Measuring Social Influence Without Bias

Annie Franco * Bobbie NJ Macdonald †

December 9, 2015

CS224W: Final Paper

The Problem

How well can statistical models disentangle the effects of social influence from homophily in observational network data, and under what conditions do they models fail?[1] Many outcomes of interest to social scientists – such as obesity, political attitudes, altruism, and drug use – are highly correlated within social networks.[2] However, behaviors and attitudes tend to be correlated in social networks for reasons other than just direct peer-to-peer influence, resulting in serious methodological challenges for making causal inferences in purely observational data (Angrist, 2014; Manski, 1993).

First, similar individuals tend to self-select into relationships with one another. Because of this sorting, it is notoriously difficult to separate the effects of *social influence* from the confounding effects of homophily (see Shalizi and Thomas, 2011). Second, outcomes might be correlated due to similar contextual or environmental effects, such as a shock that affects all members of a shared group at once. Finally, peer effects are reciprocal, as each individual contributes to the expected outcome of their alters. As a result, there is enormous debate over the extent to which results from observational studies reflect true peer-to-peer influence.

In this paper, we assess the relative performance of propensity score and kernel methods in recovering unbiased estimates of social influence in observational data. We use Monte Carlo simulation to apply the following methods to synthetic longitudinal data and measure their bias relative to regression adjustment: (1) propensity score matching, (2) inverse propensity score weighting, and (3) kernel-balancing. Our simulations show that propensity score matching is indeed problematic to use in network settings even under ideal conditions (e.g. no mis-specification or missing data). As a result, weighting methods (both parametric and non-parametric) are better suited for the type of longitudinal networks used in this study.

*Department of Political Science, Stanford University. abfranco@stanford.edu.
†Department of Political Science, Stanford University. bmacdon@stanford.edu.
[1]Throughout this paper, we use the terms *social influence* and *peer effects* interchangeably.
[2]For a review, see Christakis and Fowler (2013).

Prior Work

A commonly used strategy for circumventing the problems discussed above is the use of procedures to achieve covariate balance between units that are exposed to a social signal or behavior of interest through their peers (*treated*) and those that are not (*control*). Here, we describe two methodological approaches–propensity-score methods and kernel-balancing–in further detail, which we empirically test in a simulated network setting.

Propensity-score methods

Matched sample estimation has become a popular method for causal inference, in large part because of its perceived reduction of model dependence. To estimate a causal effect, treated units (i.e. those exposed to a social signal) are matched to control units that are most similar to them in terms of observable characteristics $\mathbf{X}$, and only these matched pairs are retained in the sample. The goal of this procedure is to achieve balance over the common support of $\mathbf{X}$, so a simple difference in means between treated and control units on the matched dataset yields an estimate of the average treatment effect of interest. There is a great amount of flexibility in choosing the conditioning variables and matching function, but arguably the most popular approach is *propensity score matching*, whereby units are matched on their propensity to have been treated at time t. Typically, the propensity scores are predictions from a regression of a treatment indicator on a vector of covariates, but predictions from more flexible models can also be employed.

Aral et al. (2009) notably employ this technique to disentangle homophily and influence in a large-scale dynamic network dataset of mobile app adoptions. For each of the observed time periods, the authors matched individuals to others with the same number of adopter friends as a function of their dynamic behaviors and attributes prior to each time period. Aral et al. (2009) compare the matched sample estimates to those computed using random matching, finding homophily appears to explain more than half of the observed diffusion of adoption behavior.

However, propensity score matching has several limitations (King and Nielsen, 2015). Treated units can systematically differ from control units on X, which not only violates the assumption of *selection on observables* (i.e. $T\ Y_0|\mathbf{X}$) but can also lead to a lack of common support for some values of $\mathbf{X}$ ($0 < P(T = 1|\mathbf{X}) = 1$). Researchers will typically discard observations that fall outside of the common support of X, but this can yield biased estimates of causal effects if the trimmed and retained units exhibit treatment effect heterogeneity. Treated units can also differ substantially from units with equivalent propensity scores, since different combinations of observed attributes X_i can yield similar score values.

An alternative to matching is inverse propensity score weighting (IPSW), which retains all comparison units. However, propensity score methods do not guarantee multivariate balance. Even if the conditioning is complete, these methods often only achieve balance on marginal univariate distributions; functions of the distribution are typically not identical across matched groups (Hazlett, 2015). The procedure might even increase balance on some covariates at the expense of greater imbalance on others. Although it is possible to include

all possible multiplicative interaction terms p, in practice this may not be feasible if observed network sizes are small. Regularization methods might help with $p <<< N$, but would not attenuate the issue of sparse or inequivalent matches.

We propose to test the performance of PSM against kernel-based methods for multivariate balancing (e.g. Scott, 2015; Ma and Hu, 2013).

Kernel-Balancing

The method we evaluate here, proposed by (Hazlett, 2015), exploits the "kernel trick" to find a set of non-negative weights for the control units such that the weighted average vector X_C for the control group equals the unweighted average vector X_T. Specifically, a kernel function is used to map a vector of features across observations in each group to produce a similarity measure that takes a value between 0 (least similar) and 1 (most similar) for every observation (Hazlett, n.d.).

Let $\mathbf{K} = k(X_T, X_C)$ be an $(N_T + N_C) \times N$ positive semi-definite matrix containing all pairwise applications of the kernel function (e.g. Gaussian kernel). We can recover a set of weights w_i for the control units by partitioning $\mathbf{K}$ into treated ($\mathbf{K_T}$) and control ($\mathbf{K_C}$), and finding a vector w_i such that the average row of $\mathbf{K_T}$, $\bar{k_T} = \sum_{i \in C} w_i k_i$. This procedure achieves mean balance in the density at each location in the covariate space. Moreover, all control units are retained, so the entire support of $\mathbf{X}$ in both treated and control groups is preserved[3]. Because the expected potential outcomes Y_0 are now equal across control and treatment groups, we can recover an unbiased estimate of the average treatment effect with a weighted DIM estimator. In sum, the kernel-balancing approach solves the problems associated with propensity score matching by (a) allowing every observation to contribute to the treatment effect estimate and (b) producing mean balance across almost any function of the covariates. This is particularly important if we have reason to suspect the influence function is non-linear, as PSM is unable to achieve balance in high-dimensional space.

Research Design

Our goal in this study is to assess the relative performance of propensity-score and kernel-balancing methods in recovering estimates of peer effects from observational network data. We do so by applying these methods across a range of conditions (i.e. varying network size and treatment definition) that produce longitudinal networks with known patterns of homophily and social influence.

Specifically, for each set of conditions, we:

(a) randomly sample a dataset of n observations with correlated covariates X;

(b) simulate 500 sparse, clustered, homophilous networks with these observations;

[3]Note, however, that control units far outside of the support of X in the treated group will be assigned very low weight. Of course, if there is total separation between treated and control units on this range, it may not be possible to implement any of the estimation procedures considered here.

(c) construct a panel dataset of outcomes Y_{it} for each simulated network according to a known data-generating process;

(d) estimate the average treatment effect on the treated (ATT) using 8 different procedures (see "Estimation Procedures"); and

(e) evaluate the estimators by computing their bias and root mean-squared error (RMSE).

We describe this procedure for the Monte Carlo analyses in more detail below.

Setup

We assess the performance of these methods on synthetic longitudinal networks, where we allow outcomes to change over time while the network structure and individual characteristics remain unchanged. For each simulated network, we sample from an exponential random graph model (ERGM) to produce an undirected graph g with n nodes $i = \{1, \ldots, n\}$. Specifically, we specify the probability of observing graph g with n nodes as:

$$p(g) = \frac{exp(\theta_1 e(g) + \theta_2 h(\mathbf{X}))}{\sum_{g' \in G} exp(\theta_1 e(g') + \theta_2 h(\mathbf{X}'))} \tag{1}$$

where G is the set of all possible networks with n nodes, $e(g)$ is the number of edges in the network, and $h(\mathbf{X})$ is the number of edges whose source and target nodes share the same $\mathbf{X}$ characteristics (e.g. age, race, education). In each network simulation, we use the `statnet` package in `R` to draw coefficients θ_1 and θ_2 to produce network structures at time 0 that capture two characteristic features of human social networks: sparsity and homophily. To achieve sparsity, we draw θ_1 to reflect a network density of approximately $0.2^{log_{10}(n)}$, such that the target number of edges for a given network of size n is $(0.2^{log_{10}(n)}) * ((n * (n-1))/2)$. To achieve homophily, we draw θ_2 such that approximately 50% of all edges fall on the diagonal of the mixing matrix for covariates $\mathbf{X}$. We then simulate 500 networks using equation (1) and the coefficients θ_1 and θ_2. Figure 1 displays one of these sampled networks.

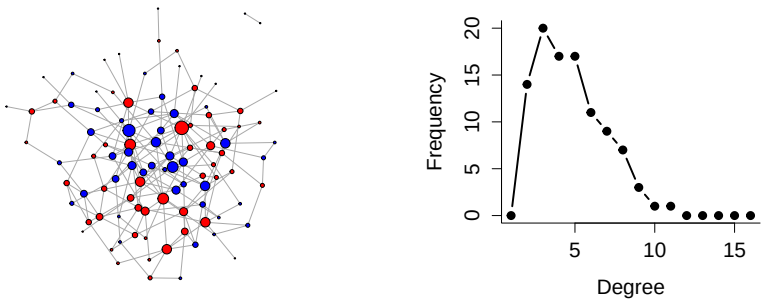
We then produce an $n \times K$ covariate matrix ($\mathbf{X}$) with $K = 4$. To simulate these covariates, we first take n draws from a multivariate normal distribution with $var(x_k) = 1$ and $cov(x_k, x_l) \sim N(0, 0.3)$ and then dichotomize each observation by assigning a value of 1 to all observations great than 0 and assigning 0 to all observations less than 0. This procedure results in a set of correlated dichotomous variables for each of the 500 networks.

Simulating Longitudinal Outcome Data. For each simulated network, we construct a panel dataset with $t = \{1, \ldots, 3\}$ periods. At $t = 0$, 10% of the nodes are randomly selected to adopt some social innovation $y \in \{0, 1\}$. In each subsequent period, nodes update their behavior as a probabilistic function of their own characteristics ($\mathbf{X}$) and the behavior of their neighbors in $t - 1$. Specifically, we determine the probability of outcome y_{it} for node i at

Figure 1: Simulated homophilous network with $N = 1000$ nodes.

(a) Nodes are colored by value of homophilous covariate x_1.

(b) Degree distribution.



time t using the logistic transformation:

$$p(y_{it}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 D_{i,t} + \boldsymbol{\beta_2} \mathbf{X}_{i,t})}} \tag{2}$$

Here, $D_{i,t}$ is an indicator for whether a node is *treated* at time t, defined as having at least one neighbor with $y = 1$ in $t - 1$ (we vary this treatment definition in the simulations, as described in `Experimentation`). The coefficients of this model have a natural interpretation: β_1 captures social influence while the vector $\boldsymbol{\beta_2}$ captures common shocks to nodes with similar attributes.

Estimation

In each simulation, we compute the *average treatment effect on the treated* (ATT) in each time period using the following methods with and without covariate adjustment: (1) (naive) logistic regression; (2) PSM within each time period; (3) inverse propensity score weighting; and (4) kernel-balancing within each time period. Hence, we compare 8 estimators in total. We assess the accuracy of each estimator by examining the bias and root mean-square error (RMSE) of the estimated ATT across the time periods:

$$Bias = (T * nSims)^{-1} \sum_{t}^{T} \sum_{j}^{nSims} \theta_{jD} - \hat{\theta_{jD}} \tag{3}$$

$$RMSE = \sqrt{(T * nSims)^{-1} \sum_{t}^{T} \sum_{j}^{nSims} (\theta_{jD} - \hat{\theta_{jD}})^2} \tag{4}$$

Experimentation

To compare the performance of these methods under various conditions, we repeat the 500 network simulations six times across different network sizes and treatment definitions. These two parameters provide us with six parameter vectors over which to conduct the Monte Carlo analyses (3 network sizes $\times$ 2 treatment definitions), resulting in a total of 3000 (500×6) simulated networks.

Specifically, we vary:

1. **Network size.** We expect that smaller network sizes will be under-powered for detecting effect sizes relative to large network sizes. This suggests that estimates of peer effects in smaller networks, if significant, will exhibit larger bias than if the estimates were computed using data from a larger network. We conduct simulations for $n \in \{100, 500, 1000\}$, where n is the number of nodes.
2. **Definition of treatment.** We examine performance under two alternative treatment definitions:
   (a) **Change definition:** Treatment status for individual i is equal to 1 if any of her neighbors adopted the social innovation between $t-2$ and $t-1$, 0 otherwise;[4]
   (b) **Threshold definition:** Treatment status for individual i is equal to 1 if the proportion of neighbors with $y = 1$ in $t-1$ was greater than or equal to 25%.

Results

Our findings are summarized in Figures 2 and 3. We omit the results for PSM + regression, as this method performed extremely poorly. This was especially true in small networks where the social innovation diffused quickly and widely.

Consistent with our expectations, we see that RMSE is decreasing with network size. Contrary to our expectations, however, bias is lowest for the medium-sized network. We are currently investigating the reason for this non-monotonic trend.

Note that we use a static data-generating model, i.e. allow outcomes to change over time while the network structure and individual characteristics remain unchanged, *and* we correctly specify both the propensity score and outcome models. These are the best possible conditions for unbiased naive estimation, and we see that indeed, logistic regression on the treatment indicator performs reasonably well compared to other methods. On average, however, naive regression and PSM do worse relative to other estimation strategies, with weighting methods (IPSW and KB) performing best in larger sample sizes.

[4]For t=1, all individuals are randomly initialized with y for t=0 and t=-1, where $y \sim Bernoulli(0.1)$ for all individuals in these two periods.

Figure 2: Estimator Bias

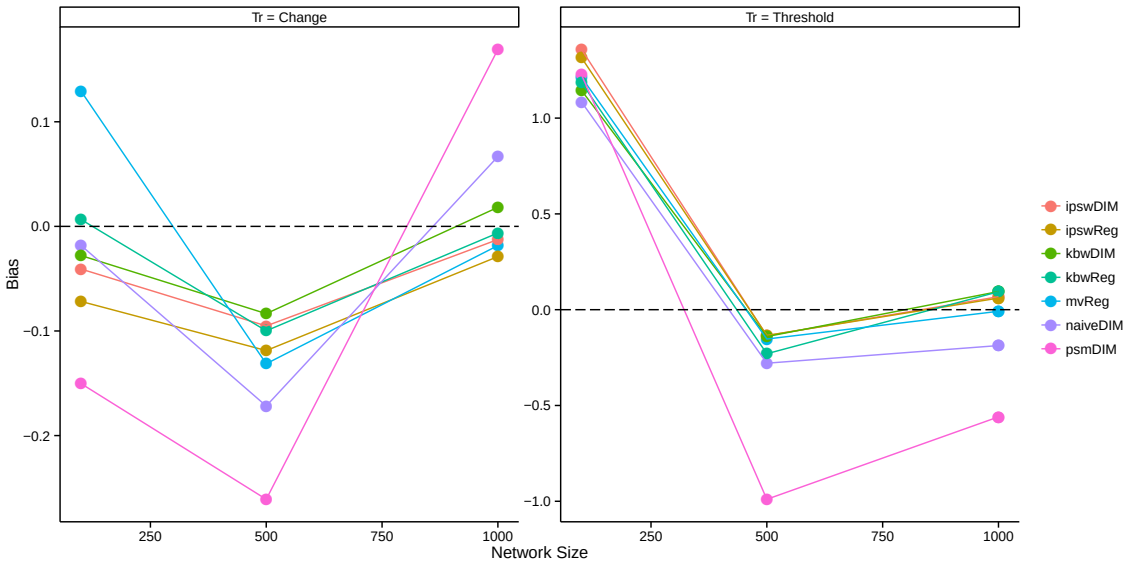


Figure 2. These plots show the average bias of each of the methods across the parameter values.

Figure 3: Estimator RMSE

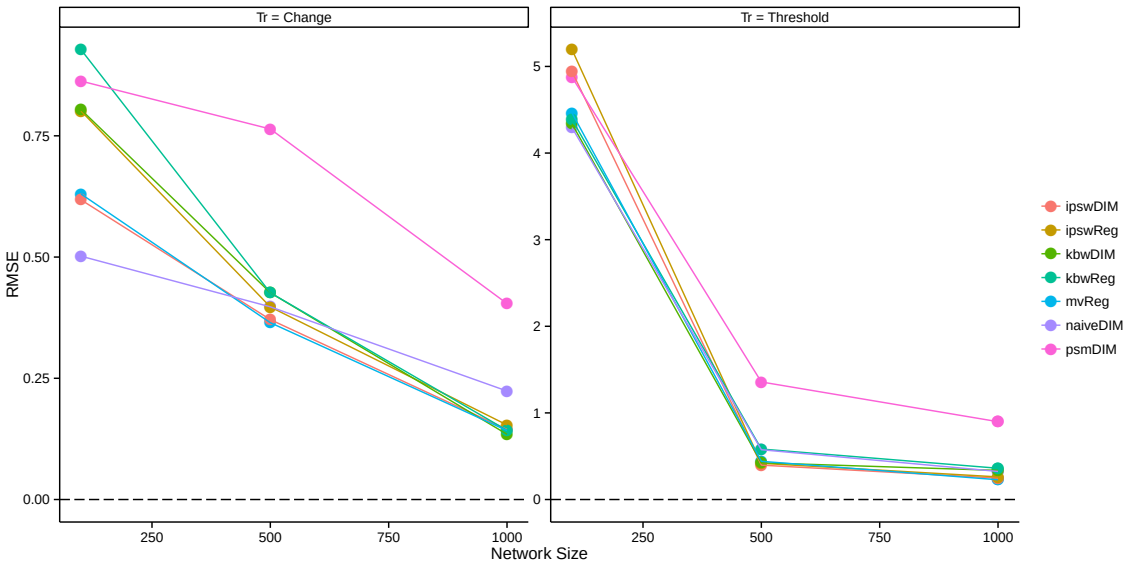


Figure 3. These plots show the RMSE of each of the methods across the parameter values.

Conclusion and Future Directions

In this analysis, we find that PSM is indeed problematic to use in network settings even under ideal conditions (e.g. no misspecification or missing data). Weighting methods, both parametric and non-parametric, are better suited for this type of data.

We plan to extend the analyses presented here to larger network sizes, longer time periods, and greater confounding. Some additional directions for future research are as follows:

1. **Dynamic Models**: Future iterations of this project will allow outcomes to be autocorrelated over time and covariates to be endogenous to treatment. We might also consider the case of endogenous network formation and evolution, whereby nodes are allowed to form new edges in each time period.
2. **Missing edges**: One difficulty in collecting observational data in offline contexts is that a substantial proportion of people's network remains unobserved. We plan to examine the impact of edge censoring on estimator performance.
3. **Alternative treatment definitions**: We will examine alternative diffusion processes (e.g. influence from intransitive peers, heterogeneous treatment effects), and the impact of mis-specification of the treatment.

References

Angrist, J. D. (2014, jun). The perils of peer effects. *Labour Economics in press*.

Aral, S., L. Muchnik, and A. Sundararajan (2009, dec). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences of the United States of America 106*(51), 21544–9.

Christakis, N. a. and J. H. Fowler (2013, feb). Social contagion theory: examining dynamic social networks and human behavior. *Statistics in medicine 32*(4), 556–77.

Hazlett, C. J. (2015). *Kernel Balancing: A flexible non-parametric weighting procedure for estimating causal effects.* Ph. D. thesis, Massachusetts Institute of Technology.

King, G. and R. Nielsen (2015). Why propensity scores should not be used for matching. *Copy at http://j. mp/1FQhySn Export BibTex Tagged XML Download Paper 452*.

Ma, Z. and F. Hu (2013). Balancing continuous covariates based on kernel densities. *Contemporary clinical trials 34*(2), 262–269.

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

Scott, D. W. (2015). *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons.

Shalizi, C. R. and A. C. Thomas (2011, may). Homophily and Contagion Are Generically Confounded in Observational Social Network Studies. *Sociological methods & research 40*(2), 211–239.