Efficiently Learning the Topology and Behavior of a Networked Dynamical System Via Active Queries

Daniel J. Rosenkrantz [1,3] Abhijin Adiga [1] Madhav V. Marathe [1,2] Zirou Qiu [1,2]
S. S. Ravi [1,3] Richard E. Stearns [1,3] Anil Vullikanti [1,2]

[1]UVA Biocomplexity Institute and Initiative, [2]University of Virginia and [3]University at Albany – SUNY

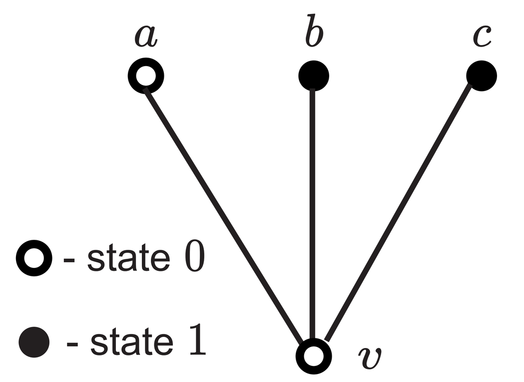
(Poster Presented at ICML-2022)

Dynamical Systems: Basics

A **Synchronous Dynamical System** (SyDS) has:

- An underlying graph $G(V, E)$.
- **Nodes:** Agents in the system.
- **Edges:** Permissible local interactions.
- State values for nodes from a finite domain (here: $\{0, 1\}$).
- A **local transition function** for each node.
- **Update mechanism: synchronous**.

Local Transition Function:



Local function f_v :

- Inputs: States of nodes in the closed neighborhood $N^+(v)$ of node v.
- Output: Next state of v.

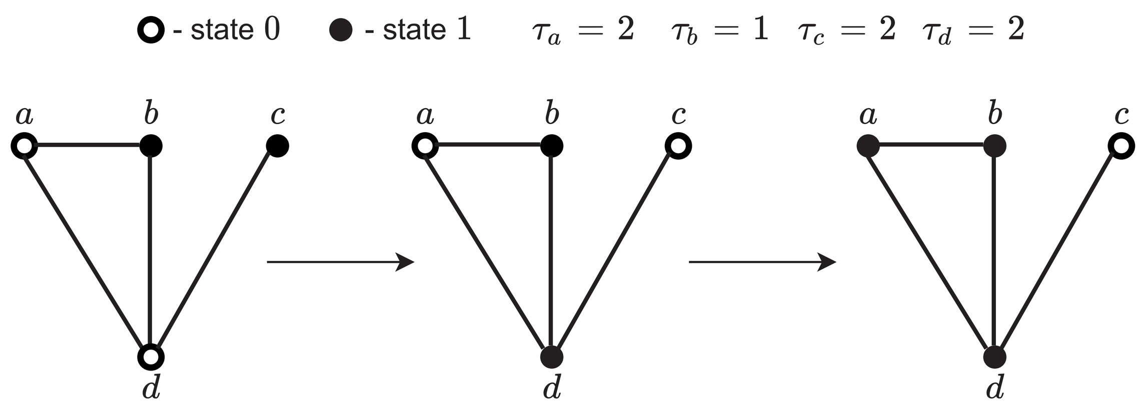
Symmetric Function:

- Each node v has a table T_v with $deg_v + 1$ rows.
- When the number of 1's in the input is i, the next state of v is given by $T_v[i]$.

Threshold Function:

- Each node v has an integer threshold τ_v.
- The next state of v is 1 iff the no. of state-1 nodes in $N^+(v)$ is at least τ_v.

Example of a Threshold-SyDS:

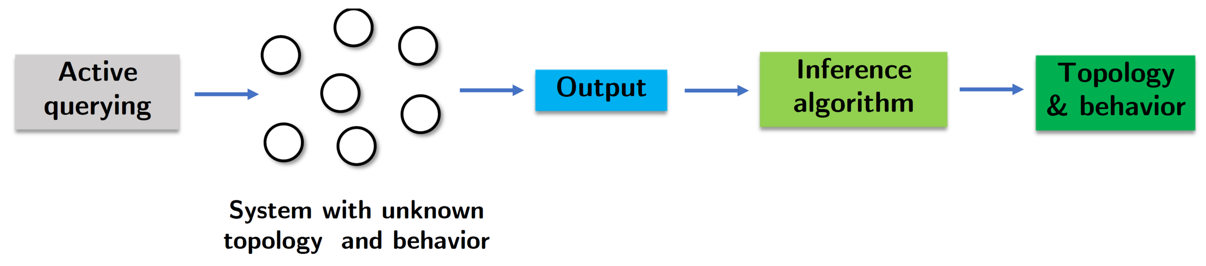


Acknowledgments

This research is supported by UVA Strategic Investment Fund award number SIF160, VDH grant VDH-21-501- 0135-1, NSF Grants OAC-1916805 (CINES), CCF-1918656 (Expeditions), IIS-1931628, IIS-1955797, and NIH grant R01GM109718.

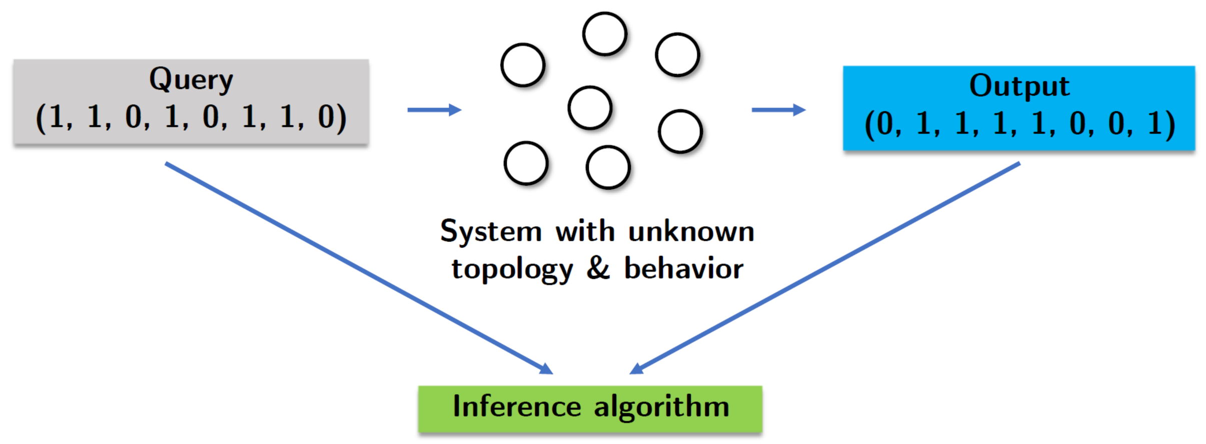
Problem Overview

Our focus: Inferring both the topology and the local functions of a SyDS.



Query Model

- Each query specifies the *input* to each of the local functions. The response provides the value of each function for the inputs specified by the query.
- **Batch mode**: Submit all queries in a *single batch*.
- **Adaptive mode**: Submit Queries in *multiple stages*; new queries may be based on the responses to the previous queries.



Previous Work (Brief)

- Active querying to predict users' choices from a known set of options [Kleinberg et al., 2017].
- Inferring influence functions for networked systems [Romero et al., 2011, Narasimhan et al., 2015].
- Inferring the network structure given the contagion propagation model [Abrahao et al., 2013, Gomez-Rodriguez et al., 2010].
- Learning the local functions of discrete dynamical systems from *observed* data [Adiga et al., 2017].
- Given the network topology, active querying to learn the local functions [Adiga et al., 2018; 2019; 2020].

Contact

Zirou Qiu (`zq5au@virginia.edu`)

Our Contributions

To infer symmetric-SyDSs: Efficient inference algorithms under batch and adaptive modes.

> **Theorem 1 (Batch: $O(n^2)$ queries)**
> In polynomial time, a *batch* query set Q with $O(n^2)$ queries can be constructed, such that Q can be used to identify the local functions and the topology of $\mathcal{S}$.

> **Theorem 2 (Adaptive: $O(n + m \log n)$ queries)**
> In polynomial time, an *adaptive* query set Q with $O(n + m \log n)$ queries can be constructed, such that Q can be used to identify the local functions and the topology of a symmetric-SyDS $\mathcal{S}$.

Remarks:

- No. of nodes = n and number of edges = m.
- The algorithms don't assume that m is known; the parameter m in the bound results from the analysis.
- For *sparse* networks (e.g., $m = O(n)$), the adaptive mode algorithm uses asymptotically fewer queries.

To infer threshold-SyDSs: A deterministic algorithm under the adaptive mode and a randomized algorithm under the batch mode.

> **Theorem 3 (Adaptive: $O(n + m \log n)$ queries)**
> In polynomial time, an *adaptive* query set Q with $O(n + m \log n)$ queries can be constructed, such that Q can be used to identify the local functions and the topology of a threshold-SyDS $\mathcal{S}$.

> **Theorem 4 (Batch: Randomized)**
> In polynomial time, a *batch* query set Q with $O(n\Delta \log n)$ queries can be constructed, such that Q can be used to identify the local functions and the topology of a threshold-SyDS $\mathcal{S}$ w.h.p.

Remark: The randomized algorithm assumes that an upper bound on the maximum degree Δ is known.

Our Contributions (cont.)

A lower bound for batch mode:

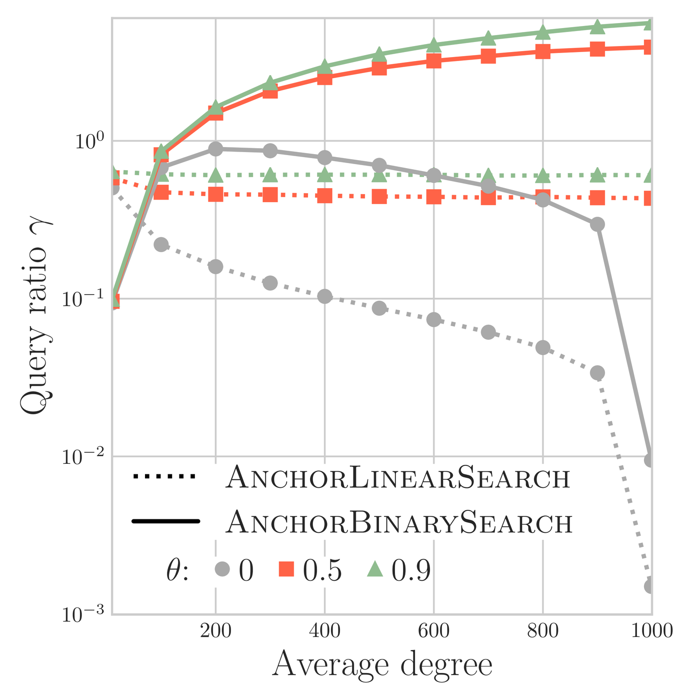
> **Theorem 5 (Batch: Lower bound)**
> Under the batch mode, any query set that can correctly learn the network topology of any threshold-SyDS must contain $\Omega(n \log n)$ queries.

Experimental Results

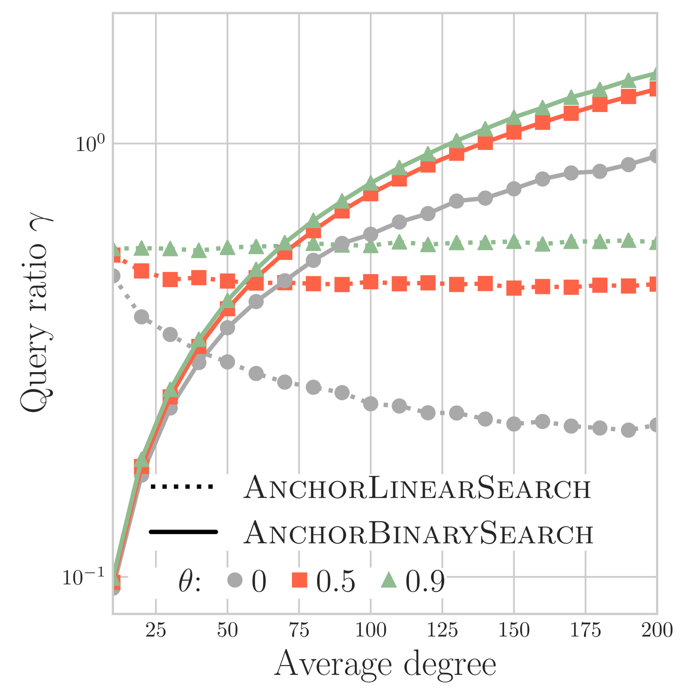
Experimental Setup:

- **Datasets**: (i) real-world networks from various domains, with the numbers of nodes ranging from $7,000+$ to $28,000+$; (ii) synthetic `gnp` and power-law networks.
- **Metric**: Query ratio $\gamma = a/b$, where a and b are the numbers of queries used by an adaptive algorithm and a batch algorithm respectively.

Experimental Results (Examples):



`Gnp` network with $10,000$ nodes and average degree varied from 10 to 999.



Power-law network with $10,000$ nodes and average degree varied from 10 to 200.