Connecting Models with Data for Coupled Online–Offline Social Systems

@ 2026 SIAM Annual Meeting

Moyi Tian
Department of Applied Mathematics
University of Colorado Boulder

July 9, 2026

Applied Math
UNIVERSITY OF COLORADO **BOULDER**

- **Fact:**

  [Online-offline spillovers] Social media is transforming how, when, and where conflicts occur.

  About 68.7% of the global population uses social media (Digital 2026 Global Overview Report by DataReportal).

  New online connections —> 7% increase in physical activity (Althoff *et al.* 2017).

  Online communication and on-site protests exhibit bidirectional influence, as illustrated by the worldwide Occupy movement in 2011 (Bastos *et al.* 2015).

- **Goal:** Understand how online and offline activities influence one another.

Two approaches: **Model-Driven** & **Data-Driven**

Modeling Coupled Online—Offline Dynamics

U: Uninterested

E: Engaged

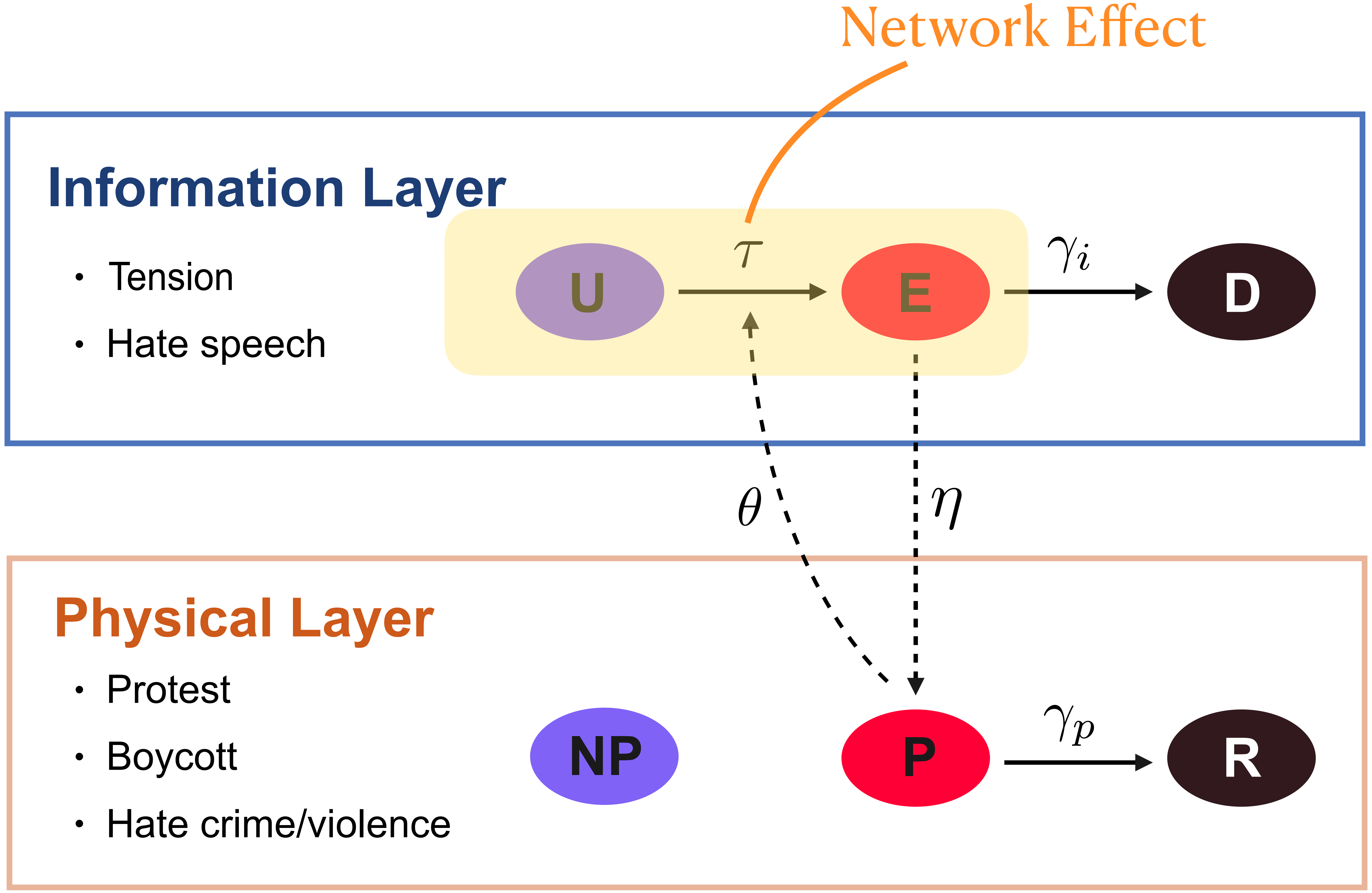
D: Disengaged

τ : transmission rate
γ_i : recovery rate
η : transmission rate
γ_p : recovery rate
θ : self-excitement

NP: Non-participating

P: Participating

R: Recovered from Participation



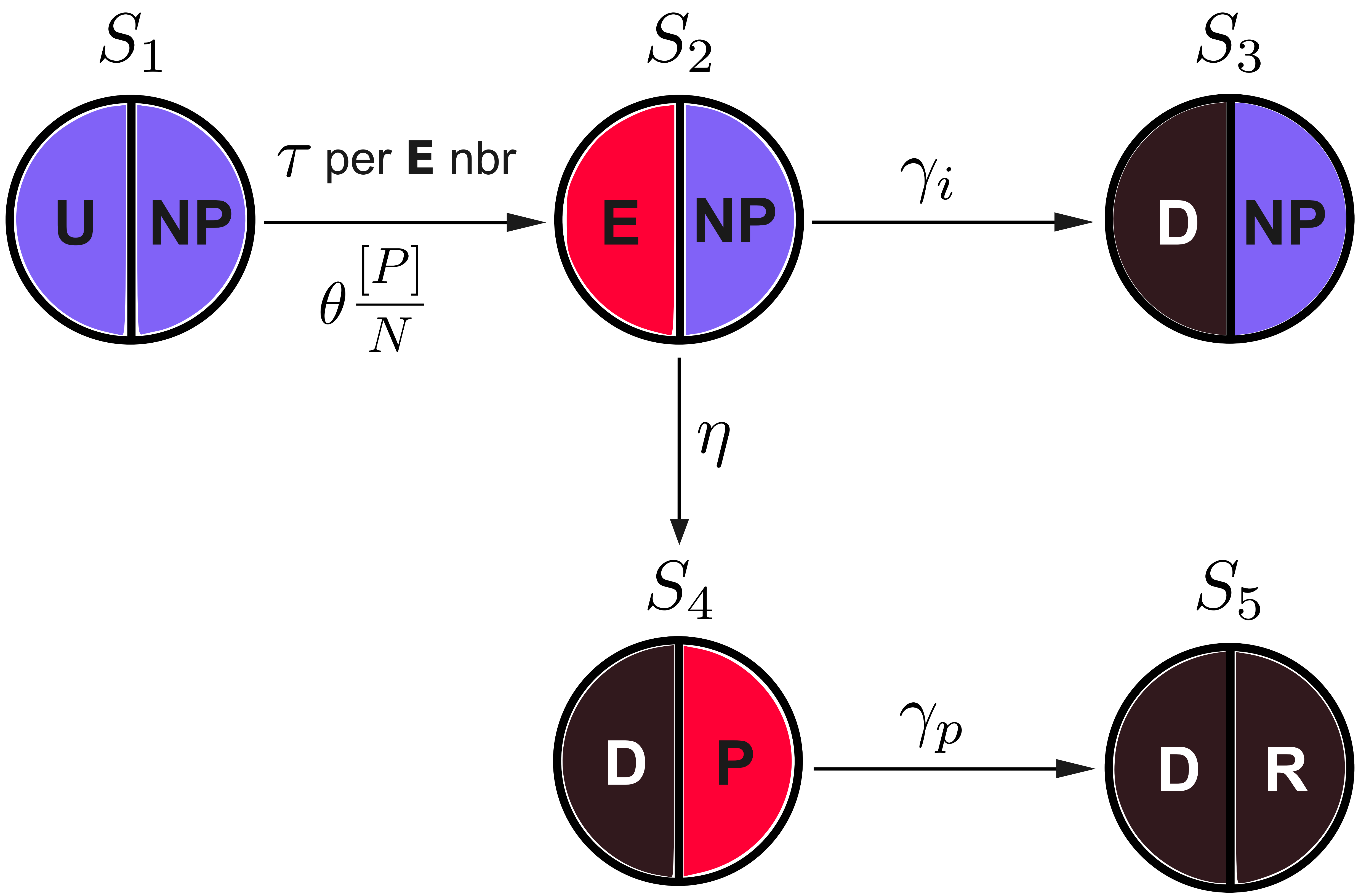
- Multi-layer network compartmental model incorporating the coupled effects of online engagement and offline protests

[Tian, Brantingham & Rodríduez, J. Complex Netw. (2026)]

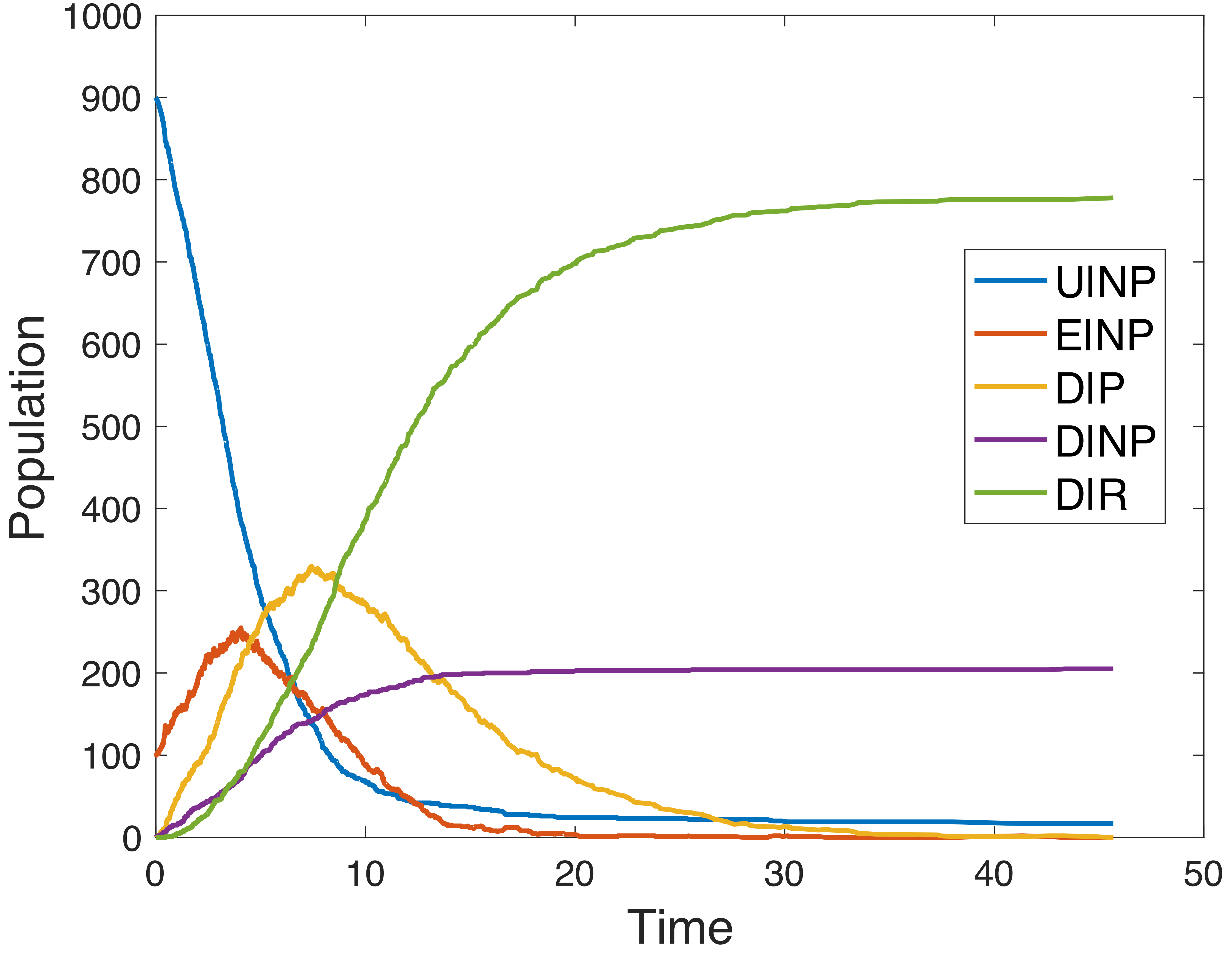
- Continuous-Time Markov chain

N: Total # of nodes

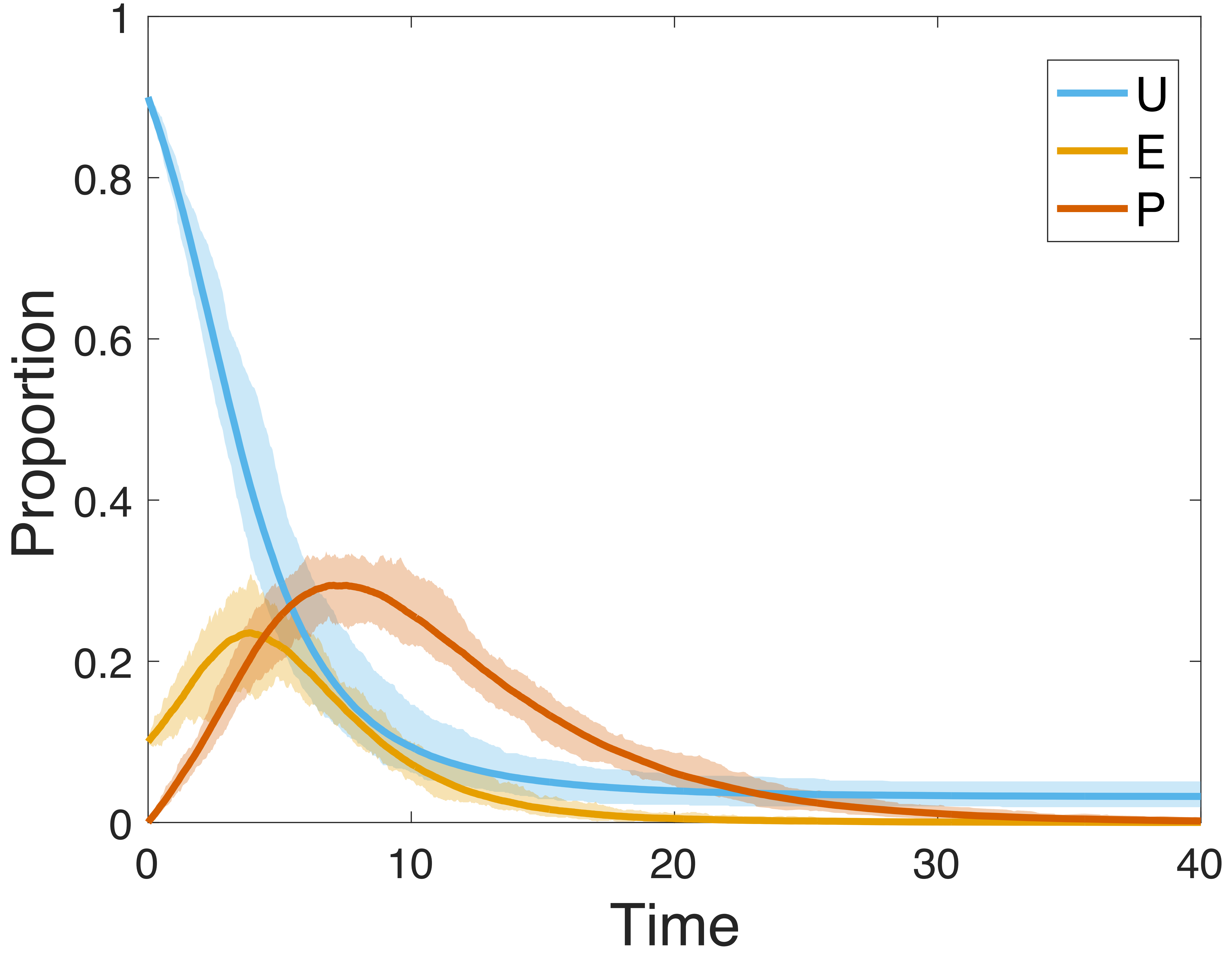
[X]: Expected # of individuals in state X



Single simulation:



Mean aggregated trajectories from multiple simulations (in ratios):



- Markovian Stochastic Process —> ODE Model

[XY]: Expected # of Y neighbors of all Xs

Online

$$\begin{cases} \dfrac{d[U](t)}{dt} = -\tau[UE] - \dfrac{\theta}{N}[U][P] \\ \dfrac{d[E](t)}{dt} = \tau[UE] + \dfrac{\theta}{N}[U][P] - (\eta + \gamma_i)[E] \\ \dfrac{d[D](t)}{dt} = (\eta + \gamma_i)[E] \end{cases}$$

Offline

$$\begin{cases} \dfrac{d[P](t)}{dt} = \eta[E] - \gamma_p[P] \\ \dfrac{d[R](t)}{dt} = \gamma_p[P] \end{cases}$$

Need to approximate the [UE] term to close the system

-> **mean-field approximation** that depends on network structures

- Mean-field models can be derived under
  - Different **assumptions on network structures** (fully-connected? Homogeneous/heterogeneous degree?...)
  - & Different **levels of approximation accuracy** (node-level, pairwise, triplet, with the truncation chosen to balance accuracy and tractability)

For Example:

- If assuming **Fully-Connected Network**: An average engaged individual has approximately $[U]/N$ uninterested neighbors —> approximate [UE] by $\frac{[U]}{N}[E]$
- If assuming **Homogeneous Network** (k-regular graph):
  - **Single-Level (Mean-Field) Approximation**: Pair counts are approximated using singleton counts.

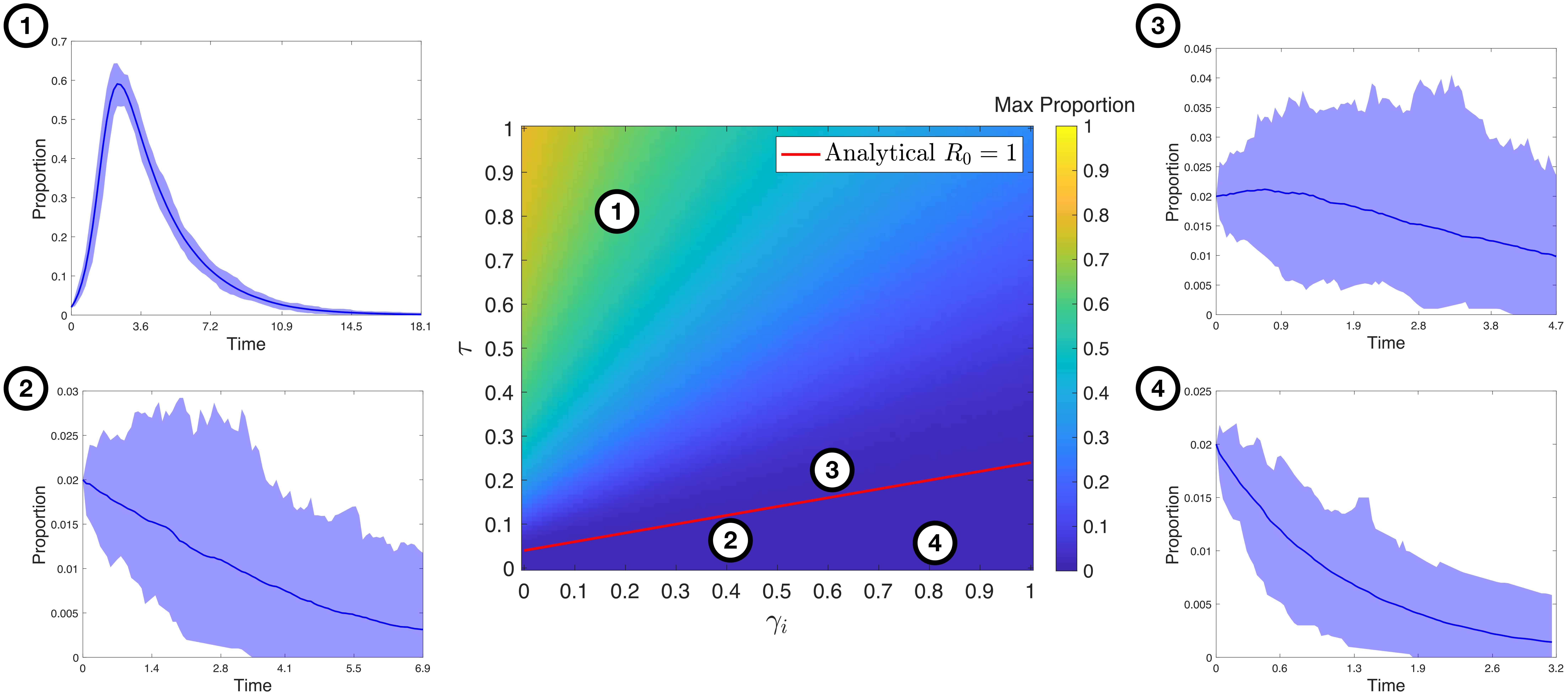
    Assume that engaged individuals are distributed randomly, so $[UE] \approx \frac{k}{N}[U][E]$
  - OR **Pairwise Approximation**: We explicitly track the evolution of pair counts, which depends on triplet counts, and close the system by approximating triplets in terms of pairs.

    Key idea: The probability of choosing X&Z to be neighbors of Y is approximately $\frac{[XY][YZ]}{k^2[Y]}$. There are $k(k-1)$ ways to choose 2 neighbors.
- If assuming **Heterogeneous Network**: Similar approach, but nodes are differentiated by degree, and state counts are tracked within each degree class

[Kiss, Miller & Simon, Mathematics of Epidemics on Networks: From Exact to Approximate Models, Springer (2017)]

Application: Reproductive Number R_0

The average number of newly engaged individuals motivated by a single engaged person in the pre-engaged online population



For Online Engagement E

* R_0 calculated from the stochastic process

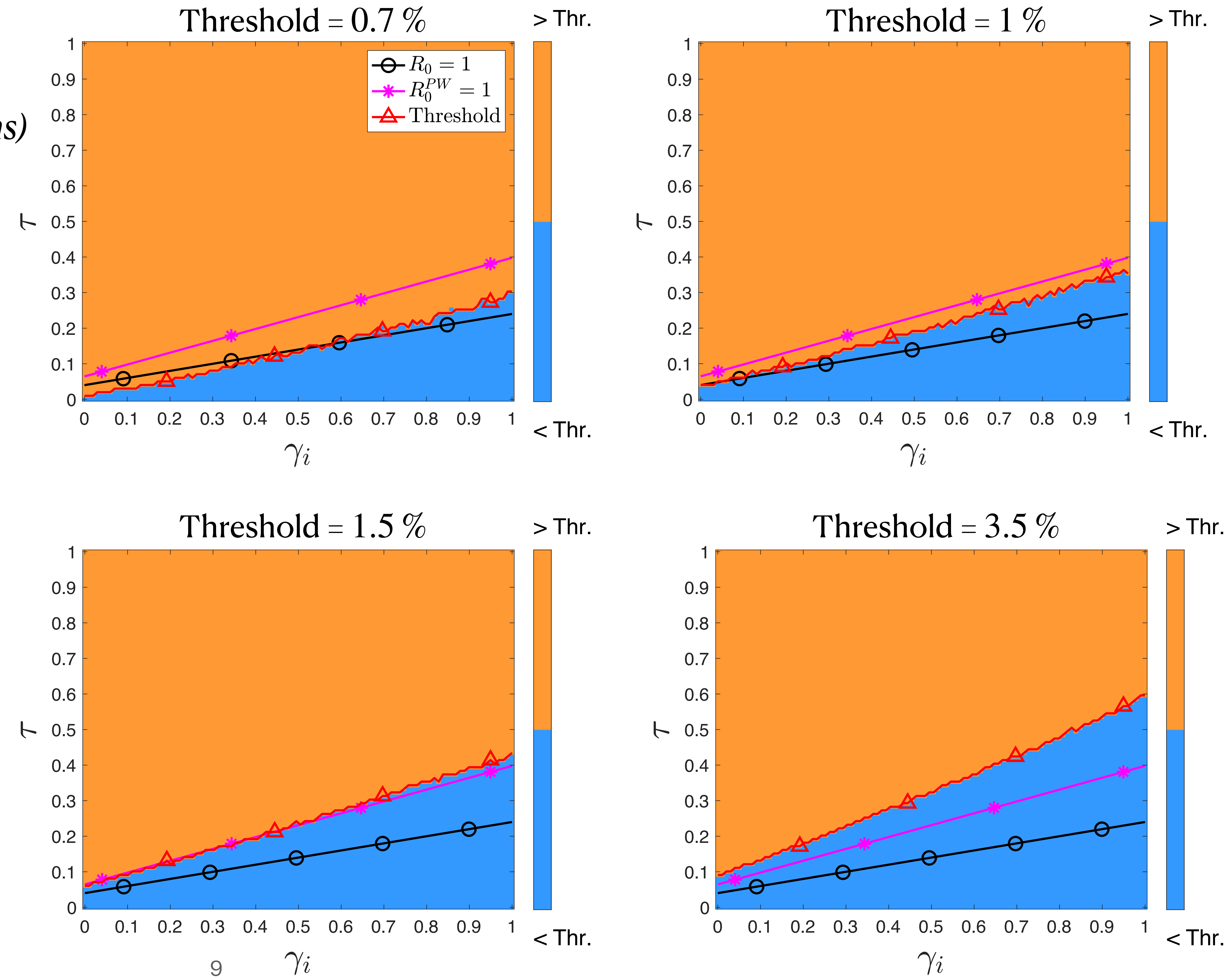
Categorization of Offline Outbursts & Comparison

* R_0^{PW} from the pairwise approximation model
(more accurate, as it incorporates inter-layer interactions)

- Categorization of offline outbursts based on threshold for the max proportion

- Protests exceeding 3.5% active participation at peak activity have historically tended to succeed (Chenoweth, 2020).

- The $R_0 = 1$ threshold is more conservative than $R_0^{PW} = 1$; however, both appear overly conservative relative to empirical observations.



- **Goal:** Understand how online and offline activities influence one another.

Two approaches: **Model-Driven** & **Data-Driven**

Two Attempts:

⇨ System Identification by Weak Form SINDy

⇨ Model Inference from Multi-Agent LLM Dynamics

System Identification by Weak Form SINDy

SINDy

Sparse Identification of Nonlinear Dynamics

Method for learning governing equations $\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}(t))$ from observed data observed at timepoints $t_1, t_2, \ldots, t_M$

$$\mathbf{X} = [\mathbf{x}(t_1) + \epsilon_1 \quad \mathbf{x}(t_2) + \epsilon_2 \quad \ldots \quad \mathbf{x}(t_M) + \epsilon_M]^T$$

using sparse regression on

$$\dot{\mathbf{X}} = \Theta(\mathbf{X})\,\mathbf{w}$$

where $\Theta(\mathbf{X}) = [f_1(\mathbf{x}) \quad f_2(\mathbf{x}) \ \cdots\ ]$ are candidate functions in the chosen library.

Solving a sparse $\mathbf{w}^*$ by minimizing $\|\dot{\mathbf{X}} - \Theta(\mathbf{X})\,\mathbf{w}\|_2$ —> Discover a model with a few terms

[Brunton, Proctor & Kutz, Proc. Natl. Acad. Sci (2016)]

WSINDy

Weak form SINDy

To avoid approximation for derivatives sensitive to noise, we can solve the generalized least squares problem $G\mathbf{w} = b$ in weak form where $G \in \mathbb{R}^{L\times J}$ and $b \in \mathbb{R}^{L\times d}$ are

$$G_{l,j} = \langle \phi_l, f_j(\mathbf{X})\rangle \quad \text{and} \quad b_{l,i} = -\langle \phi_l', \mathbf{X}_i\rangle.$$

$\{\phi_l\}_{l=1,\ldots,L}$ is a family of smooth compactly supported test functions.

When using multiple trajectory data, we solve:

$$\begin{bmatrix} G^1 \\ G^2 \\ \vdots \\ G^K \end{bmatrix} \mathbf{w} = \begin{bmatrix} b^1 \\ b^2 \\ \vdots \\ b^K \end{bmatrix}$$

where K is the number of trajectories included in the learning.

* GitHub page for WSINDy group: https://github.com/MathBioCU

[Messenger & Bortz, Multiscale Model. Simul. (2021)]

Results

* On Erdős–Rényi with $N = 1000$, $k = 5$

May suggest a better approximation model !

(particularly on networks with low average degree when mean-field performs poorly)

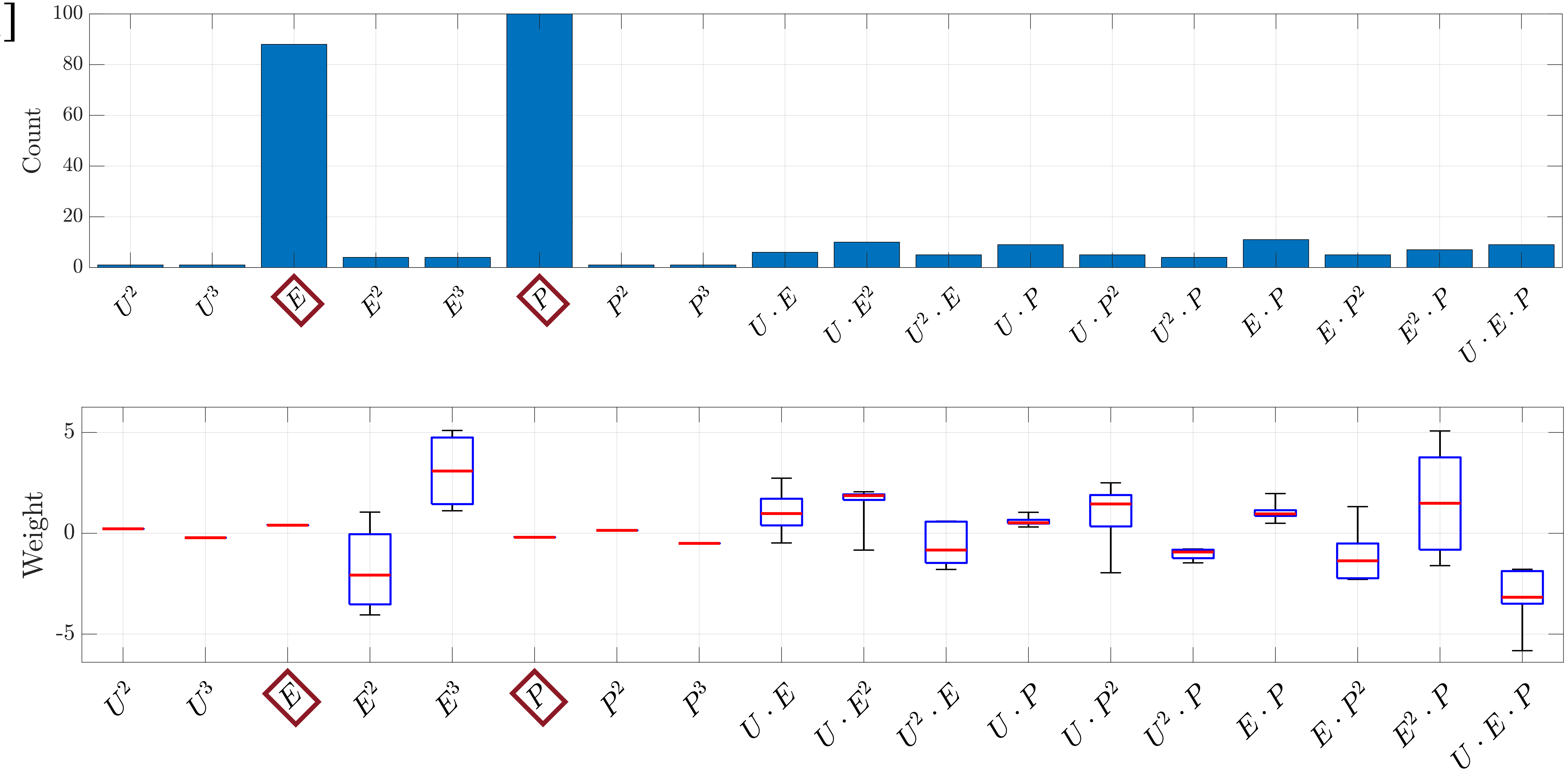
* GitHub: https://github.com/Moyi-Tian/WSINDy-NetworkDynamics

[Tian, Messenger, Dukic, Rodríduez & Bortz, arXiv (2026)]

A Bit More on the Learned Models

Terms & Weight learned across 100 stochastic data

[P Equation]

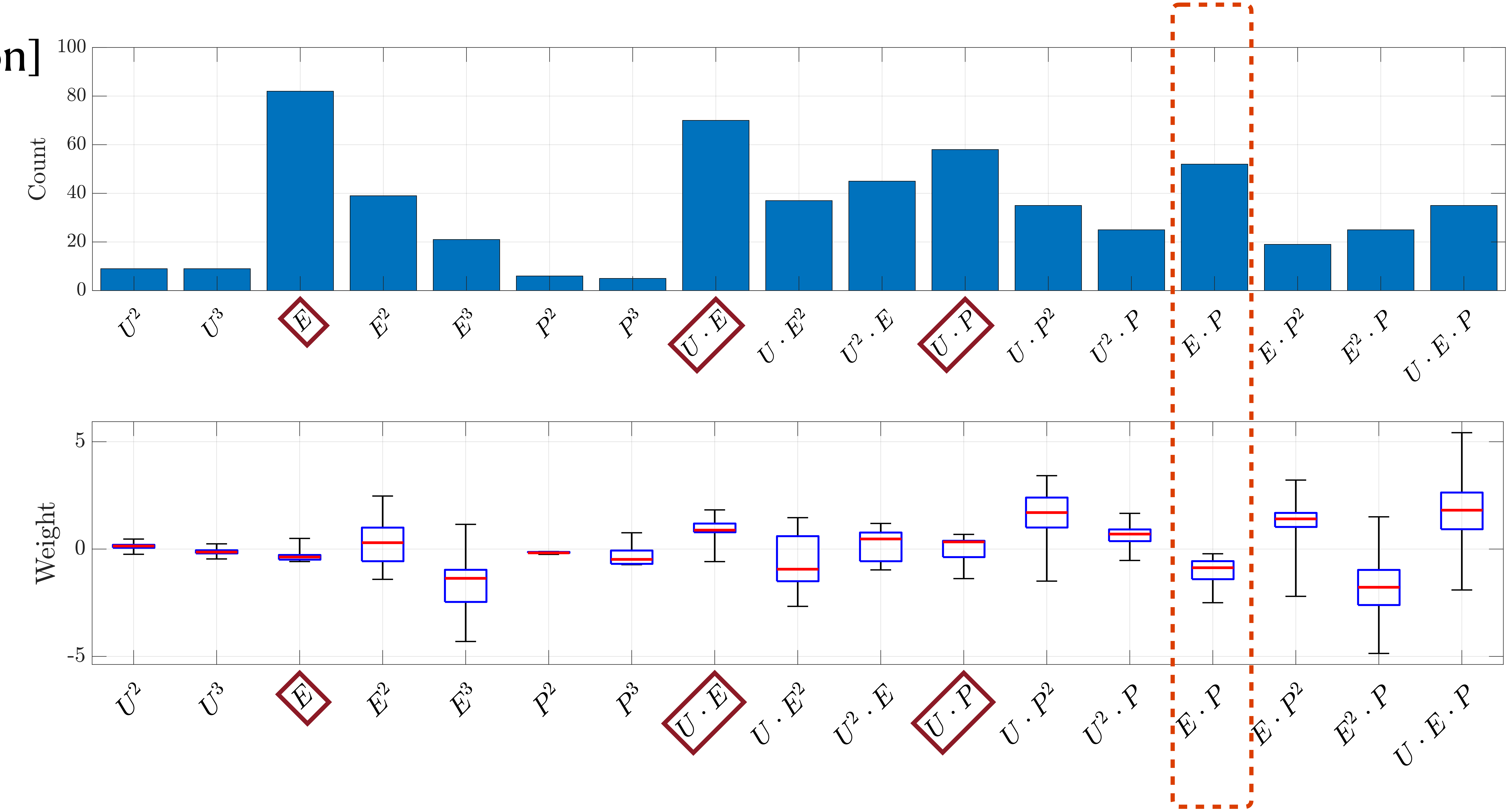


These are right on!

* Terms boxed in red are present in the mean-field

Highest-frequency non-mean-field candidate term from WSINDy

[E Equation]



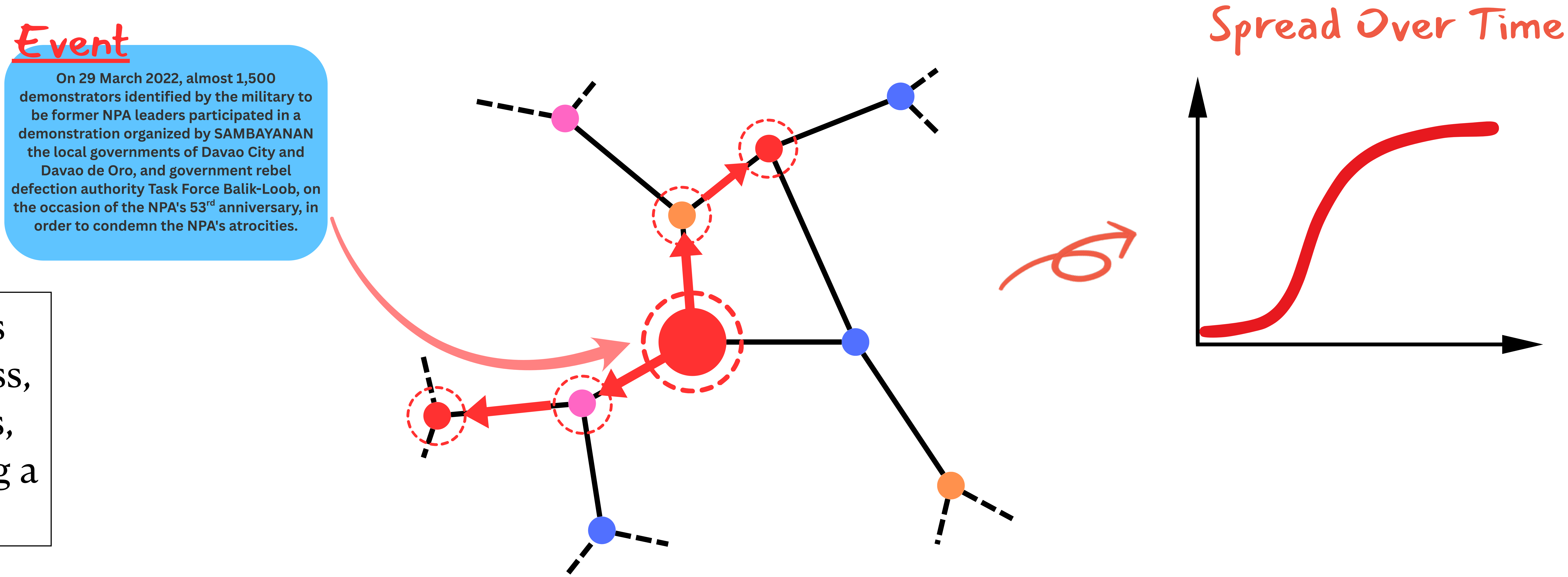
* Terms boxed in red are present in the mean-field

Model Inference from Multi-Agent LLM Dynamics

- Motivation:

More than half of the Internet traffic in 2023 was generated by bots on social media (Ng & Carley, Scientific Reports 2025).

➡ Bots are highly active online, occupying much of the traffic and exerting significant influence.

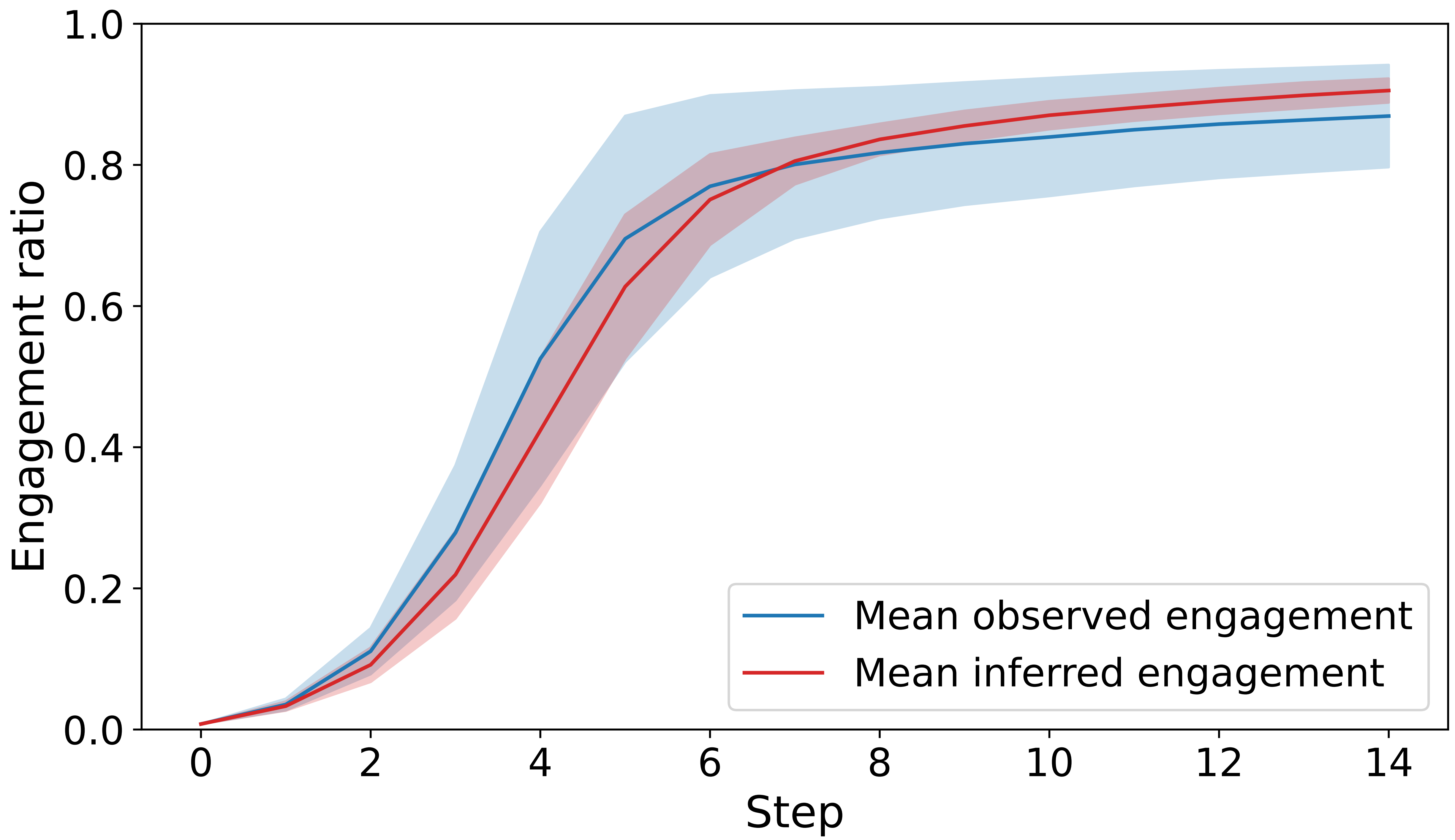


Assign agents *Big Five* traits (openness, conscientiousness, extraversion, agreeableness, and neuroticism), each taking a binary value of low or high

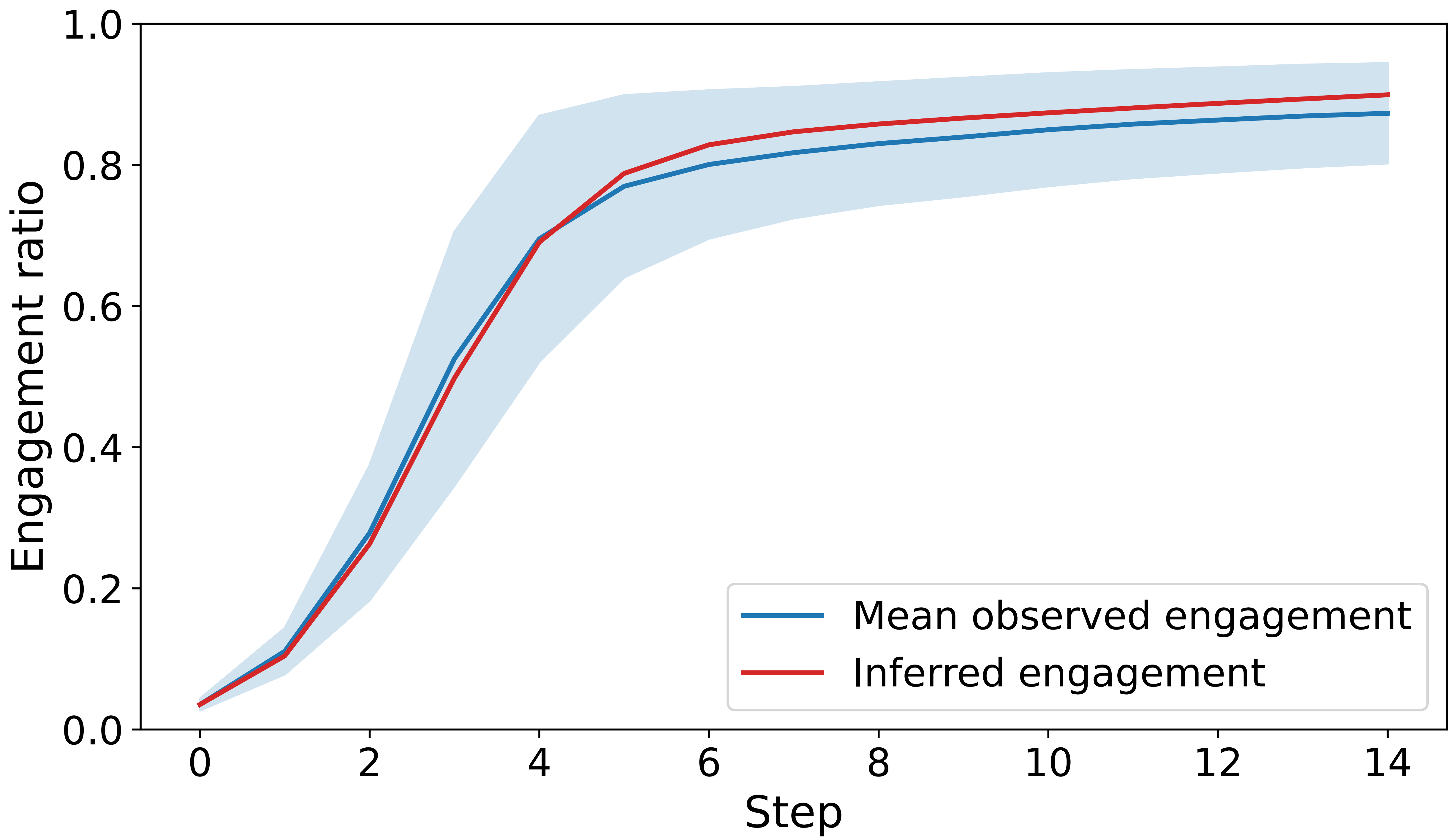
*Events taken from ACLED Philippines data (2021-02-24 ~ 2024-02-27).

[Tian, Mohler, Brantingham & Rodríduez, arXiv (2026)]

Model Fitting from Data



Logistic Model



2-group Mean-Field Model

Simpler and Better !

* GitHub: https://github.com/Moyi-Tian/Learning-LLM-Dynamics

Summary

- We developed a two-layer model for online–offline dynamics, derived corresponding mean-field approximations, and analyzed outburst behavior using the reproductive number.
- From a data-driven perspective, we investigated the learning of effective models in two complementary projects, representing initial steps toward bridging model-driven and data-driven approaches for online–offline social systems on networks.

Takeaway: ODE approximation models can effectively capture dynamics on complex networks and provide a useful framework for subsequent analysis.

Thanks to our sponsors:

- Air Force Office of Scientific Research Multidisciplinary University Research Initiative [grant FA9550-22-1-0380]
- NIH-NIGMS Division of Biophysics, Biomedical Technology and Computational Biosciences [grant R35GM149335]
- U.S. National Science Foundation Division of Mathematical Sciences [grant 2042413]

References

1. Moyi Tian, P. Jeffrey Brantingham, and Nancy Rodríguez. Modelling the spillover from online engagement to offline protest: stochastic dynamics and mean-field approximations on networks. Journal of Complex Networks. 2026.
2. Moyi Tian, Daniel A. Messenger, Vanja Dukic, Nancy Rodríguez, and David M. Bortz. Learning effective models from network dynamics data with multiple initial conditions using weak form SINDy. arXiv:2605.30432, 2026.
3. Moyi Tian, George Mohler, P. Jeffrey Brantingham, and Nancy Rodríguez. Learning dynamics from online-offline systems of LLM agents. arXiv:2602.23437, 2026.

[See paper]

2

Daniel Messenger
(LANL)

Vanja Dukic
(CU Boulder)

David Bortz
(CU Boulder)

Nancy Rodríguez
(CU Boulder)

1

P. Jeffrey Brantingham
(UCLA)

3

George Mohler
(Boston College)