Modeling and Learning Dynamics of Online–Offline Social Systems on Networks

Moyi Tian [1], Nancy Rodríguez [1], David M. Bortz [1], George Mohler [2] and P. Jeffrey Brantingham [3]

1. Department of Applied Mathematics, University of Colorado Boulder 2. Computer Science Department, Boston College. 3. Department of Anthropology, University of California, Los Angeles

Introduction

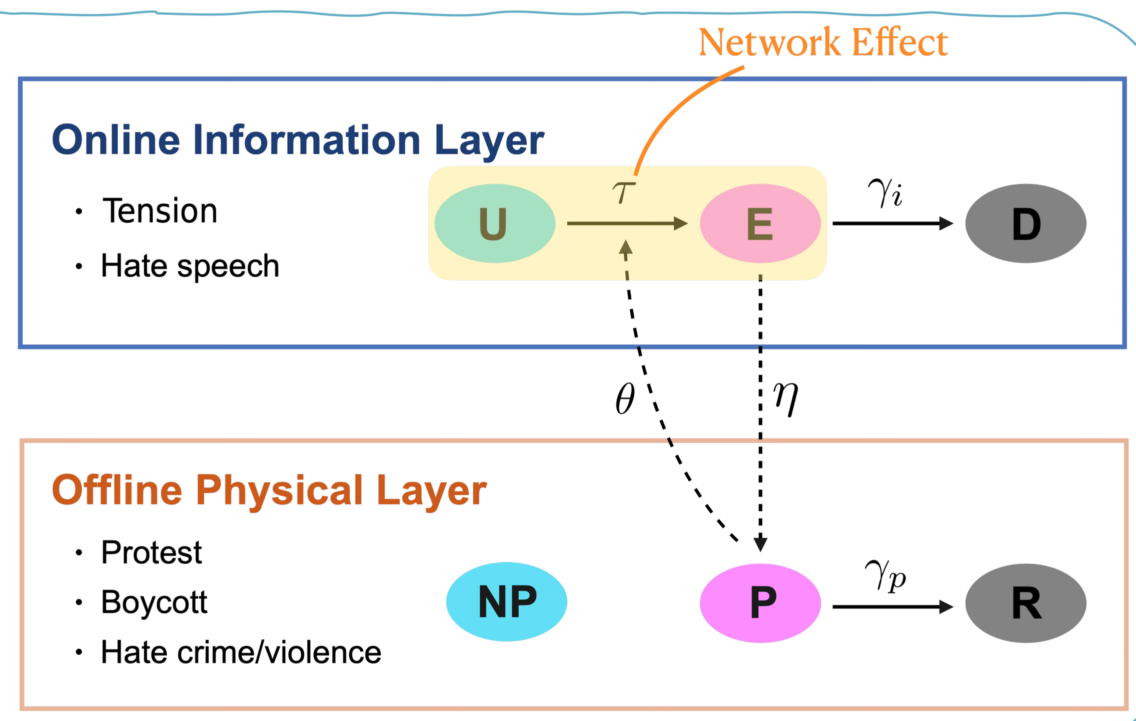
- Social media is transforming how collective action and conflict unfold.
- Active social media user identities are equivalent to about 69% of the global population [DataReportal, 2026].
- Online communication and on-site protests can shape each other, creating coupled online-offline dynamics.
- **Goal:** Model mechanisms linking online engagement to offline participation and connect the models with data.

Framework

Online States
- **U:** Uninterested
- **E:** Engaged
- **D:** Disengaged

Offline States
- **NP:** Not participating
- **P:** Participating
- **R:** Recovered

Parameters
- τ : Online transmission rate
- γ_i : Online recovery rate
- η : Offline transmission rate
- γ_p : Offline recovery rate
- θ : Self-excitement



Modeling & Analysis [1]

Main idea: Start from a stochastic network model, derive continuum ODE approximations, and use them for fast prediction and analysis.

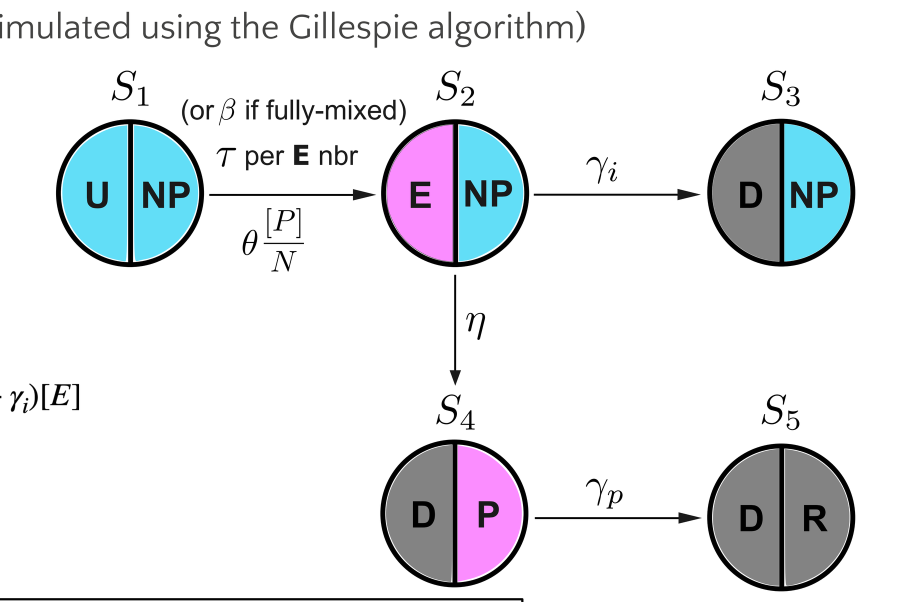
Continuous-Time Markov Chain (simulated using the Gillespie algorithm)

derive expected-count ODE

ODE Model

Online:
$$\frac{d[U](t)}{dt} = -\tau[UE] - \frac{\theta}{N}[U][P]$$
$$\frac{d[E](t)}{dt} = \tau[UE] + \frac{\theta}{N}[U][P] - (\eta+\gamma_i)[E]$$
$$\frac{d[D](t)}{dt} = (\eta+\gamma_i)[E]$$

Offline:
$$\frac{d[P](t)}{dt} = \eta[E] - \gamma_p[P]$$
$$\frac{d[R](t)}{dt} = \gamma_p[P]$$



Notation
- N: Total number of nodes
- [X]: Expected # of individuals in state X
- [XY]: Expected # of X–Y neighbor pairs

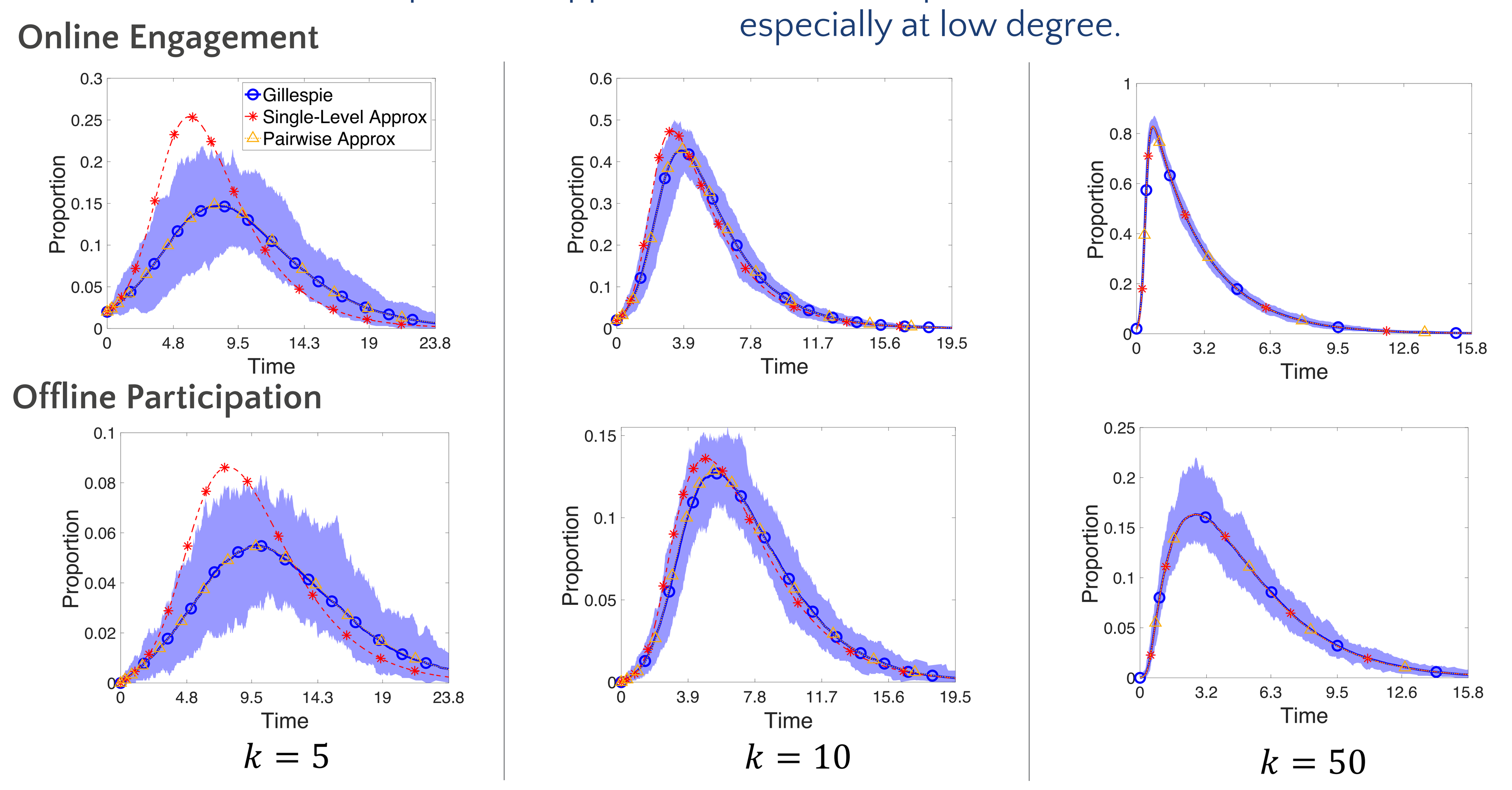
Challenge: The ODE system depends on the pair term $[UE]$, which must be approximated to close the system.

Method: Mean-field approximations close $[UE]$ using different assumptions about network structure and different levels of network information [4]. For example, assuming
- **Random mixing**, i.e., everyone is connected to everyone else, then **fully-mixed population model**
- Network is **homogeneous** (k-regular), then **single-level approximation model** – approximate pair counts by singletons
  OR **pairwise approximation model** – approximate triplet counts by pair counts
  OR even higher-order approximations (more accurate but slower)
- The network is **heterogeneous**, then also **single-level approximation** OR **pairwise approximation** OR higher-order approximation with counts by degree groups

Simulations

Stochastic simulations on Erdős–Rényi networks ($N = 1000$) with varying average degree (k) are compared with heterogeneous single-level and pairwise approximations. The pairwise approximation is more accurate, while the single-level approximation improves as k increases because random mixing becomes more valid.

The pairwise approximation better captures stochastic network simulations, especially at low degree.



Why continuum approximations?
- Fast prediction of expected population-level behavior.
- Basic reproductive numbers can be computed from the ODE models and used to classify online/offline outbursts.

Data-Driven Learning

To connect modeling with data, we study two inverse problems from complementary directions:
1. Identify effective ODE models from synthetic noisy network dynamics data using WSINDy.
2. Infer models better matching LLM-agent spreading data.

Identify Effective Models with WSINDy [2]

Question: Can we learn interpretable ODE models from stochastic network dynamics data?

Method:
WSINDy: Weak Form SINDy (Sparse Identification of Nonlinear Dynamics), a system identification method.

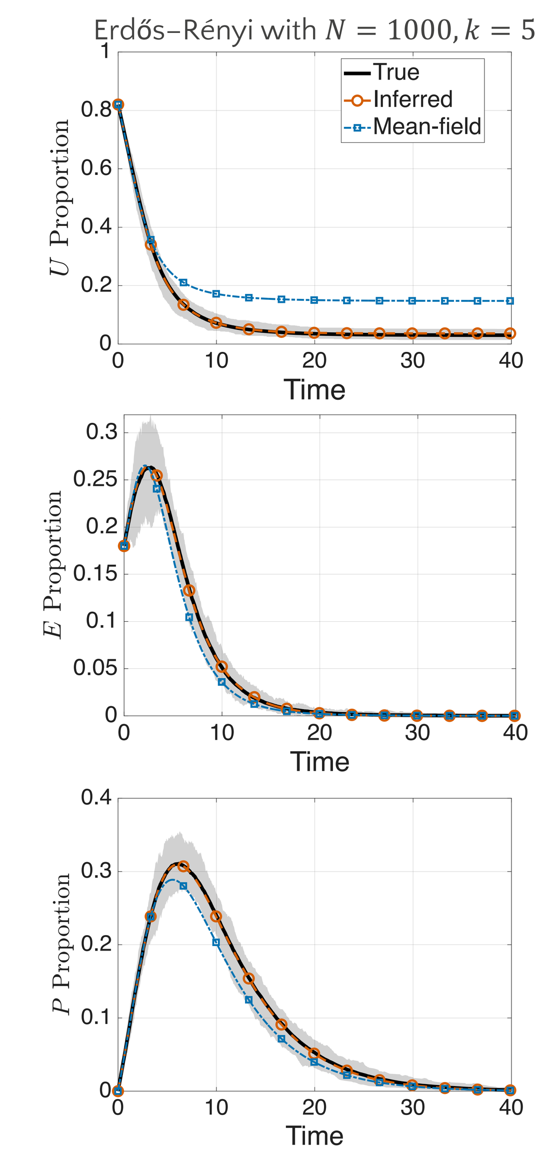
Goal: Learn governing equations $\dot{x}(t) = F(x(t))$ from noisy data measured at times $t_1, \dots, t_M$

$$X = [x(t_1)+\varepsilon_1 \ \ x(t_2)+\varepsilon_2 \ \dots \ \ x(t_M)+\varepsilon_M]^T.$$

SINDy: [Brunton, Proctor & Kutz, PNAS (2016)]
Solve the sparse regression problem, choosing as few nonzero weights as possible from a predefined library of candidate functions, so that the learned model has few terms.

WSINDy: [Messenger & Bortz, MMS (2021)]
Transform the SINDy problem into weak form by convolving with a family of smooth compactly supported test functions to avoid derivative approximation and thus reduce sensitivity to noise.

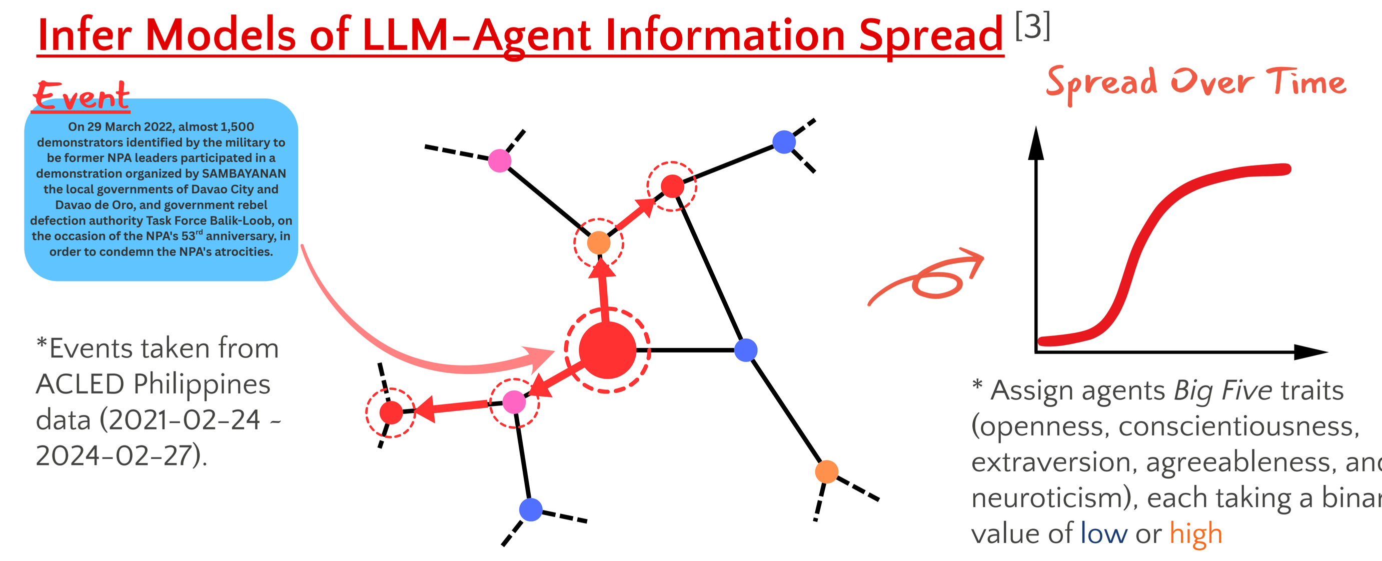
We use **multiple trajectories** with different initial conditions to **learn models** from stochastic network-based dynamics data.



Main Result: The learned model provides a better approximation than the mean-field, particularly on sparse networks.

* True = stochastic data; Mean-Field = homogeneous single-level approximation (uses comparable input information)

Infer Models of LLM-Agent Information Spread [3]



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

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

Motivation:
- Bots generated over half of Internet traffic in 2023 [Ng & Carley, Sci. Rep. (2025)].
- Their high activity can strongly shape the spread of online information.

Question: What model best matches LLM-agent spreading dynamics?

Experimental Process & Method:
- Build a random network of LLM agents, each assigned a *Big Five* personality profile.
- Choose one initial spreader for a given event.
- At each time step, non-spreader neighbors of current spreaders decide whether to forward the news to their own neighbors, based on their personality traits and the news content. Once an agent decides to spread, it remains a spreader for all later steps.
- Extract the ratio of spreaders over time and fit logistic & ODE models.

$$\begin{cases} \frac{dU_1}{dt} = -k\tau_1 U_1 E, \\ \frac{dU_2}{dt} = -k\tau_2 U_2 E, \\ \frac{dE}{dt} = \sum_{i=1,2} k\tau_i U_i E \end{cases}$$

U: % non-spreader
E: % spreader
$i = 1,2$: trait-profile groups
A homogeneous single-level model

Main Result:
A two-group mean-field model better captures aggregate spreading dynamics than a high-dimensional logistic model.

Conclusion

- We developed a two-layer network model for online-offline dynamics and derived mean-field approximations for outburst prediction and analysis.
- We used data-driven learning to infer effective models from stochastic network simulations and LLM-agent spreading data.

Takeaway: ODE approximations can capture complex network dynamics and provide useful, interpretable models for fast analysis.

References

Papers behind this poster:
[1] **M. Tian**, P. J. Brantingham, and N. Rodríguez. "Modelling the spillover from online engagement to offline protest: stochastic dynamics and mean-field approximations on networks." Journal of Complex Networks, 2026. [Published]
[2] **M. Tian**, D. A. Messenger, V. Dukic, N. Rodríguez, and D. M. Bortz. "Learning effective models from network dynamics data with multiple initial conditions using weak-form SINDy." arXiv:2605.30432, 2026. [Preprint]
[3] **M. Tian**, G. Mohler, P. J. Brantingham, and N. Rodríguez. "Learning dynamics from online-offline systems of LLM agents." arXiv:2602.23437, 2026. [Preprint]

Related background:
[4] I. Z. Kiss, J. C. Miller, and P. L. Simon. Mathematics of Epidemics on Networks: From Exact to Approximate Models. Springer, 2017.

Funding

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