A. (2013). A class of unbiased estimators of the average treatment effect in randomized experiments. *Journal of Causal Inference*, *1*(1), 135–154.

Aronow, P. M., & Samii, C. (2017). Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics*, *11*(4), 1912–1947.

Aronow, P. M., Samii, C., & Wang, Y. (2019). Design-based inference for spatial experiments with interference.

Arpino, B., Benedictis, L. D., & Mattei, A. (2017). Implementing propensity score matching with network data: The effect of the general agreement on tariffs and trade on bilateral trade. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *66*(3), 537–554.

Arpino, B., & Mattei, A. (2013). Assessing the impact of financial aids to firms: Causal inference in the presence of interference.

Assmann, S. F., Pocock, S. J., Enos, L. E., & Kasten, L. E. (2000). Subgroup analysis and other (mis) uses of baseline data in clinical trials. *The Lancet*, *355*(9209), 1064–1069.

Athey, S., Eckles, D., & Imbens, G. W. (2018a). Exact p-values for network interference. *Journal of the American Statistical Association*, *113*(521), https://doi.org/10.1080/01621459.2016.1241178, 230–240. https://doi.org/10.1080/01621459.2016.1241178

Athey, S., Eckles, D., & Imbens, G. W. (2018b). Exact p-values for network interference. *Journal of the American Statistical Association*, *113*(521), 230–240.

Athey, S., & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, *113*(27), 7353–7360.

Athey, S., & Imbens, G. W. (2015). Machine learning methods for estimating heterogeneous causal effects. *stat*, *1050*(5).

Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, *11*, 685–725.

Athey, S., Tibshirani, J., Wager, S., Et al. (2019). Generalized random forests. *The Annals of Statistics*, *47*(2), 1148–1178.

Babanezhad, M., Vansteelandt, S., & Goetghebeur, E. (2010). Comparison of causal effect estimators under exposure misclassification. *Journal of Statistical Planning and Inference*, *140*(5), 1306–1319.

Baird, S., Bohren, J. A., McIntosh, C., & Özler, B. (2018). Optimal design of experiments in the presence of interference. *Review of Economics and Statistics*, *100*(5), 844–860.

Baller, I. (2017). Specialists, party members, or national representatives: Patterns in co-sponsorship of amendments in the european parliament. *European Union Politics*, *18*(3), 469–490.

Bandiera, O., Barankay, I., & Rasul, I. (2011). Field experiments with firms. *Journal of Economic Perspectives*, *25*(3), 63–82.

Barber, M. J. (2007). Modularity and community detection in bipartite networks. *Physical Review E*, *76*(6), 066102.

Bargagli Stoffi, F. J., & Gnecco, G. (2018). Estimating heterogeneous causal effects in the presence of irregular assignment mechanisms. *In Proceedings of the 5th IEEE Conference in Data Science and Advanced Analytics*.

Bargagli Stoffi, F. J., & Gnecco, G. (2019). Causal tree with instrumental variable: An extension of the causal tree framework to irregular assignment mechanisms. *International Journal of Data Science and Analytics*.

Bargagli-Stoffi, F. J., De Witte, K., & Gnecco, G. (2019). Heterogeneous causal effects with imperfect compliance: A novel bayesian machine learning approach. *arXiv preprint arXiv:1905.12707*.

Barkley, B. G., Hudgens, M. G., Clemens, J. D., Ali, M., & Emch, M. E. (2017). Causal inference from observational studies with clustered interference. *arXiv preprint arXiv:1711.04834*.

Baron, D. P. (2006). Competitive lobbying and supermajorities in a majority-rule institution. *Scandinavian Journal of Economics*, *108*(4), 607–642.

Barthélemy, M. (2011). Spatial networks. *Physics Reports*, *499*(1-3), 1–101.

Bartlett, M. S. (1947). The use of transformations. *Biometrics*, *3*(1), 39–52.

Basse, G., & Feller, A. (2018). Analyzing two-stage experiments in the presence of interference. *Journal of the American Statistical Association*, *113*(521), 41–55.

Battaglini, M., & Patacchini, E. (2018). Influencing connected legislators. *Journal of Political Economy*, *126*(6), 2277–2322.

Battaglini, M., Sciabolazza, V. L., & Patacchini, E. (2019). Effectiveness of connected legislators. *American Journal of Political Science*.

Battiston, S., & Caldarelli, G. (2013). Systemic risk in financial networks. *Journal of Financial Management, Markets and Institutions*, *1*(2), 129–154.

Becatti, C., Caldarelli, G., & Saracco, F. (2019). Entropy-based randomization of rating networks. *Physical Review E*, *99*(2), 022306.

Becatti, C., Crimaldi, I., & Saracco, F. (2019). Collaboration and followership: A stochastic model for activities in social networks. *PloS one*, *14*(10), e0223768.

Benson, K., & Hartz, A. J. (2000). A comparison of observational studies and randomized, controlled trials. *New England Journal of Medicine*, *342*(25), 1878–1886.

Bertrand, M., Bombardini, M., Fisman, R., & Trebbi, F. (2018). *Tax-exempt lobbying: Corporate philanthropy as a tool for political influence* (tech. rep.). National Bureau of Economic Research.

Besley, T., & Coate, S. (2001). Lobbying and welfare in a representative democracy. *The Review of Economic Studies*, *68*(1), 67–82.

Bianchi, M., Buonanno, P., & Pinotti, P. (2008). Immigration and crime: An empirical analysis. *Bank of Italy Temi di Discussione (Working Paper) No*, *698*.

Bianchi, M., Buonanno, P., & Pinotti, P. (2012). Do immigrants cause crime? *Journal of the European Economic Association*, *10*(6), 1318–1347.

Bigo, D. (2002). Security and immigration: Toward a critique of the governmentality of unease. *Alternatives*, *27*(1_suppl), 63–92.

Binka, F. N., Indome, F., & Smith, T. (1998). Impact of spatial distribution of permethrin-impregnated bed nets on child mortality in rural northern ghana. *The American journal of tropical medicine and hygiene*, *59*(1), 80–85.

Blöchl, F., Theis, F. J., Vega-Redondo, F., & Fisher, E. O. (2011). Vertex centralities in input-output networks reveal the structure of modern economies. *Physical Review E*, *83*(4), 046127.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D.-U. (2006). Complex networks: Structure and dynamics. *Physics reports*, *424*(4-5), 175–308.

Bombardini, M., & Trebbi, F. (2019). *Empirical models of lobbying* (tech. rep.). National Bureau of Economic Research.

Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data, In *Handbook of econometrics*. Elsevier.

Bourdieu, P. (2011). The forms of capital.(1986). *Cultural theory: An anthology*, *1*, 81–93.

Bove, V., & Böhmelt, T. (2016). Does immigration induce terrorism? *The Journal of Politics*, *78*(2), 572–588.

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

Bratton, K. A., & Rouse, S. M. (2011). Networks in the legislative arena: How group dynamics affect cosponsorship. *Legislative Studies Quarterly*, *36*(3), 423–460.

Braun, D., Gorfine, M., Zigler, C., Dominici, F., & Parmigiani, G. (2014). Adjustment for mismeasured exposure using validation data and propensity scores.

Braun, D., Zigler, C., Dominici, F., & Gorfine, M. (2016). Using validation data to adjust the inverse probability weighting estimator for misclassified treatment. *Using Validation Data to Adjust the Inverse Probability Weighting Estimator for Misclassified Treatment*.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. CRC press.

Bridges, C. B., Thompson, W. W., Meltzer, M. I., Reeve, G. R., Talamonti, W. J., Cox, N. J., Lilac, H. A., Hall, H., Klimov, A., & Fukuda, K. (2000). Effectiveness and cost-benefit of influenza vaccination of healthy working adults: A randomized controlled trial. *Jama*, *284*(13), 1655–1663.

Brochmann, G., & Hammar, T. (1999). *Mechanisms of immigration control: A comparative analysis of european regulation policies*. Bloomsbury Academic.

Broderick, M. P., Hansen, C. J., & Russell, K. L. (2008). Exploration of the effectiveness of social distancing on respiratory pathogen transmission implicates environmental contributions. *The Journal of infectious diseases*, *198*(10), 1420–1426.

Buss, T. F. (2001). The effect of state tax incentives on economic growth and firm location decisions: An overview of the literature. *Economic Development Quarterly*, *15*(1), 90–105.

Buuren, S. v., & Groothuis-Oudshoorn, K. (2010). Mice: Multivariate imputation by chained equations in r. *Journal of statistical software*, 1–68.

C, V., & Wiseman, A. (2017). The center for effective lawmaking. https://thelawmake
Cai, J., De Janvry, A., & Sadoulet, E. (2015). Social networks and the decision to insure. *American Economic Journal: Applied Economics*, *7*(2), 81–108.
Caldarelli, G. (2007). *Scale-free networks: Complex webs in nature and technology*. Oxford University Press.
Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. M. (2006). *Measurement error in nonlinear models: A modern perspective*. CRC press.
Carroll, R., Lewis, J., Lo, J., McCarty, N., Poole, K., & Rosenthal, H. (2011). Dw-nominate scores with bootstrapped standard errors. *Available at: voteview. com/dwnomin. htm*.
Catanzaro, M., Caldarelli, G., & Pietronero, L. (2004). Assortative model for social networks. *Physical review e*, *70*(3), 037101.
Cattaneo, M. D. (2010). Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics*, *155*(2), 138–154.
Chakraborty, B., & Murphy, S. (2014). Dynamic treatment regimes. *Annual review of statistics and its application*, *1*, 447–464.
Chandrasekhar, A. (2016). Econometrics of network formation. *The Oxford Handbook of the economics of networks*, 303–357.
Chen, H., & Chen, T.-J. (2002). Asymmetric strategic alliances: A network view. *Journal of Business Research*, *55*(12), 1007–1013.
Chen, L. H. Y., & Shao, Q.-M. (2004). Normal approximation under local dependence. *The Annals of Probability*, *32*(3), 1985–2028. http://www.jstor.org/stable/3481601
Chin, A., Daysal, N. M., & Imberman, S. A. (2013). Impact of bilingual education programs on limited english proficient students and their peers: Regression discontinuity evidence from texas. *Journal of Public Economics*, *107*, 63–78.
Chua, L. O., & Yang, L. (1988). Cellular neural networks: Theory. *IEEE Transactions on circuits and systems*, *35*(10), 1257–1272.
Chuang, Y.-C., & Lin, C.-M. (1999). Foreign direct investment, r&d and spillover efficiency: Evidence from taiwan's manufacturing firms. *The Journal of Development Studies*, *35*(4), 117–137.
Cimini, G., Squartini, T., Saracco, F., Garlaschelli, D., Gabrielli, A., & Caldarelli, G. (2019). The statistical physics of real-world networks. *Nature Reviews Physics*, *1*(1), 58–71.
Clauset, A., Newman, M. E., & Moore, C. (2004). Finding community structure in very large networks. *Physical review E*, *70*(6), 066111.

Cockx, B., Lechner, M., & Bollens, J. (2019). Priority to unemployed immigrants? a causal machine learning evaluation of training in belgium. *arXiv preprint arXiv:1912.12864*.

Cohen, W. M., Goto, A., Nagata, A., Nelson, R. R., & Walsh, J. P. (2002). R&d spillovers, patents and the incentives to innovate in japan and the united states. *Research policy*, *31*(8-9), 1349–1367.

Coleman, D. (2008). The demographic effects of international migration in europe. *Oxford Review of Economic Policy*, *24*(3), 452–476.

Comment, L., Mealli, F., Haneuse, S., & Zigler, C. (2019). Survivor average causal effects for continuous time: A principal stratification approach to causal inference with semicompeting risks. *arXiv preprint arXiv:1902.09304*.

Concato, J., Shah, N., & Horwitz, R. I. (2000). Randomized, controlled trials, observational studies, and the hierarchy of research designs. *New England journal of medicine*, *342*(25), 1887–1892.

Contreras, M. G. A., & Fagiolo, G. (2014). Propagation of economic shocks in input-output networks: A cross-country analysis. *Physical Review E*, *90*(6), 062812.

Conyon, M. J., Muldoon, M. R. Et al. (2008). Ownership and control: A small-world analysis. *Advances in Strategic Management*, *25*, 31–65.

Cook, D. I., Gebski, V. J., & Keech, A. C. (2004). Subgroup analysis in clinical trials. *Medical Journal of Australia*, *180*(6), 289–291.

Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Skaaning, S.-E., Teorell, J., Altman, D., Bernhard, M., Cornell, A., Fish, M. S., Et al. (2018a). V-dem codebook v8.

Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Skaaning, S.-E., Teorell, J., Altman, D., Bernhard, M., Cornell, A., Fish, M. S., Et al. (2018b). V-dem codebook v8.

Cox, D. R. (1958). Planning of experiments.

Cranmer, S. J., & Desmarais, B. A. (2011). Inferential network analysis with exponential random graph models. *Political analysis*, *19*(1), 66–86.

Crimaldi, I., Forastiere, L., Mealli, F., & Tortú, C. (2020). The causal effect of immigration policy on income inequality. *Proceedings of the conference of the Italian Statistical Society(SIS 2020) (furthcoming)*.

Croxatto, H. B., Salvatierra, A. M., Croxatto, H. D., & Fuentealba, B. (1993). Effects of continuous treatment with low dose mifepristone throughout one menstrual cycle. *Human Reproduction*, *8*(2), 201–207.

Crucitti, P., Latora, V., & Porta, S. (2006). Centrality measures in spatial networks of urban streets. *Physical Review E*, *73*(3), 036125.

Davis, C. S. (1982). The distribution of a linear combination of chi square variables. JSTOR.

De Stefano, D., & Zaccarin, S. (2012). Exponential random graph model for multivariate networks: An application in knowledge network analysis. *XLVI Riunione Scientifica Società Italiana Di Statistica*.

de Heer, H. D., Koehly, L., Pederson, R., & Morera, O. (2011). Effectiveness and spillover of an after-school health promotion program for hispanic elementary school children. *American journal of public health*, *101*(10), 1907–1913.

Degenne, A., & Forsé, M. (1999). *Introducing social networks*. Sage.

Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and statistics*, *84*(1), 151–161.

Dekel, E., Jackson, M. O., & Wolinsky, A. (2008). Vote buying: General elections. *Journal of Political Economy*, *116*(2), 351–380.

Del Prete, D., Forastiere, L., & Leone Sciabolazza, V. (2019). Causal inference on networks under continuous treatment interference: An application to trade distortions in agricultural markets. *Available at SSRN 3363173*.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, *113*(3), 554–559.

Dempster, A. (1990). Causality and statistics. *Journal of statistical planning and inference*, *25*(3), 261–278.

Denzau, A. T., & Munger, M. C. (1986). Legislators and interest groups: How unorganized interests get represented. *The American Political Science Review*, 89–106.

Diermeier, D., & Myerson, R. B. (1999). Bicameralism and its consequences for the internal organization of legislatures. *American Economic Review*, *89*(5), 1182–1196.

DiMaggio, P. (1982). Cultural capital and school success: The impact of status culture participation on the grades of us high school students. *American sociological review*, 189–201.

DiMaggio, P., & Useem, M. (1978). The origins and consequences of class differences in exposure to the arts in america. *Theory and Society*, *5*(2), 141–161.

Dixit, A. (1996). Special-interest lobbying and endogenous commodity taxation. *Eastern Economic Journal*, *22*(4), 375–388.

Dixit, A., Grossman, G. M., & Helpman, E. (1997). Common agency and coordination: General theory and application to government policy making. *Journal of political economy*, *105*(4), 752–769.

Doove, L. L., Van Buuren, S., & Dusseldorp, E. (2014). Recursive partitioning for missing data imputation in the presence of interaction effects. *Computational Statistics & Data Analysis*, *72*, 92–104.

Drutman, L. (2011). The business of america is lobbying: Explaining the growth of corporate political activity in washington, dc.

Duchesne, P., & De Micheaux, P. L. (2010). Computing the distribution of quadratic forms: Further comparisons between the liu–tang–zhang approximation and exact methods. *Computational Statistics & Data Analysis*, *54*(4), 858–862.

Duflo, E., & Saez, E. (2003). The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. *The Quarterly journal of economics*, *118*(3), 815–842.

Eckles, D., Karrer, B., & Ugander, J. (2017). Design and analysis of experiments in networks: Reducing bias from interference. *Journal of Causal Inference*, *5*(1).

Efron, B. (1992). Bootstrap methods: Another look at the jackknife, In *Breakthroughs in statistics*. Springer.

Elliott, M., Golub, B., & Jackson, M. O. (2014). Financial networks and contagion. *American Economic Review*, *104*(10), 3115–53.

Erdős, P., & Rényi, A. (1959). On random graphs i. *Publicationes Mathematicae*, *6*(290-297), 18.

Falk, A., & Kosfeld, M. (2012). It's all about connections: Evidence on network formation. *Review of Network Economics*, *11*(3).

Falkowski, T., Bartelheimer, J., & Spiliopoulou, M. (2006). Mining and visualizing the evolution of subgroups in social networks, In *2006 ieee/wic/acm international conference on web intelligence (wi 2006 main conference proceedings)(wi'06)*. IEEE.

Fang, D., Chen, B., Hubacek, K., Ni, R., Chen, L., Feng, K., & Lin, J. (2019). Clean air for some: Unintended spillover effects of regional air pollution policies. *Science advances*, *5*(4), eaav4707.

Fellows, I., & Handcock, M. S. (2012). Exponential-family random network models. *arXiv preprint arXiv:1208.0121*.

for Lobbying, T. N. I., & Ethics. (n.d.). *What is lobbying?* https://lobbyinginstitute.com/what-is-lobbying/

Forastiere, L., Airoldi, E. M., & Mealli, F. (2020). Identification and estimation of treatment and interference effects in observational studies on networks. *Journal of the American Statistical Association*, (just-accepted), 1–49.

Forastiere, L., Lattarulo, P., Mariani, M., Mealli, F., & Razzolini, L. (2019a). Exploring encouragement, treatment, and spillover effects using principal stratification, with application to a field experiment on teens' museum attendance. *Journal of Business & Economic Statistics*, *0*(0), 1–15.

Forastiere, L., Lattarulo, P., Mariani, M., Mealli, F., & Razzolini, L. (2019b). Exploring encouragement, treatment, and spillover effects using principal stratification, with application to a field experiment on teens' museum attendance. *Journal of Business & Economic Statistics*, 1–15.

Forastiere, L., Mealli, F., & VanderWeele, T. J. (2016). 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., Mealli, F., Wu, A., & Airoldi, E. (2018). Estimating causal effects under interference using bayesian generalized propensity scores. *arXiv preprint arXiv:1807.11038*.

Fortunato, S. (2010). Community detection in graphs. *Physics reports*, *486*(3-5), 75–174.

Foster, J. C., Taylor, J. M., & Ruberg, S. J. (2011). Subgroup identification from randomized clinical trial data. *Statistics in Medicine*, *30*(24), 2867–2880.

Fouirnaies, A., & Hall, A. B. (2018). How do interest groups seek access to committees? *American Journal of Political Science*, *62*(1), 132–147.

Fouquin, M., Hugot, J. Et al. (2016). *Two centuries of bilateral trade and gravity data: 1827-2014* (tech. rep.). Universidad Javeriana-Bogotá.

Fowler, J. H. (2006a). Connecting the congress: A study of cosponsorship networks. *Political Analysis*, *14*(4), 456–487.

Fowler, J. H. (2006b). Legislative cosponsorship networks in the us house and senate. *Social Networks*, *28*(4), 454–465.

Friedman, J. H., Olshen, R. A., Stone, C. J., Et al. (1984). Classification and regression trees. *Belmont, CA: Wadsworth & Brooks*.

Fuller, W. A. (2009). *Measurement error models* (Vol. 305). John Wiley & Sons.

Furusawa, T., & Konishi, H. (2007). Free trade networks. *Journal of International Economics*, *72*(2), 310–335.

Gai, P., & Kapadia, S. (2010). Contagion in financial networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *466*(2120), 2401–2423.

Geddes, A., & Scholten, P. (2016). *The politics of migration and immigration in europe*. Sage.

Gilligan, M. J., Sergenti, E. J. Et al. (2008). Do un interventions cause peace? using matching to improve causal inference. *Quarterly Journal of Political Science*, *3*(2), 89–122.

Goldthorpe, J. H. (2001). Causation, statistics, and sociology. *European sociological review*, *17*(1), 1–20.

Grace, Y. Y. (2017). Statistical analysis with measurement error or misclassification strategy, method and application.

Graham, B. S. (2019). *Network data* (tech. rep.). National Bureau of Economic Research.

Grandjean, P., Budtz-Jørgensen, E., Keiding, N., & Weihe, P. (2004). Underestimation of risk due to exposure misclassification. *International journal of occupational medicine and environmental health*.

Granoveter, M. (1973). Thestrengthofweakties. *AmericanJournalofSociology*, *78*(6), 1360–1380.

Groseclose, T., & Snyder Jr, J. M. (1996). Buying supermajorities. *American Political Science Review*, 303–315.

Grossman, G. M., & Helpman, E. (1994). Protection for sale. *The American Economic Review*, 833–850.

Grotzer, T. (2012). Learning causality in a complex world: Understandings of consequence.

Guber, R. (2018). Instrument validity tests with causal trees: With an application to the same-sex instrument.

Gueyffier, F., Boutitie, F., Boissel, J.-P., Pocock, S., Coope, J., Cutler, J., Ekbom, T., Fagard, R., Friedman, L., Perry, M., Et al. (1997). Effect of antihypertensive drug treatment on cardiovascular outcomes in women and men: A meta-analysis of individual patient data from randomized, controlled trials. *Annals of Internal Medicine*, *126*(10), 761–767.

Guo, S., & Fraser, M. W. (2014). *Propensity score analysis: Statistical methods and applications* (Vol. 11). SAGE publications.

Hahn, P. R., Murray, J. S., Carvalho, C. M., Et al. (2020). Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects. *Bayesian Analysis*.

Halloran, M. E., & Struchiner, C. J. (1995). Causal inference in infectious diseases. *Epidemiology*, 142–151.

Hamby, D. M. (1994). A review of techniques for parameter sensitivity analysis of environmental models. *Environmental monitoring and assessment*, *32*(2), 135–154.

Hansen, B. B., & Bowers, J. (2008). Covariate balance in simple, stratified and clustered comparative studies. *Statistical Science*, 219–236.

Haynie, D. L. (2002). Friendship networks and delinquency: The relative nature of peer delinquency. *Journal of Quantitative criminology*, *18*(2), 99–134.

Helbling, M., Bjerre, L., Römer, F., & Zobel, M. (2017). Measuring immigration policies: The impic database. *European Political Science*, *16*, 79–98.

Helpman, E., & Persson, T. (2001). Lobbying and legislative bargaining. *The BE Journal of Economic Analysis & Policy*, *1*(1).

Hendrickson, B., Rosen, D., & Aune, R. K. (2011). An analysis of friendship networks, social connectedness, homesickness, and satisfaction levels of international students. *International journal of intercultural relations*, *35*(3), 281–295.

Hennessy, J., Dasgupta, T., Miratrix, L., Pattanayak, C., & Sarkar, P. (2016). A conditional randomization test to account for covariate imbalance in randomized experiments. *Journal of Causal Inference*, *4*(1), 61–80.

Hernán, M. A., & Robins, J. M. (2010). Causal inference. CRC Boca Raton, FL;

Hildebrandt, N., & McKenzie, D. J. (2005). *The effects of migration on child health in mexico*. The World Bank.

Hill, J. L. (2011). Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, *20*(1), 217–240.

Hirano, K., & Imbens, G. W. (2004). The propensity score with continuous treatments. *Applied Bayesian modeling and causal inference from incomplete-data perspectives*, *226164*, 73–84.

Hirano, K., Imbens, G. W., Rubin, D. B., & Zhou, X.-H. (2000). Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics*, *1*(1), 69–88.

Hoff, P. D. (2015). Dyadic data analysis with amen. *arXiv preprint arXiv:1506.08237*.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*(396), 945–960.

Hong, G., & Raudenbush, S. W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observa-

tional data. *Journal of the American Statistical Association*, *101*(475), 901–910.

Hoover, K. D. (2006). Causality in economics and econometrics. *Available at SSRN 930739*.

Hoover, K. D. (2012). Economic theory and causal inference. *Philosophy of economics*, *13*, 89–113.

Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, *47*(260), 663–685.

Howard, S., Omumbo, J., Nevill, C., Some, E., Donnelly, C., & Snow, R. (2000). Evidence for a mass community effect of insecticide-treated bednets on the incidence of malaria on the kenyan coast. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *94*(4), 357–360.

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

Imai, K., Jiang, Z., & Malai, A. (2018). Causal inference with interference and noncompliance in two-stage randomized experiments. *Unpublished Manuscript*.

Imai, K., Jiang, Z., & Malani, A. (2020). Causal inference with interference and noncompliance in two-stage randomized experiments. *Journal of the American Statistical Association*, (just-accepted), 1–39.

Imai, K., & Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, 243–263.

Imai, K., & Van Dyk, D. A. (2004). Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association*, *99*(467), 854–866.

Imai, K., & Yamamoto, T. (2010). Causal inference with differential measurement error: Nonparametric identification and sensitivity analysis. *American Journal of Political Science*, *54*(2), 543–560.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, *87*(3), 706–710.

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Iooss, B., & Lemaıtre, P. (2015). A review on global sensitivity analysis methods, In *Uncertainty management in simulation-optimization of complex systems*. Springer.

Jackson, C. K., Johnson, R. C., & Persico, C. (2015). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *The Quarterly Journal of Economics*, *131*(1), 157–218.

Jackson, M. O. (2010). *Social and economic networks*. Princeton university press.

Johansson, P., & Palme, M. (2002). Assessing the effect of public policy on worker absenteeism. *Journal of Human Resources*, 381–409.

Johnson, M., Cao, J., & Kang, H. (2019). Detecting heterogeneous treatment effect with instrumental variables. *arXiv preprint arXiv:1908.03652*.

Kalemli-Ozcan, S., Sorensen, B., Villegas-Sanchez, C., Volosovych, V., & Yesiltas, S. (2015). *How to construct nationally representative firm level data from the orbis global database: New facts and aggregate implications* (tech. rep.). National Bureau of Economic Research.

Kang, H., & Imbens, G. (2016). Peer encouragement designs in causal inference with partial interference and identification of local average network effects. *arXiv preprint arXiv:1609.04464*.

Keele, L., Malhotra, N., & McCubbins, C. H. (2013). Do term limits restrain state fiscal policy? approaches for causal inference in assessing the effects of legislative institutions. *Legislative Studies Quarterly*, *38*(3), 291–326.

Keely, C. B. (2000). Demography and international migration. *Migration Theory, Routledge*, 43–60.

Kelso, J. K., Milne, G. J., & Kelly, H. (2009). Simulation suggests that rapid activation of social distancing can arrest epidemic development due to a novel strain of influenza. *BMC public health*, *9*(1), 117.

Kempe, D., Kleinberg, J., & Tardos, É. (2003a). Maximizing the spread of influence through a social network. *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 137–146.

Kempe, D., Kleinberg, J., & Tardos, É. (2003b). Maximizing the spread of influence through a social network, In *Proceedings of the ninth acm sigkdd international conference on knowledge discovery and data mining*.

Kessler, D., & Krehbiel, K. (1996). Dynamics of cosponsorship. *American Political Science Review*, 555–566.

Kim, D., Hwong, A., Stafford, D., Hughes, D., O'Malley, J., Fowler, J., & Christakis, N. (2015). Social network targeting to maximise population behaviour change: A cluster randomised controlled trial. *Lancet*, *386*.

Kirkland, J. H. (2011). The relational determinants of legislative outcomes: Strong and weak ties between legislators. *The Journal of Politics*, *73*(3), 887–898.

Kisida, B., Greene, J. P., & Bowen, D. H. (2014). Creating cultural consumers: The dynamics of cultural capital acquisition. *Sociology of Education*, *87*(4), 281–295.

Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H., & Lipsitch, M. (2020). Projecting the transmission dynamics of sars-cov-2 through the postpandemic period. *Science*.

Kleinberg, S., & Hripcsak, G. (2011). A review of causal inference for biomedical informatics. *Journal of biomedical informatics*, *44*(6), 1102–1112.

Kondo, H., & Togari, A. (2011). Continuous treatment with a low-dose β-agonist reduces bone mass by increasing bone resorption without suppressing bone formation. *Calcified Tissue International*, *88*(1), 23–32.

Kosorok, M. R., & Laber, E. B. (2019). Precision medicine. *Annual Review of Statistics and Its Application*, *6*(1), 263–286.

Kovesdy, C. P., & Kalantar-Zadeh, K. (2012). Observational studies versus randomized controlled trials: Avenues to causal inference in nephrology. *Advances in chronic kidney disease*, *19*(1), 11–18.

Kramer, M. S., Moodie, E. E., Dahhou, M., & Platt, R. W. (2011). Breastfeeding and infant size: Evidence of reverse causality. *American journal of epidemiology*, *173*(9), 978–983.

Kreif, N., & DiazOrdaz, K. (2019). Machine learning in policy evaluation: New tools for causal inference. *arXiv preprint arXiv:1903.00402*.

Lattarulo, P., Mariani, M., & Razzolini, L. (2017). Nudging museums attendance: A field experiment with high school teens. *Journal of Cultural Economics*, *41*(3), 259–277.

Lechner, M. (2019). Modified causal forests for estimating heterogeneous causal effects.

Lee, K., Bargagli-Stoffi, F. J., & Dominici, F. (2020). Causal rule ensemble: Interpretable inference of heterogeneous treatment effects. *Working paper*.

Lee, K., Small, D. S., & Dominici, F. (2018). Discovering effect modification and randomization inference in air pollution studies. *arXiv preprint arXiv:1802.06710*.

Leung, M. P. (2020). Treatment and spillover effects under network interference. *The Review of Economics and Statistics*, *102*(2), 368–380. https://doi.org/10.1162/rest_a_00818

Leuven, E., Oosterbeek, H., & Van der Klaauw, B. (2010). The effect of financial rewards on students' achievement: Evidence from a randomized experiment. *Journal of the European Economic Association*, *8*(6), 1243–1265.

Lewbel, A. (2007). Estimation of average treatment effects with misclassification. *Econometrica*, *75*(2), 537–551.

Liben-Nowell, D., & Kleinberg, J. (2007). The link-prediction problem for social networks. *Journal of the American society for information science and technology*, *58*(7), 1019–1031.

Lindberg, S. I., Coppedge, M., Gerring, J., & Teorell, J. (2014). V-dem: A new way to measure democracy. *Journal of Democracy*, *25*(3), 159–169.

Linden, A., Uysal, S. D., Ryan, A., & Adams, J. L. (2016). Estimating causal effects for multivalued treatments: A comparison of approaches. *Statistics in Medicine*, *35*(4), 534–552.

Liu, L., & Hudgens, M. G. (2014). Large sample randomization inference of causal effects in the presence of interference. *Journal of the american statistical association*, *109*(505), 288–301.

Liu, L., Hudgens, M., & Becker-Dreps, S. (2016). On inverse probability-weighted estimators in the presence of interference. *Biometrika*, *103*(4), 829–842. https://doi.org/10.1093/biomet/asw047

Loh, W. W., Hudgens, M. G., Clemens, J. D., Ali, M., & Emch, M. E. (2020). Randomization inference with general interference and censoring. *Biometrics*, *76*(1), 235–245.

Long, S. J., Long, J. S., & Freese, J. (2006). *Regression models for categorical dependent variables using stata*. Stata press.

Lopez, M. J., Gutman, R. Et al. (2017). Estimation of causal effects with multiple treatments: A review and new ideas. *Statistical Science*, *32*(3), 432–454.

Lü, L., & Zhou, T. (2011). Link prediction in complex networks: A survey. *Physica A: statistical mechanics and its applications*, *390*(6), 1150–1170.

Manning, W. G., Newhouse, J. P., Duan, N., Keeler, E. B., & Leibowitz, A. (1987). Health insurance and the demand for medical care: Evidence from a randomized experiment. *The American economic review*, 251–277.

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

Manski, C. (2013). Identification of treatment response with social interactions. *Econometrics Journal*, *16*(1), S1–S23. https://doi.org/10.1111/j.1368-423X.2012.00368.x

Marsh, D., & Lewis, C. (2014). The political power of big business: A response to bell and hindmoor. *New Political Economy*, *19*(4), 628–633.

Mayer, A. K. (2011). Does education increase political participation? *The Journal of Politics*, *73*(3), 633–645.

Mayer, T., & Zignago, S. (2011). Notes on cepii's distances measures: The geodist database.

McCaffrey, D. F., Lockwood, J., & Setodji, C. M. (2013). Inverse probability weighting with error-prone covariates. *Biometrika*, *100*(3), 671–680.

Melitz, J., & Toubal, F. (2014). Native language, spoken language, translation and trade. *Journal of International Economics*, *93*(2), 351–363.

Menard, S. (2002). *Applied logistic regression analysis* (Vol. 106). Sage.

Messina, A. M. (2007). *The logics and politics of post-wwii migration to western europe*. Cambridge University Press.

Miles, C. H., Petersen, M., & van der Laan, M. J. (2019). Causal inference when counterfactuals depend on the proportion of all subjects exposed. *Biometrics*, *75*(3), 768–777.

Minhas, S., Hoff, P. D., & Ward, M. D. (2019). Inferential approaches for network analysis: Amen for latent factor models. *Political Analysis*, *27*(2), 208–222.

Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., & Bhattacharjee, B. (2007). Measurement and analysis of online social networks, In *Proceedings of the 7th acm sigcomm conference on internet measurement*.

Montes, F., Jaramillo, A. M., Meisel, J. D., Diaz-Guilera, A., Valdivia, J. A., Sarmiento, O. L., & Zarama, R. (2020). Benchmarking seeding strategies for spreading processes in social networks: An interplay between influencers, topologies and sizes. *Scientific Reports*, *10*(1), 3666.

Moore, B., & Rhodes, J. (1973). Evaluating the effects of british regional economic policy. *The Economic Journal*, *83*(329), 87–110.

Morgan, K. L., Rubin, D. B. Et al. (2012). Rerandomization to improve covariate balance in experiments. *The Annals of Statistics*, *40*(2), 1263–1282.

Munck, G. L., & Verkuilen, J. (2002). Conceptualizing and measuring democracy: Evaluating alternative indices. *Comparative political studies*, *35*(1), 5–34.

Muralidharan, K., & Sundararaman, V. (2015). The aggregate effect of school choice: Evidence from a two-stage experiment in india. *The Quarterly Journal of Economics*, *130*(3), 1011–1066.

Murphy, S. A. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*(2), 331–355.

Muthén, B., & Brown, H. C. (2009). Estimating drug effects in the presence of placebo response: Causal inference using growth mixture modeling. *Statistics in medicine*, *28*(27), 3363–3385.

Naranjo, A. J. (2010). Spillover effects of domestic law enforcement policies. *International Review of Law and economics*, *30*(3), 265–275.

Newman, M. E. (2001a). Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. *Physical review E*, *64*(1), 016132.

Newman, M. E. (2001b). The structure of scientific collaboration networks. *Proceedings of the national academy of sciences*, *98*(2), 404–409.

Newman, M. E. (2004). Analysis of weighted networks. *Physical review E*, *70*(5), 056131.

Nichol, K. L., Lind, A., Margolis, K. L., Murdoch, M., McFadden, R., Hauge, M., Magnan, S., & Drake, M. (1995). The effectiveness of vaccination against influenza in healthy, working adults. *New England Journal of Medicine*, *333*(14), 889–893.

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

Papadogeorgou, G., Mealli, F., & Zigler, C. M. (2019). Causal inference with interfering units for cluster and population level treatment allocation programs. *Biometrics*, *75*(3), 778–787.

Park, W. B., Kim, N.-H., Kim, K.-H., Lee, S. H., Nam, W.-S., Yoon, S. H., Song, K.-H., Choe, P. G., Kim, N. J., Jang, I.-J., Et al. (2012). The effect of therapeutic drug monitoring on safety and efficacy of voriconazole in invasive fungal infections: A randomized controlled trial. *Clinical infectious diseases*, *55*(8), 1080–1087.

Paster, T. (2018). How do business interest groups respond to political challenges? a study of the politics of german employers. *New Political Economy*, *23*(6), 674–689.

Pearl, J. Et al. (2009). Causal inference in statistics: An overview. *Statistics surveys*, *3*, 96–146.

Perotti, R. (2005). Estimating the effects of fiscal policy in oecd countries.
Persson, T. (1998). Economic policy and special interest politics. *The Economic Journal*, *108*(447), 310–327.
Petersen, M. L., Porter, K. E., Gruber, S., Wang, Y., & Van Der Laan, M. J. (2012). Diagnosing and responding to violations in the positivity assumption. *Statistical methods in medical research*, *21*(1), 31–54.
Polissar, L. (1980). The effect of migration on comparison of disease rates in geographic studies in the united states. *American journal of epidemiology*, *111*(2), 175–182.
Powell, E. N., & Grimmer, J. (2016). Money in exile: Campaign contributions and committee access. *The Journal of Politics*, *78*(4), 974–988.
Quattrociocchi, W., Scala, A., & Sunstein, C. R. (2016). Echo chambers on facebook. *Available at SSRN 2795110*.
Reinert, K. A., Rajan, R. S., Glass, A. J., & Davis, L. S. (2009). *The princeton encyclopedia of the world economy.(two volume set)* (Vol. 1). Princeton University Press.
Rice, S. A. (1927). The identification of blocs in small political bodies. *The American Political Science Review*, *21*(3), 619–627.
Richter, B. K., & Werner, T. (2017). Campaign contributions from corporate executives in lieu of political action committees. *The Journal of Law, Economics, and Organization*, *33*(3), 443–474.
Romer, T., & Snyder Jr, J. M. (1994). An empirical investigation of the dynamics of pac contributions. *American journal of political science*, 745–769.
Rosenbaum, P. R. (2002). Overt bias in observational studies, In *Observational studies*. Springer.
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. (1983a). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society: Series B (Methodological)*, *45*(2), 212–218.
Rosenbaum, P. R., & Rubin, D. B. (1983b). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55.
Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American statistical Association*, *79*(387), 516–524.

Routray, S. K., Sahin, G., da Rocha, J. R. F., & Pinto, A. N. (2015). Statistical analysis and modeling of shortest path lengths in optical transport networks. *Journal of Lightwave Technology*, *33*(13), 2791–2801.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, *66*(5), 688.

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

Rubin, D. B. (1986). Comment: Which ifs have causal answers. *Journal of the American Statistical Association*, *81*(396), 961–962.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American statistical Association*, *91*(434), 473–489.

Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys* (Vol. 81). John Wiley & Sons.

Rudolph, C. (2003). Security and the political economy of international migration. *American Political Science Review*, *97*(4), 603–620.

Rudolph, C. (2006). *National security and immigration: Policy development in the united states and western europe since 1945*. Stanford University Press.

Rungi, A., Morrison, G., & Pammolli, F. (2017). Global ownership and corporate control networks. *IMT Lucca EIC WP Series*, *7*.

Saito, K., Nakano, R., & Kimura, M. (2008). Prediction of information diffusion probabilities for independent cascade model, In *International conference on knowledge-based and intelligent information and engineering systems*. Springer.

Sandler, D., Wilcox, A., & Everson, R. (1985). Cumulative effects of lifetime passive smoking on cancer risk. *The Lancet*, *325*(8424), 312–315.

Saracco, F., Di Clemente, R., Gabrielli, A., & Squartini, T. (2015). Randomizing bipartite networks: The case of the world trade web. *Scientific reports*, *5*, 10595.

Sattar, N., & Preiss, D. (2017). Reverse causality in cardiovascular epidemiological research: More common than imagined? Am Heart Assoc.

Schmid, S. D., & Helbling, M. (2016). *Validating the immigration policies in comparison (impic) dataset* (tech. rep.). WZB Discussion Paper.

Schweitzer, F., Fagiolo, G., Sornette, D., Vega-Redondo, F., Vespignani, A., & White, D. R. (2009). Economic networks: The new challenges. *science*, *325*(5939), 422–425.

Segal, N. A., Hein, J., & Basford, J. R. (2004). The effects of pilates training on flexibility and body composition: An observational study. *Archives of physical medicine and rehabilitation*, *85*(12), 1977–1981.

Senn, S. (1989). Covariate imbalance and random allocation in clinical trials. *Statistics in Medicine*, *8*(4), 467–475.

Sewell, D. K., & Chen, Y. (2015). Latent space models for dynamic networks. *Journal of the American Statistical Association*, *110*(512), 1646–1657.

Shah, A. D., Bartlett, J. W., Carpenter, J., Nicholas, O., & Hemingway, H. (2014). Comparison of random forest and parametric imputation models for imputing missing data using mice: A caliber study. *American journal of epidemiology*, *179*(6), 764–774.

Shang, M.-S., Lü, L., Zhang, Y.-C., & Zhou, T. (2010). Empirical analysis of web-based user-object bipartite networks. *EPL (Europhysics Letters)*, *90*(4), 48006.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), 591–611.

Shu, D., & Yi, G. Y. (2019a). Causal inference with measurement error in outcomes: Bias analysis and estimation methods. *Statistical methods in medical research*, *28*(7), 2049–2068.

Shu, D., & Yi, G. Y. (2019b). Weighted causal inference methods with mismeasured covariates and misclassified outcomes. *Statistics in medicine*, *38*(10), 1835–1854.

Shumate, M., & Palazzolo, E. T. (2010). Exponential random graph (p*) models as a method for social network analysis in communication research. *Communication Methods and Measures*, *4*(4), 341–371.

Sinclair, B., McConnell, M., & Green, D. P. (2012). Detecting spillover effects: Design and analysis of multilevel experiments. *American Journal of Political Science*, *56*(4), 1055–1069.

Snyder Jr, J. M. (1991). On buying legislatures. *Economics & Politics*, *3*(2), 93–109.

Sobel, M. E. (2006a). What do randomized studies of housing mobility demonstrate? *Journal of the American Statistical Association*, *101*(476), 1398–1407.

Sobel, M. E. (2006b). What do randomized studies of housing mobility demonstrate? causal inference in the face of interference. *Journal of the American Statistical Association*, *101*(476), 1398–1407.

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), 20160003. https://doi.org/10.1515/jci-2016-0003

Sofrygin, O., & van der Laan, M. J. (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).

South, S. J., & Haynie, D. L. (2004). Friendship networks of mobile adolescents. *Social Forces*, *83*(1), 315–350.

Squartini, T., Fagiolo, G., & Garlaschelli, D. (2011). Randomizing world trade. i. a binary network analysis. *Physical Review E*, *84*(4), 046117.

Squartini, T., Picciolo, F., Ruzzenenti, F., & Garlaschelli, D. (2013). Reciprocity of weighted networks. *Scientific reports*, *3*(1), 1–9.

Stansfield, R. (2016). Reevaluating the effect of recent immigration on crime: Estimating the impact of change in discrete migration flows to the united kingdom following eu accession. *Crime & Delinquency*, *62*(11), 1426–1447.

Starling, J. E., Murray, J. S., Lohr, P. A., Aiken, A. R., Carvalho, C. M., & Scott, J. G. (2019). Targeted smooth bayesian causal forests: An analysis of heterogeneous treatment effects for simultaneous versus interval medical abortion regimens over gestation. *arXiv preprint arXiv:1905.09405*.

Stewart, C., & Woon, J. (2007). Congressional committee assignments data 1993-2007. Electronic files.

Stigler, G. J. (1971). The theory of economic regulation. *The Bell journal of economics and management science*, 3–21.

Stillman, S., McKenzie, D., & Gibson, J. (2007). *Migration and mental health: Evidence from a natural experiment*. The World Bank.

Strogatz, S. H. (2001). Exploring complex networks. *nature*, *410*(6825), 268–276.

Su, X., Kang, J., Fan, J., Levine, R. A., & Yan, X. (2012). Facilitating score and causal inference trees for large observational studies. *Journal of Machine Learning Research*, *13*(Oct), 2955–2994.

Tchetgen Tchetgen, E. J. (2014). Identification and estimation of survivor average causal effects. *Statistics in medicine*, *33*(21), 3601–3628.

Tchetgen, E. J. T., & VanderWeele, T. J. (2012). On causal inference in the presence of interference. *Statistical methods in medical research*, *21*(1), 55–75.

Teseo, E. (2020). What drives us corporate elites' campaign contribution behavior?

Tortù, C., Forastiere, L., Crimaldi, I., & Mealli, F. (2020). Modelling network interference with multi-valued treatments: The causal effect of immigration policy on crime rates. *arXiv preprint arXiv:2003.10525*.

Traud, A. L., Mucha, P. J., & Porter, M. A. (2012). Social structure of facebook networks. *Physica A: Statistical Mechanics and its Applications*, *391*(16), 4165–4180.

Truman, D. B. (1959). *The congressional party, a case study*. Wiley.

Valdes, G., Luna, J. M., Eaton, E., Simone II, C. B., Ungar, L. H., & Solberg, T. D. (2016). Mediboost: A patient stratification tool for interpretable decision making in the era of precision medicine. *Scientific reports*, *6*, 37854.

Valente, T. W. (2012). Network interventions. *Science*, *337*(6090), 49–53.

Van der Waerden, B. (1952). Order tests for the two-sample problem and their power, In *Indagationes mathematicae (proceedings)*. Elsevier.

van der Laan, M. J. (2014). Causal inference for a population of causally connected units. *Journal of Causal Inference J. Causal Infer.*, *2*(1), 13–74.

Van Dijk, B. (2013). Orbis database. *URL: https://orbis. bvdinfo. com*.

Vanderweele, T. J. (2012). Inference for additive interaction under exposure misclassification. *Biometrika*, *99*(2), 502–508.

VanderWeele, T. J., & Christakis, N. A. (2019). Network multipliers and public health. *International Journal of Epidemiology*, *48*(4), 1032–1037.

VanderWeele, T. J., & Tchetgen, E. J. T. (2011). Effect partitioning under interference in two-stage randomized vaccine trials. *Statistics & probability letters*, *81*(7), 861–869.

VanderWeele, T. J., Tchetgen, E. J. T., & Halloran, M. E. (2014). Interference and sensitivity analysis. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *29*(4), 687.

Victora, C. G., Habicht, J.-P., & Bryce, J. (2004). Evidence-based public health: Moving beyond randomized trials. *American journal of public health*, *94*(3), 400–405.

Vogel, D. (2003). *Fluctuating fortunes: The political power of business in america*. Beard Books.

Volden, C., Wai, J., & Wiseman, A. E. (2020). Elite education and legislative behavior in the us congress.

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *113*(523), 1228–1242.

Wang, C., Chen, W., & Wang, Y. (2012). Scalable influence maximization for independent cascade model in large-scale social networks. *Data Mining and Knowledge Discovery*, *25*(3), 545–576.

Wang, G.-J., Xie, C., He, K., & Stanley, H. E. (2017). Extreme risk spillover network: Application to financial institutions. *Quantitative Finance*, *17*(9), 1417–1433.

Wang, G., Li, J., & Hopp, W. J. (2017). Personalized health care outcome analysis of cardiovascular surgical procedures. *Ross School of Business Paper*, (1336).

Wang, G., Li, J., & Hopp, W. J. (2018). An instrumental variable tree approach for detecting heterogeneous treatment effects in observational studies. *Available at SSRN 3045327*.

Wang, P., Robins, G., & Matous, P. (2016). Multilevel network analysis using ergm and its extension, In *Multilevel network analysis for the social sciences*. Springer.

Wang, Y. J., & Wong, G. Y. (1987). Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association*, *82*(397), 8–19.

Wiedermann, W., & Von Eye, A. (2016). *Statistics and causality*. Wiley Online Library.

Wilson, N., Sloper, K., & Silverman, M. (1995). Effect of continuous treatment with topical corticosteroid on episodic viral wheeze in preschool children. *Archives of disease in childhood*, *72*(4), 317–320.

Wilson, R. K., & Young, C. D. (1997). Cosponsorship in the us congress. *Legislative Studies Quarterly*, 25–43.

Xu, K. S., & Hero, A. O. (2014). Dynamic stochastic blockmodels for time-evolving social networks. *IEEE Journal of Selected Topics in Signal Processing*, *8*(4), 552–562.

Yanagi, T. (2018). Inference on local average treatment effects for misclassified treatment. *Available at SSRN 3065923*.

Yang, J., & Zhou, Y. (2013). Credit risk spillovers among financial institutions around the global credit crisis: Firm-level evidence. *Management Science*, *59*(10), 2343–2359.

Yang, S., Imbens, G. W., Cui, Z., Faries, D. E., & Kadziola, Z. (2016). Propensity score matching and subclassification in observational studies with multi-level treatments. *Biometrics*, *72*(4), 1055–1065.

Zaccarin, S., & Rivellini, G. (2010). Modelling network data: An introduction to exponential random graph models, In *Data analysis and classification*. Springer.

Zhang, W., Le, T. D., Liu, L., Zhou, Z.-H., & Li, J. (2017). Mining heterogeneous causal effects for personalized cancer treatment. *Bioinformatics*, *33*(15), 2372–2378.

Zhao, J., Runfola, D. M., & Kemper, P. (2017). Quantifying heterogeneous causal treatment effects in world bank development finance projects, In *Joint european conference on machine learning and knowledge discovery in databases*. Springer.

Zhou, T., Lü, L., & Zhang, Y.-C. (2009). Predicting missing links via local information. *The European Physical Journal B*, *71*(4), 623–630.

Zhou, T., Ren, J., Medo, M., & Zhang, Y.-C. (2007). Bipartite network projection and personal recommendation. *Physical review E*, *76*(4), 046115.

Zigler, C. M., & Papadogeorgou, G. (2018). Bipartite causal inference with interference. *arXiv preprint arXiv:1807.08660*.