Network Experiments under Dynamic Treatment Diffusion: Design and Estimation for Bias Reduction

Qinyang Yu (MSEC' 26), Weihang Liu (MAE' 26)

COMPSCI 590 Project Final Report

April 29, 2025

Abstract

This paper proposes a comprehensive framework for designing and analyzing network experiments under treatment diffusion dynamics, extending the work of Eckles et al. (2017). We focus on improving treatment assignment design and estimation strategies to reduce bias in estimating the average treatment effect (ATE). The framework consists of five phases: initialization, treatment assignment, treatment diffusion, outcome generation, and estimation. We introduce the Independent Cascade Model (ICM) as the key innovation for modeling treatment diffusion and evaluate two treatment assignment designs—independent random assignment and graph cluster randomization—along with three estimators. Our main result shows that neighborhood-based estimators combined with clustered assignment effectively reduce bias, improve accuracy, and remains robust, particularly when treatment or spillover effects are strong. Additionally, we extend our framework by analyzing sensitivity to diffusion speed and exploring alternative diffusion patterns using the Linear Threshold Model (LTM).

Keywords: Treatment Diffusion, Network Experiments, Average Treatment Effect (ATE), Graph Cluster Randomization, Neighborhood-Based Estimators, Independent Cascade Model (ICM), Bias Reduction

1 Introduction

In the context of epidemics, governments often launch vaccine awareness campaigns to promote public health. A common approach is to randomly vaccinate a subset of individuals in a population through a pilot study. However, in densely connected communities, the impact of vaccinating one individual may extend far beyond that person due to both biological and behavioral spillovers. For instance, vaccinating a child in a school not only protects them but can also reduce disease transmission to their classmates, indirectly safeguarding the unvaccinated. Furthermore, individuals may update their beliefs or behaviors upon observing the vaccination status of others, leading to diffusion of vaccine adoption through peer influence, trust, or encouragement. Over time, as vaccinated individuals interact with others, the outcome of interest—such as infection status or health-seeking behavior—evolves dynamically, influenced by both direct protection and social contagion.

Traditional causal inference methods fail to account for these intertwined dynamics of interference and treatment diffusion, limiting their applicability in networked settings like vaccine campaigns. Existing literature primarily addresses interference in networked experiments without considering treatment diffusion. To address this gap, we propose a comprehensive framework for designing and analyzing network experiments that account for treatment diffusion dynamics, extending the work of Eckles et al. (2017).

Our framework consists of five phases: initialization, treatment assignment, treatment diffusion, outcome generation, and estimation. Beyond using a small-world network model to simulate realistic social structures and a classical linear-in-means model for outcome generation, we incorporate the Independent Cascade Model (ICM) in the experiment as our key innovation for modeling treatment diffusion. We evaluate two treatment assignment designs—independent random assignment and graph cluster randomization—along with three estimators: the individual-level difference-in-means estimator, and two neighborhood-based estimators. Extensive simulation allows us to explore how these approaches perform in reducing bias and estimating the average treatment effect (ATE).

The main contribution of our study is incorporating treatment diffusion dynamics in networked experimental design. Our core results demonstrate that combining graph cluster

randomization with neighborhood-based estimators effectively reduces bias, improves accuracy, and remains robust compared to other methods, particularly when treatment effects or spillover effects are strong. Our findings confirm and extend the robustness of Eckles et al. (2017)'s conclusions in the presence of treatment diffusion. We further extend our framework in three important ways: (i) by analyzing sensitivity to diffusion speed, and (ii) by exploring alternative diffusion patterns using the Linear Threshold Model (LTM).

This study makes significant contributions to the growing field of network-based causal inference. By providing a framework for bias reduction in the presence of both treatment diffusion and spillover effects, we aim to offer a clearer understanding of how to capture these dynamics in randomized experiments. Our work ensures that policy interventions in networked environments are evaluated accurately and meaningfully, providing more reliable evidence for decision-making in public health and similar areas.

2 Related Work

Randomized experiments in networks pose fundamental identification challenges due to *interference*, the phenomenon where one unit's treatment can influence another unit's outcome (Cox, 1958). Classical experimental designs typically rely on the Stable Unit Treatment Value Assumption (SUTVA) (Rubin, 1980), which assumes no interference between units and forms the basis for much of the standard causal inference literature. However, in networked settings, SUTVA is systematically violated, leading to potential bias or misinterpretation of traditional estimands.

A growing literature in network-based causal inference has sought to disentangle the various dimensions of interference by examining *spillover effects*—how a focal unit's outcome depends on the treatment status of its neighbors (Aronow and Samii, 2017; Basse and Feller, 2018; Sävje et al., 2017). To address this, a growing body of literature has developed methods, such as exposure mappings, partial interference assumptions, and network-based cluster randomization, to account for dependence across units in networked settings. Specifically, to formally incorporate interference into potential outcomes, Aronow and Samii (2017) introduced the framework of exposure mappings, which define potential outcomes based on

treatment configurations within a unit's neighborhood. Sävje et al. (2017) extended this by developing randomization-based estimators for average treatment effects under general interference, without assuming knowledge of the interference structure. Basse and Feller (2018) advanced a two-stage estimation strategy that first recovers exposure-specific outcomes and then aggregates them to estimate average causal effects under interference. To improve efficiency, Park and Kang (2022) introduced a semi-parametrically efficient estimator using targeted minimum loss estimation under partial interference. Papadogeorgou et al. (2019) applied marginal structural models to account for time-varying exposures and dynamic interference in longitudinal settings. On the design front, Ugander and Yin (2023) studied optimal network-based experimental designs that balance power and spillover control.

Despite these advances, most approaches assume fixed treatments and pre-specified static exposure, limiting their relevance for settings with dynamic *treatment diffusion*, where adoption spreads endogenously over time via peer influence. An (2018) studies treatment diffusion in a experimental design but view it as a concrete form of treatment interference. In our study, we explicitly distinguish between treatment spillover effects, which occur when the treatment of one individual directly affects the outcomes of others, and treatment diffusion, which emerge when individuals adopt treatment behaviors due to the influence of others within their network. This distinction is crucial for understanding how both the direct and indirect effects of treatment evolve over time and influence the overall population, particularly when the network structure itself plays a central role in how individuals interact and respond to treatment.

This paper contributes to bridging these methodological and substantive gaps by offering a unified framework that simultaneously accounts for treatment diffusion and spillover in a dynamic network setting. This integrated perspective enhances both the internal validity of causal estimates and the external validity of empirical findings across diverse application domains—from public health interventions to technology adoption in economic networks. The framework enables policymakers and researchers to predict not only immediate causal effects but also longer-term consequences as treatments propagate through social systems, thereby improving welfare calculations for networked interventions.

3 Theoretical Framework

In this section, we present the theoretical model for networked experimental design under treatment diffusion dynamics. The experimentation primarily consists of five phases: (i) initialization, (ii) treatment assignment, (iii) treatment diffusion, (iv) outcome generation, and (v) estimation.

3.1 Initialization

The initial phase includes network formation and characteristic generation before the experiment, which can be random or fixed. In this study, we generate a random network from small-world models (Watts and Strogatz, 1998), which is detailed in Section 4.

Consider a population of N agents. After the initialization, all the agents are connected through the network structure, represented as an undirected and unweighted graph $G(V, E)$, where each node $v \in V$ corresponds to an individual and each edge $e \in E$ represents a connection between two agents. Let $A \in \{0, 1\}^{n \times n}$ denote the adjacency matrix and $\mathcal{N}(i)$ denote the set of neighbors of node i.

3.2 Treatment Assignment

In our experiment design, we consider a binary (i.e., an "A/B" test) and irreversible treatment for simplicity (consider the example in the introduction). Let $t \in \{0, 1, \ldots, T\}$ be discrete time steps. Let $Z_i(t) \in \{0, 1\}$ denote the treatment status of node i at time t, where 1 represents treatment and 0 represents control. Treatment assignment determines the status of initial treatment $Z_i(0)$, which is a mapping from vertices V to the binary treatment status space $\{0, 1\}$.

We primarily consider two treatment assignment approaches in this study, following Eckles et al. (2017). The first approach is *independent random assignment*, which is commonly used in experimental design. This assignment ensures an individual's assignment is uncorrelated with others' assignments. Specifically, each individual's initial treatment is a random

variable independently drawn from a Bernoulli distribution, i.e.,

$$Z_i(0) \sim \text{Bernoulli}(p_0),$$

where p_0 is the probability of assignment to the treatment.

The second approach is *graph cluster randomization* proposed by Ugander et al. (2013), which allows assignment with network autocorrelation. Specifically, the treatment assignment is achieved at a cluster level. We partition all the vertices into M clusters $\{C_1, C_2, ..., C_M\}$ and define $C(\cdot) : \{1, ..., N\} \to \{1, ..., M\}$ as a mapping from vertex indices to cluster indices. Essentially, $C(i)$ is the index of cluster containing vertex i. In standard graph cluster randomization, each cluster C_j's initial treatment W_j is a random variable independently drawn from a Bernoulli distribution and each individual's initial treatment is subsequently determined, i.e.,

$$W_j \sim \text{Bernoulli}(q_0), \quad Z_i(0) = W_{C(i)}.$$

Following Ugander et al. (2013), the mapping $C(\cdot)$ is formed by an ϵ-net clustering, where any two nodes within the same cluster are within a network distance of at most ϵ, and every node is within ϵ of at least one cluster center. This clustering approach ensures that locally connected nodes are grouped together, allowing cluster-level treatment assignment to better align with the underlying network structure and thereby reduce interference between treated and control units. Other clustering mapping includes community detection (Fortunato, 2010), which may introduced too large variance in bias reduction and is thus avoided (Eckles et al., 2017).

To provide an intuition for graph cluster randomization, this approach addresses interference in network experiments by assigning treatment at the cluster level rather than the individual level. Instead of randomizing nodes independently—which often leaves treated and control units mixed among neighbors and introduces bias—clustered random assignment groups closely connected nodes into clusters and randomizes treatment across these groups. This ensures that treated units are mostly surrounded by other treated units, and control units by other controls, creating cleaner exposure environments. By aligning treatment assignment with network structure, clustered random assignment can sometimes substantially

reduce bias caused by spillover effects while preserving the core validity of experimental comparisons.

In later simulation, we will compare the estimation results formed by independent random assignment and clustered random assignment using ϵ-nets, respectively.

3.3 Treatment Diffusion

This section constitutes the core of our study and extends the framework proposed by Eckles et al. (2017). In their setting, treatment status is deterministic following the initial assignment and remains static throughout the experiment. In other words, $Z_i(t) = Z_i(0)$ for any $t \geq 1$. In our study, we relax this assumption of constant treatment status and introduce dynamics by allowing treatment diffusion. At time $t \geq 1$, treatment evolves over time via a transition rule

$$Z_i(t) = \max\left(Z_i(t-1),\ f_i\left(Z_{\mathcal{N}(i)}(t-1)\right)\right),$$

where $f_i(\cdot)$ is a general diffusion function. According to Kempe et al. (2003), the Independent Cascade Model (ICM) and the Linear Threshold Model (LTM) are two of the most basic and widely-studied diffusion models. Therefore, we have two candidates for the diffusion function $f_i(\cdot)$: the probabilistic diffusion function and threshold-based diffusion function.

In this study, we adopt ICM as the baseline diffusion process for two main reasons. First, ICM is a simple, interpretable, and widely-used stochastic diffusion model that captures independent neighbor-to-neighbor influence and serves as a natural starting point for introducing dynamic treatment evolution. Second, and more importantly for our experimental design, ICM allows direct control over diffusion speed through the activation probability parameter p. This property enables a clean comparison between random and clustered initial treatment assignments by varying p uniformly across the network. In contrast, the Linear Threshold Model (LTM) introduces threshold-based cumulative adoption dynamics, where diffusion speed becomes highly sensitive to the spatial arrangement of initially treated nodes. Under LTM, clustered random assignments can lead to dramatically different diffusion patterns that are harder to standardize, complicating controlled comparisons. Therefore, we use ICM as a controlled baseline model in Section 4 and extend to LTM to capture more

realistic but less easily controlled diffusion phenomena in Section 5.2.

In ICM, the probability of transition is defined as

$$p = \mathbb{P}\left(Z_i(t) = 1 \mid Z_i(t-1) = 0\right) = 1 - \prod_{j \in \mathcal{N}(i)} \left(1 - p_{ji} \cdot Z_j(t-1)\right),$$

where the parameter p_{ji} is a latent variable representing the probability that node j, if active at time $t-1$, will successfully infect node i at time t. In this case, f_i can be interpreted as a Bernoulli random variable, indicating whether node i becomes active at time t given the activity of its neighbors at $t-1$. Formally,

$$f_i\left(Z_{\mathcal{N}(i)}(t-1)\right) \sim \text{Bernoulli}(p).$$

3.4 Outcome Generation

Given the dynamic treatment diffusion, the observed outcomes also evolve over time. Let $Y_i^*(t) \in \mathbb{R}$ be the latent continuous potential outcome of node i at time t. Under a full treatment history for all nodes in the network, the continuous potential outcome depends not only on its own treatment but also on the evolving past outcomes of its neighbors (interference). Formally, the functional form of the continuous potential outcome can be specified as

$$Y_i^*(t) = h_i(Z_i(t), Y_{\mathcal{N}(i)}(t-1)).$$

To simplify the structure and achieve a concrete representation, we adopt a *linear-in-means* model (Bramoullé et al., 2009; Manski, 1993), which is a widely used and interpretable specification to capture both the direct effect of own treatment status and the indirect effect through peer influence via neighbors' past outcomes. To account for time-varying treatment dynamics, the treatment status $Z_i(t) \in \{0, 1\}$ of unit i at time t is determined endogenously via a treatment diffusion process as introduced in Section 3.3. Moreover, this linear-in-means framework also supports extensions to heterogeneous effects and more flexible functional forms (e.g., non-linear, learned exposure functions, multiple lagged terms).

Specifically, at each discrete time step t, the latent continuous outcome $Y_i^*(t)$ for node i

is generated according to the following structural equation:

$$Y_i^*(t) = \alpha + \beta Z_i(t) + \gamma \frac{A_i' Y(t-1)}{k_i} + \varepsilon_i(t),$$

where α denotes the baseline intercept term, β captures the direct effect of receiving the treatment, and γ measures the peer effect, represented by the average behavior of node i's neighbors at the previous time step $t-1$. Here, A_i' is the i-th row of the adjacency matrix A, k_i is the degree (number of neighbors) of node i, and $\varepsilon_i(t)$ is an idiosyncratic noise term independently drawn from a normal distribution $\mathcal{N}(0,1)$.

The observed outcome $Y_i(t)$ is then a deterministic transformation of the latent variable, defined as

$$Y_i(t) = a\left(Y_i^*(t)\right),$$

where $a(\cdot)$ is a link function. In the main analysis, we focus on the binary outcome case by using a threshold function $a(x) = \mathbf{1}\{x > 0\}$, where $\mathbf{1}\{\cdot\}$ is the indicator function. That is, an agent adopts the behavior if and only if the latent utility surpasses zero.

This specification corresponds to a standard probit model, where the observed binary outcome reflects whether the underlying latent utility exceeds a fixed threshold. In this setting, the stochastic component $\varepsilon_i(t)$ follows a normal distribution, and the probability of adoption at time t depends on both the treatment status and the average behavior of neighbors at the previous time step.

3.5 Estimation

We focus on estimating the average treatment effect (ATE) under the treatment diffusion dynamics described above. The ATE is defined as the expected difference in outcomes when all individuals are assigned to treatment versus when all individuals are assigned to control, namely,

$$\tau = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left[Y_i(Z=1) - Y_i(Z=0)\right],$$

where $Z = 1$ and $Z = 0$ denote the vectors corresponding to global treatment and global control assignments, respectively. This representation of the ATE is naturally of interest, as

it captures whether applying the treatment to all units would yield beneficial outcomes.

After the final time step T, the treatment diffusion process and outcome evolution are complete, and we proceed to estimate τ based on the realized treatment statuses $Z_i(T)$ and observed outcomes $Y_i(T)$. Following standard approaches in network experiments (Eckles et al., 2017), we consider three estimators that differ in how they address interference and treatment diffusion.

The first estimator is the simple *individual-level difference-in-means* estimator, which ignores network structure and computes the outcome difference between treated and untreated units. Formally, it is given by

$$\widehat{\tau}_{\text{ITR},S} = \frac{1}{|\{i : Z_i(T) = 1\}|} \sum_{i:Z_i(T)=1} Y_i(T) - \frac{1}{|\{i : Z_i(T) = 0\}|} \sum_{i:Z_i(T)=0} Y_i(T).$$

This estimator implicitly assumes an individualistic treatment response (ITR) model where SUTVA is true and may suffer from bias when interference is present.

To mitigate interference-induced bias, we consider alternative estimators leveraging a fractional neighborhood exposure model, which assumes a fractional relaxation of neighborhood treatment response (FNTR). Under this approach, an individual is considered effectively treated if a sufficient fraction λ of its neighbors are also treated, and effectively control if the fraction of treated neighbors falls below a complementary threshold. Let $\mathcal{T}(\lambda)$ and $\mathcal{C}(\lambda)$ denote the sets of effectively treated and effectively control individuals, respectively, under a threshold $\lambda \in [0, 1]$. Our second estimator is thus the *simple unweighted neighborhood-based* estimator

$$\widehat{\tau}_{\text{FNTR},S} = \frac{1}{|\mathcal{T}(\lambda)|} \sum_{i\in\mathcal{T}(\lambda)} Y_i(T) - \frac{1}{|\mathcal{C}(\lambda)|} \sum_{i\in\mathcal{C}(\lambda)} Y_i(T).$$

By focusing on individuals who are exposed to more homogeneous treatment environments, this estimator aims to reduce bias at the cost of potentially increased variance.

Finally, to further account for the fact that the probability of satisfying the effective treatment condition may vary across individuals, we employ an inverse-probability-weighted estimator based on Hajek weighting. Let $\pi_i(1)$ and $\pi_i(0)$ denote the probabilities that individual i is effectively treated or effectively control, respectively. The Hajek estimator is

given by

$$\widehat{\tau}_{\text{FNTR},H} = \left(\sum_{i=1}^{n} \frac{\mathbf{1}\{i \in \mathcal{T}(\lambda)\}}{\pi_i(1)}\right)^{-1} \sum_{i=1}^{n} \frac{Y_i(T)\mathbf{1}\{i \in \mathcal{T}(\lambda)\}}{\pi_i(1)} - \left(\sum_{i=1}^{n} \frac{\mathbf{1}\{i \in \mathcal{C}(\lambda)\}}{\pi_i(0)}\right)^{-1} \sum_{i=1}^{n} \frac{Y_i(T)\mathbf{1}\{i \in \mathcal{C}(\lambda)\}}{\pi_i(0)}.$$

This estimator re-weights observations to correct for differential exposure probabilities, thereby further mitigating bias at the expense of greater estimator variance.

Throughout our analysis, we compare the performance of these three estimators under both independent and clustered random assignments. The relative trade-offs between bias reduction and variance inflation are evaluated in Section 4.

4 Simulation Study

4.1 Simulation Procedure

The simulation study is conducted on synthetic networks generated using the Watts–Strogatz (WS) small-world model, which captures both the high clustering and short average path lengths commonly observed in real-world social networks (Eckles et al., 2017; Watts and Strogatz, 1998). We construct networks with $n = 1000$ nodes, each initially connected to $k = 10$ nearest neighbors in a ring topology. To introduce randomness, each edge is rewired with independent probability $\beta_0 = 0.4$. The choice of β_0 reflects a balance between the regularity of a lattice and the randomness of an Erdős–Rényi graph.

Initial treatments are assigned at time $t = 0$ according to either independent random assignment or graph cluster randomization. Under independent random assignment, each node independently receives treatment with probability $p_0 = 0.4$, chosen to approximate the average treatment proportion generated under clustered randomization. Under graph cluster randomization, nodes are partitioned into clusters using an ϵ-net clustering algorithm with distance parameter $\epsilon = 3$, ensuring that connected nodes are grouped together based on their proximity in the network topology (Figure 1). Treatment is then assigned at the clus-

ter level via random Bernoulli draws, inducing network autocorrelation in initial treatment patterns without fixing a priori treatment probabilities for individual nodes. Figure 2 shows the assignment results for both independent and clustered randomization. Compared to independent assignment, clustered assignment results in a more locally clustered treatment distribution, and utilizes network structure to the fullest extent, effectively separating the treatment group from the control group and minimizing the interference confounding.

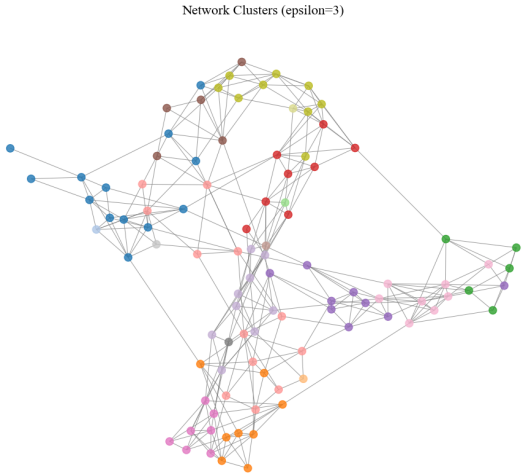

Figure 1: Network clusters generated by ϵ-Net clustering ($\epsilon = 3$)

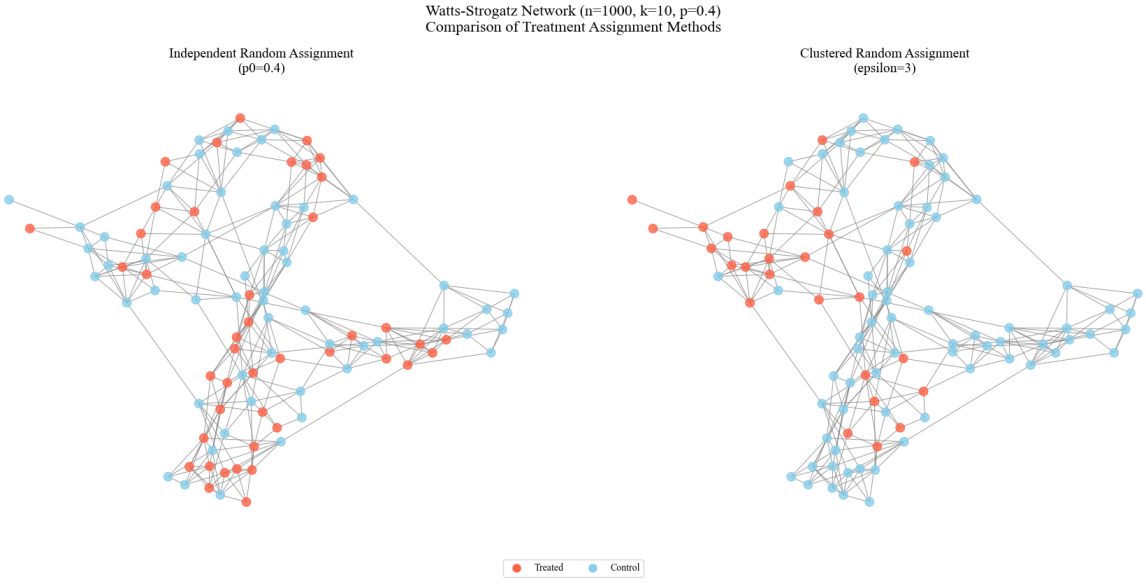

Figure 2: Treatment assignment design (Independent vs Clustered)

Following initial assignment, treatment diffusion unfolds over $T = 5$ discrete periods according to either the Independent Cascade Model (ICM) or the Linear Threshold Model (LTM). In the baseline model, we adopt the ICM, where treated nodes attempt to activate each of their untreated neighbors independently at each time step, with activation probabilities p_{ij} drawn uniformly at random from the interval $[0, 0.01]$ for each directed edge (i, j). This stochastic activation mechanism ensures heterogeneity in peer influence across the network. The choice of an upper bound of 0.01 is designed to maintain moderate diffusion intensity over the simulation horizon, with the cumulative proportion of treated nodes staying below 50% after five periods. This controlled setting facilitates clean identification of direct and spillover effects without overwhelming saturation and will be systematically varied to assess sensitivity to diffusion strength in the later extension.

Figure 3 illustrates the treatment diffusion process under ICM, comparing independent and clustered initial assignment methods. As shown, both strategies lead to a steady but gradual increase in the proportion of treated nodes over time, ultimately reaching similar treatment coverage levels by $T = 5$. This similarity in diffusion speed under ICM establishes a controlled baseline for causal effect estimation. In Section 5.2, we will explore alternative diffusion dynamics using LTM, where clustered assignment is expected to generate more rapid and concentrated diffusion patterns due to the threshold-based nature of activation.

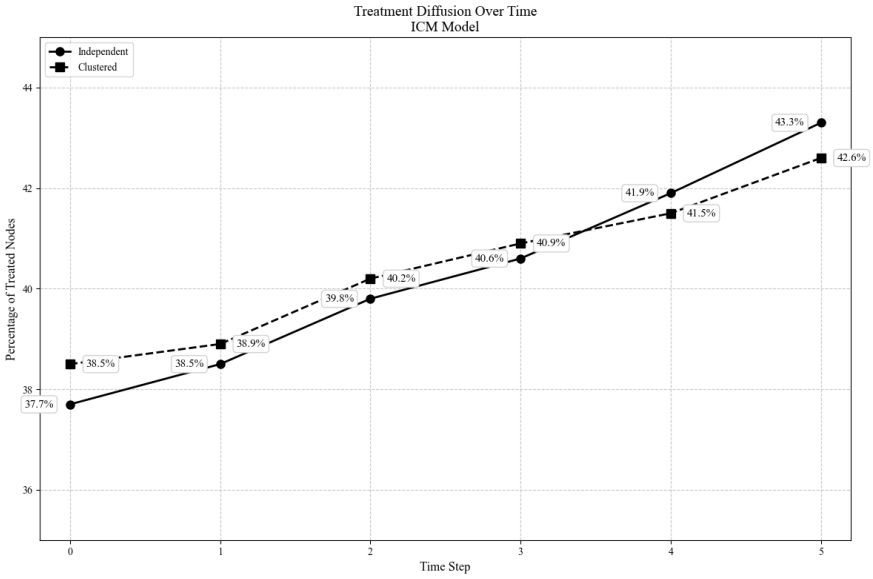

Figure 3: Treatment diffusion over time under the Independent Cascade Model (ICM)

Next, we simulate individual outcomes using a probit specification, setting the baseline intercept at $\alpha = -1.5$ to ensure relatively infrequent behavior:

$$Y_i^*(t) = -1.5 + \beta Z_i(t) + \gamma \frac{A_i' Y(t-1)}{k_i} + \varepsilon_i(t), \qquad Y_i(t) = \mathbf{1}\{Y_i^*(t) > 0\},$$

where $\varepsilon_i(t) \sim \mathcal{N}(0,1)$ denotes an idiosyncratic shock independently drawn at each time step.

Prior to treatment diffusion, all vertices are initialized with $Y_i(0) = 0$, representing an initial state of non-adoption. We conduct simulations across all combinations of $\beta \in \{0.1, 0.5, 0.9\}$ and $\gamma \in \{0.1, 0.5, 0.9\}$, allowing systematic variation in the strength of direct treatment effects and peer spillovers. Treatment diffusion is carried out over a maximum horizon of $T = 5$ periods under ICM, after which outcomes are generated according to the above structural equation.

Following the completion of treatment diffusion and outcome evolution at time T, we simulate the ground truth ATE and estimate it using three approaches. The first is the individual-level unweighted estimator ($\widehat{\tau}_{\text{ITR},S}$) which compares the average outcomes between treated and untreated nodes without adjusting for network interference. The second is the unweighted fractional neighborhood treatment response (FNTR) estimator ($\widehat{\tau}_{\text{FNTR},S}$) which classifies nodes as effectively treated or effectively control based on whether the fraction of treated neighbors exceeds a threshold $\lambda = 0.75$ following Ugander et al. (2013), thereby leveraging local network structure to mitigate bias. The third is the Hajek-weighted FNTR estimator ($\widehat{\tau}_{\text{FNTR},H}$) which further adjusts for heterogeneous probabilities of effective treatment exposure by applying inverse probability weights. For each simulation configuration, we compute all three estimators under both independent and clustered random assignment designs, performing 100 repetitions per scenario to ensure statistical stability. To approximate the ground truth ATE, we separately simulate outcomes under full population treatment and full population control, conducting 100 repetitions for each counterfactual setting.

4.2 Simulation Results

We evaluate estimator performance by examining the bias and root mean squared error (RMSE) of the estimated ATE under independent and clustered assignment designs. Figure 4

displays the evolution of estimator bias and RMSE over time steps $t = 1, \ldots, 5$ in a dynamic treatment diffusion setting. While the full set of results covers all combinations of direct and peer effects ($\beta, \gamma \in \{0.1, 0.5, 0.9\}$), we present here the representative case of $\beta = 0.5$ and $\gamma = 0.5$, which balances the influence of own treatment and neighbor spillovers. The unweighted and Hajek-weighted FNTR estimator yield nearly indistinguishable results across both bias and RMSE metrics, with their curves closely overlapping throughout all time steps.

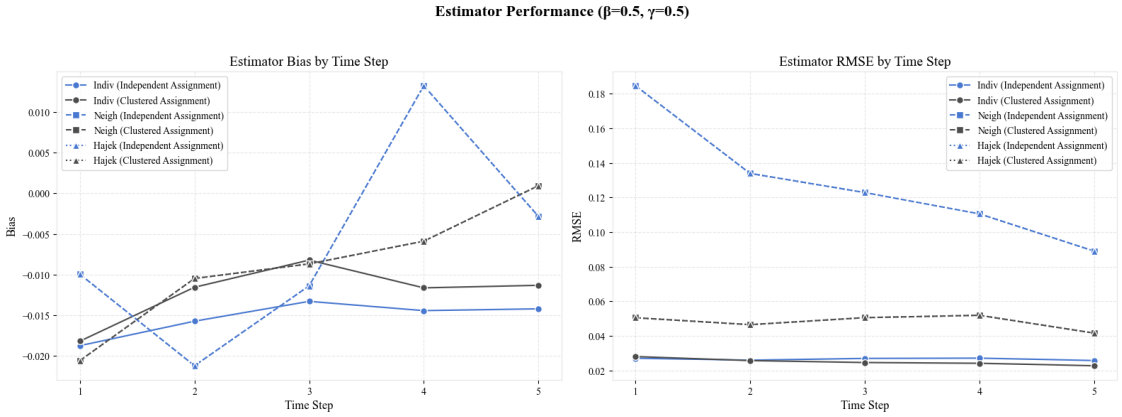

Figure 4: Estimator bias and RMSE by time step ($\beta = 0.5$, $\gamma = 0.5$)

Figure 4 provides several insights into estimator performance across the dynamic treatment diffusion process. The most salient finding is that designs combining graph cluster randomization with neighborhood-based estimators consistently achieve the lowest bias magnitudes across almost all stages of dynamic diffusion and all randomization designs considered. Clustered assignment, when paired with network exposure modeling, substantially mitigates the bias introduced by treatment interference and produces estimates that remain closer to zero throughout the diffusion process. In addition, the RMSE patterns further underscore the benefits of clustered designs for neighborhood-based estimators: independent randomization yields notably higher RMSE, while graph cluster randomization stabilizes and reduces estimation error over time. Although the simple individual-level estimator exhibits lower RMSE overall, it does so at the cost of persistent negative bias due to its failure to account for spillover effects.

In subsequent analysis, we focus on the final time step $T = 5$ as the primary outcome of interest for comparing estimator performance across designs and parameter configurations.

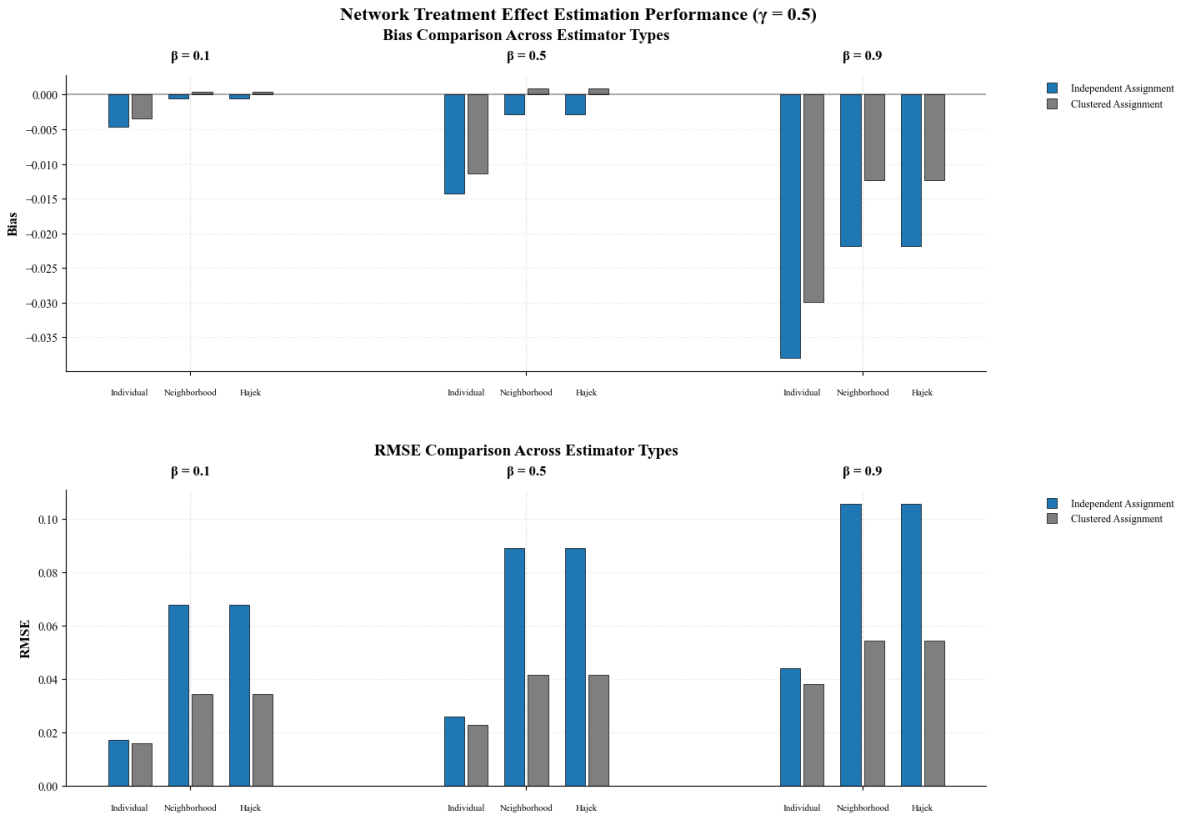

Figure 5: Estimator bias and RMSE comparison across β when $\gamma = 0.5$

Figure 5 summarizes the estimator performance across different levels of direct treatment effects (β) under a moderate peer effect ($\gamma = 0.5$). Across all values of β, neighborhood-based estimators (FNTR and Hajek) consistently achieve lower bias compared to the individual-level estimator, reflecting the advantage of explicitly modeling network exposure in the presence of interference and dynamic diffusion. Moreover, for each estimator, cluster randomization systematically improves performance relative to independent assignment, yielding lower bias magnitudes. Neighborhood-based estimators under cluster randomization successfully balance the bias-variance tradeoff, achieving both relatively low bias and reduced RMSE compared to their independently assigned counterparts. These results suggest that, under moderate peer effects, combining network exposure modeling with cluster randomization provides the most robust and accurate estimates of treatment effects across varying strengths of direct treatment influence.

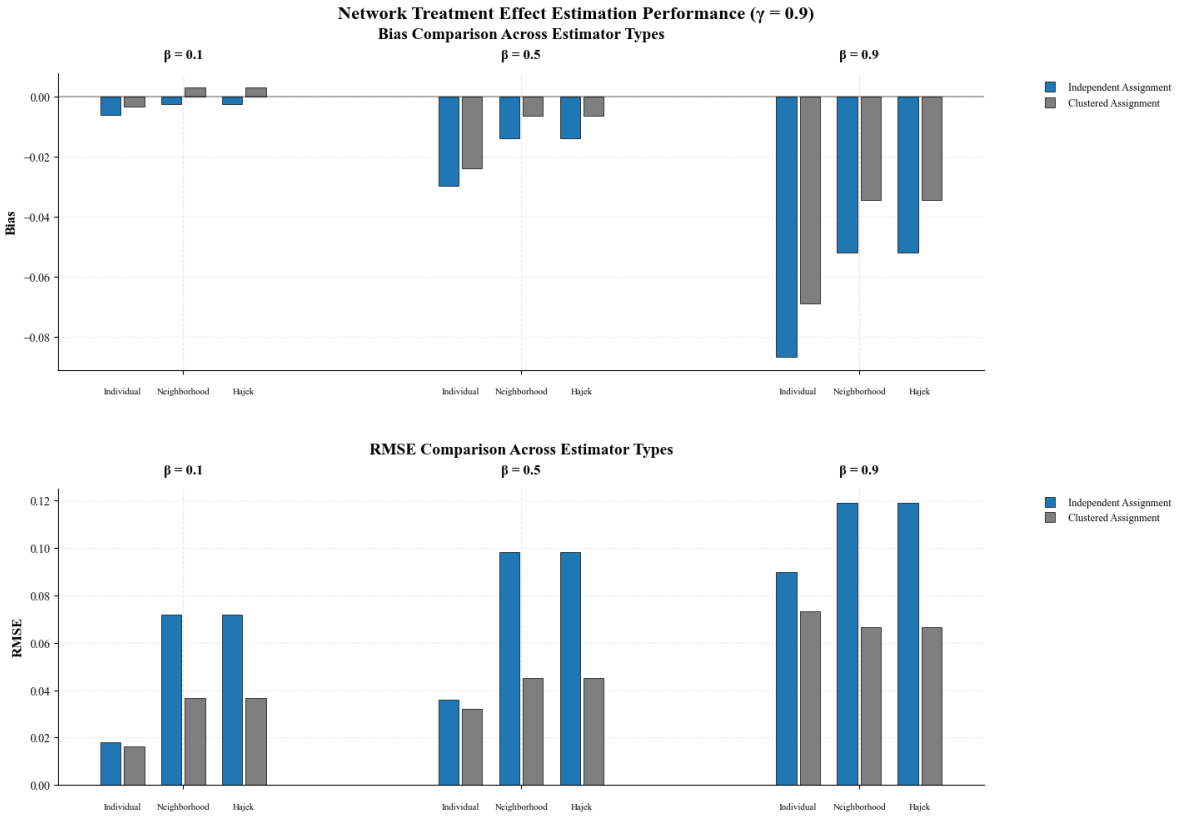

Figure 6: Estimator bias and RMSE comparison across β when $\gamma = 0.9$

Figure 6 examines estimator performance under a stronger peer effect setting ($\gamma = 0.9$), holding the same range of direct effects β. Consistent with earlier findings, neighborhood-based estimators (FNTR and Hajek) combined with cluster randomization continue to offer the best balance between bias reduction and RMSE minimization for moderate and large β values. These results demonstrate the robustness of the advantage of network exposure modeling and graph clustering randomization across different levels of peer spillover intensity under treatment diffusion. Only when $\beta = 0.1$ does the direct treatment effect becomes too weak to substantively shift outcomes, resulting in less pronounced effect of treatment diffusion and smaller outcome variation. In this regime, the superiority of the neighborhood-based estimators under clustered randomization becomes ambiguous, as biases and RMSE differences across designs are minor and unstable.

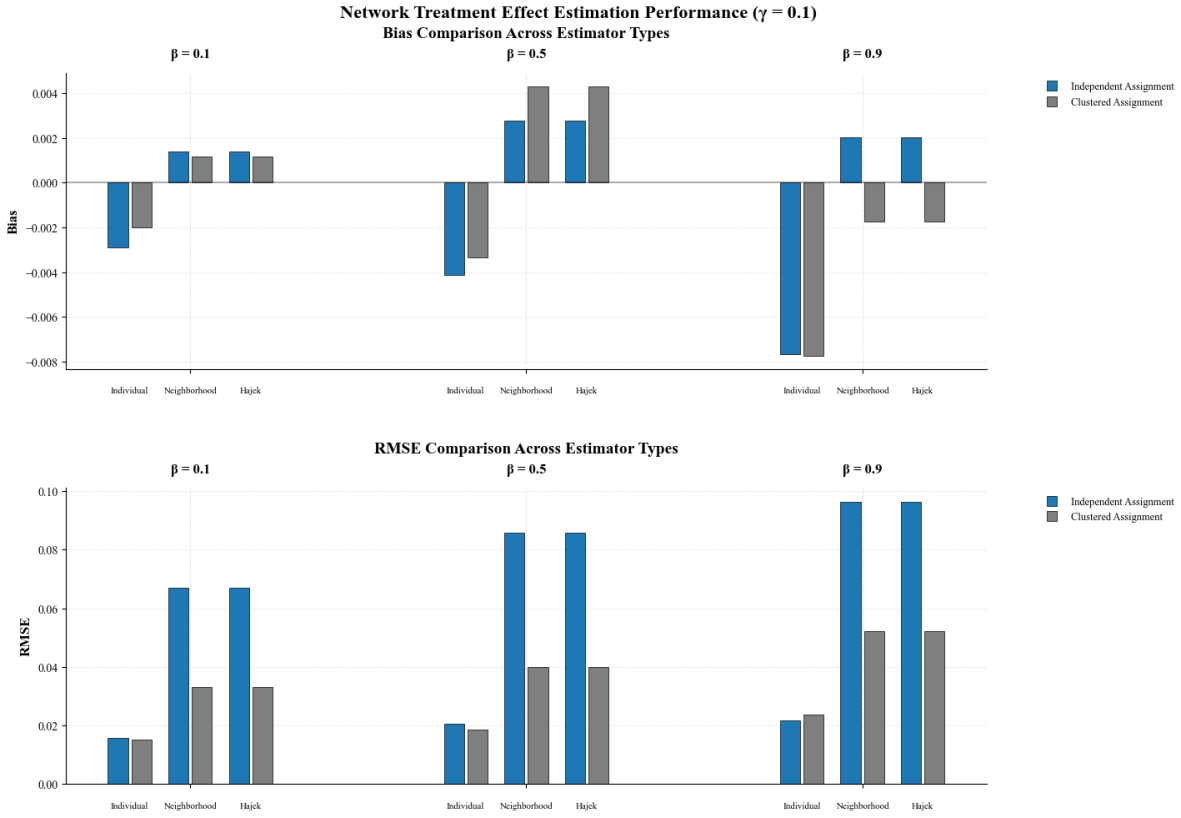

Figure 7: Estimator bias and RMSE comparison across β when $\gamma = 0.1$

Focusing on Figure 7, we observe that under a low peer effect setting ($\gamma = 0.1$), the advantage of combining neighborhood-based estimators with clustered randomization becomes less robust across all levels of direct treatment effect β. When peer influence is minimal, treated nodes exert little indirect influence on their neighbors. As a result, the benefits of modeling network exposure and structuring assignments through clusters diminished. In the bias comparison panels, the absolute biases are already very small for all estimator-design combinations, and no clear superiority emerges for the neighborhood-based approaches. These patterns indicate that when spillover effect is weak, network-based refinements to the design and estimation contribute relatively little to improving estimator performance.

By combining flexible treatment assignment mechanisms, endogenous diffusion processes, and dynamic network-dependent outcome evolution, this simulation design provides a structured yet flexible environment for evaluating the identification of causal effects under interference and essentially confirms the robustness of the results from Eckles et al. (2017) in such a dynamic network setting.

5 Extensions

5.1 Sensitivity to Diffusion Speed

In this extension, we analyze the sensitivity of estimator performance to the diffusion speed of treatment within the network. To evaluate this, we generate Figures 8, 9, and 10, which illustrate the variations in estimator bias and RMSE across different values of β and diffusion probabilities, for three distinct values of the peer effect parameter γ. These figures allow us to observe how estimator performance varies as diffusion speed changes, providing insights into the robustness of different estimators under varying levels of diffusion intensity.

Neighborhood-based estimators, represented by the dashed lines in the figures, exhibit a higher sensitivity to diffusion speed as the diffusion probability increases, leading to more dramatic fluctuations in bias and RMSE for these estimators. This sensitivity arises because neighborhood estimators rely heavily on local network dynamics. When treatment spreads rapidly, the influence of neighboring treated units intensifies, causing the estimator to adjust more abruptly to the changing treatment environment and making it harder for the estimators to maintain precision. This sensitivity is visible in all three figures.

In contrast, when neighborhood-based estimators are combined with clustered random assignment, the resulting estimators are less sensitive to diffusion speed. This stabilization effect is clearly evident by comparing the black and blue dashed lines, as shown in three figures. Clustered assignment helps stabilize estimator performance by reducing the cross-treatment interference between treated and control units. Treatment is assigned at the cluster level, ensuring that treated units are more likely to be surrounded by other treated units, and control units by other controls. This alignment of treatment assignment with network structure reduces the variance in the treatment diffusion process and mitigates the volatility in estimator performance that is seen when diffusion speed increases. As a result, neighborhood-based estimators with clustered assignment maintain a more consistent performance, with smaller fluctuations in both bias and RMSE across varying diffusion probabilities.

The third key finding is that neighborhood-based estimators with clustered assignment outperform other approaches when γ (spillover effects) and β (direct treatment effect) are

large. This is particularly noticeable when $\gamma = 0.9$ and $\beta = 0.9$ in Figure 10, where the bias and RMSE for neighborhood-based estimators with clustered assignment remain consistently low, even as the diffusion speed increases. The stronger peer effects and direct treatment effects in these settings amplify the importance of accurately capturing spillover dynamics. Since the clustered assignment helps isolate the treated and control groups more effectively, this leads to a clearer comparison between these groups, as the spillover effects are contained within clusters, reducing the risk of bias that can occur when treatment diffusion is uncontrolled. As γ and β grow larger, the performance of the neighborhood-based estimators with clustered assignment improve significantly, making it the most reliable estimator in these scenarios.

In conclusion, this extension reveals the critical role of treatment diffusion speed in networked experiments. While neighborhood-based estimators are highly sensitive to diffusion speed, combining them with clustered randomization stabilizes their performance and reduces sensitivity. Moreover, when both direct and spillover effects are strong, this combination serves as the most robust method for estimating treatment effects in dynamic networks. These findings underline the importance of considering both network structure and treatment diffusion dynamics when designing experiments in networked settings.

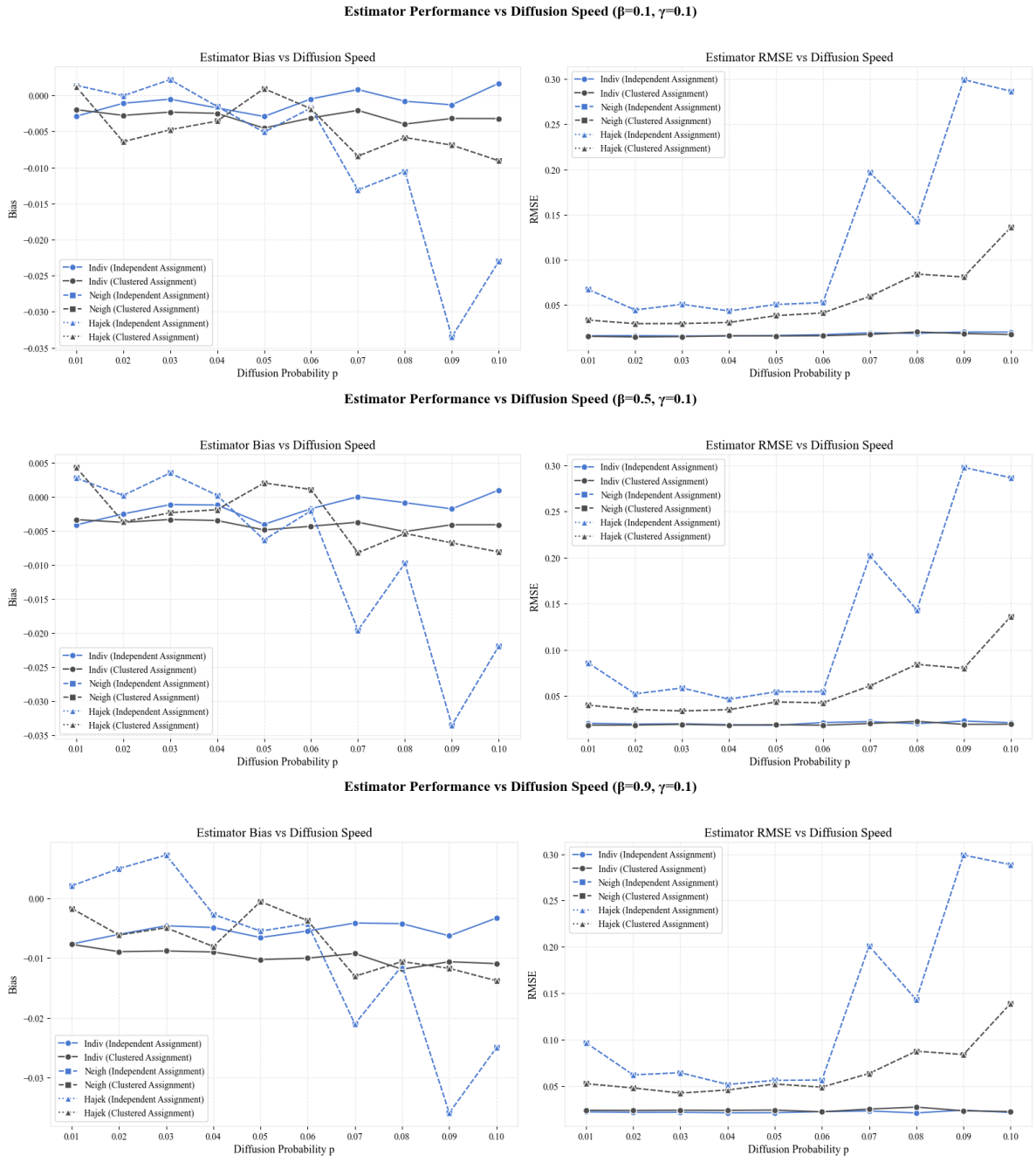

Figure 8: Estimator bias and RMSE variations across β and different diffusion probability when $\gamma = 0.1$

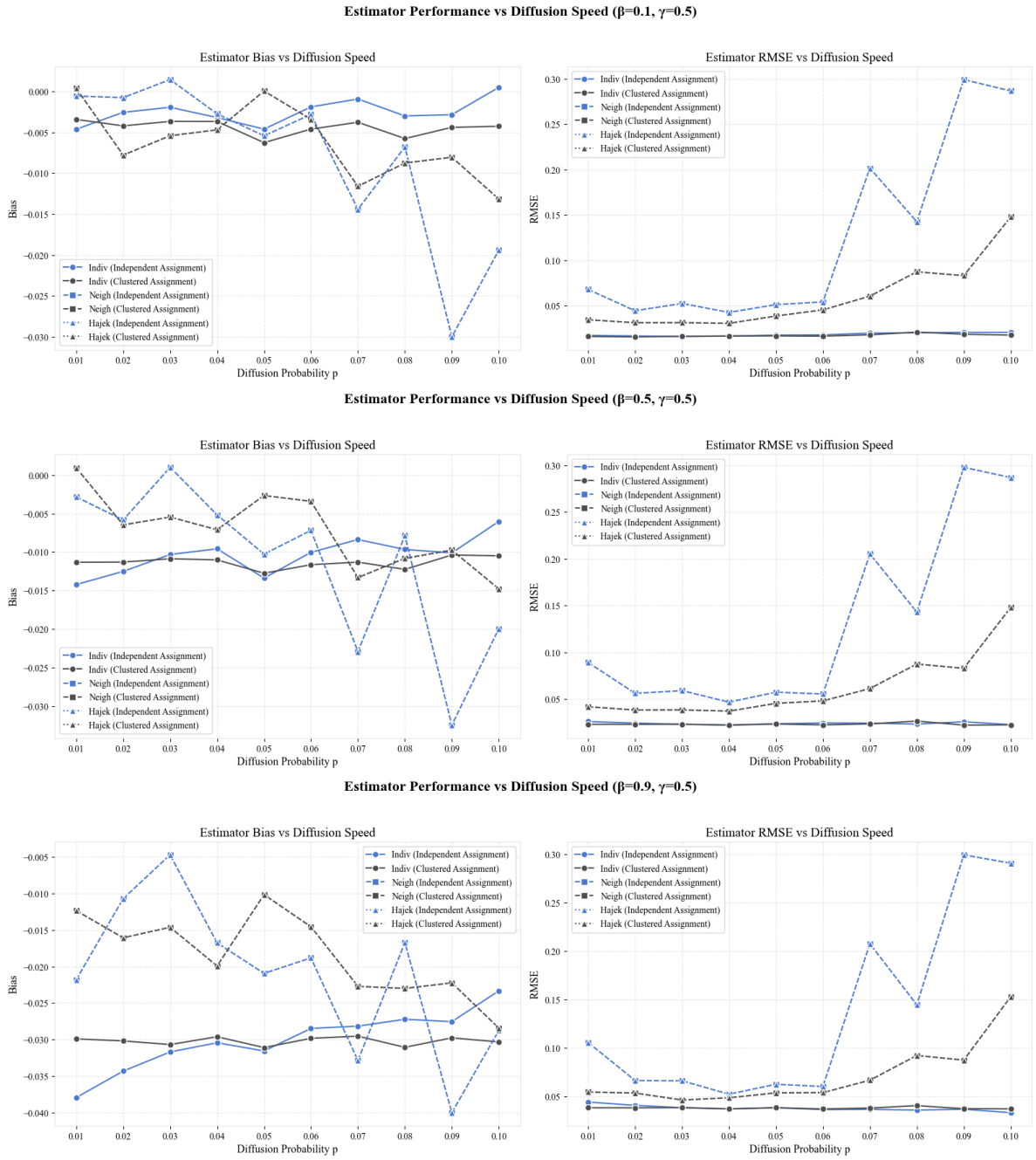

Figure 9: Estimator bias and RMSE variations across β and different diffusion probability when $\gamma = 0.5$

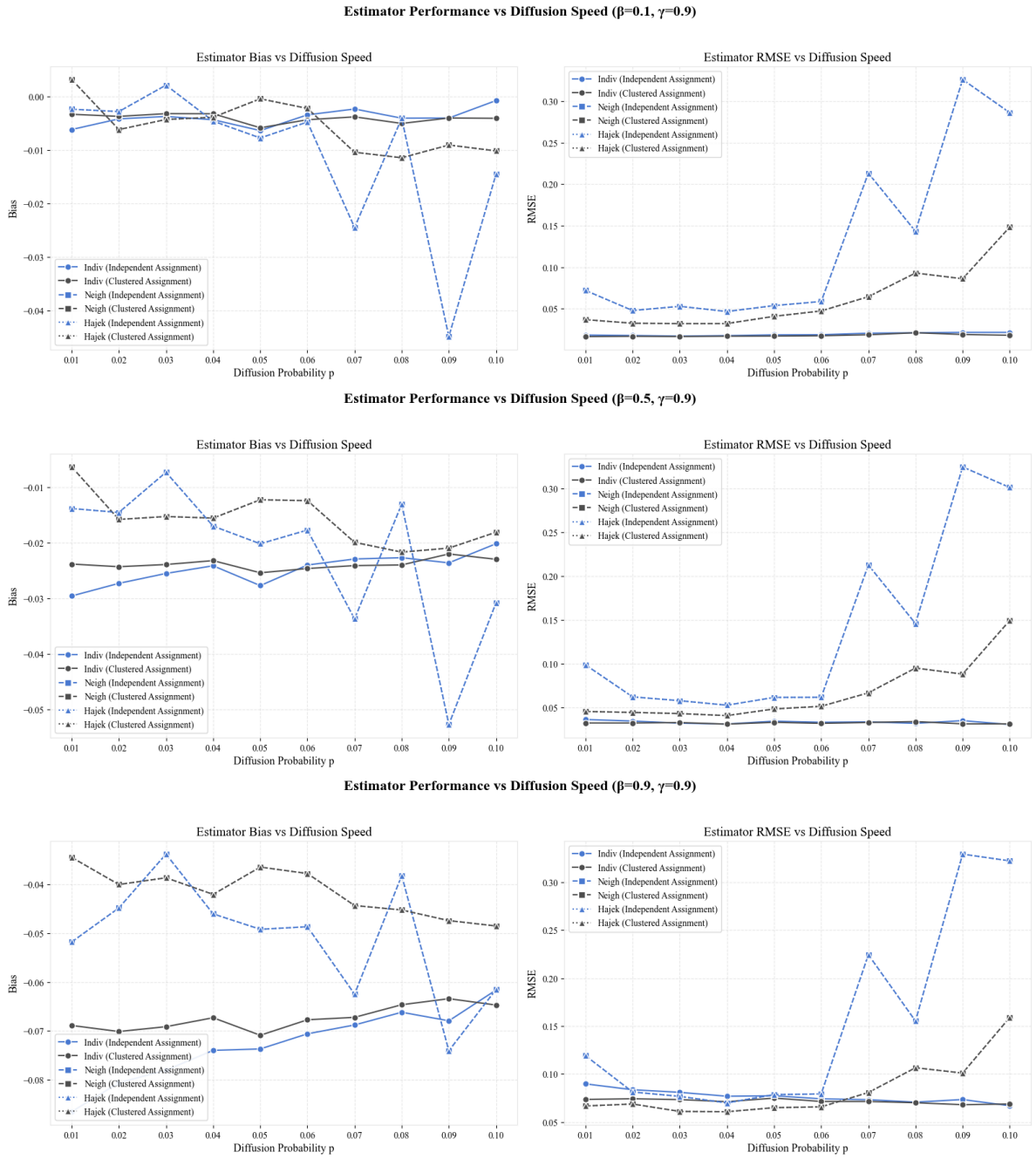

Figure 10: Estimator bias and RMSE variations across β and different diffusion probability when $\gamma = 0.9$

5.2 Alternative Diffusion Pattern

While ICM captures the randomness inherent in peer-to-peer influence, where each treated neighbor independently attempts to activate a node with a certain probability, real-world dif-

fusion dynamics often involve more complex decision mechanisms. These mechanisms might require multiple neighbors to reach a certain threshold of influence before an agent adopts a behavior. To account for such richer diffusion processes, we extend our framework to include the Linear Threshold Model (LTM), which models threshold-based adoption behavior. The LTM assumes that each node will adopt a treatment if the sum of the influences from its neighbors reaches or exceeds a certain threshold.

The LTM diffusion function is mathematically expressed as follows:

$$f_i = \mathbb{I}\left\{ \sum_{j \in \mathcal{N}(i)} w_{ji} Z_j(t-1) \geq \theta_i \right\},$$

where w_{ji} represents the influence weight from node j to node i, and $\theta_i \in [0, 1]$ is the adoption threshold for node i. The latent variables w_{ji} and θ_i can either be sampled from a simulation distribution or estimated from real-world diffusion data if available. In our simulation, the influence weight w_{ji} between nodes is assigned as the inverse of their degrees, ensuring normalized influence, while each node's adoption threshold θ_i is randomly chosen from a uniform distribution between 0.55 and 0.6, which is carefully chosen to adjust for appropriate diffusion speeds. This model captures more complex diffusion behavior where adoption does not occur unless a node is sufficiently influenced by its neighbors, rather than through independent activations as in the ICM.

Unlike ICM, the choice of treatment assignment method (independent or clustered) significantly influences the diffusion speed under LTM. Under independent assignment, treated nodes are randomly distributed across the network, leading to a faster and more widespread diffusion process. In contrast, clustered randomization groups treated nodes together, leading to a more localized and slower diffusion, as treated nodes within the same cluster are more likely to influence each other but are less likely to spread treatment beyond the cluster. This creates a slower diffusion pattern under clustered randomization, as seen in Figure 11, which illustrates the faster diffusion of treatment over time under independent assignment compared to clustered assignment.

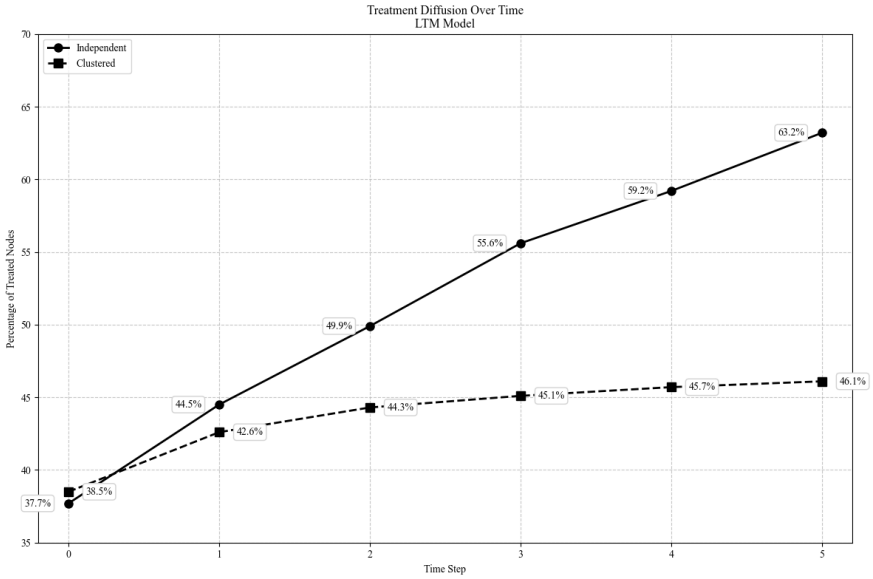

Figure 11: Treatment diffusion over time under the Linear Threshold Model (LTM)

Figure 12, 13, 14, and 15 present the results of estimator bias and RMSE by time step and across different values of β and γ under the LTM. These results further demonstrate that neighborhood-based estimators are sensitive to diffusion speed, with bias and RMSE fluctuating more as the treatment spreads faster under clustered randomization when comparing Figure 13, 14, and 15.

As with ICM, neighborhood-based estimator with clustered assignment consistently lead to lower bias and RMSE, especially when direct and spillover effects are strong. These results suggest that the choice of diffusion model does not fundamentally change the conclusions about the relative performance of treatment assignment and estimation methods and reaffirms the robustness of the main results in Section 4.

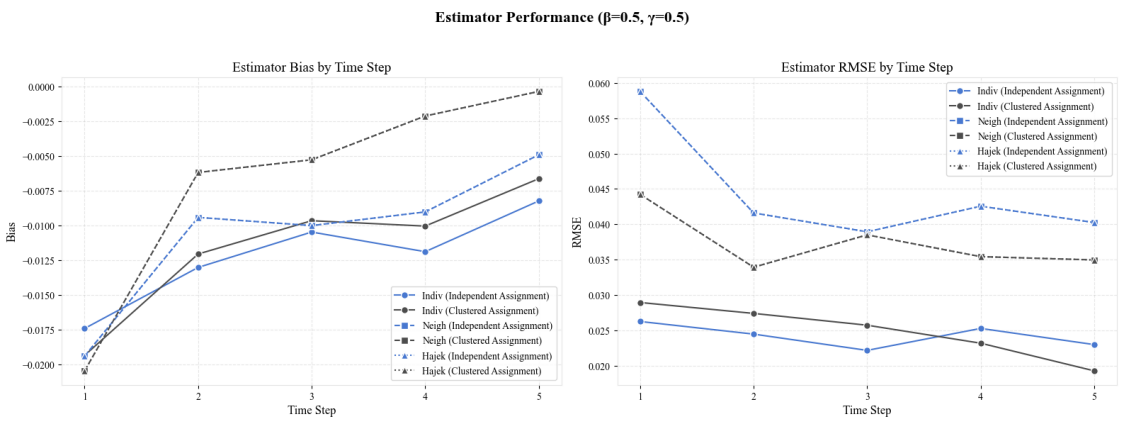

Figure 12: Estimator bias and RMSE by time step ($\beta = 0.5$, $\gamma = 0.5$) under LTM

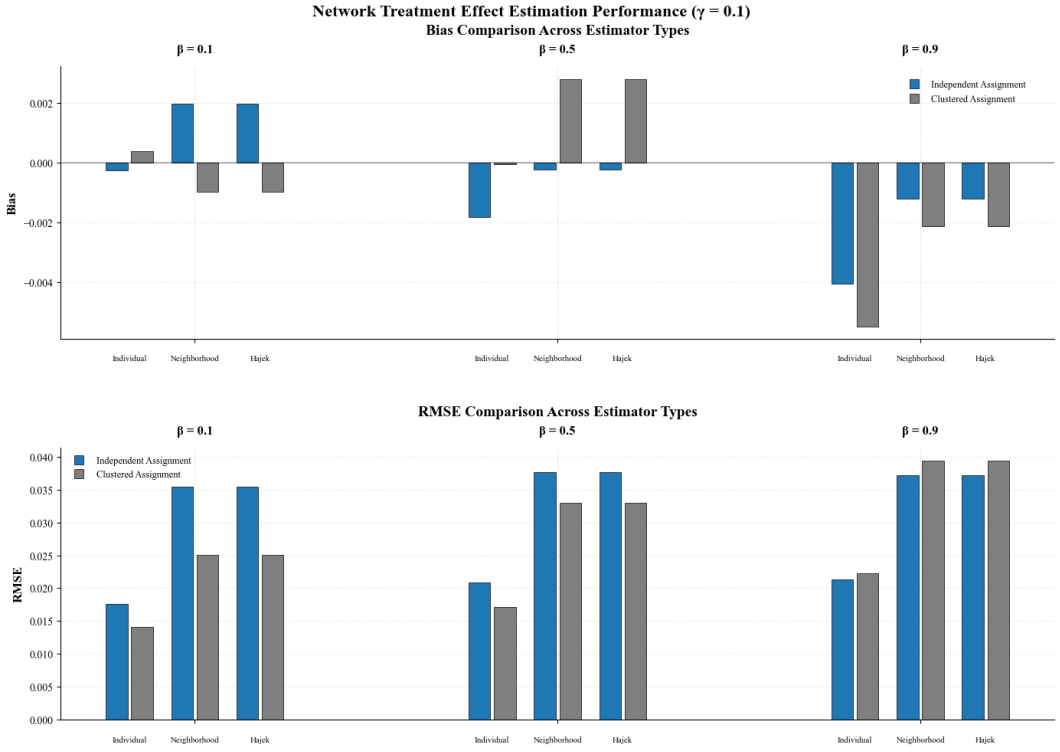

Figure 13: Estimator bias and RMSE comparison across β when $\gamma = 0.1$ under LTM

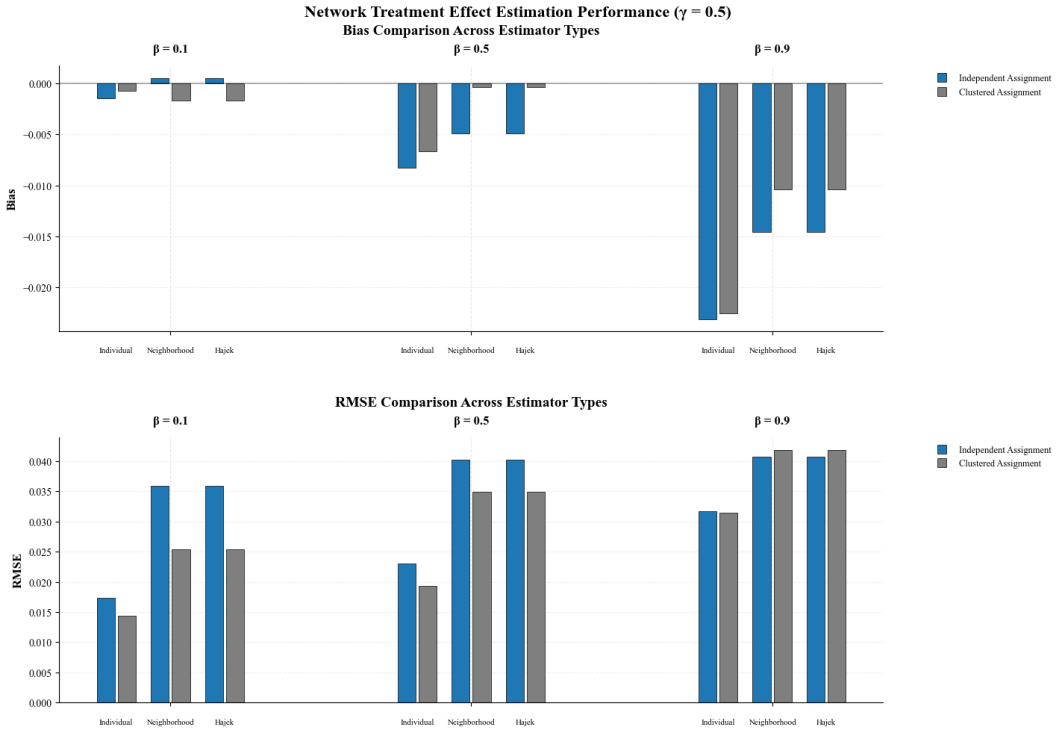

Figure 14: Estimator bias and RMSE comparison across β when $\gamma = 0.5$ under LTM

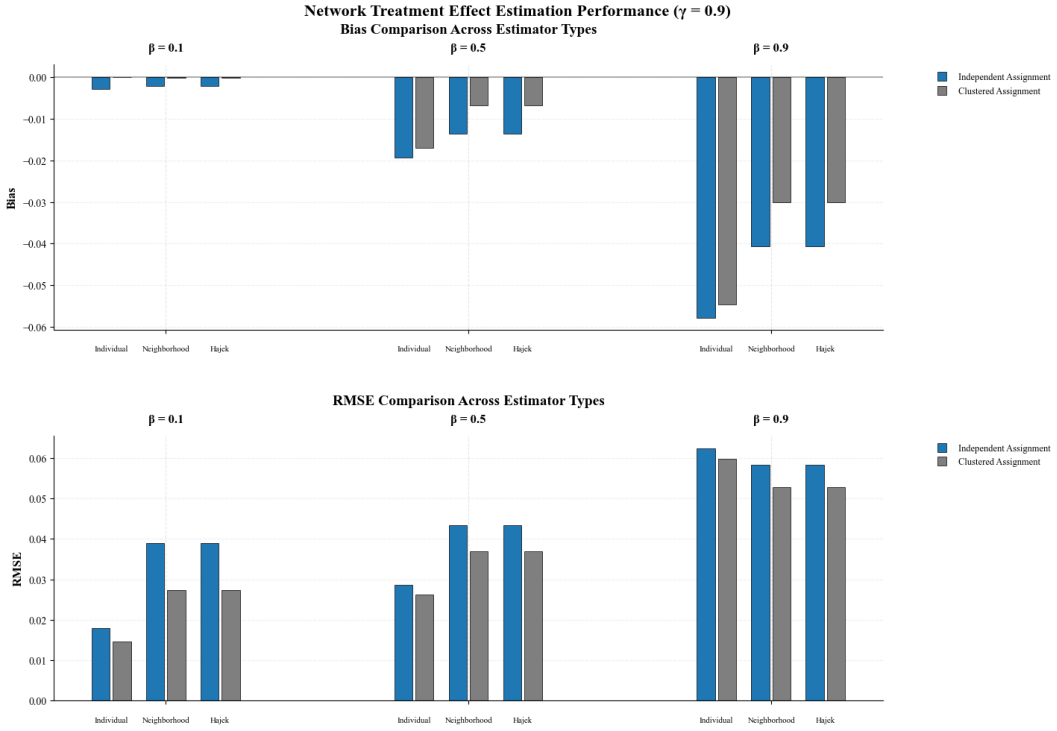

Figure 15: Estimator bias and RMSE comparison across β when $\gamma = 0.9$ under LTM

6 Discussion and Future Work

This paper proposes a comprehensive framework for designing and analyzing network experiments under treatment diffusion dynamics, extending the foundational work of Eckles et al. (2017). We introduce a five-phase experimental design—comprising initialization, treatment assignment, treatment diffusion, outcome generation, and estimation—that explicitly incorporates dynamic treatment evolution through peer influence. By employing ICM as the baseline diffusion mechanism, and extending to LTM as an alternative, we systematically evaluate how treatment assignment and estimator selection interact to affect causal identification.

Our simulation results highlight several key findings. First, combining graph cluster randomization with neighborhood-based estimators consistently reduces bias, improves accuracy, and remains robust, especially when direct and spillover effects are strong. Second, the intensity of treatment diffusion speed fundamentally shapes the identification environment, influencing both bias and variance properties of estimators. Our results demonstrate that faster or more heterogeneous diffusion patterns can amplify spillover effects, increasing neighborhood-based estimator sensitivity which can be offset by clustered assignment. Third, different treatment diffusion patterns offer similar observations in bias reduction, suggesting robustness of the results.

Nevertheless, several limitations and opportunities for future work remain. First, although we modeled networks using the Watts-Strogatz small-world structure to capture basic clustering and short-path properties, real-world networks often exhibit more complex heterogeneity. Extensions to more general network models, such as the Degree-Corrected Stochastic Block Model (DCBM) (Karrer and Newman, 2011), would allow for richer modeling of degree variation and community structures. Such extensions could improve the robustness of treatment assignment and estimator design.

Second, our framework assumes that the underlying network structure is fully observed and static throughout the experimental period. However, in many real-world applications—such as public health campaigns, online social platforms, or organizational interventions—the network is only partially observed, uncertain, or evolving over time. A critical

avenue for future work is to develop methods that jointly infer network structure and estimate treatment effects, potentially leveraging machine learning techniques such as graph neural networks (GNNs), latent space models, or network reconstruction algorithms. Incorporating uncertainty over network topology would improve the realism of experimental designs and enable credible causal inference when the true network is only imperfectly known.

Third, while our study focused on two canonical diffusion processes—ICM and LTM—future work could explore heterogeneous or time-varying diffusion rates, non-Markovian diffusion mechanisms, or learning-based diffusion processes. Incorporating heterogeneity in peer influence or modeling temporal evolution in adoption thresholds could better capture complex diffusion observed in real-world systems, such as the gradual adoption of technologies or cascading behaviors in social networks.

Fourth, although neighborhood-based estimators substantially reduce bias under dynamic diffusion, they rely on exposure mappings that may not fully capture higher-order dependencies or complex spillover patterns. Recent advances in machine learning offer promising avenues for estimator refinement. In particular, Double Machine Learning (DML) frameworks and Graph Neural Networks (GNNs) can flexibly model both treatment assignment mechanisms and outcome evolution, potentially improving estimation in high-dimensional or partially observed network settings. Further research could extend the framework by integrating representation learning techniques to predict effective exposures or to model treatment propagation pathways more accurately.

Finally, while our simulation results demonstrate effective bias reduction and improved estimator performance, the absence of formal theoretical guarantees remains an important limitation. In particular, rigorous analysis of identification conditions, consistency, and asymptotic properties under dynamic treatment diffusion is needed to complement the empirical findings. Future research should focus on establishing formal theoretical foundations for network-based estimators in dynamic settings, including deriving bias bounds, convergence rates, and conditions under which causal parameters are identifiable despite treatment propagation. Such results would strengthen the generalizability and robustness of the proposed framework across diverse network environments.

In summary, by integrating treatment diffusion into network experimental design and

demonstrating robust bias reduction through cluster randomization and neighborhood-based estimation, this study provides a foundation for more realistic and accurate causal inference in interconnected social systems. Expanding this framework to accommodate richer network structures and more complex diffusion dynamics, represents a promising direction for future research on causal inference under interference.

References

An, W. (2018). Causal inference with networked treatment diffusion. *Sociological Methodology*, 48(1):152–181.

Aronow, P. M. and Samii, C. (2017). Estimating average causal effects under general interference. *Annals of Applied Statistics*, 11(4):1912–1947.

Basse, G. W. and Feller, A. (2018). Analyzing two-stage randomized experiments in networks. *Biometrika*, 105(4):849–858.

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

Cox, D. R. (1958). Planning of experiments.

Eckles, D., Karrer, B., and Ugander, J. (2017). Design and analysis of experiments in networks: Reducing bias from interference. *Journal of Causal Inference*, 5(1):... (article pagination).

Fortunato, S. (2010). Community detection in graphs. *Physics reports*, 486(3-5):75–174.

Karrer, B. and Newman, M. E. (2011). Stochastic blockmodels and community structure in networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 83(1):016107.

Kempe, D., Kleinberg, J., and Tardos, É. (2003). Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–542.

Papadogeorgou, G., Mealli, F., and Zigler, C. M. (2019). Causal inference with interfering units for cluster and population level treatment allocation programs. *Biometrics*, 75(3):778–787.

Park, C. and Kang, H. (2022). Efficient semiparametric estimation of network treatment effects under partial interference. *Biometrika*, 109(4):1015–1031.

Rubin, D. B. (1980). Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American statistical association*, 75(371):591–593.

Sävje, F., Aronow, P. M., and Hudgens, M. G. (2017). Average treatment effects in the presence of unknown interference. *The Annals of Statistics*, 45(6):2103–2130.

Ugander, J., Karrer, B., Backstrom, L., and Kleinberg, J. (2013). Graph cluster randomization: Network exposure to multiple universes. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 329–337.

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

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442.