Proceedings of the 2025 Winter Simulation Conference
E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.

SPILLOVER-AWARE SIMULATION ANALYSIS FOR POLICY EVALUATION IN EPIDEMIC NETWORKS

Jingyuan Chou[1,2], Jiangzhuo Chen[2], Madhav Marathe[1,2]

[1]Dept. of Computer Science, University of Virginia, Charlottesville, VA, USA
[2]Biocomplexity Institute, University of Virginia, Charlottesville, VA, USA

ABSTRACT

Simulations are widely used to evaluate public health interventions, yet they often fail to quantify how interventions in one region indirectly affect others—a phenomenon known as spillover. This omission can lead to incorrect policy evaluations and misattributed effects. We propose a post-simulation framework for estimating causal spillover in spatial epidemic networks. Our method introduces a directional graph neural network (Dir-GNN) estimator that learns homophily-aware representations and estimates counterfactual outcomes under hypothetical neighbor treatments. Applied to a semi-synthetic setup built on PatchSim—a metapopulation SEIR simulator with realistic inter-county mobility—our estimator recovers spillover effects and corrects attribution errors inherent in standard evaluation. Experiments show that accounting for spillover improves treatment estimation and policy reliability.

1 INTRODUCTION

Epidemic simulations are widely used for evaluating public health interventions by modeling disease spread under various policy scenarios (Qian et al. 2025; Nitzsche and Simm 2024; Moon et al. 2024; Bhattacharya et al. 2023). These models enable counterfactual simulations of interventions such as lockdowns, travel restrictions, and vaccinations, helping policymakers assess potential impacts before implementation.

However, simulations do not quantify spillover—the indirect effects on untreated regions. (Kaminsky et al. 2019; Liu et al. 2024). This limitation presents a fundamental challenge: Simulations report total infection counts but do not explain why changes occur in untreated regions. Post-simulation analysis helps separate direct effects from spillover for better policy evaluation.

For example, suppose a vaccination campaign in one region is followed by a decline in infections in a neighboring region. Standard simulations cannot determine whether this decline is due to true spillover effects—where the intervention indirectly benefits nearby areas—or other factors such as natural epidemic decline, local behavioral changes, or demographic similarities between regions. Misinterpreting simulation outputs may lead to ineffective policy expansion or misallocation of resources. To address this, we introduce a spillover-aware simulation analysis framework that extends the decision-making utility of epidemic simulations. Instead of modifying the underlying simulation models, our approach integrates causal spillover estimation as an analytical layer to extract indirect effects from simulation outputs. This allows policymakers to: **i). Quantify spillover effects**: Separating indirect intervention impacts from direct effects and natural epidemic trends. **ii). Improve policy evaluation**: Providing a structured assessment of whether interventions should be localized, expanded, or coordinated across regions. **iii). Enhance simulation-based decision support**: Transforming traditional simulations into tools that explicitly account for indirect intervention effects.

By embedding spillover estimation into simulation workflows, our framework ensures that epidemic models predict disease dynamics and support more robust, data-driven decision-making. However, spillover estimation in epidemic simulations is particularly challenging due to: **i). Confounding in spatial networks**: Neighboring regions exhibit similar infection trends due to shared demographics, mobility structures, and

healthcare resources (Feng et al. 2021; Shalizi and Thomas 2011). Failing to account for these factors may lead to biased estimates of spillover effects. **ii). Dynamic population movement**: Unlike static social networks, epidemic networks evolve as disease transmission and mobility patterns change. Existing spillover estimation methods assume static networks, limiting their applicability to epidemic models.

This paper introduces a simulation-aware framework for estimating causal spillover effects in epidemic networks. Our key contributions are:

- We formalize the problem of causal spillover estimation in the context of epidemic simulations and show how neglecting spillover can mislead policy evaluations.
- We propose a directional GNN-based estimator that jointly models treatment and outcomes, leveraging homophily-aware node representations to adjust for unobserved confounding.
- We integrate our framework with PatchSim, a metapopulation SEIR simulator using real population and mobility data, and validate it on semi-synthetic intervention scenarios. Our experiments demonstrate that spillover-aware evaluation improves policy attribution, recovering true indirect effects that would otherwise be masked in standard simulation outputs.

By bridging the gap between epidemic forecasting and policy evaluation, our approach ensures that policymakers obtain unbiased estimates of intervention effectiveness, improving public health strategies.

2 RELATED WORK

Our work connects to three key areas: simulation-based epidemic policy evaluation, causal inference in networked settings, and spillover estimation in spatial networks.

2.1 Simulation-Based Epidemic Policy Evaluation

Epidemic simulations are widely used to model disease spread and assess intervention strategies. Compartmental metapopulation models (Venkatramanan et al. 2021; Kermack and McKendrick 1927; Chen et al. 2024) simulate aggregate-level dynamics, while agent-based models (ABMs) (Chen et al. 2024; Robertson et al. 2025; Bhattacharya et al. 2021; Hoops et al. 2021; Chen et al. 2024) capture individual mobility and behavior. These models enable counterfactual analysis but do not explicitly quantify spillover effects, making it difficult to assess whether interventions indirectly influence untreated regions (Lawson and Rotejanaprasert 2023; He et al. 2024).

Spillover effects are significant for evaluating localized interventions (e.g., regional travel restrictions or vaccination campaigns). If simulations only report total infection counts without estimating indirect effects, policymakers cannot determine whether observed reductions in untreated regions result from spillover or external epidemic trends. Our work addresses this gap by integrating spillover estimation into simulation analysis, providing structured methods to separate direct and indirect intervention effects.

2.2 Causal Inference in Networked Settings

Traditional causal inference methods assume independent treatment assignments, which fails in networked environments where interventions affect neighboring units (Manski 2013; Aronow and Samii 2017). Several methods have been developed to estimate spillover effects in networks, including inverse probability weighting (IPW) (Lee et al. 2023; Buchanan et al. 2024) and network-based counterfactual approaches (Zhang et al. 2024; McFowland and Shalizi 2021; Chang et al. 2023). Additionally, graph neural networks (GNNs) have been explored for causal effect estimation (Ma and Tresp 2020; Huang et al. 2024).

However, these methods are typically designed for static social networks and do not integrate with dynamic epidemic simulations. Unlike economic or social networks, where relationships are fixed, epidemic networks evolve with disease transmission and mobility patterns. Our approach bridges this gap by adapting graph-based spillover estimation to simulation-based forecasting.

2.3 Spillover Estimation in Spatial Networks

Spillover estimation has been widely studied in social and economic networks (Benjamin-Chung et al. 2017; Toulis and Kao 2013; Eckles et al. 2016; Fatemi and Zheleva 2020a; Laffers and Mellace 2020; van der Laan 2014; Ogburn et al. 2024; Tchetgen et al. 2021; Fatemi and Zheleva 2020b; Hoshino and Yanagi 2024), but its application to epidemiological simulations remains limited. While existing methods adjust for confounding in spatial networks (Veitch et al. 2019; Sridhar et al. 2022; Tec et al. 2024; Papadogeorgou and Samanta 2024; Jiang et al. 2023), they do not account for dynamic disease spread and intervention spillover.

Our work builds on these foundations by introducing a spillover-aware framework tailored for epidemic simulations. Rather than treating spillover as a separate problem, we extend spillover estimation into simulation workflows, enhancing their ability to provide policy-relevant insights post-simulation. Different from previous work, our approach: **i). Extends simulation analysis** by incorporating structured spillover estimation into forecasting models. **ii). Bridges causal inference and simulation** by adapting spillover estimators for dynamic epidemic settings. **iii). Improves simulation-based decision support** by providing policymakers with more interpretable intervention evaluations.

By enhancing simulation-based policy evaluation with spillover estimation, our framework ensures that epidemic simulations go beyond forecasting to directly inform data-driven public health strategies.

3 PROBLEM FORMULATION

Epidemic simulations provide a way to model disease spread under different intervention scenarios. However, standard simulations output total infection counts without explicitly separating **direct effects**, the reduction in infections due to an intervention in a treated region, and **spillover effects**, the indirect decrease in infections in untreated regions due to treatments in other regions. Furthermore, we should also note that there also exists **confounding effects** that similar infection trends across regions due to shared population characteristics (homophily) rather than true intervention spillover.

This limitation makes it difficult to assess whether observed infection declines outside the treated regions are due to true spillover effects, external factors, or inherent similarities between regions.

3.1 Simulated Epidemic Network

We represent the epidemic system as a graph $G(\mathbb{V}, E)$, where nodes correspond to spatial regions (e.g., counties), and edges encode spatial connectivity based on mobility patterns. Each node v_i has: i). A treatment assignment T_i (e.g., whether an intervention was applied). ii). A simulated infection counts Y_i, representing the number of infections in the region v_i at a given timestamp. iii). Covariates X_i, such as population density or healthcare capacity. iv). Unobserved confounders C_i, factors influence both T_i and Y_i simultaneously. Simulations evolve, updating Y_i based on transmission dynamics. However, they do not explicitly separate spillover effects from confounding effects due to homophily, potentially leading to biased policy conclusions.

3.2 Spillover Effect Estimation in Simulation Outputs

We adopt a structured spillover-aware estimation approach based on the potential outcomes framework (Rubin 2005) to quantify spillover effects. Rather than modifying the simulation process itself, we analyze simulation outputs to separate direct and spillover effects.

Following the potential outcomes framework, we define the expected infection count at each node as $\mathbb{E}[Y_i|T_i, T_i^*, G, X_i, X_i^*, C_i]$. T_i^* summarizes the treatment assignments of neighboring regions, X_i^* represents the aggregation of covariates of neighbors, C_i represents unobserved confounders, which may introduce bias by making similar regions appear to have spillover effects even when none exist, and $\mathbb{E}[\cdot]$ represents

the expected (average) infection count under given treatment conditions, accounting for randomness in disease transmission.

3.3 Decomposing Simulation Outputs into Direct and Spillover Effects

To distinguish direct and spillover effects in simulation outcomes, we decompose the observed infection count Y_i as: $Y_i = \psi_{i,D} + \psi_{i,S}$, where $\psi_{i,D}$ represents the direct effect of an intervention on v_i, and $\psi_{i,S}$ represents the spillover effect from interventions in neighboring regions.

We estimate the Average Spillover Effect (ASE) across all nodes to quantify the network-wide impact of spillovers in the simulation: $\psi_S = \frac{1}{n}\sum_{i=1}^{n} \psi_{i,S}$.

3.4 Neighboring Treatment Influence Representation

To account for the effects of interventions in neighboring regions, we define T_i^* in two ways: **i). Mean Aggregation**: $T_i^* = \frac{1}{|\mathcal{N}_i|}\sum_{j\in\mathcal{N}_i} T_j$, summarizing mean of treatments of neighbors. **ii). Vector Representation**: $T_i^* = (T_{j_1}, T_{j_2}, \ldots, T_{j_{|\mathcal{N}_i|}})$, retaining full neighbor-level treatment information. In our work, we adopt the vector representation to enable more flexible modeling of heterogeneous spillover effects.

3.5 Estimating Spillover Effects from Simulated Data

In practice, we estimate the Average Spillover Effect (ASE) ψ_S as:

$$\psi_S = \frac{1}{n}\sum_{i=1}^{n} \left(\mathbb{E}[Y_i \mid T_i, T_i^* = \mathbf{1}, G, X_i, X_i^*, C_i] - \mathbb{E}[Y_i \mid T_i, T_i^* = \mathbf{0}, G, X_i, X_i^*, C_i]\right), \tag{1}$$

where $T_i^* = \mathbf{1}$ indicates scenarios where neighbors are treated, and $T_i^* = 0$ represents untreated neighbor scenarios. The term C_i accounts for potential confounding biases due to homophily. Since C_i is unobserved, directly applying Eq. 1 is infeasible.

In the following section, we describe disentangling spillover from homophily in Section 4. By incorporating structured spillover estimation into simulation analysis, we provide policymakers with a clearer understanding of intervention effects, ensuring proper attribution of observed infection reductions.

4 SPILLOVER ESTIMATION IN SIMULATION-BASED POLICY ANALYSIS

4.1 Challenges in Estimating Spillover Effects

While simulations provide infection counts, comparing treated and untreated regions may not accurately isolate spillover effects due to two key challenges: **i) Confounding:** Similar infection patterns across neighboring regions may arise from external factors (e.g., climate, socioeconomic conditions) rather than true spillover, leading to biased estimates. **ii) Indirect Spillover:** Effects may extend beyond direct neighbors through multi-hop interactions. Capturing such higher-order effects requires models beyond standard simulation approaches.

If these challenges are not properly addressed, observed changes in untreated regions may be incorrectly attributed to spillover effects rather than underlying homophily or other external factors. By distinguishing between direct and indirect spillover effects while accounting for confounding biases, we can provide more accurate and actionable insights for policy evaluation.

4.2 Why Adjusting for Confounding is Necessary

To calculate ψ_S from Eq. 1, we introduce a confounder-adjusted spillover estimator $\psi_{S,-C}$ for approximation, and $\psi_{S,-C}$ corrects for unobserved confounding by leveraging homophily representations (λ) extracted from the simulation graph and covariates. Since both the graph structure G and node covariates X contribute to homophily, we define a learned representation λ that serves as a sufficient proxy for both G and X.

This adjustment assumes that the representation λ_i is a sufficient summary of both observed covariates and network-induced confounding effects, thereby making C_i redundant in the estimation process.

Since C_i is unobserved, we use λ_i, learned from G and X, as a confounding proxy. Thus, we have:

$$\psi_{S,-C} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[Y_i | T_i, T_i^*, \lambda_i, \lambda_i^*, G], \tag{2}$$

where Eq. 2 allows us to adjust for hidden biases by leveraging the network structure, and λ_i is referred to as a confounder-adjusted representation since it accounts for both observed and unobserved confounding effects through homophily. This relies on the homophily assumption that connected nodes with similar attributes tend to exhibit similar unobserved characteristics, a phenomenon widely observed in networked systems(Mcpherson et al. 2001).

4.3 How Confounder Adjustment Works in Spillover Estimation

To estimate $\psi_{S,-C}$, we replace the raw infection counts in the simulation with a potential outcome function $m(\cdot)$, which takes (T,G,X) as input. Since we assume λ captures sufficient information given (G,X), we can express $\psi_{S,-C}$ as:

$$\psi_{S,-C} = \frac{1}{n}\sum_{i=1}^{n} \left(\mathbb{E}[m(T_i, T_i^*, \lambda_i, \lambda_i^*)] - \mathbb{E}[m(T_i, \mathbf{0}, \lambda_i, \lambda_i^*)]\right). \tag{3}$$

The directed nature of the Dir-GNN captures the asymmetric mobility patterns critical for spillover estimation, ensuring that the learned representation remains robust to confounding biases. Eq. 3 approximates ψ_S using λ from (G,X). Next, we describe how to estimate ψ_S from simulations.

5 ESTIMATION PROCEDURE

To empirically estimate $\psi_{S,-C}$, we compute:

$$\tilde{\psi} := \frac{1}{n}\sum_{i=1}^{n} (m(T_i, T_i^*, \lambda_i, \lambda_i^*) - m(T_i, \mathbf{0}, \lambda_i, \lambda_i^*)). \tag{4}$$

Here, $m(\cdot)$ models the expected infection count under different intervention scenarios. Since direct measurement of the true spillover effect is infeasible, we rely on $m(\cdot)$ to infer the expected infection count, making it possible to estimate spillover effects using the learned representations. λ_i is a learned representation that adjusts for unobserved confounding, and $\tilde{\psi}$ serves as an empirical estimate of spillover effects.

We estimate m and λ using observed simulation outputs (T,X,Y) and the epidemic network G, where λ is obtained through the Dir-GNN framework to account for confounding effects. This approach enables us to separate spillover from confounding, providing more accurate and interpretable policy evaluations.

5.1 Graph-Based Representation Learning for Confounder Adjustment

Spillover effects in simulation outputs are influenced by unobserved confounders C_i, which may bias estimates. While direct measurement of C_i is infeasible, prior research (Mcpherson et al. 2001) suggests that confounding effects are embedded in the network structure (homophily). We use this property to learn a confounder-adjusted representation λ_i.

To extract λ, we employ a Directed Graph Neural Network (Dir-GNN), which captures directional and weighted epidemic transmission patterns:

$$\tilde{\lambda} = \text{Dir-GNN}(G,X), \tag{5}$$

where Dir-GNN generates node embeddings that approximate unobserved confounders.

5.2 Directed Graph Neural Network for Mobility-Aware Spillover Estimation

Epidemic transmission follows population movement, making directionality crucial for estimating spillover. Existing methods often assume undirected graphs (Cristali and Veitch 2022; Ma et al. 2021), which fail to model directed disease spread. To incorporate mobility-aware transmission, we modify the Dir-GCN framework (Rossi et al. 2023).

We define message-passing matrices for in- and out-neighbors:

$$\mathbb{M}_{\rightarrow} = D_{\rightarrow}^{-\frac{1}{2}} A D_{\leftarrow}^{-\frac{1}{2}}, \quad (\mathbb{M}_{\rightarrow})_{ij} = \frac{w_{ij} A_{ij}}{\sqrt{d_i^{\rightarrow} d_j^{\leftarrow}}}, \tag{6}$$

where w_{ij} represents the normalized mobility intensity from region i to j. Also, $d_i^{\rightarrow}$ and $d_j^{\leftarrow}$ denote the out-degree of node i and the in-degree of j, respectively. Node embeddings are updated as:

$$\lambda^k = \sigma(\mathbb{M}_{\rightarrow} \lambda^{k-1} W_{k-1}^{\rightarrow} + \mathbb{M}_{\rightarrow}^T \lambda^{k-1} W_{k-1}^{\leftarrow}). \tag{7}$$

This ensures that $\tilde{\lambda}$ encodes confounding patterns embedded in the simulation network.

5.3 Joint Learning of Treatment and Outcome Models

To estimate $\tilde{\psi}$, we simultaneously learn: **i). Treatment Model:** Predicts T_i to approximate unobserved confounders. **ii.) Outcome Model:** Predicts Y_i to estimate spillover effects. Correspondingly, the learning objective is:

$$L(\lambda, \gamma) = \frac{1}{n} \sum_{i=1}^{n} (Y_i - m(T_i, T_i^*, \lambda_i, \lambda_i^*; \gamma))^2 + \frac{1}{n} \sum_{i=1}^{n} [T_i \log p_i + (1 - T_i) \log(1 - p_i)] + \varepsilon \|\Theta\|^2. \tag{8}$$

Where the first term minimizes the mean squared error (MSE) for outcome prediction, the second term applies cross-entropy loss for treatment prediction, ensuring that confounder embeddings correctly predict T_i, and the third term is L2 regularization to prevent overfitting. In this loss, p_i is the predicted probability of treatment for node i, and Θ includes all learnable parameters in both outcome and treatment models.

5.4 Final Estimation of Spillover Effects

After training the model by minimizing:

$$\tilde{\lambda}, \tilde{\gamma} = \arg\min_{\lambda, \gamma} \mathbb{E}[L(\lambda, \gamma) | G, X], \tag{9}$$

we define:

$$\tilde{m}(T_i, T_i^*, \tilde{\lambda}_i, \tilde{\lambda}_i^*) = m(T_i, T_i^*, \lambda_i, \lambda_i^*; \tilde{\gamma}). \tag{10}$$

Finally, we estimate the spillover effect as:

$$\tilde{\psi} = \frac{1}{n} \sum_{i=1}^{n} (\tilde{m}(T_i, T_i^*, \tilde{\lambda}_i, \tilde{\lambda}_i^*) - \tilde{m}(T_i, \mathbf{0}, \tilde{\lambda}_i, \tilde{\lambda}_i^*)). \tag{11}$$

This provides an empirical estimate of causal spillover effects in simulation-based policy analysis.

6 SPILLOVER ESTIMATION AND SIMULATION-BASED POLICY DESIGN

While our framework operates as a post-simulation analysis tool, its output has direct implications for how simulation results are interpreted and used in policy planning. By quantifying spillover effects, we offer a layer of insight that traditional simulation outputs cannot provide, enabling more informed decision-making in the following ways:

- **Interpreting Networked Intervention Effects:** Our estimator separates direct and spillover effects, improving intervention attribution. This decomposition enables policymakers to avoid misattributing effects and to better understand how interventions propagate through a mobility network.
- **Guiding Intervention Scope:** When spillover effects are large, this indicates that local interventions produce meaningful indirect benefits—supporting the case for regionally coordinated policies. Conversely, minimal spillover suggests that targeted, node-specific interventions may be sufficient.
- **Improving Policy Attribution in Simulations:** Epidemic simulations are often used to evaluate hypothetical interventions. However, without spillover-aware analysis, these evaluations may conflate direct and indirect effects. Our framework adds interpretability by clarifying the true drivers of observed benefits within the simulated environment.
- **Supporting Resource Prioritization:** By quantifying both direct and spillover effects, our method allows policymakers to identify nodes that not only benefit from intervention but also produce large positive externalities for others. This supports more efficient resource allocation across the network.

7 EXPERIMENT

We present the simulation and experimental setup in the context of epidemiology, where we aim to evaluate not only the ability of our estimator to capture spillover effects, but also how **accounting for spillover alters the evaluation of policy effectiveness**. Our central hypothesis is that **spillover effects meaningfully influence intervention outcomes**, and **ignoring them can lead to misleading policy conclusions**.

7.1 Metapopulation Simulation

Metapopulation models are widely used in epidemiology to capture spatial heterogeneity in disease dynamics, balancing individual-based and population-based approaches. A notable implementation is PatchSim (Venkatramanan et al. 2021), which uses the SEIR framework to generate population counts for each node at different timestamps. The publicly available PatchSim code serves as the foundation for our simulation setup:

- **Duration**: The simulation runs for 100 timestamps.
- **Nodes**: We model 133 Virginia counties, each with realistic population sizes.
- **Confounder and Treatment**: Following (Ma et al. 2023), we simulate unobserved confounders as $C_i \sim \mathcal{N}(0, \mu \mathbf{I})$, where $\mu = 20$. Treatment assignment follows $BI(\cdot)$, a sigmoid function, to convert the input to a probability and then samples the output using the Bernoulli distribution. Specifically, it takes both self-node and neighboring nodes into consideration:

$$T_i = BI(\theta_{t,x}^T X_i + \theta_{t,ci}^T C_i + \theta_{t,cj}^T C_j + \varepsilon_t), \tag{12}$$

  where $\theta_{t,x} \in \mathbb{R}^{d_x}$, $\theta_{t,ci}, \theta_{t,cj} \in \mathbb{R}^{d_c}$ are parameters drawn from $\mathcal{N}(0, 0.5^2)$, and ε_t is the random Gaussian noise $\varepsilon_t \sim \mathcal{N}(0, 0.01^2)$. Covariates X_i come from (Zhang et al. 2023), which include 23 features related to demographics and economic indicators.
- **Graph Construction**: Edges are generated based on confounder similarity, where dis is the Euclidean distance function, thus we can simulate a county-county graph based on the latent confounding between nodes, and we selected *scale* to be 0.1:

$$P(A_{ij} = 1) = e^{-\frac{\text{dis}(c_i, c_j)}{\text{scale}}}, \quad \text{scale} = 0.1. \tag{13}$$

- **Treatment Mechanism**: The treated counties are assumed to have a reduced (halved in our simulation) exposure rate, resulting in fewer infections. The treatment is considered effective during a specific time interval, after which it is discontinued and the exposure rate is restored.
- **Initial Cases**: We assign each county 10 initial cases as "seed" cases.

- **Commuting on Graphs**: We model county-level mobility by assuming that 95% of a county's population remains local, while the remaining 5% is evenly distributed to its neighbors. Due to population size differences, the number of commuters between counties $i \to j$ and $j \to i$ varies, leading to asymmetric edge weights. This 95%/5% split is a simplified assumption and can be adjusted as needed.
- **Other Parameters**: We set the exposure, infection, and recovery rates to 0.65, 0.67, and 0.4, respectively, to create a realistic epidemic scenario for meaningful spillover analysis. These values are illustrative and can be adjusted for customized settings.

7.2 Spillover Estimation and Metrics

Following the prior evaluation strategies in semi-synthetic causal studies (Ma et al. 2023), we simulate 50 randomized treatment settings and evaluate across multiple timestamps. To estimate spillover, we consider counterfactual outcomes where neighboring treatments are removed. The average spillover effect is:

$$\psi_S = \frac{1}{n}\sum_{i=1}^{n}(Y_{i,c} - Y_{i,a}), \tag{14}$$

where $Y_{i,a}$ is the factual outcome for node i, and $Y_{i,c}$ is the counterfactual outcome where its neighbors are untreated. The estimated counterfactual outcome is produced by our model:

$$\tilde{Y}_{i,c} = \tilde{m}(t_i, \mathbf{0}, \tilde{\lambda}_i, \tilde{\lambda}_i^*), \quad \tilde{Y}_c = \frac{1}{n}\sum_{i=1}^{n}\tilde{Y}_{i,c}.$$

We evaluate spillover estimation quality using the following metrics: **Spillover Estimation Ratio (SER):** $\mathrm{SER}(i) = \frac{(\tilde{Y}_{i,c} - Y_{i,a}) \cdot 1[\tilde{Y}_{i,c} > Y_{i,a}]}{Y_{i,c} - Y_{i,a}}$.. We report $\mathrm{ASER} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{SER}(i)$, where values closer to 1 indicate accurate estimation. Values less than 1 suggest underestimation, while values greater than 1 indicate overestimation.

Additionally, we also design another metric **Coverage Rate (CR)**: $\mathrm{CR} = \frac{\sum_i^N ((\tilde{Y}_{i,c} - Y_{i,a}) > 0) \wedge ((Y_{i,c} - Y_{i,a}) > 0)}{N}$, measuring the proportion of counties with consistent spillover sign between ground truth and prediction.

7.3 Results and Interpretation

Table 1 summarizes model performance. We observe that spillover effects are most accurately captured between +15 and +30 timestamps post-treatment, as indicated by ASER and CR values. Performance drops slightly at later timestamps, which can be attributed to diffusion saturation, a state where the majority of susceptible individuals have already been exposed or infected, limiting the further spread of the intervention's effect. This saturation effect makes it harder to differentiate spillover from natural epidemic decline, leading to a slight overestimation.

Table 1: Experiment Performance Table

		Evaluation Timestamp					
Treatment Period	Metrics	+1	+10	+15	+20	+30	+40
0-5	ASER	0.24 ± 0.11	0.79 ± 0.18	$\mathbf{0.95 \pm 0.09}$	1.12 ± 0.18	1.32 ± 0.21	1.77 ± 0.37
	CR	0.31 ± 0.14	0.53 ± 0.09	$\mathbf{0.94 \pm 0.04}$	0.91 ± 0.07	0.83 ± 0.11	0.76 ± 0.20
0-10	ASER	0.22 ± 0.12	0.91 ± 0.13	$\mathbf{1.01 \pm 0.09}$	1.02 ± 0.13	1.30 ± 0.18	2.12 ± 0.35
	CR	0.29 ± 0.15	0.89 ± 0.09	$\mathbf{0.97 \pm 0.03}$	0.89 ± 0.09	0.85 ± 0.13	0.62 ± 0.27
0-15	ASER	0.61 ± 0.19	0.92 ± 0.09	0.94 ± 0.16	$\mathbf{1.02 \pm 0.10}$	$\mathbf{1.02 \pm 0.05}$	1.89 ± 0.23
	CR	0.51 ± 0.08	0.95 ± 0.03	$\mathbf{0.97 \pm 0.03}$	0.94 ± 0.03	0.95 ± 0.05	0.71 ± 0.16

7.4 Policy Evaluation and Attribution Effects

Conventional policy evaluations often assess intervention effectiveness by measuring infection reduction in treated regions alone, implicitly assuming no spillover. However, in epidemic settings with inter-region

Table 2: Comparison between Baseline Methods

Treatment	Evaluation Timestamp									
0-5	+10		+15		+20		+30		+40	
Method	ASER	CR	ASER	CR	ASER	CR	ASER	CR	ASER	CR
Ours	**0.79**	53%	**0.95**	**94%**	**1.12**	**91%**	**1.32**	**83%**	1.77	**76%**
PIE	0.62	**56%**	0.94	88%	1.54	81%	1.78	74%	2.46	60%
PSM	0.23	19%	0.48	61%	0.62	64%	0.51	51%	**0.41**	32%
CEVAE	0.42	39%	0.49	66%	0.56	73%	1.69	79%	2.94	51%
DNDC	1.71	31%	1.39	73%	1.53	79%	1.80	69%	2.84	45%
Netdeconf	0.58	49%	1.10	91%	1.40	75%	1.99	70%	3.01	53%
LR-Net	0.39	27%	0.47	27%	0.48	25%	0.41	29%	0.40	28%

mobility, interventions frequently yield benefits beyond directly treated nodes. Ignoring this spillover can result in incorrect causal attribution and misleading estimates of policy impact.

To illustrate this, consider a simplified example. Suppose node i would experience 200 infections if left untreated. If only node i is treated while its neighbors remain untreated, the infection count drops to 100. However, if both node i and its neighbors are treated, the observed outcome is 30. Of 170 cases prevented, 100 are direct, and 70 are due to spillover, yet simulations credit all to direct intervention. A traditional evaluation that ignores this distinction would falsely credit the full reduction to node i's treatment, overestimating its effectiveness.

This misattribution is not hypothetical—it appears in our simulation as well. Under the treatment window [0,5] and evaluation timestamp +15, the average factual outcome for treated nodes is **446**, while our model estimates that the counterfactual outcome—had neighboring treatments been removed—would be **637**. This yields an average spillover effect of **191** infections per node. Without this decomposition, policy evaluations would risk drawing incorrect conclusions about which nodes are most responsive to treatment, potentially leading to ineffective or inefficient intervention strategies.

Our estimator enables this causal decomposition, providing a more accurate attribution of treatment outcomes. Spillover-aware evaluation is thus essential for designing scalable and effective public health policies, particularly in networked and spatially connected systems like epidemics.

7.5 Comparison with Baselines

We compare our method with several representative spillover estimation methods: **PIE** (Cristali and Veitch 2022): Relational ERM-based estimator for peer effects on undirected networks. **DNDC** (Ma et al. 2021) and **NetDeconf** : Learns latent confounders via historical and structural data. **CEVAE** (Louizos et al. 2017), **Propensity Score Models (PSM)**, **Linear Regression on Network Features (LR-Net)**: Standard causal estimation baselines.

All models are evaluated on the same treatment window (0–5 timestamps) and tested at multiple future timestamps. As shown in Table 2, our model consistently achieves the most accurate spillover estimates across both ASER and CR, particularly at +15 and +30. While PIE performs comparably at +10, its design for undirected graphs limits its consistency.

Baselines such as CEVAE and PSM underperform in our setup because they assume static treatment-response relationships or unconfounded treatment assignment. These assumptions break down in epidemic simulations where dynamic mobility and interference violate standard ignorability.

8 CONCLUSION

We introduce a simulation-centered framework for estimating causal spillover effects in spatial epidemic networks. Built on top of PatchSim, a metapopulation SEIR (Susceptible–Exposed–Infectious–Recovered) simulator with county-level mobility, our approach applies a directional GNN-based estimator to distinguish direct effects from spillover while adjusting for homophily and latent confounding. Experiments on semi-synthetic simulations demonstrate that our estimator accurately recovers counterfactual outcomes and yields

more reliable attribution of treatment impact. Critically, we show that standard simulation outputs can mislead policy evaluations when spillover is ignored. By enabling post-simulation causal analysis, our framework bridges the gap between epidemic simulation and causal inference, offering a more interpretable and policy-relevant perspective on intervention effectiveness.

REFERENCES

Aronow, P. M., and C. Samii. 2017. "Estimating Average Causal Effects Under General Interference, with Application to a Social Network Experiment". *The Annals of Applied Statistics* 11(4).

Benjamin-Chung, J., B. F. Arnold, D. Berger, S. P. Luby, E. Miguel, J. M. Colford Jr *et al*. 2017. "Spillover Effects in Epidemiology: Parameters, Study Designs and Methodological Considerations". *International Journal of Epidemiology* 47(1):332–347.

Bhattacharya, P., J. Chen, S. Hoops, D. Machi, B. Lewis, S. Venkatramanan, *et al*. 2023. "Data-driven Scalable Pipeline Using National Agent-based Models For Real-time Pandemic Response and Decision Support". *The International Journal of High Performance Computing Applications* 37(1):4–27.

Bhattacharya, P., D. Machi, J. Chen, S. Hoops, B. Lewis, H. Mortveit, *et al*. 2021. "AI-Driven Agent-Based Models to Study the Role of Vaccine Acceptance in Controlling COVID-19 Spread in the US". In *2021 IEEE International Conference on Big Data (Big Data)*, 1566–1574.

Buchanan, A. L., R. U. Hernández-Ramírez, J. J. Lok, S. H. Vermund, S. R. Friedman, L. Forastiere *et al*. 2024. "Assessing Direct and Spillover Effects of Intervention Packages in Network-Randomized Studies". *Epidemiology* 35(4):481–488.

Chang, S., D. Vrabac, J. Leskovec, and J. Ugander. 2023. "Estimating Geographic Spillover Effects of COVID-19 Policies from Large-Scale Mobility Networks". *Proceedings of the AAAI Conference on Artificial Intelligence* 37(12):14161–14169.

Chen, J., P. Bhattacharya, S. Hoops, D. Machi, A. Adiga, H. Mortveit, *et al*. 2024. "Role of Heterogeneity: National Scale Data-driven Agent-based Modeling for the US COVID-19 Scenario Modeling Hub". *Epidemics* 48:100779.

Chen, J., S. Hoops, H. S. Mortveit, B. L. Lewis, D. Machi, P. Bhattacharya, *et al*. 2024. "Epihiper—A High Performance Computational Modeling Framework to Support Epidemic Science". *PNAS Nexus* 4(1):pgae557.

Chen, Y., Y. R. Gel, M. V. Marathe, and H. V. Poor. 2024. "A Simplicial Epidemic Model for COVID-19 Spread Analysis". *Proceedings of the National Academy of Sciences* 121(1):e2313171120.

Cristali, I., and V. Veitch. 2022. "Using Embeddings for Causal Estimation of Peer Influence in Social Networks". In *Advances in Neural Information Processing Systems*, Volume 35, 15616–15628.

Eckles, D., R. Kizilcec, and E. Bakshy. 2016. "Estimating Peer Effects in Networks with Peer Encouragement Designs". *Proceedings of the National Academy of Sciences* 113:7316–7322.

Fatemi, Z., and E. Zheleva. 2020a. "Minimizing Interference and Selection Bias in Network Experiment Design". *Proceedings of the International AAAI Conference on Web and Social Media* 14:176–186.

Fatemi, Z., and E. Zheleva. 2020b. "Network Experiment Design for Estimating Direct Treatment Effects". In *KDD International Workshop on Mining and Learning with Graphs (MLG 2020)*.

Feng, F., W. Huang, X. He, X. Xin, Q. Wang, and T.-S. Chua. 2021. "Should Graph Convolution Trust Neighbors? A Simple Causal Inference Method". In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1208–1218.

He, C.-Q., B.-H. Sun, W.-T. Yu, S.-Y. An, B.-J. Qiao, and W. Wu. 2024. "Evaluating The Impact of COVID-19 Outbreak on Hepatitis B and Forecasting the Epidemiological Trend in Mainland China: A Causal Analysis". *BMC Public Health* 24(1):47.

Hoops, S., J. Chen, A. Adiga, B. Lewis, H. Mortveit, H. Baek, *et al*. 2021. "High Performance Agent-Based Modeling to Study Realistic Contact Tracing Protocols". In *2021 Winter Simulation Conference (WSC)*, 1–12 https://doi.org/10.1109/WSC52266.2021.9715382.

Hoshino, T., and T. Yanagi. 2024. "Causal Inference with Noncompliance and Unknown Interference". *Journal of the American Statistical Association* 119(548):2869–2880.

Huang, Z., J. Hwang, J. Zhang, J. Baik, W. Zhang, D. Wodarz, *et al*. 2024. "Causal Graph ODE: Continuous Treatment Effect Modeling in Multi-agent Dynamical Systems". In *Proceedings of the ACM on Web Conference 2024*, WWW '24, 4607–4617.

Jiang, S., Z. Huang, X. Luo, and Y. Sun. 2023. "CF-GODE: Continuous-Time Causal Inference for Multi-Agent Dynamical Systems". In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '23, 997–1009.

Kaminsky, J., L. T. Keegan, C. J. E. Metcalf, and J. Lessler. 2019. "Perfect Counterfactuals for Epidemic Simulations". *Philosophical Transactions of the Royal Society B: Biological Sciences* 374(1776):20180279.

Kermack, W. O., and A. G. McKendrick. 1927. "A Contribution to the Mathematical Theory of Epidemics". *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115(772):700–721.

Laffers, L., and G. Mellace. 2020. "Identification of the average Treatment Effect when SUTVA is Violated". Discussion Papers on Economics 3/2020, University of Southern Denmark, Department of Economics.

Lawson, A., and C. Rotejanaprasert. 2023. "Bayesian Spatio-Temporal Prediction and Counterfactual Generation: An Application in Non-Pharmaceutical Interventions in COVID-19". *Viruses* 15(2).

Lee, T., A. L. Buchanan, N. Katenka, L. Forastiere, M. E. Halloran, and G. Nikolopoulos. 2023. "Evaluating Spillover Effects in Network-Based Studies in the Presence of Missing Outcomes". *arXiv preprint arXiv:2304.09361*. Preprint.

Liu, H., X. Xu, X. Deng, Z. Hu, R. Sun, J. Zou, *et al*. 2024. "Counterfactual Analysis of the 2023 Omicron XBB Wave in China". *Infectious Disease Modelling* 9(1):195–203.

Louizos, C., U. Shalit, J. M. Mooij, D. Sontag, R. Zemel, and M. Welling. 2017. "Causal Effect Inference with Deep Latent-Variable Models". *Advances in neural information processing systems* 30.

Ma, J., C. Chen, A. Vullikanti, R. Mishra, G. Madden, D. Borrajo *et al*. 2023. "A Look into Causal Effects under Entangled Treatment in Graphs: Investigating the Impact of Contact on MRSA Infection". In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 4584–4594.

Ma, J., R. Guo, C. Chen, A. Zhang, and J. Li. 2021. "Deconfounding with Networked Observational Data in a Dynamic Environment". *Proceedings of the The Fourteenth ACM International Conference on Web Search and Data Mining*.

Ma, Y., and V. Tresp. 2020. "Causal Inference under Networked Interference and Intervention Policy Enhancement". In *International Conference on Artificial Intelligence and Statistics*.

Manski, C. F. 2013. "Identification of Treatment Response with Social Interactions". *The Econometrics Journal* 16(1):S1–S23.

McFowland, E., and C. R. Shalizi. 2021. "Estimating Causal Peer Influence in Homophilous Social Networks by Inferring Latent Locations". *Journal of the American Statistical Association* 118(541):707–718.

Mcpherson, M., L. Smith-Lovin, and J. Cook. 2001. "Birds of a Feather: Homophily in Social Networks". *Annu Rev Sociol* 27:415–444.

Moon, S. A., J. Chen, B. Espinoza, B. Lewis, M. Marathe, J. Outten, *et al*. 2024. "Agent-Based Simulation Framework for Multi-Variant Surveillance". In *2024 Winter Simulation Conference (WSC)*, 276–287 https://doi.org/10.1109/WSC63780.2024.10838920.

Nitzsche, C., and S. Simm. 2024. "Agent-Based Modeling to Estimate the Impact of Lockdown Scenarios and Events on A Pandemic Exemplified on SARS-CoV-2". *Scientific Reports* 14(1):13391.

Ogburn, E. L., O. Sofrygin, I. Díaz, and M. J. van der Laan. 2024. "Causal Inference for Social Network Data". *Journal of the American Statistical Association* 119(545):597–611.

Papadogeorgou, G., and S. Samanta. 2024. "Spatial Causal Inference in the Presence of Unmeasured Confounding and Interference". *arXiv preprint arXiv:2303.08218*.

Qian, Y., É. Marty, A. Basu, E. B. O'Dea, X. Wang, S. Fox, *et al*. 2025. "Physics-Informed Deep Learning for Infectious Disease Forecasting". *arXiv preprint arXiv:2501.09298*.

Robertson, C., C. Safta, N. Collier, J. Ozik, and J. Ray. 2025. "Bayesian Calibration of Stochastic Agent Based Model via Random Forest". *Statistics in Medicine* 44(6):e70029.

Rossi, E., B. Charpentier, F. Di Giovanni, F. Frasca, S. Günnemann, and M. Bronstein. 2023. "Edge Directionality Improves Learning on Heterophilic Graphs". *arXiv preprint arXiv:2305.15398*.

Rubin, D. B. 2005. "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions". *Journal of the American Statistical Association* 100(469):322–331.

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

Sridhar, D., C. De Bacco, and D. Blei. 2022. "Estimating Social Influence from Observational Data". In *Conference on Causal Learning and Reasoning*, 712–733.

Tchetgen, E. J. T., I. R. Fulcher, and I. Shpitser. 2021. "Auto-G-Computation of Causal Effects on a Network". *Journal of the American Statistical Association* 116(534):833–844.

Tec, M., A. Trisovic, M. Audirac, S. M. Woodward, J. K. Hu, N. Khoshnevis *et al*. 2024. "SpaCE: The Spatial Confounding Environment". In *The Twelfth International Conference on Learning Representations*.

Toulis, P., and E. Kao. 2013. "Estimation of Causal Peer Influence Effects". In *Proceedings of the 30th International Conference on Machine Learning*, Volume 28, 1489–1497.

van der Laan, M. J. 2014. "Causal Inference for a Population of Causally Connected Units". *Journal of Causal Inference* 2:13–74.

Veitch, V., Y. Wang, and D. M. Blei. 2019. "Using Embeddings to Correct for Unobserved Confounding in Networks". In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*.

Venkatramanan, S., A. Sadilek, A. Fadikar, C. L. Barrett, M. Biggerstaff, J. Chen, *et al*. 2021. "Forecasting Influenza Activity Using Machine-Learned Mobility Map". *Nature Communications* 12(1):726.

Zhang, L., A. Darzi, S. Ghader, M. L. Pack, C. Xiong, M. Yang, *et al*. 2023. "Interactive COVID-19 Mobility Impact and Social Distancing Analysis Platform". *Transportation Research Record* 2677(4):168–180.

Zhang, Z., H. Wang, Z. Fan, X. Song, and R. Shibasaki. 2024, November. "Assessing the Spatial-Temporal Causal Impact of COVID-19-Related Policies on Epidemic Spread". *ACM Trans. Knowl. Discov. Data* 19(1).

AUTHOR BIOGRAPHIES

JINGYUAN CHOU is a Ph.D. in Computer Science at the Biocomplexity Institute, University of Virginia, US. His background is focused on causal inference and machine learning. His interests include causal effect estimation and graph neural networks. His email address is jc2wv@virginia.edu.

JIANGZHUO CHEN is a research associate professor in Biocomplexity Institute, University of Virginia, US. Chen received his BA in Economics from Nanjing University, MA. in Economics from Boston College, and PhD in Computer Science from Northeastern University. Dr. Chen's dissertation explores confluent flows. Chen was a post-doctoral researcher and a senior research associate in the Network Dynamics and Simulation Science Laboratory at the Biocomplexity Institute of Virginia Tech. Chen's work has been published in journals and peer-reviewed conferences, such as the Journal of the ACM, ACM Transactions on Modeling and Computer Simulation, PLOS ONE, Computational Economics, JEBO, STOC, and ICS. His e-mail address is jc3qw@virginia.edu.

MADHAV MARATHE is an endowed Distinguished Professor of Biocomplexity, Executive Director of the Biocomplexity Institute, and a tenured Professor of Computer Science at the University of Virginia. Dr. Marathe is a passionate advocate and practitioner of transdisciplinary team science. During his 30-year professional career, he has established and led many large transdisciplinary projects and groups. His areas of expertise include digital twins, network science, artificial intelligence, multi-agent systems, high-performance computing, computational epidemiology, biological and socially coupled systems, and data analytics. He is a Fellow of the American Association for the Advancement of Science (AAAS), Society for Industrial and Applied Mathematics (SIAM), Association for Computing Machinery (ACM), and Institute of Electrical and Electronics Engineers (IEEE). His email address is mvm7hz@virginia.edu.