Higher-Order Causal Message Passing for Experimentation with Complex Interference

Mohsen Bayati[1] **Yuwei Luo**[1] **William Overman**[1] **Sadegh Shirani**[1] **Ruoxuan Xiong**[2]

[1] Stanford Graduate School of Business [2] Emory University
{bayati, yuweiluo, wpo, sshirani}@stanford.edu, ruoxuan.xiong@emory.edu

Abstract

Accurate estimation of treatment effects is essential for decision-making across various scientific fields. This task, however, becomes challenging in areas like social sciences and online marketplaces, where treating one experimental unit can influence outcomes for others through direct or indirect interactions. Such interference can lead to biased treatment effect estimates, particularly when the structure of these interactions is unknown. We address this challenge by introducing a new class of estimators based on causal message-passing, specifically designed for settings with pervasive, unknown interference. Our estimator draws on information from the sample mean and variance of unit outcomes and treatments over time, enabling efficient use of observed data to estimate the evolution of the system state. Concretely, we construct non-linear features from the moments of unit outcomes and treatments and then learn a function that maps these features to future mean and variance of unit outcomes. This allows for the estimation of the treatment effect over time. Extensive simulations across multiple domains, using synthetic and real network data, demonstrate the efficacy of our approach in estimating total treatment effect dynamics, even in cases where interference exhibits non-monotonic behavior in the probability of treatment.

1 Introduction

Randomized experiments are widely recognized as a reliable method in data-driven decision-making for determining the causal effects of new interventions, such as medical treatments or upgrades of market products. The conventional approach involves administering the new treatment to a randomly selected subset of the observation units (e.g., patients, products, or geographical areas), referred to as the *treatment* group, and comparing their outcomes with those units who received no treatment, the *control* group. However, the validity of these methods substantially relies on the assumption that treating a group of units does not interfere with the outcomes of the control units, known as the Stable Unit Treatment Value Assumption (SUTVA) [Cox, 1958, Rubin, 1978, Manski, 1990, Imbens and Rubin, 2015, Sussman and Airoldi, 2017].

In many social science and online marketplace scenarios, treating one unit impacts not only its outcome but also the outcomes of units that directly or indirectly interact with the treated unit [Bond et al., 2012, Blake and Coey, 2014, Holtz et al., 2020, Johari et al., 2022, Bright et al., 2022]. This *interference* of treatments and outcomes makes estimating the causal effect of the treatment particularly challenging. Considering the network of interactions, when a unit is treated, its interactions with neighboring units lead to subsequent changes in their outcomes. These interactions continue over the experimental time horizon and may display complex behaviors. For example, as the treatment is expanded to a larger population, the interference effect may intensify or diminish. This necessitates efficient data usage and robust estimators to capture and adapt to such intricacies.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

 https://doi.org/10.52202/079017-2600

Given the complexity of analyzing interference phenomena, research on network interference often relies on a series of simplifying assumptions. One common assumption is to ignore variations over time and assume outcomes are observed at equilibrium, which discards valuable information before the system reaches equilibrium. To reduce the complexity of the analysis, further assumptions are imposed on the nature and level of interference [Choi, 2017, Cortez et al., 2022, Li and Wager, 2022a], such as the neighborhood interference assumption or assumptions on the maximum degree of the network. Additionally, a frequently made assumption to help estimate treatment effects is that the interference network is observed [Chen et al., 2024, Agarwal et al., 2022, Jia et al., 2024], which is impractical in some settings, such as under pervasive interference. For example, in large-scale online platforms, units may interact through competing platforms, making it difficult to account for all sources of interference. Our aim in this paper is to relax these assumptions.

The impact of network interference can be intricate, particularly when considering interactions among units over time. For example, applying the treatment to one unit can have *spillover effects* on some control units, or one unit's outcome can directly exert *peer effects* on other units' outcomes. Simultaneously, treatments with long-lasting effects can have *carryover effects* to future time periods, and units' outcomes can be serially correlated or have *autocorrelation* over time. Consequently, whenever SUTVA fails to hold, the number of potential outcomes grows exponentially with the population size and the time horizon of the experiment. This renders the estimation of causal effects under general interference structures impossible due to non-identifiability challenges [Manski, 2013, Aronow and Samii, 2017, Basse and Airoldi, 2018, Karwa and Airoldi, 2018, Forastiere et al., 2022].

Recently, Shirani and Bayati [2024] introduced a new framework called *Causal Message-Passing* (CMP) to address the challenge of causal effect estimation under unobserved pervasive interference. Their methodology relies on observing outcomes over time and is rooted in statistical physics [Mezard et al., 1986, Mezard and Montanari, 2009] and approximate message passing (AMP) [Donoho et al., 2009, Bayati and Montanari, 2011] from high dimensional statistics. Instead of investigating the complex relationships among units, which requires knowledge of the network, CMP focuses on the dynamics of one-dimensional quantities, such as the sample mean and sample variance of units' outcomes over time. These one-dimensional equations, also known as *state evolution* equations, can help track how the administered intervention propagates through the network of units over time, which enables the estimation of counterfactual scenarios. However, it remains underexplored how to use state evolution to estimate causal effects.

In this work, we propose to utilize machine learning to learn a mapping that updates key parameters of the distribution of outcomes over time for causal effect estimation. This is achieved by introducing a set of non-linear feature functions that act on the observed outcomes, creating a "basis" for the learning task. By training a properly designed machine learning model on this extracted basis, we estimate the Total Treatment Effect (TTE), also known as the Global Treatment Effect (GTE) or Global Average Treatment Effect (GATE), which measures the causal effect of altering the treatment scenario from treating no one to treating everyone. The result is a family of estimators that allow one to extract more information from the experimental data, thereby ensuring efficient use of the data.

To be more specific, this work builds on the foundation established by Shirani and Bayati [2024], extending their method in two directions by introducing Higher-Order Causal Message Passing (HO-CMP) algorithms. First, HO-CMP incorporates higher-order moments of unit outcomes, unlike Shirani and Bayati [2024]'s approach, which only employs the first moments for estimation. Second, while Shirani and Bayati [2024] focus solely on two-stage experiments with two different probabilities of treatment, our work leverages the additional data provided by having more than two experimental stages with multiple probabilities of treatment. Thus, our work aligns with the common practice in the tech industry of rolling out treatments through a sequence of experiments [Kohavi et al., 2020].

We then validate the performance of HO-CMP by simulating multiple experimental settings, encompassing both linear and non-linear outcome specifications and various types of interference, such as synthetic random geometric networks and real-world networks. Specifically, we introduce a *Non-LinearInMeans* outcome specification, where the spillover effect is non-monotone in the fraction of treated neighbors; as an example of a complex treatment effect structure, we demonstrate how HO-CMP successfully estimates the total treatment effect by effectively utilizing higher-order moments of unit outcomes.

Simulating the experiments also allows us to calculate the ground truth value of the TTE, which remains unknown in real experiments, enabling us to compare the performance of HO-CMP to

the ground truth TTE. Additionally, we benchmark HO-CMP against standard approaches such as difference-in-means and Horvitz-Thompson estimators, a recent technique of Cortez et al. [2022], and a first-order CMP estimation, like the one by Shirani and Bayati [2024]. We emphasize that a large body of recent estimators, e.g., Jia et al. [2024], requires knowledge of the interference network and is not applicable in our setting. The results showcase HO-CMP outperforming the benchmarks in estimating the TTE over time and its flexibility to cover different outcome specifications and interference structures.

Related causal inference literature. The primary objective of research on causal inference in the context of network interference is to estimate causal effects while relaxing SUTVA. For this purpose, various assumptions and methods have been proposed. We briefly discuss the predominant ones.

A common approach to relax SUTVA is partial interference. Under this assumption, units are divided into disjoint clusters and interference is assumed only within the same cluster [Sobel, 2006, Rosenbaum, 2007, Hudgens and Halloran, 2012, Tchetgen and VanderWeele, 2012, Liu and Hudgens, 2014, Kang and Imbens, 2016, Viviano, 2020b, Bhattacharya et al., 2020, Qu et al., 2021, Auerbach and Tabord-Meehan, 2021, Candogan et al., 2023, Ugander and Yin, 2023]. When interference extends across clusters, standard estimators become biased. To address this, Eckles et al. [2016] propose a cluster-randomized approach that randomizes treatment assignment across clusters, reducing bias. However, it requires knowledge of the clusters.

The other assumption to replace SUTVA is the Neighborhood Interference Assumption (NIA). NIA states that outcomes are only influenced by the treatments of neighboring units in the network. This assumption is commonly imposed in the literature that relaxes the SUTVA [Sussman and Airoldi, 2017]. Some recent studies combine the NIA with the availability of either a fully or partially observed interference structure [Leung, 2020, Viviano, 2020a, Agarwal et al., 2022, Belloni et al., 2022, Li and Wager, 2022b]. Without prior knowledge of the interference structure, Cortez et al. [2022] consider low-degree polynomial interactions among units in the network. Leung [2022] also introduces a weaker version of the NIA, where the interference between two units located far away from each other is allowed to be nonzero, but negligible.

Another approach is to facilitate the estimation of causal effects by setting restrictions on the network structure [Chin, 2018, Jagadeesan et al., 2020, Wang et al., 2020, Li and Wager, 2022a, Agarwal et al., 2022, Jagadeesan et al., 2020, Leung, 2022]. These restrictions include bounding the largest node degree of the interference graph, limiting the degree of the dependency graph, observing specific patterns in the network, locally constrained interference structures, and restricting the topology of the interference network.

Driven by applications in marketplace platforms and two-sided marketplaces, several recent works have examined specific interference patterns [Holtz et al., 2020, Wager and Xu, 2021, Munro et al., 2021, Johari et al., 2022, Harshaw et al., 2022, Farias et al., 2022, Bright et al., 2022, Farias et al., 2023]. For example, Farias et al. [2022] study experiments in Markovian systems where interference effects propagate through constraints like limited inventory.

From another perspective, most of the existing literature on network interference focuses on the case of single-time point observation [Hudgens and Halloran, 2012, Aronow and Samii, 2017, Basse et al., 2019, Jackson et al., 2020, Sävje et al., 2021]. These studies have provided insightful results on spatial interference effects, but they often overlook temporal variations of the treatment effect. Recently, there has been a shift to consider settings with multiple-time observations [Li and Wager, 2022a, Boyarsky et al., 2023]. However, the problem of considering the dynamics of units' outcomes remains understudied [Arkhangelsky and Imbens, 2023].

2 Setup and Foundation

Consider a system of N units indexed by $i \in [N] := \{1, \cdots, N\}$ subject to a randomized experiment. The units are observed over a time horizon of $T+1$ periods and for each $t \in \{0, 1, \ldots, T\}$, we let W_t^i denote the treatment status of unit i during time period t. For simplicity, we consider a Bernoulli randomized design such that $W_t^i \sim \text{Bernoulli}(\pi_t)$. That is, at time t unit i receives the *treatment* with a probability of π_t, corresponding to $W_t^i = 1$. Otherwise, unit i belongs to the *control* group and $W_t^i = 0$. In this context, we collectively define $\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots, \pi_T)$ as the *experimental design*. Then, following the potential outcome framework [Imbens and Rubin, 2015], let $Y_t^i(\boldsymbol{W})$ represent

 https://doi.org/10.52202/079017-2600

the potential outcome of unit i at time t, where $\boldsymbol{W}$ denotes the entire treatment allocation matrix, with W_t^i as the entry in row t and column i.

Administering the treatment of unit i at time t according to w_t^i (as one realization of the random variable W_t^i), we use $\boldsymbol{w}$ (as one realization of $\boldsymbol{W}$) to show the matrix that captures the treatments of all units throughout the experiment; accordingly, we let $y_t^i = Y_t^i(\boldsymbol{W} = \boldsymbol{w})$ be the observed outcome of unit i at time t under the treatment assignment $\boldsymbol{w}$:

$$\boldsymbol{w} = \begin{bmatrix} w_0^1 & w_0^2 & \dots & w_0^N \\ w_1^1 & w_1^2 & \dots & w_1^N \\ \vdots & \vdots & \ddots & \vdots \\ w_T^1 & w_T^2 & \dots & w_T^N \end{bmatrix}, \qquad \boldsymbol{y} = \begin{bmatrix} y_0^1 & y_0^2 & \dots & y_0^N \\ y_1^1 & y_1^2 & \dots & y_1^N \\ \vdots & \vdots & \ddots & \vdots \\ y_T^1 & y_T^2 & \dots & y_T^N \end{bmatrix}.$$

Observing $(\boldsymbol{w}, \boldsymbol{y})$, we are interested in estimating the TTE of the intervention, defined as below:

$$\mathrm{TTE}_t = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left[Y_t^i(\mathbf{1}) - Y_t^i(\mathbf{0}) \right], \quad t = 0, 1, \dots, T, \tag{1}$$

where $\mathbf{1}$ and $\mathbf{0}$ are matrices of all 1 and all 0 of appropriate dimensions (in this case, $T+1$ by N). Intuitively, the TTE measures the average effect of changing the treatment for the entire population. This is a common estimand in the network interference literature and provides important insights into the efficacy of the treatment for decision-makers [Jia et al., 2024, Chen et al., 2024, Viviano et al., 2023, Yu et al., 2022, Cortez et al., 2022].

Deriving a practical and efficient estimator for the TTE is challenging due to the fact that we can observe the population only under one treatment scenario [Holland, 1986]. Indeed, in Eq. (1), we can observe at most one of $Y_t^i(\mathbf{1})$ or $Y_t^i(\mathbf{0})$, and often, neither.[1] In the following sections, we address this challenge by proposing a new class of estimators grounded in the CMP framework. These estimators rely on the efficient use of experimental data, $\boldsymbol{y}$ and $\boldsymbol{w}$, yielding accurate causal estimation under unknown network interference.

2.1 Potential outcome specification and state evolution of the experiment

In this section we provide a summary of the outcome specification and results of Shirani and Bayati [2024] that we utilize in the remaining. For $t = 0, 1, \dots, T-1$, we let $g_t : \mathbb{R} \times \mathbb{R}^{T+1} \mapsto \mathbb{R}$ be an unknown measurable function. We also use $\vec{W}^i = \left(W_0^i, \dots, W_T^i\right)^\top$ to denote the treatment assignment of unit i during the experiment. Accordingly, the treatment allocation matrix $\boldsymbol{W}$ is a $T+1$ by N matrix with columns equal to $\vec{W}^i$. Given potential outcomes $Y_t^j(\boldsymbol{W})$ at time t and $j \in [N]$, their outcomes in time period $t+1$ are specified by

$$Y_{t+1}^i(\boldsymbol{W}) = \sum_{j=1}^{N} \mathrm{G}^{ij} g_t\left(Y_t^j(\boldsymbol{W}), \vec{W}^j\right) + \epsilon_t^i, \qquad t = 0, 1, \dots, T-1, \tag{2}$$

where G^{ij} quantifies the impact of unit j on unit i at time t and ϵ_t^i is a zero-mean Gaussian noise with a variance of σ_e^2, accounting for measurement errors. In addition, we let $\mathbf{G} = [\mathrm{G}^{ij}]_{i,j\in[N]}$ and refer to it as the *interference matrix*. Then, according to Eq. (2), the function g_t captures the impact of past outcomes and treatment assignments of other units on the current outcome of unit i.

Now, fixing t, we define

$$\nu_t(\boldsymbol{W}) := \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} Y_t^i(\boldsymbol{W}), \qquad \rho_t(\boldsymbol{W})^2 := \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} Y_t^i(\boldsymbol{W})^2 - \nu_t(\boldsymbol{W})^2. \tag{3}$$

Then, as shown by Shirani and Bayati [2024], whenever the elements of the interference matrix G^{ij} are i.i.d. Gaussian random variables with mean μ/N and variance σ^2/N, under mild moment conditions on initial values Y_0^i, we have

$$\begin{aligned} \nu_{t+1}(\boldsymbol{W}) &\overset{\text{a.s.}}{=} \mu \mathbb{E}\left[g_t\big(\nu_t(\boldsymbol{W}) + \rho_t(\boldsymbol{W}) Z_t, \vec{W}\big) \right], \\ \rho_{t+1}(\boldsymbol{W})^2 &\overset{\text{a.s.}}{=} \sigma^2 \mathbb{E}\left[g_t\big(\nu_t(\boldsymbol{W}) + \rho_t(\boldsymbol{W}) Z_t, \vec{W}\big)^2 \right] + \sigma_e^2, \end{aligned} \tag{4}$$

[1] We can only observe $\boldsymbol{y}$ for one of exponentially many realizations of $\boldsymbol{w}$.

where $Z_t \sim \mathcal{N}(0,1)$ is independent from $\vec{W} \sim \text{Bernoulli}(\boldsymbol{\pi})$ (that is, $W_t \sim \text{Bernoulli}(\pi_t)$ and $\vec{W} = (W_0, W_1, \ldots, W_T)^\top$) and the equalities hold almost surely. We note that the theory behind this result is rooted in the AMP literature, going back to Bolthausen [2014], Bayati and Montanari [2011]. However, as Shirani and Bayati [2024] note, there is a major distinction between the AMP literature and the above setting: in the AMP literature, the matrix $\mathbf{G}$ is observed, and the aim is to construct proper functions g_t for a completely different objective, which is studying the high-dimensional asymptotics of first-order algorithms. However, in the current context, the matrix $\mathbf{G}$ and functions g_t are *unknown* and the goal is to estimate them.

Considering Eq. (3), the equations in (4) determine the dynamics of the sample mean and sample variance of unit outcomes over time in large sample asymptotics, and are denoted by the State Evolution (SE) equations of the experiment [Shirani and Bayati, 2024]. In the next section, we present an efficient algorithm to learn the state evolution dynamics outlined in Eq. (4). This method enables us to accurately estimate the TTE defined in Eq. (1) and its corresponding confidence interval.

3 Algorithm

In this section, we introduce *Higher-order Causal Message-passing* (HO-CMP) for estimating the TTE over the entire time horizon of the experiment. Briefly speaking, HO-CMP directly estimates the update function in the state evolution equations (4), thereby estimating counterfactual quantities while accounting for the impact of unknown network interference. To this end, by Eqs. (1) and (3), we rewrite the TTE as the difference of the sample means in the large limits:

$$\text{TTE}_t = \nu_t(\mathbf{1}) - \nu_t(\mathbf{0}).$$

That means the problem of estimating the TTE is equivalent to estimating $\nu_t(\mathbf{1})$ and $\nu_t(\mathbf{0})$ using the observed data, denoted by $(\boldsymbol{w}, \boldsymbol{y})$. On the other hand, considering the state evolution equations in (4), the *system state* at time $t+1$, denoted by $(\nu_{t+1}(\boldsymbol{w}), \rho_{t+1}(\boldsymbol{w})^2)$, is a (nonlinear) function of the system state distribution at time t, characterized by $(\nu_t(\boldsymbol{w}), \rho_t(\boldsymbol{w})^2)$ and $\vec{W}$, encompassing the sample mean and variance of observed outcomes as well as the design of the experiment. However, because the exact functional form and parameters of equations in (4) are unknown, one cannot directly apply the SE to track the evolution of states. Therefore, we propose to estimate the unknown update functions in SE equations, utilizing the observed data $(\boldsymbol{w}, \boldsymbol{y})$. For this purpose, we fix the treatment assignment matrix $\boldsymbol{w}$ and define

$$\hat{\nu}_t(\boldsymbol{w}) := \frac{1}{N}\sum_{i=1}^{N} y_t^i, \qquad \hat{\rho}_t(\boldsymbol{w})^2 := \frac{1}{N}\sum_{i=1}^{N} \left(y_t^i - \hat{\nu}_t(\boldsymbol{w})\right)^2,$$
$$\bar{w}_t := \frac{1}{N}\sum_{i=1}^{N} w_t^i, \qquad \vec{\bar{w}} := (\bar{w}_0, \ldots, \bar{w}_T)^\top.$$

In addition, let $\vec{\phi} = (\phi_k)_{k\in[K]}$ be a prespecified vector of measurable feature functions of current estimates of the sample mean $\hat{\nu}_t(\boldsymbol{w})$, sample variance $\hat{\rho}_t(\boldsymbol{w})^2$, and the design $\boldsymbol{w}$. We define $\boldsymbol{x}_t$ to represent the *feature vector* as follows:

$$\boldsymbol{x}_t = \vec{\phi}\Big(\hat{\nu}_t(\boldsymbol{w}), \hat{\rho}_t(\boldsymbol{w}), \boldsymbol{w}\Big) := \Big[\phi_1\Big(\hat{\nu}_t(\boldsymbol{w}), \hat{\rho}_t(\boldsymbol{w}), \boldsymbol{w}\Big), \ldots, \phi_K\Big(\hat{\nu}_t(\boldsymbol{w}), \hat{\rho}_t(\boldsymbol{w}), \boldsymbol{w}\Big)\Big]^\top.$$

Then, we formally propose learning the mapping $f_{\boldsymbol{\theta}}(\cdot)$ defined by,

$$(\hat{\nu}_{t+1}(\boldsymbol{w}), \hat{\rho}_{t+1}(\boldsymbol{w})^2) = f_{\boldsymbol{\theta}}(\boldsymbol{x}_t) \tag{5}$$

We summarize the method in Algorithm 1. Note in our experiment design we begin with all units under control by setting $\pi_0 = 0$, meaning no units receive treatment in period 0. Additionally, to avoid non-identifiability issues, the experiment requires at least two stages, which corresponds to having at least two distinct values in the set $\{\pi_1, \ldots, \pi_T\}$.

The proposed HO-CMP method encompasses a rich family of estimators, offering flexibility through the selection of feature functions $\{\phi_k\}_{k\in[K]}$ and model $f_{\boldsymbol{\theta}}(\cdot)$. Specifically, incorporating proper feature (basis) functions, with examples shown in Table 1, facilitates the extraction of informative patterns for learning the unknown nonlinear dynamics of the system throughout the experiment. In

 https://doi.org/10.52202/079017-2600

Table 1: Two examples of feature functions

Algorithms	Feature functions $\{\phi_k(\hat{\nu}_t(\boldsymbol{w}), \hat{\rho}_t(\boldsymbol{w})^2, \boldsymbol{w})\}_{k\in[K]}$	$f_{\boldsymbol{\theta}}(\cdot)$
FO-CMP	$\{\hat{\nu}_t(\boldsymbol{w}), \bar{w}_{t+1}, \hat{\nu}_t(\boldsymbol{w}) \cdot \bar{w}_t\}$	linear regression
HO-CMP	$\left\{\hat{\nu}_t(\boldsymbol{w}), \bar{w}_{t+1}, \hat{\nu}_t(\boldsymbol{w}) \cdot \bar{w}_t, \hat{\rho}_t(\boldsymbol{w})^2, \bar{w}_{t+1}^2\right\}$	linear regression

practice, one could choose these basis functions based on heuristics, domain knowledge, and prior information about the dynamics.

Specifically, in this paper, we consider the following estimators, as summarized in Table 1.

FO-CMP (First-Order Causal Message-Passing): This corresponds to the simple setting where $\nu_{t+1}(\boldsymbol{w})$ is assumed to be a function of the previous sample mean $\nu_t(\boldsymbol{w})$, the sample mean of the current treatment $\bar{w}_{t+1}$, and an additional term to model the interaction of the dynamics and previous treatments $\nu_t(\boldsymbol{w})\bar{w}_t$. Consequently, this model is irrelevant of the variance $\rho_{t+1}(\boldsymbol{w})^2$. This is true when g_t takes a simple nonlinear form $g_t(y_t, \vec{w}) = \alpha y_t + \beta w_{t+1} + \gamma y_t w_t$. We remark that FO-CMP essentially uses the first state evolution equation in (4) and fails to extract informative signals from the second evolution equation.

HO-CMP (Higher-Order Causal Message-Passing): HO-CMP further introduces the second-order terms $(\bar{w}_{t+1})^2$ and $\hat{\rho}_t(\boldsymbol{w})^2$ to model the nonlinear treatment effects. It improves data efficiency by utilizing both state evolution equations. It also allows estimation of higher order terms in Taylor series of g_t.

While FO-CMP extends the estimation algorithm in Shirani and Bayati [2024] to accommodate experiments with more than two stages, HO-CMP introduces a new dimension to the estimation problem by incorporating second-order terms. This inclusion enhances data utilization, resulting in higher estimation efficiency in HO-CMP compared to FO-CMP.

Algorithm 1: Higer-Order Causal Message Passing (HO-CMP)

Data: Observed data $(\boldsymbol{w}, \boldsymbol{y})$, feature functions $\vec{\phi} = (\phi_k)_{k\in[K]}$, machine learning model $f_{\boldsymbol{\theta}}(\cdot)$

Step 1: Data processing

for $t \leftarrow 0$ ***to*** T **do**

$\hat{\nu}_t(\boldsymbol{w}) \leftarrow \frac{1}{N}\sum_{i=1}^{N} y_t^i$,

$\hat{\rho}_t(\boldsymbol{w})^2 \leftarrow \frac{1}{N}\sum_{i=1}^{N}(y_t^i - \hat{\nu}_t(\boldsymbol{w}))^2$,

$\boldsymbol{x}_t \leftarrow \vec{\phi}\left(\hat{\nu}_t(\boldsymbol{w}), \hat{\rho}_t(\boldsymbol{w})^2, \boldsymbol{w}\right)$

end

Step 2: Model Estimation

Estimate $f_{\boldsymbol{\theta}}$ from data $\left\{\left(\boldsymbol{x}_t, \left(\hat{\nu}_{t+1}(\boldsymbol{w}), \hat{\rho}_{t+1}(\boldsymbol{w})^2\right)\right)\right\}_{t\in[T-1]}$, guided by (5).

Step 3: Counterfactual Estimation

$\hat{\nu}_0(\mathbf{0}) \leftarrow \hat{\nu}_0(\boldsymbol{w})$, $\hat{\nu}_0(\mathbf{1}) \leftarrow \hat{\nu}_0(\boldsymbol{w})$, $\hat{\rho}_0(\mathbf{0})^2 \leftarrow \hat{\rho}_t(\boldsymbol{w})^2$, $\hat{\rho}_0(\mathbf{1})^2 \leftarrow \hat{\rho}_t(\boldsymbol{w})^2$, $\widehat{\mathrm{TTE}}_0 \leftarrow 0$

for $t \leftarrow 0$ ***to*** $T-1$ **do**

Compute the features and predict the counterfactuals

$\boldsymbol{x}_t(\mathbf{0}) \leftarrow \vec{\phi}\left(\hat{\nu}_t(\mathbf{0}), \hat{\rho}_t(\mathbf{0})^2, \mathbf{0}\right)$, $\boldsymbol{x}_t(\mathbf{1}) \leftarrow \vec{\phi}\left(\hat{\nu}_t(\mathbf{1}), \hat{\rho}_t(\mathbf{1})^2, \mathbf{1}\right)$

$\left(\hat{\nu}_{t+1}(\mathbf{0}), \hat{\rho}_{t+1}(\mathbf{0})^2\right) \leftarrow f_{\boldsymbol{\theta}}(\boldsymbol{x}_t(\mathbf{0}))$, $\left(\hat{\nu}_{t+1}(\mathbf{1}), \hat{\rho}_{t+1}(\mathbf{1})^2\right) \leftarrow f_{\boldsymbol{\theta}}(\boldsymbol{x}_t(\mathbf{1}))$

Estimate the TTE

$\widehat{\mathrm{TTE}}_{t+1} \leftarrow \hat{\nu}_{t+1}(\mathbf{1}) - \hat{\nu}_{t+1}(\mathbf{0})$

end

Result: $\left\{\widehat{\mathrm{TTE}}_t\right\}_{t\in[T]}$

4 Experiments

In this section, we use synthetic experiments under simulated and real-world network interference patterns, to compare the performance of FO-CMP and HO-CMP estimators, outlined in Table 1 and Algorithm 1, with several benchmarks. First, we introduce the experimental design, benchmark estimators, interference patterns, and outcome specifications.

Experimental design. We primarily focus on the staggered rollout design with L distinct treated probabilities, denoted by $\pi^{(1)}, \cdots, \pi^{(L)}$, where $\pi^{(\ell)}$ increases monotonically with $\ell \in \{1, \ldots, L\}$. In the first $T^{(1)}$ periods, $\pi^{(1)} \times 100\%$ of units are in the treatment group. From $T^{(1)}$ to $T^{(2)}$ periods, $\pi^{(2)} \times 100\%$ of units are in the treatment group, and so forth. In the staggered rollout design, once a unit is allocated to treatment, it remains in the treatment group until the experiment concludes [Xiong et al., 2024]. In the appendix, we also consider the Bernoulli randomized design, where the treatment is re-randomized at every time period, allowing units to switch between the treatment and control groups throughout the experiment. We use two values of $T = 40,\ 200$ and set $L = 4$, with $(\pi^{(1)}, \pi^{(2)}, \pi^{(3)}, \pi^{(4)}) = (0.1, 0.2, 0.4, 0.5)$. In the appendix, we show the impact of increasing L or the maximum treatment probability $\pi^{(L)}$.

Benchmark estimators. We first present two benchmark estimators commonly used for treatment effect estimation, both in settings with and without network interference. The final estimator is designed specifically for settings with unknown network interference [Cortez et al., 2022].

The first benchmark estimator is the standard difference-in-means (DM) estimator given by

$$\widehat{\mathrm{TTE}}_t^{\mathrm{dm}} = \frac{\sum_{j=1}^N y_t^j w_t^j}{\sum_{j=1}^N w_t^j} - \frac{\sum_{j=1}^N y_t^j (1 - w_t^j)}{\sum_{j=1}^N (1 - w_t^j)},$$

which is the difference in average outcomes between treated and control units at each time period t.

The second benchmark is the standard Horvitz and Thompson [1952] (HT) estimator given by

$$\widehat{\mathrm{TTE}}_t^{\mathrm{ht}} = \frac{1}{N} \sum_{j=1}^N \left[\frac{y_t^j w_t^j}{\pi_t} - \frac{y_t^j (1 - w_t^j)}{1 - \pi_t} \right],$$

which weights observed outcomes by the inverse propensity score (i.e., $1/\pi_t$ or $1/(1 - \pi_t)$).

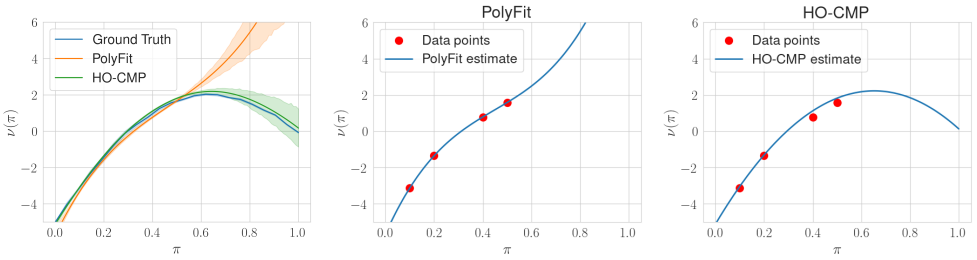
The third benchmark estimator is the polynomial interpolation estimator (PolyFit) introduced by Cortez et al. [2022]. PolyFit operates by obtaining estimates for the average of outcomes at equilibrium for L treated probabilities $\pi^{(1)}, \cdots, \pi^{(L)}$, denoted by $\nu_{\mathrm{equil}}(\pi^{(1)}), \ldots, \nu_{\mathrm{equil}}(\pi^{(L)})$, then it utilizes Lagrange interpolation method and obtains a degree-L polynomial approximation for the function $\nu_{\mathrm{equil}} : [0, 1] \to \mathbb{R}$ which can be used to estimate the equilibrium values under global control and treatment, $\hat{\nu}_{\mathrm{equil}}(0)$ and $\hat{\nu}_{\mathrm{equil}}(1)$. Finally, TTE is estimated by

$$\widehat{\mathrm{TTE}}_t^{\mathrm{polyfit}} = \hat{\nu}_{\mathrm{equil}}(1) - \hat{\nu}_{\mathrm{equil}}(0).$$

On the one hand, PolyFit does not need any knowledge of the interference network; however, it comes at the expense of having to grapple with two challenges. First, it may incur a high variance as L increases due to fitting a high-degree polynomial. The second challenge is that it needs accurate estimates for each $\nu_{\mathrm{equil}}(\pi^{(\ell)})$, which requires treating $\pi^{(\ell)}$ fraction of units for a long enough number of periods so that the outcomes reach an equilibrium. This can be achieved if the staggered roll-out design is performed over a long enough horizon T with each $T^{(\ell)}$ sufficiently large, and then estimating each $\nu_{\mathrm{equil}}(\pi^{(\ell)})$ by sample average of outcomes at time $T^{(\ell)}$. However, when such a lengthy experiment is not feasible, the estimates for $\nu_{\mathrm{equil}}(\pi^{(\ell)})$ will be less accurate.

Interference networks. We consider two networks (graphs). The first graph is a simulated random geometric graph model, studied by Leung [2022]. The second graph is a social network of Twitch users [Rozemberczki and Sarkar, 2021]. In either scenario, we denote the adjacency matrix of the graph by $E \in \{0, 1\}^{N \times N}$. For any i and j, E_{ij} equals 1 if j is a neighbor of i and 0 otherwise.

https://doi.org/10.52202/079017-2600



(a) $\nu_{\text{equil}}(\pi)$ vs π (b) Curve fitting of polyfit (c) Curve fitting of HO-CMP

Figure 1: (a) $\nu_{\text{equil}}(\pi)$ with PolyFit and HO-CMP estimates across runs (Non-LinearInMeans). (b) and (c) show one sample estimates with observed data points.

Outcome generating processes. We consider two outcome specifications to model monotone and non-monotone interference patterns. Specifically, for both settings, we generate outcomes using the following specification:

$$Y_{t+1}^i = \alpha + \beta \frac{\sum_{j=1}^N E_{ij} Y_t^j}{\sum_{j=1}^N E_{ij}} + \delta \cdot g\left(\frac{\sum_{j=1}^N E_{ij} W_{t+1}^j}{\sum_{j=1}^N E_{ij}}\right) + \gamma W_{t+1}^i + \epsilon_{t+1}^i \,,$$

where in the first setting, $g(\cdot)$ is taken to be the identity function, i.e., $g(x) = x$ for any x. Therefore, Y_{t+1}^i depends linearly on the fraction of treated neighbors, and we refer to this setting as the *LinearInMeans* outcome setting. This setting is widely studied in the causal inference literature [Cai et al., 2015, Eckles et al., 2016, Leung, 2022].

In the second setting, $g(\cdot)$ is specified by a periodic function, i.e., $g(x) = \sin(\pi x)$ for any x. Therefore, Y_{t+1}^i, on average, first increases and then decreases with the fraction of treated neighbors, as visualized by the Ground Truth curve in panel (a) of Figure 1. We refer to this setting as the *Non-LinearInMeans* outcome setting.

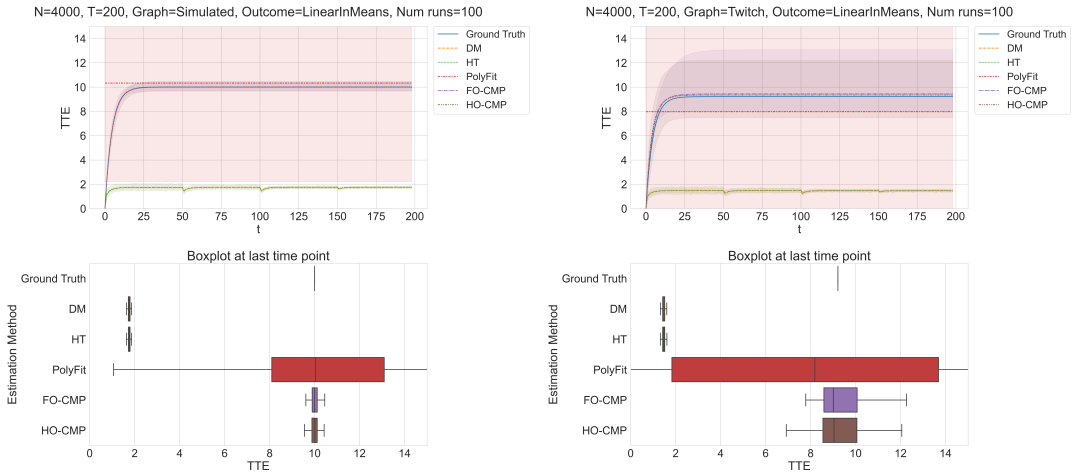
Results. We compare FO-CMP and HO-CMP with the three benchmarks for estimating the TTE across the aforementioned outcome specifications and interference networks for long ($T = 200$) and short ($T = 40$) horizons. In each scenario, we perform 100 simulations of the synthetic experiment. The resulting distributions of ground truth and estimated TTEs are shown in Figures 2-5. All experiments were conducted on a MacBook Air with an Apple M1 chip and 16 GB of memory, with each setting taking about 15 minutes for 100 iterations. The key takeaways are as follows.

First, the DM and HT estimators exhibit significant bias across all cases. This is intuitive, as they estimate the TTE without accounting for the network interference.

Second, in the *LinearInMeans* outcome setting, FO-CMP and HO-CMP achieve low estimation error and minimal bias. This holds for both long experiment durations ($T = 200$), where outcomes reach equilibrium, and short experiment durations ($T = 40$), where outcomes have not yet reached equilibrium, as shown in Figures 2 and 3, respectively.

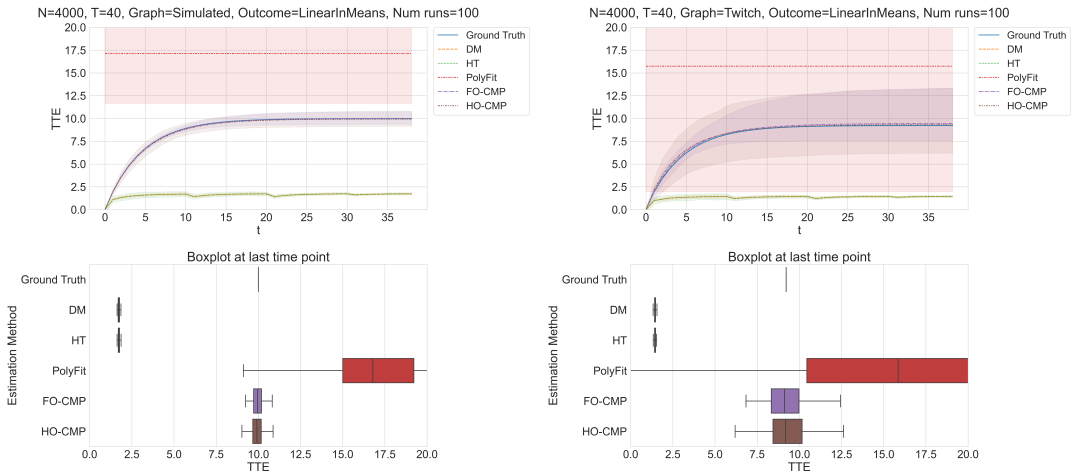
Third, as expected, PolyFit's dependence on accurate estimates for each $\nu_{\text{equil}}(\pi^{(\ell)})$ requires a large T to reduce estimation bias. This is evident when comparing Figures 2 and 3: with a smaller T, PolyFit shows bias. This is also demonstrated in panel (b) of Figure 1, where the red points—which represent sample averages of outcomes at $T^{(1)}, \ldots, T^{(4)}$—have not yet converged and are slightly lower than their ground truth (equilibrium) values. This causes PolyFit's estimation of $\hat{\nu}_{\text{equil}}(1)$ to be inaccurate, leading to a large bias. In contrast, HO-CMP, as shown in panel (c) of Figure 1, is immune to this problem as it is designed to work with off-equilibrium data. Even with a larger T, the degree-L polynomial estimation costs PolyFit with higher variance than both FO-CMP and HO-CMP, as shown in Figure 2. Overall, this underscores the more efficient data utilization of FO-CMP and HO-CMP through their ability to leverage off-equilibrium data.

Fourth, in the *Non-LinearInMeans* outcome setting, HO-CMP achieves substantially lower estimation error compared to FO-CMP, as shown in Figures 4 and 5. This makes intuitive sense, as the higher-



(a) Random graph (b) Twitch graph

Figure 2: *LinearInMeans with* $T = 200$. $L = 4$, with $(\pi^{(1)}, \pi^{(2)}, \pi^{(3)}, \pi^{(4)}) = (0.1, 0.2, 0.4, 0.5)$ and $T^{(\ell)} = 50\ell$ for all $\ell \in \{1, 2, 3, 4\}$. Shaded regions show 95% percentile intervals of mean.



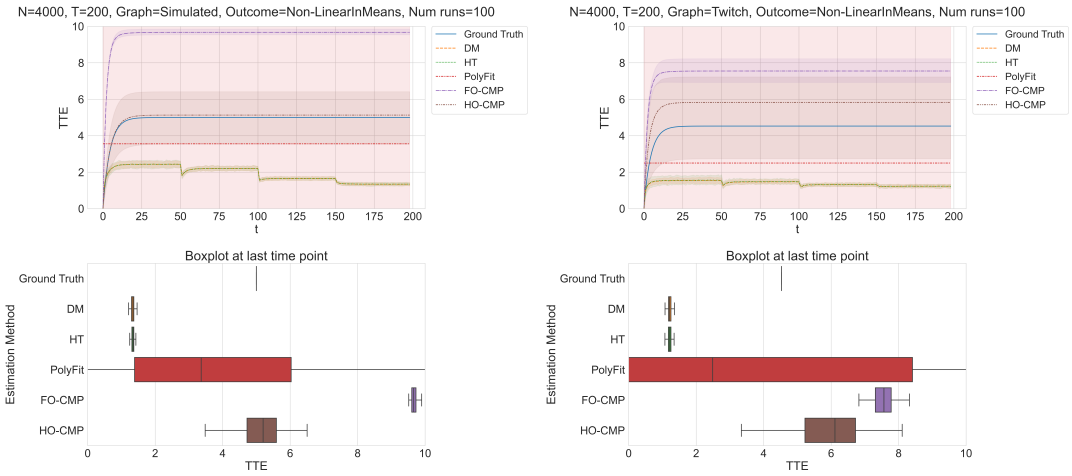
(a) Random graph (b) Twitch graph

Figure 3: *LinearInMeans with* $T = 40$. $L = 4$, with $(\pi^{(1)}, \pi^{(2)}, \pi^{(3)}, \pi^{(4)}) = (0.1, 0.2, 0.4, 0.5)$ and $T^{(\ell)} = 10\ell$ for all $\ell \in \{1, 2, 3, 4\}$. Shaded regions show 95% percentile intervals of mean.

order terms in HO-CMP better capture the nonlinearity of $\nu_{\text{equil}}(\pi)$ in π, while leveraging the additional data on sample variance dynamics over time, thereby enhancing the estimation accuracy.

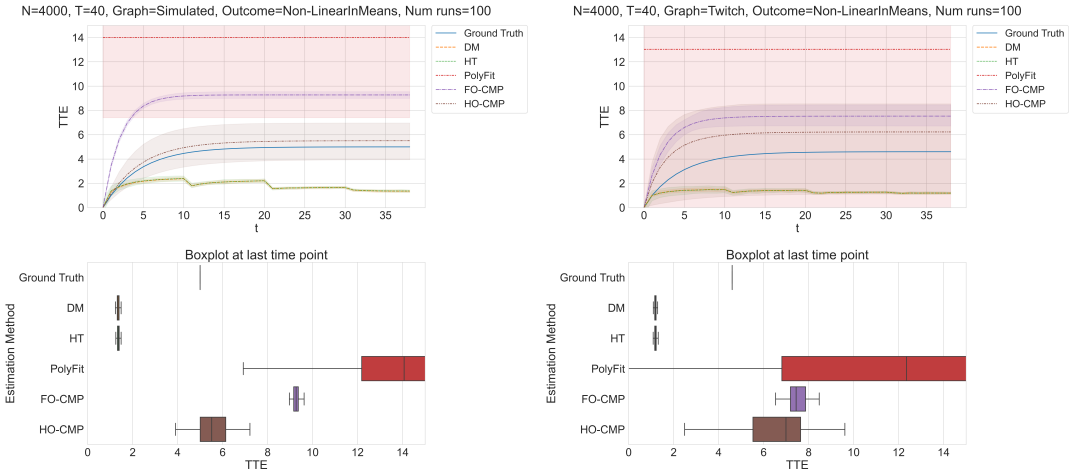
Finally, the proposed estimation method demonstrates robustness across different experimental setups, including both *LinearInMeans* and *Non-LinearInMeans* outcome specifications. Additionally, robustness to graph structure—random versus Twitch graph—is evident from comparing the left and right plots in Figures 2-5. In Figure 6 of Appendix A, we also demonstrate the robustness of the proposed methods to various parameters: the number of treatment probabilities L, the maximum treatment probability $\pi^{(L)}$, and the choice of experimental design (staggered rollout versus Bernoulli randomization).

https://doi.org/10.52202/079017-2600



(a) Random graph (b) Twitch graph

Figure 4: *Non-LinearInMeans with* $T = 200$. $L = 4$, with $(\pi^{(1)}, \pi^{(2)}, \pi^{(3)}, \pi^{(4)}) = (0.1, 0.2, 0.4, 0.5)$, and $T^{(\ell)} = 50\ell$ for all ℓ. Shaded regions show 95% percentile intervals of mean.



(a) Random graph (b) Twitch graph

Figure 5: *Non-LinearInMeans with* $T = 40$. $L = 4$, with $(\pi^{(1)}, \pi^{(2)}, \pi^{(3)}, \pi^{(4)}) = (0.1, 0.2, 0.4, 0.5)$, and $T^{(\ell)} = 10\ell$ for all ℓ. Shaded regions show 95% percentile intervals of mean.

5 Conclusion

Estimating causal effects under pervasive interference presents significant challenges [Sussman and Airoldi, 2017]. Building on the causal message-passing framework of Shirani and Bayati [2024], we incorporate higher-order moments of observed outcomes and treatment probabilities to estimate the total treatment effect, without requiring knowledge of the interference network. Our approach leverages machine learning techniques to extract informative patterns from these higher moments, enabling our estimator to capture complex counterfactual behaviors, including non-monotonic trends in outcome means relative to treatment proportions. While we demonstrate strong performance across various outcome specifications and network structures, the framework's applicability may be limited when multiple outcome observations are unavailable.

References

Anish Agarwal, Sarah Cen, Devavrat Shah, and Christina Lee Yu. Network synthetic interventions: A framework for panel data with network interference. *arXiv preprint arXiv:2210.11355*, 2022.

Dmitry Arkhangelsky and Guido Imbens. Causal models for longitudinal and panel data: A survey. Technical report, National Bureau of Economic Research, 2023.

Peter M Aronow and Cyrus Samii. Estimating average causal effects under general interference, with application to a social network experiment. 2017.

Eric Auerbach and Max Tabord-Meehan. The local approach to causal inference under network interference. *arXiv preprint arXiv:2105.03810*, 2021.

Guillaume W Basse and Edoardo M Airoldi. Limitations of design-based causal inference and a/b testing under arbitrary and network interference. *Sociological Methodology*, 48(1):136–151, 2018.

Guillaume W Basse, Avi Feller, and Panos Toulis. Randomization tests of causal effects under interference. *Biometrika*, 106(2):487–494, 2019.

Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57(2):764–785, 2011.

Alexandre Belloni, Fei Fang, and Alexander Volfovsky. Neighborhood adaptive estimators for causal inference under network interference. *arXiv preprint arXiv:2212.03683*, 2022.

Rohit Bhattacharya, Daniel Malinsky, and Ilya Shpitser. Causal inference under interference and network uncertainty. In *Uncertainty in Artificial Intelligence*, pages 1028–1038. PMLR, 2020.

Thomas Blake and Dominic Coey. Why marketplace experimentation is harder than it seems: the role of test-control interference. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, EC '14, page 567–582, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450325653. doi: 10.1145/2600057.2602837. URL `https://doi.org/10.1145/2600057.2602837`.

Erwin Bolthausen. An iterative construction of solutions of the tap equations for the sherrington–kirkpatrick model. *Communications in Mathematical Physics*, 325(1):333–366, 2014.

Robert M Bond, Christopher J Fariss, Jason J Jones, Adam DI Kramer, Cameron Marlow, Jaime E Settle, and James H Fowler. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415):295–298, 2012.

Ariel Boyarsky, Hongseok Namkoong, and Jean Pouget-Abadie. Modeling interference using experiment roll-out. *arXiv preprint arXiv:2305.10728*, 2023.

Ido Bright, Arthur Delarue, and Ilan Lobel. Reducing marketplace interference bias via shadow prices. *arXiv preprint arXiv:2205.02274*, 2022.

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

Ozan Candogan, Chen Chen, and Rad Niazadeh. Correlated cluster-based randomized experiments: Robust variance minimization. *Management Science*, 2023.

Qianyi Chen, Bo Li, Lu Deng, and Yong Wang. Optimized covariance design for ab test on social network under interference. *Advances in Neural Information Processing Systems*, 36, 2024.

Alex Chin. Central limit theorems via stein's method for randomized experiments under interference. *arXiv preprint arXiv:1804.03105*, 2018.

David Choi. Estimation of monotone treatment effects in network experiments. *Journal of the American Statistical Association*, 112(519):1147–1155, 2017.

 https://doi.org/10.52202/079017-2600

Mayleen Cortez, Matthew Eichhorn, and Christina Yu. Staggered rollout designs enable causal inference under interference without network knowledge. In *Advances in Neural Information Processing Systems*, 2022.

David Roxbee Cox. Planning of experiments. 1958.

David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

Dean Eckles, Brian Karrer, and Johan Ugander. Design and analysis of experiments in networks: Reducing bias from interference. *Journal of Causal Inference*, 5(1):20150021, 2016.

Vivek Farias, Andrew Li, Tianyi Peng, and Andrew Zheng. Markovian interference in experiments. *Advances in Neural Information Processing Systems*, 35:535–549, 2022.

Vivek F Farias, Hao Li, Tianyi Peng, Xinyuyang Ren, Huawei Zhang, and Andrew Zheng. Correcting for interference in experiments: A case study at douyin. *arXiv preprint arXiv:2305.02542*, 2023.

Laura Forastiere, Fabrizia Mealli, Albert Wu, and Edoardo M Airoldi. Estimating causal effects under network interference with bayesian generalized propensity scores. *Journal of Machine Learning Research*, 23(289):1–61, 2022.

Christopher Harshaw, Fredrik Sävje, David Eisenstat, Vahab Mirrokni, and Jean Pouget-Abadie. Design and analysis of bipartite experiments under a linear exposure-response model. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, EC '22, page 606, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450391504. URL `https://doi.org/10.1145/3490486.3538269`.

Paul W Holland. Statistics and causal inference. *Journal of the American statistical Association*, 81 (396):945–960, 1986.

David Holtz, Ruben Lobel, Inessa Liskovich, and Sinan Aral. Reducing interference bias in online marketplace pricing experiments. *arXiv preprint arXiv:2004.12489*, 2020.

Daniel G Horvitz and Donovan J Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, pages 663–685, 1952.

Michael G Hudgens and M Elizabeth Halloran. Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842, 2012.

Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.

Matthew O Jackson, Zhongjian Lin, and Ning Neil Yu. Adjusting for peer-influence in propensity scoring when estimating treatment effects. *Available at SSRN 3522256*, 2020.

Ravi Jagadeesan, Natesh S Pillai, and Alexander Volfovsky. Designs for estimating the treatment effect in networks with interference. 2020.

Su Jia, Nathan Kallus, and Christina Lee Yu. Clustered switchback experiments: Near-optimal rates under spatiotemporal interference, 2024.

Ramesh Johari, Hannah Li, Inessa Liskovich, and Gabriel Y Weintraub. Experimental design in two-sided platforms: An analysis of bias. *Management Science*, 68(10):7069–7089, 2022.

Hyunseung Kang and Guido Imbens. Peer encouragement designs in causal inference with partial interference and identification of local average network effects. *arXiv preprint arXiv:1609.04464*, 2016.

Vishesh Karwa and Edoardo M Airoldi. A systematic investigation of classical causal inference strategies under mis-specification due to network interference. *arXiv preprint arXiv:1810.08259*, 2018.

Ron Kohavi, Diane Tang, and Ya Xu. *Trustworthy online controlled experiments: A practical guide to a/b testing*. Cambridge University Press, 2020.

Michael P Leung. Treatment and spillover effects under network interference. *Review of Economics and Statistics*, 102(2):368–380, 2020.

Michael P Leung. Causal inference under approximate neighborhood interference. *Econometrica*, 90 (1):267–293, 2022.

Shuangning Li and Stefan Wager. Network interference in micro-randomized trials. *arXiv preprint arXiv:2202.05356*, 2022a.

Shuangning Li and Stefan Wager. Random graph asymptotics for treatment effect estimation under network interference. *The Annals of Statistics*, 50(4):2334–2358, 2022b.

Lan Liu and Michael G Hudgens. Large sample randomization inference of causal effects in the presence of interference. *Journal of the american statistical association*, 109(505):288–301, 2014.

Charles F Manski. Nonparametric bounds on treatment effects. *The American Economic Review*, 80 (2):319–323, 1990.

Charles F Manski. Identification of treatment response with social interactions. *The Econometrics Journal*, 16(1):S1–S23, 2013.

M Mezard, G Parisi, and M Virasoro. *Spin Glass Theory and Beyond, An Introduction to the Replica Method and Its Applications*. World Scientific, Paris, Roma, November 1986. doi: 10.1142/0271.

Marc Mezard and Andrea Montanari. *Information, physics, and computation*. Oxford University Press, 2009.

Evan Munro, Stefan Wager, and Kuang Xu. Treatment effects in market equilibrium. *arXiv preprint arXiv:2109.11647*, 2021.

Zhaonan Qu, Ruoxuan Xiong, Jizhou Liu, and Guido Imbens. Efficient treatment effect estimation in observational studies under heterogeneous partial interference. *arXiv preprint arXiv:2107.12420*, 2021.

Paul R Rosenbaum. Interference between units in randomized experiments. *Journal of the american statistical association*, 102(477):191–200, 2007.

Benedek Rozemberczki and Rik Sarkar. Twitch gamers: a dataset for evaluating proximity preserving and structural role-based node embeddings, 2021.

Donald B Rubin. Bayesian inference for causal effects: The role of randomization. *The Annals of statistics*, pages 34–58, 1978.

Fredrik Sävje, Peter Aronow, and Michael Hudgens. Average treatment effects in the presence of unknown interference. *Annals of statistics*, 49(2):673, 2021.

Sadegh Shirani and Mohsen Bayati. Causal message-passing for experiments with unknown and general network interference. *Proceedings of the National Academy of Sciences*, 121(40):e2322232121, 2024.

Michael E Sobel. What do randomized studies of housing mobility demonstrate? causal inference in the face of interference. *Journal of the American Statistical Association*, 101(476):1398–1407, 2006.

Daniel L Sussman and Edoardo M Airoldi. Elements of estimation theory for causal effects in the presence of network interference. *arXiv preprint arXiv:1702.03578*, 2017.

Eric J Tchetgen Tchetgen and Tyler J VanderWeele. On causal inference in the presence of interference. *Statistical methods in medical research*, 21(1):55–75, 2012.

Johan Ugander and Hao Yin. Randomized graph cluster randomization. *Journal of Causal Inference*, 11(1):20220014, 2023.

Davide Viviano. Experimental design under network interference. *arXiv preprint arXiv:2003.08421*, 2020a.

 https://doi.org/10.52202/079017-2600

Davide Viviano. Policy design in experiments with unknown interference. *arXiv preprint arXiv:2011.08174*, 2020b.

Davide Viviano, Lihua Lei, Guido Imbens, Brian Karrer, Okke Schrijvers, and Liang Shi. Causal clustering: design of cluster experiments under network interference. *arXiv preprint arXiv:2310.14983*, 2023.

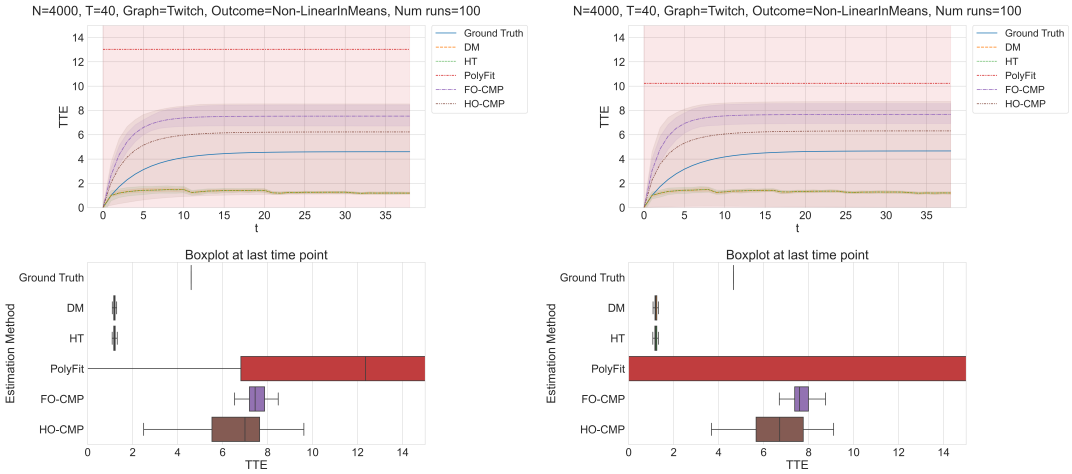
Stefan Wager and Kuang Xu. Experimenting in equilibrium. *Management Science*, 67(11):6694–6715, 2021.

Ye Wang, Cyrus Samii, Haoge Chang, and PM Aronow. Design-based inference for spatial experiments with interference. *arXiv preprint arXiv:2010.13599*, 2020.

Ruoxuan Xiong, Susan Athey, Mohsen Bayati, and Guido Imbens. Optimal experimental design for staggered rollouts. *Management Science*, 70(8):5317–5336, 2024.

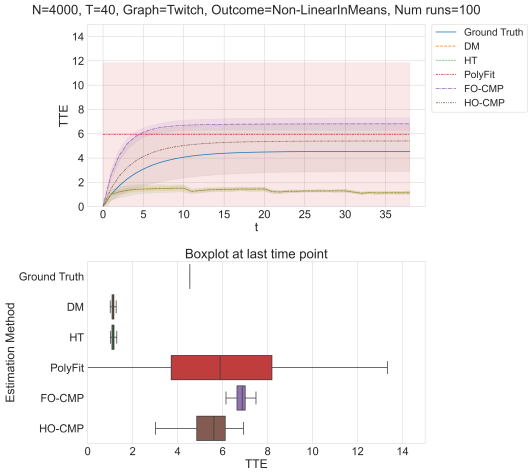
Christina Lee Yu, Edoardo M Airoldi, Christian Borgs, and Jennifer T Chayes. Estimating the total treatment effect in randomized experiments with unknown network structure. *Proceedings of the National Academy of Sciences*, 119(44):e2208975119, 2022.

A Supplementary Experiments

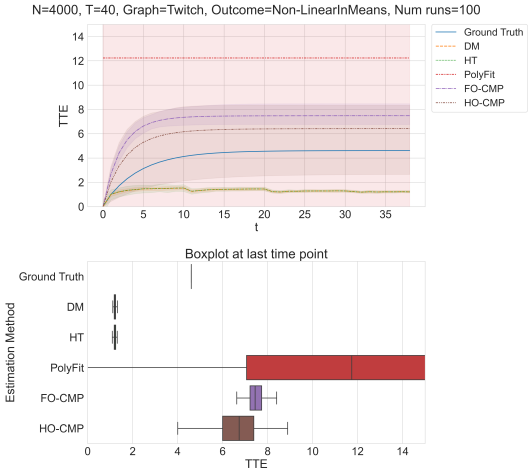


(a) The same setting as Figure 5b

(b) Increasing number of π's



(c) Increasing the largest π



(d) Bernoulli randomized design

Figure 6: *Robustness check, under the Non-LinearInMeans with Twitch graph and* $T = 40$*, i.e., setting of Figure 5b.* (a): Original Figure 5b. (b): Increasing L: i.e., $\pi^{(\ell)} = 0.1\ell$ and $T^{(\ell)} = 8\ell$ for all $\ell \in \{1, \ldots, 5\}$. (c): Increasing treatment probabilities, i.e., $(\pi^{(1)}, \pi^{(2)}, \pi^{(3)}, \pi^{(4)}) = (0.1, 0.2, 0.4, 0.6)$. (d): Using Bernoulli randomized design.

https://doi.org/10.52202/079017-2600

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification:

   Guidelines: Yes, the theoretical results in Sections 2 and 3 and the empirical results in Section 4 and Appendix A support the main claims made the abstract and introduction.

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss the limitations of the work in Section 5.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: All the assumptions are clearly stated in Sections 2 and 3.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: All necessary information for reproducing the main experimental results are stated in Section 4.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

https://doi.org/10.52202/079017-2600

Answer: [Yes]

Justification: We will provide open access to the data and code.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: All the details of the experiments are provided in Section 4.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification:

   Guidelines: All of our experiments are replicated for 100 times and the error bars are reported.

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).

- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We provide the details of the computer resources in Section 4.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification:

   Guidelines: The research conducted in the paper adheres to the NeurIPS Code of Ethics.

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: We discuss broader impacts in Section 1.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

https://doi.org/10.52202/079017-2600

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer:[NA]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We cite the real network graph in Section 4.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: We will release the code with documentation.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

https://doi.org/10.52202/079017-2600