

# A model for treatment effect estimation for randomized experiments on networks

Wenfei Du

December 10, 2017

## Abstract

In social and biological networks, the results of experiments are difficult to interpret since outcomes for each observation may not be independent. In this paper, we provide a model and reduced form that can be used to estimate average direct and peer treatment effects on networks. We assume a randomized design over individual treatment assignment and neighboring treatment assignment.

## 1 Introduction

The literature on treatment effects in the randomized experiment setting has focused on cases where the potential outcomes for each observation are independent of the treatment assignment of other observations. In this case, the average treatment effect is defined as the population average in the difference of potential outcomes. This can be estimated by subtracting the mean observed outcomes in treated and control groups. Inference can be done using a two sample test.

Recently, there has been interest in analyzing experiments on networks, where the assumption of independence between observations is unlikely to be met. For instance, on social network platforms, the outcome for one individual may depend on the treatment given to their friends due to influencing behavior. More generally, any experiment with linkage between units has the same problem. Since the potential outcomes are now dependent on possibly the treatment assignments of all units, the average treatment effect is not well defined. There are multiple possible quantities of interest. The total effect, for instance, is the average in the difference of potential outcomes for the cases where everyone is treated and where no one is treated. In [7], the authors proposed experimental design and analysis using graph cluster randomization, where all of the individuals in each cluster were given the same treatment assignment. This estimation is valuable for web experiments, where one might

be interested in the overall effect if a new feature is deployed to all users.

In this paper, we focus on estimating individual effects and decomposing them into direct effects and indirect effects through peers. The interest is not only in what happens if one individual is treated but also what happens if some proportion of his friends, friends of friends, and so on are treated as well. Since we only observe one network, additional assumptions need to be imposed for estimation. Under a simple model, we propose a reduced form regression based estimation method to identify these effects.

## 2 Literature review

One part of the network experimentation literature focuses on identifying peer effects. In this setting, a problem with observational studies is that it is difficult to distinguish between homophily and peer influence. Individuals who form connections tend to have similar preferences, so when one person adopts a new product after his friend, the underlying cause may not be peer influence. Many papers, including [2], have attempted to separate these causes by controlling for demographic and behavioral characteristics. In [6], the authors compared experimental with observational results and showed that adding controls can still lead to large overestimates of peer effects. Only high dimensional adjustment involving many prior behaviors was close to eliminating bias.

Thus, experiments are needed and are often done by exposing only randomly selected individuals to peer behaviors. In [5], the authors estimated the effect of peer influence by removing a piece of shared information from users' Facebook feeds. They found a larger probability of sharing for those receiving the link on their feed. Users with more friends sharing or with stronger ties to a friend sharing, whether they were exposed to it on their feed or not, were more likely to share it themselves. However, the relative gain from being exposed was larger when few friends

shared and when the sharing friend had weak ties.

In the context of product adoption, [3] and [4] randomly enabled and disabled notifications about usage and referral messages for an application. They found that the level of influence differed depending on the background of the individual and his peers, such as age, sex, and marital status. Also, they showed that the effect on use for the original individual was positively associated with treatment. One possibility was that the added features in the treated group promoted higher usage. However, this effect was insignificant when the number of peer adopters was included in the model, suggesting that the effect was likely due to externalities. This feedback loop was also discussed in [8]. They used an encouragement approach to induce treated peers to interact with the Facebook comment box. Using treatment assignment as an instrument, they estimated an increase in comments and likes by individuals in response to an increase in their peers’ comments and likes.

These papers show that in networks like Facebook, peer effects can be significant when experimenting with a new feature. Even if a user is in the control group, the change in behavior by his friends in the treated group can affect his own behavior. If a user is treated, there may still be magnifying effects from his friends’ actions, as shown in [4] and [8]. In other words, average treatment effect estimation under the independence assumption is likely invalid. This stands in contrast to web experiments on functions like search or ads, where interaction with the feature is not social and independence between user behaviors may be a reasonable assumption. The papers, including [5] and [3], also find factors that may influence how large the peer effect may be, including background characteristics, number of peers, and strength of ties to peers.

Given the results from this literature, we propose a potential outcome model that accounts for peer factors and allows for estimation of effects of peer treatments and individual treatment. This follows [10], where the authors estimated the direct treatment effect under a similar model. We discuss the assumptions for these models in detail in the next section.

### 3 Model and assumptions

We build on the potential outcomes framework for causal inference. Suppose that there are  $N$  individuals and each individual  $i$  is randomly assigned a treatment  $T_i \in \{0, 1\}$  and has two potential outcomes  $Y_i(0)$  and  $Y_i(1)$  corresponding to control and treatment. In this setting, one goal is to estimate the

population average treatment effect

$$ATE = \mathbb{E}[Y_i(1) - Y_i(0)]$$

Since we only observe either  $Y_i(0)$  or  $Y_i(1)$  for each individual, we cannot compute  $ATE$ , but we can estimate it by taking the sample means in the two groups. This procedure is valid under the stable unit treatment value assumption, which states that no other individual’s treatment assignment has an effect on a given individual’s potential outcomes.

In a network setting, this can be violated as shown in [4] and [8], who hypothesized feedback loops between users. Since the treatment assignments of the entire network can possibly affect an individual’s outcome, we associate each person with potential outcomes  $Y_i(T)$  for every permutations of treatments  $T$  across all individuals. The  $ATE$  is not well defined anymore. At the individual level, we would need to estimate  $2^N - 1$  unobserved outcomes instead of one to find the effect of each assignment. To simplify the situation, we follow the literature and focus on a few average effects. Letting  $e_i$  be the  $i$ th unit vector, we define

$$ATE_{direct} = \mathbb{E}[Y_i(e_i) - Y_i(0)]$$

which is the average direct effect when the treatment is only applied to one individual. More generally, we may be interested in the effect when treatment is applied to a subset of the user’s friends, friends of friends, and so on. For first degree connections, letting  $v_i$  take values in  $\{0, 1\}$  for  $j$  neighboring  $i$  and zero otherwise, we define

$$ATE_{v_i} = \mathbb{E}[Y_i(v_i) - Y_i(0)]$$

$$ATE_{v_i} = \mathbb{E}[Y_i(v_i + e_i) - Y_i(0)]$$

In the extreme case, where we consider all users in the network, we have

$$ATE_{total} = \mathbb{E}[Y_i(1) - Y_i(0)]$$

which defines the total effect.

The estimation of these effects still requires further assumptions on the potential outcome model, because we only observe one treatment assignment across the network. One assumption by [10] and others was that spillover effects are limited to treatment assignments of first degree connections and that these effects are symmetric. This allowed estimation of the direct effect while controlling for peer effects in the experimental design.

For our purpose, we estimate peer effects as defined by the effect on outcome of number of neighbors treated, assuming symmetry of connections so

that this effect is well defined. Then, we can write each potential outcome as  $Y_i(T_i, N_{iT})$ , where  $T_i$  is the individual treatment and  $N_{iT}$  is the number of treated neighbors. Other possible effects include proportion of connections treated or difference in number of treated and control connections. We choose number treated based on previous studies on peer effects that varied this factor such as [5].

To estimate average effects, we assume an experimental design where we stratify by degree as in [10] and assign individual treatments and number of neighbors treated randomly in each group. Over all users, this design is unrealistic since the numbers for nearby users will correlate unless the network is sparse. However, we continue under this assumption since for large graphs, we can consider a sample of distant users such that we can randomly assign individual and neighbor treatments for each user. Under this design, we have randomized both  $T_i$  and  $N_{iT}$ . Therefore, similar to estimation of the typical average treatment effect, we can take the difference in sample means for individuals with  $T_i = 1, N_{iT} = 0$  and  $T_i = N_{iT} = 0$  to estimate  $ATE_{direct}$  for instance. The other effects of interest above can be estimated similarly by noting that  $Y_i(v_i) = Y_i(0, N_{iT})$  and  $Y_i(v_i + e_i) = Y_i(1, N_{iT})$ , where  $N_{iT}$  is equal to the number of ones in  $v_i$ .

We note that there is a selection issue due to our stratification, since larger degree users likely have more neighbors treated. Instead, we can have estimates conditional on strata. Alternatively, we can weight each conditional mean by the inverse probability of assignment, so that the larger degree users are weighted more for small  $N_{iT}$ . We cannot do this for large  $N_{iT}$ , since those with low degree have zero probability of having a large number of treated neighbors. Thus, we can generally only estimate a  $ATT$ , or average treatment effect on the treated. However, it is likely that the marginal effect decreases with a larger number of treated friends, so that we can consider all those with  $N_{iT} > n$  for some small  $n$  in the same treatment as those with  $N_{iT} = n$ . This allows approximate estimation of  $ATE$  via weighting for those in the population with degree at least  $n$ . We assume in the regressions below that an adjustment based on the probability weights is made.

To start, we can write the above model and estimate the same average effects using

$$Y_i = \alpha_0 + \alpha_1 T_i + \sum_n \beta_n \mathbb{I}(N_{iT} = n) + \sum_n \gamma_n T_i \mathbb{I}(N_{iT} = n) + \epsilon_i$$

where  $\epsilon_i$  represents the error from randomized treatment assignments,  $\alpha_1$  is the baseline treatment effect when  $N_{iT} = 0$ ,  $\alpha_t^n$  is the baseline increase from adding the  $n$ th neighbor to treatment when individual  $i$  is in control, and  $\beta_t^n$  is the increase in this baseline when individual  $i$  is in treatment. For the population of interest,  $ATE_{direct}$ ,  $ATE_{total}$ , and the other average effects in between correspond to parameters or sums of parameters. Since  $n$  may be large, a functional form assumption can be imposed on  $N_{iT}$  for estimation.

To generalize this model, we assume that beyond the first degree connection effects above, individuals can be influenced by others' outcomes in a way that is independent of treatment. This is related to the generative model described in [7]. They assumed a dynamic model where the effect of others' treatment on an individual outcome at time  $t$  is completely mediated by previous outcomes at time  $t - 1$ . Instead, our model ignores the dynamic element and uses the additional effect to capture any deviations from the symmetry assumption as well as spillovers from the rest of the network. We can write the potential outcomes of interest as  $Y_i(T_i, N_{iT}, \{Y_j\}_{j \text{ neighbor } i})$ . Here each  $Y_j$  is itself a function of the treatments and potential outcomes of the neighbors of  $j$  and so on. Under independence, we have

$$Y_i = \alpha_0 + \alpha_1 T_i + \sum_n \beta_n \mathbb{I}(N_{iT} = n) + \sum_n \gamma_n T_i \mathbb{I}(N_{iT} = n) + f_\pi(\{\epsilon_j\}) + \epsilon_i$$

To estimate, the same functional form assumptions as above can be used for  $N_{iT}$ . To eliminate the complexity from modeling the errors, we assume they are drawn from a correlated distribution. The reduced model becomes

$$Y_i = \alpha_0 + \alpha_1 T_i + \sum_n \beta_n \mathbb{I}(N_{iT} = n) + \sum_n \gamma_n T_i \mathbb{I}(N_{iT} = n) + \epsilon_i = X_i \beta + \epsilon_i$$

$$\epsilon \sim (0, \Sigma_\epsilon)$$

where  $\Sigma_\epsilon$  models correlations between outcomes. One simplifying assumption to reduce the number of parameters in the covariance matrix is that users have zero partial correlation with other users who are not neighbors. This means that conditional on neighboring outcomes, the individual outcome does not depend on the outcomes for the rest of the network.

Then, the corresponding elements in the inverse covariance matrix are zero for users who are not connected. In case there are many parameters, a structural form can be assumed or regularization can be used.

We use maximum likelihood assuming a Gaussian error distribution to estimate the mean and covariance parameters. This can be viewed as generalized least squares with an estimated covariance. We first consider a constraint that  $\Sigma_\epsilon^{-1}$  belongs to a set  $C$ , where  $C$  is a lower dimensional subset of covariance matrices with the particular sparsity pattern corresponding to the network. We minimize

$$\begin{aligned} & -1/2 \log \det \Sigma_\epsilon^{-1} + 1/2(Y - X\beta)^t \Sigma_\epsilon^{-1} (Y - X\beta) \\ & \text{st } \Sigma_\epsilon^{-1} \in C \end{aligned}$$

This may not be convex, depending on  $C$ , but we iteratively estimate using

$$\begin{aligned} \hat{\beta} &= (X^t \hat{\Sigma}_\epsilon^{-1} X)^{-1} (X^t \hat{\Sigma}_\epsilon^{-1} Y) \\ \hat{\Sigma}_\epsilon^{-1} &= \arg \min -1/2 \log \det \Sigma_\epsilon^{-1} + 1/2 \hat{\epsilon}^t \Sigma_\epsilon^{-1} \hat{\epsilon} \\ & \text{st } \Sigma_\epsilon^{-1} \in C \end{aligned}$$

The simplest constraint is to allow a common partial correlation  $\alpha$  among nodes that are connected.

We also consider adding regularization. We minimize

$$-1/2 \log \det \Sigma_\epsilon^{-1} + 1/2(Y - X\beta)^t \Sigma_\epsilon^{-1} (Y - X\beta) + p_\lambda(\Sigma_\epsilon^{-1})$$

where  $p_\lambda$  is a penalty function. This is biconvex if we take  $p_\lambda$  to be the  $l_1$  or  $l_2$  norms, so we iteratively estimate using

$$\hat{\beta} = (X^t \hat{\Sigma}_\epsilon^{-1} X)^{-1} (X^t \hat{\Sigma}_\epsilon^{-1} Y)$$

$$\hat{\Sigma}_\epsilon^{-1} = \arg \min -1/2 \log \det \Sigma_\epsilon^{-1} + 1/2 \hat{\epsilon}^t \Sigma_\epsilon^{-1} \hat{\epsilon} + p_\lambda(\Sigma_\epsilon^{-1})$$

The second optimization can be solved analytically for the  $l_2$  norm. For the  $l_1$  norm, [9] proposed the graphical lasso algorithm. This penalty produces sparse estimates and may be preferred under our assumption that there are exact zeros in the inverse covariance matrix corresponding to users that are not connected.

With only one observation of the network outcome, imposing strict structural assumptions may be necessary to obtain results, since regularization approaches do poorly with low numbers of observations. However, regularization methods can be generally used when we have multiple observations of the same network over time, allowing more precise estimates of the covariance matrix. Stacking the samples across time

and assuming the average treatment effect remains constant, we have the model

$$\begin{aligned} Y_{it} &= \alpha_0 + \alpha_1 T_i + \sum_n \beta_n \mathbb{I}(N_{iT} = n) \\ &+ \sum_n \gamma_n T_i \mathbb{I}(N_{iT} = n) + \epsilon_{it} \\ &= X_i \beta + \epsilon_{it} \\ \epsilon &\sim N(0, \Sigma_\epsilon) \end{aligned}$$

where  $\Sigma_\epsilon$  in this case models correlations between outcomes and across times. To reduce dimension, we can assume  $\Sigma_\epsilon = \Sigma_{\text{network}} \otimes \Sigma_{\text{time}}$ , the Kronecker product of the covariance corresponding to network structure that we estimated above and the covariance corresponding to time. This may impose a strong assumption on how individual outcomes change based on their network and time step. However, estimating a covariance matrix for the network and time dimension incorporates the relationships we are interested in. We again place structural constraints on each of  $\Sigma_{\text{network}}$  and  $\Sigma_{\text{time}}$  or regularize them. Under regularization, [1] proposed an iterative graphical lasso based algorithm.

The idea of introducing a joint model for mean and covariance is similar to time correlated models and generalized least squares estimation strategies. In this setting, [11] and [12] introduced one possible parametrization that allowed estimation. Our goal is to generalize by modeling a network explicitly in the covariance structure in addition to possible time components in order to obtain more efficient estimates of mean parameters and treatment effects. This in addition gives an estimate of possible spillovers occurring beyond those accounted for in the mean model.

## 4 Simulation results

In our baseline simulations, we assume the form of the generating model is known and find the errors in our parameter estimates. Even though the assumptions made here are unrealistic, this allows us to check our methodology and estimation strategy. We set the true direct treatment effect to one and the functional form on the number of treated neighbors to be linear with slope one for numbers less than two and constant for numbers greater than two. For simplicity, there are no interaction effects. The outcome model is

$$\begin{aligned} Y_i &= T_i + \mathbb{I}(N_{iT} == 1) + 2\mathbb{I}(N_{iT} >= 2) + \epsilon \\ \epsilon &\sim N(0, \Sigma_\epsilon) \end{aligned}$$

For each sample, to generate  $\Sigma_\epsilon$ , we generate a network with  $N = 100$  nodes and choose edges either at random or in cluster formations with each node having degree  $1/25N$ . These models are labeled 2 and 0 respectively in the below tables. The corresponding inverse covariance matrix has common partial correlations of 0.7 for connected node outcomes. The diagonal elements are either 1 or 10. Finally, we vary the number of time steps the model is observed over, setting  $T$  equal to 1 or 10. Here, we assume the errors are uncorrelated in time, so each network observation is independent.

The tables below show the mean squared error of the least squares estimator(1), the generalized least squares estimator given the true covariance(2), and the generalized least squares estimators using the known structural form(3) and using graphical lasso(4). For (3), the structural form is assumed to be constant partial correlations for connected nodes. For (3) and (4) in the 10 observation case, the correlation between time steps is assumed zero. These assumptions correspond to our generating model. The generalized least squares estimator with the true covariance is efficient in this case, since the errors are Gaussian. By comparing our results to (2), we see that empirically we are able to get accurate estimates of the direct and peer treatment effects for all methods. When the edges are random, using the structured covariance model improves performance with either 1 or 10 observations. Graphical lasso does not perform as well unless there are 10 observations. This is expected, since there is too much noise to determine unknown structure from only 1 observation.

model	T	1	2	3	4
1 0	1	0.089	0.072	0.082	0.281
1 0	10	0.009	0.008	0.008	0.013
1 2	1	0.479	0.042	0.045	1.419
1 2	10	0.073	0.004	0.005	0.007
10 0	1	0.011	0.009	0.010	0.015
10 0	10	0.001	0.001	0.001	0.001
10 2	1	0.033	0.006	0.005	0.028
10 2	10	0.004	0.000	0.000	0.001

Table 1: Direct treatment effect MSE (100 samples)

Our next simulation uses the same process for network generation as above but uses a dynamic generating model to determine the covariance matrix for individual outcomes. This dynamic model follows from [7], except we propagate the error terms instead of the outcomes over the network to have independence between the error terms and the terms in the mean model. Using the same mean model as above, we

model	T	1	2	3	4
1 0	1	0.027	0.018	0.021	0.219
1 0	10	0.003	0.002	0.002	0.003
1 2	1	0.285	0.032	0.032	0.775
1 2	10	0.051	0.002	0.003	0.005
10 0	1	0.003	0.002	0.002	0.004
10 0	10	0.000	0.000	0.000	0.000
10 2	1	0.017	0.002	0.002	0.013
10 2	10	0.002	0.000	0.000	0.000

Table 2: Peer treatment effect MSE (100 samples)

model	T	1	2	3	4
1 0	1	0.490	0.000	0.104	0.517
1 0	10	0.490	0.000	0.099	0.477
1 2	1	0.490	0.000	0.209	0.504
1 2	10	0.490	0.000	0.205	0.447
10 0	1	0.490	0.000	0.104	0.519
10 0	10	0.490	0.000	0.098	0.490
10 2	1	0.490	0.000	0.212	0.505
10 2	10	0.490	0.000	0.205	0.463

Table 3: Correlation MSE (100 samples)

have

$$Y_{i0} = T_i + \mathbb{I}(N_{iT} == 1) + 2\mathbb{I}(N_{iT} >= 2) + \epsilon_{i0}$$

$$\epsilon_{i0} \sim N(0, I)$$

$$Y_{it} = T_i + \mathbb{I}(N_{iT} == 1) + 2\mathbb{I}(N_{iT} >= 2) + A\epsilon_{t-1}/d_i + \epsilon_{it}$$

$$\epsilon_{it} \sim N(0, I)$$

for  $t > 0$ . Here,  $A$  is the adjacency matrix and  $d_i$  is the degree of  $i$ . Again, we consider either 1 sample at the last time step or all 10 samples.

The mean squared errors are given below for the least squares estimator(1) and the generalized least squares estimators using the structural form(3) and using graphical lasso(4). Again, we assume the structural form to be constant partial correlations for connected nodes in (3) and zero between time steps in (3) and (4). However, in this simulation, the generating model does not follow this structure. We see that the joint mean and covariance model does not improve performance, likely because the structural assumptions are not met, but it does about the same as the mean modeling. The only exception is again when there is 1 observation and graphical lasso is used to find the structure.

From these simulations, we conclude that under a typical model for individual behavior, covariance modeling is unlikely to outperform just mean modeling unless the errors follow a structural model that is known. With 1 observation, it is difficult to obtain

model	T	1	3	4
1 0	1	0.046	0.046	0.241
1 0	10	0.015	0.015	0.016
1 2	1	0.061	0.060	0.172
1 2	10	0.015	0.015	0.016

Table 4: Direct treatment effect MSE (100 samples)

model	T	1	3	4
1 0	1	0.044	0.044	0.158
1 0	10	0.017	0.017	0.013
1 2	1	0.067	0.067	0.183
1 2	10	0.012	0.013	0.011

Table 5: Peer treatment effect MSE (100 samples)

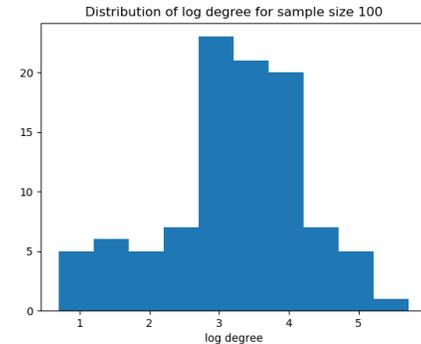
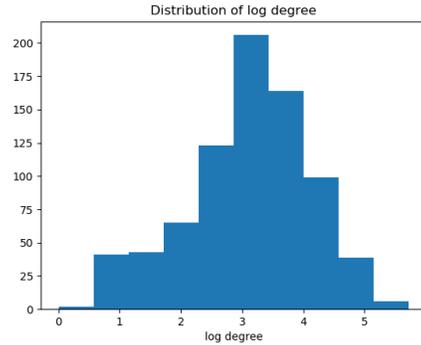
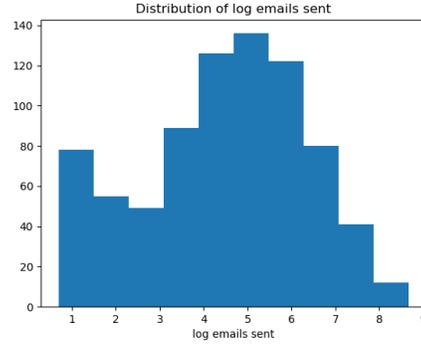
better estimates. However, we hypothesize that with 10 observations and without the time independence assumption, using the Kronecker model to decompose the covariance into time and network components may improve performance. We do not provide results for this estimation procedure in this paper.

## 5 Email simulation results

Suppose that we consider a simulated experiment on top of an temporal email network: <http://snap.stanford.edu/data/email-Eu-core-temporal.html>. For example, this could represent a messaging addition or change made to the email service. To measure the effect on number of emails sent, suppose we choose a stratified random sample of users to introduce the change to. We divide the time period into 10 steps and assume that the given data represents the potential outcome under no treatment. We make up the potential outcome under treatment following the above dynamic model of user behavior for 10 steps. We take log of emails sent as the outcome since the distribution of emails sent is skewed. The distribution of log emails is plotted below.

We present some differences between our simulated data and the email network. In the simulations, we only considered 100 nodes with constant degrees. The connections were either formed randomly or in exact clusters. For the data, we choose all users with at least one email sent during our chosen time period, giving 788 nodes. We then subsample 100 nodes for each sample and define a connection if there was an email sent between two nodes. The overall degree distribution for all 788 nodes and an example for 100 nodes is shown below.

We test how well each of our procedures captures the structure in this network and the mean squared



errors are tabled below. The graphical lasso performs well in this case for 10 time steps. This may be because the skew in the degree distribution varies the correlation between connected nodes. This may bias the results when the observations are assumed independent or correlated with the same factor.

## 6 Conclusion

In this paper, we consider the problem of direct and peer treatment effect estimation. Under strict assumptions of no spillovers beyond first degree connec-

T	1	3	4
1	2.037	2.037	2.704
10	0.942	0.942	0.107

Table 6: Direct treatment effect MSE (100 samples)

T	1	3	4
1	1.154	1.157	2.019
10	1.377	1.378	0.961

Table 7: Peer treatment effect MSE (100 samples)

tions, average effects can be computed by taking difference in means if the sample is randomized for treatment and number of peers treated. If there are potential spillovers from the rest of the network, we introduce and analyze the performance of some reduced form estimation strategies. Ordinary least squares performs best when there are no repeated observations. If observations are taken at different times, it may be useful to model the covariance matrix and use a maximum likelihood estimation approach. This is under the assumption that the mean model is constant over time, so each individual’s treatment and treatment effect does not change.

## References

- [1] Genevera I. Allen and Robert Tibshirani. “Transposable regularized covariance models with an application to missing data imputation”. In: *The Annals of Applied Statistics* 4.2 (2010), pp. 764–790.
- [2] Sinan Aral, Lev Muchnik, and Arun Sundararajan. “Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks”. In: *PNAS* 106.51 (2009), pp. 21544–21549.
- [3] Sinan Aral and Dylan Walker. “Creating social contagion through viral product design: A randomized trial of peer influence in networks”. In: *Management Science* 57.9 (2011), pp. 1623–1639.
- [4] Sinan Aral and Dylan Walker. “Identifying influential and susceptible members of social networks”. In: *Science* 337.6092 (2012), pp. 337–341.
- [5] Eytan Bakshy et al. “The role of social networks in information diffusion”. In: *arXiv preprint arXiv: 1201.4145* (2012).
- [6] Dean Eckles and Eytan Bakshy. “Bias and high-dimensional adjustment in observational studies of peer effects”. In: *arXiv preprint arXiv: 1706.04692* (2017).
- [7] Dean Eckles, Brian Karrer, and Johan Ugander. “Design and analysis of experiments in networks: Reducing bias from interference”. In: *arXiv preprint arXiv: 1404.7530* (2014).
- [8] Dean Eckles, Rene F. Kizilcec, and Eytan Bakshy. “Estimating peer effects with peer encouragement designs”. In: *PNAS* 113.27 (2016), pp. 7316–7322.
- [9] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. “Sparse inverse covariance estimation with the graphical lasso”. In: *Biostatistics* 9.3 (2008), pp. 432–441.
- [10] Ravi Jagadeesan, Natesh Pillai, and Alexander Volfovsky. “Designs for estimating the treatment effect in networks with interference”. In: *arXiv preprint arXiv: 1705.08524* (2017).
- [11] Mohsen Pourahmadi. “Joint mean-covariance models with applications to longitudinal data: unconstrained parametrisation”. In: *Biometrika* 86.3 (1999), pp. 677–690.
- [12] Mohsen Pourahmadi. “Joint mean-covariance models with applications to longitudinal data: unconstrained parametrisation”. In: *Biometrika* 87.2 (2000), pp. 425–435.