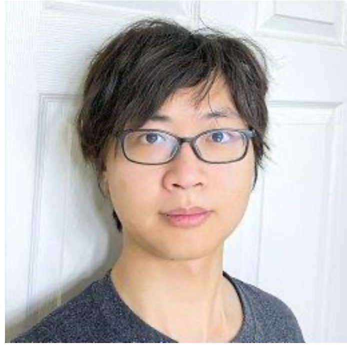
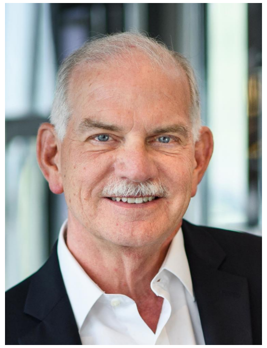
Reconstructing Graph Diffusion History from a Single Snapshot



Ruizhong Qiu
UIUC
(Presenter)

Dingsu Wang
UIUC

Lei Ying
UMich

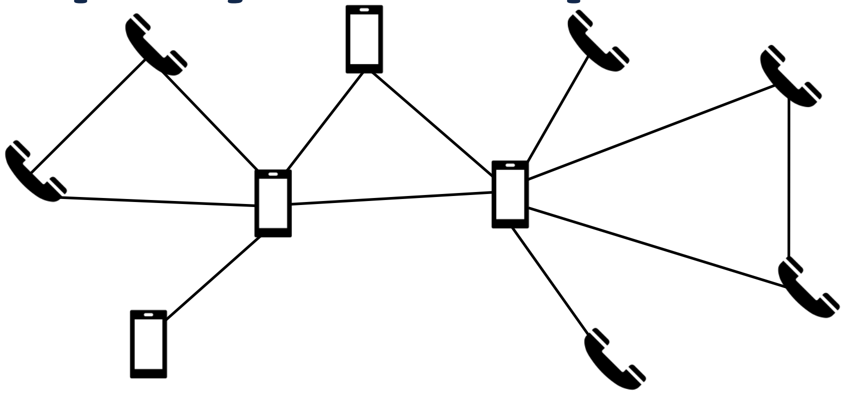
H. Vincent Poor
Princeton

Yifang Zhang
C3.ai DTI

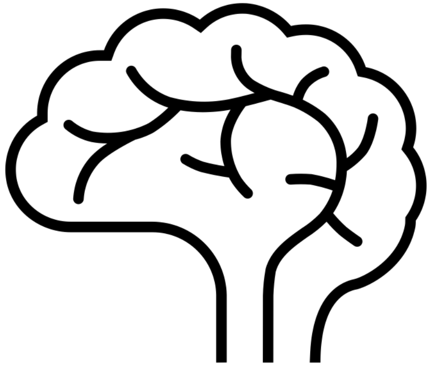
Hanghang Tong
UIUC

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN
UNIVERSITY OF MICHIGAN
PRINCETON UNIVERSITY
C3.ai Digital Transformation Institute

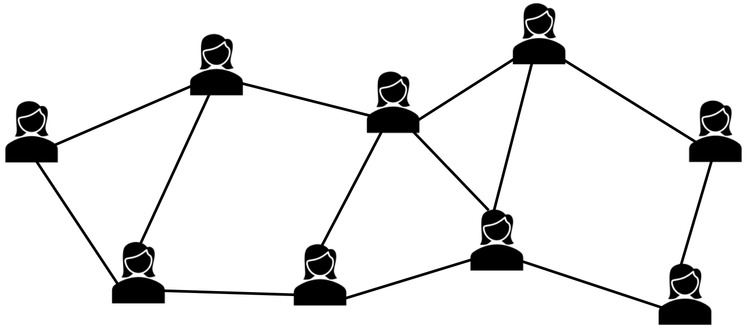
Ubiquity of Graphs



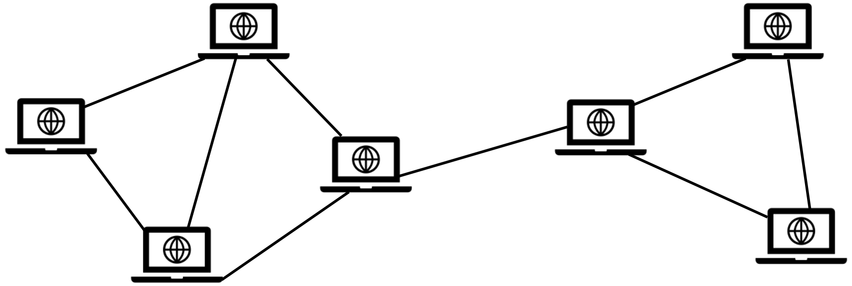
Sociology:
Communication Network [1]



Neuroscience:
Brain Network [2]



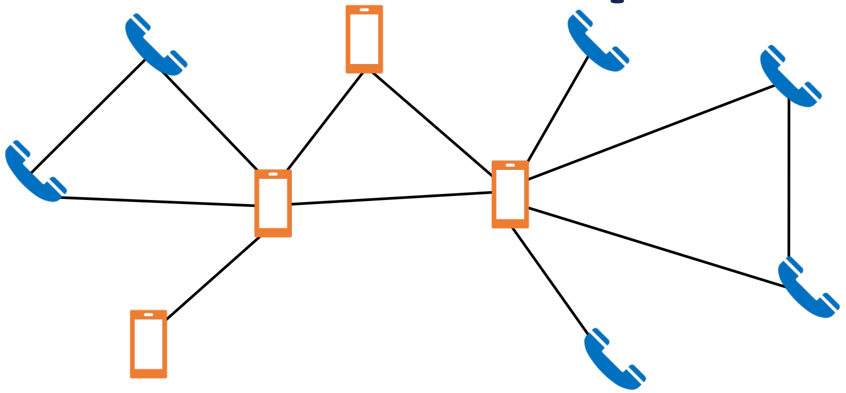
Epidemiology:
Contact Network [3]



Cybersecurity:
Computer Network [4]

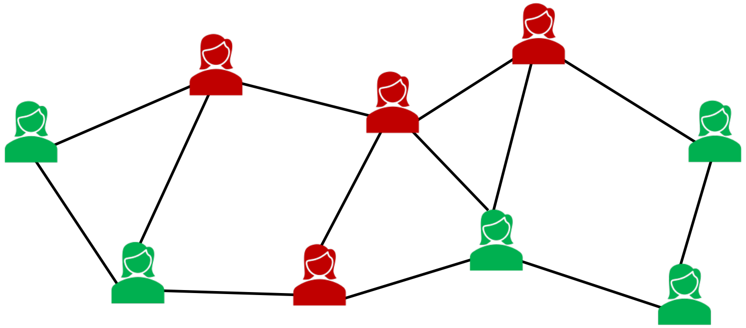
[1] Valente. *Network Models of the Diffusion of Innovations* (2nd edition, 1995). Hampton Press.
[2] Avena-Koenigsberger et al. Communication dynamics in complex brain networks. *Nature Reviews Neuroscience* 19, 1 (2018), 17–33.
[3] Klovdahl. Social networks and the spread of infectious diseases: The AIDS example. *Social Science & Medicine* 21, 11 (1985), 1203–1216.
[4] Wang et al. Understanding the spreading patterns of mobile phone viruses. *Science* 324, 5930 (2009), 1071–1076.

Diffusion on Graphs

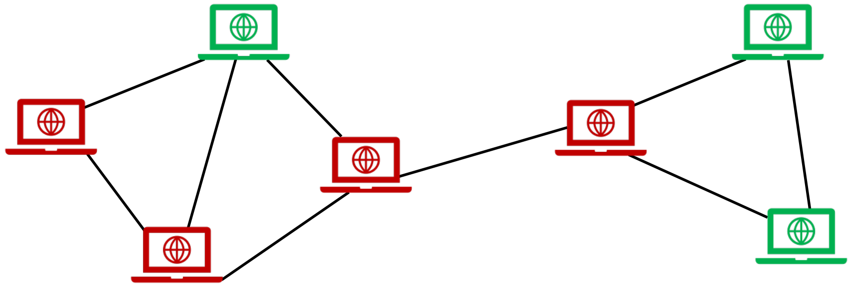


Sociology:
Diffusion of Innovations [1]

Neuroscience:
Activation Cascading [2]



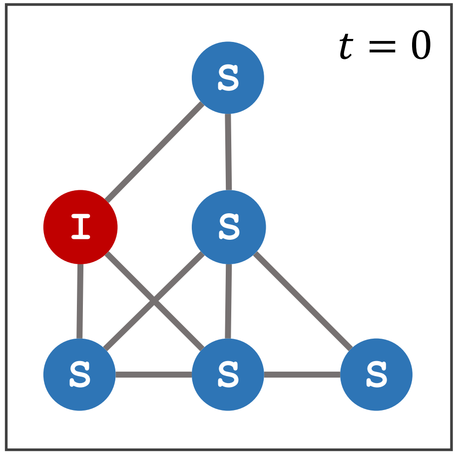
Epidemiology:
Disease Contagion [3]



Cybersecurity:
Malware Spreading [4]

[1] Valente. *Network Models of the Diffusion of Innovations* (2^{nd} edition, 1995). Hampton Press.
[2] Avena-Koenigsberger et al. Communication dynamics in complex brain networks. *Nature Reviews Neuroscience* 19, 1 (2018), 17–33.
[3] Klovdahl. Social networks and the spread of infectious diseases: The AIDS example. *Social Science & Medicine* 21, 11 (1985), 1203–1216.
[4] Wang et al. Understanding the spreading patterns of mobile phone viruses. *Science* 324, 5930 (2009), 1071–1076.

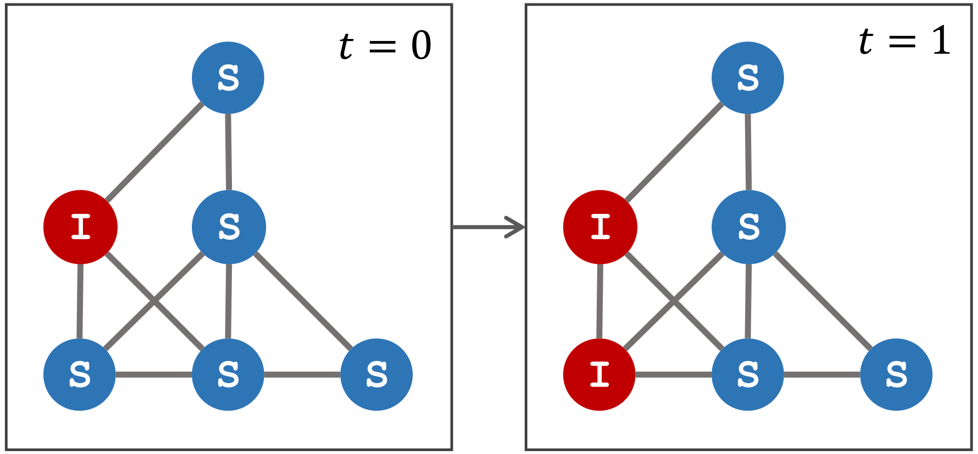
Example: Disease Contagion



$\mathcal{X} = \{$ S Susceptible I Infected R Recovered $\}$

- At time $t = 0$, only one node was infected.

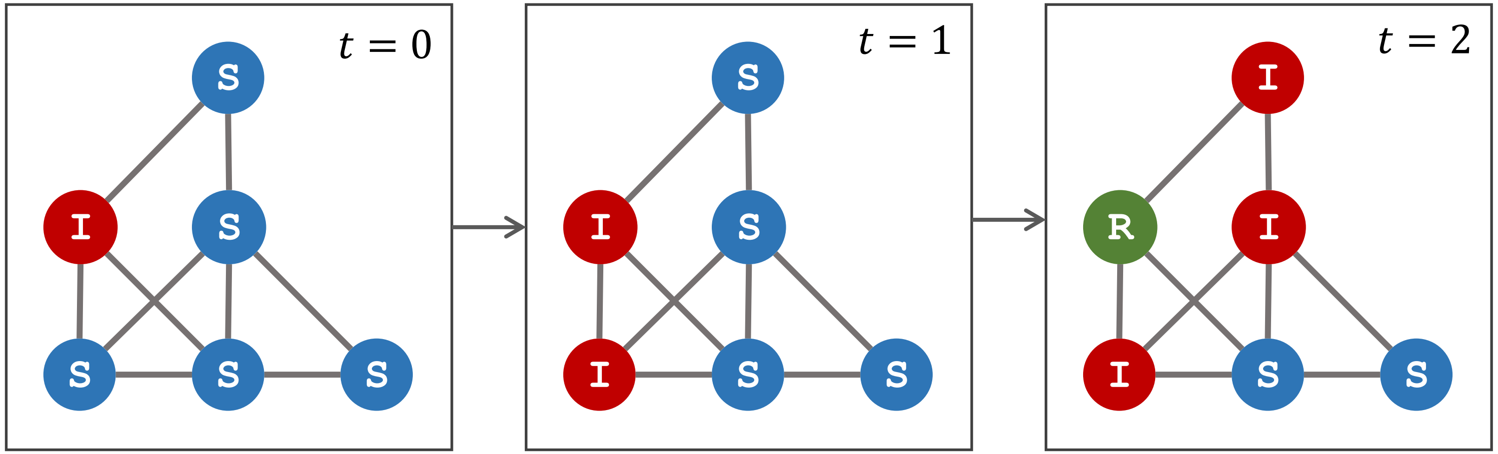
Example: Disease Contagion



$\mathcal{X} = \{$ S Susceptible I Infected R Recovered $\}$

- At time $t = 0$, only one node was infected.
- At time $t = 1$, a neighbor got infected, but the other did not.

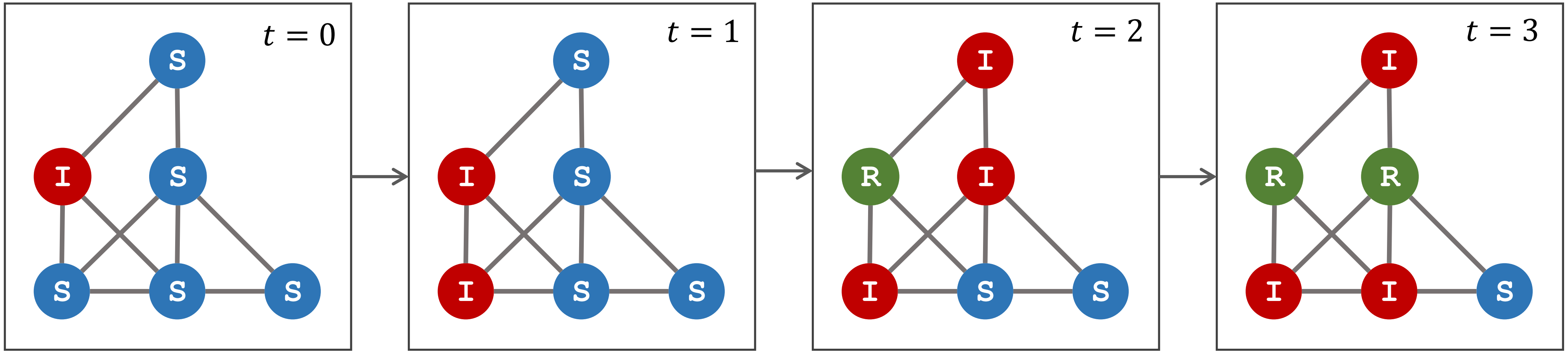
Example: Disease Contagion



$\mathcal{X} = \{$ S Susceptible, I Infected, R Recovered $\}$

- At time $t = 0$, only one node was infected.
- At time $t = 1$, a neighbor got infected, but the other did not.
- At time $t = 2$, two nodes got infected, and an infected node recovered.

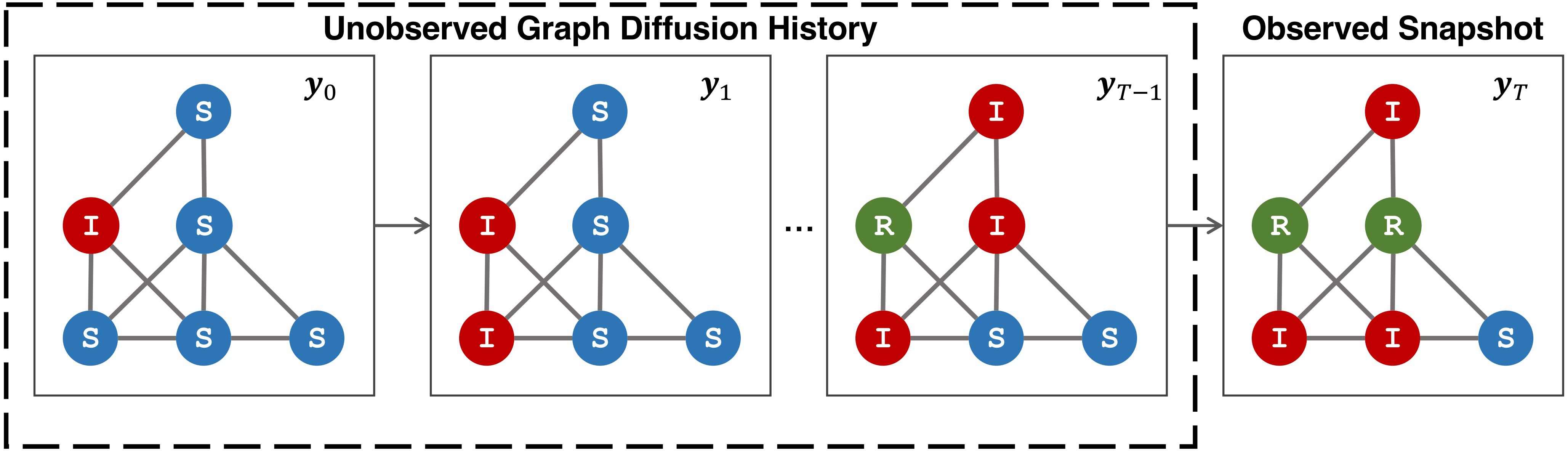
Example: Disease Contagion



$\mathcal{X} = \{$ S Susceptible, I Infected, R Recovered $\}$

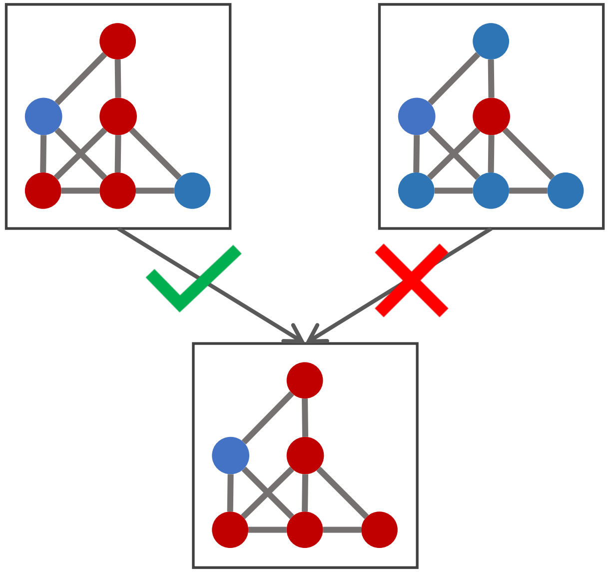
- At time $t = 0$, only one node was infected.
- At time $t = 1$, a neighbor got infected, but the other did not.
- At time $t = 2$, two nodes got infected, and an infected node recovered.
- At time $t = 3$, a new node got infected, and one more node recovered.

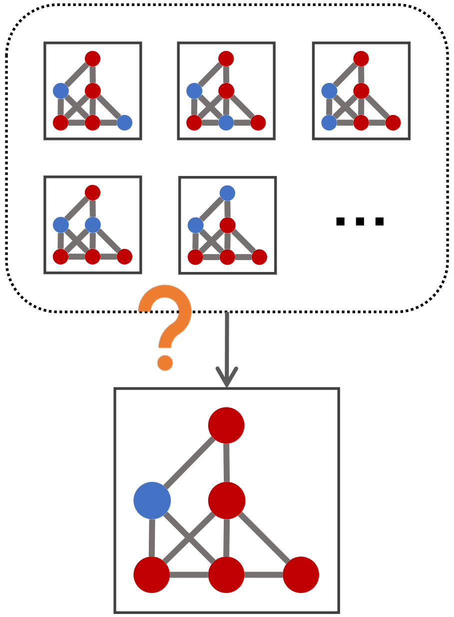
Problem Definition

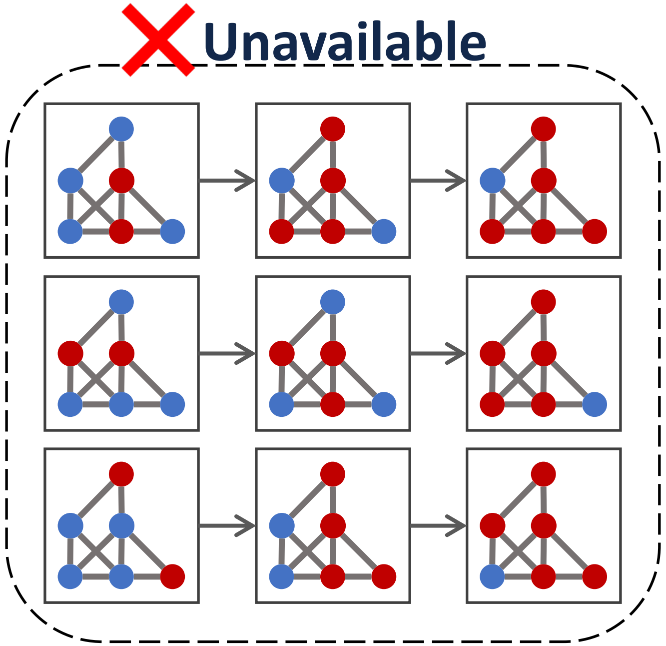


- **Problem** (DASH): *Reconstructing Diffusion history from A single SnapsHot.*
  - ***Input:*** *(i) graph* $(\mathcal{V}, \mathcal{E})$*;* *(ii) timespan* T *of interest;* *(iii) final snapshot* $\boldsymbol{y}_T \in \mathcal{X}^{\mathcal{V}}$*;* *(iv) initial distribution* $P[\boldsymbol{y}_0]$.
  - ***Output:*** *reconstructed complete diffusion history* $\widehat{\boldsymbol{Y}} = [\widehat{\boldsymbol{y}}_0, \dots, \widehat{\boldsymbol{y}}_{T-1}, \boldsymbol{y}_T]^{\mathrm{T}} \in \mathcal{X}^{\mathcal{T} \times \mathcal{V}}$.
    - ➢We **do not** assume knowing **true** diffusion parameters.

Challenges of the DASH Problem





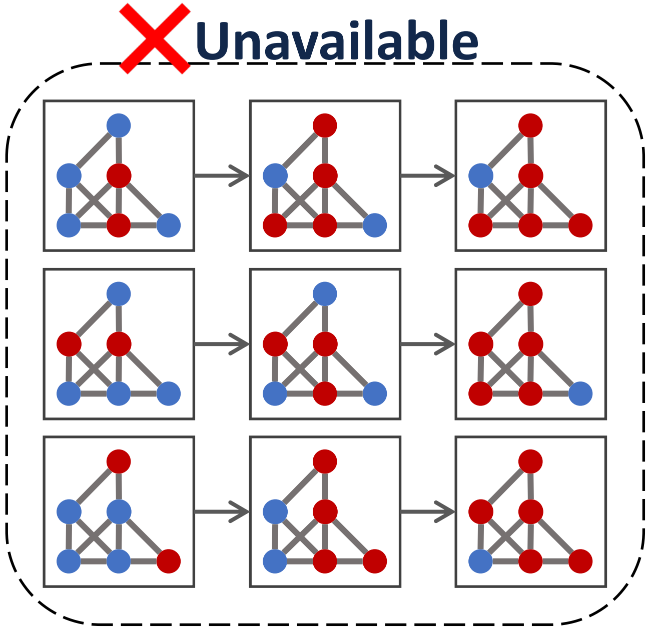


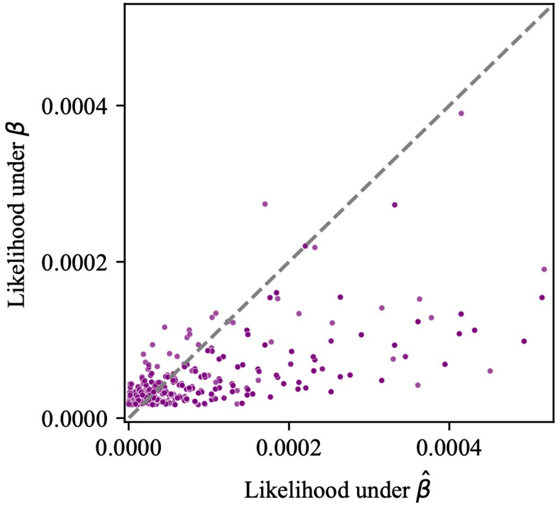
C1: Ill-posedness
Need appropriate inductive bias

C2: Explosive search space
Exponentially many possibilities

C3: Scarcity of training data
Few history data in practice

Previous Methods & Their Limitations





(a) $P_{\boldsymbol{\beta}}[Y]$ vs $P_{\widehat{\boldsymbol{\beta}}}[Y]$ in the MLE formulation.

- *Supervised time series imputation* (e.g., [1, 2])
  - ➢**Impractical** due to the scarcity of training data.
- *Maximum likelihood estimation* (MLE; e.g., [3, 4])
  - ➢**Sensitive** to estimation error of diffusion parameters (our Theorems 1 & 2).

[1] Cini et al. Filling the G_ap_s: Multivariate time series imputation by graph neural networks. *International Conference on Learning Representations* (2022).
[2] Marisca et al. Learning to reconstruct missing data from spatiotemporal graphs with sparse observations. *Advances in Neural Information Processing Systems* (2022).
[3] Sefer & Kingsford. Diffusion archeology for diffusion progression history reconstruction. *Knowledge and Information Systems* 49, 2 (2016), 403–427.
[4] Chen et al. Detecting multiple information sources in networks under the SIR model. *IEEE Transactions on Network Science and Engineering* 3, 1 (2016), 17–31.

Summary of Main Results

- **Theoretical insights:** Fundamental limitation of the MLE formulation.
  - Theorems 1 & 2 $\Longrightarrow$ **Unavoidable** estimation error of diffusion parameters.
  - Theorem 3 $\Longrightarrow$ The MLE formulation is **sensitive** to that estimation error.

- **Problem formulation:**
  - A novel ***barycenter formulation*** based on *hitting times*.
  - Provably **stable** against estimation error of diffusion parameters.

- **Proposed method:** *<u>DI</u>ffusion hi<u>T</u>ting <u>T</u>imes with <u>O</u>ptimal proposal* (DITTO).
  - Reducing the problem to estimating *posterior expected hitting times* via M–H MCMC;
  - Using a GNN to learn an **optimal** proposal to accelerate convergence of M–H MCMC.

GNN = Graph Neural Network
M–H MCMC = Metropolis–Hastings Markov Chain Monte Carlo

Contents

- ✓Introduction
- ➢**Revisiting Diffusion History MLE**
- Proposed Method: DITTO
- Main Experiments
- Conclusion

NP-Hardness of Diffusion Parameter Estimation

- To estimate diffusion parameters $\boldsymbol{\beta}$, a conventional approach is MLE:

$$\max_{\widehat{\boldsymbol{\beta}}} P_{\widehat{\boldsymbol{\beta}}}[\boldsymbol{y}_T]. \quad (\star)$$

- **Theorem 1** (informal): *Computing the probability* $P_{\widehat{\boldsymbol{\beta}}}[\boldsymbol{y}_T]$ *is NP-hard.*
  - ➢ $O\left(\binom{T+1}{2}^n (n+m)\right)$ time.

❖ Think deeper: Is there an algo for $\widehat{\boldsymbol{\beta}}$ MLE **without** computing $P_{\widehat{\boldsymbol{\beta}}}[\boldsymbol{y}_T]$?

- **Theorem 2** (informal): *Diffusion parameter MLE* $(\star)$ *is NP-hard.*

$\Longrightarrow$ Implication: Estimation error of $\boldsymbol{\beta}$ is **unavoidable**.

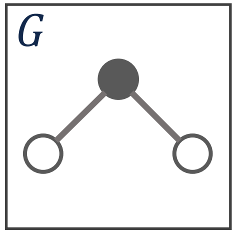
Proof Sketch of Theorem 1

- By reduction from **Minimum Dominating Set** (MDS; NP-complete).
- Suppose an algo that can compute $p \approx P_{\hat{\boldsymbol{\beta}}}[\boldsymbol{y}_T]$ up to relative error ϵ.
- Given any MDS instance G, we can construct an SI instance such that

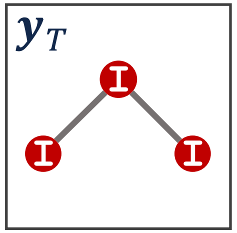
$$\mathrm{MDS}(G) = n - \left\lfloor \log_{\left(\frac{1+\epsilon}{1-\epsilon}2^n\right)} \left(\frac{\left(1 + \frac{1+\epsilon}{1-\epsilon}2^n\right)^n}{1-\epsilon} p \right) \right\rfloor.$$

  ➢ n = # nodes in G

- Example:



$\xrightarrow{\epsilon = 0}$



$T = 1 \quad \hat{\beta}^{\mathrm{I}} = 1$

$P[\boldsymbol{y}_0] = \dfrac{8^{3-n^{\mathrm{I}}(\boldsymbol{y}_0)}}{729} \Rightarrow$

$p = P_{\hat{\boldsymbol{\beta}}}[\boldsymbol{y}_T] = 89/729$

$3 - \lfloor \log_8(729p) \rfloor = 1 = \mathrm{MDS}(G)$

$\mathrm{MDS}(G) = 1$

SI instance

- Remark: Need **arbitrary-precision** arithmetics with $\mathrm{poly}(n, \log 1/\epsilon)$ bits.

Sensitivity to Estimation Error of Diffusion Parameters

- ***MLE formulation*** for diffusion history reconstruction:

$$\max_{\widehat{Y} \in \mathrm{supp}(P|y_T)} P_{\widehat{\boldsymbol{\beta}}}[\widehat{\boldsymbol{Y}}].$$

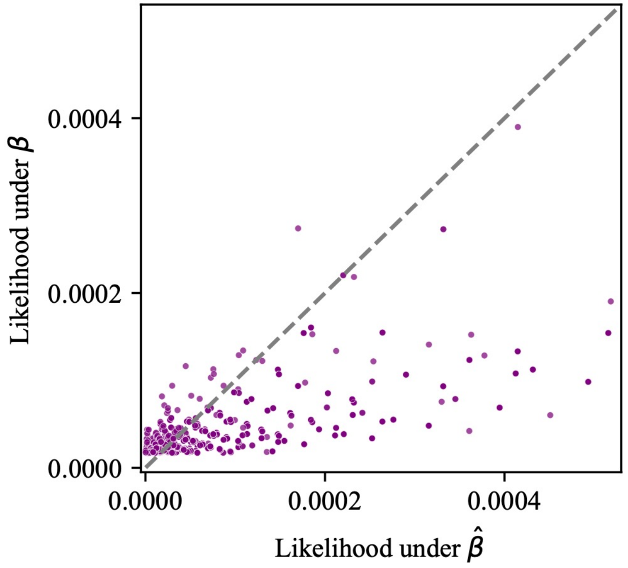
- **Theorem 3**. *Under the SIR model and mild conditions, for all possible history $\boldsymbol{Y}$, we have:*

$$\frac{\partial}{\partial \beta^{\mathrm{I}}} P_{\boldsymbol{\beta}}[\boldsymbol{Y}] = \Theta\left(\frac{1}{\beta^{\mathrm{I}}}\right) P_{\boldsymbol{\beta}}[\boldsymbol{Y}],$$

$$\frac{\partial}{\partial \beta^{\mathrm{R}}} P_{\boldsymbol{\beta}}[\boldsymbol{Y}] = \Theta\left(\frac{1}{\beta^{\mathrm{R}}}\right) P_{\boldsymbol{\beta}}[\boldsymbol{Y}].$$

➢ **Large** for small $\boldsymbol{\beta}$

❖Real-world diffusion typically has **small** true $\boldsymbol{\beta}$ [1, 2].



(a) $P_{\boldsymbol{\beta}}[Y]$ vs $P_{\widehat{\boldsymbol{\beta}}}[Y]$ in the MLE formulation.

$\Longrightarrow$ Implication: MLE formulation is **sensitive** to estimation error of $\widehat{\boldsymbol{\beta}}$.

[1] O'Brien et al. The epidemiology of nontuberculous mycobacterial diseases in the United States: results from a national survey. *American Review of Respiratory Disease* 135, 5 (1987), 1007–1014.
[2] Gardner et al. Inferring contagion patterns in social contact networks using a maximum likelihood approach. *Natural Hazards Review* 15, 3 (2014), 04014004.

Proof Sketch of Theorem 3

- Follows from our **fine-grained** characterization of $P_{\boldsymbol{\beta}}[\boldsymbol{Y}]$:
- **Lemma 8.** *Every possible history* $\boldsymbol{Y}$ *has*

$$P_{\boldsymbol{\beta}}[\boldsymbol{Y}] = \omega_{\boldsymbol{Y}}\left(\beta^{\mathrm{I}}\right)^{\overbrace{n^{\mathrm{IR}}(\boldsymbol{y}_T)-n^{\mathrm{IR}}(\boldsymbol{y}_0)}^{\#\text{ new infections}}}\left(\beta^{\mathrm{R}}\right)^{\overbrace{n^{\mathrm{R}}(\boldsymbol{y}_T)-n^{\mathrm{R}}(\boldsymbol{y}_0)}^{\#\text{ new recoveries}}}\left(1+\mathrm{O}(\|\boldsymbol{\beta}\|)\right)$$

for some constant number $\omega_{\boldsymbol{Y}} > 0$ *independent of* $\boldsymbol{\beta}$.

➢ <u>Intuition</u> of Lemma 8:

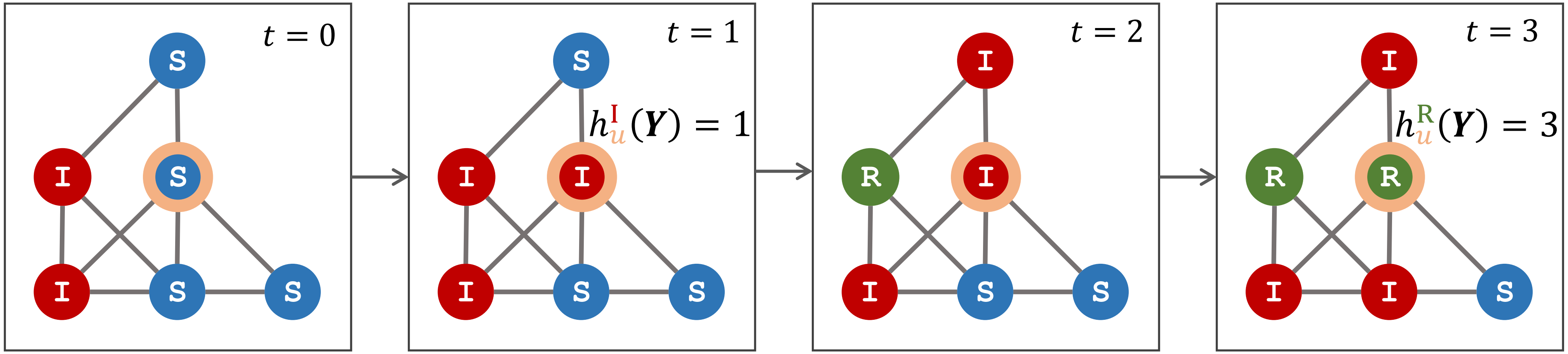
- $\mathrm{S} \to \mathrm{I}: \propto \beta^{\mathrm{I}}\left(1+\mathrm{O}(\|\boldsymbol{\beta}\|)\right);$
- $\mathrm{I} \to \mathrm{R}: \propto \beta^{\mathrm{R}}\left(1+\mathrm{O}(\|\boldsymbol{\beta}\|)\right);$
- $\mathrm{S} \to \mathrm{R}: \propto \beta^{\mathrm{I}}\beta^{\mathrm{R}}\left(1+\mathrm{O}(\|\boldsymbol{\beta}\|)\right);$
- $\mathrm{S} \to \mathrm{S}, \mathrm{I} \to \mathrm{I}, \mathrm{R} \to \mathrm{R}: \propto 1+\mathrm{O}(\|\boldsymbol{\beta}\|).$

Details in the appendix of our paper

Contents

Hitting Times

- ***Hitting times*** for a node u in a history $\boldsymbol{Y}$:
  - First infection time: $h_u^{\mathrm{I}}(\boldsymbol{Y}) \coloneqq \min\{T+1, \min\{t \geq 0: y_{t,u} \geq \mathrm{I}\}\}$.
  - First recovery time: $h_u^{\mathrm{R}}(\boldsymbol{Y}) \coloneqq \min\{T+1, \min\{t \geq 0: y_{t,u} \geq \mathrm{R}\}\}$.



State transition order: S < I < R

Stability of Posterior Expected Hitting Times

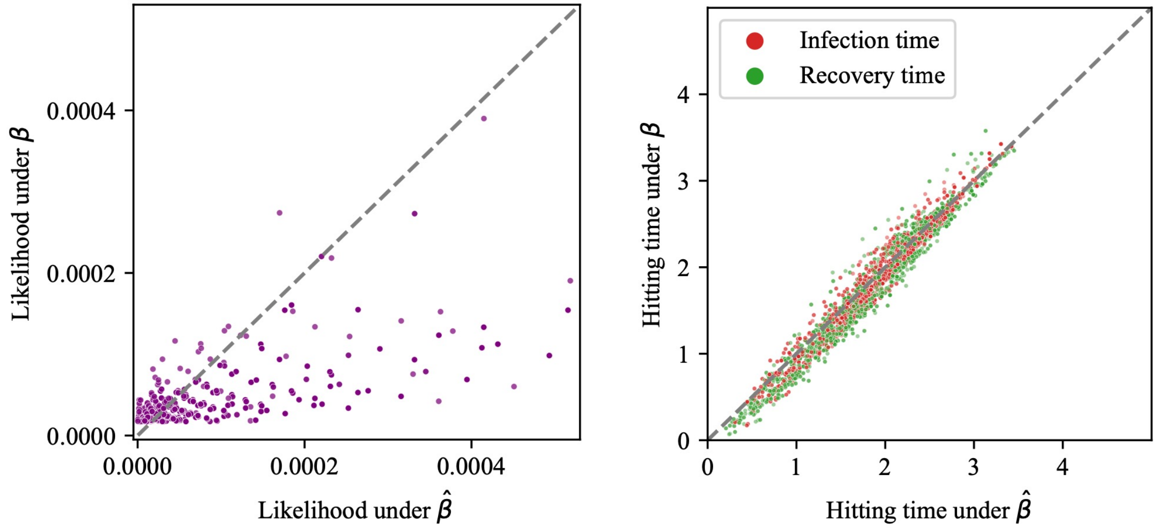
(**Key** theoretical observation)

- **Theorem 4**. *Under SIR model and mild conditions, for any possible snapshot* $\boldsymbol{y}_T$,

$$\nabla_{\boldsymbol{\beta}} \mathop{\mathbb{E}}_{\boldsymbol{Y}\sim P_{\boldsymbol{\beta}}|\boldsymbol{y}_T}[h_u^{\mathrm{I}}(\boldsymbol{Y})] = O(1),$$

$$\nabla_{\boldsymbol{\beta}} \mathop{\mathbb{E}}_{\boldsymbol{Y}\sim P_{\boldsymbol{\beta}}|\boldsymbol{y}_T}[h_u^{\mathrm{R}}(\boldsymbol{Y})] = O(1).$$

➢ **Stable** even for small $\boldsymbol{\beta}$



(a) $P_{\boldsymbol{\beta}}[Y]$ vs $P_{\widehat{\boldsymbol{\beta}}}[Y]$ in the MLE formulation.

(b) $\mathop{\mathbb{E}}_{P_{\boldsymbol{\beta}}|\boldsymbol{y}_T}[h_u^x(Y)]$ vs $\mathop{\mathbb{E}}_{P_{\widehat{\boldsymbol{\beta}}}|\boldsymbol{y}_T}[h_u^x(Y)]$ in the barycenter formulation.

Figure 2: Sensitivity of the MLE formulation vs stability of the barycenter formulation.

❖ Proof Idea: Use Lemma 8 again to characterize $P_{\boldsymbol{\beta}}[\boldsymbol{Y}]$ and $P_{\boldsymbol{\beta}}[\boldsymbol{y}_T]$.

MLE Formulation → Barycenter Formulation

- Recall:
  - Estimation error of $\boldsymbol{\beta}$ is **unavoidable**.

  ✗ The MLE formulation is **sensitive** to estimation error of $\boldsymbol{\beta}$.

  Posterior expected *hitting times* are **stable** against estimation error of $\boldsymbol{\beta}$.

- Our solution:
  - ➢ A novel ***barycenter formulation*** based on *hitting times.*

Barycenter Formulation

- ***History distance*** d**:**

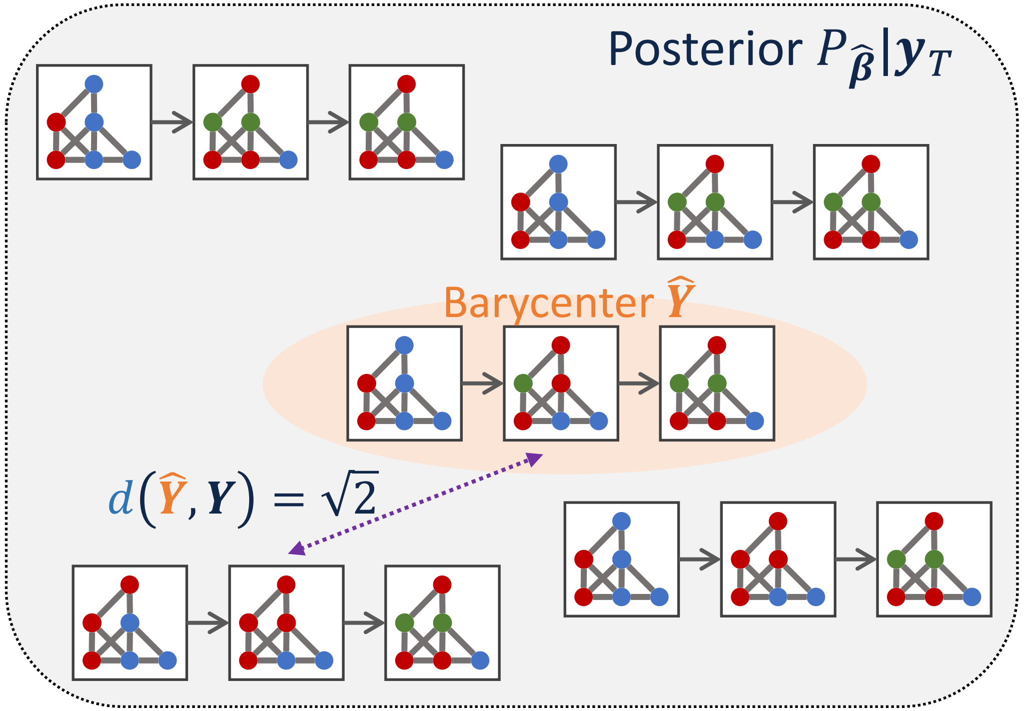
  (Euclidean distance with *hitting times* as coordinates)

$$d(\widehat{\boldsymbol{Y}}, \boldsymbol{Y}) := \sqrt{\sum_{u \in \mathcal{V}} \sum_{x=\mathrm{I,R}} \left(h_u^x(\widehat{\boldsymbol{Y}}) - h_u^x(\boldsymbol{Y})\right)^2}.$$

- ***Barycenter formulation:***

  (Finding the *barycenter* $\widehat{\boldsymbol{Y}}$ of the posterior distribution $P_{\widehat{\boldsymbol{\beta}}} | \boldsymbol{y}_T$ w.r.t. the history distance d)

$$\min_{\widehat{\boldsymbol{Y}}} \mathop{\mathbb{E}}_{\boldsymbol{Y} \sim P_{\widehat{\boldsymbol{\beta}}} | \boldsymbol{y}_T} \left[d(\widehat{\boldsymbol{Y}}, \boldsymbol{Y})^2 \right].$$



Solution to the Barycenter Formulation

- Bias–variance decomposition:

➢ Posterior expected hitting times

$$\mathop{\mathbb{E}}_{\boldsymbol{Y}\sim P_{\widehat{\boldsymbol{\beta}}}|\boldsymbol{y}_T}\left[d\left(\widehat{\boldsymbol{Y}},\boldsymbol{Y}\right)^2\right]=\sum_{u\in\mathcal{V}}\sum_{x=\mathrm{I,R}}\left(\left(h_u^x(\widehat{\boldsymbol{Y}})-\mathop{\mathbb{E}}_{\boldsymbol{Y}\sim P_{\widehat{\boldsymbol{\beta}}}|\boldsymbol{y}_T}[h_u^x(\boldsymbol{Y})]\right)^2+\mathop{\mathrm{Var}}_{\boldsymbol{Y}\sim P_{\widehat{\boldsymbol{\beta}}}|\boldsymbol{y}_T}[h_u^x(\boldsymbol{Y})]\right).$$

- Variances are **constant** w.r.t. $\widehat{\boldsymbol{Y}}$ $\Longrightarrow$ Optimal solution $\widehat{\boldsymbol{Y}}$:

$$h_u^x(\widehat{\boldsymbol{Y}})=\mathrm{round}\left(\mathop{\mathbb{E}}_{\boldsymbol{Y}\sim P_{\widehat{\boldsymbol{\beta}}}|\boldsymbol{y}_T}[h_u^x(\boldsymbol{Y})]\right),\qquad x=\mathrm{I,R};$$

$$\hat{y}_{t,u}=\begin{cases}\mathrm{S}, & \text{for } 0\le t<h_u^{\mathrm{I}}(\widehat{\boldsymbol{Y}});\\ \mathrm{I}, & \text{for } h_u^{\mathrm{I}}(\widehat{\boldsymbol{Y}})\le t<h_u^{\mathrm{R}}(\widehat{\boldsymbol{Y}});\\ \mathrm{R}, & \text{for } h_u^{\mathrm{R}}(\widehat{\boldsymbol{Y}})\le t\le T.\end{cases}$$

➢Now our problem **reduces** to estimating $\mathop{\mathbb{E}}_{\boldsymbol{Y}\sim P_{\widehat{\boldsymbol{\beta}}}|\boldsymbol{y}_T}[h_u^x(\boldsymbol{Y})]$.

M–H MCMC for Posterior Expectation Estimation

- How to estimate $\mathbb{E}_{\boldsymbol{Y}\sim P_{\widehat{\boldsymbol{\beta}}}|\boldsymbol{y}_T}[h_u^x(\boldsymbol{Y})]$? ➢Recall: **Intractable** to compute $P_{\widehat{\boldsymbol{\beta}}}[\boldsymbol{Y}|\boldsymbol{y}_T]$.
- Our solution: ***M–H MCMC*** [1, 2].
  1. Design a ***proposal*** distribution $Q_{\boldsymbol{\theta}}(\boldsymbol{y}_T)[\cdot]$ over **possible** histories.
  2. Each step of M–H MCMC samples L histories $\boldsymbol{Y}^{(s,i)} \sim Q_{\boldsymbol{\theta}}(\boldsymbol{y}_T)$, $i = 1, \dots, L$.
  3. Each previous history $\boldsymbol{Y}^{(s-1,i)}$ is replaced by the new history $\boldsymbol{Y}^{(s,i)}$ with probability:

$$\min\left\{1, \frac{P_{\widehat{\boldsymbol{\beta}}}\left[\boldsymbol{Y}^{(s,i)}\middle|\boldsymbol{y}_T\right]Q_{\boldsymbol{\theta}}(\boldsymbol{y}_T)\left[\boldsymbol{Y}^{(s-1,i)}\right]}{P_{\widehat{\boldsymbol{\beta}}}\left[\boldsymbol{Y}^{(s-1,i)}\middle|\boldsymbol{y}_T\right]Q_{\boldsymbol{\theta}}(\boldsymbol{y}_T)\left[\boldsymbol{Y}^{(s,i)}\right]}\right\} = \min\left\{1, \frac{P_{\widehat{\boldsymbol{\beta}}}\left[\boldsymbol{Y}^{(s,i)}\right]Q_{\boldsymbol{\theta}}(\boldsymbol{y}_T)\left[\boldsymbol{Y}^{(s-1,i)}\right]}{P_{\widehat{\boldsymbol{\beta}}}\left[\boldsymbol{Y}^{(s-1,i)}\right]Q_{\boldsymbol{\theta}}(\boldsymbol{y}_T)\left[\boldsymbol{Y}^{(s,i)}\right]}\right\}.$$

➢**Tractable** to compute

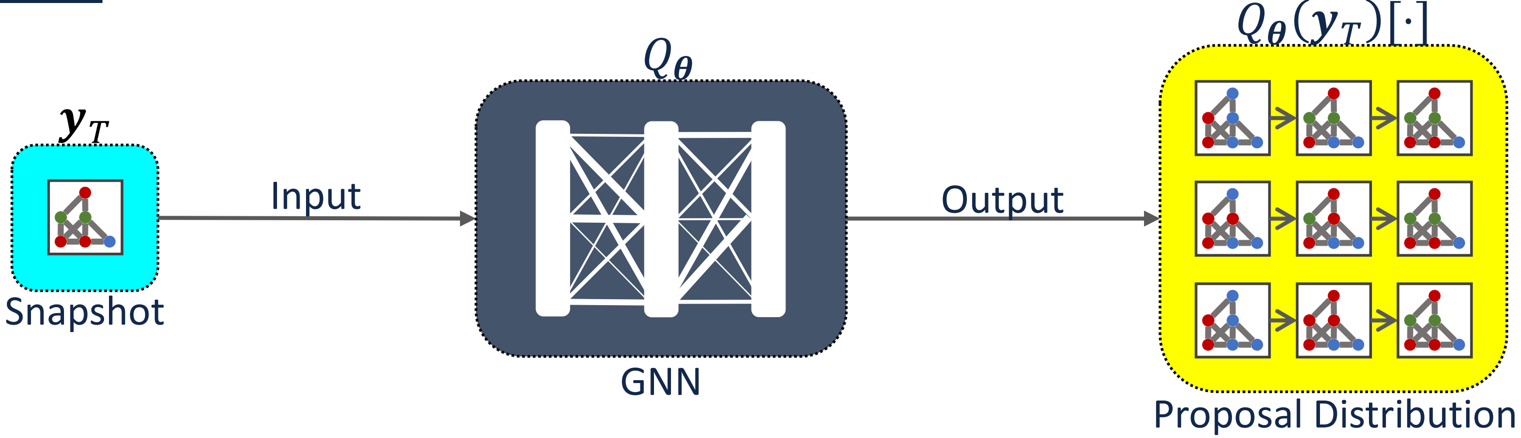
➢The Markov chain $\langle \boldsymbol{Y}^{(s,i)} \rangle$ **provably converges** to the posterior distribution $P_{\widehat{\boldsymbol{\beta}}}|\boldsymbol{y}_T$ [2].

$$\mathbb{E}_{\boldsymbol{Y}\sim P_{\widehat{\boldsymbol{\beta}}}|\boldsymbol{y}_T}[h_u^x(\boldsymbol{Y})] \approx \frac{1}{L}\sum_{i=1}^{L} h_u^x\left(\boldsymbol{Y}^{(s,i)}\right), \qquad s \to +\infty.$$

[1] Metropolis et al. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* 21, 6 (1953), 1087–1092.
[2] Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 1 (1970), 97–109.

Learning an Optimal Proposal for M–H MCMC

- The ***convergence rate*** of M–H MCMC depends critically on the proposal $Q_{\boldsymbol{\theta}}$.
  - ➢The proposal $Q_{\boldsymbol{\theta}}(\boldsymbol{y}_T)$ **closer** to $P_{\hat{\boldsymbol{\beta}}}|\boldsymbol{y}_T \Longrightarrow$ **Higher** rate of convergence [1].
- Our solution: Use a GNN to learn an **optimal** proposal.



- We want $Q_{\boldsymbol{\theta}}(\boldsymbol{y}_T)$ to approximate $P_{\hat{\boldsymbol{\beta}}}|\boldsymbol{y}_T$ well $\Longrightarrow$ Objective function:

$$\min_{\boldsymbol{\theta}} \mathop{\mathbb{E}}_{\boldsymbol{Y}\sim P_{\hat{\boldsymbol{\beta}}}} \left[\left(\log Q_{\boldsymbol{\theta}}(\boldsymbol{y}_T)[\boldsymbol{Y}] - \log P_{\hat{\boldsymbol{\beta}}}[\boldsymbol{Y}|\boldsymbol{y}_T]\right)^2\right]. \qquad (*)$$

[1] Mengersen & Tweedie. Rates of convergence of the Hastings and Metropolis algorithms. *The Annals of Statistics* 24, 1 (1996), 101–121.

Equivalent Objective for the Proposal GNN

- Vanilla objective function:

$$\min_{\boldsymbol{\theta}} \mathbb{E}_{\boldsymbol{Y} \sim P_{\widehat{\boldsymbol{\beta}}}} \left[\left(\log Q_{\boldsymbol{\theta}}(\boldsymbol{y}_T)[\boldsymbol{Y}] - \log P_{\widehat{\boldsymbol{\beta}}}[\boldsymbol{Y} | \boldsymbol{y}_T] \right)^2 \right]. \qquad (*)$$

❖Recall: **Intractable** to compute $P_{\widehat{\boldsymbol{\beta}}}[\boldsymbol{Y} | \boldsymbol{y}_T]$. Solution?

- **Theorem 5** (Equivalent objective). *Under mild conditions, for any* **strictly convex** *function* $\psi: \mathbb{R}_+ \to \mathbb{R}$, *the vanilla objective Eq.* $(*)$ *is equivalent to*

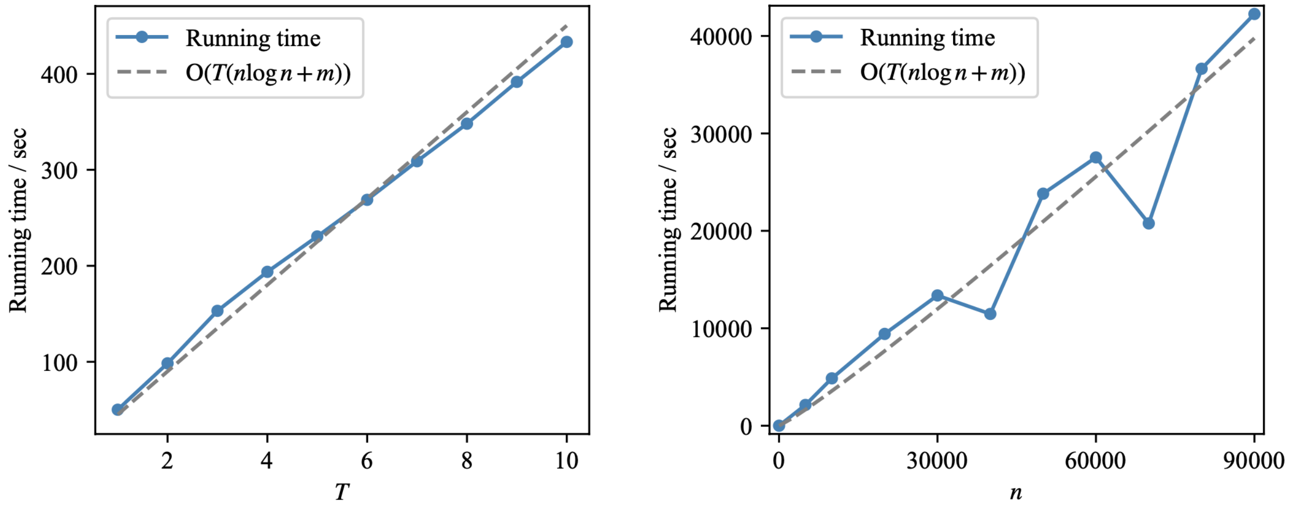
$$\min_{\boldsymbol{\theta}} \mathbb{E}_{\boldsymbol{Y} \sim P_{\widehat{\boldsymbol{\beta}}}} \left[\psi \left(\frac{Q_{\boldsymbol{\theta}}(\boldsymbol{y}_T)[\boldsymbol{Y}]}{P_{\widehat{\boldsymbol{\beta}}}[\boldsymbol{Y}]} \right) \right]. \qquad (**)$$

➢Implication: **Intractable** $P_{\widehat{\boldsymbol{\beta}}}[\boldsymbol{Y} | \boldsymbol{y}_T]$ in Eq. $(*)$ → **Tractable** $P_{\widehat{\boldsymbol{\beta}}}[\boldsymbol{Y}]$ in Eq. $(**)$.

- In this work, we use $\psi(z) \coloneqq -\log z$, and this objective Eq. $(**)$ instantiates as

$$\min_{\boldsymbol{\theta}} \mathbb{E}_{\boldsymbol{Y} \sim P_{\widehat{\boldsymbol{\beta}}}} \left[\log P_{\widehat{\boldsymbol{\beta}}}[\boldsymbol{Y}] - \log Q_{\boldsymbol{\theta}}(\boldsymbol{y}_T)[\boldsymbol{Y}] \right] \Leftrightarrow \min_{\boldsymbol{\theta}} \mathbb{E}_{\boldsymbol{Y} \sim P_{\widehat{\boldsymbol{\beta}}}} \left[-\log Q_{\boldsymbol{\theta}}(\boldsymbol{y}_T)[\boldsymbol{Y}] \right].$$

Time Complexity



(a) Running time vs T.

(b) Running time vs n.

Figure 3: Running time (training time + testing time).

- Overall time complexity: $O\big(T(n\log n + m)\big)$.
  - ➢ **Nearly linear** w.r.t. the output size $\Theta(Tn)$.

Contents

Experimental Setting

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

- Datasets:
  - Synthetic graph + synthetic diffusion
  - Real graph + synthetic diffusion
  - Real graph w/ real diffusion ➢ Not exactly SI/SIR; unknown true $\boldsymbol{\beta}$
- Baselines:
  - Supervised methods: GCN [1], GIN [2], BRITS [3], GRIN [4], SPIN [5]
  - MLE-based methods: DHREC [6], CRI [7]
- Evaluation metrics:
  - F-1 score of node states
  - Normalized rooted mean squared error (NRMSE) of hitting times

[1] Kipf & Welling. Semi-supervised classification with graph convolutional networks. *International Conference on Learning Representations* (2017).
[2] Xu et al. How powerful are graph neural networks? *International Conference on Learning Representations* (2019).
[3] Cao et al. BRITS: Bidirectional recurrent imputation for time series. *Advances in Neural Information Processing Systems 31* (2018).
[4] Cini et al. Filling the G_ap_s: Multivariate time series imputation by graph neural networks. *International Conference on Learning Representations* (2022).
[5] Marisca et al. Learning to reconstruct missing data from spatiotemporal graphs with sparse observations. *Advances in Neural Information Processing Systems 35* (2022).
[6] Sefer & Kingsford. Diffusion archeology for diffusion progression history reconstruction. *Knowledge and Information Systems* 49, 2 (2016), 403–427.
[7] Chen et al. Detecting multiple information sources in networks under the SIR model. *IEEE Transactions on Network Science and Engineering* 3, 1 (2016), 17–31.

Performance for Real-World Diffusion

- *BrFarmers* is very close to SI [1].

Table 4: Results for real-world diffusion. "OOM" indicates "out of memory."

Type	Method	BrFarmers F1↑	BrFarmers NRMSE↓	Pol F1↑	Pol NRMSE↓	Covid F1↑	Covid NRMSE↓	Hebrew F1↑	Hebrew NRMSE↓
Supervised (w/ estimated $\widehat{\boldsymbol{\beta}}$)	GCN	.5409	.6660	.4458	.4946	.3162	.5214	.3350	.6070
	GIN	.4548	.6565	.5203	.4767	.3226	.4951	.3704	.7816
	BRITS	.5207	.3995	OOM		.3524	.5333	.3120	.6584
	GRIN	.8003	.2425	.6518	.3731	.5448	.3040	.5916	**.2212**
	SPIN	**.8268**	**.2084**	OOM		.5917	.2932	.5178	.3330
MLE	DHREC	.6131	.4150	.7023	.3398	.3540	.6023	.6251	.4169
	CRI	.6058	.4444	.7468	.2942	.4170	.5487	.5344	.3552
Barycenter	DITTO (ours)	.8206	.2142	**.7471**	**.2903**	**.6240**	**.2637**	**.6411**	.2983

DITTO: Consistently strong performance across all datasets.

MLE/Supervised: Bad when real diffusion deviates from SI/SIR.

[1] Valente. 1995. *Network Models of the Diffusion of Innovations* (2nd Edition). Hampton Press, Cresskill, NJ.

Comparison with MLE-Based Methods

Table 5: Comparison with MLE-based methods on synthetic SI and SIR diffusion. *We use GRIN trained with true β as the ideal performance and calculate *Gap* w.r.t. this ideal performance.

Type	Method	BA-SI F1↑	BA-SI Gap↓	BA-SI NRMSE↓	BA-SI Gap↓	ER-SI F1↑	ER-SI Gap↓	ER-SI NRMSE↓	ER-SI Gap↓	Oregon2-SI F1↑	Oregon2-SI Gap↓	Oregon2-SI NRMSE↓	Oregon2-SI Gap↓	Prost-SI F1↑	Prost-SI Gap↓	Prost-SI NRMSE↓	Prost-SI Gap↓
Ideal	GRIN	.8404*	—	.2123*	—	.8317*	—	.2166*	—	.8320*	—	.2249*	—	.8482*	—	.2155*	—
MLE	DHREC	.6026	28.30%	.4644	118.75%	.6281	24.48%	.4495	107.53%	.6038	27.43%	.4101	82.35%	.6558	22.68%	.4138	92.02%
	CRI	.7502	10.73%	.3012	41.87%	.7797	6.25%	.2744	26.69%	.8183	1.65%	.2438	8.40%	.8083	4.70%	.2491	15.59%
Barycenter	DITTO (ours)	**.8384**	**0.24%**	**.2139**	**0.75%**	**.8269**	**0.58%**	**.2225**	**2.72%**	**.8280**	**0.48%**	**.2289**	**1.78%**	**.8327**	**1.83%**	**.2317**	**7.52%**

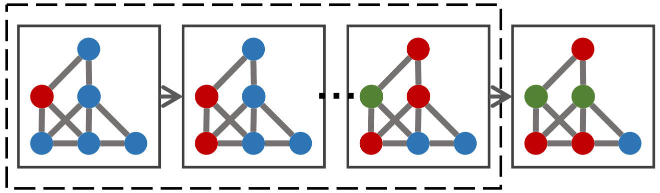
Type	Method	BA-SIR F1↑	BA-SIR Gap↓	BA-SIR NRMSE↓	BA-SIR Gap↓	ER-SIR F1↑	ER-SIR Gap↓	ER-SIR NRMSE↓	ER-SIR Gap↓	Oregon2-SIR F1↑	Oregon2-SIR Gap↓	Oregon2-SIR NRMSE↓	Oregon2-SIR Gap↓	Prost-SIR F1↑	Prost-SIR Gap↓	Prost-SIR NRMSE↓	Prost-SIR Gap↓
Ideal	GRIN	.7867*	—	.1692*	—	.7626*	—	.2484*	—	.8024*	—	.1651*	—	.8067*	—	.1652*	—
MLE	DHREC	.5080	35.43%	.4722	179.08%	.5500	27.88%	.4423	78.06%	.6044	24.68%	.4478	171.23%	.6268	22.30%	.4326	161.86%
	CRI	.5994	23.81%	.3356	98.35%	.6129	19.63%	.3109	25.16%	.5761	28.20%	.3576	116.60%	.5738	28.87%	.3406	106.17%
Barycenter	DITTO (ours)	**.7783**	**1.07%**	**.1633**	**−3.49%**	**.7734**	**−1.42%**	**.1679**	**−32.41%**	**.7928**	**1.20%**	**.1707**	**3.39%**	**.7929**	**1.71%**	**.1690**	**2.30%**

DITTO: Stably achieves the strongest performance.

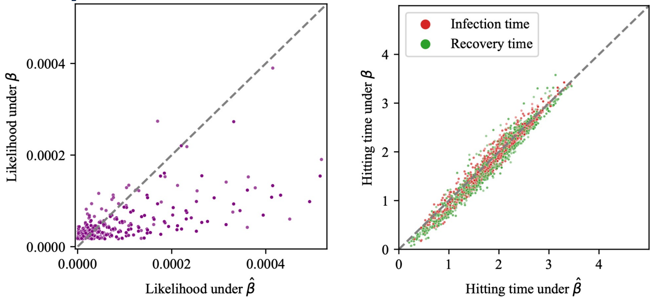
MLE: Performance varies largely across datasets due to **sensitivity**.

Contents

Conclusion

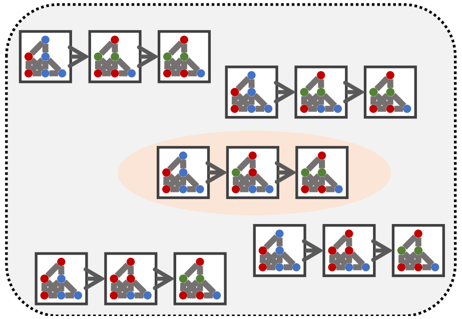


- PROBLEM: Reconstructing Diffusion history from A single SnapsHot (**DASH**).



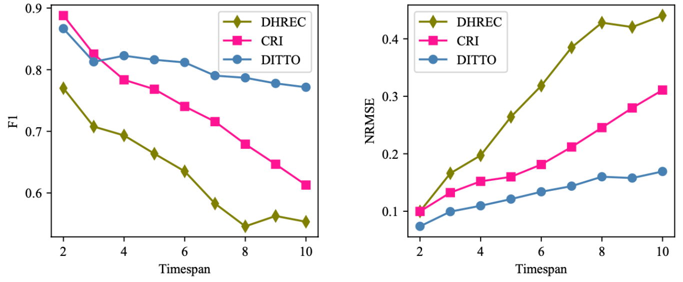
- THEORETICAL INSIGHTS: Fundamental limitation of the MLE formulation.
  - ➢ Estimation error of $\boldsymbol{\beta}$ is **unavoidable**.
  - ➢ The MLE formulation is **sensitive** to estimation error of $\boldsymbol{\beta}$.

- NOVEL FORMULATION: ***Barycenter formulation*** with provable **stability**.



- PROPOSED METHOD: *DIffusion hiTting Times with Optimal proposal* (**DITTO**).
  - ➢ Reducing DASH to estimating *posterior expected hitting times* via M–H MCMC;
  - ➢ Using a GNN to learn an **optimal** proposal to accelerate convergence of M–H MCMC.

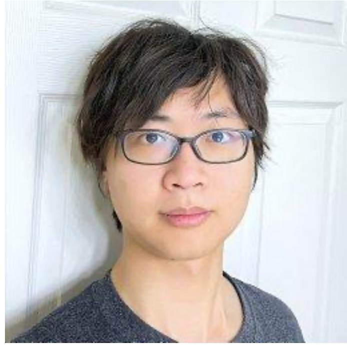
- EXPERIMENTAL RESULTS:
  - ➢ Outperforms both **supervised** and **MLE-based** methods.
  - ➢ Strong performance for both **synthetic** and **real-world** diffusion.



(a) F1 vs timespan T. **(b) NRMSE vs timespan T.**

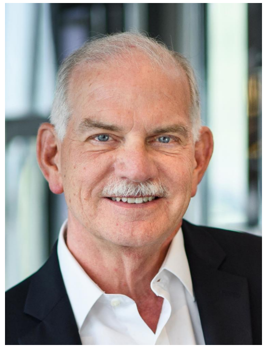
Thanks for attending

Reconstructing Graph Diffusion History from a Single Snapshot


Ruizhong Qiu
UIUC
(Presenter)


Dingsu Wang
UIUC


Lei Ying
UMich


H. Vincent Poor
Princeton


Yifang Zhang
C3.ai DTI


Hanghang Tong
UIUC

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN
UNIVERSITY OF MICHIGAN
PRINCETON UNIVERSITY
C3.ai Digital Transformation Institute

ACKNOWLEDGEMENTS

NSF amazon DARPA USDA NIFA DEVCOM ARMY RESEARCH LABORATORY C3.ai Digital Transformation Institute IBM