Graph Diffusion History Reconstruction via Feasibility-Aware Markov Chain Monte Carlo Estimation

Yijing Zuo*
yzuo7@illinois.edu
University of Illinois
Urbana–Champaign, IL, USA

Ruizhong Qiu*
rq5@illinois.edu
University of Illinois
Urbana–Champaign, IL, USA

Lingjie Chen
lingjie7@illinois.edu
University of Illinois
Urbana–Champaign, IL, USA

Hanghang Tong
htong@illinois.edu
University of Illinois
Urbana–Champaign, IL, USA

Abstract

Diffusion dynamics on graphs arise across many fields including information spreading and rumor cascades in online platforms, propagation of cascading outages in power and transportation infrastructures, diffusion of behaviors and product adoption in social networks, and transmission of shocks in financial and supply-chain systems. Graph diffusion provides a compact representation of how states propagate through interacting entities, yet in many applications the diffusion history is not fully observed. Typically, only a small set of snapshots are available while all other states are missing. Diffusion history reconstruction is challenging due to explosive search space, complex combinatorial constraints, and scarcity of training data. To address these challenges, we propose a new method called HERMES. HERMES has two main stages: (i) diffusion parameter estimation and (ii) diffusion history reconstruction. The first stage is to estimate the unknown diffusion parameters from the observed snapshots. To bypass the intractable maximum likelihood estimation of diffusion parameters, we instead propose a tractable mean-field approximation to estimate diffusion parameters. Second, based on the estimated diffusion parameters, we theoretically reduce history reconstruction to expected hitting time estimation through a bias–variance decomposition and estimate the expected hitting times via Metropolis–Hastings Markov chain Monte Carlo (M–H MCMC). The core component of M–H MCMC is the proposal distribution, and our proposal distribution handles the complex combinatorial constraints via a dynamic reachability mechanism that ensures compatibility with all observed snapshots. Moreover, to further enhance M–H MCMC, we parameterize the proposal using a graph neural network (GNN) and train the GNN to match the posterior distribution. Extensive experiments demonstrate that HERMES consistently outperforms existing methods on 12 synthetic and real-world datasets. Theoretical proofs are deferred to Appendix B.

*Equal contribution.

KDD '26, Jeju Island, Republic of Korea

ACM ISBN 979-8-4007-2259-2/2026/08
https://doi.org/10.1145/3770855.3818096

CCS Concepts

• **Mathematics of computing** → **Graph algorithms**; • **Computing methodologies** → *Neural networks*.

Keywords

Graph Diffusion, History Reconstruction, Markov Chain Monte Carlo (MCMC), Graph Neural Network (GNN)

ACM Reference Format:
Yijing Zuo, Ruizhong Qiu, Lingjie Chen, and Hanghang Tong. 2026. Graph Diffusion History Reconstruction via Feasibility-Aware Markov Chain Monte Carlo Estimation. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD '26), August 09–13, 2026, Jeju Island, Republic of Korea.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3770855.3818096

1 Introduction

Diffusion dynamics on graphs offer a compact way to describe how states propagate through interacting entities, and they arise across many fields. Typical examples include information spreading and rumor cascades in online platforms [65], propagation of cascading outages in power and transportation infrastructures [19], diffusion of behaviors and product adoption in social networks [5], and transmission of shocks in financial and supply-chain systems [22]. In these applications, *complete* diffusion histories are desired because they support tasks such as uncovering latent spreading patterns [37], assessing mitigation strategies [53], and forecasting the impact of interventions [50].

However, complete histories are rarely available, and typically only a few intermittent snapshots are accessible. This is because: (i) early stages are easy to miss; (ii) continuous sensing is expensive; and (iii) fine-grained tracing can be privacy-sensitive [10, 54]. Motivated by such practical considerations, we study the problem of diffusion history reconstruction from sparse observations: given only a few observed snapshots, we aim to reconstruct the complete diffusion history. Compared with the large body of work on forward diffusion tasks (e.g., [8, 16, 23, 32, 47, 57]), diffusion history reconstruction has received far less attention despite its critical importance. There are three major challenges in this problem: **(i) Explosive search space:** The number of possible histories is exponentially large w.r.t. the graph size and the timespan. **(ii) Complex combinatorial constraints:** The multiple observed snapshots impose complex combinatorial constraints on the unobserved history.

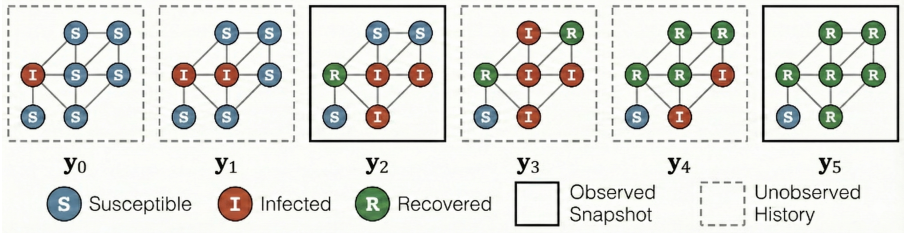


Figure 1: Diffusion history reconstruction problem: Given multiple observed snapshots (i.e., $\mathbf{y}_2$ and $\mathbf{y}_5$), we aim to reconstruct the unobserved diffusion history (i.e., $\mathbf{y}_0, \mathbf{y}_1, \mathbf{y}_3, \mathbf{y}_4$).

(iii) Scarcity of training data: Supervised methods for time series imputation (e.g., [14, 45, 61]) require complete diffusion histories as training data, but they are rarely available in practice.

To address the aforementioned challenges, we propose *History rEconstRuction from MultiplE Snapshots* (HERMES). HERMES has two main stages: (i) diffusion parameter estimation and (ii) diffusion history reconstruction. The first stage is to estimate the unknown diffusion parameters from the observed snapshots. Since the exact maximum likelihood estimation of diffusion parameters is known to be intractable (NP-hard, [49]), we propose a tractable mean-field approximation as pseudolikelihood and estimate diffusion parameters by maximizing the joint pseudolikelihood. Second, based on the estimated diffusion parameters, we theoretically reduce history reconstruction to expected hitting time estimation through a bias–variance decomposition and estimate the expected hitting times via Metropolis–Hastings Markov chain Monte Carlo (M–H MCMC). The core component of M–H MCMC is the proposal distribution, and our proposal distribution handles the complex combinatorial constraints via a dynamic reachability mechanism that ensures compatibility with all observed snapshots. Moreover, to further enhance M–H MCMC, we parameterize the proposal using a graph neural network (GNN) and train the GNN to match the posterior distribution. Experiments on 12 synthetic and real-world datasets show that HERMES significantly outperforms existing methods and remains scalable as the graph size and time horizon grow.

Main contributions. The main contributions of this paper are:

- **Inverse problem.** We study an important yet challenging inverse problem: diffusion history reconstruction. In this problem, only a small number of snapshots are observed, and we need to reconstruct the complete diffusion history.
- **Novel methodology.** First, we propose a tractable mean-field approximation to bypass the intractable maximum likelihood estimation of diffusion parameters. Additionally, we design a nontrivial proposal distribution for M–H MCMC such that proposed histories are theoretically guaranteed to be feasible and compatible with observed snapshots. Moreover, to further enhance M–H MCMC, we parameterize the proposal using a GNN and train the GNN to match the posterior distribution.
- **Theoretical guarantee.** Our proposal distribution is theoretically guaranteed to be compatible with observed snapshots. We also show that the complexity of our proposed HERMES is polynomial w.r.t. the graph size and linear w.r.t. the timespan.
- **Empirical performance.** We conduct extensive experiments on synthetic and real-world datasets. The results show that our proposed HERMES significantly outperforms existing methods and remains scalable as the graph size and time horizon grow.

Table 1: Nomenclature.

Symbol	Definition
$\mathcal{X}$	the set of diffusion states
$\mathcal{V}$	the set of nodes
$\mathcal{E}$	the set of edges
$\mathcal{N}_u$	the set of neighbors of node u
T	the timespan of interest
$\mathcal{T}$	the set of times
$\mathcal{T}_{\text{obs}}$	the set of observation times, $\mathcal{T}_{\text{obs}} \subseteq \mathcal{T}$
$y_{t,u}$	the state of node u at time t
$\mathbf{y}_t$	a snapshot of diffusion at time t
$[0, t_1]$, $(t_1, t_2]$, etc.	time segments
$\mathbf{y}_{\mathcal{T}_{\text{obs}}}$	the observed snapshots $\{\mathbf{y}_t\}_{t \in \mathcal{T}_{\text{obs}}}$
$\mathbf{Y}$	a complete diffusion history
$\widehat{\mathbf{Y}}$	the reconstructed diffusion history
$\mathcal{V}^{\mathcal{X}_0}(\mathbf{y}_t)$, $n^{\mathcal{X}_0}(\mathbf{y}_t)$	the set / number of nodes with state in $\mathcal{X}_0 \subseteq \mathcal{X}$ in snapshot $\mathbf{y}_t$
$\mathtt{S}, \mathtt{I}, \mathtt{R}$	states in the SIR model
$\beta^{\mathtt{I}}, \beta^{\mathtt{R}}$	the infection rate and the recovery rate
$\boldsymbol{\beta}$	true diffusion parameters
$\widehat{\boldsymbol{\beta}}$	estimated diffusion parameters
$P_{\boldsymbol{\beta}}$	the probability measure of the diffusion model
$P_{\boldsymbol{\beta}} \mid \mathbf{y}_{\mathcal{T}_{\text{obs}}}$	the posterior given the observed snapshots
$\text{supp}(P)$	the set of possible histories
$\text{supp}(P \mid \mathbf{Y}_{\mathcal{T}_{\text{obs}}})$	the set of histories consistent with $\mathbf{y}_{\mathcal{T}_{\text{obs}}}$
$\mathbb{E}$	the expectation operator
$\mathbb{I}$	the indicator operator
O	the asymptotic notation

2 Problem Definition

This section formally defines the problem setting. In this paper, we use calligraphic letters for sets (e.g., $\mathcal{E}$); plain uppercase letters for constants and probability measures (e.g., T and P); plain lowercase letters for indices and scalar-valued functions (e.g., t and f); boldface lowercase/uppercase letters for vectors/matrices (e.g., $\boldsymbol{\beta}$ and $\mathbf{Y}$); monospaced font for diffusion states (e.g., $\mathtt{S}$); and a hat for estimated quantities (e.g., $\widehat{\boldsymbol{\beta}}$). Notations are summarized in Table 1.

2.1 Preliminaries

Diffusion on graphs. We consider discrete-time diffusion on an undirected graph $(\mathcal{V}, \mathcal{E})$. Let $\mathcal{T} := \{0, 1, \ldots, T\}$ be the time index set, and let $\mathcal{X}$ denote the set of diffusion states. The graph has

$|\mathcal{V}| = n$ nodes and $|\mathcal{E}| = m$ edges. For each node $u \in \mathcal{V}$, we write $\mathcal{N}_u := \{v \in \mathcal{V} : (u, v) \in \mathcal{E}\}$ for its neighbors.

Let $y_{t,u} \in \mathcal{X}$ be the state of node u at time $t \in \mathcal{T}$. A *diffusion process* [59] on graph $(\mathcal{V}, \mathcal{E})$ is a spatiotemporal stochastic process $(y_{t,u})_{t\in\mathcal{T},u\in\mathcal{V}}$ such that, for every $t > 0$ and $u \in \mathcal{V}$, the variable $y_{t,u}$ depends only on the previous-step states $\{y_{t-1,v} : v \in \{u\} \cup \mathcal{N}_u\}$. This locality implies the Markov property on the sequence of snapshots. A *snapshot* at time t is the vector $\mathbf{y}_t := (y_{t,u})_{u\in\mathcal{V}} \in \mathcal{X}^{\mathcal{V}}$. A diffusion history (or simply a history) is the trajectory

$$\mathbf{Y} := (\mathbf{y}_0, \ldots, \mathbf{y}_T)^\top = (y_{t,u})_{t\in\mathcal{T},u\in\mathcal{V}} \in \mathcal{X}^{\mathcal{T}\times\mathcal{V}}. \tag{1}$$

We call a history $\mathbf{Y}$ feasible iff it occurs with nonzero probability under the diffusion model.

Graph diffusion models. We focus on two classic diffusion models on graphs: the Susceptible–Infected (SI) model and the Susceptible–Infected–Recovered (SIR) model [33]. We start from SIR as SI is a special case of SIR. We use the total order $\mathtt{S} < \mathtt{I} < \mathtt{R}$.

In the SIR model, recovered nodes obtain permanent protection and cannot be infected again. The state space is $\mathcal{X} := \{\mathtt{S}, \mathtt{I}, \mathtt{R}\}$. The probability measure $P_{\boldsymbol{\beta}}$ is parameterized by $\boldsymbol{\beta} := (\beta^{\mathtt{I}}, \beta^{\mathtt{R}}) \in (0, 1)^2$, where $\beta^{\mathtt{I}}$ and $\beta^{\mathtt{R}}$ are the infection rate and the recovery rate, respectively. For any $\mathcal{X}_0 \subseteq \mathcal{X}$, let $\mathcal{V}^{\mathcal{X}_0}(\mathbf{y}_t)$ denote the set of nodes whose states belong to $\mathcal{X}_0$ at time t, and let $n^{\mathcal{X}_0}(\mathbf{y}_t) := |\mathcal{V}^{\mathcal{X}_0}(\mathbf{y}_t)|$; for instance, $n^{\mathtt{IR}}(\mathbf{y}_t)$ counts infected or recovered nodes. The SI model is a special case of SIR with $\beta^{\mathtt{R}} := 0$, which captures diffusion processes with irreversible infection.

By the Markov property, the probability of a history factorizes over time as

$$P_{\boldsymbol{\beta}}[\mathbf{Y}] := P[\mathbf{y}_0] \prod_{t=0}^{T-1} P_{\boldsymbol{\beta}}[\mathbf{y}_{t+1} \mid \mathbf{y}_t]. \tag{2}$$

Moreover, conditioned on $\mathbf{y}_t$, the transition probability decomposes over nodes:

$$P_{\boldsymbol{\beta}}[\mathbf{y}_{t+1} \mid \mathbf{y}_t] := \prod_{u\in\mathcal{V}} P_{\boldsymbol{\beta}}[y_{t+1,u} \mid \mathbf{y}_t]. \tag{3}$$

For each node u, the single-node transitions are

$$P_{\boldsymbol{\beta}}[y_{t+1,u} := \mathtt{S} \mid \mathbf{y}_t] := \begin{cases} \prod_{v\in\mathcal{N}_u \wedge y_{t,v}=\mathtt{I}} (1-\beta^{\mathtt{I}}), & \text{if } y_{t,u} = \mathtt{S}, \\ 0, & \text{if } y_{t,u} \in \{\mathtt{I}, \mathtt{R}\}; \end{cases}$$

$$P_{\boldsymbol{\beta}}[y_{t+1,u} = \mathtt{I} \mid \mathbf{y}_t] := \begin{cases} \Big(1 - \prod_{v\in\mathcal{N}_u \wedge y_{t,v}=\mathtt{I}} (1-\beta^{\mathtt{I}})\Big)(1-\beta^{\mathtt{R}}), & \text{if } y_{t,u} = \mathtt{S}, \\ 1-\beta^{\mathtt{R}}, & \text{if } y_{t,u} = \mathtt{I}, \\ 0, & \text{if } y_{t,u} = \mathtt{R}; \end{cases}$$

$$P_{\boldsymbol{\beta}}[y_{t+1,u} = \mathtt{R} \mid \mathbf{y}_t] := \begin{cases} \Big(1 - \prod_{v\in\mathcal{N}_u \wedge y_{t,v}=\mathtt{I}} (1-\beta^{\mathtt{I}})\Big)\beta^{\mathtt{R}}, & \text{if } y_{t,u} = \mathtt{S}, \\ \beta^{\mathtt{R}}, & \text{if } y_{t,u} = \mathtt{I}, \\ 1, & \text{if } y_{t,u} = \mathtt{R}. \end{cases} \tag{4}$$

For instance, in Figure 1, the initially infected node with state $\mathtt{I}$ in $\mathbf{y}_0$ infects a neighbor from state $\mathtt{S}$ to state $\mathtt{I}$ in $\mathbf{y}_1$ and then recovers to state $\mathtt{R}$ in $\mathbf{y}_2$.

Let $\mathrm{supp}(P) := \{\mathbf{Y} \in \mathcal{X}^{\mathcal{T}\times\mathcal{V}} : P_{\boldsymbol{\beta}}[\mathbf{Y}] > 0\}$ be the set of possible histories. For observed snapshots $\mathbf{Y}_{\mathcal{T}_{\mathrm{obs}}} \in \mathcal{X}^{K\times\mathcal{V}}$, let

$$\mathrm{supp}(P \mid \mathbf{Y}_{\mathcal{T}_{\mathrm{obs}}}) := \{\mathbf{Y} \in \mathcal{X}^{\mathcal{T}\times\mathcal{V}} : P_{\boldsymbol{\beta}}[\mathbf{Y} \mid \mathbf{Y}_{\mathcal{T}_{\mathrm{obs}}}] > 0\} \tag{5}$$

denote the set of histories consistent with $\mathbf{Y}_{\mathcal{T}_{\mathrm{obs}}}$. Since supp does not depend on $\boldsymbol{\beta}$, we omit $\boldsymbol{\beta}$ in its notation.

2.2 Problem Statement

We study an inverse problem for discrete-time diffusion on graphs, where only a few snapshots of the process are observed and the full spatiotemporal evolution is hidden. Let $\mathcal{T}_{\mathrm{obs}} = \{t_1 < t_2 < \cdots < t_K = T\}$ be a set of observed times within the timespan $\mathcal{T} := \{0, 1, \ldots, T\}$, and suppose we are given the corresponding snapshots $\{\mathbf{y}_{t_i}\}_{i=1}^K$. Our goal is to recover a complete diffusion history $\widehat{\mathbf{Y}} = (\widehat{\mathbf{y}}_0, \widehat{\mathbf{y}}_1, \ldots, \widehat{\mathbf{y}}_T)^\top$ such that (i) it is feasible under the SI/SIR dynamics, and (ii) it exactly matches every observation, i.e., $\widehat{\mathbf{y}}_{t_k} = \mathbf{y}_{t_k}$ for all $k = 1, \ldots, K$. To simplify notation, we let $t_0 := 0$, but it is not in $\mathcal{T}_{\mathrm{obs}}$.

In many applications, neither ground-truth diffusion histories nor calibrated diffusion parameters are available. Accordingly, we do not assume access to a database of histories, and we do not assume knowing the true diffusion parameters $\boldsymbol{\beta}$. Since sparse observations alone do not determine the diffusion family, we assume the underlying diffusion model (SI or SIR) as domain knowledge while treating its parameters as unknown, which is a standard assumption in prior work (e.g., [49]). Furthermore, we assume that the source nodes are unknown, and we only have rough knowledge about the prior initial distribution $P[\mathbf{y}_0]$. Specifically, we only assume knowing a rough number $n_0^{\mathtt{I}}$ of initially infected nodes. If true diffusion parameters are unknown, then such a prior is necessary for estimating diffusion parameters [49]. We assume no initially recovered nodes, since they can be removed from the graph. Hence, we define the prior initial distribution as:

$$P_{\boldsymbol{\beta}}[\mathbf{y}_0] \propto \exp\big(-\gamma\, |n^{\mathtt{I}}(\mathbf{y}_0) - n_0^{\mathtt{I}}| - \gamma\, n^{\mathtt{R}}(\mathbf{y}_0)\big), \tag{6}$$

where $\gamma > 0$ is a hyperparameter reflecting our confidence in $n_0^{\mathtt{I}}$. We do not treat $n_0^{\mathtt{I}}$ as a hard constraint since it is typically only an approximate estimate rather than an exact count.

Our problem is formally defined in Problem 2.1 and illustrated in Figure 1: given two observed snapshots $\mathbf{y}_2$ and $\mathbf{y}_5$, we aim to reconstruct the unobserved diffusion history $\mathbf{y}_0, \mathbf{y}_1, \mathbf{y}_3, \mathbf{y}_4$.

Problem 2.1 (diffusion history reconstruction). *Under the SI/SIR model, reconstruct the complete diffusion history from multiple observed snapshots without knowing true diffusion parameters.* ***Input:*** *(i) graph $(\mathcal{V}, \mathcal{E})$; (ii) timespan T of interest; (iii) observed snapshots $\{\mathbf{y}_{t_i}\}_{i=1}^K \subset \mathcal{X}^{\mathcal{V}}$ at times $\mathcal{T}_{\mathrm{obs}} = \{t_1 < \cdots < t_K = T\}$; (iv) initial distribution $P[\mathbf{y}_0]$.* ***Output:*** *reconstructed complete diffusion history $\widehat{\mathbf{Y}} \in \mathcal{X}^{\mathcal{T}\times\mathcal{V}}$.*

3 Diffusion Parameter Estimation via Mean-Field Approximation

Exact likelihood-based estimation is intractable because the snapshot likelihood marginalizes over an exponential number of feasible histories. To address the intractability, we propose to estimate diffusion parameters by maximizing a tractable mean-field pseudolikelihood. To leverage multiple observed snapshots, we further factorize this pseudolikelihood into time segments separated by observation times. In Sec. 3.1, we introduce the notation and the segmented maximum-pseudolikelihood objective. In Sec. 3.2, we derive the

Algorithm 1 HERMES Reconstruction Procedure

Input: (i) graph $G = (\mathcal{V}, \mathcal{E})$; (ii) timespan T, observed times $\mathcal{T}_{\text{obs}} \subseteq \mathcal{T}$, and observed snapshots $\mathbf{Y}_{\mathcal{T}_{\text{obs}}} = \{\mathbf{y}_t\}_{t\in\mathcal{T}_{\text{obs}}}$; (iii) estimated diffusion parameters $\widehat{\boldsymbol{\beta}}$; (iv) trained proposal $Q_{\boldsymbol{\theta}}(\cdot)$; (v) batch size L, MCMC steps S, and moving-average hyperparameter η.

Output: reconstructed diffusion history $\widehat{\mathbf{Y}}$.

sample L histories $\mathbf{Y}^{(0,1)}, \dots, \mathbf{Y}^{(0,L)} \sim Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})$

initialize the hitting time estimates: for each $u \in \mathcal{V}$

$\widehat{h}^{\mathsf{I}}_u \leftarrow \frac{1}{L}\sum_{i=1}^{L} h^{\mathsf{I}}_u(\mathbf{Y}^{(0,i)}), \quad \widehat{h}^{\mathsf{R}}_u \leftarrow \frac{1}{L}\sum_{i=1}^{L} h^{\mathsf{R}}_u(\mathbf{Y}^{(0,i)})$

for $s = 1, \dots, S$ **do**

  sample L histories $\widetilde{\mathbf{Y}}^{(s,1)}, \dots, \widetilde{\mathbf{Y}}^{(s,L)} \sim Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})$

  generate $\xi^{(s,1)}, \dots, \xi^{(s,L)} \sim \text{Uniform}[0,1)$

  update MCMC by the M–H rule: for each $i = 1, \dots, L$,

$$\mathbf{Y}^{(s,i)} \leftarrow \begin{cases} \widetilde{\mathbf{Y}}^{(s,i)} & \text{if } \xi^{(s,i)} < \frac{P_{\widehat{\boldsymbol{\beta}}}[\widetilde{\mathbf{Y}}^{(s,i)}]\, Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[\mathbf{Y}^{(s-1,i)}]}{P_{\widehat{\boldsymbol{\beta}}}[\mathbf{Y}^{(s-1,i)}]\, Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[\widetilde{\mathbf{Y}}^{(s,i)}]} \\ \mathbf{Y}^{(s-1,i)} & \text{otherwise} \end{cases}$$

  update the hitting time estimates: for each $u \in \mathcal{V}$,

$$\widehat{h}^{\mathsf{I}}_u \leftarrow \eta\widehat{h}^{\mathsf{I}}_u + \frac{1-\eta}{L}\sum_{i=1}^{L} h^{\mathsf{I}}_u(\mathbf{Y}^{(s,i)}), \quad \widehat{h}^{\mathsf{R}}_u \leftarrow \eta\widehat{h}^{\mathsf{R}}_u + \frac{1-\eta}{L}\sum_{i=1}^{L} h^{\mathsf{R}}_u(\mathbf{Y}^{(s,i)})$$

end for

reconstruct the diffusion history $\widehat{\mathbf{Y}}$: for each $u \in \mathcal{V}$

$$\widehat{y}_{t,u} \leftarrow \begin{cases} \mathsf{S} & \text{for } 0 \le t < \text{round}(\widehat{h}^{\mathsf{I}}_u) \\ \mathsf{I} & \text{for } \text{round}(\widehat{h}^{\mathsf{I}}_u) \le t < \text{round}(\widehat{h}^{\mathsf{R}}_u) \\ \mathsf{R} & \text{for } \text{round}(\widehat{h}^{\mathsf{R}}_u) \le t \le T \end{cases}$$

return $\widehat{\mathbf{Y}}$

per-segment computation (initialization and mean-field updates), and Sec. 3.3 presents the optimization procedure and complexity.

3.1 Maximum Pseudolikelihood Estimation with Segment Factorization

We estimate the diffusion parameters $\boldsymbol{\beta} = (\beta^{\mathsf{I}}, \beta^{\mathsf{R}})$ from multiple observed snapshots. Let $\mathcal{T}_{\text{obs}} = \{t_1, \dots, t_K\}$ with $0 = t_0 < t_1 < \cdots < t_K = T$, and let $\mathbf{y}_{\mathcal{T}_{\text{obs}}} := \{\mathbf{y}_t\}_{t\in\mathcal{T}_{\text{obs}}} = \{\mathbf{y}_{t_k}\}_{k=1}^{K}$ be the observed snapshots. Each snapshot $\mathbf{y}_t = (y_{t,u})_{u\in\mathcal{V}} \in \mathcal{X}^{\mathcal{V}}$ assigns every node's diffusion state at time t, where the state space is $\mathcal{X} := \{\mathsf{S}, \mathsf{I}, \mathsf{R}\}$. (For the SI model, we set $\beta_{\mathsf{R}} = 0$ and drop the recovered state R throughout; the estimator below remains unchanged in form.)

Under the SI/SIR Markov diffusion model, the *exact* likelihood of the observed snapshots requires marginalizing over exponentially many unobserved intermediate states and has been shown to be NP-hard [49]. To bypass the intractable likelihood, our key idea is to derive a tractable *mean-field* approximation [59] that we call *pseudolikelihood* and estimate diffusion parameters by maximizing the pseudolikelihood.

Using the Markov property of SI/SIR, we can factorize the probability of a complete history $\mathbf{Y}$ into K conditionally independent segments, where each segment $k = 1, \dots, K$ is from t_{k-1} to t_k:

$$P_{\boldsymbol{\beta}}[\mathbf{Y}] = P_{\boldsymbol{\beta}}[\mathbf{Y}_{[0,t_1]}] \prod_{k=2}^{K} P_{\boldsymbol{\beta}}[\mathbf{Y}_{(t_{k-1},t_k]} \mid \mathbf{y}_{t_{k-1}}]. \tag{7}$$

Hence, for each segment k, time $t \in [t_{k-1}, t_k]$, node $u \in \mathcal{V}$, and state $x \in \mathcal{X}$, we compute a mean-field pseudolikelihood $f^{x}_{k,t,u}(\boldsymbol{\beta})$ (to be introduced in Section 3.2) that approximates:

$$f^{x}_{k,t,u}(\boldsymbol{\beta}) \approx \begin{cases} P_{\boldsymbol{\beta}}[y_{t,u} = x], & \text{for } k = 1, \\ P_{\boldsymbol{\beta}}[y_{t,u} = x \mid \mathbf{y}_{t_{k-1}}], & \text{for } k = 2, \dots, K. \end{cases} \tag{8}$$

Finally, we estimate the diffusion parameters $\widehat{\boldsymbol{\beta}}$ by maximizing the following joint log-pseudolikelihood, which decomposes into the sum of log-pseudolikelihoods of each single node in each observed snapshot:

$$\widehat{\boldsymbol{\beta}} := \operatorname*{argmax}_{\boldsymbol{\beta}} \sum_{k=1}^{K} \sum_{u\in\mathcal{V}} \log f^{\,y_{t_k,u}}_{k,t_k,u}(\boldsymbol{\beta}). \tag{9}$$

We will use the estimated diffusion parameters $\widehat{\boldsymbol{\beta}}$ in our diffusion history reconstruction algorithm in Section 4.

3.2 Pseudolikelihood in Each Segment

In this subsection, we describe how to compute the marginal pseudolikelihoods $f^{x}_{k,t,u}(\boldsymbol{\beta})$ in each segment $k = 1, \dots, K$.

Left-end initialization. For the first segment $k = 1$, since the initial snapshot $\mathbf{y}_0$ is not observed, we initialize the pseudolikelihoods at $t_0 = 0$ using the (rough) number of initial infections n^{I}_0. Assuming that the set of n^{I}_0 initially infected nodes is uniformly drawn from all $\binom{n}{n^{\mathsf{I}}_0}$ possible sets, the probability that a node is initially infected is $\binom{n-1}{n^{\mathsf{I}}_0-1}/\binom{n}{n^{\mathsf{I}}_0} = n^{\mathsf{I}}_0/n$. Hence, for each node $u \in \mathcal{V}$, we let

$$f^{\mathsf{S}}_{1,0,u}(\boldsymbol{\beta}) := 1 - \frac{n^{\mathsf{I}}_0}{n}, \quad f^{\mathsf{I}}_{1,0,u}(\boldsymbol{\beta}) := \frac{n^{\mathsf{I}}_0}{n}, \quad f^{\mathsf{R}}_{1,0,u}(\boldsymbol{\beta}) := 0. \tag{10}$$

For later segments $k \ge 2$, since the left-end snapshot $\mathbf{y}_{t_{k-1}}$ is observed, then we initialize the pseudolikelihoods at t_{k-1} according to the observation:

$$f^{x}_{k,t_{k-1},u}(\boldsymbol{\beta}) := \mathbb{I}[y_{t_{k-1},u} = x], \qquad u \in \mathcal{V},\ x \in \mathcal{X}. \tag{11}$$

Mean-field updates. Starting from t_{k-1}, we propagate the pseudolikelihoods forward step by step until t_k. For $t = t_{k-1}+1, \dots, t_k$, the mean-field approximation $\mu_{k,t,u}(\boldsymbol{\beta})$ of the probability that a node $u \in \mathcal{V}$ is not infected by its infected neighbor is:

$$\mu_{k,t,u}(\boldsymbol{\beta}) := \prod_{v\in\mathcal{N}_u} \left(1 - \beta^{\mathsf{I}} \cdot f^{\mathsf{I}}_{k,t-1,v}(\boldsymbol{\beta})\right). \tag{12}$$

Then, a node is susceptible at time t iff it is susceptible at time $t-1$ and is not infected by its infected neighbors:

$$f^{\mathsf{S}}_{k,t,u}(\boldsymbol{\beta}) := f^{\mathsf{S}}_{k,t-1,u}(\boldsymbol{\beta}) \cdot \mu_{k,t,u}(\boldsymbol{\beta}); \tag{13}$$

a node is infected at time t if it is infected at time $t-1$, or if it is susceptible at time $t-1$ and gets infected, and if it has not recovered:

$$f^{\mathsf{I}}_{k,t,u}(\boldsymbol{\beta}) := \left(f^{\mathsf{I}}_{k,t-1,u}(\boldsymbol{\beta}) + f^{\mathsf{S}}_{k,t-1,u}(\boldsymbol{\beta}) \cdot (1 - \mu_{k,t,u}(\boldsymbol{\beta}))\right) \cdot (1 - \beta^{\mathsf{R}}); \tag{14}$$

and a node is recovered at time t in all other cases:

$$f^{\mathsf{R}}_{k,t,u}(\boldsymbol{\beta}) := 1 - f^{\mathsf{S}}_{k,t,u}(\boldsymbol{\beta}) - f^{\mathsf{I}}_{k,t,u}(\boldsymbol{\beta}). \tag{15}$$

Finally, at the right-end time t_k, we plug the pseudolikelihoods into Eq. (9) to estimate $\widehat{\boldsymbol{\beta}}$. Since our pseudolikelihoods are differentiable, we use gradient-based optimization to find the optimal $\widehat{\boldsymbol{\beta}}$.

3.3 Complexity Analysis

Proposition 3.1 (time complexity). *The time complexity of computing the joint pseudolikelihood Eq.* (9) *is* $O(T(n+m))$.

Proposition 3.1 shows that our diffusion parameter estimation is computationally efficient. The time complexity scales with the history size $O(Tn)$ and the graph size $O(n+m)$.

4 MCMC-Based History Reconstruction

Likelihood-based history reconstruction is both computationally intractable and brittle to parameter error, so we instead aim for a robust posterior summary of feasible histories. Sec. 4.1 represents each history by node-wise infection/recovery hitting times, defines a barycenter objective in this coordinate system, and reduces reconstruction to estimating posterior expected hitting times via Metropolis–Hastings MCMC. Sec. 4.2 then develops a learned proposal whose support exactly matches the multi-snapshot feasible set. Finally, Sec. 4.3 concludes the resulting computational complexity.

4.1 Reduction to MCMC-Based Hitting Time Estimation

Reduction to estimating expected hitting times. To compactly represent a complete history, we introduce *hitting times*: for each node $u \in \mathcal{V}$, let

$$h_u^{\mathtt{I}}(\mathbf{Y}) := \min\{T+1,\ \min\{t \geq 0 :\ y_{t,u} \in \{\mathtt{I},\mathtt{R}\}\}\}, \tag{16}$$

$$h_u^{\mathtt{R}}(\mathbf{Y}) := \min\{T+1,\ \min\{t \geq 0 :\ y_{t,u} = \mathtt{R}\}\}. \tag{17}$$

Here, $h_u^x(\mathbf{Y}) = T+1$ ($x \in \{\mathtt{I},\mathtt{R}\}$) represents that node u is never infected/recovered within timespan T. Then, we define a distance D between two histories $(\widehat{\mathbf{Y}}, \mathbf{Y})$ using hitting times as coordinates:

$$D(\widehat{\mathbf{Y}}, \mathbf{Y}) := \sqrt{\sum_{u\in\mathcal{V}} ((h_u^{\mathtt{I}}(\widehat{\mathbf{Y}}) - h_u^{\mathtt{I}}(\mathbf{Y}))^2 + (h_u^{\mathtt{R}}(\widehat{\mathbf{Y}}) - h_u^{\mathtt{R}}(\mathbf{Y}))^2)}. \tag{18}$$

Based on this distance, we propose to reconstruct a history $\widehat{\mathbf{Y}}$ that is closest to all possible histories w.r.t. distance D under posterior distribution $P_{\widehat{\boldsymbol{\beta}}} \mid \mathbf{Y}_{\mathcal{T}_{\text{obs}}}$:

$$\min_{\widehat{\mathbf{Y}}} \mathop{\mathbb{E}}_{\mathbf{Y}\sim P_{\widehat{\boldsymbol{\beta}}}\mid \mathbf{Y}_{\mathcal{T}_{\text{obs}}}} [D(\widehat{\mathbf{Y}}, \mathbf{Y})^2]. \tag{19}$$

In fact, Eq. (19) admits a simple closed-form solution. Using bias–variance decomposition, we can see that the optimal estimates are rounding each expected hitting time to the nearest integer:

$$h_u^x(\widehat{\mathbf{Y}}) := \text{round}\left(\mathop{\mathbb{E}}_{\mathbf{Y}\sim P_{\widehat{\boldsymbol{\beta}}}\mid y_{\mathcal{T}_{\text{obs}}}} [h_u^x(\mathbf{Y})]\right), \quad x \in \{\mathtt{I},\mathtt{R}\}. \tag{20}$$

Consequently, our formulation Eq. (19) avoids the *intractable* posterior likelihood computation [49] and reduces the complex combinatorial optimization problem Eq. (19) to a *tractable* statistical estimation problem Eq. (20). Finally, given estimated hitting times $h_u^x(\widehat{\mathbf{Y}})$, we reconstruct the diffusion history $\widehat{\mathbf{Y}}$ as follows: for each node $u \in \mathcal{V}$ and time $0 \leq t \leq T$,

$$\widehat{y}_{t,u} := \begin{cases} \mathtt{S}, & 0 \leq t < h_u^{\mathtt{I}}(\widehat{\mathbf{Y}}), \\ \mathtt{I}, & h_u^{\mathtt{I}}(\widehat{\mathbf{Y}}) \leq t < h_u^{\mathtt{R}}(\widehat{\mathbf{Y}}), \\ \mathtt{R}, & h_u^{\mathtt{R}}(\widehat{\mathbf{Y}}) \leq t \leq T. \end{cases} \tag{21}$$

Estimating expected hitting times via MCMC. Nevertheless, we still cannot straightforwardly compute $\mathbb{E}_{\mathbf{Y}\sim P_{\widehat{\boldsymbol{\beta}}}\mid y_{\mathcal{T}_{\text{obs}}}}[h_u^x(\mathbf{Y})]$ because sampling directly from the posterior $P_{\widehat{\boldsymbol{\beta}}} \mid \mathbf{y}_{\mathcal{T}_{\text{obs}}}$ is still intractable.

To address the intractability, our key idea is that if we can design a parametric proposal distribution $Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[\cdot]$ (where $\boldsymbol{\theta}$ denotes its parameters) that (i) is tractable to sample from and (ii) satisfies

$$\text{supp}(Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})) = \text{supp}(P \mid \mathbf{Y}_{\mathcal{T}_{\text{obs}}}), \tag{22}$$

then we can employ the Metropolis–Hastings Markov chain Monte Carlo (M–H MCMC) method [25, 44] to estimate the expected hitting times, which constructs a Markov chain whose stationary distribution is the desired posterior $P_{\widehat{\boldsymbol{\beta}}} \mid \mathbf{y}_{\mathcal{T}_{\text{obs}}}$. Specifically, M–H uses $Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[\cdot]$ as the so-called *proposal* distribution and maintains a history $\mathbf{Y}$; in each step, M–H MCMC proposes a new history $\widetilde{\mathbf{Y}} \sim Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})$ and replaces $\mathbf{Y}$ with $\widetilde{\mathbf{Y}}$ with probability

$$\min\left\{1,\ \frac{P_{\widehat{\boldsymbol{\beta}}}[\widetilde{\mathbf{Y}}]\, Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[\mathbf{Y}]}{P_{\widehat{\boldsymbol{\beta}}}[\mathbf{Y}]\, Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[\widetilde{\mathbf{Y}}]}\right\}. \tag{23}$$

This defines a Markov chain of histories, which converges to the desired posterior $P_{\widehat{\boldsymbol{\beta}}} \mid \mathbf{y}_{\mathcal{T}_{\text{obs}}}$ after sufficiently many steps. Hence, we can sample in parallel multiple histories using M–H MCMC and use their average to estimate the expected hitting times. In the next Sec. 4.2, we will design a proposal distribution $Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[\cdot]$ that satisfies Eq. (22).

4.2 Feasibility-Aware MCMC Proposal

In this subsection, we design an MCMC proposal $Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[\cdot]$ such that $\text{supp}(Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})) = \text{supp}(P \mid \mathbf{Y}_{\mathcal{T}_{\text{obs}}})$. With a little abuse of notation, let $\mathbf{Y}$ denote the proposed history. Inspired by the Markov property of graph diffusion, we define $Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[\cdot]$ by factorizing the timespan T according to the observed segments: $Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[\mathbf{Y}] :=$

$$Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[\mathbf{Y}_{[0,t_1]} \mid \mathbf{y}_{t_1}] \prod_{k=2}^{K} Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[\mathbf{Y}_{(t_{k-1},t_k]} \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}_{t_k}]. \tag{24}$$

That is, we define $Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[\cdot]$ over each segment k separately.

Feasibility-aware proposal. We need to design $Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[\cdot]$ to ensure feasibility over every segment. We first present an algorithm for $k = 2, \ldots, K$ and then adapt it to $k = 1$. Our following Lemma 4.1 characterizes a sufficient and necessary condition of feasibility.

Lemma 4.1 (feasibility characterization). *For $k = 2, \ldots, K$ and $t_{k-1} < t < t_k$, given complete snapshots $\mathbf{y}_{t_{k-1}}, \mathbf{y}_{t+1} \in \mathcal{X}^{\mathcal{V}}$ with $P_{\boldsymbol{\beta}}[\mathbf{y}_{t+1} \mid \mathbf{y}_{t_{k-1}}] > 0$ and incomplete snapshot $\mathbf{y}_t \in (\mathcal{X} \cup \{?\})^{\mathcal{V}}$ with $\mathbf{y}_{t_{k-1}} \leq \mathbf{y}_t \leq \mathbf{y}_{t+1}$ (where ? means "undefined"; the $\leq$ comparison ignores undefined nodes), let*

$$\mathcal{V}_t := \{u \in \mathcal{V} : y_{t_{k-1},u} \neq \mathtt{R},\ y_{t,u} \neq \mathtt{S},\ y_{t+1,u} \neq \mathtt{S}\}, \tag{25}$$

with the convention that $? \neq \mathtt{S}$; and let

$$\mathcal{W}_t := \{u \in \mathcal{V}_t : \exists v \in \mathcal{V}^{\mathtt{I}}(\mathbf{y}_{t_{k-1}}) \text{ s.t. } d_{\mathcal{V}_t}(v,u) \leq t - t_{k-1}\}, \tag{26}$$

where $d_{\mathcal{V}_t} : \mathcal{V} \times \mathcal{V} \to \mathbb{N}_{\geq 0}$ denotes the shortest-path distance when only visiting nodes in $\mathcal{V}_t$. Then, $P_{\boldsymbol{\beta}}[\mathbf{y}_{t,\neq ?} \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}_{t+1}] > 0$ (where $\mathbf{y}_{t,\neq ?}$ excludes undefined *nodes in $\mathbf{y}_t$) if and only if for every node*

$v \in \mathcal{V}^{\mathrm{IR}}(\mathbf{y}_{t+1}) \setminus \mathcal{V}^{\mathsf{R}}(\mathbf{y}_{t_{k-1}})$, *we have*

$$v \in \begin{cases} \bigcup_{u \in \mathcal{W}_t \setminus \mathcal{V}^{\mathsf{R}}(\mathbf{y}_t)} \mathcal{N}_u, & \text{if } y_{t,v} = \mathsf{S}, \\ \mathcal{W}_t, & \text{if } y_{t,v} = \mathtt{I} \text{ or } \mathsf{R}, \\ \mathcal{W}_t \cup \bigcup_{u \in \mathcal{W}_t \setminus \mathcal{V}^{\mathsf{R}}(\mathbf{y}_t)} \mathcal{N}_u, & \text{if } y_{t,v} = ?. \end{cases} \tag{27}$$

Therefore, it suffices to ensure $\mathbf{y}_{t_{k-1}} \le \mathbf{y}_t \le \mathbf{y}_{t+1}$ and Eq. (27), which can be efficiently checked via breath-first search.

Our proposal $Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\mathrm{obs}}})[\mathbf{Y}_{(t_{k-1},t_k]} \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}_{t_k}]$ is designed as follows. We generate the snapshots inductively in the reverse temporal order of $t = t_k - 1, \ldots, t_{k-1} + 1$. Suppose that we have a complete $\mathbf{y}_{t+1}$ and want to generate $\mathbf{y}_t$. Initially, we let $\mathbf{y}_t = (?)_{u \in \mathcal{V}}$ (i.e., undefined). We process the undefined nodes one by one. For each undefined node u, let $\mathcal{X}_{t,u} \in \{\{\mathsf{S}\}, \{\mathsf{S}, \mathtt{I}\}, \{\mathsf{S}, \mathtt{I}, \mathsf{R}\}, \{\mathtt{I}\}, \{\mathtt{I}, \mathsf{R}\}, \{\mathsf{R}\}\}$ denote the set of states such that setting $y_{t,u}$ to the state satisfies both $\mathbf{y}_{t_{k-1}} \le \mathbf{y}_t \le \mathbf{y}_{t+1}$ and Eq. (27). Our parametric proposal distribution predicts two probabilities $0 < q^{\mathtt{I}}_{t,u}, q^{\mathsf{S}}_{t,u} < 1$, and we then define two *feasibility-aware* probabilities:

$$p^{\mathtt{I}}_{t,u} := \begin{cases} 0, & \text{if } \mathcal{X}_{t,u} = \{\mathsf{R}\}, \\ 1, & \text{if } \mathsf{R} \notin \mathcal{X}_{t,u}, \\ q^{\mathtt{I}}_{t,u}, & \text{otherwise;} \end{cases} \tag{28}$$

$$p^{\mathsf{S}}_{t,u} := \begin{cases} 0, & \text{if } \mathsf{S} \notin \mathcal{X}_{t,u}, \\ 1, & \text{if } \mathcal{X}_{t,u} = \{\mathsf{S}\}, \\ q^{\mathsf{S}}_{t,u}, & \text{otherwise.} \end{cases} \tag{29}$$

Finally, we decide $y_{t,u}$ using $(p^{\mathtt{I}}_{t,u}, p^{\mathsf{S}}_{t,u})$:

$$y_{t,u} := \begin{cases} \mathsf{R}, & \text{with probability } 1 - p^{\mathtt{I}}_{t,u}, \\ \mathtt{I}, & \text{with probability } p^{\mathtt{I}}_{t,u}(1 - p^{\mathsf{S}}_{t,u}), \\ \mathsf{S}, & \text{with probability } p^{\mathtt{I}}_{t,u} p^{\mathsf{S}}_{t,u}. \end{cases} \tag{30}$$

For segment $k = 1$, we can use $\mathcal{W}_t := \mathcal{V}_t := \{u \in \mathcal{V} : y_{t_{k-1},u} \neq \mathsf{R},\ y_{t,u} \neq \mathsf{S},\ y_{t+1,u} \neq \mathsf{S}\}$ to similarly define $Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\mathrm{obs}}})[\mathbf{Y}_{[0,t_1]} \mid \mathbf{y}_{t_1}]$.

Theoretical guarantee of feasibility. We have the following Theorem 4.2 showing that our proposal distribution $Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\mathrm{obs}}})[\cdot]$ is guaranteed to be compatible with the observed snapshots $\mathbf{Y}_{\mathcal{T}_{\mathrm{obs}}}$.

Theorem 4.2 (feasibility). *Our MCMC proposal distribution $Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\mathrm{obs}}})[\cdot]$ described in Sec. 4.2 satisfies*

$$\mathrm{supp}(Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\mathrm{obs}}})) = \mathrm{supp}(P \mid \mathbf{Y}_{\mathcal{T}_{\mathrm{obs}}}). \tag{31}$$

4.3 GNN-Based Proposal

To further boost the reconstruction accuracy, we parameterize the proposal distribution $Q_{\boldsymbol{\theta}}(\cdot)$ via an Anisotropic GNN with edge gating [6, 31, 48]. Let $\mathbf{g}^{\ell,1}_u$ and $\mathbf{g}^{\ell,2}_{u,v}$ denote the node and edge embeddings at layer ℓ associated with node u and edge (u, v), respectively. The node inputs $\mathbf{g}^{0,1}_u$ are initialized by feeding $\mathbf{y}_{\mathcal{T}_{\mathrm{obs}},u}$ into a linear layer. The edge inputs $\mathbf{g}^{0,2}_{u,v}$ are learnable parameters. The embeddings at the next layer is propagated with an anisotropic message passing scheme with SiLU activation [20]:

$$\begin{aligned} \mathbf{g}^{\ell+1,1}_u &:= \mathbf{g}^{\ell,1}_u + \mathrm{SiLU}(\mathrm{BN}(\mathbf{W}^{\ell,1}\mathbf{g}^{\ell,1}_u + \underset{v \in \mathcal{N}_u}{\mathrm{Mean}}(\sigma(\mathbf{g}^{\ell,2}_{u,v}) \odot (\mathbf{W}^{\ell,2}\mathbf{g}^{\ell,1}_v)))), \\ \mathbf{g}^{\ell+1,2}_{u,v} &:= \mathbf{g}^{\ell,2}_{u,v} + \mathrm{SiLU}(\mathrm{BN}(\mathbf{W}^{\ell,3}\mathbf{g}^{\ell,2}_{u,v} + \mathbf{W}^{\ell,4}\mathbf{g}^{\ell,1}_u + \mathbf{W}^{\ell,5}\mathbf{g}^{\ell,1}_v)). \end{aligned} \tag{32}$$

where $\mathbf{W}^{\ell,1}, \ldots, \mathbf{W}^{\ell,5}$ are learnable parameters of layer ℓ, BN is batch normalization [29], Mean is mean aggregation, σ is the sigmoid function, and $\odot$ is entry-wise multiplication. We append a Multi-Layer Perceptron (MLP) after the GNN to predict $(q^{\mathtt{I}}_{t,u}, q^{\mathsf{S}}_{t,u})$ for all nodes $u \in \mathcal{V}$ at all times $0 \le t < T$. Since no training data is available, we train our proposal by minimizing the following cross entropy over randomly simulated histories $\mathbf{Y}'$ under estimated diffusion parameters $\widehat{\boldsymbol{\beta}}$:

$$\min_{\boldsymbol{\theta}} \mathop{\mathbb{E}}_{\mathbf{Y}' \sim P_{\widehat{\boldsymbol{\beta}}}} [-\log Q_{\boldsymbol{\theta}}(\mathbf{Y}'_{\mathcal{T}_{\mathrm{obs}}})[\mathbf{Y}']]. \tag{33}$$

4.4 Complexity Analysis

Proposition 4.3 (sampling cost). *Sampling a complete history from our proposal $Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{obs}})[\cdot]$ takes $O(Tn(n+m))$ time.*

Proposition 4.3 shows that generating candidate histories with M–H MCMC using our proposal runs in polynomial time.

5 Experiments

We conduct extensive experiments on both synthetic and real-world datasets to evaluate HERMES for multi-snapshot diffusion history reconstruction, and answer the following research questions:

- **RQ1** (effectiveness): How accurate does our HERMES reconstruct diffusion histories?
- **RQ2** (timespan): How robust is our HERMES against increase in timespan T?
- **RQ3** (snapshot count): How does the reconstruction quality vary with the number K of observed snapshots?
- **RQ4** (diffusion parameters): How accurate are estimated diffusion parameters $\widehat{\boldsymbol{\beta}}$?
- **RQ5** (ablation): How much does the learned proposal $Q_{\boldsymbol{\theta}}(\cdot)$ in M–H MCMC improve reconstruction quality?

5.1 Experimental Settings

Due to the page limit, please refer to Appendix A for additional experimental settings.

Datasets. We evaluate on 12 datasets that cover both synthetic and real graphs, as well as synthetic and real diffusion. Following a common taxonomy: (D1) synthetic diffusion on synthetic graphs: Barabási–Albert (BA) [3] and Erdős–Rényi (ER) [21]; (D2) synthetic diffusion on real graphs: Oregon2 [38] and Prost [38]; and (D3) real diffusion on real graphs: BrFarmers [52, 58], Pol [15], Covid [49], and Hebrew [4]. Dataset statistics are summarized in Table 3. For D1 and D2, we consider both SI and SIR dynamics, yielding 8 datasets; together with the 4 datasets in D3, this totals 12 datasets.

Baselines. We compare HERMES with two groups of baselines: (B1) supervised methods: GNNs (GCN [34], GIN [62]) and time series imputation methods (BRITS [7], GRIN [14], SPIN [43]); (B2) statistical methods for SI/SIR: DHREC [54], CRI [11], and DITTO [49]. When diffusion parameters are required, baselines use the same estimated parameters as HERMES. DITTO is evaluated in its native single-snapshot setting, it uses only the final snapshot $\mathbf{y}_T$ and discards the intermediate snapshot $\mathbf{y}_{t_{\mathrm{obs}}}$. Comparison with DITTO quantifies what can be achieved when the intermediate observation is unavailable and directly justifies the value of multiple snapshots; a comparison with two-snapshot segment-and-stitch version of DITTO is reported in Table 4 in Section 5.7.

Table 2: Comparison of reconstruction quality. Our proposed HERMES consistently achieves the best average ranks on both SI and SIR datasets. "OOM" indicates "out of memory."

Type	Method	BA-SI		ER-SI		Oregon2-SI		Prost-SI		BrFarmers (SI)		Pol (SI)		Avg.
		F1↑	NRMSE↓	F1↑	NRMSE↓	F1↑	NRMSE↓	F1↑	NRMSE↓	F1↑	NRMSE↓	F1↑	NRMSE↓	Rank↓
Supervised	GCN	0.8602	0.2085	0.8469	0.2131	0.7811	0.3172	0.8134	0.2626	0.8512	0.1709	0.6847	0.3572	5.42
	GIN	0.8422	0.2188	0.8416	0.2104	0.7634	0.3106	0.8451	0.2226	0.7659	0.2652	0.7648	0.2878	5.17
	BRITS	0.5932	0.2237	0.6135	0.2233	0.5579	0.4818	0.5616	0.5197	0.5617	0.2181	0.5127	0.4707	8.58
	GRIN	0.8939	0.1541	0.8993	0.1323	0.8784	0.1578	0.6732	0.4147	0.9043	**0.1268**	0.8477	**0.1752**	3.08
	SPIN	0.8917	0.1377	0.8996	0.1291	0.8941	**0.1328**	0.7157	0.3751	0.8967	0.1315	OOM		4.08
Statistical	DHREC	0.7406	0.2353	0.7537	0.2245	0.5686	0.2612	0.8204	0.2299	0.6366	0.2766	0.8176	0.2096	6.50
	CRI	0.7686	0.2622	0.8019	0.2256	0.7953	0.2409	0.8523	0.1984	0.7523	0.2681	0.8465	0.1958	5.33
	DITTO	0.8384	0.2139	0.8269	0.2225	0.8280	0.2289	0.8327	0.2317	0.8206	0.2142	0.7471	0.2903	5.17
	HERMES (ours)	**0.9012**	**0.1284**	**0.9011**	**0.1269**	**0.8964**	0.1380	**0.9056**	**0.1308**	**0.8980**	0.1491	**0.8487**	0.1823	**1.75**
Type	**Method**	BA-SIR		ER-SIR		Oregon2-SIR		Prost-SIR		Covid (SIR)		Hebrew (SIR)		Avg.
		F1↑	NRMSE↓	F1↑	NRMSE↓	F1↑	NRMSE↓	F1↑	NRMSE↓	F1↑	NRMSE↓	F1↑	NRMSE↓	Rank↓
Supervised	GCN	0.6890	0.1905	0.6754	0.1769	0.5994	0.2466	0.6091	0.2180	0.6008	0.3226	0.5805	0.2263	5.58
	GIN	0.6435	0.2134	0.6428	0.2046	0.5443	0.2844	0.5974	0.2378	0.5660	0.2831	0.5960	0.2909	6.67
	BRITS	0.5660	0.2191	0.5817	0.2118	0.3483	0.6243	0.3528	0.6156	0.4751	0.3485	0.4297	0.5518	8.50
	GRIN	0.8633	**0.0981**	0.8631	**0.0962**	0.8706	**0.1034**	0.6079	0.2785	0.7635	0.1647	0.5582	0.1653	3.25
	SPIN	0.7458	0.1137	0.7400	0.1471	0.6857	0.1225	0.5796	0.2619	0.7656	0.2480	0.5582	0.1653	4.50
Statistical	DHREC	0.7493	0.2511	0.7610	0.2349	0.7631	0.2695	0.7822	0.2653	0.6794	0.4226	0.7718	0.1501	5.17
	CRI	0.6417	0.2537	0.6594	0.2297	0.5993	0.2834	0.6240	0.2635	0.5171	0.4991	0.7258	0.1466	6.08
	DITTO	0.7783	0.1633	0.7734	0.1679	0.7928	0.1707	0.7929	**0.1690**	0.6240	0.2637	0.6411	0.2983	3.75
	HERMES (ours)	**0.8640**	0.1509	**0.8657**	0.1307	**0.8719**	0.1874	**0.8609**	0.1737	**0.7819**	**0.1589**	**0.8356**	**0.1223**	**1.50**

Table 3: Summary of datasets.

Dataset	#Nodes	#Edges	Timespan	Graph	Diffusion
BA	1,000	3,984	10	Synthetic	Synthetic
ER	1,000	3,987	10	Synthetic	Synthetic
Oregon2	11,461	32,730	15	Real	Synthetic
Prost	15,810	38,540	15	Real	Synthetic
BrFarmers	82	230	16	Real	Real SI
Pol	18,470	48,053	40	Real	Real SI
Covid	344	2,044	10	Real	Real SIR
Hebrew	3,521	18,064	9	Real	Real SIR

Evaluation metrics. We report macro-F1 (F1) of $\widehat{\mathbf{Y}}_{[0,T)}$ and normalized RMSE (NRMSE) of hitting times:

$$\mathrm{NRMSE}(\mathbf{Y}, \widehat{\mathbf{Y}}) := \sqrt{\frac{\sum_{u \in \mathcal{V}} ((h_u^{\mathrm{I}}(\mathbf{Y}) - h_u^{\mathrm{I}}(\widehat{\mathbf{Y}}))^2 + (h_u^{\mathrm{R}}(\mathbf{Y}) - h_u^{\mathrm{R}}(\widehat{\mathbf{Y}}))^2)}{2|\mathcal{V}|(T+1)^2}}. \quad (34)$$

Observed snapshots. Although HERMES is designed for a general multi-snapshot setting, in the main experiments we adopt a consistent two-snapshot setup ($|\mathcal{T}_{\mathrm{obs}}| = 2$) across all datasets: (i) we always observe the final snapshot at time T, and (ii) we additionally observe one intermediate snapshot at time $t_{\mathrm{obs}} = \lfloor T/2 \rfloor$. That is, $\mathcal{T}_{\mathrm{obs}} := \{t_{\mathrm{obs}}, T\}$.

Implementation details. Since our original algorithm in Sec. 4.2 is less efficient on GPUs, we implement a modified version that runs more efficiently on GPUs but might not ensure feasibility in some edge cases. Our code is publicly available at https://github.com/Yijing-Zuo/KDD26-HERMES. We use the same random seed 123456789 for all methods. Unless otherwise stated, methods are executed on an Nvidia B200 GPU, except that CRI runs on CPU.

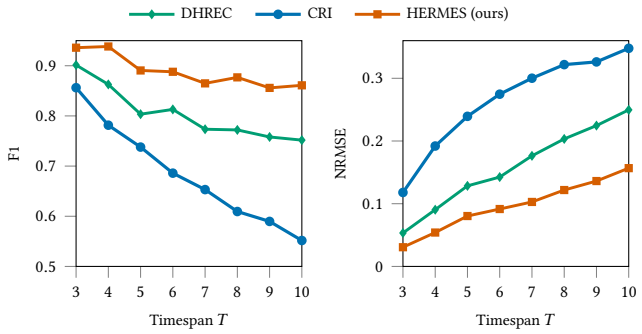
5.2 Reconstruction Quality

To evaluate how accurate our HERMES reconstructs complete diffusion histories and answer RQ1, we compare with baseline methods on the 12 datasets. The results are summarized in Table 2. Notably, HERMES achieves the best in F1 on 11 datasets. The gains are clearest on real diffusion. For instance, on Covid, HERMES attains 0.7819 F1 and 0.1589 NRMSE, which significantly improves over the best F1 0.7656 among supervised methods and reduces the best NRMSE 0.2637 among statistical methods by 39.7%; on Hebrew, our HERMES reaches 0.8356 F1 while supervised methods peak at 0.5960 F1 and statistical methods at 0.7718 F1. Besides that, the best supervised method SPIN is OOM on the larger Pol while our HERMES consistently consumes a practical amount of memory.

Additionally, to compare with a segment-and-stitch version of DITTO that uses all observed snapshots, our Table 4 in Section 5.7 shows that HERMES still achieves higher F1 than this version on all four representative synthetic datasets.

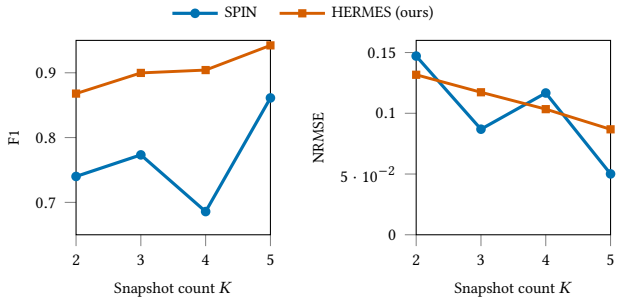
5.3 Robustness to Timespan Increase

As the timespan T increases, the inverse problem becomes more uncertain since a much larger set of diffusion histories can explain the same observed snapshots. To examine how robust our HERMES is against this effect and answer RQ2, we vary timespan T on BA-SIR from 3 to 10 and test our HERMES. Figure 2 compares HERMES with MLE-based baselines CRI and DHREC. While the performance of all methods degrades as T increases as expected, our HERMES remains consistently best and degrades more gracefully. For instance, under $T = 10$, HERMES achieves F1 0.8610 and NRMSE 0.1577 while DHREC / CRI decrease F1 to 0.7517 / 0.5517 and increase NRMSE to 0.2496 / 0.3478, respectively. Overall, across $3 \le T \le 10$, the F1 of our HERMES decreases by only 0.0750, which is significantly better than 0.1497 for DHREC and 0.3046 for CRI. These results



(a) F1 ↑. (b) NRMSE ↓.

Figure 2: Performance v.s. timespan T. Our HERMES is the most robust against timespan increase.



(a) F1 ↑. (b) NRMSE ↓.

Figure 3: Performance of our HERMES consistently improves as the number K of observed snapshots increases.

indicate that our HERMES better handles the increased uncertainty introduced by longer timespans.

5.4 Effect of the Number of Observed Snapshots

To answer RQ3, we evaluate the effect of the number K of observed snapshots on reconstruction quality by varying the snapshot count $K = 2, 3, 4, 5$ while fixing the timespan $T = 10$. We choose $\mathcal{T}_{\text{obs}}$ by uniformly partitioning the timespan ($K = 2$: $\{5, 10\}$; $K = 3$: $\{3, 6, 10\}$; $K = 4$: $\{2, 5, 7, 10\}$; $K = 5$: $\{2, 4, 6, 8, 10\}$). Figure 3(a)(b) report F1 and NRMSE of our HERMES and the strongest supervised baseline SPIN. The results show that our HERMES benefits consistently from additional observations. For instance, as K increases from 2 to 5, F1 improves from 0.8679 to 0.9423 and NRMSE decreases from 0.1317 to 0.0868. In contrast, SPIN exhibits a less stable trend.

5.5 Accuracy of Diffusion Parameter Estimation

To answer RQ4, we evaluate our diffusion parameter estimator on synthetic diffusion datasets, where the true diffusion parameters are known by construction. Figure 4 reports a histogram of the estimation errors $\widehat{\beta}_\cdot - \beta_\cdot$ of diffusion parameters from our method. We see that all estimation errors are very close to zero, significantly

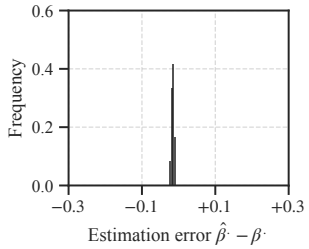
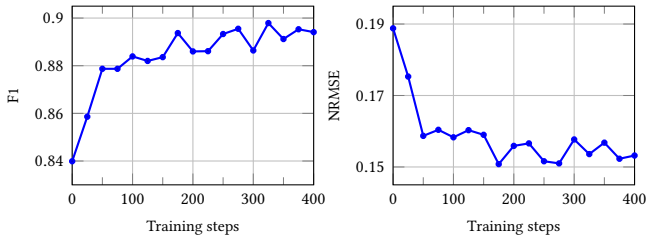


Figure 4: Histogram of estimation errors of β. All estimation errors of our HERMES are very close to zero.



(a) F1 ↑. (b) NRMSE ↓.

Figure 5: Ablation study on the learned proposal. Our proposal training significantly improves reconstruction quality.

smaller than 0.1. The results suggest that our HERMES yields accurate estimates of diffusion parameters, which supports the use of $\widehat{\boldsymbol{\beta}}$ in our history reconstruction pipeline.

5.6 Ablation Study on Proposal Training

To verify the efficacy of our proposal training (Sec. 4.3) and answer RQ5, we vary the number of proposal training steps from 0 to 400 on BrFarmers and report F1 and NRMSE in Figure 5. We see that the performance of our HERMES improves substantially as training progresses. In particular, F1 increases from 0.8399 to 0.8979 and NRMSE drops from 0.1888 to 0.1510. Furthermore, the convergence is rapid: both metrics start to plateau after only 150 training steps.

5.7 Comparison with Segment-and-Stitch DITTO

To assess whether the two-snapshot setting can be reduced to repeated single-snapshot reconstruction, we evaluate a direct segment-and-stitch adaptation of DITTO [49] (which we call DITTO-seg). It runs DITTO on $[0, t_{\text{obs}}]$ using $\mathbf{y}_{t_{\text{obs}}}$ as the observed snapshot, runs DITTO again on $(t_{\text{obs}}, T]$ using $\mathbf{y}_T$ as the observed snapshot, and stitches the two segments as the reconstructed history. Table 4 reports the results on four representative synthetic datasets. DITTO-seg improves over the original single-snapshot DITTO baseline, while HERMES achieves the highest F1 on all four datasets. This suggests that stitching single-snapshot reconstructions is less effective than our joint multi-snapshot modeling in HERMES.

5.8 Analysis of MCMC Iterations

The default HERMES configuration uses $S = 100$ MCMC iterations. To test sensitivity to this choice, we fix the estimated diffusion

Table 4: Comparison with DITTO and segment-and-stitch DITTO (DITTO-seg).

Method	BA-SI		ER-SI		BA-SIR		ER-SIR	
	F1	NRMSE	F1	NRMSE	F1	NRMSE	F1	NRMSE
DITTO	0.8384	0.2139	0.8269	0.2225	0.7783	0.1633	0.7734	0.1679
DITTO-seg	0.8810	0.1565	0.8631	0.1483	0.8121	**0.1116**	0.8115	**0.1080**
HERMES (ours)	**0.9012**	**0.1284**	**0.9011**	**0.1269**	**0.8640**	0.1509	**0.8657**	0.1307

Table 5: Sensitivity analysis w.r.t. the number of MCMC iterations S (mean ± standard deviation over four MCMC seeds).

Dataset	Metric	$S = 25$	$S = 50$	$S = 100$	$S = 200$
BA-SIR	F1	0.8626 ± 0.0006	0.8625 ± 0.0009	0.8634 ± 0.0010	0.8632 ± 0.0007
	NRMSE	0.1504 ± 0.0028	0.1494 ± 0.0030	0.1503 ± 0.0021	0.1510 ± 0.0029
ER-SIR	F1	0.8682 ± 0.0009	0.8679 ± 0.0008	0.8685 ± 0.0010	0.8686 ± 0.0008
	NRMSE	0.1296 ± 0.0014	0.1293 ± 0.0014	0.1303 ± 0.0027	0.1307 ± 0.0023

Table 6: Sensitivity analysis w.r.t. initial infection count n_0^{I}.

Dataset	F1 Range	NRMSE Range	Max $\lvert\Delta$F1$\rvert$
BA-SI	0.8994–0.9032	0.1285–0.1304	0.0035
ER-SI	0.8970–0.9025	0.1267–0.1309	0.0054
BA-SIR	0.8616–0.8653	0.1433–0.1558	0.0024
ER-SIR	0.8659–0.8699	0.1240–0.1336	0.0020

parameters and the trained proposal model, and vary the number of MCMC iterations over $S \in \{25, 50, 100, 200\}$. Each setting is repeated with four MCMC seeds. Table 5 shows that the final reconstruction quality is stable over this 8× range of S on representative SIR datasets.

5.9 Robustness to Initial Infection Prior

HERMES uses the rough initial infected count n_0^{I} as a soft prior rather than as a hard constraint. To evaluate sensitivity to this prior, we vary the assumed initial infected count from 0.5× to 1.5× of the true count on four representative synthetic datasets. The results in Table 6 show that reconstruction quality changes only mildly across this range. Relative to the 1.0× setting, the maximum absolute change is 0.0054 in F1 and 0.0097 in NRMSE across all tested settings. As a remark, the estimated diffusion parameters also move in the expected compensatory direction. For example, on BA-SI, the estimated $\widehat{\beta^{\mathrm{I}}}$ decreases from 0.0972 to 0.0715 as the assumed initial infected count increases from 0.5× to 1.5×. On BA-SIR, the estimated $(\widehat{\beta^{\mathrm{I}}}, \widehat{\beta^{\mathrm{R}}})$ changes from $(0.1000, 0.0821)$ to $(0.0751, 0.0774)$.

6 Related Work

Graph diffusion works broadly split into forward problems and inverse problems [17]. Forward studies cover epidemic thresholds, diffusion optimization (e.g., influence maximization, immunization), and diffusion operators for graph learning [8, 16, 23, 32, 47]. With observed traces, inverse methods estimate diffusion parameters and infer latent diffusion networks from cascades, including settings that reduce reliance on precise timestamps [24, 28, 51, 60]. When histories are missing, work centers on source localization and history reconstruction. Source localization infers initial infected nodes from a snapshot using structural/sample-path ideas or learned inference [9, 27, 39–41, 55, 67]. History reconstruction targets the full latent trajectory and is largely likelihood-based (often assuming known parameters/timestamps), with single-snapshot variants and time-series imputation baselines also studied [1, 7, 10, 14, 18, 30, 43, 46, 46, 49, 54, 63, 64]. Such imputers are typically supervised and do not enforce diffusion-feasibility constraints [12, 35, 42, 56, 66]. Accordingly, we compare with supervised GNN predictors and likelihood-based DHREC/CRI, which highlight parameter dependence and distribution mismatch under multi-snapshot constraints [9, 11, 13, 17, 27, 34, 36, 42, 54, 62]. The multi-snapshot setting is more challenging because multiple fixed snapshots impose coupled segment-wise constraints rather than a single endpoint constraint. DITTO [49] supports only a single observed snapshot and shares with HERMES the high-level mean-field, hitting-time, and M–H MCMC perspective. However, extending this perspective to multiple snapshots requires new machinery: a segmented mean-field pseudolikelihood over observed intervals, a segment-wise proposal enforcing right-end feasibility and left-end extendability through a dynamic reachability mask, and the support-matching guarantee in Theorem 4.1. Thus, HERMES is a dedicated multi-snapshot estimator and support-correct proposal framework.

7 Conclusion & Discussion

In this work, we study diffusion history reconstruction from multiple observed snapshots on graphs. We estimate unknown diffusion parameters with a mean-field pseudo-likelihood that aggregates evidence from all observation times. We reduce history reconstruction to estimating posterior expected hitting times via M–H MCMC and design a nontrivial proposal distribution for M–H MCMC such that proposed histories are theoretically guaranteed to be feasible and compatible with observed snapshots. Moreover, we parameterize the proposal through a GNN to further enhance performance of M–H MCMC. Extensive experiments further demonstrate the superior performance of HERMES.

This work also has a few limitations to be addressed by future work. First, we study multi-snapshot history reconstruction for SI/SIR dynamics only (with unknown parameters) and assume that the diffusion model is given by domain knowledge. If the diffusion model is not given, a statistical test [2] can be performed to determine the diffusion model. Extending HERMES to a new diffusion model such as SEIR [26] would require adjustments to our algorithm. Besides that, while our complexity analysis shows that our HERMES runs in polynomial time, it is currently not optimized for real-time processing of million-scale graphs. Further engineering efforts are needed to further improve scalability.

Acknowledgments

This work is supported by NSF (2324770). The content of the information in this document does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

References

[1] Nourhan Ahmed, Vijaya Krishna Yalavarthi, and Lars Schmidt-Thieme. 2025. Motif-aware Graph Neural Networks for Networked Time Series Imputation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 11409–11417.

[2] Hirotugu Akaike. 1974. A new look at the statistical model identification. *IEEE transactions on automatic control* 19, 6 (1974), 716–723.

[3] Albert-László Barabási and Réka Albert. 1999. Emergence of scaling in random networks. *science* 286, 5439 (1999), 509–512.

[4] Alon Bartal, Nava Pliskin, and Oren Tsur. 2020. Local/Global contagion of viral/non-viral information: Analysis of contagion spread in online social networks. *Plos one* 15, 4 (2020), e0230811.

[5] Rushi Bhatt, Vineet Chaoji, and Rajesh Parekh. 2010. Predicting product adoption in large-scale social networks. In *Proceedings of the 19th ACM international conference on Information and knowledge management*. 1039–1048.

[6] Xavier Bresson and Thomas Laurent. 2018. An experimental study of neural networks for variable graphs. In *Workshop of the 6th International Conference on Learning Representations*.

[7] Wei Cao, Dong Wang, Jian Li, Hao Zhou, Lei Li, and Yitan Li. 2018. Brits: Bidirectional recurrent imputation for time series. *Advances in neural information processing systems* 31 (2018).

[8] Deepayan Chakrabarti, Yang Wang, Chenxi Wang, Jurij Leskovec, and Christos Faloutsos. 2008. Epidemic thresholds in real networks. *ACM Transactions on Information and System Security (TISSEC)* 10, 4 (2008), 1–26.

[9] Hongyi Chen, Jingtao Ding, Xiaojun Liang, Yong Li, and Xiao-Ping Zhang. 2025. Structure-prior Informed Diffusion Model for Graph Source Localization with Limited Data. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*. 250–259.

[10] Zhen Chen, Hanghang Tong, and Lei Ying. 2019. Inferring full diffusion history from partial timestamps. *IEEE Transactions on Knowledge and Data Engineering* 32, 7 (2019), 1378–1392.

[11] Zhen Chen, Kai Zhu, and Lei Ying. 2016. Detecting multiple information sources in networks under the SIR model. *IEEE Transactions on Network Science and Engineering* 3, 1 (2016), 17–31.

[12] Chaoran Cheng, Boran Han, Danielle C Maddix, Abdul Fatir Ansari, Andrew Stuart, Michael W Mahoney, and Yuyang Wang. 2024. Gradient-free generation for hard-constrained systems. *arXiv preprint arXiv:2412.01786* (2024).

[13] Le Cheng, Peican Zhu, Yangming Guo, Chao Gao, Zhen Wang, and Keke Tang. 2025. SourceDetMamba: A Graph-aware State Space Model for Source Detection in Sequential Hypergraphs. *arXiv preprint arXiv:2505.12910* (2025).

[14] Andrea Cini, Ivan Marisca, and Cesare Alippi. 2021. Filling the g_ap_s: Multivariate time series imputation by graph neural networks. *arXiv preprint arXiv:2108.00298* (2021).

[15] Michael Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. 2011. Political polarization on twitter. In *Proceedings of the international aaai conference on web and social media*, Vol. 5. 89–96.

[16] Wen Cui, Xiaoqing Gong, Chen Liu, Dan Xu, Xiaojiang Chen, Dingyi Fang, Shaojie Tang, Fan Wu, and Guihai Chen. 2016. Node immunization with time-sensitive restrictions. *Sensors* 16, 12 (2016), 2141.

[17] Giovanni De Felice, Andrea Cini, Daniele Zambon, Vladimir V Gusev, and Cesare Alippi. 2024. Graph-based virtual sensing from sparse and partial multivariate observations. *arXiv preprint arXiv:2402.12598* (2024).

[18] Leyan Deng, Chenwang Wu, Defu Lian, and Enhong Chen. 2024. Learning from highly sparse spatio-temporal data. *Advances in Neural Information Processing Systems* 37 (2024), 94022–94046.

[19] Leonardo Duenas-Osorio and Srivishnu Mohan Vemuru. 2009. Cascading failures in complex infrastructure systems. *Structural safety* 31, 2 (2009), 157–167.

[20] Stefan Elfwing, Eiji Uchibe, and Kenji Doya. 2018. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Networks* 107 (2018), 3–11.

[21] P Erdös, A Rényi, and B Bollobás. 1959. Publicationes mathematicae debrecen. *Random Graphs I* 6 (1959), 290–297.

[22] Hubert Escaith and Fabien Gonguet. 2009. International trade and real transmission channels of financial shocks in globalized production networks. *Available at SSRN 1408584* (2009).

[23] Johannes Gasteiger, Stefan Weißenberger, and Stephan Günnemann. 2019. Diffusion improves graph learning. *Advances in neural information processing systems* 32 (2019).

[24] Manuel Gomez-Rodriguez, Jure Leskovec, and Andreas Krause. 2012. Inferring networks of diffusion and influence. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 5, 4 (2012), 1–37.

[25] Wilfred Keith Hastings. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 1 (1970), 97–109.

[26] Herbert W Hethcote. 2000. The mathematics of infectious diseases. *SIAM review* 42, 4 (2000), 599–653.

[27] Dongpeng Hou, Yuchen Wang, Chao Gao, and Xianghua Li. 2025. A generalized diffusion framework with learnable propagation dynamics for source localization. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*. 2919–2927.

[28] Keke Huang, Ruize Gao, Bogdan Cautis, and Xiaokui Xiao. 2024. Scalable continuous-time diffusion framework for network inference and influence estimation. In *Proceedings of the ACM Web Conference 2024*. 2660–2671.

[29] Sergey Ioffe and Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine Learning*, Vol. 37. PMLR, 448–456.

[30] Baoyu Jing, Dawei Zhou, Kan Ren, and Carl Yang. 2024. Causality-aware spatiotemporal graph neural networks for spatiotemporal time series imputation. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. 1027–1037.

[31] Chaitanya K Joshi, Quentin Cappart, Louis-Martin Rousseau, and Thomas Laurent. 2020. Learning TSP requires rethinking generalization. arXiv:2006.07054 (2020).

[32] David Kempe, Jon Kleinberg, and Éva Tardos. 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*. 137–146.

[33] William Ogilvy Kermack and Anderson 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 (1927), 700–721.

[34] TN Kipf. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).

[35] Lingkai Kong, Yuanqi Du, Wenhao Mu, Kirill Neklyudov, Valentin De Bortoli, Dongxia Wu, Haorui Wang, Aaron Ferber, Yi-An Ma, Carla P Gomes, et al. 2024. Diffusion models as constrained samplers for optimization with unknown constraints. *arXiv preprint arXiv:2402.18012* (2024).

[36] K Krishnamoorthy, Avishek Mallick, and Thomas Mathew. 2009. Model-based imputation approach for data analysis in the presence of non-detects. *Annals of Occupational Hygiene* 53, 3 (2009), 249–263.

[37] Takeshi Kurashima, Tomoharu Iwata, Noriko Takaya, and Hiroshi Sawada. 2014. Probabilistic latent network visualization: inferring and embedding diffusion networks. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. 1236–1245.

[38] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. 2005. Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*. 177–187.

[39] Ziqi Li, Chaoyi Shi, Qi Zhang, and Tianguang Chu. 2024. Inferring the source of diffusion in networks under weak observation condition. *Physica A: Statistical Mechanics and its Applications* 637 (2024), 129581.

[40] Chen Ling, Tanmoy Chowdhury, Jie Ji, Sirui Li, Andreas Züfle, and Liang Zhao. 2024. Source Localization for Cross Network Information Diffusion. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 5419–5429.

[41] Chen Ling, Junji Jiang, Junxiang Wang, and Zhao Liang. 2022. Source localization of graph diffusion via variational autoencoders for graph inverse problems. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*. 1010–1020.

[42] Manuel Madeira, Clement Vignac, Dorina Thanou, and Pascal Frossard. 2024. Generative modelling of structurally constrained graphs. *Advances in Neural Information Processing Systems* 37 (2024), 137218–137262.

[43] Ivan Marisca, Andrea Cini, and Cesare Alippi. 2022. Learning to reconstruct missing data from spatiotemporal graphs with sparse observations. *Advances in neural information processing systems* 35 (2022), 32069–32082.

[44] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. 1953. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* 21, 6 (1953), 1087–1092.

[45] Xiaoye Miao, Yangyang Wu, Jun Wang, Yunjun Gao, Xudong Mao, and Jianwei Yin. 2021. Generative semi-supervised learning for multivariate time series imputation. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35. 8983–8991.

[46] Tong Nie, Guoyang Qin, Wei Ma, Yuewen Mei, and Jian Sun. 2024. ImputeFormer: Low rankness-induced transformers for generalizable spatiotemporal imputation. In *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining*. 2260–2271.

[47] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. 2015. Epidemic processes in complex networks. *Reviews of modern physics* 87, 3 (2015), 925–979.

[48] Ruizhong Qiu, Zhiqing Sun, and Yiming Yang. 2022. DIMES: A differentiable meta solver for combinatorial optimization problems. In *Advances in Neural Information Processing Systems*.

[49] Ruizhong Qiu, Dingsu Wang, Lei Ying, H Vincent Poor, Yifang Zhang, and Hanghang Tong. 2023. Reconstructing graph diffusion history from a single snapshot. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 1978–1988.

[50] Mrinmoy Ray, Anil Rai, KN Singh, Amrender Kumar, et al. 2017. Technology forecasting using time series intervention based trend impact analysis for wheat

yield scenario in India. *Technological Forecasting and Social Change* 118 (2017), 128–133.
[51] Manuel Gomez Rodriguez, Jure Leskovec, David Balduzzi, and Bernhard Schölkopf. 2014. Uncovering the structure and temporal dynamics of information propagation. *Network Science* 2, 1 (2014), 26–65.
[52] Everett M Rogers, Joseph R Ascroft, and Niels G Röling. 1970. *Diffusion of innovations in Brazil, Nigeria, and India.* Vol. 24. Department of Communication, Michigan State University.
[53] Patrick Schramowski, Manuel Brack, Björn Deiseroth, and Kristian Kersting. 2023. Safe latent diffusion: Mitigating inappropriate degeneration in diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 22522–22531.
[54] Emre Sefer and Carl Kingsford. 2016. Diffusion archeology for diffusion progression history reconstruction. *Knowledge and information systems* 49, 2 (2016), 403–427.
[55] Devavrat Shah and Tauhid Zaman. 2011. Rumors in a network: Who's the culprit? *IEEE Transactions on information theory* 57, 8 (2011), 5163–5181.
[56] Bowen Song, Soo Min Kwon, Zecheng Zhang, Xinyu Hu, Qing Qu, and Liyue Shen. 2023. Solving inverse problems with latent diffusion models via hard data consistency. *arXiv preprint arXiv:2307.08123* (2023).
[57] Ayush Tewari, Tianwei Yin, George Cazenavette, Semon Rezchikov, Josh Tenenbaum, Frédo Durand, Bill Freeman, and Vincent Sitzmann. 2023. Diffusion with forward models: Solving stochastic inverse problems without direct supervision. *Advances in Neural Information Processing Systems* 36 (2023), 12349–12362.
[58] Thomas W Valente. 1996. Network models of the diffusion of innovations.
[59] Yang Wang, Deepayan Chakrabarti, Chenxi Wang, and Christos Faloutsos. 2003. Epidemic spreading in real networks: An eigenvalue viewpoint. In *22nd International Symposium on Reliable Distributed Systems, 2003. Proceedings.* IEEE, 25–34.
[60] Yuchen Wang, Dongpeng Hou, Chao Gao, Xianghua Li, and Zhen Wang. 2024. Inferring information diffusion networks without timestamps. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management.* 2453–2461.
[61] Henry Webel, Lili Niu, Annelaura Bach Nielsen, Marie Locard-Paulet, Matthias Mann, Lars Juhl Jensen, and Simon Rasmussen. 2024. Imputation of label-free quantitative mass spectrometry-based proteomics data using self-supervised deep learning. *Nature Communications* 15, 1 (2024), 5405.
[62] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 2018. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826* (2018).
[63] Xinyu Yang, Yu Sun, Xinyang Chen, Ying Zhang, and Xiaojie Yuan. 2025. Graph Structure Learning for Spatial-Temporal Imputation: Adapting to Node and Feature Scales. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 959–967.
[64] Zhaoyang Zhang, Ziqi Chen, Qiao Liu, Jinhan Xie, and Hongtu Zhu. 2025. Sampling-guided Heterogeneous Graph Neural Network with Temporal Smoothing for Scalable Longitudinal Data Imputation. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2.* 3912–3920.
[65] Fan Zhou, Xovee Xu, Goce Trajcevski, and Kunpeng Zhang. 2021. A survey of information cascade analysis: Models, predictions, and recent advances. *ACM Computing Surveys (CSUR)* 54, 2 (2021), 1–36.
[66] Xinkai Zhou, Qiang Heng, Eric C Chi, and Hua Zhou. 2024. Proximal MCMC for Bayesian inference of constrained and regularized estimation. *The American Statistician* 78, 4 (2024), 379–390.
[67] Kai Zhu and Lei Ying. 2014. Information source detection in the SIR model: A sample-path-based approach. *IEEE/ACM Transactions on Networking* 24, 1 (2014), 408–421.

A Experimental Settings (Cont'd)

A.1 Datasets

We follow the 12 datasets from [49], spanning synthetic and real graphs, and synthetic and real diffusion. Our benchmarks cover: **(D1) Synthetic graphs + synthetic diffusion.** We generate diffusion histories on two synthetic graph models: BA [3] and ER [21] (each with 1,000 nodes). **(D2) Real graphs + synthetic diffusion.** We simulate diffusion on two real networks: Oregon2 [38] (11,461 nodes) and Prost [38] (15,810 nodes). **(D3) Real graphs + real diffusion.** We use four real diffusion datasets: BrFarmers [52, 58] (SI-like), Pol [15] (SI-like), Covid (SIR-like) [49], and Hebrew (SIR-like) [4]. For D1 and D2, we consider both SI and SIR diffusion, yielding 8 datasets: BA-SI, BA-SIR, ER-SI, ER-SIR, Oregon2-SI, Oregon2-SIR, Prost-SI, Prost-SIR. For D3, we use the real diffusion type indicated above, giving 4 additional datasets: BrFarmers-SI, Pol-SI, Covid-SIR, Hebrew-SIR (total 12). Dataset statistics are summarized in Table 3.

A.2 Baselines

Supervised methods. These methods treat diffusion history reconstruction as time-series imputation on graphs: given partially observed node states over time, they learn to fill in the missing snapshots. We include representative and widely-used models: GCN [34], GIN [62], BRITS [7], GRIN [14], and SPIN [43]. To make these methods applicable in our setting (where real complete histories are scarce), we follow standard practice in this line of work: supervised baselines are trained using diffusion histories generated from an estimated diffusion model, and then tested under the same two-snapshot masking pattern (i.e., only $\mathbf{y}_{t_{\text{obs}}}$ and $\mathbf{y}_T$ are observed).

Statistical methods. These methods reconstruct histories by optimizing (or approximating) the likelihood under an assumed diffusion process. We include two established baselines: DHREC [54] and CRI [11]. In our setting, true diffusion parameters are not assumed known. When a baseline requires diffusion parameters, it uses the same diffusion-parameter estimation routine and configuration as in our pipeline (so that differences are due to reconstruction quality rather than parameter access). Also, to isolate the benefit of leveraging additional snapshots, we include a single-snapshot baseline DITTO [49] that reconstructs the history using only the final snapshot $\mathbf{y}_T$ and discarding $\mathbf{y}_{t_{\text{obs}}}$.

B Theoretical Proofs

B.1 Proof Sketch of Proposition 3.1

Proof sketch. In each diffusion-parameter estimation iteration, we evaluate the segmented mean-field recursion over all node–time pairs. For each step t and node u, the only non-constant operation is the neighbor aggregation, whose cost is $O(|\mathcal{N}_u|)$. Summing over all $u \in \mathcal{V}$ gives $O\big(\sum_{u\in\mathcal{V}} |\mathcal{N}_u|\big) = O(m)$, plus $O(n)$ for the remaining pointwise updates. Hence, each time step costs $O(n+m)$. Since $K \le T$ and $t_1 + \sum_{k=2}^{K}(t_k - t_{k-1}) = T$, the forward evaluation of the objective is $O(T(n+m))$. Since backpropagation through this recursion has the same complexity as the forward pass, each gradient-based optimization iteration is also $O(T(n+m))$. □

B.2 Proof of Lemma 4.1

Proof. We need to establish that $P_{\boldsymbol{\beta}}[\mathbf{y}_{t,\neq ?} \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}_{t+1}] > 0$ if and only if the given condition Eq. (27) holds. By the Markov property of the SIR model, this probability is positive if and only if there exists at least one fully specified intermediate snapshot $\mathbf{y}_t^* \in \mathcal{X}^{\mathcal{V}}$ such that:

- $\mathbf{y}_t^*$ exactly matches $\mathbf{y}_t$ on all defined nodes: $y^*_{t,u} = y_{t,u}$ for all $u \in \mathcal{V}$ where $y_{t,u} \neq ?$;
- The state monotonicity is maintained: $\mathbf{y}_{t_{k-1}} \le \mathbf{y}_t^* \le \mathbf{y}_{t+1}$;
- The sequence is causally feasible: $P_{\boldsymbol{\beta}}[\mathbf{y}_t^* \mid \mathbf{y}_{t_{k-1}}] > 0$ and $P_{\boldsymbol{\beta}}[\mathbf{y}_{t+1} \mid \mathbf{y}_t^*] > 0$.

Part I ("only if"). We have $P_{\boldsymbol{\beta}}[\mathbf{y}_{t,\neq ?} \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}_{t+1}] > 0$. Let $\mathbf{y}_t^*$ be a valid complete snapshot corresponding to $\mathbf{y}_t$. Take any node $v \in \mathcal{V}^{\text{IR}}(\mathbf{y}_{t+1}) \setminus \mathcal{V}^{\text{R}}(\mathbf{y}_{t_{k-1}})$. We evaluate the necessary conditions based on its state in the incomplete snapshot $\mathbf{y}_t$.

Table 7: Distribution of proposed $y_{t,u}$.

$\mathcal{X}_{t,u}$	S	I	R
{S}	1	0	0
{S, I}	$q^{\mathtt{S}}_{t,u}$	$1-q^{\mathtt{S}}_{t,u}$	0
{S, I, R}	$q^{\mathtt{I}}_{t,u}q^{\mathtt{S}}_{t,u}$	$q^{\mathtt{I}}_{t,u}(1-q^{\mathtt{S}}_{t,u})$	$1-q^{\mathtt{I}}_{t,u}$
{I}	0	1	0
{I, R}	0	$q^{\mathtt{I}}_{t,u}$	$1-q^{\mathtt{I}}_{t,u}$
{R}	0	0	1

Case I.1: $y_{t,v} = \mathtt{S}$. Since $\mathbf{y}^*_t$ matches $\mathbf{y}_t$, we have $y^*_{t,v} = \mathtt{S}$. Because $v \in \mathcal{V}^{\mathtt{IR}}(\mathbf{y}_{t+1})$, node v transitions from S at time t to either I or R at time $t+1$. Under the SIR model dynamics, transitioning from S to either I or R strictly requires v to have at least one infected neighbor at time t. Thus, there exists $u \in \mathcal{N}_v$ such that $y^*_{t,u} = \mathtt{I}$. For u to be infected by time t, there must exist a causal infection path from the initially infected nodes $\mathcal{V}^{\mathtt{I}}(\mathbf{y}_{t_{k-1}})$ to u of length at most $t - t_{k-1}$. All nodes along this path must be infected by time t, meaning they belong to $\mathcal{V}_t$. Consequently, $u \in \mathcal{W}_t$. Furthermore, since $y^*_{t,u} = \mathtt{I}$, we know $y_{t,u} \neq \mathtt{R}$, yielding $u \in \mathcal{W}_t \setminus \mathcal{V}^{\mathtt{R}}(\mathbf{y}_t)$. Thus, $v \in \bigcup_{u \in \mathcal{W}_t \setminus \mathcal{V}^{\mathtt{R}}(\mathbf{y}_t)} \mathcal{N}_u$.

Case I.2: $y_{t,v} \in \{\mathtt{I}, \mathtt{R}\}$. This implies $y^*_{t,v} \in \{\mathtt{I}, \mathtt{R}\}$, meaning v itself was already infected by time t. By the same causal graph reasoning as above, there must exist a valid infection propagation path from $\mathcal{V}^{\mathtt{I}}(\mathbf{y}_{t_{k-1}})$ to v entirely within $\mathcal{V}_t$ of length $d \le t - t_{k-1}$. Therefore, $v \in \mathcal{W}_t$.

Case I.3: $y_{t,v} = ?$. In the fully specified snapshot $\mathbf{y}^*_t$, the actual state $y^*_{t,v}$ must be either S, I, or R. If $y^*_{t,v} = \mathtt{S}$, the logic of Case I.1 applies. If $y^*_{t,v} \in \{\mathtt{I}, \mathtt{R}\}$, the logic of Case I.2 applies. Because v must fall into one of these two disjoint scenarios, it unconditionally follows that $v \in \mathcal{W}_t \cup \bigcup_{u \in \mathcal{W}_t \setminus \mathcal{V}^{\mathtt{R}}(\mathbf{y}_t)} \mathcal{N}_u$.

Part II ("if"). Suppose that the piecewise condition holds for every $v \in \mathcal{V}^{\mathtt{IR}}(\mathbf{y}_{t+1}) \setminus \mathcal{V}^{\mathtt{R}}(\mathbf{y}_{t_{k-1}})$. We construct a fully specified candidate snapshot $\mathbf{y}^*_t$ as follows: for each $u \in \mathcal{V}$,

- If $y_{t,u} \neq ?$, set $y^*_{t,u} = y_{t,u}$.
- If $y_{t,u} = ?$, set $y^*_{t,u} = \mathtt{I}$ if $u \in \mathcal{W}_t$, and set $y^*_{t,u} = \mathtt{S}$ otherwise.

First, we verify $P_{\boldsymbol{\beta}}[\mathbf{y}^*_t \mid \mathbf{y}_{t_{k-1}}] > 0$. We must demonstrate a feasible sequence of events from t_{k-1} to t yielding $\mathbf{y}^*_t$. By definition of $\mathcal{W}_t$, every node u assigned $y^*_{t,u} \in \{\mathtt{I}, \mathtt{R}\}$ has a path in $\mathcal{V}_t$ connecting it to an initially infected node in $\mathcal{V}^{\mathtt{I}}(\mathbf{y}_{t_{k-1}})$ of length $\le t - t_{k-1}$. A valid infection propagation schedule can be constructed via *breadth-first search*, where each layer d is sequentially infected exactly at time $t_{k-1} + d$. For nodes requiring a direct transition to R at time t, the defined SIR dynamics uniquely permit a node to simultaneously infect its neighbors and transition to R within the exact same time step. Thus, the temporal trajectory reaching $\mathbf{y}^*_t$ is entirely feasible without violating any transition constraints.

Second, we verify $P_{\boldsymbol{\beta}}[\mathbf{y}_{t+1} \mid \mathbf{y}^*_t] > 0$. Consider any node v that must become infected at time $t+1$ (i.e., $y^*_{t,v} = \mathtt{S}$ and $y_{t+1,v} \in \{\mathtt{I}, \mathtt{R}\}$). By our construction, $y^*_{t,v} = \mathtt{S}$ implies $v \notin \mathcal{W}_t$. According to the assumed condition, v must therefore belong to $\bigcup_{u \in \mathcal{W}_t \setminus \mathcal{V}^{\mathtt{R}}(\mathbf{y}_t)} \mathcal{N}_u$. This guarantees the existence of a neighbor u such that $u \in \mathcal{W}_t$ and $u \notin \mathcal{V}^{\mathtt{R}}(\mathbf{y}_t)$. Because $u \in \mathcal{W}_t \subseteq \mathcal{V}_t$, then $y_{t,u} \neq \mathtt{S}$. Since $u \notin \mathcal{V}^{\mathtt{R}}(\mathbf{y}_t)$, then $y_{t,u} \neq \mathtt{R}$. If $y_{t,u}$ is defined, it must be I. If $y_{t,u} = ?$, our construction explicitly assigns $y^*_{t,u} = \mathtt{I}$ because $u \in \mathcal{W}_t$. In either case, $y^*_{t,u} = \mathtt{I}$, ensuring v has an infected neighbor at time t, which guarantees a positive probability for the transition to $\mathbf{y}_{t+1}$.

Since both $P_{\boldsymbol{\beta}}[\mathbf{y}^*_t \mid \mathbf{y}_{t_{k-1}}] > 0$ and $P_{\boldsymbol{\beta}}[\mathbf{y}_{t+1} \mid \mathbf{y}^*_t] > 0$ hold, then by Bayes' theorem and the Markov property of SIR,

$$P_{\boldsymbol{\beta}}[\mathbf{y}_{t,\neq ?} \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}_{t+1}] \tag{35}$$

$$\ge P_{\boldsymbol{\beta}}[\mathbf{y}^*_t \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}_{t+1}] \tag{36}$$

$$= \frac{P_{\boldsymbol{\beta}}[\mathbf{y}^*_t \mid \mathbf{y}_{t_{k-1}}] P_{\boldsymbol{\beta}}[\mathbf{y}_{t+1} \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}^*_t]}{P_{\boldsymbol{\beta}}[\mathbf{y}_{t+1} \mid \mathbf{y}_{t_{k-1}}]} \tag{37}$$

$$= \frac{P_{\boldsymbol{\beta}}[\mathbf{y}^*_t \mid \mathbf{y}_{t_{k-1}}] P_{\boldsymbol{\beta}}[\mathbf{y}_{t+1} \mid \mathbf{y}^*_t]}{P_{\boldsymbol{\beta}}[\mathbf{y}_{t+1} \mid \mathbf{y}_{t_{k-1}}]} > 0. \qquad \square$$

B.3 Proof of Theorem 4.2

Proof sketch. To show $\mathrm{supp}(Q_{\boldsymbol{\theta}}(\mathrm{Y}_{\mathcal{T}_{\mathrm{obs}}})) = \mathrm{supp}(P \mid \mathrm{Y}_{\mathcal{T}_{\mathrm{obs}}})$, we will show that $\mathrm{supp}(Q_{\boldsymbol{\theta}}(\mathrm{Y}_{\mathcal{T}_{\mathrm{obs}}})) \subseteq \mathrm{supp}(P \mid \mathrm{Y}_{\mathcal{T}_{\mathrm{obs}}})$ and that $\mathrm{supp}(Q_{\boldsymbol{\theta}}(\mathrm{Y}_{\mathcal{T}_{\mathrm{obs}}})) \supseteq \mathrm{supp}(P \mid \mathrm{Y}_{\mathcal{T}_{\mathrm{obs}}})$ separately.

Part I ($\subseteq$). Take any history $\mathrm{Y} = (\mathbf{y}_0, \ldots, \mathbf{y}_T)$ with $Q_{\boldsymbol{\theta}}(\mathrm{Y}_{\mathcal{T}_{\mathrm{obs}}})[\mathrm{Y}] > 0$. For any $k = 2, \ldots, K$ and any $t = t_{k-1}+1, \ldots, t_k - 1$, the probability distribution of $y_{t,u}$ is shown in Table 7; note that $y_{t,u} \in \mathcal{X}_{t,u}$ with probability 1. Then by Lemma 4.1, we have ensured that

$$P_{\boldsymbol{\beta}}[\mathbf{y}_t \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}_{t+1}] > 0. \tag{38}$$

Hence, by Bayes' theorem, the Markov property of SIR, and a telescoping product,

$$P_{\boldsymbol{\beta}}[\mathrm{Y}_{(t_{k-1},t_k]} \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}_{t_k}] \tag{39}$$

$$= \frac{P_{\boldsymbol{\beta}}[\mathrm{Y}_{(t_{k-1},t_k]} \mid \mathbf{y}_{t_{k-1}}]}{P_{\boldsymbol{\beta}}[\mathbf{y}_{t_k} \mid \mathbf{y}_{t_{k-1}}]} = \frac{\prod_{t=t_{k-1}}^{t_k-1} P_{\boldsymbol{\beta}}[\mathbf{y}_{t+1} \mid \mathbf{y}_t]}{P_{\boldsymbol{\beta}}[\mathbf{y}_{t_k} \mid \mathbf{y}_{t_{k-1}}]} \tag{40}$$

$$= \frac{P_{\boldsymbol{\beta}}[\mathbf{y}_{t_{k-1}+1} \mid \mathbf{y}_{t_{k-1}}]}{P_{\boldsymbol{\beta}}[\mathbf{y}_{t_k} \mid \mathbf{y}_{t_{k-1}}]} \prod_{t=t_{k-1}+1}^{t_k-1} P_{\boldsymbol{\beta}}[\mathbf{y}_{t+1} \mid \mathbf{y}_t] \tag{41}$$

$$= \left(\prod_{t=t_{k-1}+1}^{t_k-1} \frac{P_{\boldsymbol{\beta}}[\mathbf{y}_t \mid \mathbf{y}_{t_{k-1}}]}{P_{\boldsymbol{\beta}}[\mathbf{y}_{t+1} \mid \mathbf{y}_{t_{k-1}}]} \right) \left(\prod_{t=t_{k-1}+1}^{t_k-1} P_{\boldsymbol{\beta}}[\mathbf{y}_{t+1} \mid \mathbf{y}_t] \right) \tag{42}$$

$$= \prod_{t=t_{k-1}+1}^{t_k-1} \frac{P_{\boldsymbol{\beta}}[\mathbf{y}_t \mid \mathbf{y}_{t_{k-1}}] P_{\boldsymbol{\beta}}[\mathbf{y}_{t+1} \mid \mathbf{y}_t]}{P_{\boldsymbol{\beta}}[\mathbf{y}_{t+1} \mid \mathbf{y}_{t_{k-1}}]} \tag{43}$$

$$= \prod_{t=t_{k-1}+1}^{t_k-1} \frac{P_{\boldsymbol{\beta}}[\mathbf{y}_t \mid \mathbf{y}_{t_{k-1}}] P_{\boldsymbol{\beta}}[\mathbf{y}_{t+1} \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}_t]}{P_{\boldsymbol{\beta}}[\mathbf{y}_{t+1} \mid \mathbf{y}_{t_{k-1}}]} \tag{44}$$

$$= \prod_{t=t_{k-1}+1}^{t_k-1} P_{\boldsymbol{\beta}}[\mathbf{y}_t \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}_{t+1}] \tag{45}$$

$$> 0. \tag{46}$$

Similarly, for $k = 1$, we also have

$$P_{\boldsymbol{\beta}}[\mathrm{Y}_{[0,t_1]} \mid \mathbf{y}_{t_1}] > 0. \tag{47}$$

It follows from the Markov property that

$$P_{\boldsymbol{\beta}}[\mathrm{Y} \mid \mathrm{Y}_{\mathcal{T}_{\mathrm{obs}}}] = P_{\boldsymbol{\beta}}[\mathrm{Y}_{[0,t_1]} \mid \mathbf{y}_{t_1}] \prod_{k=2}^{K} P_{\boldsymbol{\beta}}[\mathrm{Y}_{(t_{k-1},t_k]} \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}_{t_k}] > 0.$$

Therefore, $\mathrm{supp}(Q_{\boldsymbol{\theta}}(\mathrm{Y}_{\mathcal{T}_{\mathrm{obs}}})) \subseteq \mathrm{supp}(P \mid \mathrm{Y}_{\mathcal{T}_{\mathrm{obs}}})$.

Part II ($\supseteq$). Take any history Y with $P_{\boldsymbol{\beta}}[\mathrm{Y} \mid \mathrm{Y}_{\mathcal{T}_{\mathrm{obs}}}] > 0$. For any $k = 2, \ldots, K$, any $t = t_k - 1, \ldots, t_{k-1} + 1$, suppose that nodes are processed in the order of $u_1, \ldots, u_n$. Let $\mathbf{y}^0_t$ denote a snapshot of

all ?'s (i.e., undefined); and for each node u_i ($i = 1, \ldots, n$), let $\mathbf{y}_t^i$ denote an incomplete version of $\mathbf{y}_t$ where nodes processed after u_i are marked as ?. Since $P_{\boldsymbol{\beta}}[\mathbf{Y} \mid \mathbf{Y}_{\mathcal{T}_{\text{obs}}}] > 0$, then by Eq. (45), we have $P_{\boldsymbol{\beta}}[\mathbf{y}_t \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}_{t+1}] > 0$. This implies

$$P_{\boldsymbol{\beta}}[\mathbf{y}_{t,\neq ?}^i \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}_{t+1}] \geq P_{\boldsymbol{\beta}}[\mathbf{y}_t \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}_{t+1}] > 0. \tag{48}$$

By Lemma 4.1, $y_{t,u_i}^i \in \mathcal{X}_{t,u_i}$. Since $0 < q_{t,u}^{\mathrm{I}}, q_{t,u}^{\mathrm{S}} < 1$, then by Table 7,

$$\operatorname{supp}(Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[y_{t,u_i}^i = \cdot \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}_t^{i-1}, \mathbf{y}_{t_k}]) = \mathcal{X}_{t,u_i}. \tag{49}$$

This implies that

$$Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[y_{t,u_i}^i \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}_t^{i-1}, \mathbf{y}_{t_k}] > 0. \tag{50}$$

Hence, by the definition of $Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[\mathbf{Y}_{(t_{k-1},t_k]} \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}_{t_k}]$,

$$\begin{aligned}
& Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[\mathbf{Y}_{(t_{k-1},t_k]} \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}_{t_k}] && (51)\\
&= \prod_{t=t_k-1}^{t_{k-1}+1} Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[\mathbf{y}_t \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}_{t_k}] && (52)\\
&= \prod_{t=t_k-1}^{t_{k-1}+1} Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[\mathbf{y}_t^n \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}_{t_k}] && (53)\\
&= \prod_{t=t_k-1}^{t_{k-1}+1} Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[\mathbf{y}_t^n \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}_t^0, \mathbf{y}_{t_k}] && (54)\\
&= \prod_{t=t_k-1}^{t_{k-1}+1} \prod_{i=1}^{n} Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[\mathbf{y}_t^i \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}_t^{i-1}, \mathbf{y}_{t_k}] && (55)\\
&= \prod_{t=t_k-1}^{t_{k-1}+1} \prod_{i=1}^{n} Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[y_{t,u_i}^i \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}_t^{i-1}, \mathbf{y}_{t_k}] && (56)\\
&> 0. && (57)
\end{aligned}$$

Similarly, for $k = 1$, we also have

$$Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[\mathbf{Y}_{[0,t_1]} \mid \mathbf{y}_{t_1}] > 0. \tag{58}$$

It follows from the definition of $Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[\mathbf{Y}]$ that

$$\begin{aligned}
& Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[\mathbf{Y}]\\
&= Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[\mathbf{Y}_{[0,t_1]} \mid \mathbf{y}_{t_1}] \prod_{k=2}^{K} Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})[\mathbf{Y}_{(t_{k-1},t_k]} \mid \mathbf{y}_{t_{k-1}}, \mathbf{y}_{t_k}]\\
&> 0. && (59)
\end{aligned}$$

Therefore, $\operatorname{supp}(Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})) \supseteq \operatorname{supp}(P \mid \mathbf{Y}_{\mathcal{T}_{\text{obs}}})$.

Conclusion. Combining Parts I & II, we conclude that

$$\operatorname{supp}(Q_{\boldsymbol{\theta}}(\mathbf{Y}_{\mathcal{T}_{\text{obs}}})) = \operatorname{supp}(P \mid \mathbf{Y}_{\mathcal{T}_{\text{obs}}}). \qquad \square$$

B.4 Proof Sketch of Proposition 4.3

Proof sketch. The history $\mathbf{Y}$ has $O(Tn)$ entries $y_{t,u}$ to sample. For each entry $y_{t,u}$, we can run breadth-first search in $O(n+m)$ time to check whether Eq. 27 is satisfied. Therefore, the total time complexity is

$$O(Tn) \cdot O(n+m) = O(Tn(n+m)). \qquad \square$$