UC Berkeley
CEGA Working Papers

Title

Estimating spillovers using imprecisely measured networks

Permalink

Authors

Hardy, Morgan
Heath, Rachel M
Lee, Wesley
et al.

Publication Date

2019-05-01

Estimating spillovers using imprecisely measured networks

Morgan Hardy, Rachel M. Heath, Wesley Lee, Tyler H. McCormick

CEGA

Center for Effective Global Action

Working Paper Series

Center for Effective Global Action

University of California

eScholarship
University of California

Estimating spillovers using imprecisely measured networks

Morgan Hardy
New York University - Abu Dhabi

Rachel M. Heath
University of Washington

Wesley Lee
University of Washington

Tyler H. McCormick*
University of Washington

Abstract

In many experimental contexts, whether and how network interactions impact the outcome of interest for both treated and untreated individuals are key concerns. Networks data is often assumed to perfectly represent these possible interactions. This paper considers the problem of estimating treatment effects when measured connections are, instead, a noisy representation of the true spillover pathways. We show that existing methods, using the potential outcomes framework, yield biased estimators in the presence of this mismeasurement. We develop a new method, using a class of mixture models, that can account for missing connections and discuss its estimation via the Expectation-Maximization algorithm. We check our method's performance by simulating experiments on real network data from 43 villages in India. Finally, we use data from a previously published study to show that estimates using our method are more robust to the choice of network measure.

1 Introduction

Interactions between peers are of interest in many economic settings, such as health (Oster and Thornton, 2011; Godlonton and Thornton, 2012), education (Angelucci *et al.*, 2010; Duflo *et al.*, 2011), job search (Magruder, 2010; Wang, 2013; Heath, 2018), personal finance (Bursztyn *et al.*, 2014), agriculture (Cai *et al.*, 2015; BenYishay and Mobarak, 2018; Beaman *et al.*, 2018; Vasilaky and Leonard, 2018), and microenterprises (Hardy and McCasland, 2016). Moreover, even when spillovers to non-treated peers are not of direct interest, the possibility of treatment spillovers to the control group violates the stable unit treatment

---

*Contact email: tylermc@uw.edu

value assumption (SUTVA) needed to identify causal treatment effects (Rubin, 1974). In both cases, knowing the group of peers who are potentially affected by a treatment allows researchers to accurately estimate peer effects and assess potential SUTVA violations.

However, measuring social networks is challenging. It is expensive to collect data on an individual's entire social network (Breza et al., 2017), leading researchers to use data on proxies for networks such as geography (Foster and Rosenzweig, 1995; Miguel and Kremer, 2004; Bayer et al., 2008; Godlonton and Thornton, 2012), or sharing a common nationality (Beaman, 2011), language (Bandiera et al., 2009, 2010), ethnic group (Fafchamps, 2003), religion (Munshi and Myaux, 2006), or caste (Munshi and Rosenzweig, 2006, 2016). Even if researchers collect network data,[1] the set of individuals potentially affected by a given treatment may not entirely correspond to the elicited network if networks are truncated due to concerns about survey fatigue[2], or the experiment changes the network itself (Comola and Prina; Banerjee et al., 2018; Stein, 2018). It is also difficult to ask respondents to specify the precise set of individuals potentially affected by a given treatment by asking about either past interactions or hypothetical future interactions (Hardy and McCasland, 2016).

In this paper, we focus on the setting where the observed network represents a corrupted version of the true network of treatment interference, allowing for both unreported spillover pathways and misreported links over which no spillovers could occur. We use a local network exposure approach (Aronow and Samii, 2017; Ugander et al., 2013) that defines a set of individuals whose treatment status can potentially affect each subject. We first show missing links and misreported links in the network can cause mismeasured treatment exposure conditions and hence biased estimators. We develop a class of mixture models that can model the distribution of the latent true exposure conditions and discuss parameter estimation using the Expectation-Maximization (EM) algorithm. These models rely on parametric

---

[1]There is a substantial empirical literature that addresses the reliability of network data elicited through surveys, with emphasis on the type and salience of relationships being surveyed, temporal dependence, and how links are elicited. See Bell et al. (2007); Marsden (2016) for reviews.

[2]Griffith (2017a) shows how limiting the number of peers a subject can report in the data from the National Adolescent Health Project can bias estimates in the linear-in-means model.

assumptions about the distribution of missing links conditional on the observed network data as well as parametric assumptions on the behavior of outcomes within each treatment exposure condition. Under a linear regression model for the latter assumptions, we prove the mixture model is identifiable and the maximum likelihood estimator from the EM algorithm is consistent.

We evaluate our method with both simulations and replication of an existing study. We simulate experiments on real networks of Indian households from 47 villages. We are able to recover accurate estimates of direct and indirect treatment effects when state-of-the-art methods fail. Finally, we implement our method using networks data from a randomized evaluation of insurance information sessions with rural farmers in China (Cai *et al.*, 2015). We find that our method produces more consistent estimates of direct and indirect treatment effects across various choices of network measures.

Our results are relevant to many experimental contexts where subject's behavior or outcome may be influenced by other subjects' treatment assignment in addition to their own. A common approach in these cases is to randomize treatment at a geographic or organizational level that plausibly contains each treated individual's network of potential spillovers, such as a village (in isolated, rural settings), and then compare treated individuals to "pure controls" in non-treated units. However, even if this is possible, comparing treated to control subjects still confounds treatment effects and spillovers on these treated subjects.[3] Moreover, in other cases, a pure control is not feasible, because the experiment must be implemented within a single firm (Bandiera *et al.*, 2009; Bloom *et al.*, 2014; Adhvaryu *et al.*, 2016) or market (Conlon and Mortimer, 2013), or it is not possible to leave a large enough buffer between treated and control areas to render spillovers unimportant. Potential SUTVA violations could then introduce both upward and downward bias in direct treatment effect estimates.[4]

---

[3]An exception would be if the treated individuals are a small enough fraction of treated units that they are unlikely to know treated subjects. Comparing treatment to control individuals would then identify the average direct treatment effect by construction. However, this would likely require a large enough number of units to be impractical or prohibitively expensive in many settings. Treatment effects in such contexts are also not particularly informative about the results from scaling up a treatment to an entire population.

[4]The reduction in exposure to disease from directly treated school children in Kenya may indirectly

The local network exposure approach (Aronow and Samii, 2017; Ugander *et al.*, 2013) that we use contrasts with linear-in-means models (Manski, 1993; Bramoullé *et al.*, 2009) in its assumed avenues of treatment interference. Local network exposure models assume the avenues of interference for each subject are limited to the treatment assignments of other subjects in their local neighborhood. Discrete treatment "exposure conditions" are defined based on a subject's and their connections' treatment assignments, and are used within Rubin's potential outcome framework as a set of potential treatment conditions (as opposed to using the levels of treatment). On the other hand, linear-in-means models (Manski, 1993; Bramoullé *et al.*, 2009) hypothesize indirect treatment effects manifest as a linear relationship between a subject's outcome and the average treatment and average outcome of that subject's peers. The dependence between outcomes of connected subjects allows for the a subject's outcome to be influenced by any other subject to which they are directly or indirectly connected to in the social network, with the amount of influence being modulated by their distance in the network. While we use a local network exposure model that allows us to focus on the effect of exposure to at least one treated subject, in section 2.3, we explain how our approach can allow for similar dynamics as a linear-in-means model if we increase the number of indirect treatment bins assumed to influence an individual.

Our approach is related to a growing literature in economics, political science, sociology, and statistics on network sampling and mismeasurement. Fundamentally, we draw a distinction between the observed network data and the true latent network. That is, we view the observed graph as a corrupted version of a true, unobserved, network. As in Handcock and Gile (2010) and Newman (2018), we relate the two via a probabilistic model and, given a set of model parameters, construct a distribution over the true network conditional on the observed graph. Existing methodologies consider cases of sampled networks without measurement error, where some of the potential links are intentionally not elicited. In these

---

improve the health outcomes of school children who did not directly receive the treatment, biasing downward naively estimated benefits of deworming pills (Miguel and Kremer, 2004). In contrast, increased police patrolling on the streets of Bogota, Colombia, may merely push crime "around the corner", biasing upward the estimated impact on crime rates (Blattman *et al.*, 2017).

4

cases, estimating model parameters relating the observed graph to the true latent network can be relatively straightforward.[5] In effect, the sampled graph can act as a training set from which to learn a probabilistic model that characterizes the full network. In contrast, our paper considers the case where all potential links are measured, but may contain some error. In this case, even a full graph cannot be used to train probabilistic models, because of the potential for error on every link (and non-link). This creates an inability to learn the parameters of the corruption process. For example, the observed network data does not inform the proportion of true links missing from the observed graph.

This paper proceeds as follows. In Section 2 we discuss existing methods for estimating direct and indirect treatment effects, with focus on a model from the potential outcomes framework. We characterize this model's limitations in the presence of missing links. In Section 3 we propose a mixture model to estimate treatment effects that can account for latent ties between subjects. We discuss when this model is identified, how to estimate model parameters and treatment effects, and examine model performance using simulations. We apply our methodology to an experiment in Section 4, and conclude in Section 5 with a discussion.

2 Measuring network spillovers in experiments

2.1 Notation

In this section we introduce some basic notation that we will use throughout the rest of the paper. Let $i \in \{1, 2, \cdots, N\}$ index the subjects in the study, with corresponding observed outcomes y_i, which we vectorize as $\mathbf{y}$. For simplicity, suppose treatment is binary with levels "treatment" (1) and "control" (0), and the treatment assignment mechanism is random and explicitly known. Denote the vector of treatment assignment with $\mathbf{t} \in \{0, 1\}^N$, in which

---

[5]For example, Chandrasekhar and Lewis (2011) shows how egocentrically sampled network data can be used to predict the "full" network in a process they term graphical reconstruction. See Williams (2016) and Griffith (2017b) for sample applications.

the treatment of individual i is t_i. Suppose the true influence network G is directed and binary, with the edge $i \to j$, representing individual i's influence on individual j, encoded by $G_{ij} = 1$. Let G_j denote the jth column of G, indicating the influencers of individual j, so $\mathbf{1}'G_j$ is the the number of influencers or in-degree of j. For now, let us assume G is observed. Finally, let $\overline{G_j}$ denote the jth column of G normalized to sum to 1 ($\mathbf{1}'\overline{G_j} = 1$) unless $1'G_j = 0$, in which case $\overline{G_j} = G_j = \mathbf{0}$.

2.2 Local network exposure model

Aronow and Samii (2017) and Ugander et $al.$ (2013) propose estimators for average direct and indirect treatment effects by building on the Rubin causal model (Rubin, 1974). In the context of experiments, each subject has a set of "potential outcomes" $(Y_i(t_i = 0), Y_i(t_i = 1))$ corresponding to the possible outcomes under each treatment. The inference task is to estimate the average treatment effect, defined to be the difference between the average outcome of the population if the entire population was treated and the average outcome if the entire population was in the control:

$$ATE(1,0) = \frac{1}{N} \sum_{i=1}^{N} \left[Y_i(t_i = 1) - Y_i(t_i = 0) \right]. \tag{1}$$

This quantity is not observed since we cannot observe the full set of potential outcomes for each subject, but assuming completely random assignment can be estimated by the difference in sample means:

$$\widehat{ATE(1,0)} = \frac{1}{N_1} \sum_{i=1}^{N} y_i \mathbf{1}[t_i = 1] - \frac{1}{N_0} \sum_{i=1}^{N} y_i \mathbf{1}[t_i = 0], \tag{2}$$

where N_k is the number of subjects in treatment k. A crucial assumption in the Rubin causal model is SUTVA, which states than a subject's potential outcomes are unaffected by the treatments of other subjects. In experiments on networks, SUTVA is violated if the

treatments of peers influence the outcomes for an individual.

Aronow and Samii (2017) considers a violation of SUTVA by allowing for individuals to be systematically affected by the treatment assignments of their peers. By making assumptions that restrict the nature of these influences, they induce mappings of the treatment vector $\mathbf{t}$ to distinct "exposure conditions", or what Manski (2013) terms "effective treatments." In a simple instance of their framework, which we borrow for our model in Section 3, individuals are affected by whether or not any of their influencers in G are treated, inducing a random assignment into one of four exposure conditions, corresponding to levels of direct and indirect exposure to treatment:

$$
C_i \equiv C_i(t, G_i) =
\begin{cases}
c_{00} & \text{(No Exposure)}: & t_i = 0 \text{ and } \mathbf{t}'G_i = 0 \\
c_{01} & \text{(Indirect Exposure)}: & t_i = 0 \text{ and } \mathbf{t}'G_i > 0 \\
c_{10} & \text{(Direct Exposure)}: & t_i = 1 \text{ and } \mathbf{t}'G_i = 0 \\
c_{11} & \text{(Full Exposure)}: & t_i = 1 \text{ and } \mathbf{t}'G_i > 0.
\end{cases}
\tag{3}
$$

In this model, both direct and indirect effects are taken to be binary, with an individual being indirectly exposed to treatment if one or more of their influencers are (directly) treated. Each subject i would have four potential outcomes $(Y_i(C_i = c_{00}), Y_i(C_i = c_{01}), Y_i(C_i = c_{10}), Y_i(C_i = c_{11}))$, one for each exposure condition. Note this setup assumes that an individual can only be influenced by a first-order connection in the network and the number of connections treated does not have an effect beyond the presence or absence of at least one. The choice of indirect exposure can be related to diffusion models of information and disease in which "contagion" can occur given a single source of exposure (Centola and Macy, 2007). Both of these assumptions could be relaxed, however, simply by adding additional potential outcomes.

The primary quantities of interest would then be given by average treatment effects akin to equation (1):

$$
ATE(c, d) = \frac{1}{N} \sum_{i=1}^{N} \left[Y_i(C_i = c) - Y_i(C_i = d) \right].
\tag{4}
$$

The average direct treatment effect would be given by $ATE(c_{10}, c_{00})$, while the average indirect treatment effect when not directly treated would be $ATE(c_{01}, c_{00})$. Estimating these quantities is equivalent to estimating the mean outcomes of the entire population under each exposure condition:

$$\mu_c = \frac{1}{N} \sum_{i=1}^{N} Y_i(C_i = c), \tag{5}$$

so we focus on the latter for this section and the next, with the additional assumption that each subject is assigned to treatment with some constant and known probability independently of other subjects. Note if certain subjects have zero probability of being placed in certain exposure conditions, e.g. when a subject has no influencers, estimation must be restricted to the sub-population of individuals with non-zero probability of being placed in every condition. In contrast to the case when the SUTVA assumption is satisfied, we cannot estimate these means using just their sample counterparts. Variability in the in-degrees of individuals causes variation in the probabilities of assignment to each exposure condition. Namely, individuals with high in-degree are more likely to be indirectly exposed to the treatment since they have more influencers who potentially may be treated. This selection bias could affect the mean estimates if there is heterogeneity in the outcomes within exposure conditions associated with in-degree. Horvitz-Thompson estimators use inverse probability weighting in order to take varying exposure probabilities into account to produce unbiased estimators of these mean outcomes:

$$\widehat{\mu}_{c,HT} = \frac{1}{N} \sum_{i=1}^{N} \frac{y_i \cdot \mathbf{1}_{[C_i=c]}}{P(C_i = c)}. \tag{6}$$

Note this estimator is equivalent to the sample mean if the probability of assignment to a cluster is constant among subjects. These estimators are unbiased regardless of the form of the heterogeneity between the outcomes and network degrees, but can have high variance when the exposure conditions are highly imbalanced on in-degree. This would arise when the probabilities $P(C_i = c)$ are small for some i, yielding large weights $\frac{1}{P(C_i=c)}$.

Explicitly modeling the relationship between potential outcomes and network degrees can result in more stable estimators at the cost of additional assumptions about the validity of these relationships. For example, suppose we believe that for each exposure condition c, the relationship for the in-degree $(\mathbf{1}'G_i)$ and the potential outcome $Y_i(C_i = c)$ can be modeled with

$$Y_i(C_i = c) \sim f\left(\cdot; \theta_c, \mathbf{1}'G_i\right), \tag{7}$$

where $\mathbf{1}'G_i$ is the in-degree of individual i and θ_c are model parameters. Assuming this model accurately characterizes the relationship between the potential outcomes and in-degrees, the distribution of potential outcomes is conditionally independent of the exposure assignment (induced by the treatment assignment) vector given the in-degrees of the subjects, such that the exposure assignment mechanism can be "ignored" during inference of the means (Rubin, 1974). The estimate of the mean outcome under exposure condition c would then be given by

$$\widehat{\mu}_{c,R} = \frac{1}{N}\sum_{i=1}^{N}\widehat{Y}_i(C_i = c) = \frac{1}{N}\sum_{i=1}^{N} E_{f\left(\cdot; \widehat{\theta}_c, \mathbf{1}'G_i\right)}[y_i], \tag{8}$$

provided an estimate of model parameters $\widehat{\theta}_c$. Parametric models $f\left(\cdot; \theta_c, \mathbf{1}'G_i\right)$ of the outcomes under condition c and in-degree d are necessary for likelihood-based approaches to estimation and are used in the model we propose in Section 3. A common model familiar to many economists is

$$f\left(\cdot; \theta_c \equiv (\alpha_c, \beta_c, \sigma^2), \mathbf{1}'G_i\right) = \mathrm{N}\left(\cdot; \alpha_c + \beta_c\mathbf{1}'G_i, \sigma^2\right), \tag{9}$$

which corresponds to a linear model with different intercepts and slopes for each exposure condition (but common variance). In this case, the estimates of mean outcome would be given by $\widehat{\mu}_{c,R} = \widehat{\alpha}_c + \widehat{\beta}_c \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}'G_i$.

2.3 Relation to the linear-in-means model

Although we focus on the local network exposure model in the subsequent sections, let us first compare it to another popular approach for accounting for and measuring treatment spillovers: Manski linear-in-means models (Manski, 1993; Bramoullé *et al.*, 2009). In the context of experiments without additional covariates, a basic form of these models would be

$$y_i = \alpha + \beta y' \overline{G_i} + \gamma t_i + \delta t' \overline{G_i} + \epsilon_i, \tag{10}$$

In addition to allowing for "endogenous effects" ($\beta \neq 0$), which would account for second- and higher-order indirect treatment effects (e.g. how an individual is affected by their peers' peers' outcomes), the linear-in-means-model differs from the described local network exposure model by placing a different set of assumptions on the mechanism of indirect treatment effects. Rather than partitioning indirect effects into varying magnitudes based on indirect exposure conditions, the linear-in-means model assumes the size of indirect effects vary linearly with the proportion of peers treated. Note a similar assumption could be used with the local network exposure setting by characterizing indirect exposure with proportions of peers treated instead of the presence of any peers treated, albeit these proportions would have to be arranged into discrete bins. Similarly, we could introduce a non-linear dependence on the number of treated peers in the linear-in-means framework by adding multiple indicator variables to Equation 10. In both cases, introducing more discrete potential outcome categories will likely lead to small cell counts in practice. This observation highlights the importance of the linearity assumption in the linear-in-means model.

2.4 Local network exposure model under mismeasurement

In this section we consider the potential for bias in Horvitz-Thompson estimators (6) if using a corrupted network, $\widetilde{G}$, to estimate exposure conditions instead of the true network G. We allow $\widetilde{G}$ to be corrupted such that there are either links present in $\widetilde{G}$ that are not in G or

vice-versa. Suppose our treatment assignment mechanism $\mathbf{t}$ is constructed such that each subject i has positive probability of being placed in treatment and positive probability of being placed in control. We can break the impact of using $\widetilde{G}$ in estimation into two distinct factors. First, recall that the Horvitz-Thompson estimator can only be used for subjects with non-zero probability of being placed in each exposure condition. Namely, subjects with zero in-degree must be excluded, reflecting the idea that a potential outcome under indirect exposure only makes sense if the subject could be indirectly exposed to treatment under some hypothetical treatment assignment. When we observe a corrupted version of the network, we may not be able to accurately identify which subjects should be excluded. Certain individuals who have positive in-degree in G may be observed to have zero in-degree in $\widetilde{G}$ and thus would be incorrectly excluded for estimation. At the same time, certain individuals with $\mathbf{1}'G_i = 0$ may be observed to have positive in-degree and thus be included during estimation. If either of these situations arose, our estimated average outcomes would then represent a different subpopulation than the true population of subjects with non-zero in-degree.

Second, even if we are able to accurately identify all subjects with non-zero in-degree, bias in mean estimates may be induced by distorted observed exposure conditions. Subjects who are in truth indirectly exposed to treatment would not be observed to be indirectly exposed if all connections to influencers who are treated are unobserved (and no false links to other treated individuals are observed). Similarly, subjects not indirectly exposed to treatment may be falsely observed to be indirectly exposed. The corrupted exposure conditions are able to correctly identify the level of direct treatment for each subject but not necessarily the level of indirect treatment. Mathematically, observing $\widetilde{C}_i \equiv C_i(\mathbf{t}, \widetilde{G}_i) = c_{kl}$ for any $k, l \in \{0, 1\}$ may correspond to either $C_i(\mathbf{t}, G_i) = c_{k0}$ or $C_i(\mathbf{t}, G_i) = c_{k1}$. Recall that in this notation the first subscript denotes the direct treatment condition (whether i is directly treated or not) and the second subscript denotes the indirect treatment (whether at least one member of i's network was treated). The Horvitz-Thompson estimators for each treatment exposure

condition c under the corrupted network $\widetilde{G}$ are given by

$$\widehat{\mu}_{c,HT,\widetilde{G}} = \frac{1}{N} \sum_{i=1}^{N} \frac{y_i \cdot \mathbf{1}_{[\widetilde{C}_i = c]}}{P(\widetilde{C}_i = c)}. \tag{11}$$

where observed $y_i = \sum_c Y_i(C_i = c)\mathbf{1}_{[C_i=c]}$ is dependent on the true exposure condition and the probabilities are taken over possible treatment assignments $\mathbf{t}$. Holding the observed and true networks fixed and taking the expectation of the estimators $\widehat{\mu}_{c,HT,\widetilde{G}}$ over the possible treatment assignments $\mathbf{t}$ we have:

$$E\left[\widehat{\mu}_{kl,HT,\widetilde{G}}\right] = \frac{1}{N} \sum_{i=1}^{N} \frac{E\left[y_i \cdot \mathbf{1}_{[\widetilde{C}_i = c_{kl}]}\right]}{P(\widetilde{C}_i = c_{kl})} \tag{12}$$

$$= \frac{1}{N} \sum_{i=1}^{N} \frac{\sum_c Y_i(C_i = c)P\left(\widetilde{C}_i = c_{kl}, C_i = c\right)}{P(\widetilde{C}_i = c_{kl})} \tag{13}$$

$$= \frac{1}{N} \sum_{i=1}^{N} Y_i(C_i = c_{k0})P\left(C_i = c_{k0}|\widetilde{C}_i = c_{kl}\right) + Y_i(C_i = c_{k1})P\left(C_i = c_{k1}|\widetilde{C}_i = c_{kl}\right). \tag{14}$$

We find the mean estimate of the c_{kl} conditioned on $\widehat{\mu}_{kl,HT,\widetilde{G}}$ under the corrupted network $\widetilde{G}$ tends to lie between the mean outcomes under the two exposure conditions corresponding to the same level of direct treatment: μ_{k0} and μ_{k1}. As an example, if we assume the probability of observing an incorrect exposure condition is constant for all i, then
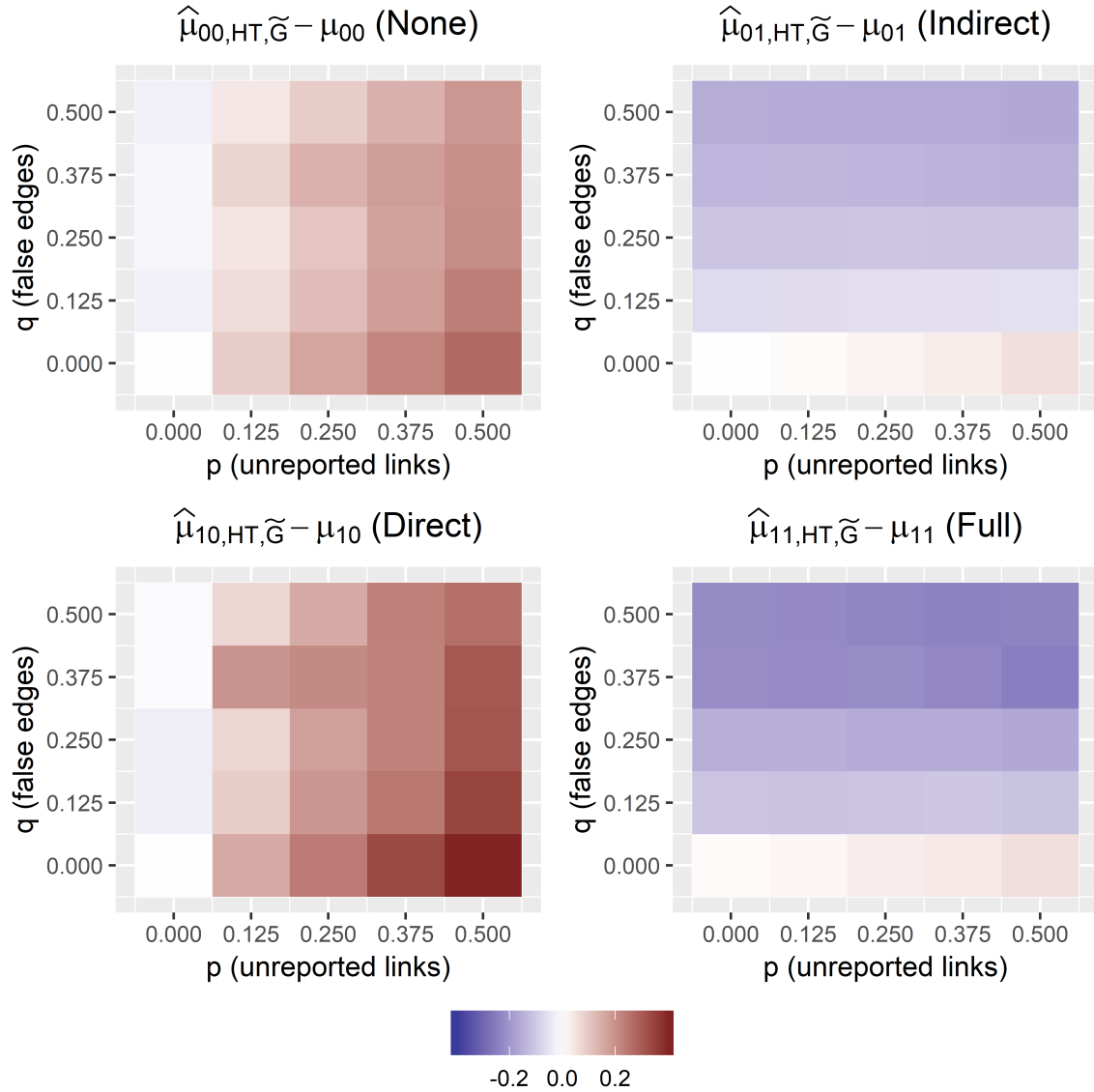
$$E\left[\widehat{\mu}_{kl,HT,\widetilde{G}}\right] = \frac{1}{N} \sum_{i=1}^{N} w_{kl}Y_i(C_i = c_{k0}) + (1 - w_{kl})Y_i(C_i = c_{k1}) \tag{15}$$

$$= w_{kl}\mu_{kl} + (1 - w_{kl})\mu_{k(1-l)} \tag{16}$$

for some $w_{kl} \in [0, 1]$ and our estimate is a linear combination of these two mean outcomes.

In Figure 1, we illustrate the bias of the Horvitz-Thompson estimators in our simulated experiments on the networks of 75 Indian villages using data from Banerjee *et al.* (2013). Our simulation scenarios vary the proportion of true links which are unreported (p) and the

proportion of falsely observed links[6] (q), where the links being added/dropped are completely at random (an assumption we relax later on). Further details of the simulation are provided in Section 3.4, but we show results here as an empirical counterpart to the finding above.

$$\widehat{\mu}_{00,HT,\widetilde{G}} - \mu_{00} \text{ (None)}$$

$$\widehat{\mu}_{01,HT,\widetilde{G}} - \mu_{01} \text{ (Indirect)}$$

$$\widehat{\mu}_{10,HT,\widetilde{G}} - \mu_{10} \text{ (Direct)}$$

$$\widehat{\mu}_{11,HT,\widetilde{G}} - \mu_{11} \text{ (Full)}$$

Figure 1: *Bias of the Horvitz-Thompson estimates for the mean outcome under each exposure condition, for various mismeasurement scenarios (indexed by p and q). Intensity of color is used to denote the magnitude of these biases, with upwards biases colored red and downwards biases colored blue.*

First, consider the behavior of our estimators when we fix q and vary the proportion of

---

[6]More specifically, the probability of observing a link which should not be observed, scaled by the density of true network.

unreported true edges p. As p increases, more links to treated subjects are dropped, leading to a larger proportion of subjects with indirect treatment to be falsely classified as not indirectly treated. The bias in the Horvitz-Thompson estimators for no exposure $\widehat{\mu}_{00,HT,\widetilde{G}}$ and direct exposure $\widehat{\mu}_{10,HT,\widetilde{G}}$ increase upwards, as these estimators trend towards the mean outcomes under the indirect and full exposure conditions respectively (in these simulations, $\mu_{00} < \mu_{01}$ and $\mu_{10} < \mu_{11}$). The bias in the HT estimators for indirect exposure and full exposure also increase upwards, but to a smaller magnitude. Since true links are dropped independently regardless of the treatment status of the subjects involved, the reduced number of subjects observed to be indirectly treated is accounted for by a decrease in the probabilities in the denominator of Equation 11. Rather, the additional bias for these estimators arises solely from low-degree subjects being removed from the estimation procedure (in these simulations, higher degree subjects tend to have higher outcomes). This effect is more pronounced at low q, when fewer (if any) false edges are being added and we are more likely to observe zero-degree subjects.

Similar patterns of behavior emerge when we fix p and vary q. As q increases, more subjects with no true connections to treated individuals are falsely observed to be indirectly treated. As a result, the HT estimators for indirect exposure and full exposure bias further downwards, towards the mean outcomes for no exposure and direct exposure respectively. Additionally, individuals with zero-degree are more likely to be included in the HT estimation procedure, biasing all four estimates downwards. Note this effect is not attenuated at higher p, since zero-degree individuals have no true links to drop to begin with.

3   Latent Variable Model for Network Spillovers

In this section we propose a latent variable approach to estimating average treatment effects when the network observed is a noisy representation of the true network of interest. We assume that each true edge $G_{ij} = 1$ is not observed $\widetilde{G}_{ij} = 0$ with probability p, non edges

$G_{ij} = 0$ are observed with probability q, and edges are observed/not observed independently of one another. These corruption mechanisms assume the observed edges are a random subset of the true edges and the false edges are a random subset of the non-edges, and notably ignores any differences in saliency between edges. We have relaxed this assumption to include mechanisms which allow for adding/subtracting edges to depend on observed covariates, though we do not include these results in this draft.

3.1   Latent Variable Model

Suppose the true network of interest G is unobserved and we only observe a corrupted network $\widetilde{G}$. Furthermore, assume the effects of treatment can be characterized with the four exposure conditions defined in (3). For individual i, we observe corrupted exposure condition $C_i(t, \widetilde{G}_i)$ and in-degree $\mathbf{1}'\widetilde{G}_i$. The statistical problem is then to model the relationship between these corrupted statistics and their true, latent, counterparts. Given a distribution over the true exposure condition $C_i(t, G_i)$ and in-degree $\mathbf{1}'G_i$, we can use models like those in equations (7) and (8) to estimate mean outcomes for each exposure category. For notational simplicity, let $\widetilde{C}$ represent the vector of corrupted exposure conditions, $\mathbf{1}'\widetilde{G}$ the vector of corrupted in-degree, and C and $\mathbf{1}'G$ the corresponding latent terms.

Consider subject i, who has exposure condition $C_i(t, G_i) \equiv c_{kl}$, degree $\mathbf{1}'G_i \equiv d$, and $t'G_i \equiv d_t$ connections with treated subjects, but for whom we observe exposure condition $C_i(t, \widetilde{G}_i) \equiv c_{\tilde{k}\tilde{l}}$, degree $\mathbf{1}'\widetilde{G}_i \equiv \tilde{d}$, and $t'\widetilde{G}_i \equiv \tilde{d}_t$ connections with treated individuals instead. Holding the treatment assignments to be fixed, we can separately model the number of connections to treated subjects d_t and the number of connections to not-treated subjects $d - d_t$, from which we can derive the induced exposure conditions. Note this procedure works for any indirect exposure conditions entirely characterized by the number of treated connections and the number of total connections (e.g. ratio of treated connections) and not just (3). Following Bayes' rule and noting we observe $\tilde{d}_t$ treated connections when x of the d_t actual treated connections are dropped and another $\tilde{d}_t - d_t + x$ false connections to treated

individuals are observed,

$$P(t'G_i = d_t | t, t'\widetilde{G}_i = \tilde{d}_t; p, q) \propto P(t'\widetilde{G}_i = \tilde{d}_t | t, t'G_i = d_t; p, q) P(t'G_i = d_t) \tag{17}$$

$$\propto \sum_{x=0}^{d_t} \text{Bin}(x; d_t, p) \text{Bin}(\tilde{d}_t - d_t + x; \mathbf{1}'t - t_i - d_t, q) P(t'G_i = d_t) \tag{18}$$

where $\text{Bin}(x; n, p)$ is the probability of x successes from a binomial distribution with n attempts and success probability p. Similarly for connections for non-treated subjects,

$$P((1-t)'G_i = d_{nt} | t, (1-t)'\widetilde{G}_i = \tilde{d}_{nt}; p, q) \propto \sum_{x=0}^{d_{nt}} \text{Bin}(x; d_{nt}, p)$$

$$\times \text{Bin}(\tilde{d}_{nt} - d_{nt} + x; N - 1 - \mathbf{1}'t + t_i - d_{nt}, q)$$

$$\times P((1-t)'G_i = d_{nt}) \tag{19}$$

Both sets of equation require a (prior) model over the number of true connections to treated and un-treated subjects. Assuming no additional information about the structure of the true network, one of the most simplistic models would be to model the true graph as an Erdos-Renyi graph. Under this model, the number of connections to treated/un-treated subjects could be modeled with binomial distributions. However, in many real-world networks we find that the degree distribution demonstrates extra-binomial variation, and thus in the following sections we prefer to use a beta-binomial model. With a beta-binomial distribution, we can think of each degree as being sampled from a binomial distribution $d \sim \text{Binom}(N-1, p)$, where p is independently sampled from a Beta distribution $p \sim \text{Beta}(\alpha, \beta)$ for each draw. Rather than α and β, we find it helpful to characterize beta-binomial distributions in terms of an average probability of success μ and an overdispersion parameter ρ. The variance of this beta-binomial would be given by $(N-1)\mu(1-\mu)(1+(N-2)\rho)$, compared to $(N-1)\mu(1-\mu)$ for a binomial distribution with parameter μ. We leave the these parameters to be chosen

16

on a application-by-application basis, noting that the choice of these parameters are more influential when there is high corruption in the network and hence higher uncertainty over the true degrees [7].

Using the above equations, we can express the relationship between the true exposure condition and degree and their observed counterparts:

$$\tau_i(c_{kl}, d; p, q) \equiv P(C_i(t, G_i) = c_{kl}, \mathbf{1}'G_i = d | \widetilde{G}_i, t; p, q) \tag{20}$$

$$= \begin{cases} P(t'G_i = 0, \mathbf{1}'G_i = d | t, \widetilde{G}_i; p, q), & l = 0 \\ P(t'G_i > 0, \mathbf{1}'G_i = d | t, \widetilde{G}_i; p, q), & l = 1 \end{cases} \tag{21}$$

$$= \begin{cases} P(t'G_i = 0 | t, \widetilde{G}_i; p, q) P((1 - t)'G_i = d | t, \widetilde{G}_i; p, q), & l = 0 \\ \sum_{d_t=1}^{d} P(t'G_i = d_t | t, \widetilde{G}_i; p, q) P((1 - t)'G_i = d - d_t | t, \widetilde{G}_i; p, q), & l = 1 \end{cases} \tag{22}$$

Equation (22) defines a distribution over the true, unobserved exposure condition and in-degree, conditional on the treatment vector and the number of observed treated and non-treated connections for individual i. When coupled with a parametric model $f(\cdot; \theta_c, d)$ (see 7) for the potential outcomes under each (true) exposure condition c and in-degree d, we can model the observed outcome y_i as arising from a mixture of the $f(\cdot; \theta_c, d)$ with weights corresponding to the probabilities $\tau_i(c_{kl}, d; p, q)$ over the unobserved quantities.[8] The log-likelihood of the parameters $\Theta \equiv \{\theta_{00}, \theta_{01}, \theta_{10}, \theta_{11}, p, q\}$ given y is

$$l(\Theta) = \sum_{i=1}^{N} \log \left[\sum_c \sum_{d=0}^{N-1} \tau_i(c, d; p, q) f(y_i; \theta_c, d) \right]. \tag{23}$$

Estimation of the parameters Θ are provided using maximum likelihood estimation via the Expectation-Maximization algorithm, details of which are provided in Section 3.3. Note

---

[7]Via simulations, we find that a good choice of μ, which governs the overall density of the true network, is more important for our model to recover unbiased estimates.

[8]One downside of the Horvitz-Thompson estimator (6) is that it does not model individual potential outcomes and thus is less amenable to likelihood-based approaches.

that likelihood estimation is only justified if the observed outcomes are representative of the potential outcomes under each exposure condition, conditional on the true in-degrees. In general, it is sufficient to have a random treatment mechanism, such as randomly choosing n out of N subjects to treat[9].

Provided an estimate of the model parameters $\widehat{\Theta}$, estimating the mean outcome under exposure condition c (recall equation (5)) is straightforward and given by the expectation of the potential outcome under exposure c for each subject averaged across the population. We estimate μ_c with the following plug-in estimator:

$$\widehat{\mu}_c = \frac{1}{N} \sum_{i=1}^{N} E\left[Y_i(C_i = c)|y, t, \widetilde{G}\right] \tag{24}$$

$$= \frac{1}{N} \sum_{i=1}^{N} \sum_d P(\mathbf{1}'G = d|y, t, \tilde{G}) E_{f\left(\cdot; \widehat{\theta}_c, d\right)}[y_i] \tag{25}$$

$$= \frac{1}{N} \sum_{i=1}^{N} \sum_d \left(\sum_{c'} \tau_i(c', d; \widehat{p}, \widehat{q}) \right) E_{f\left(\cdot; \widehat{\theta}_c, d\right)}[y_i]. \tag{26}$$

3.2   Identification

Before we discuss estimation strategies for our mixture model (23), we will (partially) characterize the conditions under which this model is identifiable. Without model identifiability, estimation may be unstable and parameters estimates uninterpretable. In this section, we assume $f\left(\cdot; \theta_c, d\right)$ arise from a common univariate family of distributions parameterized by $\eta \equiv \eta(\theta_c, d)$.

In general, mixture models are trivially unidentifiable since relabeling components yields different parameterizations of a model with the same marginal distribution (see Chapter 1.5 of McLachlan and Basford (1988)). This identifiability issue is of particular concern in our setting, where the labeling of the components is inherently meaningful; for example, being unable to disentangle clusters corresponding to no treatment and indirect treatment would

---

[9]Stratified sampling based on known covariates could also be addressed by directly introducing these covariates into the model.

leave us unable to estimate the direction of any indirect treatment effect. We are able to leverage the structure from our corruption model and the linear relationships between mixture components with the same exposure condition to prevent such relabeling from occurring.

Following (Frühwirth-Schnatter, 2006), we use "generic identifiability" to refer to identifiability problems not solved by permuting component labels. Generic identifiability holds for mixtures of Gaussians and many other univariate continuous distributions, with the major exceptions being the binomial and uniform distributions. For the binomial distribution, generic identifiability only holds under a lower bound on the number of trials/observations per subject, dependent on the number of components. See (Frühwirth-Schnatter, 2006) for a review of generic identifiability issues.

Unfortunately, generic identifiability of the mixture model (23) does not directly follow from the generic identifiability of the family f, as (Hennig, 2000) showed in the case of mixtures of linear regression models. For example, in a mixture of simple linear regressions with two distinct covariate values $0, 1$ and common variance σ^2, an equal mixture of $f(x) = x$ and $f(x) = 1 - x$ yields the same model as a equal mixture of $f(x) = 0$ and $f(x) = 1$. Observations from a third covariate value would yield generic identifiability. While not immediately applicable to our class of models since in-degree (our covariate) is also latent, (Hennig, 2000) and (Grün and Leisch, 2007) define conditions under which mixtures of linear and generalized linear models are generically identifiable.

Next, we explicitly prove the identifiability of our mixture model under the regression model (9) for f. Results are readily generalizable to other f that arise from generically identifiable families provided that distinct values of d would allow for the identification of our model parameters θ_c from the distribution parameters $\eta(\theta_c, d)$.

Proposition 1. *Let f be defined as in (9) and τ_i as in (22). Assume $p, q, p', q' < 1$[10] and*

---

[10]Both of these edge cases are relatively uninteresting, as when $p = 1$ all true edges are not observed and when $q = 1$ all non-edges are falsely observed.

that indirect exposure has some effect (i.e. $\theta_{00} \neq \theta_{01}$ and $\theta_{10} \neq \theta_{11}$). Then

$$\sum_c \sum_{d=0}^{N-1} \tau_i(c, d; p, q) f(y_i; \theta_c, d) = \sum_c \sum_{d=0}^{N-1} \tau_i(c, d; p', q') f(y_i; \theta'_c, d) \qquad (27)$$

for all given $t'\widetilde{G}_i = \tilde{d}_t$ and $(\mathbf{1}-t)'\widetilde{G}_i = \tilde{d}_{nt}$ implies $\{\theta_{00}, \theta_{01}, \theta_{10}, \theta_{11}\} = \{\theta'_{00}, \theta'_{01}, \theta'_{10}, \theta'_{11}\}$ as long as there exists two distinct d such that we have subjects under each direct treatment status with observed degree $\tilde{d}_t + \tilde{d}_{nt} = d$, and, of these subjects, some have treated connections $\tilde{d}_t > 0$ while others do not, with $\tilde{d}_t = 0$.

Proof. It suffices to show identifiability of $\{\theta_{00}, \theta_{01}\} = \{\theta'_{00}, \theta'_{01}\}$, since we assume direct treatment status can always be accurately ascertained. The exposure conditions $\{c_{00}, c_{01}\}$ are only corrupted with one another, as are $\{c_{10}, c_{11}\}$.

Let us begin with the most general case, when both $\{p, q\} \in (0, 1)$. In this situation, the probabilities τ_i are positive over all feasible true exposure conditions and degrees, regardless of the pair of observed degrees $(\tilde{d}_t, \tilde{d}_{nt})$. The only restriction on the support of these probabilities are that, under no indirect treatment, degree cannot be larger than $N - 1 - 1't + t_i \equiv N_{nt,i}$ (otherwise there would have to exist a connection to a treated subject), and degree must be at least one for an individual to be indirectly treated. Mathematically, $\tau_i(c_{00}, d; p, q) > 0$ for any d satisfying $d \leq N_{nt,i}$ and $\tau_i(c_{01}, d; p, q) > 0$ for any $d \geq 1$.

At the other extreme, when there is no corruption ($p = 0$ and $q = 0$), then the true exposure condition and degree match their observed counterparts. Mathematically, $\tau_i(c, d; p, q) > 0$ only for $d = \tilde{d}_t + \tilde{d}_{nt}$ and either $c = c_{00}$ if $\tilde{d}_t = 0$ or $c = c_{01}$ if $\tilde{d}_t > 0$. When exactly one kind of corruption exists, the support of τ_i is limited, but to a lesser extent that when neither types of corruption exist. When $p > 0$ but $q = 0$, true edges can be dropped but all observed edges also exist in the true network. Namely, any observed connection to a treated subject must exist in the true graph. For subjects with at least one of these connections $\tilde{d}_t > 0$, the support of τ_i is limited to $c = c_{01}$ and $d \geq \tilde{d}_t + \tilde{d}_{nt}$. If instead we have $\tilde{d}_t = 0$, τ_i is positive for $\tilde{d}_{nt} \leq d \leq N_{nt,i}$ when $c = c_{00}$ and $d \geq \tilde{d}_{nt} + 1$ when $c = c_{01}$.

20

Lastly, when $q > 0$ and $p = 0$, the observed connections is a superset of the links in the true graph. Thus, when we observe no connections to treated subjects $\tilde{d}_t = 0$, τ_i is only positive for $c = c_{00}$ and $d \leq \tilde{d}_{nt}$. When such an connection is observed, τ_i is positive for $c = c_{00}$ and $d \leq \tilde{d}_t + \tilde{d}_{nt} - 1$ or $c = c_{01}$ and $1 \leq d \leq \tilde{d}_t + \tilde{d}_{nt}$.

Case 1: $p > 0$, $q > 0$, and $\beta_{10} \neq 0$

For any pair of $\left(\tilde{d}_t, \tilde{d}_{nt} \right)$, the LHS is a mixture of normal distributions that includes $N-1$ distinct components with means $\alpha_{01} + \beta_{01}d$ and variance σ^2 for any d from $\{1, ..., N-1\}$. There are at most $N_{nt,i} + 1 < N - 1$ other mixture components corresponding to the c_{00} terms. Following the generic identifiability of finite normal mixtures, the same component normals must exist on the RHS, with the same weights. For there to be at least $N-1$ distinct components on the RHS for both $\tilde{d}_t = 0$ and $\tilde{d}_t > 0$, we must have $p' > 0$ and $q' > 0$. On the LHS, we have $N - 1$ components which are evenly spaced $|\beta_{01}|$ apart, while on the RHS we have $N - 1$ components evenly spaced $|\beta'_{01}|$ apart. Since there are fewer than $N - 1$ other components on either side, these $N - 1$ components must match, with $|\beta_{01}| = |\beta'_{01}|$. This leads to two possibilities: we must have either $\alpha'_{01} = \alpha_{01}$ and $\beta'_{01} = \beta_{01}$ or $\alpha'_{01} = \alpha_{01} + N\beta_{01}$ and $\beta'_{01} = -\beta_{01}$. The latter cannot occur due to would-be inconsistencies in the weights. For example, consider weights for the component with mean $\alpha_{01} + \beta_{01}$ under this scenario. On the LHS, the weight would correspond to the probability $\tau_i(c_{01}, 1; p, q)$, while on the RHS, the weight would correspond to the probability $\tau_i(c_{01}, N - 1; p', q')$. The former quantity changes with $\tilde{d}_t$ if holding the total observed degree $\tilde{d}_t + \tilde{d}_{nt}$ fixed, since the observed treated degree would affect the probability of a true treated connection, but the latter does not since for very large true degree $d > N_{nt,i}$ we will always have a treated connection. Thus, we have $\alpha'_{01} = \alpha_{01}$ and $\beta'_{01} = \beta_{01}$.

We can then use our identification of the c_{01} components to isolate the remaining, un-explained components, which must correspond to c_{00}. If $\beta_{00} \neq 0$, the LHS has $N_{nt,i} + 1$ remaining components, while if $\beta_{00} = 0$, the LHS has one component. The same holds for β'_{00} and the RHS. Thus, when $\beta_{00} = 0$, $\beta'_{00} = 0$ and we must have $\alpha'_{00} = \alpha_{00}$. On the other

21

hand, if $\beta_{00} \neq 0$, both sides consist of $N_{nt,i} + 1$ components, spaced $|\beta_{00}|$ and $|\beta'_{00}|$ apart respectively. We must have either $\alpha'_{00} = \alpha_{00}$ and $\beta'_{00} = \beta_{00}$ or $\alpha'_{00} = \alpha_{00} + N_{nt,i}\beta_{00}$ and $\beta'_{00} = -\beta_{00}$. Following similar logic as above for the c_{01} components, we can use would-be inconsistencies in the weights to eliminate the second scenario. Namely, consider the weights for the α_{00} component, which is $\tau_i(c_{00}, 0; p, q)$ for the LHS and $\tau_i(c_{00}, N_{nt,i}; p', q')$ for the RHS. For fixed $\tilde{d}_t + \tilde{d}_{nt}$, $\tau_i(c_{00}, 0; p, q)$ is unaffected by varying $\tilde{d}_t$ as all observed connections regardless of treatment status must be falsely observed, while the treatment status of the observed connections will effect the probability of having a treated connection given $N_{nt,i}$ true connections.

Case 2: $p > 0$, $q > 0$, and $\beta_{10} = 0$

Next, let us consider the scenario when we have $\beta_{10} = 0$, but $\beta_{00} \neq 0$. For any pair of $\left(\tilde{d}_t, \tilde{d}_{nt}\right)$, the LHS is a normal mixture including $N_{nt,i} + 1$ or $N_{nt,i} + 2$ components with means α_{01} and $\alpha_{00} + \beta_{00}d$ and variance σ^2 for any d from $\{0, ..., N_{nt,i}\}$. Since the number of components does not change for any pair of observed degrees, we have $p' > 0$, $q' > 0$, and $\beta'_{10} = 0$. Following the same logic used in case 1 but reversing the order in which we consider the c_{00} and c_{01} components, we can show $\{\theta_{00}, \theta_{01}\} = \{\theta'_{00}, \theta'_{01}\}$.

The alternate scenario involves the case $\beta_{00} = 0$ and $\beta_{01} = 0$. Since we assume there is a non-zero indirect treatment effect $(\theta_{00} \neq \theta_{01})$, the LHS consists of a mixture of two normals with means α_{00} and α_{01}. Following the generic identifiability of normal mixtures, the RHS must consist of two normals with the same means. In order for the RHS to have two mixture components regardless of observed degree $\left(\tilde{d}_t, \tilde{d}_{nt}\right)$, we must have $\beta'_{00} = 0$ and $\beta'_{01} = 0$ as well as non-zero corruption in both p' and q'. For $q' = 0$, $\tilde{d}_t > 0$ would yield just one mixture component, and similarly with $\tilde{d}_t = 0$ for $p' = 0$. Thus, either $\alpha'_{00} = \alpha_{00}$ and $\alpha'_{01} = \alpha_{01}$ or $\alpha'_{00} = \alpha_{01}$ and $\alpha'_{01} = \alpha_{00}$. If the latter is the case, the weight of the α_{00} component is the probability of no indirect treatment $\sum_d \tau_i(c_{00}, d; p, q)$ on the LHS and the probability of indirect treatment $\sum_d \tau_i(c_{01}, d; p', q')$ on the RHS. These weights must be the same for any pair of $\left(\tilde{d}_t, \tilde{d}_{nt}\right)$. However, when holding $\tilde{d}_t + \tilde{d}_{nt}$ fixed and increasing the number of

observed connections to treated individuals $\tilde{d}_t$, the weight of the LHS decreases while the weight of the RHS increase. Thus, we must have $\alpha'_{00} = \alpha_{00}$ and $\alpha'_{01} = \alpha_{01}$.

Case 3: $p > 0$ and $q = 0$

First, consider an observation i with at least one observed connection to a treated subject $\tilde{d}_t > 0$. The mixture on the LHS consists of components with means $\alpha_{01} + \beta_{01}d$ corresponding to $\tau_i(c_{01}, d; p, q)$ for any d satisfying $d \geq \tilde{d}_t + \tilde{d}_{nt}$. If $\beta_{01} = 0$, the LHS will just be one component, while if $\beta_{01} \neq 0$, the LHS will have $N - (\tilde{d}_t + \tilde{d}_{nt})$ distinct components.

Suppose for now the latter is true. Then increasing total observed degree $\tilde{d}_t + \tilde{d}_{nt}$ decreases the number of components on the LHS. Changing total observed degree has no effect on the number of distinct components when $p, q > 0$ or $p = q = 0$, while the case $p = 0$ and $q > 0$ would imply an increase in the number of distinct components. Thus, to match the behavior on the RHS, we must have $p' > 0$ and $q' = 0$. For the components on both sides to have the same set of means, we must have either $\alpha'_{01} = \alpha_{01}$ and $\beta'_{01} = \beta_{01}$ or $\alpha'_{01} = \alpha_{01} + (N - 1 + \tilde{d}_t + \tilde{d}_{nt})\beta_{01})$ and $\beta'_{01} = -\beta_{01}$. We can again invalidate the second case by examining would-be inconsistencies in the weights τ_i, but in this case we can also simply note that the latter scenario cannot be simultaneously valid across multiple choices of $\tilde{d}_t + \tilde{d}_{nt}$. Having established $\alpha'_{01} = \alpha_{01}$ and $\beta'_{01} = \beta_{01}$, we can consider observations with $\tilde{d}_t = 0$ and isolate the remaining $\tau_i(c_{00}, d; p, q)$ components on the LHS, of which there would be either 1 (if $\beta_{00} = 0$) or $N_{nt,i} - \tilde{d}_{nt} + 1$ (if $\beta_{00} \neq 0$) components. Matching these components on the RHS across multiples values of $\tilde{d}_{nt}$ will avoid the potential case where $\beta'_{00} = -\beta_{00}$ and yield $\alpha'_{00} = \alpha_{00}$ and $\beta'_{00} = \beta_{00}$.

Let us now return to the case where $\beta_{01} = 0$. While we could still find $\beta'_{01} = 0$ and $\alpha'_{01} = \alpha_{01}$, examining the number of components when $\tilde{d}_t > 0$ is not sufficient to imply $p' > 0$ and $q' = 0$. However, we can attempt to ascertain whether or not this must be the scenario by examining observations with $\tilde{d}_t = 0$. If $\beta_{00} \neq 0$, there would be $N_{nt,i} - \tilde{d}_{nt} + 1$ distinct components on the LHS. A decreasing number of components for these observations as $\tilde{d}_{nt}$ increase is only consistent with $p' > 0$ and $q' = 0$. From here, we can use the

equal spacing of these components as well as the structure imposed by the weights to show $\alpha'_{00} = \alpha_{00}$ and $\beta'_{00} = \beta_{00}$.

Lastly, when both $\beta_{00} = 0$ and $\beta_{01} = 0$, we observe one mixture component with mean α_{01} when $\tilde{d}_t > 0$ and two mixture components with means α_{00} and α_{01} when $\tilde{d}_t = 0$. Returning to the logic used in the counterpart scenario in case 2, the LHS can only be matched when $p' > 0$ and $q' = 0$. Then the RHS will have one mixture component when $\tilde{d}_t > 0$ and two components when $\tilde{d}_t == 0$, and we will have $\{\theta_{00}, \theta_{01}\} = \{\theta'_{00}, \theta'_{01}\}$.

Case 4: $p = 0$ and $q > 0$

This case follows identical logic as case 3 but switching the roles of the c_{00} and c_{01} components. Namely, observations with $\tilde{d}_t = 0$ will isolate the c_{00} components, which in turn can be used to inform observations with $\tilde{d}_t > 0$ to match the c_{01} components.

Case 5: $p = 0$ and $q = 0$

For any pair of $\left(\tilde{d}_t, \tilde{d}_{nt}\right)$, the LHS will consist of consist of a single normal distribution. If $p' > 0$ or $q' > 0$, this behavior could only arise if $\beta_{00} = \beta_{01} = 0$ and $\alpha_{00} = \alpha_{01}$. However, we require $\theta'_{00} \neq \theta'_{11}$, so we must have $p' = 0$ and $q' = 0$. Observations from two distinct values of $\tilde{d}_t + \tilde{d}_{nt}$ for each of $\tilde{d}_t = 0$ and $\tilde{d}_t > 0$ will uniquely identify the model parameters θ_{00} and θ_{01} respectively.

$\square$

3.3   Estimation

Maximizing the log-likelihood (23) with respect to the parameters Θ cannot be done in closed form due to the summations inside the logarithmic terms. However, if we had directly observed the latent variables $\{C, \mathbf{1}'G\}$, the log-likelihood of the parameters Θ given y, C and $\mathbf{1}'G$ would be given by

$$l(\Theta) = \sum_{i=1}^{N} \log\left[\tau_i(c, d; p, q) f\left(y_i; \theta_c, d\right)\right]. \tag{28}$$

This would be substantially easier to work with, due to the lack of summation inside the logarithmic terms. The EM algorithm (Dempster *et al.*, 1977) is a well-established technique for maximum likelihood estimation in the presence of latent variables that leverages this disparity between the two log-likelihood expressions. Given some set of initial parameter values $\widehat{\Theta}^0$, the algorithm alternates between estimating posterior distribution of latent variables for each subject given the current parameter values (E-step) and updating the parameter values given these posterior probabilities (M-step). Explicitly working with the latent variables in the M-step yields simpler maximization problems. Each iteration of the algorithm increases the log-likelihood, leading to a local optima, and the algorithm is run from multiple initialization values in order to achieve a higher quality final estimate $\widehat{\Theta}_{EM}$.

Suppose at iteration t we have parameter estimates $\widehat{\Theta}^{(t)}$. In the E-step, we compute the posterior probabilities over the latent variables using the current parameter estimates. These probabilities, or "responsibilities," are given by

$$\widehat{\gamma}_{icd}^{(t+1)} \equiv P\left(C_i(t, G_i) = c, \mathbf{1}'G_i = d | y_i, t, \widetilde{G}_i; \widehat{\Theta}^{(t)}\right) \tag{29}$$

$$\propto \tau_i^{(t)}(c, d; \widehat{p}^{(t)}, \widehat{q}^{(t)}) f\left(y_i; \widehat{\theta}_c^{(t)}, d\right). \tag{30}$$

In the M-step, we use these responsibilities to maximize the expectation of the complete likelihood 28 under these posterior probabilities

$$\widehat{\Theta}^{(t+1)} = \underset{\Theta}{\operatorname{argmax}}\ E_{C, \mathbf{1}'G | y, t, \widetilde{G}, \widehat{\Theta}^{(t)}} \left[\sum_{i=1}^{N} \log\left[\tau_i(c, d; p, q) f\left(y_i; \theta_c, d\right)\right] \right] \tag{31}$$

$$= \underset{\Theta}{\operatorname{argmax}}\ \sum_{i=1}^{N} \sum_{c} \sum_{d=0}^{N-1} \widehat{\gamma}_{icd}^{(t+1)} \log\left[\tau_i(c, d; p, q) f\left(y_i; \theta_c, d\right)\right]. \tag{32}$$

For example, under the linear model 9, we can compute closed form updates for the regression

parameters:

$$\widehat{\beta}_c^{(t+1)} = \frac{\overline{dy}_c - \overline{d}_c\overline{y}_c}{\overline{d^2}_c - \overline{d}_c^2} \tag{33}$$

$$\widehat{\alpha}_c^{(t+1)} = \overline{y}_c - \widehat{\beta}_c^{(t+1)}\overline{d}_c \tag{34}$$

$$\widehat{\sigma}^{2(t+1)} = \frac{1}{N}\sum_{i=1}^{N}\sum_{c,d}\left[\widehat{\gamma}_{icd}^{(t+1)}\left(y - \widehat{\alpha}_c^{(t+1)} - \widehat{\beta}_c^{(t+1)}d\right)^2\right] \tag{35}$$

where $\overline{d}_c = \frac{\sum_{i=1}^{N}\widehat{\gamma}_{icd}^{(t+1)}d}{\sum_{i=1}^{N}\widehat{\gamma}_{icd}^{(t+1)}}$, $\overline{y}_c = \frac{\sum_{i=1}^{N}\widehat{\gamma}_{icd}^{(t+1)}y_i}{\sum_{i=1}^{N}\widehat{\gamma}_{icd}^{(t+1)}}$, $\overline{d^2}_c = \frac{\sum_{i=1}^{N}\widehat{\gamma}_{icd}^{(t+1)}d^2}{\sum_{i=1}^{N}\widehat{\gamma}_{icd}^{(t+1)}}$, and $\overline{dy}_c = \frac{\sum_{i=1}^{N}\widehat{\gamma}_{icd}^{(t+1)}dy_i}{\sum_{i=1}^{N}\widehat{\gamma}_{icd}^{(t+1)}}$. Note the similarly of our linear model estimators to those obtained from weighted least squares. Updates for the corruption parameters p and q cannot be computed in closed form but can be solved for using a general optimizer such as `optim` in R. Estimates of the mean outcomes under each exposure condition (26) are functions of the model parameters and can be calculated accordingly.

Given the likelihood (28) is bounded and satisfies mild smoothness conditions, as well as sufficiently many runs of the EM algorithm, we should be able to find the global optima and obtain the MLE $\widehat{\Theta}_{EM} = \widehat{\Theta}_{MLE}$ (McLachlan and Basford, 1988). We can consider the consistency of $\widehat{\Theta}_{MLE}$ under the scenario we had access to comparable experiments on many networks, and that the outcomes for each experiment arise from the mixture model (23) with the same set of parameters Θ^*. The major condition from Wald (1949) needed to ensure consistency of $\widehat{\Theta}_{MLE}$ is the identifiability of the mixture model on a non-zero probability set of the subjects in these experiments. In the case of our linear model for the potential outcomes (9), consistency requires regular variation in the observed in-degrees and exposure conditions. Building off of Proposition 1, a sufficient condition for the consistency of $\widehat{\Theta}_{MLE}$ would be to observe infinitely many subjects with at least two distinct observed in-degrees $\tilde{d}$ under each observed exposure condition.

While the EM algorithm does not provide standard errors and confidence intervals for our estimate $\widehat{\Theta}_{EM}$ and functions thereof (such as the mean outcome estimates), bootstrap meth-

ods have been used in the context of other mixture models to approximate these quantities (Basford *et al.*, 1997). In particular, we consider the parametric bootstrap, which consists of the following steps:

1. Generate samples $\{y^{(1)}, y^{(2)}, .., y^{(m)}\}$ from the fitted model given by (23) with parameters $\widehat{\Theta}_{EM}$, holding the treatment assignment vector t and observed network $\widetilde{G}$ fixed.

2. Estimate m sets of parameters $\{\widehat{\Theta}_{EM}^{(1)}, \widehat{\Theta}_{EM}^{(2)}, .., \widehat{\Theta}_{EM}^{(m)}\}$ using the EM algorithm on the generated samples $\{y^{(1)}, y^{(2)}, .., y^{(m)}\}$.

3. Calculate Monte-Carlo estimates of the standard errors and/or confidence intervals for $\widehat{\Theta}_{EM}$ using the parameters estimates $\{\widehat{\Theta}_{EM}^{(1)}, \widehat{\Theta}_{EM}^{(2)}, .., \widehat{\Theta}_{EM}^{(m)}\}$.

3.4 Simulation Study

In this section, we apply our mixture model to simulated experiments run over the households of 75 Indian villages, collected by and described in detail in (Banerjee *et al.*, 2013). Within each village, we consider an experiment with two treatment levels ("treatment" and "control") and treatment interference on the outcome of interest through the household network of borrowing and lending money.

The 75 villages range in size from 77 to 356 households, with an average of slightly under 200. We take the true spillover network in each village to be the reported network of borrowing and lending, with a link between two households if there is any monetary borrowing or lending between the two. While these networks may themselves be mismeasurements of the actual borrowing/lending networks for each village, they exhibit useful properties one may expect in networks, such as small world phenomenon, significant clustering, and substantial variation in degrees. When excluding households with zero degree, the average number of households in a village is about 170.

We run 10 simulated experiments on each village, each of which consists of the following steps:

1. Independently assign each household to treatment with probability 0.25.

2. Calculate the exposure condition C_i for each household i using the true network G.

3. Generate outcome of interest y_i for each household according to (9) with parameters $\alpha = (0, 0.25, 0.5, 1)$, $\beta = (0.05, 0.1, 0.05, 0.1)$, and $\sigma^2 = 0.25$.

4. Observe a corrupted network $\widetilde{G}$, where each edge in the true network is observed independently with probability $1 - p$, while false edges in the network are observed independently with probability qd, where d is the density of true graph G[11]. Simulations are repeated for every pair of corruption probabilities (p, q) with $p \in \{0, \frac{1}{8}, \frac{1}{4}, \frac{3}{8}, \frac{1}{2}\}$ and $q \in \{0, \frac{1}{8}, \frac{1}{4}, \frac{3}{8}, \frac{1}{2}\}$.

Across the 10 simulations and 75 villages, we average 45 households with no exposure to treatment (under the true network), 82 households with indirect exposure to treatment, 15 households with direct exposure to treatment, and 28 households with full exposure to treatment. The challenge is to derive estimates of the mean outcome under each exposure condition (5) despite observing a corrupted network $\widetilde{G}$ and thus corrupted exposure conditions. Some of our exposure conditions can be quite challenging; for example, when $p = q = 0.5$, on average half of the true edges are not observed, but instead are "replaced" with roughly the same number of false edges.

Estimation proceeds as described in Section 3.3. We choose the parameters of the beta-binomial distribution over the true degrees (recall the discussion following equations 18 and 19) by taking μ to be the density of the true network and choosing dispersion parameter ρ such that the second moment of the beta-binomial distributions matches that of the the observed degree distribution. This simulation scenario is consistent with a setting where we have *a priori* expectations on the density of the true network of spillovers. In Figure 6 located in Appendix A, we present results where μ and ρ are chosen solely by matching

---

[11]This specification for the corruption parameter governing the observance of false edges is easier to compare across networks with varying densities. $q = 0.5$ implies observing a number of false edges equal to about half the actual number of true edges.

the first two moments of the observed degree distribution. Our results are slightly worse under the purely empirical specification but still represent a substantial improvement over the Horvitz-Thompson approach not accounting for any corruption.

In Figures 2 and 3, we present heat maps to compare the estimates obtained by the EM algorithm to the Horvitz-Thompson estimates that do not take into account missingness in the network. Positive deviations from the true mean outcomes are denoted in red while negative deviations are denoted in blue, with the intensity of the colors corresponding to the size of the deviation. The estimates of mean outcome given by the EM algorithm are quite reasonable across the varying levels of corruption p and q considered, and represent a substantial improvement over the Horvitz-Thompson estimates. Differences in performance across the various corruption levels are much more muted, at least across the different levels of corruption considered in our simulations.

To examine these results in more detail, we focus on the two cases presented in Figures 4 and 5, where we fix $p = 0.5$ and vary q and fix $q = 0.5$ and vary p respectively. Estimates from our method are presented in cyan while estimates from the Horvitz-Thompson are presented in red. In general, estimates from our method exhibit considerably less bias and are simultaneously have less variance. We find that our method has slightly higher levels of bias and variance for higher levels of corruption, which is consistent with the idea that for higher p and q there is larger uncertainty over the true network (22) and thus our results are more dependent on the assumed beta-binomial model over the true degree distribution. Even if μ and ρ are chosen to match the true degree distribution, the beta-binomial model still represents a (higher-order) deviation from the true degree distribution for real-life networks. The direction of the bias, upwards for $\widehat{\mu}_0$ and $\widehat{\mu}_D$ towards μ_I and μ_F respectively and the reverse for $\widehat{\mu}_I$ and $\widehat{\mu}_F$, are a product of imperfectly learning the latent exposure conditions, especially in these higher uncertainty settings.
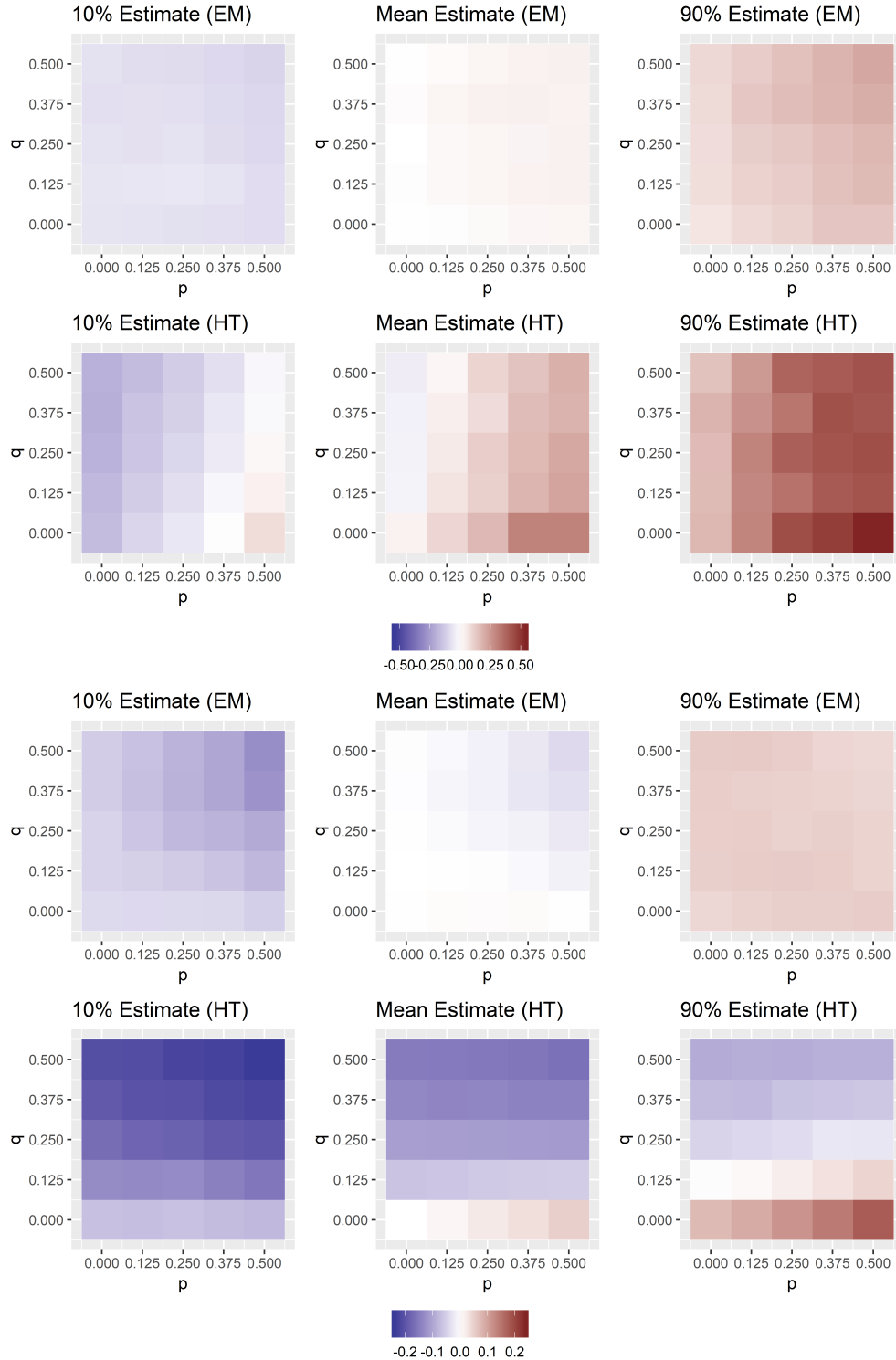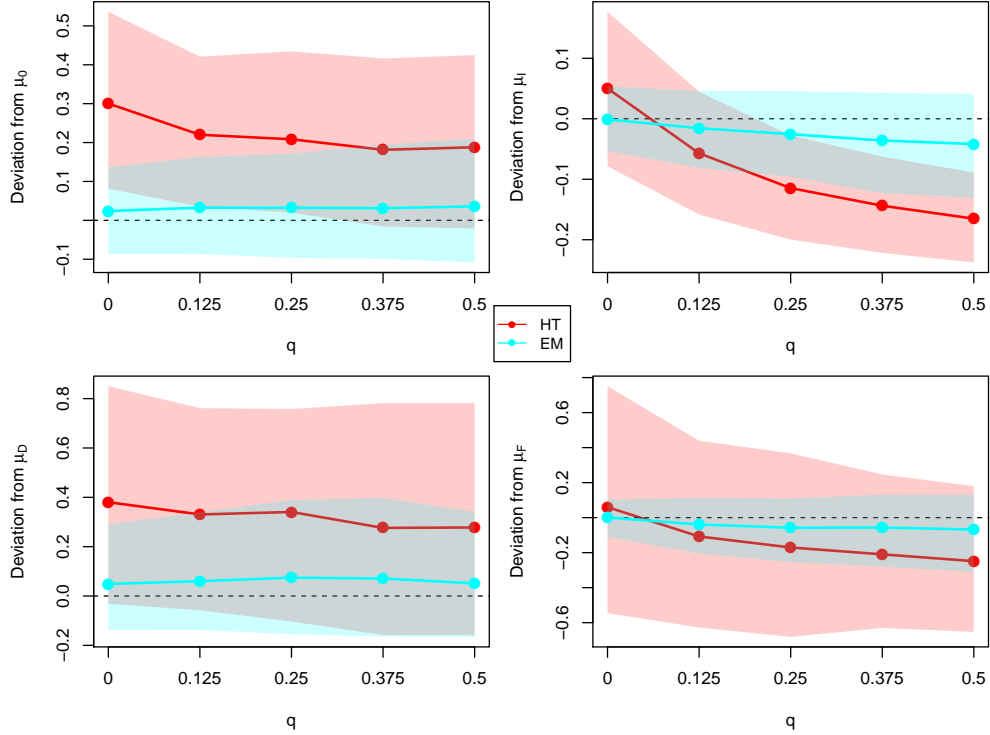
Figure 2: *Estimates of the mean outcome of the no exposure (top) and indirect exposure (bottom) conditions from their true values under varying corruption levels (p,q) for the network. Estimates obtained from our model using the EM algorithm are compared against estimates from the Horvitz-Thompson estimators assuming no corruption in the network. The 0.1 and 0.9 quantiles are provided for both methods to give a sense of the variability in these estimates. Note the color gradient scales are different for the two exposure conditions.*
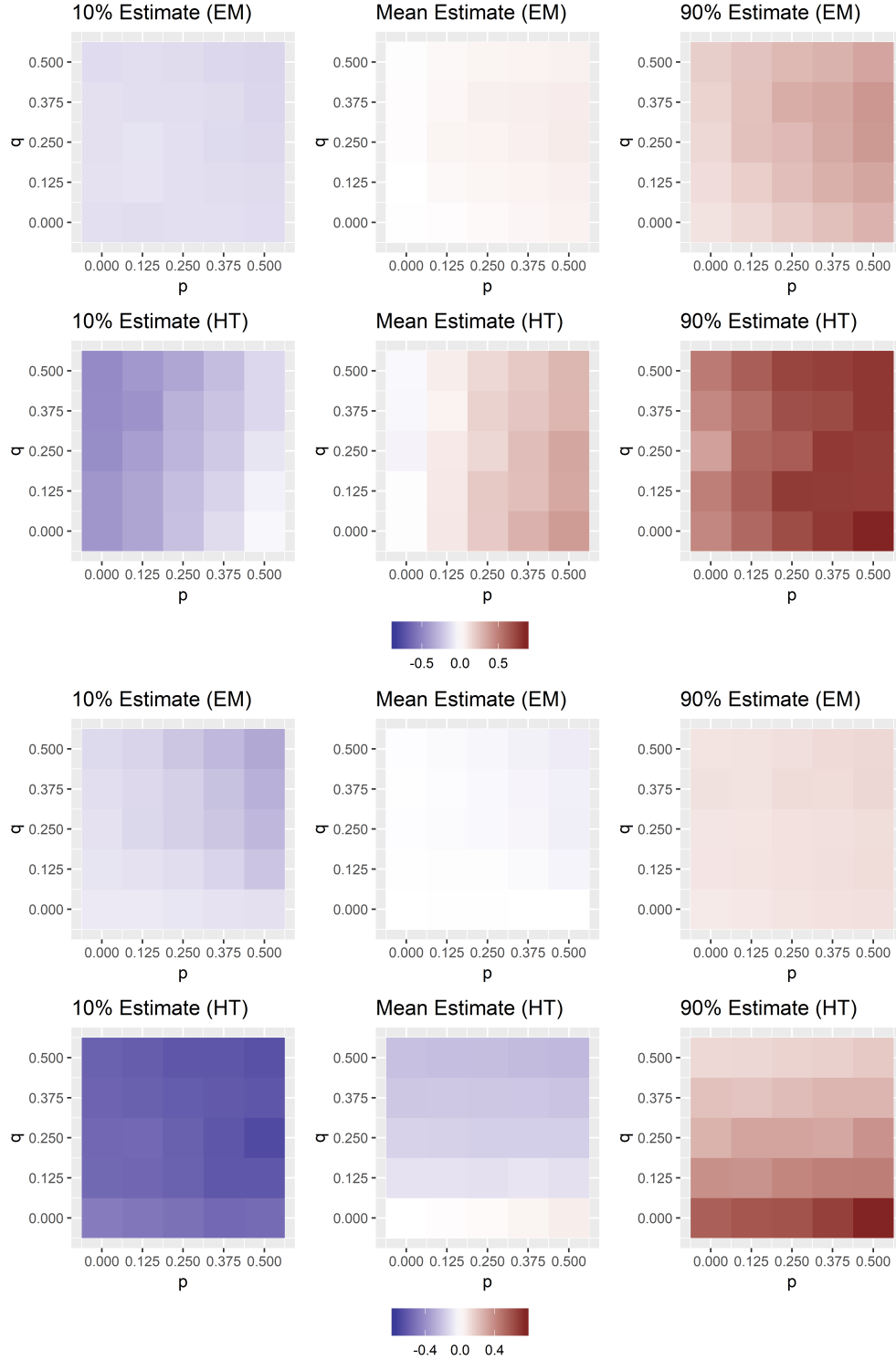
Figure 3: *Estimates of the deviation of the mean outcome of the direct exposure (top) and full exposure (bottom) conditions from their true values under varying corruption levels (p,q) for the network. Estimates obtained from our model using the EM algorithm are compared against estimates from the Horvitz-Thompson estimators assuming no corruption in the network. The 0.1 and 0.9 quantiles are provided for both methods to give a sense of the variability in these estimates. Note the color gradient scales are different for the two exposure conditions.*
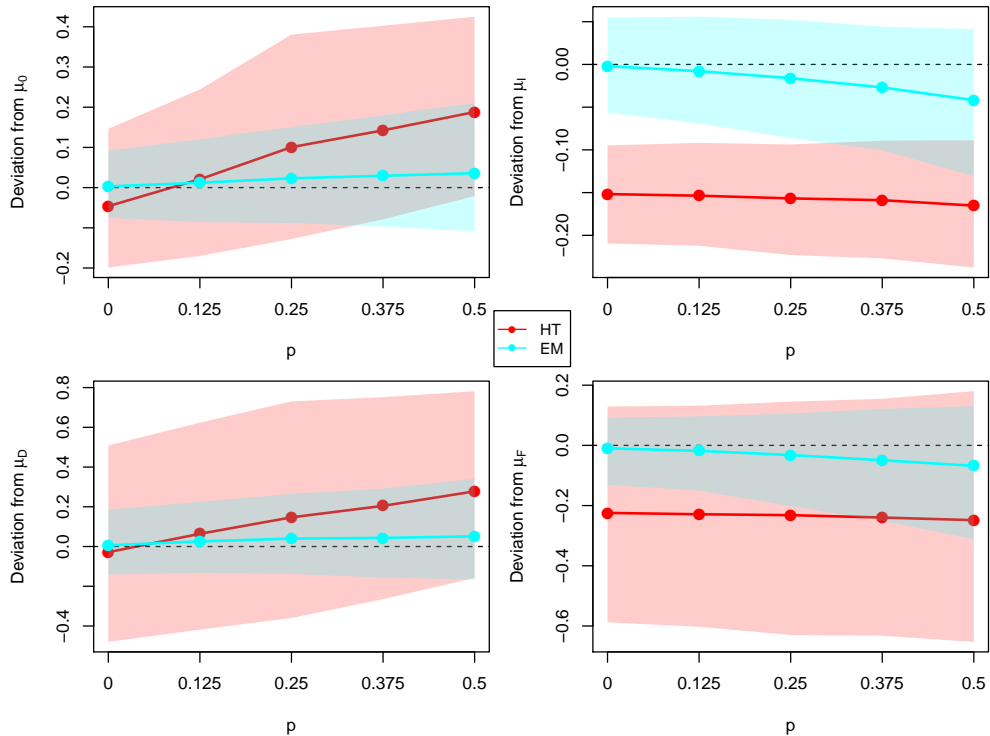
Figure 4: *Estimates of the deviation of the mean outcome of each exposure conditions (top left: no exposure, top right: indirect exposure, bottom left: direct exposure, bottom right: full exposure) from their true values under $p = 0.5$ and varying q from 0 to 0.5. Estimates obtained from our model using the EM algorithm are compared against estimates from the Horvitz-Thompson estimators assuming no corruption in the network. The 0.1 and 0.9 quantiles are shaded for both methods to give a sense of the variability in these estimates.*

4   Diffusion of insurance information between farmers

Cai *et al.* (2015) study the adoption decisions of rice farmers in rural China in regards to weather insurance, which typically exhibit low take-up rates even in the presence of heavy government subsidies. The authors were interested in how difficulties communicating the benefits of the product may be modulated by the dissemination of information via a farmer's peers. In conjunction with the introduction of a new weather insurance product, researchers randomized about 5000 households across 185 rural villages into information sessions about the new insurance product. Sessions were held in two rounds three days apart, and could either be "simple" sessions just describing the product or longer "intensive" sessions which also emphasized the expected benefits from insurance. Drop out was not a major issue in

Figure 5: *Estimates of the deviation of the mean outcome of each exposure conditions (top left: no exposure, top right: indirect exposure, bottom left: direct exposure, bottom right: full exposure) from their true values under $q = 0.5$ and varying p from 0 to 0.5. Estimates obtained from our model using the EM algorithm are compared against estimates from the Horvitz-Thompson estimators assuming no corruption in the network. The 0.1 and 0.9 quantiles are shaded for both methods to give a sense of the variability in these estimates.*

this experiment, with an overall session attendance rate of about 90%.

One specific question the authors were interested in was how insurance take-up and knowledge for households assigned to second round sessions were affected by whether or not they had friends assigned to first round intensive sessions. Before the experiment, each household was asked to list five friends whom they most frequently discussed production or financial issues with. About 96% list 5 friends included in the study, with the rest listing between one to four friends. In general, prompting respondents to list five friends can censor the number of connections for individuals with high in-degree, as well as cause the reported network to contain some weaker connections that would otherwise be unreported. However, this concern may be relatively mild in this case, as the authors conducted a pilot study in

two villages where the number of friends was uncensored and found 96% of farmers reported either four or five connections. Most of the paper's results use this reported network, which, borrowing their language, we will term "general network measure."

Since insurance take-up is a binary outcome measure, which our method cannot handle due to the unidentifiability of mixtures of Bernoulli variables, we focus on the network effect on insurance knowledge. To measure insurance knowledge, each household completed a five question test after the experiment and were scored from 0-5. The authors found that, for households assigned to second round sessions, having a friend who was assigned a first round intensive session had the same (statistically significant) benefit for insurance knowledge as personally being assigned an intensive session in the second round. Furthermore, households who had a friend who assigned to a first round intensive session derived no additional benefit in terms of insurance knowledge from being personally assigned to an intensive session.

First, we seek to replicate their finding with the following simple generative model:

$$\text{score} \sim \text{Bin}(5, \text{expit}(\alpha_c + \beta_c d)) \tag{36}$$

We model each household's score on the insurance test as arising from a binomial with five independent questions and a probability of getting a question right depending on the household's treatment exposure condition as well as their degree in the network. To produce comparable estimates to the linear specification (2) presented in Table 5 of their paper, we estimate the mean outcome under each exposure condition and calculate various contrasts using these means. The effect of personally being invited to a intensive session can be calculated as $\widehat{\mu}_D - \widehat{\mu}_0$, the effect of having a friend invited to a first round intensive session (denoted in our table as "Network Intensive") is calculated as $\widehat{\mu}_I - \widehat{\mu}_0$, and the interaction of these effects is given by $\widehat{\mu}_0 + \widehat{\mu}_F - \widehat{\mu}_I - \widehat{\mu}_D$. We present the results under the assumption the general network measure accurately characterizes the network of spillovers in Table 1, and compare those results to those obtained when allowing for potential corruption in the

reported network[12]. The two methods yield very similar estimates, as our method finds the outcomes consistent with a relatively low amount of corruption in the observed network.

Network measure	Intensive Session	Network Intensive	Interaction
No corruption			
General measure	0.205 (0.016)	0.198 (0.016)	-0.241 (0.023)
Strong measure	0.100 (0.012)	0.120 (0.036)	-0.188 (0.054)
Weak measure	0.157 (0.061)	0.072 (0.044)	-0.095 (0.063)
EM method			
General measure	0.177 (0.025)	0.229 (0.028)	-0.224 (0.040)
Strong measure	0.299 (0.037)	0.279 (0.040)	-0.577 (0.052)
Weak measure	0.160 (0.032)	0.259 (0.032)	-0.171 (0.048)
EM + covariates			
General measure	0.158 (0.035)	0.295 (0.039)	-0.322 (0.055)
Strong measure	0.177 (0.081)	0.305 (0.070)	-0.479 (0.116)
Weak measure	0.175 (0.095)	0.293 (0.067)	-0.155 (0.117)

Table 1: *Differences in insurance knowledge for farmers assigned to second round sessions based on (1) whether they attended an intensive session, (2) whether they had a friend attend a first round intensive session, and (3) the interaction of these two terms. We compare our method to results if we assume there is no corruption in the network for three network measures.*

Next, we consider replicating the results using alternative network specifications. In particular, we consider two other network measures mentioned in Cai *et al.* (2015). The authors define a "strong" network measure where non-reciprocal connections are dropped, as well as a "weak" network measure which adds second-order connections ("friends of friends") to the general network measure. Both specifications differ quite drastically from the general network measure and are indicative of low rates of reciprocity and transitivity in the reported network; farmers average a single connection under the strong network measure (with a mode of 0) and 16 connections under the weak network measure. We repeat the estimation of direct and indirect treatment effects using these alternative network measures both when

---

[12]The beta-binomial distribution over the true degrees is initialized with the same mean as the observed network (reflecting the results from the pilot study) and overdispersion parameter 0.0005. This overdispersion parameter was chosen based on examining the variation of the degrees in the Indian village data used in the simulations. Note the observed degree-distribution in farmers' network exhibits considerably less variation than even a binomial distribution, and thus is not particularly informative for choosing our prior distribution.

1) assuming the network measure correctly specifies the spillover pathways and 2) allowing for corruption in the network and report the results in Table 1. In contrast to the results assuming the network is correctly specified, our method gives pretty similar results under the general and weak measures, despite large differences in the networks being used. We perform quite differently for the strong measure, although perhaps this should be expected given the sparsity of network information in this case.

Finally, we consider an extension of our model (22) that allows for different levels of corruption in the connections between farmers depending on whether or not they reside in the same village. About 99.4% of reported connections are between farmers in the same village[13], while the remaining 0.6%, so separately modeling the true degree within-village and out-of-village may lead to more accurate results. We introduce distinct parameters for in-village degree (μ_{in} and ρ_{in}) and out-of-village degree (μ_{out} and ρ_{out}), along with respective corruption parameters p_{in}, q_{in}, p_{out}, and q_{out}. Note that there is a potential variance trade-off when introducing additional parameters to our model, so sample size concerns must also be considered. Estimation proceeds as described in Section 3.3, with the additional complication that the general purpose optimizer must maximize over all four corruption parameters at once. Estimates from this extension are largely similar to those obtained from our method ignoring the difference between in-village and out-of-village ties, perhaps due to the lack of out-of-village ties. However, the corresponding standard errors are substantially larger.

5   Discussion

Experimental inference on social networks present distinct challenges; not only are subjects' outcomes affected by the treatment assignments of other subjects, but this treatment interference is often of direct interest. Existing methodology for estimating treatment effects in this setting requires a precise measurement of the network of interest, which can be a

---

[13]There is substantial variation in the size of each village, which is entirely not reflected under the network measures considered

difficult assumption given the many decisions inherent in the data gathering process as well as imposing a large financial burden. In this paper, we present a class of mixture models that can accurately estimate treatment effects when the network of interest is not accurately measured, assuming the noise in the network is (conditionally) random and relying on additional assumptions about the parametric form for the treatment exposure conditions and the density of the true, latent network.

References

A. Adhvaryu, N. Kala, and A. Nyshadham. Management and shocks to worker productivity. *Mimeograph, University of Michigan Working Paper*, 2016.

M. Angelucci, G. De Giorgi, M. A. Rangel, and I. Rasul. Family networks and school enrolment: Evidence from a randomized social experiment. *Journal of public Economics*, 94(3-4):197–221, 2010.

P. M. Aronow and C. Samii. Estimating Average Causal Effects Under General Interference, with Application to a Social Network Experiment. *Annals of Applied Statistics*, 2017.

O. Bandiera, I. Barankay, and I. Rasul. Social connections and incentives in the workplace: Evidence from personnel data. *Econometrica*, 77(4):1047–1094, 2009.

O. Bandiera, I. Barankay, and I. Rasul. Social incentives in the workplace. *The review of economic studies*, 77(2):417–458, 2010.

A. Banerjee, A. G. Chandrasekhar, E. Duflo, and M. O. Jackson. The diffusion of microfinance. *Science*, 341(6144):1236498, 2013.

A. V. Banerjee, A. G. Chandrasekhar, E. Duflo, and M. O. Jackson. Changes in social network structure in response to exposure to formal credit markets. *Available at SSRN 3245656*, 2018.

K. E. Basford, D. R. Greenway, G. J. Mclachlan, and D. Peel. Standard Errors of Fitted Component Means of Normal Mixtures. *Computational Statistics*, 12(1):1–17, 1997.

P. Bayer, S. L. Ross, and G. Topa. Place of work and place of residence: Informal hiring networks and labor market outcomes. *Journal of Political Economy*, 116(6):1150–1196, 2008.

L. Beaman, A. BenYishay, J. Magruder, and A. M. Mobarak. Can network theory-based targeting increase technology adoption? Technical report, National Bureau of Economic Research, 2018.

L. A. Beaman. Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the us. *The Review of Economic Studies*, 79(1):128–161, 2011.

D. C. Bell, B. Belli-McQueen, and A. Haider. Partner Naming and Forgetting: Recall of Network Members. *Social Networks*, 29(2):279–299, 2007.

A. BenYishay and A. M. Mobarak. Social learning and incentives for experimentation and communication. *Review of Economic Studies, forthcoming*, 2018.

C. Blattman, D. Green, D. Ortega, and S. Tobón. Place-based interventions at scale: The direct and spillover effects of policing and city services on crime. Technical report, National Bureau of Economic Research, 2017.

N. Bloom, J. Liang, J. Roberts, and Z. J. Ying. Does working from home work? evidence from a chinese experiment. *The Quarterly Journal of Economics*, 130(1):165–218, 2014.

Y. Bramoullé, H. Djebbari, and B. Fortin. Identification of peer effects through social networks. *Journal of Econometrics*, 150(1):41–55, 2009.

E. Breza, A. G. Chandrasekhar, T. H. McCormick, and M. Pan. Using aggregated relational data to feasibly identify network structure without network data. Technical report, National Bureau of Economic Research, 2017.

L. Bursztyn, F. Ederer, B. Ferman, and N. Yuchtman. Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions. *Econometrica*, 82(4):1273–1301, 2014.

J. Cai, A. De Janvry, and E. Sadoulet. Social networks and the decision to insure. *American Economic Journal: Applied Economics*, 7(2):81–108, 2015.

D. Centola and M. Macy. Complex Contagions and the Weakenss of Long Ties. *American Journal of Sociology*, 113(3):702–734, 2007.

A. G. Chandrasekhar and R. Lewis. Econometrics of Sampled Networks. *Networks*, pages 270–275, 2011.

M. Comola and S. Prina. Treatment Effects Accounting for Network Changes. pages 106–112.

C. T. Conlon and J. H. Mortimer. Efficiency and foreclosure effects of vertical rebates: empirical evidence. Technical report, National Bureau of Economic Research, 2013.

A. P. Dempster, N. M. Laird, and D. B. Rubin. *Maximum likelihood from incomplete data via the EM algorithm*, volume 39. 1977.

E. Duflo, P. Dupas, and M. Kremer. Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. *American Economic Review*, 101(5):1739–74, 2011.

M. Fafchamps. Ethnicity and networks in african trade. *Contributions in Economic Analysis & Policy*, 2(1), 2003.

A. D. Foster and M. R. Rosenzweig. Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of political Economy*, 103(6):1176–1209, 1995.

S. Frühwirth-Schnatter. *Finite mixture and Markov switching models*. 2006.

S. Godlonton and R. Thornton. Peer effects in learning hiv results. *Journal of Development Economics*, 97(1):118–129, 2012.

A. Griffith. How Many Friends Do You Have ? An Empirical Investigation into Censoring-Induced Bias in Social Network Data. 2017.

A. Griffith. Random Assignment with Non-Random Peers: A Structural Approach to Counterfactual Treatment Assessment. 2017.

B. Grün and F. Leisch. Finite mixtures of generalized linear regression models. *Recent Advances in Linear Models and Related Areas*, (013):205–230, 2007.

M. S. Handcock and K. J. Gile. Modeling Social Networks From Sampled Data. *Annals of Applied Statistics*, 4(1):5–25, 2010.

M. Hardy and J. McCasland. It Takes Two : Experimental Evidence on the Determinants of Technology Diffusion. pages 1–52, 2016.

R. Heath. Why do firms hire using referrals? evidence from bangladeshi garment factories. *Journal of Political Economy*, 126(4):1691–1746, 2018.

C. Hennig. Identifiablity of models for clusterwise linear regression, 2000.

J. R. Magruder. Intergenerational networks, unemployment, and persistent inequality in south africa. *American Economic Journal: Applied Economics*, 2(1):62–85, 2010.

C. F. Manski. Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, 60(3):531, 1993.

C. F. Manski. Identification of treatment response with social interactions. *Econometrics Journal*, 16(1):1–23, 2013.

P. V. Marsden. Survey Methods for Network Data. In *The SAGE Handbook of Social Network Analysis*, pages 370–388. 2016.

G. J. McLachlan and K. E. Basford. *Mixture Models: Inference and Applications to Clustering.* 1988.

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

K. Munshi and J. Myaux. Social norms and the fertility transition. *Journal of development Economics*, 80(1):1–38, 2006.

K. Munshi and M. Rosenzweig. Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy. *American Economic Review*, 96(4):1225–1252, 2006.

K. Munshi and M. Rosenzweig. Networks and misallocation: Insurance, migration, and the rural-urban wage gap. *American Economic Review*, 106(1):46–98, 2016.

M. E. Newman. Network structure from rich but noisy data. *Nature Physics*, 14(June):1–4, 2018.

B. E. Oster and R. Thornton. Menstruation, Sanitary Products, and School Attendance: Evidence from a Randomized Evaluation. 3(January):91–100, 2011.

D. B. Rubin. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66(5):688–701, 1974.

M. Stein. Know-how and know-who: Effects of a randomized training on network changes between small urban entrepreneurs. 2018.

J. Ugander, B. Karrer, L. Backstrom, and J. Kleinberg. Graph cluster randomization: network exposure to multiple universes. 2013.

K. N. Vasilaky and K. L. Leonard. As good as the networks they keep? improving outcomes through weak ties in rural uganda. *Economic Development and Cultural Change*, 66(4):755–792, 2018.

A. Wald. Note on the Consistency of the Maximum Likelihood Estimate. *The Annals of Mathematical Statistics*, 20(4):595–601, 1949.

S.-Y. Wang. Marriage networks, nepotism, and labor market outcomes in china. *American Economic Journal: Applied Economics*, 5(3):91–112, 2013.

K. M. Williams. Economic Analyses of Educational Achievement. *ProQuest Dissertations and Theses*, page 144, 2016.

A   Additional Figures
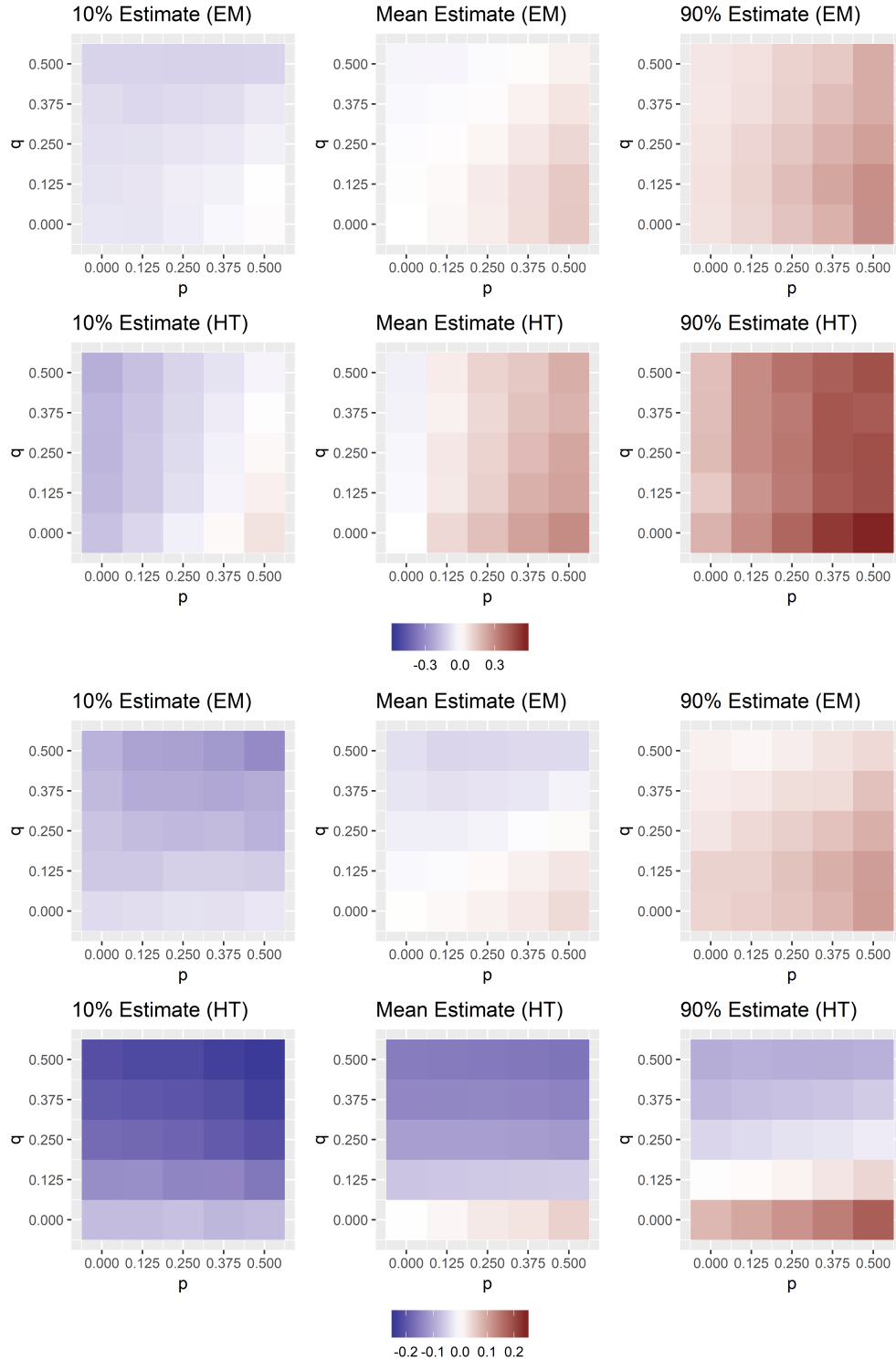
Figure 6: *Estimates of the mean outcome of the no exposure (top) and indirect exposure (bottom) conditions from their true values under varying corruption levels (p,q) for the network. Estimates obtained from our model using the EM algorithm are compared against estimates from the Horvitz-Thompson estimators assuming no corruption in the network. The 0.1 and 0.9 quantiles are provided for both methods to give a sense of the variability in these estimates. Note the color gradient scales are different for the two exposure conditions.*