Estimating Causal Effects on the Formation of Network Structures

by
Siddarth Viswanathan

Advised by
Edoardo Airoldi and
Alexander D'Amour

Presented to the Department of Statistics in
Partial Fulfillment of the Requirements for a
Bachelor of Arts Degree with Honors

Harvard College
Cambridge, Massachusetts
April 2014

Abstract

Causal inference plays an important role in shaping decisions made by a variety of organizations and in aiding discoveries in the computational, natural, and social sciences. A recently developing literature combines causal inference and network science by leveraging techniques from causal inference to estimate micro-level node causal effects from a variety of observational network data. We introduce a potentially ripe area of focus in the causal-inference-on-networks field by presenting a matching-based methodology for estimating macro-level causal effects on generating mechanisms of entire network structures. Our method is suitable for observational studies assessing the impact of interventions on interaction propensities between nodes in networks. We define the idea of sparsity for finite-population networks and show that not accounting for network sparsity biases causal estimates. To overcome this issue we propose a matched differences-in-differences estimator which adjusts for network sparsity. The matching portion of our method presents matching on pairs of nodes (dyads) instead of individual nodes, introduces network structure as an important matching covariate, and discusses quasi-exact matching on time for optimal matching on dyads. We empirically illustrate our method with a longitudinal co-patent dataset by harnessing a natural experiment in Michigan to estimate the impact of non-compete laws on inventor collaboration propensities.

Contents

List of Figures

1 Introduction and Background

A public health official is interested in assessing the impact of an HIV education awareness program on the propensity for sexual interaction in an HIV-infested South African village. The official believes that reducing the propensity for sexual interaction among villagers can effectively contain the spread of HIV and therefore tailors the education program to this goal. Suppose a massive South African government study provides complete sexual interaction survey data, leading to a sexual network, for a number of villages over a time period which includes the education program implementation. Precisely estimating the impact of the education program on the sexual interaction propensity within the village network of interest could lend support to implementation, or removal, of the program across other South African villages. However, current statistics and econometrics literatures have not yet developed observational data methods specifically for estimating the impact of programs on network structure formation parameters, such as interaction propensity. These methods would be useful for analyzing the increasing number of publicly available longitudinal (time-series) social network datasets including disease networks, academic co-citation networks, and Twitter and Facebook networks. This thesis develops a method to infer from such datasets valuable causal estimates of treatments (such as the education program) on network structure formation parameters (such as sexual interaction propensity). These causal estimates, as they potentially deal with large populations over long time periods, can significantly impact decisions made by a variety of organizations including schools, businesses, and governments[1].

1.1 A Birds-Eye View

If given village sexual network data describing all sexual interactions a few years before and after the introduction of an education program to a single village, what steps should a researcher take to assess the program's impact on the sexual interaction propensity within the village network of interest? Ideally, the researcher would have run a randomized experiment by randomly selecting treatment and control networks of villagers and introducing the education program to only the treatment network. Comparing each experimental group's

[1] We note that the method we introduce, and the causal-inference-on-networks field in general, requires more development before such causal estimates are reliable.

sexual interaction propensity could then provide an unbiased estimate of the education program's impact. Such randomized experiments, though the gold standard of social science research, are often expensive or unethical and therefore observational data techniques must be developed.

Since the data is longitudinal the researcher knows that the treatment effect (impact of the education program) will be confounded by other events which happen before or after the time of treatment implementation. She therefore decides to control for these confounding events by comparing the village which experienced the education program treatment, with a separate but similar village which did not experience the treatment. She calls the village of interest the treatment group and the separate village the control group. After assuming that these potentially confounding events impact both villages similarly she can now attribute the change in treatment group sexual interaction propensity to the education program alone. The researcher is satisfied with this quasi-experimental design[2] and to proceed with analysis must first decide on the unit of analysis.

The researcher considers summarizing village sexual interaction propensity by calculating the proportion of observed interactions to the number of potential interactions[3] and therefore realizes that *the unit of analysis is the interaction*, which is characterized by the individual-pair, or *dyad*. But since there are numerous potential pairs of interactions between the villagers[4] the researcher must decide whether to use all of these potentially interacting dyads in her analysis or only those dyads which are *at risk of interacting*. Essentially the question is whether to analyze dyads which are not at risk of sexual interaction, e.g., family members, or those too distant in age or location. Analysis performed using dyads not at risk of interaction will downwardly bias the causal estimate; it is therefore crucial to analyze only those dyads at risk of interacting.

However, social networks, such as sexual interaction networks, are known to be sparse. Network sparsity implies that not every dyad is at risk of an interaction[5]. As social networks become larger they also become sparser since it is less likely that all pairs of individuals

[2] This design implies a differences-in-differences estimator to adjust for group-level differences.

[3] We develop a more practical summary of interaction propensity in Section 5 but use this example here for illustration purposes.

[4] If there are N villagers then there are $\binom{N}{2}$ potential interactions between the villagers

[5] Interactions can take many different forms e.g., collaborating on homework, friendships or sexual interactions

are at risk of interaction. The researcher must therefore deal with the impossible task of defining which dyads are considered at risk of interaction. *Without knowing the complete social scenario of how likely each dyad in the network is to interact, the researcher is unable to select which dyads are at risk of interaction* and therefore cannot successfully estimate the causal effect. We propose an estimator which overcomes this dyad selection problem when dealing with sparse network data.

Supposing the researcher is able to overcome the dyad selection problem described above there is still another major problem preventing successful estimation of the education program effect. The treatment and control group villages must contain dyads with similar properties and structures. For example, the control group villagers may be on average more social than the experimental village and therefore intrinsically more likely to interact. By selecting dyads in the treatment and control group villages which are individually and structurally similar, the researcher adjusts for estimate-biasing village differences and essentially performs analysis on a control village equally as social as the treatment village. To adjust for group-level dyad differences the researcher decides to use the popular technique of matching dyad variables, called covariates. However, the researcher cannot proceed since *the matching literature does not address two crucial issues: i) selecting dyad covariates to control for network structure ii) matching on dyads.* We address these matching-related issues. With the ideas proposed in this thesis, given an appropriate dataset, the researcher could estimate the impact of the education program on village sexual interaction propensity. How to best use and how much to trust this estimate are more difficult questions which we hand off to future research.

To develop a methodology for estimating causal effects on dyad interaction propensity this thesis addresses the following issues:

1. *Selecting the set of nodes of the network object* (Section 3.1). Since our networks change over time it is essential to fix the set of nodes. Restricting the node set to a fixed sample makes causal effects estimation feasible with the inference approaches currently available.
2. *Defining the estimand* (Section 5.1). What are we interested in knowing? We are interested in measuring the extent to which a treatment impacts the interaction propensity between dyads at risk of interaction.

3. *Defining sparsity for finite-population networks* (Section 4.2). We define sparse networks as those where the risk set is not complete (not all dyads are at risk) and therefore where the social scenario is unknowable.
4. *Defining the estimator and inference procedure* (Section 5.3). Our definition of network sparsity implies the need for an estimator which is independent of the social scenario and thus independent of the true risk set. We develop a differences-in-differences estimator to adjust for network sparsity and group-level differences.
5. *Selecting dyad sets to perform matching and estimation over* (Section 6.2). Select dyads for matching which are known to be at risk of interaction (those observed to have an interaction).
6. *Selecting covariates to perform matching over* (Section 6.3). Select treatment and control group structural, edge, and node covariates to control for variables impacting the interaction propensity outcome.
7. *Matching on pre-treatment covariates between experimental and control groups* (Section 6.3). Use matching to restrict dyad sets in the treatment and control groups to those with similar covariate distributions. Matching reduces bias of the differences-in-differences estimator of the treatment effect.
8. *Estimating the causal treatment effect* (Section 5.3). Implement the matched differences-in-differences estimator using the inference method from Step 4.

This thesis integrates ideas from two distinct fields, network science and causal inference. The following section provides background for both fields.

1.2 Background

Network science and causal inference are rapidly growing fields with powerful applications to the natural, computational, and social sciences. Networks are complicated web-like structures which are useful for modeling systems consisting of links (called edges) between units (called nodes). For instance, the World Wide Web is a network where webpages are linked by hyperlinks; sociologists and economists often model human interaction using social networks where human nodes are linked by relationships; geneticists often view gene regulatory pathways as networks of interactions between DNA segments and other cell materials. The

success of networks in describing a variety of systems has prompted investigation of the mechanisms by which network structures form and evolve (Albert & Barabasi 2002). This thesis develops a methodology to estimate the causal impact of treatments on the propensity for edge formation within networks using the powerful mathematical techniques and conceptual frameworks of causal inference. Since network structure consists of the collection of edges between nodes, *interaction propensity*[6] *can be viewed as a parameter representing how the network structure forms and evolves.*

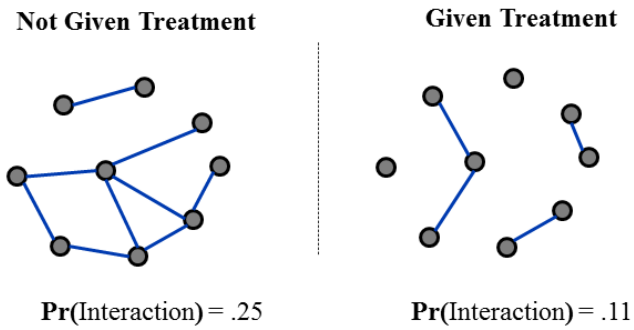


Figure 1.1: The relationship between interaction propensity and network structure. In this figure the treatment has lowered dyad interaction propensity and thus the number of observed edges. Note that **Pr**(Interaction) is not the only measure of interaction propensity, but is used here for illustration purposes. We propose an estimator in Section 5 which summarizes interaction propensity for dynamic networks using a regression model.

Causal inference aims to show that a treatment (e.g., a pill or policy) is not just correlated with an outcome (e.g., a patient's health or economic growth) but causes it. Causal questions are important for a variety of problems in fields ranging from epidemiology to economics: pharmaceutical companies and government agencies may be concerned about whether a drug causes the requisite change in a patient's health; economists may be concerned about whether a certain policy causes the desired change in unemployment rates. Causal inference techniques, originally developed in the 1970s by Donald Rubin and collaborators, have recently experienced an explosion of further development in economics

[6] For dynamic networks interaction propensity is best represented as the instantaneous probability of an interaction between two nodes. Given two nodes i and j, where $Y_{ij}(t)$ represents a counting process of the number of interactions between nodes i and j at time t, the instantaneous probability of interaction is given by $[\mathbf{Pr}(Y_{ij}(t+\epsilon) - Y_{ij}(t)) = 1]/\epsilon$. This instantaneous probability is difficult to calculate, so we consider alternative summaries of interaction propensity in Section 5.

(Imbens, 2004), political science (Ho et al., 2007), and statistics (Rosenbaum, 2002) among other fields. In this thesis we utilize the Rubin Causal Model (RCM) which views causation as a comparison of a unit's potential outcomes[7]. Potential outcomes are the outcomes of interest for a unit under different levels of treatment. For instance, the potential outcomes in a simple clinical trial are the observed health of a patient if given the placebo or the observed health of the same patient if given the drug[8]. The "fundamental problem of causal inference" (Holland, 1986) is that only one of these outcomes can be observed for each unit. Causal inference is therefore a missing data problem (Rubin, 1976) where the unobserved potential outcomes must be predicted to estimate the treatment effect.

The RCM, when used for both experimental and observational studies relies on a crucial assumption without which causal inference is difficult (Rubin 1974), the stable unit treatment value assumption (SUTVA): treatments are identical and the treatment status of a unit does not interfere with the potential outcomes of other units. To illustrate, SUTVA would be violated if there are various intensities of a drug in a clinical trial, or if unit a's health when given the placebo is somehow impacted by unit b's being given the drug[9]. Though both causal inference and network science are burgeoning fields, their intersection, causal inference on networked populations, is still in its infancy, especially when dealing with observational studies.

This thesis demonstrates the potential power of matching for improving causal effects estimation on edge formation propensities from sparse observational network data. Matching is a commonly used method for reducing bias in observational studies. Randomized experiments are preferred to observational studies when measuring causal effects since all latent and observed covariates are known to be only randomly different between treatment and control groups; thus, all difference in the outcome of interest between treatment and

[7] We note the existence of another framework to think about causation–the causal graphical models approach (Pearl, 2000). This approach is not ideal for our study since it is easier to formulate a clear causal effect in scientifically interpretable terms with the RCM. Further, the RCM does a clean job of identifying the causal problem for finite-populations while it is more difficult to think about finite population causal effects using Pearl's approach.

[8] In the birds-eye-view example, the two potential outcomes would be the sexual interaction propensity between a dyad if the dyad experienced the education program and if the same dyad did not experience the education program.

[9] In the birds-eye-view example, SUTVA would be violated if a dyad that experiences the education program influences the sexual interaction propensity of other dyads. Further research is required to understand the plausibility of SUTVA when dyads are the unit of analysis.

control groups can be causally attributed to the treatment. When it is not feasible to run randomized experiments researchers must rely on observational studies to make causal estimates. Matching aims to replicate the randomized experiment setting for observational studies by balancing the covariate distributions of treatment and control groups (Stuart, 2010). Section 6 includes a review of matching and illustrates its advantages over other covariate-adjustment techniques for use in our methodology.

1.3 Motivation and Contributions

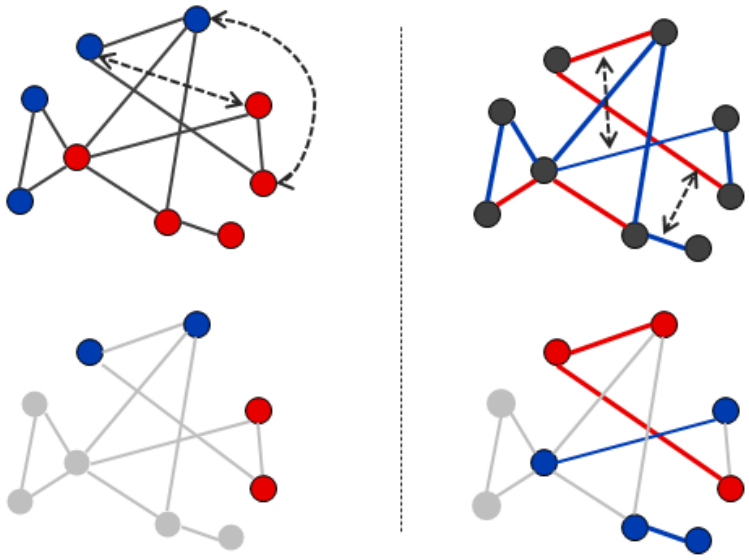
VanderWeele & An (2013) note that nearly all causal analysis of networked populations has focused on estimating causal effects on individual nodes, e.g., estimating how an individual's propensity to adopt a mobile application is impacted by the introduction of the mobile application to this individual's friend group. There are difficulties preventing convincing causal inference at the node-level of which the most important is handling interference between interacting nodes (VanderWeele & An, 2013)[10]. Solving this problem is currently a main focus of causal-inference-on-networks research under the broad heading 'causal inference with interference', which is difficult since interference between units violates SUTVA. Matching, a potential-outcomes-based method, has been proposed as a technique to control for interference effects such as homophily and peer effects; however, a majority of studies performing causal inference on networks focus on producing regression-type estimates based only on realized outcomes. Imbens & Woolridge (2009) note a number of advantages of the potential outcomes framework over a realized outcomes framework[11]. *Therefore the currently dominant regression approach for estimating causal effects on networked populations leaves room for more careful studies fully elucidating the potential outcomes approach.* Our work uses the potential outcomes framework to propose an area of focus within the causal-inference-on-networks field which is not hindered by the difficulty of identifying and

[10] For example, in a social network it is difficult to pinpoint a perceived outcome to a specific treatment when there are many environmental (e.g., company or classroom-specific variables) and peer effects (e.g., influence by friends) that are interfering with the outcome of interest.

[11] The potential outcomes framework: i) allows for definition of causal effects without reliance on any modeling assumptions ii) clarifies uncertainty of estimators iii) makes interpretation of causal estimates simpler.

removing confounding node-level interference effects [12]. *Whereas previous attempts at causal inference on networks have focused on estimating causal effects on nodes with the network structure given, we propose a method for estimating causal effects on the formation of network structures with node properties given.*

In non-networked observational studies and current node-level network studies, matching is done at the unit level, e.g., people are matched with people. In studies estimating causal effects on interaction propensities as we do, matching should be done at the dyad level, e.g., pairs of people are matched with pairs of people. A dyad is a two-node pair that *may or may not* be connected by a relationship. Causal studies utilizing matching at the node level match similar nodes together to see how a treatment impacts a sample of nodes, e.g., testing how a drug impacts the health of a group of patients. Causal studies at the dyad-level are analogously interested in seeing how a treatment impacts a sample of dyads. Since the structure of a network is merely a collection of dyads, by matching on dyads we are able to examine causal effects on network structure formation.



[12] Dyad-level interference is present on the network structures we consider. For this thesis we make the strong assumption that SUTVA can still hold even if dyads are the unit of analysis. The impact of dyad-level interference on causal estimates must be further studied.

Figure 1.2: (Left) Matching on nodes. Red and blue nodes represent treatment and control group nodes respectively. Black dashed lines represent matches done based on node covariates. On the bottom-left we see that analysis is performed using only those nodes which were matched, grayed portions are excluded from the analysis. (Right) Matching on dyads. Red and blue edges represent treatment and control group dyads respectively. Dyad covariates include node, edge, and structural covariates (explained further in Section 6.3). Black dashed lines represent matches from comparing dyad covariates. Post-matching analysis is done on the matched dyads with grayed edges and nodes excluded.

Potential applications of our method include estimating:

- how a new elementary school policy impacts the propensity for friendship formation within the school.
- how an IPO of a Facebook-competitor impacts the propensity for tie (friendship) formation within the Facebook network .
- how a financial crisis impacts the propensity for nations to invest in each other in the international financial network.
- how an HIV education program impacts the propensity for sexual interactions within members of a South African village sexual network.

The sparseness of many real-world networks has recently been identified as a major problem hindering successful estimation of interaction-related estimands (D'Amour & Airoldi, 2014). Network sparsity has been formally defined for the superpopulation (infinite node) context, but not yet in the finite-population context. Since our method is for use with observational data and hence finite populations of nodes, in Section 4.2 we rigorously define sparsity in the finite-population context. This definition is inspired by the idea that in sparse networks, such as social networks, not all dyads are at risk of interaction and therefore without knowing the social scenario it is impossible to specify which dyads are actually at risk of interaction. We present an estimator which is independent of this often unobservable social scenario and which therefore adjusts for network sparsity.

We can now state this thesis's contributions:

1. Using the potential outcomes framework we introduce an observational data methodology for estimating causal effects on interaction propensities and therefore network structures.

2. We illustrate the suitability of a novel class of network models[13] for causal effects estimation on sparse networks. Using this model framework we present a causal estimand for finite-population networks which we term the effect of the treatment on the treated at risk of interaction (ETTR). We define sparsity in the finite-population context to motivate the construction of a sparsity-adjusted DID estimator for this estimand.
3. To our knowledge, we present the first use of matching on dyads instead of nodes. We discuss matching across time to further reduce estimate bias and also present the first instance of using network structure as a matching covariate.
4. We present the first empirical study estimating causal effects on interaction propensities within networks.

This thesis contributes towards the distant yet important goal of trustworthy causal inference from observational longitudinal network data. Achieving precise causal inference on networks can have enormous implications. Gutman & Rubin (2012) state, "analyses that inform policy decisions are, *de facto* causal." Social media such as Facebook, Yelp, and Twitter are one example of numerous rapidly growing massive networked datasets which can yield valuable information such as the topological structure of society, the diffusion of information through communities, and the spread of disease through society. With the rise of numerous social network datasets, academics, businesses, and policy-makers will demand stronger methods for estimating causal effects on sparsely networked populations. This thesis endeavors to develop a method, which with further development, can potentially be a powerful tool for the investigator interested in examining causal effects on the formation of network structures, especially for sparse networks[14].

1.4 Outline

The remainder of the thesis is organized as follows: Section 2 reviews the recent causal-inference-on-networks literature and positions our method in context. Section 3 presents the

[13] These models are called Conditionally Independent Relationship (CIR) models (D'Amour & Airoldi, 2014) and are formally introduced in Section 5.

[14] Causal inference on networked populations is a young field reliant on many strong assumptions, and we therefore recommend that investigators interested in using our method interpret resulting causal estimates cautiously.

motivating co-patent dataset and natural experiment we harness to illustrate our method. Section 4 through 6 present the theoretical and methodological contributions of this thesis. Section 7 presents an empirical contribution. Section 4 reviews the Rubin Causal Model and provides a definition of sparsity for finite-population networks motivated by a recently developed class of network models. Section 5 presents our estimand and sparsity-adjusted differences-in-differences estimator and inference approach. Section 6 reviews the relevant literature on matching and presents matching on dyads. Section 7 empirically illustrates our method using the longitudinal co-patent dataset discussed in Section 3. Section 8 presents the limitations of our work and discusses necessary future research.

1.5 A Note on Undirected Networks

The networks considered in this thesis are undirected meaning edges between two nodes are identical for both nodes. Undirected edges include friendships and co-authorships, while directed edges include citations and internet hyperlinks. To aggregate all dyad information we therefore can take sums or products over pairs of nodes in the network. These sums or products are written over ij where i and j are nodes. Since the networks considered are undirected, this notation in actuality means sums or products over $i < j \leq n$, if n is the size of the node set, in order to avoid double counting dyads. Further, we use the terms tie, edge, link, and interaction interchangeably, and the terms node, ego, and vertex interchangeably.

2 A Review of the Causal-Inference-on-Networks Literature

VanderWeele & An (2013) note three kinds of network effects: relational effects, positional effects, and structural effects. Matching has recently been used in relational effect studies which address micro-level network effects such as how a treatment on a node's neighbors impacts the node's outcome (Aral et al., 2011; Carmi et al., 2012; Christakis & Fowler, 2007; Hill et al., 2006). There has been little work with observational studies for testing positional and structural effects of networks. Positional effects refer to a change in a node-level outcome caused by changing a node's position in a network while structural effects examine how treatments impact parameters of network structure such as a network's

edge generating propensity. There has been recent experimental work estimating causal structural network effects (Centola 2010), but a matching-based method has not yet been proposed for testing structural network effects using observational data. We now review the literature on matching for causal inference on node-level network effects to position our method in context.

Though network ideas have been pervasive since the 1960s when Stanley Milligram (1967) identified the "six degrees of separation" concept, the first notable use of matching on networked populations appeared only in 2009 by Aral et al. This study attempts to distinguish contagion effects (i.e., whether one person influences another) from homophily effects (i.e., whether similar people are more likely to influence each other). Aral et al. examine the impact of the adoption of Go, a mobile application, by an individual's friends on the adoption decisions of the individual over restricted time frames. Those individuals with friends adopting Go within a specified time frame are considered assigned to treatment while individuals without friends adopting Go within this time frame are considered assigned to control. Aral et al. utilize propensity score matching to match treated and untreated nodes on 46 covariates. They argue that impacts from homophily are removed by matching and then attribute the estimated treatment effect (difference in product adoption rates) to contagion effects[15].

A small literature inspired by Aral et al. (2009) uses matching on observational network data for testing node-level causal effects. Carmi et al. (2012) utilize propensity score matching to test the impact of a book review by Oprah on demand for related books. Hill et al. (2006), like Aral et al., use matching to remove homophily effects to support the hypothesis that peer effects directly impact product adoption. Oh et al. (2011) match on the network position of social media users to study what causes runaway YouTube video hits. Why has matching been so rarely used for causal inference on node-level network

[15] Aral et al. take the network structure as given by performing analysis on fixed 14-day interval samples of the network; in their analysis they therefore do not account for differences in network structures between 14-day intervals. As a result, node covariates, such as number of friends, are only defined for the 14-day interval and do not more accurately consider larger time periods. Holding network structure fixed in this way therefore biases the final estimate but must be done to make analysis simpler. It is possible to simultaneously adjust for changes in network structure within 14-day intervals, but the analysis would be far more difficult. Analogously our method requires that we consider the node set as given across the analysis time interval. We therefore hold the node set constant and calculate changes in interaction propensities for this fixed node set–the procedure for selecting the fixed node set is described in Section 3.1. Though by ignoring new node entries across time the estimate may be potentially biased, inference becomes far simpler. An extension of our method would allow for dynamic node sets.

effects?

Causal inference for networks is difficult mainly because interactions between nodes violate the stable unit treatment value assumption (SUTVA) that the response of a particular unit is only dependent on the treatment of the particular unit and not the treatments of surrounding units. A growing literature has focused on developing causal inference with interactions between units (Hudgens & Halloran, 2008), but even in basic social network settings with small amounts of nodes, causal inference with interacting units is challenging and generally not successful (Tchetgen & VanderWeele, 2012). VanderWeele (2013) states "when outcomes ... are associated, it is essentially impossible to distinguish whether this is due to social influence, homophily, or environmental confounding." Though matching has been proposed as a viable solution to adjust for homophily, social influence, and environmental confounding, a promising direction in the network modeling literature is seen as more useful for combating these issues. Recent work, which develops pathbreaking work by Snijders (2001), uses stochastic actor-oriented models that model homophily and node interaction effects directly. These models make the strong assumption that actors are fully knowledgeable about their surrounding networks but have been successful for modeling interaction data (see Perry & Wolfe, 2011)[16].

Since we are concerned with dyad interaction propensity, a parameter governing edge-formation of the entire network, our analysis is less affected by problematic node-specific issues of homophily and peer influence. Homophily and peer influence confound node-level casual estimates since they are individual-specific. We make the key observation that the treatment effects we are interested in summarize features of the *entire* network and therefore are less impacted by *individual-specific* confounding factors. Aggregate node-level homophily and peer influence effects therefore impact our analysis, but are not confounding as in the studies reviewed above. Estimators summarizing entire features of networks, such as interaction propensity, are essentially summaries of various aggregate node-level effects; these estimators are therefore less affected by confounding node-level effects.

[16] As stated earlier these models focus on producing regression-type estimates using realized outcomes and therefore leave room for more careful analysis using a potential outcomes framework.

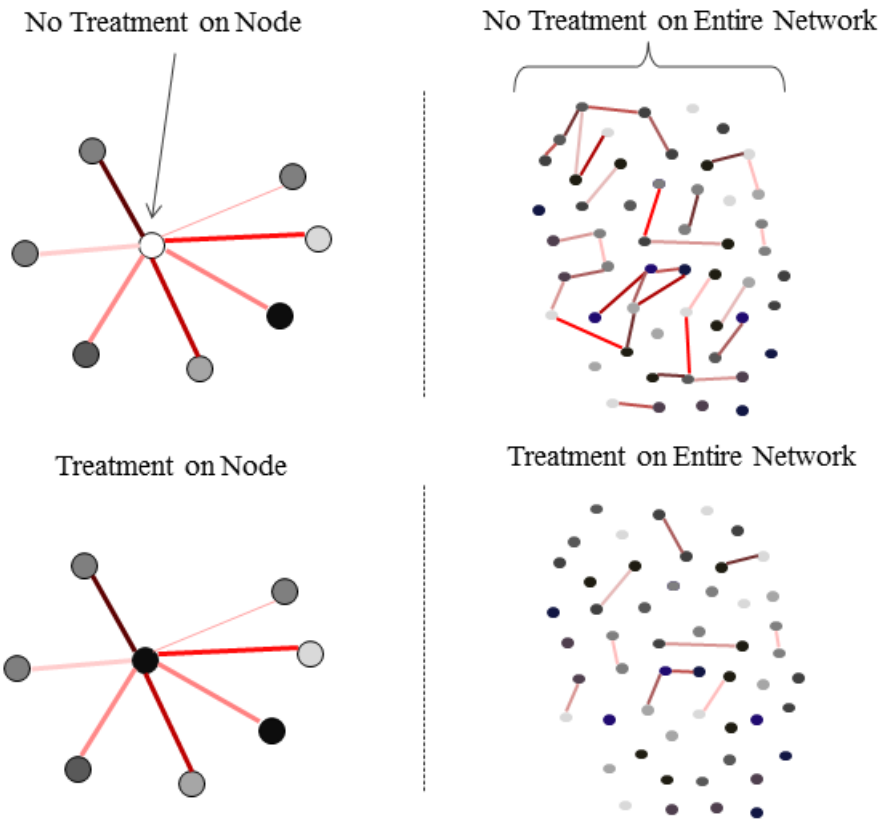


Figure 2.1: Node-level Interference does not Confound Network Structure Effects. Let similarity in node color represent homophily effects and similarity in edge color represent peer effects. (Left) Homophily and peer effects are significant since the treatment impacts an individual node (Right) Homophily and peer effects are still present when considering treatments applied to the entire network, but they are not confounding since analysis focuses on estimating a property of the entire network.

When performing causal inference it is difficult to know whether SUTVA truly holds–for instance, *we make the strong assumption that interference between dyads is not present on network structures*–but it is still useful to develop methodologies for the situation where SUTVA is assumed to hold. We therefore develop a methodology in a simplified situation which can be further developed by investigating and adjusting for dyad-level interference effects[17]. Assuming SUTVA there are still issues which need to be addressed when estimating causal effects on the formation of network structures, notably: i) accounting for network sparsity by dealing with the latent variable of which dyads are 'at-risk' of an interaction ii)

[17] Dyads sharing endpoints represent just one of many types of dyad-level interference.

reducing the bias from comparing dyads across time in dynamic networks iii) controlling for structural features of networks. The estimation and matching aspects of our methodology present strong first steps towards resolving these issues.

3 The MARA Natural Experiment

In this section we lay out our motivating dataset and natural experiment which are well-suited to illustrate our methodology.

Lai et al. (2010) use a Bayesian supervised learning approach to extract a co-patent network from the U.S. utility patent database for years 1975-2010. The data contains over eight million inventor-patents and career histories for 540,780 patenting inventors. In this co-patent network nodes represent inventors and edges represent collaborations between inventors on a patent. The data includes numerous node- and edge-specific covariates. Node covariates include: inventor location (city, state, zipcode, country), name, and company. Edge covariates include: patent owner (the assignee, whether an individual or a company), patent technology class, patent technology subclass, year and date of patent application and year and date of patent grant.

We are interested in utilizing this longitudinal network dataset to estimate the impact of non-compete laws on inventor patent collaboration propensities. To this end we exploit the MARA natural experiment which was first uncovered by Marx et al. (2009) and has been used often in studies of industrial organization and innovation. In 1905 the Michigan legislature passed statute 445.761: "All agreements and contracts by which any person ... agrees not to engage in any avocation or employment are hereby declared to be against public policy and illegal and void" which likely aided the auto industry boom in Michigan from 1905-1915 (Klepper 2002). On March 27, 1985 the Michigan Antitrust Reform Act (MARA) inadvertently repealed statute 445 and thus allowed the enforcement of non-compete laws in Michigan[18]. Marx et al. (2009) show that the reversal of Michigan's non-compete laws was inadvertent[19]. They note that MARA was modeled on the Uniform State Antitrust

[18] Non-compete laws are standard in employment contracts (Stuart and Sorenson 2003) and are used by firms to prevent employees from not joining competitive companies in similar professions. For instance, suppose Google and Apple are both in a state allowing non-compete laws; Google could potentially make employees sign a contract barring them from working with Apple for a few years after leaving Google.

[19] Meyer (1995) notes that strong natural experiments are spontaneous and inadvertent.

act of the early 1970s and argue that the non-compete law reversal was an unintended consequence of anti-trust reform. They also note that law firms and scholars quickly noted this change in non-compete laws and disseminated the information to corporate clients and imply that knowledge of the non-compete law reversal would have been widespread by 1986.

Marx et al. (2009) use a matched differences-in-differences (DID) approach[20] to argue that MARA resulted in an exodus of inventors from Michigan. Their study implies that allowing the enforcement of non-competes has negative effects on innovation.

The DID approach requires the use of a control group similar to the treatment group across the time period of analysis. Marx et al. note that strong control states for Michigan include medium-sized midwestern states with declining economies that do not enforce non-compete laws such as Ohio, Illinois, and Pennsylvania[21]. Marx et al.'s approach did not require network analysis since they were not concerned with the impact of MARA on collaboration propensity between inventors but instead on a non-interaction related outcome–inventor location. Motta and Roende (2002) suggest that non-compete laws reduce collaborative productivity among inventors since non-competes prevent inventors from seeking external opportunities and therefore from exploring their true value on the inventor market. We therefore expect that the enforcement of non-compete laws generally worsens the pace of innovative activity in a region and hypothesize that MARA will lower interaction propensities among inventors in Michigan.

Before rigorously introducing our methodology, we now specifically outline the steps, laid out generally in Section 1.1, for estimating the causal effects of MARA on patent collaboration propensities between inventors in Michigan.

1. *Selecting the set of nodes of the network object* (Section 3.1). We restrict our node set to only those inventors who remain in the group (state) for the entire time period of analysis and who have patented before the beginning of the time period of analysis.
2. *Defining the estimand* (Section 5.1). We are interested in measuring the extent to which MARA impacts collaboration propensities between inventor-dyads in Michigan

[20] Differences-in-differences is a common approach for estimating impacts of programs when using quasi-experimental designs with multiple groups and time periods.

[21] *Appendix J* presents graphics summarizing the patenting landscape (variables such as patent technology class, patent company, # patents per inventor) for Michigan and our focus control group state, Ohio. These graphics summarize the co-patent dataset and also show that Ohio is a strong control state since distributions of pre-treatment (pre-MARA) variables are similar.

at risk of interaction.

3. *The sparsity problem for finite-population networks* (Section 4.2) Estimates will be biased if inventor pairs not at risk of interacting, such as those patenting in distant firms or technology classes, are included in analysis.

4. *Defining the estimator and inference procedure* (Section 5.3). The true set of inventors-dyads at risk of interaction is unknown for Ohio and Michigan inventor-dyads. We propose a DID estimator for the estimand, and introduce the use of conditionally independent relationship models where inference adjusts for network sparsity and is independent of true risk set. This DID estimator adjusts for network sparsity and also for treatment effect bias caused by state-level differences.

5. *Selecting dyad sets to perform matching and estimation over* (Section 6.2). Select only inventor-pair dyads for which we observe an interaction (a patent).

6. *Selecting covariates to perform matching over* (Section 6.3). Select edge covariates (e.g., patent date, company owning the patent), node covariates (e.g., inventor zip-code, inventor firm), and structural covariates (e.g., how important, or central, each dyad is to the network) to remove confounding covariates which may impact inventor interaction propensity.

7. *Matching on pre-treatment covariates between experimental and control groups* (Section 6.3). Select similar matched inventor-dyad sets in pre-treatment Michigan and Ohio by performing propensity score matching.

8. *Estimating the causal treatment effect* (Section 5.3). Use the sparsity-adjusted inference procedure to calculate the differences-in-differences estimator.

3.1 Selection of Node Sets

Before we can begin the exposition of our methodology, we must first identify the type of network our method is suited for. Our analysis requires a network consisting of a fixed number of nodes (inventors) whose node-properties and connections can change over time. In the MARA scenario we select the fixed ego set by considering only inventors with at least one patent (can be a non-collaborative patent) before the beginning of the analysis time interval and who remain in a state for the duration of the analysis time interval. For a three-year time frame around MARA (from 1982-1988), the initial sample of inventors

includes inventors with at least one patent before 1982 (since MARA occurred in 1985). We call the resulting dynamic network an ego network[22].

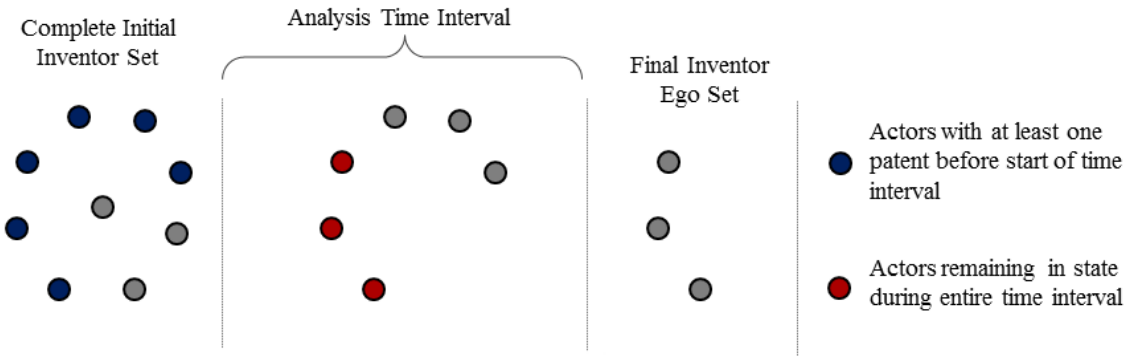


Figure 3.1: Selection Process for Inventor Ego Sets. (Left) The initial inventor set consisting of all inventors in the entire dataset. Inventors who have not had at least one patent before the start of the time interval considered (e.g., in the three-year case this is before 1982) are not considered in the final ego set. (Center) Inventors not remaining in the state during the analysis time interval are removed. (Left) The final node set consists of inventors with at least one patent before the start of the analysis time interval who also remain in their respective state (i.e., their respective experimental group) during the analysis time interval.

This method of node sampling does not include inventors who began their careers after 1982. We therefore potentially ignore information about all possible collaborations by inventors in the ego set . We restrict our network sample to a fixed number of inventors since all current models providing interaction-propensity-related estimators require fixed vertex sets[23]. Having identified the type of network object our analysis is performed on (an ego network) we can now introduce our methodology. Before we introduce our estimand and matched DID estimator, which are central to our methodology, we must review the theoretical foundations behind the potential outcomes approach to causal inference and must define network sparsity for finite populations.

[22] *Appendix K* presents graphics summarizing properties of the patents in these ego networks for Michigan and Ohio. Again we see that for these ego sets Ohio is a strong control state since distributions of pre-treatment patent-related variables are similar between states.

[23] Modeling interaction propensities with time-varying vertex sets can reduce bias of interaction propensity related estimators but has not yet been developed. It is unclear whether estimates from the ego network can reasonably be considered estimates for the entire network and not just the ego network.

4 The Rubin Causal Model and Sparsity

In this section we go over key aspects of modern causal inference for observational studies, namely the ideas of potential outcomes and ignorable assignment mechanisms, based on the work by Donald Rubin and collaborators. This discussion, draws from Rubin (1974, 1978), Imbens & Woolridge (2009), and Imbens & Rubin (2013). We also formally introduce a definition of sparsity in the finite-population context to justify our choice of estimand and estimator.

4.1 The Rubin Causal Model

The unit of analysis is the dyad. A dyad is a pair of nodes and thus represented by ij where i and j are nodes. Given a fixed set of N nodes units are indexed by $ij = 1, ..., \binom{N}{2}$. Some dyad-units are assigned to treatment and others to control. Let W_{ij} indicate whether dyad ij is given the treatment: $W_{ij} = \begin{cases} 1 & \text{if assigned to treatment group} \\ 0 & \text{if assigned to control group} \end{cases}$. $\mathbf{W}$ represents the $\binom{N}{2}$-vector with ij-th element being W_{ij}. For each unit we observe a K-dimensional row vector of covariates, X_{ij}, with $\mathbf{X}$ denoting the $\binom{N}{2} \times K$ matrix with ij-th row equal to X_{ij}.

4.1.1 Potential Outcomes

Formalized in Rubin (1974), estimating causal effects, for both randomized experiments and observational studies, is inherently a comparison between a unit's potential outcomes. Dyad ij's causal effect is the comparison of dyad ij's potential outcome if given the treatment, $Y_{ij}(1)$, and dyad ij's potential outcome if in the control group and thus not given the treatment, $Y_{ij}(0)$. The causal estimate summarizes a comparison of the potential outcomes. For each unit only one of the potential outcomes is observed since each unit can receive either the treatment or control and not both.

A main assumption of the Rubin Causal Model is the Stable Unit Treatment Value Assumption (SUTVA), which we must make for the rest of this thesis.

Assumption 4.1. (SUTVA) (Imbens & Rubin, 2013)
i) The potential outcomes for a dyad are not impacted by treatment assignment status of other dyads. ii) Treatment assignment is homogeneous and has identical impact on dyad potential outcomes[24].

SUTVA results in well-defined potential outcomes. If we accept SUTVA then potential outcomes relate only to each dyad's actions and therefore causal inference is reduced to imputing the missing potential outcomes. Only one potential outcome is observed for each unit. For unit $ij \in \left\{1, ..., \binom{N}{2}\right\}$, let Y_{ij}^{obs} represent this observed outcome. Let Y_{ij}^{mis} denote the missing potential outcome. We have that $Y_{ij}^{obs} = \begin{cases} Y_{ij}(0) & if\ W_{ij} = 0 \\ Y_{ij}(1) & if\ W_{ij} = 1 \end{cases}$ and $Y_{i}^{mis} = \begin{cases} Y_{ij}(1) & if\ W_{ij} = 0 \\ Y_{ij}(0) & if\ W_{ij} = 1 \end{cases}$. Even having accepted SUTVA it is still not possible to infer the causal impact of the treatment. The key missing piece of information is the assignment mechanism, or how each unit received its corresponding treatment level (treatment or control).

4.1.2 Assignment Mechanism

The assignment mechanism is defined as the conditional probability of receiving the treatment as a function of potential outcomes and observed covariates. In randomized experiments the assignment mechanism is random and known; this ensures covariate balance between treated and control groups. In observational studies the assignment mechanism determining which individuals are assigned to treatment is often unknown. With observational studies there is one key assumption which is required for unbiased estimates:

Assumption 4.2. (Strongly ignorable treatment assignment) (Rosenbaum & Rubin, 1983)
i) given the covariates the treatment assignment is independent of the potential outcomes ii) there is a positive probability of being assigned to treatment for all covariate values.

[24] If we make the strong assumption that there is no dyad-level interference then SUTVA intuitively holds in the MARA scenario because treatment is dependent on location e.g., our treatment group consists of inventor-dyads in Michigan, while our control group consists of inventor-dyads in Ohio. Therefore, the interaction propensities between inventor-dyads in Michigan are not impacted by the Ohio dyads. To satisfy part ii of SUTVA we make the assumption that the treatment effect (impact of MARA) is homogenous on all inventors in the treatment group. Further work is required to investigate interference effects between dyads to discuss the validity of SUTVA when dyads are the unit of analysis.

Assumption 4.2. asserts that all variables which need to be adjusted for have been controlled for by the researcher and therefore that the assignment mechanism, which is usually unknown for complicated nonexperimental studies, does not need to be known for unbiased causal estimates. Formally, $W_{ij} \parallel (Y_{ij}(0), Y_{ij}(1)) \mid X$, where $A \parallel B|C$ denotes conditional independence of A and B given C. This assumption, also called unconfoundedness, states that adjusting for differences in observed covariates can remove bias in comparisons between treatment and control units [25]. Assumption 4.2. crucially allows us to reduce estimate bias by adjusting for differences in observed covariates.

Assuming we have adjusted for confounding time trends and removed all biases between treated and control units the ultimate object of interest is the causal estimand. Before we can define our causal estimand more precisely we first discuss the issue of sparsity which influences our choice of estimand and estimator.

4.2 Network Sparsity

Like most social network data the co-patent dataset we use is sparse. Sparsity has been formalized in the superpopulation context (D'Amour & Airoldi, 2014), but not yet in the finite-population context. The finite-population context is relevant to our method since we focus on observational networks with fixed node sets. This section has three objectives: i) reviewing a definition of sparsity in the superpopulation context ii) reviewing a class of network models to provide a framework to think about sparsity in the finite-population context iii) proposing a definition of sparsity for the finite-population context to justify our choice of estimand and estimator which appear in Section 5.

4.2.1 Sparsity in the Superpopulation Context

The following definitions (Shalizi and Rinaldo, 2013; D'Amour & Airoldi, 2014) define sparsity asymptotically in the superpopulation context. First define the sample space of a generalized random graph. A generalized random graph is similar to a traditional random

[25] Assumption 4.2 holds in the MARA scenario. Since the mechanism by which inventors choose to live in a certain state is for all purposes random, dyads essentially are randomly assigned to treatment (MI) and control (OH) groups, thus satisfying part ii of Assumption 4.2. Since dyads in the treatment group automatically receive the treatment, it follows that the interaction propensity between inventor-pairs does not impact the location (treatment assignment) of the inventor pairs.

graph but allows edge random variables to take values for all possible combinations of pairs of elements in the vertex set.

Definition. 4.1 (Generalized Random Graph). *A* generalized random graph G_V *is a graph-like object consisting of an ordered pair* (V, Y) where V *is a finite set of vertices, and* Y *is a* $\binom{|V|}{2}$*-dimensional multivariate random object whose components correspond to the unique pairs of elements in* V*.*

Sample spaces are drawn from a population. A population graph is defined as an infinite stochastic process of generalized random graph.

Definition. 4.2 (Generalized Random Graph Process). *A* generalized random graph process G_V *is a stochastic process indexed by a countably infinite vertex set* $\mathbb{V}$ *whose finite-dimensional distribution for any finite subset* $V \subset \mathbb{V}$ *defines a generalized random graph* G_V *with vertex set* V*.*

A generalized random graph process allows for a generalized random graph sample to be defined in terms of a superpopulation. Sparsity can therefore be asymptotically defined for the infinite vertex superpopulation.

Definition. 4.3 (Density Operator). *Let* G_V *be a generalized random graph with vertex set* V *and let* Y_V *be its generalized adjacency matrix. Fix an element in the support of the components of* Y_V *to be zero, denoted by* $\mathbf{0}$*, and define the indicator random variables* $A_{ij} = \mathbb{I}_{\{Y_{ij} \neq 0\}}$ *for each* $i < j \in V$*.*

The density operator D *with respect to the element* $\mathbf{0}$ *maps a generalized random graph to the interval* [0,1] *by* $D(G_V) = \sum_{ij} A_{ij} / \binom{|V|}{2}$*, giving the proportion of components of* Y_V *that are non-zero.*

Intuitively, a population process is sparse if the proportion of dyads with edges, $D(G_V)$, converges to 0 as the number of vertices V increases. Formally,

Definition. 4.4 (Sparse Generalized Random Graph Process). *Let* G_V *be a generalized random graph process on* $\mathbb{V}$*.* G_V *is* sparse *if and only if for any* $\epsilon > 0$ *and any* $\delta > 0$*, there exists an* N *such that for any subset of vertices* $V \in \mathbb{V}$ *with* $|V| > N$ *the corresponding finite dimensional generalized random graph* G_V *has the property* $Pr(D(G_V) > \epsilon) < \delta$*.*

We can now intuitively define sparsity in the superpopulation context: as the number of vertices in network samples drawn from the superpopulation increases the proportion of dyads observed to interact compared to the potential number of interactions approaches 0. Before we define sparsity in the finite-population context we review a novel class of network models, developed by D'Amour & Airoldi (2014) which provides an appealing network generative process to think about the 'risk set' and 'observed interaction set' which guide our finite-population definition and our choice of estimand and estimator.

4.2.2 Conditionally Independent Relationship Models

The Conditionally Independent Relationship (CIR) model family $\mathcal{P}$ is parameterized by Q and β, where Q is a sparse network process which accounts for network sparsity by representing how unobservable relationships form between actors. β is a parameter summarizing the population process where a network sample is drawn from. Define the risk set as a relationship graph R_n where n is the number of nodes. The number of observed interactions of each dyad at time t is represented by the counting process $Y_{ij}(t)$. These observed interactions are drawn independently from R_n according to a distribution parameterized by β and dyad covariates X_{ij} [26]. R_n is generated from its marginal distribution Q_n which characterizes the social process defining which dyads are at risk of interaction [27]. Intuitively $R_{ij} = 0$ implies $Y_{ij} = 0$ since two actors cannot interact if they are not at risk of interaction. Similarly $Y_{ij} > 0$ at some time implies $R_{ij} = 1$. Define $A_{ij} = 1$ if $Y_{ij} > 0$ at some time, i.e., $A_{ij} = \mathbb{I}_{\{Y_{ij} \neq 0\}}$. Intuitively A_n (the adjacency matrix) summarizes the dyads in which we observe an interaction and therefore know must be at risk of interaction.

[26] D'Amour & Airoldi (2014) assume that dyads are *independently* drawn from a zero-inflated process, a process which allows for frequent zero-valued observations. In sparse networks zero-interaction pairs are frequent since as the number of nodes N increases the number of potential interactions $\binom{N}{2}$ increases quicker.

[27] To clarify, imagine students are at risk of interaction (e.g., collaborating on a homework assignment) depending on a variety of factors such as whether they share a course or live nearby. For two students i and j Q_{ij} represents the process generating whether students ij are at risk of interaction. R_{ij} is the indicator representing whether the students actually are at risk of interaction, and Y_{ij} represents the observed outcome of the number of times the students collaborate.

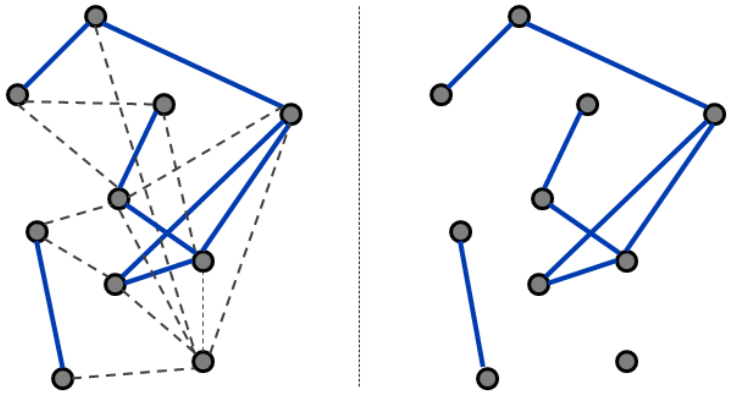


Figure 4.1: The Risk Graph and Observed Interaction Graph. (Left) Dotted gray ties represent dyads which are at risk of interaction. These are generally unobservable. Blue ties represent observable interactions, where pairs of actors have an observed interaction. (Right) The observed network sample has only observed interactions. The goal is to condition on the risk set in our estimates of interaction propensity between actors.

4.2.3 Sparsity in the Finite-population Context

By using the appealing generation process described above, where Y_V is dependent on R_V and R_V is in turn drawn from a marginal distribution Q_V, we are not required to define sparsity in terms of outcome graph density, as was required in the superpopulation context, and can now define sparsity in terms of the *process* generating the outcome graph.

In the finite-population context we cannot use an asymptotic definition of sparsity[28]. It is more pragmatic to define sparsity in terms of the number dyads actually at risk of interacting compared to the number of potential dyads [29]. *We therefore define a finite-population network as sparse if not all of the dyads in the network are at risk of interacting.* Intuitively, larger networks are more likely to be sparse. This definition implies that some dyads are too dissimilar or too distant to ever interact.

[28] We could simply but naively define a finite-population network as sparse if the proportion of observed edges to potential edges would decrease if a new node were added. This definition, however, would imply that a social network with 1000 people and only 2 friendships is not sparse if the addition of a gregarious new individual could increase the number of friendships to 5 (since $\frac{1001}{5} > \frac{1000}{2}$).

[29] Consider again the social network with 1000 people and only 2 observed friendships. With this view there are likely so few friendships formed because the number of potential dyads at risk of interactions is extremely low, e.g., only 5 of the 1000 people in the network attend the same school (and out of these 5, 2 are actually friends), while the other 998 attend 998 different schools.

Definition. 4.5 (Finite-Population Potential Outcomes Graph). *A* finite-population potential outcomes graph G_V *consists of a finite set of* $|V|$ *vertices and is represented by* $(Y_V(0), Y_V(1), R_V, X_V)$ *where* $Y_V(1)$ *and* $Y_V(0)$ *represent the potential outcome graphs if assigned to treatment or control respectively.* R_V *represents the risk graph and* X_V *represents dyad covariates.*

We make the strong assumption that R_V does not change over time. We also assume that R_{ij} is independent of treatment assignment and therefore that $R_{ij}(0) = R_{ij}(1)$. Similarly, $R_{ij} = 0$ if and only if $Y_{ij}(0) = Y_{ij}(1) = 0$.

Definition. 4.6 (Sparse Finite-Population Potential Outcomes Graph). *Let* G_V *be a finite-population potential outcomes graph with* $|V|$ *vertices.*G_V *is sparse if and only if for some dyad* ij, $R_{ij} = 0$.

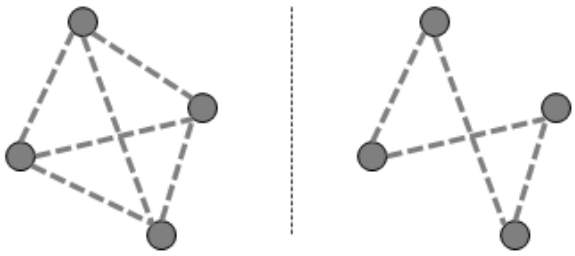


Figure 4.2: Illustration of Sparse Finite-population Network. Dashed gray lines represent dyads at risk of interaction. (Right) Non-sparse network. All nodes are at risk of interaction. (Left) Sparse network. Not all nodes are at risk of interaction, and therefore the social scenario is unknown.

Definition 4.6 states that a network is sparse if there exists at least one dyad in the network not at risk of interaction. The definition may appear trivial, but it is more useful when stated alternatively: *the social process of interest is unobservable.* Definition 4.6 provides intuition for how to define an estimator for interaction propensities. For finite-population sparse networks, with or without large numbers of nodes, we cannot assume that all dyads are at risk of interaction, e.g., not all inventors can potentially collaborate with every other inventor. The social process of interest becomes more difficult to identify

as the network becomes more sparse[30]. *Analysis is performed using only those dyads at risk of interaction, but since the social process generating the risk set is unknown for sparse networks we are required to use an inference method which is independent of the social process and the true risk set.* In the following section we use the ideas of potential outcomes and finite-population network sparsity to introduce our estimand, estimator, and inference approach.

5 Estimand, Estimator, and Inference

This section introduces our estimand which summarizes interaction propensity for dyads at risk of interaction. Dyads not at risk of interaction have zero interaction propensity and therefore will bias estimates downward if included in analysis. A limitation of this estimand is that it conditions on the true risk set, which by Definition 4.6 is unobservable for sparse networks. Our two-group design also leads to group-level confounding differences. This section proceeds to develop a differences-in-differences (DID) estimator for this estimand which does not require knowledge of the true risk set. This DID estimator adjusts for finite-population network sparsity and group-level confounding of the treatment effect. Matching is also introduced to further lessen estimate bias.

5.1 Estimand

Let there be a population of dyad-units, indexed by $ij = 1, ..., \binom{N}{2}$ where there are N nodes and therefore $\binom{N}{2}$ dyads. Each dyad is part of some experimental group$W_{ij} = \begin{cases} 1 & \text{if in treatment group} \\ 0 & \text{if in control group} \end{cases}$. There are also two time periods $t_{ij} = \begin{cases} 1 & \text{if post-intervention} \\ 0 & \text{if pre-intervention} \end{cases}$.

Let assignment to the treatment be given by $M_{ij} = W_{ij}t_{ij} = \begin{cases} 1 & \text{if given treatment} \\ 0 & \text{if not given treatment} \end{cases}$.

Let $\mathbf{W}$, $\mathbf{t}$, and $\mathbf{M}$ represent the vectors containing group, time period, and assignment-to-treatment indicators for all dyads. For each dyad-unit we observe a K-dimensional row vector of covariates, X_{ij}, with $\mathbf{X}$ denoting the $\binom{N}{2} \times K$ matrix with i-th row equal to X_{ij}.

[30] By our definition 'more sparse' now means that the number of dyads at risk of interaction becomes a smaller proportion of the total number of potential interactions.

Let $Y_{W_{ij},t_{ij},M_{ij}}$ represent the number of interactions between actors i and j in time period t_{ij}, in group W_{ij}, with treatment assignment M_{ij}[31]. *We are interested in the effect of the treatment on the treated at risk of interaction* (ETTR), e.g., how MARA impacts interaction propensities among inventors in Michigan compared to interaction propensities if the same inventors were not impacted by MARA. Our general ETTR causal estimand is given by:

$$\tau_{treated,risk} = f\left(\mathbf{Y^R_{1,1,1}}\right) - f\left(\mathbf{Y^R_{1,1,0}}\right) \tag{5.1}$$

where$\mathbf{Y^R_{1,1,1}}$ and $\mathbf{Y^R_{1,1,0}}$ are sets containing the potential outcomes for post-intervention treatment group dyads at risk of interaction if assigned to treatment and control respectively. We also define $\mathbf{Y^R_{1,0,0}}$ and $\mathbf{Y^R_{0,0,0}}$ similarly for pre-intervention treatment and control groups. $f(\cdot)$ represents a general summary of interaction propensity computed on each set of dyad potential outcomes[32]. $f(\cdot)$ is computed on all the dyads in each $\mathbf{Y^R_{W,t,M}}$. Intuitively, the ETTR is the difference between the interaction propensity among dyads at risk of interaction with and without assignment to treatment[33]. We return to CIR models to select the interaction propensity summary of choice.

D'Amour & Airoldi (2014) develop a convenient summary for interaction propensity. Covariate effect estimation problems generally use generalized linear models (GLMs), such as Poisson regressions or Cox proportional hazard regressions, to specify the expected number of interactions as a linear combination of covariates (for example, Perry & Wolfe, 2013). To this end we use a Cox proportional hazards model, which is a natural choice of model for counting processes on networks (Perry & Wolfe, 2013). We use the standard Cox proportional hazard regression where $Y_{ij}(t)$ is a counting process of interactions between actors i and j representing how many total times i and j have interacted at time t, and $X_{ij}(t)$ are

[31] In the case of the MARA natural experiment, we have N inventors and $\binom{N}{2}$ inventor-pairs. The treatment group is Michigan, the control group is some state similar to Michigan in economy, size, and region. t represents either the period post-MARA or pre-MARA, and only those dyads in Michigan post-MARA receive the treatment. Thus $Y_{1,1,1}$ would represent the number of interactions between two actors in post-intervention Michigan who experience MARA.

[32] This summary could include the instantaneous probability, average number of interactions, or impact of covariates on interaction rates among other summaries.

[33] If our summary of interest is the average number of interactions among dyads, then the average effect of the treatment on the treated at risk of interaction (ATTR) is given by $\tau_{treated,risk} = \frac{\sum_{ij:R_{ij}=1} Y^{(1)}(1)|W_{ij}=1 - Y^{(1)}_{ij}(0)|W_{ij}=1}{\sum R_{ij}}$. *Appendix A* details why we do not consider the ATTR.

covariates changing over time. The CIR model is given by:

$$Y_{ij}(t)\,|X_{ij}(t),\mathcal{F}(t) \quad \sim \quad \begin{cases} CP\left(\lambda_{ij}(t)\right) & \text{if } R_{ij}=1 \\ log\lambda_{ij}(t) \equiv log\lambda_0(t) + X_{ij}(t)\,\beta & \\ 0 & \text{if } R_{ij}=0 \end{cases}, \tag{5.2}$$

where $\mathcal{F}(t)$ represents the filtration up to time t which essentially is all the information up to time t, λ_0 represents the baseline instantaneous risk of an interaction occurring at a certain time, and $\lambda_{ij}(t)$ represents the instantaneous hazard of interaction between actors i and j at time t. β summarizes how actors interact with other actors they already know based on their covariates and is therefore a convenient summary of interaction propensity. To use this model we must assume that each unit increase of a covariate proportionately scales the hazard.

Inference on these models estimates $\hat{\beta}$ which is a parameter vector of Cox proportional hazard regresion coefficients governing the propensity for the observed interactions and thus governing the network structure. $\hat{\beta}$ is a vector which represents the instantaneous hazard rate at which regression variables impact dyad interaction propensity. Cochran (1963) notes that linear regression coefficients such as $\hat{\beta}$ are reasonable as population estimands in the finite-population context. If there are M variables used in the Cox proportional hazard regression, then $\hat{\beta} \equiv \left(\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_M\right)$where $\hat{\beta}_0$ is a baseline interaction rate while $\hat{\beta}_1, ..., \hat{\beta}_M$ represents the instantaneous log hazard rate change in interaction propensity for a unit change in the variable [34]. *$\hat{\beta}$ summarizes how dyad covariates impact interaction propensities* and can therefore be used in place of the causal estimand previously written. We write our new ETTR causal estimand as:

$$\tau_{treated,risk} = \hat{\beta}\left(\mathbf{Y^R_{1,1,1}}\right) - \hat{\beta}\left(\mathbf{Y^R_{1,1,0}}\right), \tag{5.3}$$

The notation $\hat{\beta}(\cdot)$ represents that $\hat{\beta}$ is computed on all dyads in each $\mathbf{Y^R_{W,t,M}}$. We therefore use $\hat{\beta}$ as an estimand for summarizing interaction propensities in sparse networks.

[34] The 'hazard' we are concerned with is an interaction between a dyad. For example, if binary covariate M represents whether two inventors work for the same firm, then the coefficient $\hat{\beta}_M$ represents the instantaneous change in log interaction rate between the inventors if they were to begin working for the same firm.

5.2 Sparsity-Adjusted Estimator

The proposed estimand, Equation 5.3, requires full knowledge of the risk set. However, Definition 4.6 implies that we cannot observe the true risk set, and therefore must use an inference method which estimates $\hat{\beta}$ without knowledge of the true risk set. $\hat{\beta}$ is found by maximizing the following full likelihood when applied to the finite-population of interest (D'Amour & Airoldi, 2014):

$$\begin{aligned} P_\beta\left(Y_n|X_n\right) &= \sum_{\mathbf{R}} P_Q\left(R_n|X_n\right) P_\beta\left(Y_n|X_n, R_n\right) \\ &= \sum_{\mathbf{R}} P_Q\left(R_n|X_n\right) \prod P\left(Y_{ij}|R_{ij}=1, X_{ij}\right)^{R_{ij}}. \end{aligned}$$

In the CIR model framework, the social framework is represented by the marginal distribution, Q_n, of the risk graph R_n. The likelihood for Y_n, the observed outcome graph, is given by the joint likelihood of R_n and Y_n with unobserved elements–those dyads in the risk set but not observed to have interacted–removed. This likelihood is called the full likelihood, since it includes all the information about the marginal distribution, risk set, and observed interactions.

For complicated networks researchers are often oblivious to the true functional form that Q should take. D'Amour & Airoldi (2014) note that the interpretation and inference for $\hat{\beta}$ is sensitive to the specification of Q and therefore use a partial likelihood estimation approach to obtain an estimator for $\hat{\beta}$ which is independent of the functional form of Q[35]. This partial likelihood estimation method is termed 'truncated inference' by D'Amour & Airoldi (2014). The maximum partial likelihood estimator of $\hat{\beta}$ is notated $\hat{\beta}_{tr}$ and is given

[35] In the general case letting the full parameter vector ϕ to be $\left(\hat{\beta}, Q\right)$, where $\hat{\beta}$ is the parameter vector we wish to estimate and Q is the marginal distribution of the risk graph representing the social process, the full likelihood (D'Amour & Airoldi, 2014) is

$$P_\phi\left(Y|X\right) = \left[\prod_{ij}\sum_{R_{ij}} P_Q\left(R_{ij}|A_{kl<ij}, X\right) P_\beta\left(A_{ij}|R_{ij}, X_{ij}\right)^{R_{ij}}\right]\left[\prod_{ij} P_\beta\left(Y_{ij}|A_{ij}, X_{ij}\right)^{A_{ij}}\right].$$

The second factor, $\left[\prod_{ij} P_\beta\left(Y_{ij}|A_{ij}, X_{ij}\right)^{A_{ij}}\right]$, is independent of the social process Q since the risk graph only impacts the number of interactions,Y_{ij}, through the observed interactions A_{ij}. Intuitively, this represents that an observed interaction, A_{ij}, is independent of the entire interaction distribution, Q. Conditional on the observed interaction graph A_n the outcome graph Y_n is independent of Q. The partial likelihood thus adjusts for sparsity by using only those dyads which are known to have an interaction.

by

$$\hat{\beta}_{tr} = \text{argmax}_{\beta} \prod_{ij} P_{\beta} \left(Y_{ij} | A_{ij}, X_{ij} \right)^{A_{ij}}. \tag{5.4}$$

We see that the partial likelihood is independent of any modeling of the risk set or the social scenario, and therefore is independent of network sparsity. Solely by using the observed interactions $ij \; s.t \; A_{ij} = 1$, we specify which dyads we know are at risk of interaction and can estimate $\hat{\beta}_{tr}$. Note also that the regression estimator $\hat{\beta}_{tr}$ adjusts for dyad covariates. D'Amour & Airoldi (2014) show that this estimator is not sensitive to the sparseness or size of the network, and is computationally efficient[36]. We can now rewrite our estimand, Equation 5.3, as a sparsity-adjusted estimator:

$$\tau_{treated,risk} = \hat{\beta}_{tr} \left(\mathbf{Y}^{\mathbf{A}}_{\mathbf{1,1,1}} \right) - \hat{\beta}_{tr} \left(\mathbf{Y}^{\mathbf{A}}_{\mathbf{1,1,0}} \right). \tag{5.5}$$

where **A** represents that *analysis is now only performed on dyads observed to have an interaction.* We have crucially shown that inference of the estimand adjusts for network sparsity in the finite-population context by using only dyads observed to interact. This estimator therefore does not require any modeling of the risk set.

5.3 Matched Differences-in-Differences Estimator

Differences-in-Differences (DID) is a common method to handle quasi-experimental designs involving multiple time periods and groups and has been increasingly used since Ashenfelter and Card (1985). In this section we construct a DID estimator to remove group-level confounding in the sparsity-adjusted estimator (Equation 5.5). Since we use a two-group quasi-experimental design to evaluate the impact of a treatment, the treatment effect estimate will be biased by group-level differences. We can more rewrite the treatment effect estimator more generally by including dyad-covariate-induced bias as

$$\tau = \hat{\beta}_{tr} \left(\mathbf{Y}^{\mathbf{A}}_{\mathbf{1,1,1}} \right) - \hat{\beta}_{tr} \left(\mathbf{Y}^{\mathbf{A}}_{\mathbf{1,1,0}} \right) + b \left(\mathbf{X}^{\mathbf{A}}_{\mathbf{1,1}}, \, \mathbf{X}^{\mathbf{A}}_{\mathbf{1,1}} \right) \tag{5.6}$$

[36] Full likelihood inference requires marginal maximization by the EM algorithm or Monte Carlo integration (D'Amour & Airoldi, 2014). Inference for the partial likelihood is far easier and can be found using standard MLE methods.

where $\mathbf{X^A_{W,t}}$ contains the covariate vectors for post-intervention dyads observed to interact that are in group W and time period t, and $b(\cdot,\cdot)$ is a general bias term accounting for dyad covariate differences between the groups. $b\left(\mathbf{X^A_{1,1}},\, \mathbf{X^A_{1,1}}\right) = 0$ so Equation 5.6 is identical to Equation 5.5. Note that under Assumption 4.2 covariates are independent of treatment assignment and we can therefore leave out the $\mathbf{M}$ indicator, assignment to treatment or control, for each dyad. Let the group-level effect be

$$\delta = \hat{\beta}_{tr}\left(\mathbf{Y^A_{1,1,0}}\right) - \hat{\beta}_{tr}\left(\mathbf{Y^A_{0,1,0}}\right) + b\left(\mathbf{X^A_{1,1}},\, \mathbf{X^A_{0,1}}\right). \tag{5.7}$$

δ represents how the interaction propensity summary would differ between the groups if no treatment occurred[37]. Out of the sets of outcomes, only two are observed post-treatment, $\mathbf{Y^A_{1,1,1}}$ and $\mathbf{Y^A_{0,1,0}}$. With the observed data it is only possible to calculate the effect of the treatment by comparing the outcomes of the treatment and control groups with and without treatment[38]. We see that

$$\hat{\beta}_{tr}\left(\mathbf{Y^A_{1,1,1}}\right) - \hat{\beta}_{tr}\left(\mathbf{Y^A_{0,1,0}}\right) + b\left(\mathbf{X^A_{1,1}},\, \mathbf{X^A_{0,1}}\right) = \tau + \delta \tag{5.8}$$

by adding Equations 5.6 and 5.7. Equation 5.8 shows that the effect we can calculate using the data is biased by the group-level difference and covariate differences between post-intervention dyads in treatment and control groups. We can create a minimally biased estimator of the treatment effect τ in two steps. First we make a strong assumption:

Assumption 5.1. (Parallel Trend)

$$\begin{aligned}&\hat{\beta}_{tr}\left(\mathbf{Y^A_{1,1,0}}\right) - \hat{\beta}_{tr}\left(\mathbf{Y^A_{0,1,0}}\right) + b\left(\mathbf{X^A_{1,1}},\, \mathbf{X^A_{0,1}}\right) = \\ &\hat{\beta}_{tr}\left(\mathbf{Y^A_{1,0,0}}\right) - \hat{\beta}_{tr}\left(\mathbf{Y^A_{0,0,0}}\right) + b\left(\mathbf{X^A_{1,0}},\, \mathbf{X^A_{0,0}}\right) = \delta.\end{aligned}$$

Assumption 5.1 states that the potential outcomes for the treated and control groups would follow parallel paths in the absence of treatment. In the MARA case, Assumption 5.1. states that had MARA not occurred, the interaction propensities of inventors pre- and

[37] In the MARA scenario δ represents how the log interaction rate for a unit change in a variable of interest differs if the same set of inventors not experiencing MARA are in different states.

[38] In the MARA scenario, with the observed data we can only calculate the impact of MARA by comparing the potential outcomes of Ohio inventors pre-MARA with the outcomes of Michigan inventors post-MARA. This causes state-level confounding of the treatment effect.

post-MARA would change at identical rates for Michigan and Ohio. Marx et al. (2009) note that the assumption is plausible in the MARA context.

Under Assumption 5.1 the ETTR, $\hat{\beta}\left(\mathbf{Y}^{\mathbf{A}}_{\mathbf{1},\mathbf{1},\mathbf{1}}\right) - \hat{\beta}\left(\mathbf{Y}^{\mathbf{A}}_{\mathbf{1},\mathbf{1},\mathbf{0}}\right)$, can be expressed as a DID estimator:

$$\left\{\hat{\beta}\left(\mathbf{Y}^{\mathbf{A}}_{\mathbf{1},\mathbf{1},\mathbf{1}}\right) - \hat{\beta}\left(\mathbf{Y}^{\mathbf{A}}_{\mathbf{0},\mathbf{1},\mathbf{0}}\right)\right\} - \left\{\hat{\beta}\left(\mathbf{Y}^{\mathbf{A}}_{\mathbf{1},\mathbf{0},\mathbf{0}}\right) - \hat{\beta}\left(\mathbf{Y}^{\mathbf{A}}_{\mathbf{0},\mathbf{0},\mathbf{0}}\right)\right\} = \tau - b\left(\mathbf{X}^{\mathbf{A}}_{\mathbf{1},\mathbf{1}},\, \mathbf{X}^{\mathbf{A}}_{\mathbf{0},\mathbf{1}}\right) + b\left(\mathbf{X}^{\mathbf{A}}_{\mathbf{1},\mathbf{0}},\, \mathbf{X}^{\mathbf{A}}_{\mathbf{0},\mathbf{0}}\right) \quad . \tag{5.9}$$

The second difference, $\left\{\hat{\beta}\left(\mathbf{Y}^{\mathbf{A}}_{\mathbf{1},\mathbf{0},\mathbf{0}}\right) - \hat{\beta}\left(\mathbf{Y}^{\mathbf{A}}_{\mathbf{0},\mathbf{0},\mathbf{0}}\right)\right\}$, removes the biasing group (state-level) effect from the estimate[39]. If the two groups are identical then we expect this difference to be 0. The DID estimator conveniently provides an estimate of the treatment effect using only the observed potential outcomes. The estimator, however, is still biased due to pre- and post-intervention covariate differences between treatment and control groups. Group-level covariate differences must therefore be adjusted for.

Abadie (2005) notes that pre-treatment covariate differences between treatment and control groups can lead to non-parallel outcome trends between the groups thereby violating Assumption 5.1. Under the unconfoundedness assumption (Assumption 4.2) bias between treatment and control groups can be removed by adjusting for pre-treatment differences in covariates. Removing pre-treatment covariate differences yields $b\left(\mathbf{X}^{\mathbf{A}}_{\mathbf{1},\mathbf{0}},\, \mathbf{X}^{\mathbf{A}}_{\mathbf{0},\mathbf{0}}\right) = 0$; however, the DID estimator, Equation 5.9, is still biased by the post-treatment covariates $b\left(\mathbf{X}^{\mathbf{A}}_{\mathbf{1},\mathbf{1}},\, \mathbf{X}^{\mathbf{A}}_{\mathbf{0},\mathbf{1}}\right)$. We provide two arguments to why matching on pre-treatment covariates is still useful:

1. Reducing distance between pre-treatment covariates also reduces distance between post-treatment covariates. Define a time trend $t(\cdot,\cdot)$ which represents the change in pre-treatment covariates, $t\left(\mathbf{X}^{\mathbf{A}}_{\mathbf{W},\mathbf{0}}\right) = \mathbf{X}^{\mathbf{A}}_{\mathbf{W},\mathbf{1}}$. Intuitively, identical dyads are expected to change similarly over time, and therefore the time trend will impact dyads with similar pre-treatment covariates similarly. Using matching to adjust for pre-treatment covariates leads to post-matching $\mathbf{E}\left[b\left(t\left(\mathbf{X}^{\mathbf{A}}_{\mathbf{MI},\mathbf{0}}\right), t\left(\mathbf{X}^{\mathbf{A}}_{\mathbf{OH},\mathbf{0}}\right)\right)\right] = \mathbf{E}\left[b\left(\mathbf{X}^{\mathbf{A}}_{\mathbf{MI},\mathbf{1}},\, \mathbf{X}^{\mathbf{A}}_{\mathbf{OH},\mathbf{1}}\right)\right]$

[39] In the MARA scenario the second difference represents the difference in interaction propensity between Michigan and the control state if the same inventors were in both states.

being $< b\left(\mathbf{X}^{\mathbf{A}}_{\mathbf{MI},\mathbf{1}},\ \mathbf{X}^{\mathbf{A}}_{\mathbf{OH},\mathbf{1}}\right)$pre-matching.

2. The DID estimator is based on a regression estimator for $\hat{\beta}_{tr}$ which adjusts for differences in post-treatment covariates. We therefore expect the regression estimator to be invariant to small post-treatment covariate differences. Following argument 1, matching results in more similar post-treatment covariate distributions and therefore allows for less reliance on the covariate adjustment process during inference for $\hat{\beta}_{tr}$.

In the MARA scenario our matching goal is therefore: $\forall i, j$ in MI, find a k, l in the control state OH such that $X_{ij} = X_{kl}$. This results in $b\left(\mathbf{X}^{\mathbf{A}}_{\mathbf{MI},\mathbf{0}},\ \mathbf{X}^{\mathbf{A}}_{\mathbf{OH},\mathbf{0}}\right) = 0$ and $\mathbf{E}\left[b\left(\mathbf{X}^{\mathbf{A}}_{\mathbf{MI},\mathbf{1}},\ \mathbf{X}^{\mathbf{A}}_{\mathbf{OH},\mathbf{1}}\right)\right]$ post-matching $< b\left(\mathbf{X}^{\mathbf{A}}_{\mathbf{MI},\mathbf{1}},\ \mathbf{X}^{\mathbf{A}}_{\mathbf{OH},\mathbf{1}}\right)$ pre-matching and defines a best-case DID estimator for the ETTR.

We have constructed a matched DID estimator which adjusts for network sparsity, removes biasing group-level effects, and adjusts for covariate differences. It is inevitable that this estimate will be biased, but we have argued that matching can reduce this bias. For an optimal estimate exact matching on pre-treatment covariates is required, e.g., $b\left(\mathbf{X}^{\mathbf{A}}_{\mathbf{1},\mathbf{0}},\ \mathbf{X}^{\mathbf{A}}_{\mathbf{0},\mathbf{0}}\right) = 0$. Exact matching is not feasible with large numbers of covariates; therefore, matching should find optimal balance between dyads in the treatment and control states to minimize $b\left(\mathbf{X}^{\mathbf{A}}_{\mathbf{1},\mathbf{0}},\ \mathbf{X}^{\mathbf{A}}_{\mathbf{0},\mathbf{0}}\right)$ and $b\left(\mathbf{X}^{\mathbf{A}}_{\mathbf{1},\mathbf{1}},\ \mathbf{X}^{\mathbf{A}}_{\mathbf{0},\mathbf{1}}\right)$. To this end, Section 6 provides background on propensity score matching and presents matching on dyads.

6 Dyad Selection and Matching on Dyads

We match treatment and control group dyads to remove pre-treatment covariate differences. There exist alternative methods to balance covariates which include regression models, structural equation models or selection models. Matching methods have key advantages over these other approaches (Imbens & Rubin, 2013). Matching methods, unlike structural equation or selection models, do no conflict with regression estimators, such as the Cox regression estimators we use, and are used optimally in combination. Selection and regression models do not perform well when there is minimal overlap between covariate distributions in treatment and control groups while matching methods perform well without significant overlap (Deheija and Wahba, 1999). Imbens & Rubin (2013) advocate the use of matching

since it does not rely on modeling unknown functions and instead, more pragmatically, finds direct comparisons for each unit. Ho et al., (2007) also advocate matching stating that matching methods are less sensitive to outcome model specifications and therefore are more flexible.

Matching methods have been used since sociological work in the 1940s (Greenwood, 1945). The theoretical basis for matching, which focused on matching units with one covariate, was developed by Cochran & Rubin (1973) and Rubin (1973). Since it is difficult to find exact matches between units with many covariates, matching on multiple covariates is more problematic. A landmark paper by Rosenbaum & Rubin (1983) introduced the propensity score and showed that the propensity score efficiently reduces dimensionality for large numbers of covariates; thus, matching on the propensity score, if the propensity score can be estimated well, leads to matched samples which can reduce estimation bias for high-dimensional observational studies.

6.1 Propensity Score Matching

Define the propensity score (Rosenbaum & Rubin, 1983) as the conditional probability of receiving the treatment given the covariates,

$$e\left(X_{ij}\right) \equiv Pr\left(W_{ij}=1|X_{ij}\right) = \mathbf{E}\left[W_{ij}|X_{ij}\right].$$

The propensity score summarizes all of the covariates into one scalar conveniently representing the probability of being treated. Rosenbaum & Rubin (1983) suggest the use of the propensity score to reduce the dimensionality of the matching problem with many covariates.

If we accept Assumption 4.2, the propensity score has two key properties: i) it is a balancing score. The distribution of covariates is the same among treated and control group units with identical propensity scores. Matching individuals with similar propensity scores replicates the randomized experiment scenario where group distributional differences can be attributed to randomness. ii) Treatment assignment is ignorable given the propensity score if treatment assignment is ignorable given covariates. Thus, matching on the true propensity

score is simpler, but equally as effective, as matching on the full set of covariates. However, it is often difficult to identify the true propensity score–the true assignment mechanism. Imbens and Rubin (2013) state that the estimated propensity score is often preferable to the true propensity score since better covariate balance, the ultimate goal of matching, is often achieved when using the estimated sample propensity score[40]. In this thesis we use propensity score matching as it is the most widely practiced matching technique but note that our method can easily be applied to Mahalanobis matching (Rubin 1980) or the innovative recent work on Coarsened Exact Matching (Iacus et al., 2011) and Genetic Matching (Diamond & Sekhon, 2013).

We now introduce matching on dyads and present: i) pre-matching dyad sample selection and its relation to network sparsity ii) identifying and controlling for covariates related to network structure. Section 6.2 presents a simple method for dyad sample selection. Section 6.3 presents ideas relating to dyad covariate selection, structural covariates, and matching on dyads. Section 6.4 summarizes our methodology.

6.2 Dyad Selection–Restriction Step

As explained in Section 5, the investigator matches dyads in pre-intervention treatment and control groups. Let $A_{ij} = \mathbb{I}_{\{Y_{ij}(T)\neq 0\}}$ where T represents the end of the analysis time period. Intuitively, A_{ij} represents which dyads are observed to interact, and therefore known to be at risk of interaction, during the analysis time period. Since our estimator adjusts for sparsity by considering only dyads observed to be at risk of interaction we select dyads for matching if $A_{ij} = 1$ and remove all dyads where $A_{ij} = 0$. Since our outcome of interest deals with interaction propensities, a confounding latent variable is whether or not dyads are actually at risk of interaction. Only using dyads which we know are at risk of interaction (since they have an observable interaction) controls for this confounding latent variable and also adjusts for network sparsity. Propensity score matching is performed on this restricted set of dyads.

[40] The propensity score is commonly estimated by using a logit of treatment status on the covariates and all significant interaction and quadratic terms.

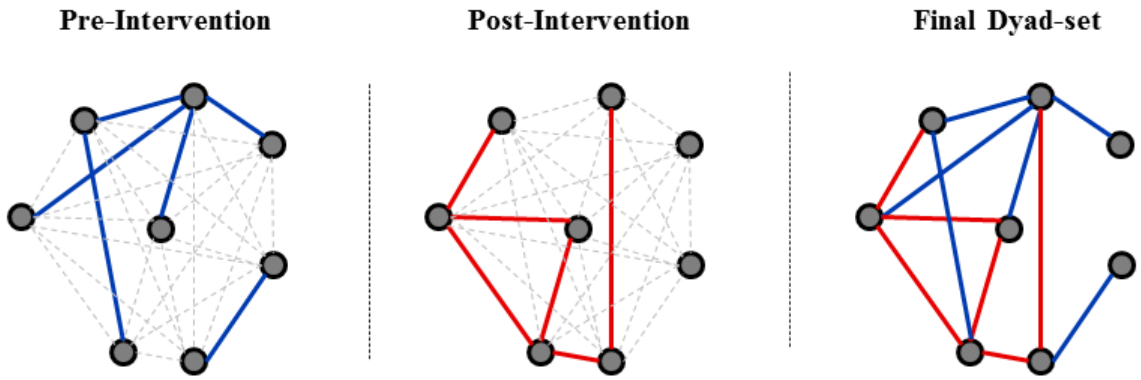


Figure 6.1: Restriction step to select dyad set before matching. (Left) Dotted gray lines represent the potential pairs of dyads. Blue lines represent dyads observed to have an interaction during the pre-intervention time period. (Center) Red lines represent dyads which do not interact during the pre-intervention period but have at least one interaction during the post-intervention period. (Right) After all the non-interacting dyads are removed (i.e., all dyads not known to be at risk of interaction are removed), the final dyad-set to perform matching on includes all the dyads observed to have an interaction. Note that differences in edge colors are purely for illustration purposes.

Selecting dyads in this manner is innovative because it considers a crucial post-treatment dyad covariate. Specifically, we select dyads which are included in $\mathbf{Y^A_{1,0,0}}$ and $\mathbf{Y^A_{0,0,0}}$ but peek ahead at future outcomes $\mathbf{Y^A_{1,1,1}}$ and $\mathbf{Y^A_{0,1,0}}$ *to include all dyads observed to be at risk of interaction.* It is often cautioned that matching should only be done using pre-treatment variables since post-treatment variables would be impacted by the treatment and could not be controlled for by the experimenter in the randomized scenario. If we consider whether a dyad is at risk of interaction as a binary covariate then by peeking ahead and including post-intervention dyads in matching we also consider post-treatment covariates. However, we review that estimation of each $\hat{\beta}_{tr}$ in the DID estimator is given by (Equation 5.4)

$$\hat{\beta}_{tr} = \text{argmax}_{\beta} \prod_{ij} P_{\beta} \left(Y_{ij} | A_{ij}, X_{ij}\right)^{A_{ij}}.$$

This regression estimator adjusts for dyad covariates (X_{ij}). Therefore, though we include a post-treatment covariate in our analysis we adjust our analysis accordingly by conditioning on this additional covariate. Calculation of $\hat{\beta}_{tr}$ requires we peek ahead and include all dyads observed to be at risk of interaction post-MARA in the analysis. Without peeking ahead, the estimate would be further biased since not all ij required for accurate calculation of $\hat{\beta}_{tr}$

would be included.

Note that the restriction step results in two separate types of dyads with different numbers of covariates. Since matching is performed only using pre-treatment covariates, dyads which are selected from post-intervention periods do not possess covariates defined in the pre-intervention period. Post-intervention dyads which are selected for matching do not interact at all in the pre-intervention period. Therefore, post-intervention dyads include node-covariates but do not include interaction-related covariates. We explain how to effectively perform matching for the two types of dyads (pre- and post-intervention dyads) in Section 6.3.

6.3 Matching on Dyads

After dyad selection the investigator must select edge and node covariates to match on, create structural covariates, adjust for matching across time, and consider matching process specifics. We address these issues in turn.

6.3.1 Edge, Node, and Structural Covariates

During matching each dyad must contain a single covariate vector; therefore node and edge covariates must be collapsed into a single dyad-metric. The restriction step (Section 6.2) is useful since it removes all dyads which do not interact during the analysis time period. Therefore, edge covariates (properties of the interactions) can be observed for each pre-intervention dyad; these are inherently properties relating to the dyad and do not need to be collapsed[41]. Node covariates are collapsed into binary dyad-covariates by considering covariate similarities, e.g., each inventor's company is a node-covariate; whether or not the inventor's work for the same company is a dyad-covariate. It is also important to consider network structure covariates to control for dyad placement and importance in the network.

Structural dyad covariates are included by collapsing node-specific structural covariates into dyad covariates. Since the outcome of interest deals with interactions, *it is necessary to control for the relationships between collaborating nodes.* Power and experience relationships

[41] If we did not restrict the set of dyads and performed matching over all $\binom{N}{2}$ potential dyads then it is not guaranteed that all dyads have edge covariates. This makes matching more difficult.

between actors in social networks will influence actors' interaction propensities; thus, controlling for covariates related to the relative importance of nodes is essential. For example, it is unlikely for a novice inventor to patent with a senior inventor, while it is more likely that two inventors of the same experience and influence in the inventor-network patent together. Dyad relationships and influence can be simply modeled with node centrality and cohesion measures and can be considered covariates that control for network structure. *Appendix B* reviews definitions of the centrality and cohesion measures used in this thesis.

A network's structure consists of the edges which define the landscape of node interactions. Centrality and cohesion measures represent the structure of a network by summarizing the relationships between the network's edges and nodes. For instance, the node degree distribution summarizes the relationship between the number of edges connected to each node. Similarly, the eigenvector centrality distribution summarizes the influence of each node on the network which when combined with the degree distribution provides additional insight into network structure. We argue that *combinations of structural node covariates can therefore control for network structure between groups of dyads.* Network structure confounds interaction-related outcomes since, intuitively dyads that have different roles (e.g., influence or centrality) in the network structure should have different interaction propensities. Node centrality and cohesion measures should therefore be used as a proxy for network structure.

The individual node structure measures can be collapsed into a single dyad-metric by adding and subtracting the node properties. Suppose node i has numeric covariate centrality or cohesion measure a and is connected to node j with numeric covariate centrality or cohesion measure b. We define the "experience-metric" of dyad ij as $a + b$ and the "relationship-metric" of dyad ij as $|a - b|$. The experience-metric represents the total dyad influence, while the relationship-metric represents the relationship between the nodes–a larger experience-metric implies that a dyad is more influential while a larger relationship-metric implies a larger imbalance in influence between nodes[42].

With dynamic networks, all pre-intervention covariates should be selected by considering only pre-interaction events, e.g., if a connection between two nodes i and j occurred at time

[42] In the MARA scenario a large experience-metric implies that two inventors are an influential pair, while a large relationship-metric represents that it is surprising that these two inventors are working together.

t then the dyad degree relationship-metric should be the difference in the total number of connections made by i and j prior to t. Selecting covariates based on pre-connection values accurately represents the dyad at the time of connection, *e.g.,* if a dyad consists of two highly influential inventors in 1984, but two novice inventors in 1981, then the 1981 dyad should be matched with a similarly unexperienced dyad even though the 1981 dyad later becomes highly influential.

For pre-intervention dyads all edge, node, and structural covariates should be matched over. For post-intervention dyads structural and edge covariates should not be included since the dyad has not interacted pre-intervention. Therefore post-intervention dyads should be matched only on collapsed node-covariates.

6.3.2 Quasi-exact Matching on Time

We recommend including the time of dyad interaction as a matching covariate. This reduces the likelihood of matches between dyads with similar covariates yet distant in time. For instance, a 1978 inventor-collaboration in Michigan and a 1984 inventor-collaboration in Ohio with similar covariates should not be matched. Though these dyads may appear similar, they should not be matched together since unobserved changes in baseline covariate values makes comparisons across time difficult[43].

Ideally, this issue could be resolved by performing exact matching on time, e.g., only permitting matches with dyads interacting at the same time; however, such exact matching would lead to small matched sample sizes. We recommend *quasi-exact matching on the time covariate* to maintain useful matched sample sizes. Quasi-exact matching on time disallows matches of interacting dyads which are separated by more than some investigator-specified

[43] The problem is outlined more formally here. Suppose there are identical two dyads ij and kl each with K continuous covariates, all of identical value between the dyads. We represent this identity by stating that the K-covariate vector for each dyad is drawn from the same $K-$variate probability distribution with PDF P. These dyads are separated in time, however. Let there be a random time trend $t(K)$ which modifies the $K-$variate PDF. In order for these two dyads to actually be considered exact matches the random time trend must have no impact on the initial PDF P. We are interested in the scenario where the some distance measure $D(\cdot,\cdot)$ between two multi-variate PDFs is 0. We seek $D(K, t(K)) = 0$. From change of variables we can rewrite as $D\left(P, P\left|\frac{\partial P}{\partial t}\right|\right)$ where $\left|\frac{\partial P}{\partial t}\right|$ represents the determinant of the Jacobian matrix. $D\left(P, P\left|\frac{\partial P}{\partial t}\right|\right) = 0$ iff $\left|\frac{\partial P}{\partial t}\right| = 1$. Intuitively, this states that the exact matching between these two dyads is correct only in the unlikely scenario that the time trend has not had any notable impact on dyad covariates. We therefore note that time should be included as a covariate to match over in order to adjust for this unobserved time trend.

time t_{max} regardless of covariate similarity. Since baseline covariate values change over time, when interactions are distantly separated in time it is not pragmatic to compare dyad covariate values. The investigator should choose t_{max} by considering how various t_{max} values impact matched sample size and covariate balance. We recommend using the lowest t_{max} value which results in a large matched sample size and strong balance. Quasi-exact matching on time is not required for post-intervention dyads since post-intervention dyads have not interacted in the pre-intervention period where quasi-exact matching on time would be performed.

6.3.3 Matching Specifics

We match using nearest neighbor propensity score matching with calipers disallowing matches between dyads which have a propensity score difference greater than .2 standard deviations of the linear propensity score. This caliper value is supported by Rosenbaum & Rubin (1985) and Austin (2011)[44]. We match 1 to 1 and also match with replacement, which often decreases bias since dyads in the control group similar to many treated dyads can be matched more than once (Stuart, 2010).

Truncated matching results in two types of dyads, those from the post-intervention period and those from the pre-intervention period. As explained in Section 6.2 post-intervention dyads have only node covariates while pre-intervention dyads have node, edge, and structural covariates. Exact matching should therefore be done on a constructed binary covariate $= \begin{cases} 1 & \text{if pre-intervention dyad} \\ 0 & \text{if post-intervention dyad} \end{cases}$ to disallow matches between pre- and post-intervention dyads.

6.4 Methodology Summary

Our methodology for estimating causal effects on interaction propensity has now been fully outlined. There are six steps which an investigator should follow:

[44] Austin (2011) runs simulations to show that using calipers of .2 standard deviations minimizes the MSE of the estimated treatment effect and also eliminates at least 98% of the bias of the estimator.

1. *Select ego sets* based on some criteria which defines a fixed set of nodes over the analysis time period.
2. *Restrict dyad set* to dyads which have had at least one observed interaction. Two distinct dyad groups for matching arise: pre- and post-treatment dyads.
3. *Select edge, node, and structural covariates* to match over. Collapse node and structural covariates into a single dyad-metric.
4. *Implement matching* on edge, node, and structural covariates for pre-treatment dyads, and on node covariates for post-treatment dyads. Use quasi-exact matching on time for pre-treatment covariates.
5. *Choose Conditionally Independent Relationship Model* regression and variables of interest to perform the regression over.
6. Using matched dyad samples *calculate the partial likelihood* from the CIR regression model for each of the four $\hat{\beta}_{tr}$ in the matched DID estimator to estimate the causal effect.

We now illustrate our methodology using a longitudinal co-patent network dataset.

7 Empirical Illustration

We use the MARA natural experiment to estimate the causal effect of non-compete laws on patent collaboration propensities of Michigan inventors. We first assess our use of DID for investigating the impact of the MARA natural experiment by highlighting the threats to validity of DID estimators.

7.1 Threats to the Validity of Differences-in-Differences Estimators

Campbell (1969) and Meyer (1995) identify multiple threats to the validity of DID estimators[45]. We address relevant threats and note their importance to our study of the MARA natural experiment:

[45] A more modern critique of DID estimators is levied by Bertrand et al. (2004). We address this critique in Section 7.6.

1. Omitted variables: events occurring between pre-treatment and post-treatment may also impact results. Marx et al. (2009) run placebo regressions placing the MARA reform in 1984 and 1986 and find significant effects in hypothesized inventor mobility only if treatment occurs during MARA's actual implementation year (1985). Marx et al. posit that inventor mobility and inventor interaction propensity are strongly related. We therefore assume that no other significant events skew results.
2. Trends in outcomes: processes impacting the outcome of interest change with time. This threat states that Assumption 5.1, the parallel trend assumption, is impractical. In *Appendices J and K*, which summarize the co-patent dataset, we note that pre-treatment covariate distributional summaries are similar for Michigan and Ohio for 3,5, and 7 year periods pre-MARA. Further, Marx et al. (2009) note that Michigan and Ohio had similar, declining economies in the 1980s and we therefore continue to follow Assumption 5.1.
3. Interaction of history and treatment: the treatment effect may differ across time periods. We estimate the impact of the treatment on 3,5, and 7 year time-intervals around MARA to address this issue. It is unclear whether the impact or knowledge of MARA changed over time. Marx et al. (2009) do not find any change in their treatment effect of interest when varying time-interval size around MARA. This implies that the MARA treatment could be homogenous across time, but the answer is unclear.
4. Attrition: loss of units from treatment and control groups. Marx et al. (2009) show that MARA resulted in a mass exodus of inventors from Michigan to other states, mainly California. This attrition is not problematic since, as explained in Section 3, we restrict our inventor set to those inventors who remain in the state pre- and post-1985. There is therefore no loss of nodes, and therefore dyads, from treatment and control states.

We make many assumptions to avoid these validity threats and therefore must be careful when interpreting the DID estimate. The rest of this section illustrates the main steps of our methodology outlined in Section 6.4.2. These include: i) node and dyad sample selection ii) covariate selection iii) matching implementation and balance assessment iv) regression model identification v) estimation of the treatment effect.

7.2 Node and Dyad Selection

For this empirical illustration Michigan is the treatment state and Ohio is the control state[46]. We consider 5-year intervals around MARA, which took place in mid-1985. Since we consider 5-year intervals our analysis is performed using the years 1980-1990. We restrict the node set to inventors who had at least one patent prior to 1980 in MI and OH and did not patent outside of the respective state from 1980-1990. Marx et al. (2009) state that the impacts of MARA would begin only in 1986; we therefore consider January 1980-December 1984 as the pre-intervention period and January 1986-December 1990 as the post-intervention period. We illustrate the specifics of matching on dyads for the 5-year surrounding interval but report our final causal estimates by using our methodology on 3-,5-, and 7-year intervals pre- and post-MARA[47].

To satisfy to required input of the sparsity-adjusted estimator we perform the restriction step (Section 6.2) to restrict the pre-matching dyad sample to dyads which interacted at least once during the analysis time interval.

	Pre-restriction	**Post-restriction**		
		Pre-intervention	Post-intervention	Total
MI	$\binom{3629}{2}$	1412	1903	3315
OH	$\binom{4465}{2}$	2117	1498	3615

Table 1: Initial numbers of dyads, and dyads from pre- and post-intervention periods remaining post-restriction step

Table 1 illustrates that the "peeking ahead" aspect of the restriction step significantly increases the sample size of pre-treatment dyads by including all dyads observed to be at risk of interaction. The restriction step also leads to a pre-matching set of dyads much

[46] *Appendices J and K* show that Ohio is a strong choice of control state.

[47] A 3-year interval results in an analysis period from 1982-1988. A 5-year interval results in an analysis period from 1980-1990, and a 7-year interval results in an analysis period from 1978-1992. Nodes are restricted based on whether they have at least one patent prior to 1982, 1980, and 1978 for the 3-,5-, and 7-year interval case respectively. Inventors also must not patent outside of the respective state during the respective analysis time period. *Appendix C* summarizes the 3-,5-, and 7-year interval networks before and after matching.

smaller than the initial set of potential dyads thereby removing bias caused by including dyads not at risk of interaction during analysis.

7.3 Covariate Selection

Edge covariates include properties of the patents and edge betweenness centrality. Patent properties represent edge covariates along with edge betweenness. We collapse node covariates by noting similar characteristics shared by inventors, e.g., same company or same zip code. For structural covariates we consider all the cohesion and centrality measures defined in *Appendix B*[48]. When collapsing structural covariates we are left with two dyad-metrics for each centrality and cohesion measure considered–the 'relationship' and 'experience' metric as defined in 6.3.1. *Appendix D* presents the four node covariates, three edge covariates, and eight structural covariates we use for matching dyads.

Pre- and post-intervention dyads are matched separately. Pre-intervention dyads contain edge, node, and structural covariates. Post-intervention dyads contain only node covariates. Exact matches are found for post-intervention dyads since the number of covariates is small. We go in depth into matching for the pre-intervention dyads in the following section.

7.4 Matching

For matching on pre-intervention dyads we go over three main steps: i) selecting which quadratic and interaction terms to add to the propensity score model ii) estimating and evaluating the linear propensity score iii) choosing optimal t_{max} value for quasi-exact matching.

7.4.1 Selecting Quadratic and Interaction Terms

Interaction terms and quadratic terms should be specified in the propensity score model. We use an algorithm, described in Imbens & Rubin (2013), to select interaction and quadratic terms. The algorithm yields five interaction and quadratic terms, which in addition to the

[48] Betweenness centrality and closeness centrality values were negligible for most nodes and are therefore not considered.

15 node, edge, and structural covariates previously selected, yields 20 total terms for use in the propensity score model. *Appendix E* presents the 20 linear, quadratic, and interaction terms used in the propensity score model.

7.4.2 Estimating the Propensity Score

We run the logit regression of treatment assignment (dyad location) against the 20 linear, interaction, and quadratic covariates to estimate the propensity score model. *Appendix F* summarizes the propensity score model fit.

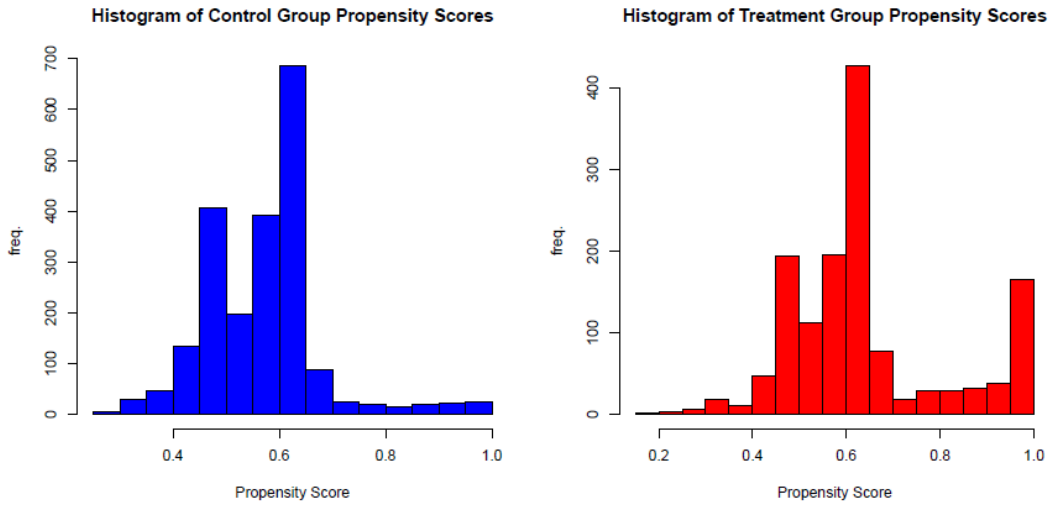


Figure 7.1: Histograms of propensity scores for control (OH) and treatment groups (MI)

The treatment and control group propensity scores overlap heavily between propensity score values of .4-.6. This overlap is notable since it allows us to focus directly on the data and not rely on model specification for extrapolation (Stuart, 2010).

7.4.3 Choosing t_{max} for quasi-exact matching

To select the optimal t_{max} value, the value which defines the maximum time distance allowed between matched interacting dyads, we vary $t_{max}$to examine its impact on post-matching sample size and covariate balance. For the covariate balance tests we examine QQ-statistics

of maximum QQ-difference, median QQ-difference, and mean QQ-difference. *Appendix G* presents the impact of varying t_{max} on QQ-statistics when performing propensity score matching with calipers of .2 standard deviations of the propensity score. We find that the t_{max} value has no significant impact on the QQ-statistics and hence no significant impact on covariate balance.

We next test the impact of varying t_{max} on the matched treatment group dyad sample size.

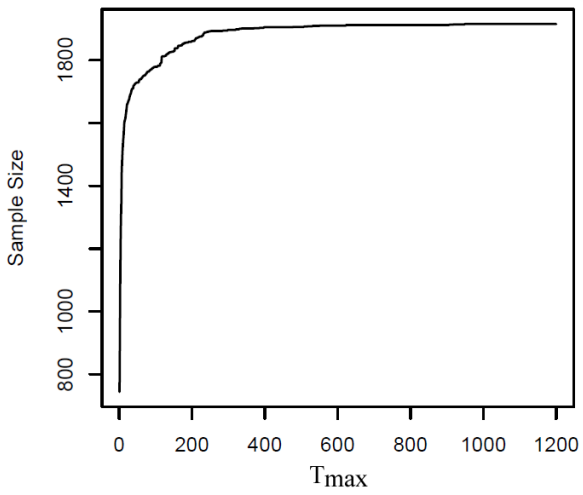


Figure 7.2: Treatment Group Matched Dyad Sample Size for Various t_{max} Values

Sample size is significantly impacted by the value of t_{max}. t_{max} of approximately 250 is optimal since this is the lowest t_{max} value for which the sample size is close to the asymptotic sample size. For our analysis we therefore use a t_{max} value of 250 days during matching.

Using a t_{max} value of 250 days, we can now check if a caliper of .2[49] standard deviations is appropriate. Statistics proposed by Rubin (2001) provide a quick summary of covariate balance post-matching: the standardized difference of means (SDM) and the variance ratio (VR) of treatment and control group propensity scores. Let the propensity score of dyads in treatment group W be given by $e_W(X)$ then $SDM = \frac{e_1(X)-e_0(X)}{\sqrt{\frac{var(e_1(X))+var(e_2(X))}{2}}}$ and $VR =$

[49] Recommended by Rosenbaum & Rubin (1985) and Austin (2011).

$\frac{var(e_1(X))}{var(e_0(X))}$. Rubin (2001) notes that if $.5 < VR < 2$ and $0 < |SDM| < .25$ then balance can be considered good. Figure 7.3 examines how changing caliper size impacts the SDM and VR.

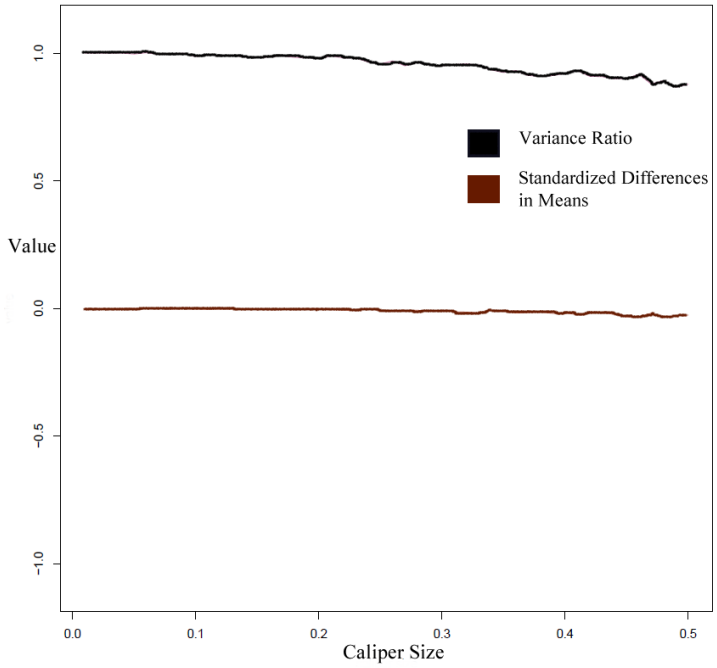


Figure 7.3: Impact of Caliper Size on Covariate Balance (VR and SDM)

Match balance is strong for all caliper sizes from .0-.5; however, the variance ratio becomes more distant from the ideal value of 1 as caliper size increases above .25. The .2-.25 standard deviation caliper recommended by Rosenbaum & Rubin (1985) and Austin (2011) is therefore appropriate for this analysis.

Using the Matching package in R (Sekhon, 2008) we perform 1 to 1 propensity score matching with replacement. We utilize a caliper of .2 standard deviations, perform exact matching on a constructed binary covariate of whether dyads were drawn pre- or post-intervention, and perform quasi-exact matching on time disallowing matches between dyads separated by more than the t_{max} of 250 days.

7.5 Match Balance Assessment

Assessing the quality of the resulting matched units is the most important step with matching methods (Stuart, 2010). Although commonly used for assessing balance, hypothesis tests (e.g., Kolmogorov-Smirnov tests and t-tests) should not be used as balance measures (Austin, 2007; Imai, King, and Stuart, 2008) since they are dependent on sample sizes and therefore do not distinguish between changes in balance and changes in statistical power. For continuous covariates we therefore examine balance using the absolute standardized difference in means (ASDM), variance ratio (VR), and QQ-plots as supported by Rubin (2001) and Ho et al. (2007). For binary covariates we examine balance using the ASDM and VR.

Appendix H presents the ASDM and VR for each covariate and the propensity scores for MI and OH, and the number of matched samples determining each statistic. Rubin (2001) notes that balance is strong if $0 < |SDM| < .25$ and $.5 < VR < 2$. Based on these criteria we see that balance on the propensity score and all covariates, both continuous and binary, is strong.

Appendix I presents QQ-plots[50] and QQ-statistics for the continuous covariates. The QQ-plots imply strong distributional similarity for all covariates, except for edge betweenness. However, all covariates, including edge betweenness, have low QQ-statistics.

Seeing that the balance is strong the treatment effect can now be estimated using the matched dyad samples.

7.6 Treatment Effect Estimation

We use a Cox proportional hazards regression and a Cox partial likelihood for inference of $\hat{\beta}_{tr}$ and standardly assume that assume that the Cox model has a constant baseline hazard. The Cox proportional hazards model is a natural choice of model for counting processes on networks (Perry & Wolfe, 2013). We use the standard Cox proportional hazard regression where $Y_{ij}(t)$ is a counting process of interactions between actors i and j representing the number of interactions between i and j at time t. The truncated partial likelihood if using

[50] The QQ-plot is interpreted as follows: if the points fall on the 45^o line then the distribution of the covariate for the groups is identical.

a general Poisson Process as an approximation to the Cox regression is given by[51]:

$$\begin{aligned} L^{tr}\left(\mathbf{Y}|\beta\right) &= \prod_{ij:Y_{ij}(T)>0} \frac{exp\left(-\int_0^T \lambda_{ij}\left(s\right)ds\right)\prod_{k=1}^{Y_{ij}(T)}\lambda_{ij}\left(t_{ij}^{(k)}\right)}{1-exp\left(-\int_0^T \lambda_{ij}\left(s\right)ds\right)} \\ &= \prod_{ij:R_{ij}=1} \frac{P\left(Y_{ij}|R_{ij}=1\right)}{1-P\left(Y_{ij}\left(T\right)=0|R_{ij}=1\right)}, \end{aligned}$$

where T is the length of the time interval each $\hat{\beta}_{tr}\left(\cdot\right)$ in Equation 5.9 is computed over, and $\lambda_{ij}\left(t\right)$ is the instantaneous interaction rate between dyads i and j at time t.

We calculate the DID estimator with Ohio as a control state for 3,5, and 7 year intervals around MARA. We calculate each $\hat{\beta}_{tr}\left(\cdot\right)$ in the DID estimator by maximizing the truncated Cox partial likelihood. We run the Cox regression with three time-dependent binary variables which summarizes interaction propensity between inventors.

- *Int*: Whether two inventors have patented previously, i.e., $Y_{ij}\left(t\right)>0$. $\hat{\beta}_{tr,Int}$ represents the instantaneous log change in interaction rate if two inventors have patented previously
- *Comp*: Whether two inventors work for the same company. $\hat{\beta}_{tr,Comp}$ represents the instantaneous log change in interaction rate if two inventors begin to work for the same company
- *Zip*: Whether two inventors share the same zip code. $\hat{\beta}_{tr,Zip}$ represents the instantaneous log change in interaction rates if two inventors begin to share the same zip code.

The CIR regression model is given by:

$$Y_{ij}\left(t\right)|X_{ij}\left(t\right),\mathcal{F}\left(t\right) \sim CP\left(\lambda_{ij}\left(t\right)\right), \tag{7.1}$$

$$log\lambda_{ij}\left(t\right) \equiv \beta_0 + \beta_{tr,Int}\cdot Int_{ij}\left(t\right) + \beta_{tr,Comp}\cdot Comp_{ij}\left(t\right) + \beta_{tr,Zip}\cdot Zip_{ij}\left(t\right)$$

We assume that coefficients are independent from each other and therefore can add variances to calculate the variance and standard error of the DID estimator. Truncated

[51] We present the general poisson process likelihood merely to illustrate the likelihood form. The actual likelihood we maximize takes the form of a truncated Cox partial likelihood.

inference using the CIR regression modle yields the following results:

		pre-MARA		post-MARA		**DID Estimate**
		MI	OH	MI	OH	
3-Year	$\hat{\beta}_{tr,Int}$(S.E)	-.033 (.18)	.096 (.26)	.53 (.17)***	-.064 (.25)	.72 (.44)
	$\hat{\beta}_{tr,Comp}$(S.E)	.21 (.22)	.40 (.31)	.095 (.16)	.087 (.27)	.20 (.49)
	$\hat{\beta}_{tr,Zip}$(S.E)	.036 (.18)	-.35 (.26)	-.22 (.14)	.051 (.24)	-.66 (.42)
5-Year	$\hat{\beta}_{tr,Int}$(S.E)	-.081 (.13)	-.34 (.16)**	.018 (.17)	.63 (.26)**	-.87 (.37)**
	$\hat{\beta}_{tr,Comp}$(S.E)	.17 (.16)	-.056 (.19)	-.019 (.13)	-.31 (.19)*	.065 (.34)
	$\hat{\beta}_{tr,Zip}$(S.E)	.18 (.11)*	-.10 (.16)	-.030 (.14)	-.14 (.24)	-.17 (.34)
7-Year	$\hat{\beta}_{tr,Int}$(S.E)	-.81 (.08)***	-.52 (.10)***	.25 (.14)*	.51 (.29)*	.030 (.35)
	$\hat{\beta}_{tr,Comp}$(S.E)	-.14 (.081)*	-.037 (.11)	-.089 (.15)	.26 (.22)	-.25 (.29)
	$\hat{\beta}_{tr,Zip}$(S.E)	.17 (.073)**	.17 (.093)*	-.19 (.15)	-.073 (.27)	-.26 (.33)

*** Significant at 1 %. ** Significant at 5%. * Significant at 10%.

Table 2: Regression Results from Truncated Inference

$\hat{\beta}_{tr,Int} = -.87$ for the 5-year interval (time interval 1980-1990) scenario is the only significant DID estimate. This result can be interpreted as follows: the passage of MARA, which allowed the enforcement of non-compete laws, reduced the propensity for collaboration among Michigan inventors who had previously worked together by a factor of $e^{.87} \approx 2.4$. Interestingly, when considering 3-year and 7-year intervals pre- and post-MARA, there is no significant impact on interaction propensity among inventors who had worked together previously. It is interesting that the DID estimator for $\hat{\beta}_{tr,Comp}$ is not significantly positive for any analysis time interval. We would expect the enforcement of non-compete laws to decrease inter-firm mobility thereby increasing interaction propensity among inventors within the same firm.

We interpret these results cautiously for a number of reasons.

1. Our analysis is dependent on patent data, which is merely a proxy for inventor collaborations. Many inventors will work together and share ideas before a patent is released; patent data does not note all of these interactions.

2. To calculate the DID estimator we make the strong assumption that the coefficients in the Cox regression are independent of each other. In actuality, there will be overlapping effects. For instance, inventors working for the same firm are more likely to have worked together previously.
3. Bertrand et al., (2004) note that DID estimators are often not trustworthy since not considering lagged correlation between outcomes leads to biased standard errors. They note that collapsing the data to adjust for time-series variation can solve this problem.
4. We have not tested whether the proportional hazards assumption is viable for this data.
5. There is a loss of statistical power from both the matching procedure and truncated inference.

We urge researchers in fields of industrial organization and innovation to apply our method on the co-patent dataset with time-series-adjusted data and more controls to propose more confident interpretation of results.

8 Discussion and Future Work

In this thesis we have identified a notable gap in the literature at the intersection of network science and causal inference–the lack of observational studies or methods for identifying causal effects on dyad interaction propensity–and have illustrated the suitability of matching to this problem. We intend for this thesis to contribute to causal-inference-on-networks research in four ways:

1. Inspired by the idea of sparsity misspecification for network models (D'Amour & Airoldi, 2014) we have defined sparsity in the finite population context and introduced a matched differences-in-differences estimator which adjusts for this sparsity.
2. We have introduced matching on dyads and have presented a justification and procedure for including network structure as an essential covariate to match over.
3. We have illustrated the suitability of a novel class of models (CIR) for causal effects estimation on sparse networks.

4. We have presented the first empirical study testing causal effects on interaction propensities within networks and mentioned many practical causal questions suited to our methodology.

Our method is based on a simple idea which accounts for network sparsity–perform analysis and matching only on dyads which have had an interaction previously and therefore are known to be at risk of interacting. Therefore, as the causal inference and network literatures develop we believe our method will still present relevant ideas for observational studies estimating interaction-related causal effects on sparse networks. Still, there are many open areas where our method requires further development:

1. We have assumed that interference effects are not present between dyads. We also assume that the risk set is constant and not impacted by exposure to treatment. These are strong assumption and ways to relax them should be investigated. A first step could be to extend CIR models to allow risk graphs to change with time.
2. Simulations and theory can lend more concrete advice about which t_{max} values are optimal.
3. We have assumed that the structural network measures chosen adequately control for network structure. Theoretical work about combinations of such network measures could prevent potential overmatching and could also reveal which network measures can optimally control for network structure.
4. In this thesis we focus on propensity score matching. Improved covariate balance and thus less biased estimates may potentially be achieved by alternative matching techniques. Perhaps there exists heightened suitability of a few matching schemes for the networked context. For example, Genetic Matching (Sekhon & Diamond, 2013) optimizes for covariate balance using a genetic algorithm. Genetic Matching may be able to optimize over time differences between dyads thereby removing the need for quasi-exact matching on time.
5. It is unclear whether estimates from fixed node sets are applicable to the population of nodes the fixed node sets are drawn from.
6. It is likely that causal estimates on interaction propensities can be further improved by using alternative, more intricate quasi-experimental designs.

Causal inference on networks is new and must be performed with caution. Especially when dealing with observational data all confidence about the applicability of methods or the preciseness of results, should be, and often is, quickly dismissed. Through e-mail correspondence, Donald Rubin, the pioneer of modern causal inference, told me, "I think that some huge percentage of the stuff that is written in the field of [causal inference on networks] is as bad as the stuff on simple causal inference was three decades ago (and some still hasn't improved, but most has), and so there appears to be a longish road ahead to clarity." Further theoretical and empirical work is required in this area before researchers can justifiably include these methods in their toolkits. The dream of precise causal inference on networks is distant, but the main problem areas are slowly becoming clearer. We hope to have elucidated a new area of focus that is ripe with interesting theoretical and empirical problems to make the "longish road ahead" look slightly more walkable.

9 Appendices

Appendix A: Problems with Using the ATTR

The ATTR is given by $\tau_{treated,risk} = \frac{\sum_{ij:R_{ij}=1} Y^{(1)}(1)|W_{ij}=1 - Y_{ij}^{(1)}(0)|W_{ij}=1}{\sum R_{ij}}$. Noting that

$$\begin{aligned} E\left[Y_{ij}|R_{ij}=1\right] &= P\left(A_{ij}=1|R_{ij}=1\right)\mathbf{E}\left[Y_{ij}|A_{ij}=1\right] \\ &= P\left(A_{ij}=1|R_{ij}=1\right)\mathbf{E}\left[Y_{ij}|A_{ij}=1\right] + P\left(A_{ij}=0|R_{ij}=1\right)\mathbf{E}\left[Y_{ij}|A_{ij}=0\right] \end{aligned}$$

we can rewrite the ATTR as

$$\frac{\sum_{ij:R_{ij}=1}\left(Y_{ij}^{(1)}(1)|W_{ij}=1\right)A_{ij}P(A_{ij}=1|R_{ij}=1)}{\sum R_{ij}} - \frac{\sum_{ij:R_{ij}=1}\left(Y_{ij}^{(1)}(0)|W_{ij}=1\right)A_{ij}P(A_{ij}=1|R_{ij}=1)}{\sum R_{ij}} .$$

Unfortunately, the $P\left(A_{ij}=1|R_{ij}=1\right)$ term is impossible to know. Like Q it represents the probability of an actual interaction occurring between a dyad at risk of an interaction. For instance, knowing $P\left(A_{ij}=1|R_{ij}=1\right)$ would be akin to knowing how likely two children in the same classroom are of becoming friends. Like Q, $P\left(A_{ij}=1|R_{ij}=1\right)$ represents the social scenario. It is unclear whether truncated inference for the ATTR–performing estimation using only those dyads with observed interactions–leads to reasonable estimates. To bypass this issue by performing inference independent of the social scenario we refer to D'Amour & Airoldi (2014).

Appendix B: Network Centrality and Cohesion Measures

The centrality of a vertex measures its importance in the network. The main measures of centrality include: degree, betweenness, closeness, and eccentricity. We also define the 2-degree as a measure of cohesiveness.

Definition. 6.1. (Degree Centrality). *The degree centrality of vertex i in graph $G = (V, E)$ with $|V|$ vertices and $|E|$ edges is given by $C_D(i) = deg(i) = \sum_j^{|V|} G_{ij}$.*

Degree centrality is the number of ties a node has and is often is interpreted as how likely a node is to receive or capture flows (e.g., information or disease) through a network.

Definition. 6.2. (Closeness Centrality). *The closeness centrality of vertex i in graph G is given by $C_C(i) = closeness(i) = \frac{1}{\sum_j d_{ij}}$ where d_{ij} is the shortest distance between nodes i and j.*

Closeness centrality is a measure of how easily information can spread from a node.

Definition. 6.3. (Betweenness centrality). *The betweenness centrality of vertex i in graph G is given by $C_B(i) = betweenness(i) = \sum_{s \neq i \neq t \in V} \frac{\sigma_{st}(i)}{\sigma_{st}}$ where σ_{st} represents the shortest paths from node s to node t, and $\sigma_{st}(v)$ represents the number of these shortest paths which pass through node i.*

Betweenness centrality measures how likely a node is to act as a bridge between information transfers between other nodes, e.g., to what extent the node acts as a connector between other nodes.

Definition. 6.4. (Eccentricity). *The eccentricity of vertex i in graph G is given by $\epsilon(i) = eccentricity(i) = max\, d(i,j) \forall i < j \in V$ where $d(i,j)$ represents the distance between vertices i and j.*

Vertex eccentricity is another centrality measure. Nodes with lower eccentricities are considered more central to the network.

Cohesion is a crucial sociological concept which represents how closely bonded a group is. We recommend modeling this using the 2-degree, the sum of the degrees of an individual's connections.

Definition. 6.5. (2-degree). *The 2-degree of vertex i in graph G is given by* $deg_2(i) = \sum_j^{|V|} G_{ij} deg(j)$.

The 2-degree measures how well connected an actor's connections are. If the 2-degree is higher then an individual is intuitively part of a more cohesive group.

We recommend edge betweenness as an additional edge covariate. Edge betweenness is defined as the number of shortest paths between all nodes in a network which pass through a certain edge. Intuitively, edge betweenness represents how important an edge is for connecting communities of nodes. For instance, a certain interdisciplinary patent may act as an influential bridge connecting groups of inventors working on patents in separate disciplines [52].

[52] Node centrality measures define a node's popularity or influence within the network while edge betweenness defines an edge's (e.g., a patent's) influence within the network.

Appendix C: Summary of 3,5,7 Year MI and OH Networks

		# Vertices | #Edges	
# Years around MARA		Pre-Matching	Post-Matching
3-Year	MI	3295 | 2301	3295 | 1896
	OH	4032 | 3102	4032 | 1658
5-Year	MI	3629 | 2812	3629 | 1924
	OH	4465 | 3615	4465 | 1750
7-Year	MI	3519 | 2913	3519 | 2476
	OH	4354 | 2560	4354 | 2095

Table 3: Number of Vertices and Edge Pre- and Post-Matching for Michigan and Ohio 3,5,7 Year Networks.

Appendix D: Covariates Used for Matching

Covariate Name	**Description**
Node/Edge Covariates	
SameCity	Binary: inventors lived in same city at time of patent
SameZip	Binary: inventors shared same zipcode at time of patent
ShareClass	Binary: inventors patented previously in same tech class
ShareCompany	Binary: inventors patented previously for same company
EdgeBet	Edge Betweenness
Co-Auth	# of previous co-authors shared
PatentTime	Time of patent in days (1 being day of first patent in time period)
Structural Covariates	
E-PPI	Experience-metric of # Patents per inventor
R-PPI	Relationship-metric of # Patents per inventor
E-Eccent	Experience-metric of node eccentricity
R-Eccent	Relationship-metric of node eccentricity
E-Deg	Experience-metric of node degree
R-Deg	Relationship-metric of node degree
E-Deg_2	Experience-metric of node 2-degree (cohesion measure)
R-Deg_2	Relationship-metric of node 2-degree (cohesion measure)

Table 4: Description of all Edge, Node, and Structural Covariates Used

Appendix E: Algorithm for Determining Interaction Terms (Imbens & Rubin, 2013)

The treatment and control group dyads consist of a K-component vector of covariates X_{ij}. The goal is to select a subset of these K covariates as linear covariates, as well as a subset of the $K(K+1)/2$ possible second order terms including both quadratic and interaction terms. In *step 1* K_B basic covariates are selected because the researcher believes they significantly impact treatment assignment. We ignore this stage since our lack of huge numbers of initial covariates mean all of our covariates are considered crucial basic covariates. In *step 2* all significant quadratic and interaction terms between the K_B step 1 covariates are selected, e.g., decide which of the $K_B(K_B+1)/2$ quadratic and interaction terms involving these K_B covariates should be included. If at some point K_Q of the $K_B(K_B+1)/2$ possible interaction terms have been included, then estimate $K_B(K_B+1)/2 - K_Q$ logistic regressions, each of which includes the intercept, the K_B basic linear terms and the K_Q second-order terms already selected, and one of the remaining $K_B(K_B+1)/2 - K_Q$ terms. Calculate the likelihood ratio statistic with the null hypothesis that the most recently added second order term has a zero coefficient . If this likelihood ratio is significant at the 95% level then compare this ratio it to all of the other significant likelihood ratio tests of the $K_B(K_B+1)/2 - K_Q$ interaction terms and continue including the term with the largest likelihood ratio statistic until all remaining likelihood ratio statistics are not significant at the 95% level.

Performing this algorithm we select the following interaction & quadratic terms.

(E-Eccent)^2
(R-Deg$_2$)^2
(E-PPI)*(R-Eccent)
(PatentTime)*(EdgeBet)
(E-PPI)*(EdgeBet)

Table 5: Interaction & Quadratic Terms Selected

Appendix F: Output from Propensity Score Specification

Terms	Estimate (S.E.)
Linear Terms	
(Intercept)	.347 (.41)
SameCity	-1.88 (.64)**
SameZip	1.32 (.63)**
PatentTime	0 (0)
ShareClass	.949 (.61)
ShareCompany	.145 (.12)
EdgeBet	.0567 (.011)***
E-PPI	.0357 (.016)**
R-PPI	-.059 (.019)***
E-Eccent	-.0628 (.021)***
R-Eccent	-.236 (.091)**
E-Deg	-.063 (.021)***
R-Deg	.046 (.026)*
E-Deg_2	.0071 (0)***
R-Deg_2	-.0071 (.22)***
Co-Auth	.136 (.042)***
Interaction+Quadratic Terms	
$(\text{E-Eccent})^2$	.0494 (5.31E-03)***
$(\text{R-Deg}_2)^2$	6.29E-05 (1.57E-05)***
(E-PPI)*(R-Eccent)	.0298 (1.10E-02)***
(PatentTime)*(EdgeBet)	8.14E-05 (2.69E-05)***
(E-PPI)*(EdgeBet)	3.17E-03 (1.67E-03)*

*** Significant at 1 %. ** Significant at 5%. * Significant at 10%.

Appendix G: The Impact of t_{max} on Covariate Balance (QQ-statistics)

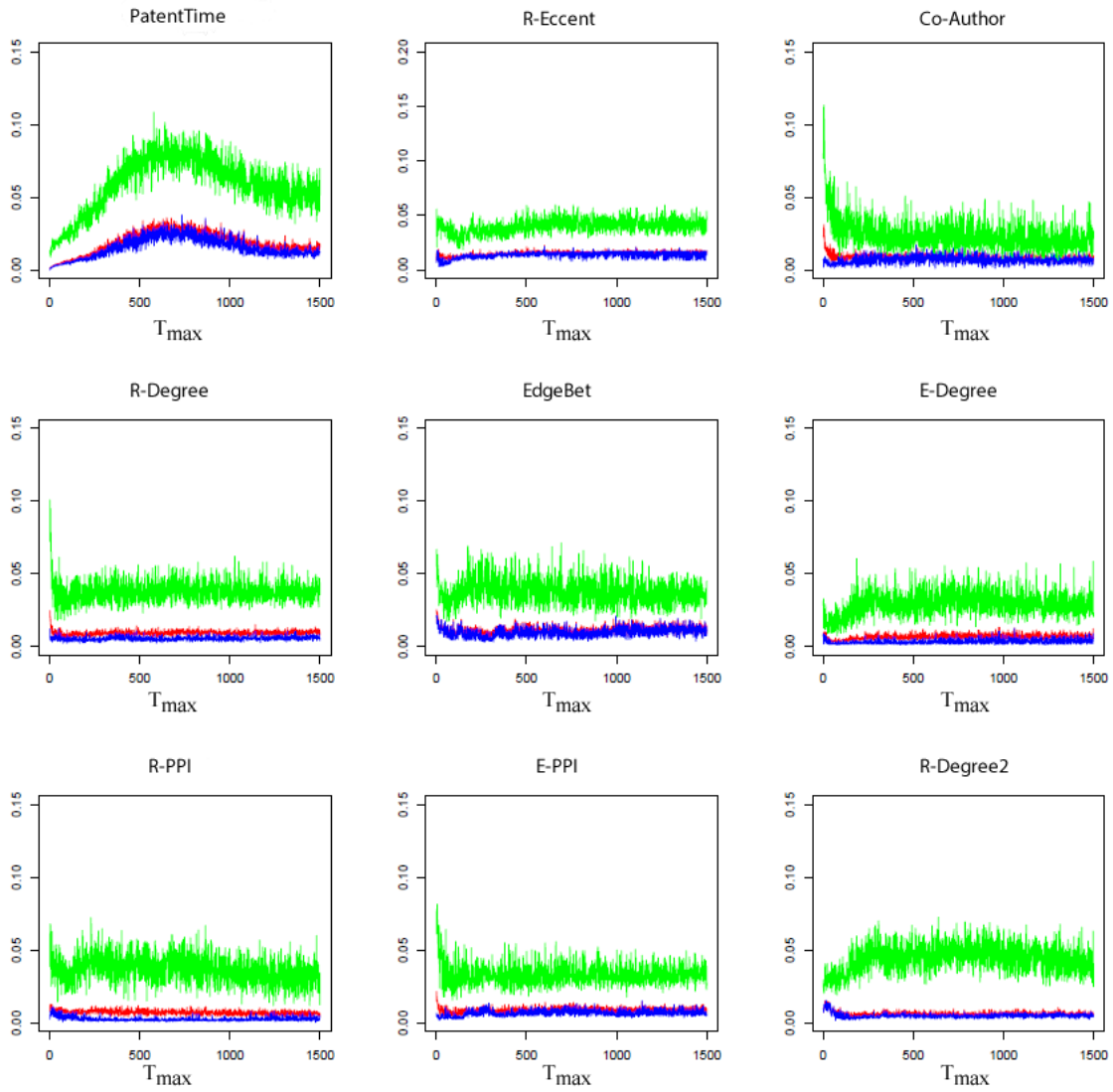
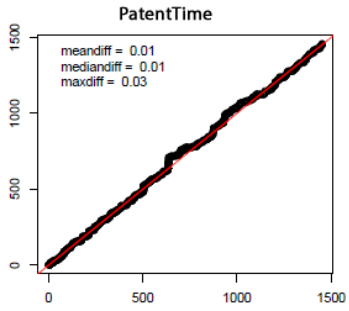


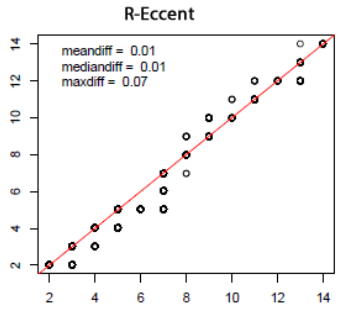
Figure 9.1: QQ-statistics for continuous covariates. (Green) Maximum QQ-difference (Blue) Mean QQ-difference (Red) Median QQ-Difference. We see that changing the value for t_{max} has no impact on covariate balance.

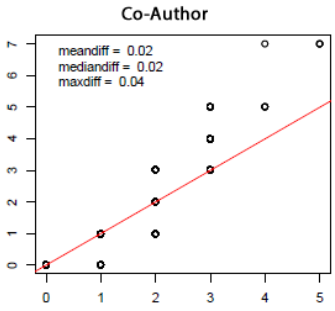
Appendix H: Standardized Difference in Means and Variance Ratios for Propensity Score and Covariates

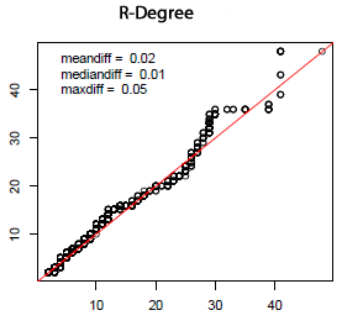
Covariate Name	VR	SDM	Number Dyads (MI|OH)
Node/Edge Covariates			
SameCity	1.04	.029	1924|1750
SameZip	1.06	.044	1924|1750
ShareClass	.923	.0079	1924|1750
ShareCompany	1.04	.0041	1924|1750
EdgeBet	1.33	.053	994|986
Co-Auth	1.13	.0091	994|986
PatentTime	.954	.0013	994|986
Structural Covariates			
E-PPI	1.15	.015	994|986
R-PPI	1.02	.032	994|986
E-Eccent	.83	.011	994|986
R-Eccent	.969	.028	994|986
E-Deg	1.07	.0039	994|986
R-Deg	.829	.029	994|986
E-Deg_2	1.04	.00024	994|986
R-Deg_2	.745	.011	994|986
Propensity Score	.986	.0056	1924|1750

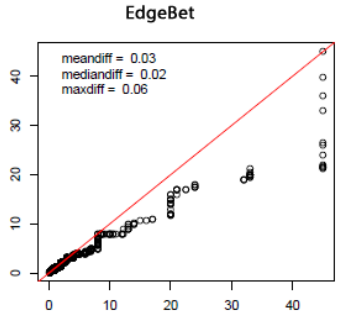
Appendix I: QQ-plots for Continuous Covariates

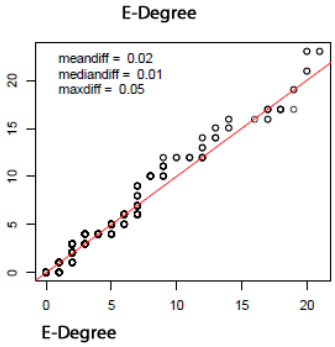


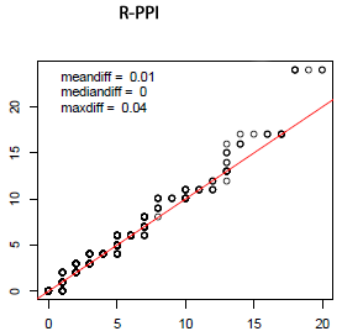


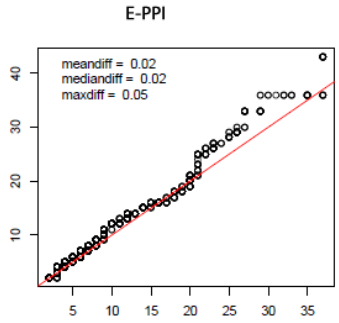


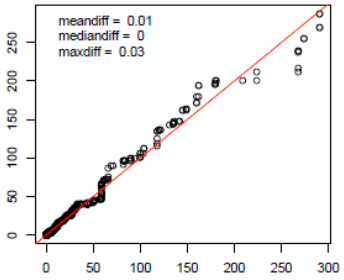












Appendix J: Patent Dataset Summary

The following graphics summarize the data for Michigan and Ohio pre- and post-MARA and show that Ohio is a strong control state for Michigan.

J1: # Patents by Corporate Assignee

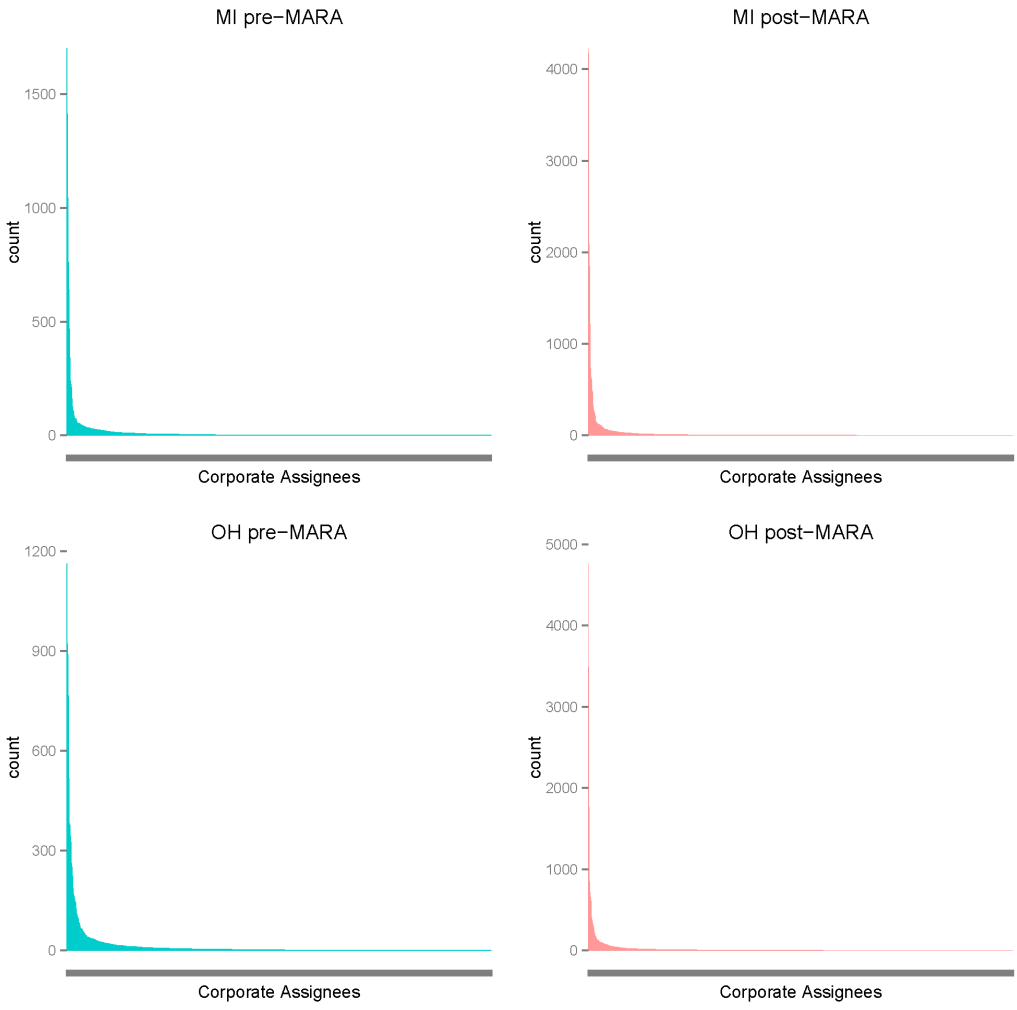


Figure 9.2: # Patents by Corporate Assignee (Company) in MI and OH pre- and post-MARA. Both states appear equally right-skewed pre-treatment. Notice that the number of patents increases greatly post-MARA, though the distribution of corporate assignees appears similar.

J2: # Patents by City

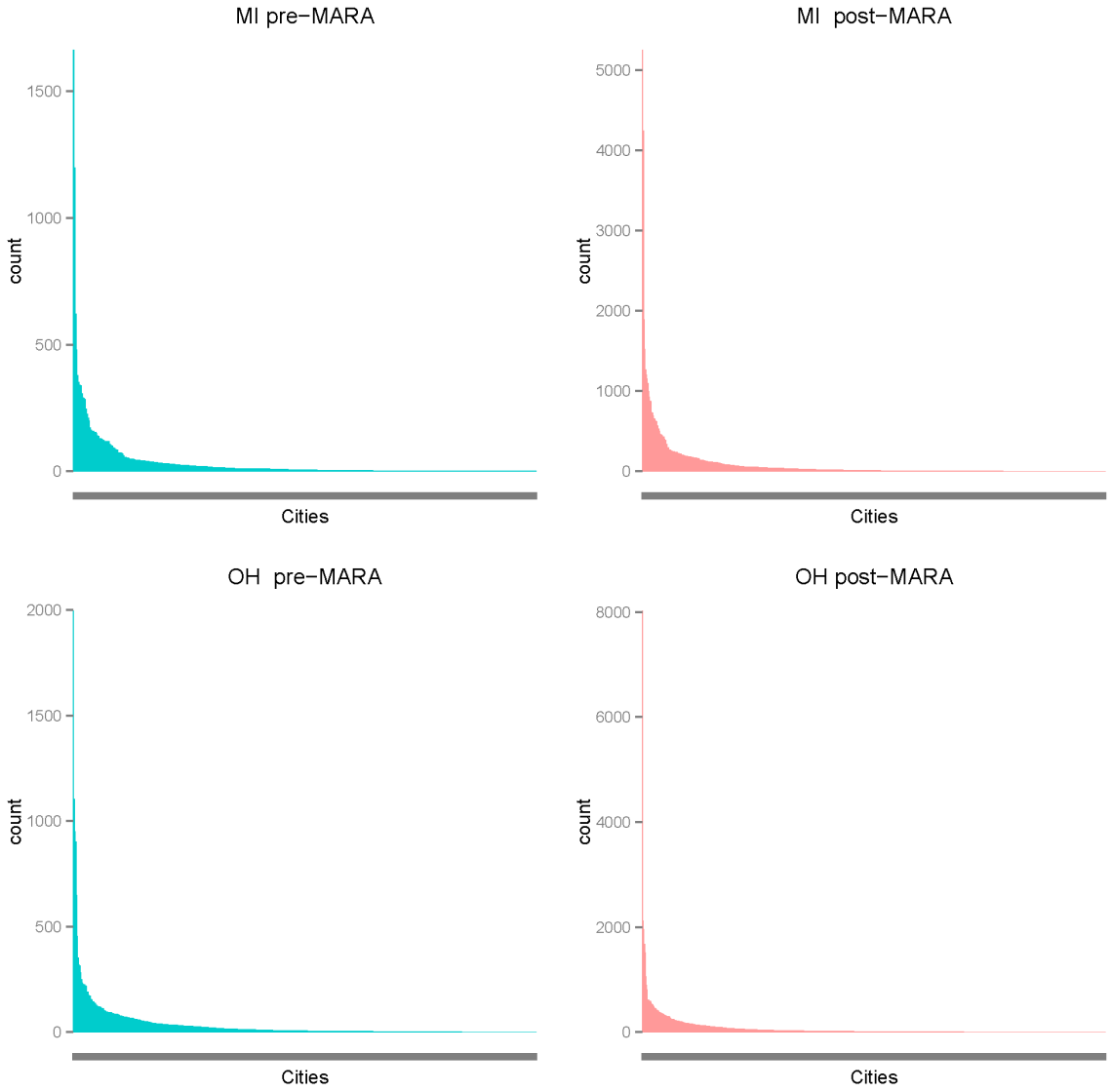


Figure 9.3: # Patents from each City in MI and OH pre- and post-MARA. Though the # of patents is far greater post-MARA than pre-MARA in both MI and OH, the distributions appear similar in both states pre-treatment.

J3: # Patents by Patent Technology Class

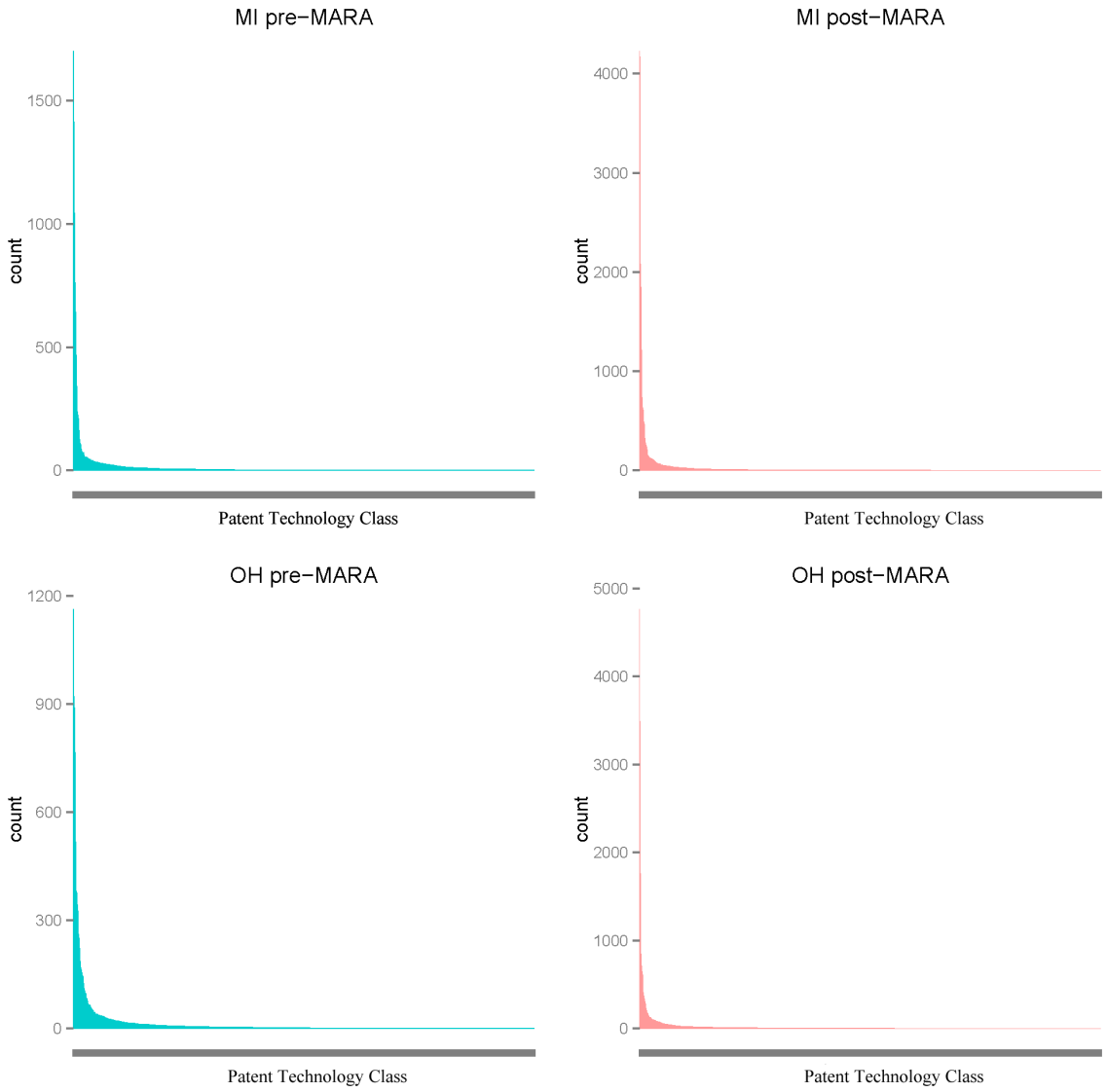


Figure 9.4: # Patents per Patent Technology Class in MI and OH pre- and post-MARA. The distribution of patents by technology class also appear similar for MI and OH both pre- and post-MARA. Both states appear to be focused on one producing one major class of product, for MI it is likely auto-related technologies, while for OH it is likely energy-related technologies.

J4: # Patents per Inventor

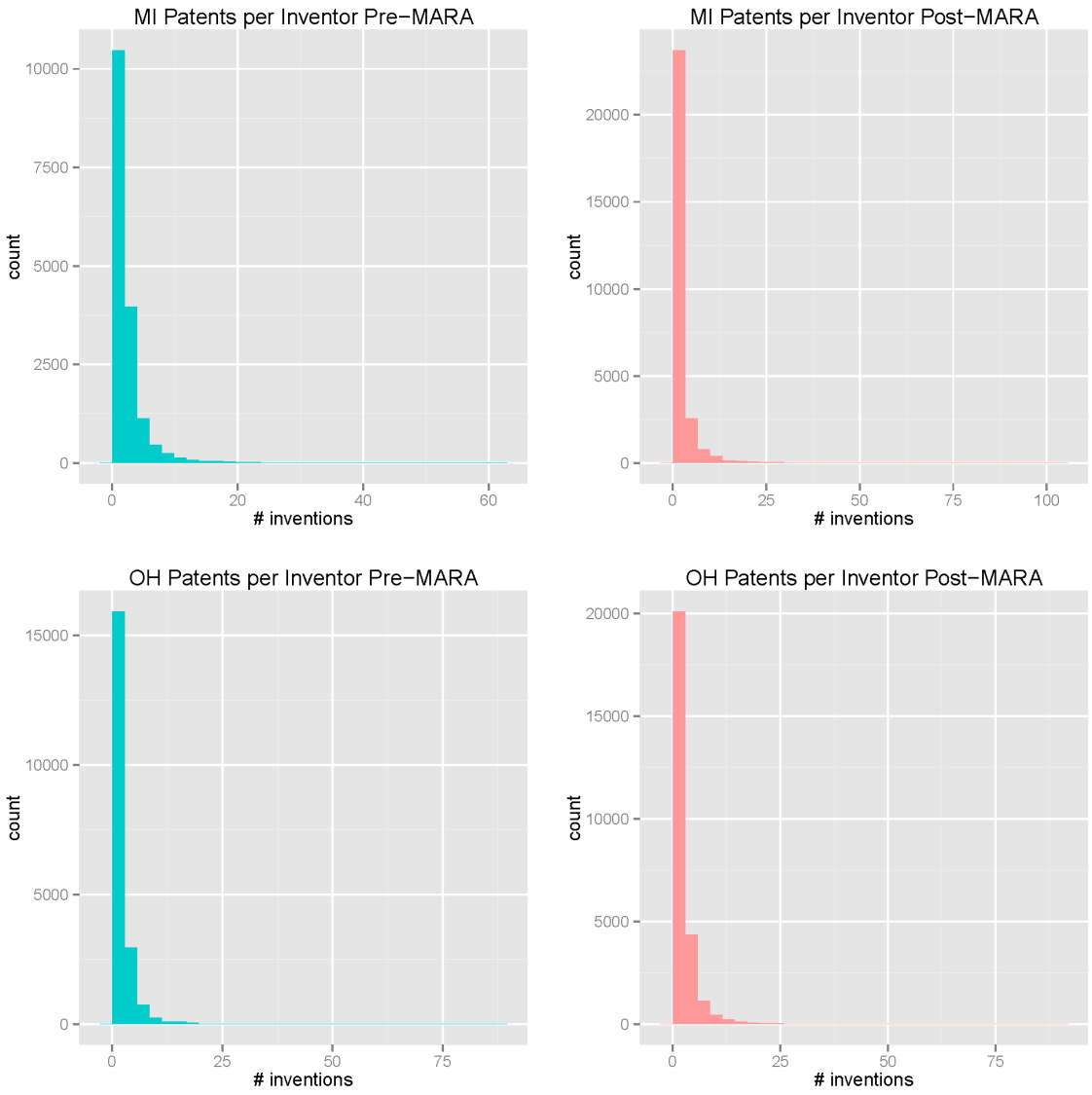


Figure 9.5: # Patents per Inventor in MI and OH pre- and post-MARA. Most inventors in both states produce very few patents, while a few produce very large numbers of patents. The distributions appear similar, though interestingly it appears that post-MARA there is a higher proportion of inventors with very few patents than pre-MARA.

Appendix K: Ego-network Summaries

The following graphics summarize the ego networks which represent interaction and patent properties of a fixed set of inventors.

K1: # Patents by Class (Ego Network)

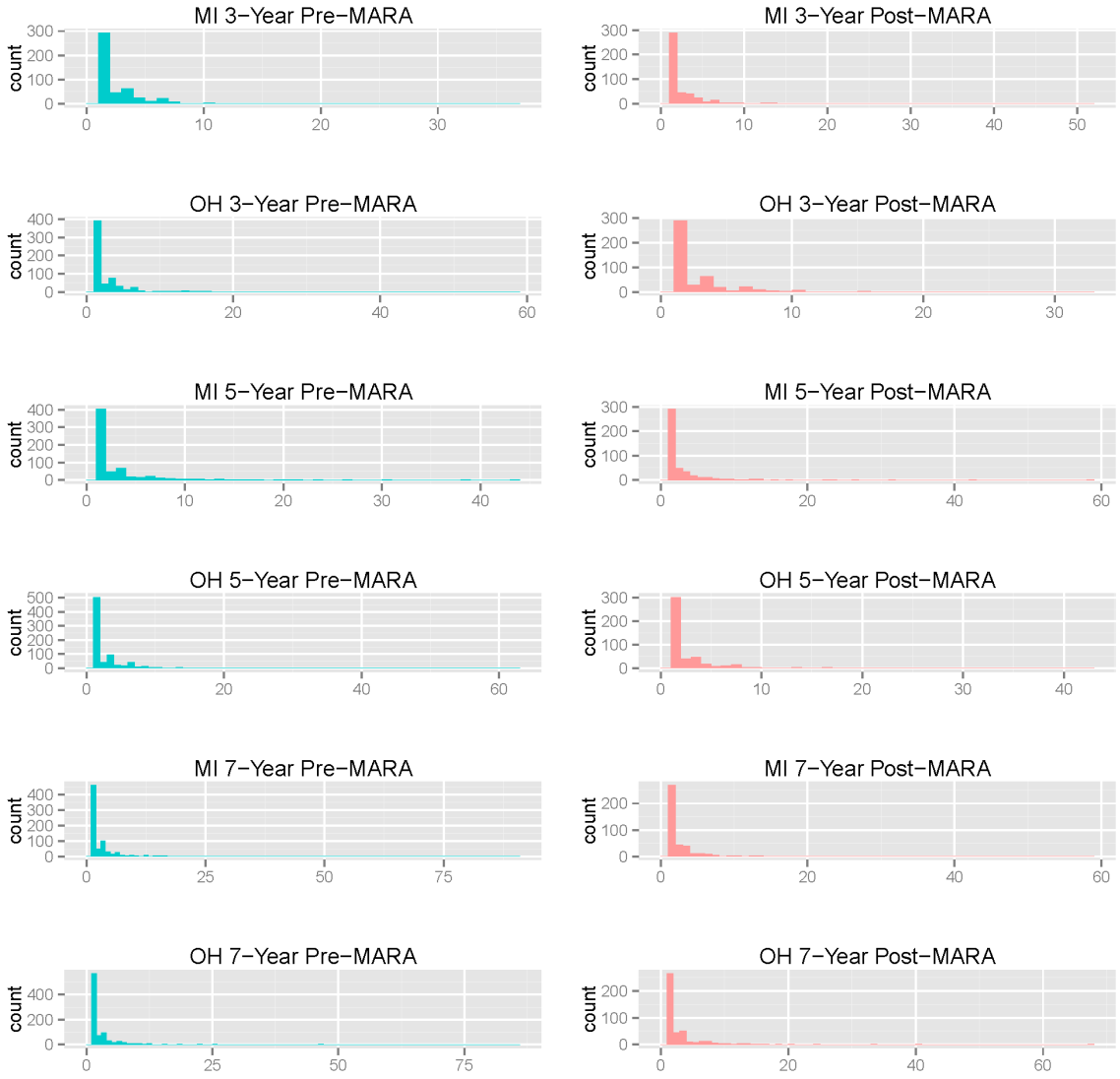


Figure 9.6: The distributions of patents by technology class pre-MARA appear similar. Very few patents are produced in most technology classes, while large amounts of patents are produced in very few technology classes. This trend appears to stays constant among the ego set pre- and post-MARA for both states.

K2: # Patents per Inventor (Ego Network)

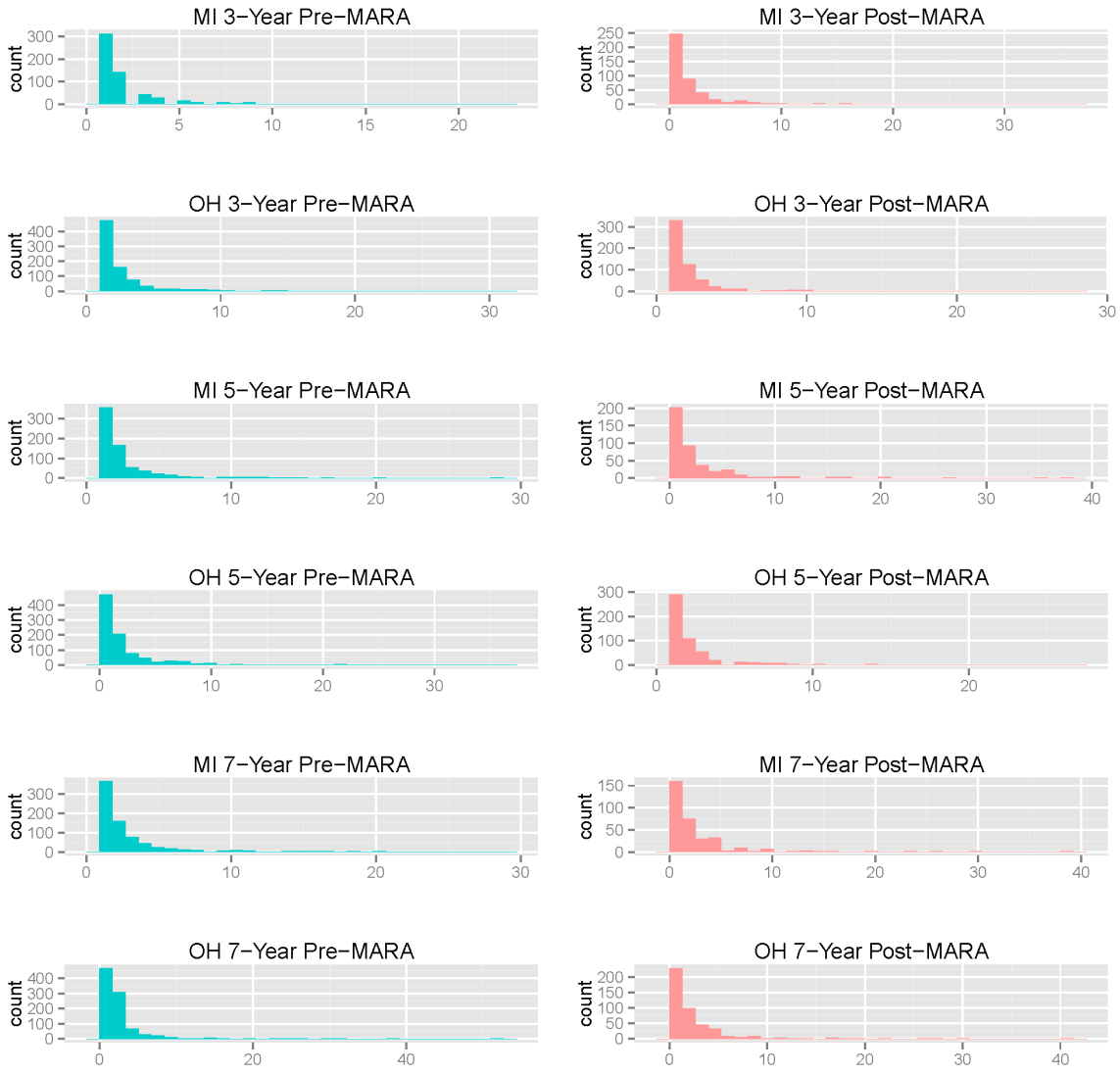


Figure 9.7: The distributions of patents per inventor appear similar. Very few patents are produced by most inventors, while large amounts of patents are produced by very few inventors.

K3: # Patents by Year Over Entire Ego Network

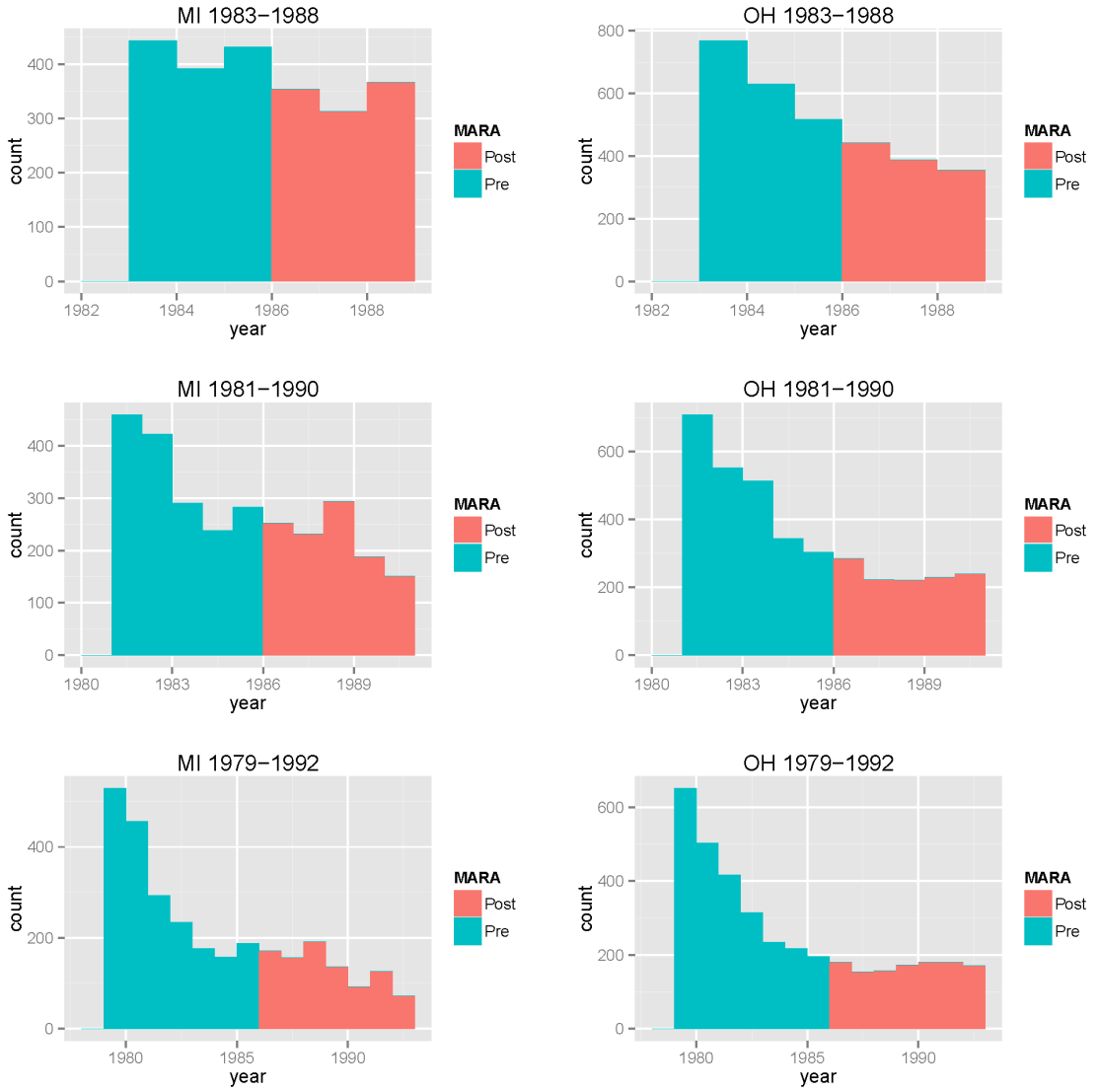


Figure 9.8: Over time the ego set produces fewer patents. This occurs because we restrict the ego set to those inventors who have at least one patent before the beginning time period. Therefore, these inventors are on average farther along in their careers than if we were to include all inventors who began patenting within the time period. As a result, their number of patents decreases over time. Without controlling for trends such as changes in # of patents among the ego set over time, the estimate of the impact of MARA on interaction propensities would be skewed since the ego set patents less over time and therefore also has a lower propensity to collaborate.

References

[1] Abadie, A. (2005). Semiparametric Differences-in-Differences Estimators. *Review of Economic Studies* **72** 1-19.

[2] Albert, R., and Barabasi, A. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics* **74** 47-97.

[3] Aral, S., Muchnik, L. and Sundararajan, A. (2009). Distinguishing influence based contagion from homophily driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences* **106** 21544-21549.

[4] Ashenfelter, O. and Card, D. (1985). Using the longitudinal structure of earnings to estimate the effect of training programs. *The Review of Economics and Statistics* **67** 648-660.

[5] Austin, P.C. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharm. Statist.* **10** 150-161.

[6] Bertrand, M., Duflo, E., Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics* **119** 249-275.

[7] Campbell, D.T. (1969). Reforms as experiments. *American Psychologist* **24** 409-429.

[8] Carmi, E., Oestreicher-Singer, G., and Sundararajan, A. (2012). Is Oprah contagious? Identifying demand spillovers in online networks. (Working Paper No. 10-18). Retrieved from Social Science Research Network website : http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1694308

[9] Centola, D. (2010). The spread of behavior in an online social experiment. *Science* **329** 1194-1197.

[10] Christakis, N.A., and Fowler, J.H. (2007). The spread of obesity in a large social network over 32 years. *N Engl J Med* **357** 370-379.

[11] Cochran, W.G. (1963). *Sampling Techniques.* John Wiley & Sons.

[12] Cox, D.R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)* **34** 187-220.

[13] D'Amour, A., and Airoldi, E. (2014). Sparsity misspecification and robust covariate effect estimation for sparse social networks.

[14] Deheija, R.H. and Wahba, S. (1999). Causal effects in non-experimental studies: Re-evaluating the evaluation of training programs. *J. Amer. Statist. Assoc.* **94** 1053-1062.

[15] Gutman, R., and Rubin, D.B. (2012). Analyses that inform policy decisions. *Biometrics* **68** 671-675.

[16] Hill, S., Provost, F., Volinsky, C. (2006). Network-based marketing: Identifying likely adopters via consumer networks. *Statistical Science* **21** 256-276.

[17] Ho, D.E., Imai, K., King, G. and Stuart, E.A. (2007). Matching as nonparametric preprocessing for reduced model dependence in parametric causal inference. *Political Analysis* **15** 199-236.

[18] Holland, P.W. (1986). Statistics and causal inference. *J. Amer. Statist. Assoc.* **81** 945-960.

[19] Hudgens, M.G., and Halloran, M.E. (2008). Toward causal inference with interference. *J Am Stat Assoc.* **103** 832-842.

[20] Iacus, S.M., King, G. and Porro, G. (2011). Causal inference without balance checking: coarsened exact matching. *Political Analysis* (forthcoming).

[21] Imbens, G.W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A Review. *Review of Economics and Statistics* **86** 4-29.

[22] Imbens, G.W., and Rubin, D.B. (2013). An introduction to causal inference in statistics, biomedical and social sciences.

[23] Imbens, G.W., and Wooldridge, J.M. (2009). Recent developments in the econometrics of program evaluation." *Journal of Economic Literature* **47** 5-86.

[24] Klepper, S. (2002). The capabilities of new firms and the evolution of the US automobile industry. *Industrial and Corporate Change* **11** 645-666.

[25] Lai, R., D'Amour, A., Yu, A., Sun, Y., Torvik, V., and Fleming, L. (2011). Disambiguation and Co-authorship Networks of the U.S. Patent Inventor Database. Harvard Institute for Quantitative Social Science, Cambridge, MA 02138.

[26] Marx, M., Strumsky, D., and Fleming, L. (2009). Mobility, skills and the Michigan non-compete experiment. *Management Science* 55 875-889.

[27] Meyer, B.D. (1995). Natural and Quasi-Experiments in Economics. *Journal of Business & Economic Statistics* **13** 151-161.

[28] Milligram, S. (1967). The Small-World Problem. *Psychology Today* **1** 61-67.

[29] Motta, M., and Roende, R. (2002). Trade secret laws, labour mobility and innovations CEPR Discussion Paper 3615, Centre for Economic Policy Research, London.

[30] Normand S.L. T., Landrum M.B., Guadagnoli E., Ayanian J.Z., Ryan T.J., Cleary P.D., and McNeil B.J. (2001). Validating recommendations for coronary angiography following an acute myocardial infarction in the elderly: A matched analysis using propensity scores. *Journal of Clinical Epidemiology* **54** 387–398.

[31] Oh, J., Susarla, A., Tan, Y., 2011. Informational Cascades and Contagion in Online Social Networks," forthcoming in *Information Systems Research.*

[32] Pearl, J. (2000). Causality: models, reasoning, and inference. Cambridge University Press New York.

[33] Perry, P.O. and Wolfe, P.J. (2013). Point process modelling for directed interaction networks. *J.R. Statist. Soc. B.* **75** 821-849.

[34] Rosenbaum, P. R. (2002). Observational Studies, 2nd ed. Springer, New York.

[35] Rosenbaum, P.R. and Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **7**0 41-55.

[36] Rosenbaum, P.R. and Rubin, D.B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Amer. Statist.* **39** 33-38.

[37] Rubin, D.B. (1973). Matching to remove bias in observational studies. *Biometrika* **29** 159-184.

[38] Rubin, D.B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66** 688-701.

[39] Rubin, D.B. (1976). Inference and missing data (with discussion). *Biometrics* **63** 581-592.

[40] Rubin, D.B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics* **6** 34-58.

[41] Rubin, D.B. (1980). Bias Reduction Using Mahalanobis-Metric Matching. *Biometrics* **36** 293-298.

[42] Rubin, D.B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services & Outcomes Research Methodology* **2** 169-188.

[43] Rubin, D.B. (2007). The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials. *Stat. Med.* **26** 20-36.

[44] Rubin, D.B. and Thomas, N. (1996). Matching using estimated propensity scores, relating theory to practise. *Biometrics* **52** 249-264.

[45] Sekhon, J. (2008). Multivariate and propensity score matching software with aitomated balance optimization: the matching package for R. *Journal of Statistical Software* **42** 1-52.

[46] Sekhon, J., and Diamond, A. (2013). Genetic matching for estimating causal effects: a general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics* **95** 932-945.

[47] Shalizi, C.R., and Rinaldo, A. (2013). Consistency under sampling of exponential random graph models. *The Annals of Statistics* **42** 508-535.

[48] Snijders, T.A.B. (2001). The statistical evaluation of social network dynamics. *Sociological Methodology* **31** 361-395.

[49] Stuart, Elizabeth A. (2010). Matching Methods for Causal Inference: A Review and a Look Forward. *Statist. Sci.* **25** 1-21.

[50] Stuart, T., and Sorenson, O. (2003). The geography of opportunity: Spatial heterogeneity in founding rates and the performance of biotechnology firms. *Research Policy* **32** 229-253.

[51] Tchetgen, T., Eric, J., & VanderWeele, T.J. (2012). On causal inference in the presence of interference. *Statistical Methods in Medical Research–Special Issue on Causal Inference* **21** 55-75.

[52] VanderWeele, J. T., An, W. (2013). Social Networks and Causal Inference. In Morgan, S.L. (ed.). Handbook of Causal Analysis for Social Research (pp. 353-374). Springer Netherlands.