Benchmarking Structural Inference Methods for Interacting Dynamical Systems with Synthetic Data

NeurIPS 2024

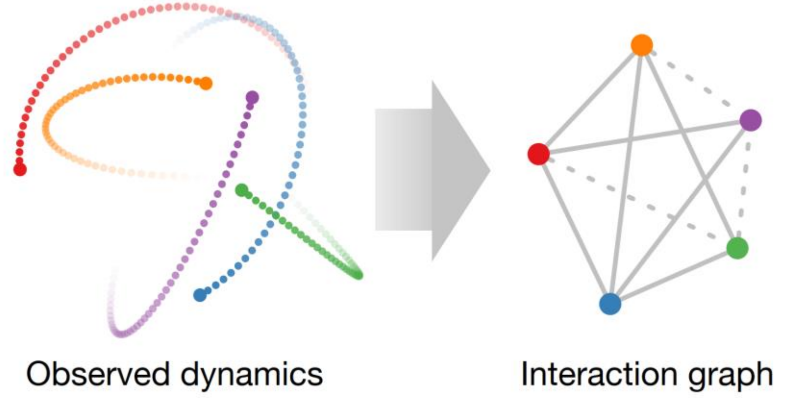
Aoran Wang, Tsz Pan Tong, Andrzej Mizera, Jun Pang

UNIVERSITÉ DU LUXEMBOURG

IDEAS NCBR

Structural inference in dynamical systems

- In a dynamic system, the topological structure of interaction may be unknown
  - → Structural Inference



- Observed data: node-level trajectories
- Example (charged particles)

Nodes	Particles
Node features	Position (x_1, x_2), Velocity (v_1, v_2)
Edges	Charges
Interaction	Electrostatics force

Problem

- Existing methods are evaluated on:
  - distinct datasets,
  - specific graph types
  - tailored to different research domains
  - unique underlying assumptions

→ Urges for a unified, systematic benchmarks across different fields

Methods for Structural Inference

Classical statistics

ppcor TIGRESS

Deep learning

NRI iSIDG
ACD RCSI
MPM

Information theory

ARACNe PIDC
CLR Scribe

Tree method

dynGENIE3 XGBGRN

Our contributions

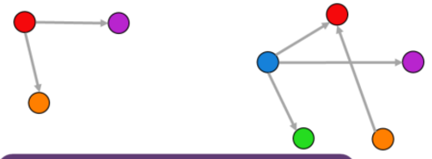
- Dataset for Structural Inference (DoSI)
    - 11 types of interaction graphs
    - Graph size from 15 to 250
    - 3 dynamic functions
    - 231 distinct graphs
    - 213,445 trajectories

Types of Interaction Graphs

BN	Brain Networks
CRNA	Chemical Reactions in Atmosphere
FW	Food Webs
GCN	Gene Co-expression Networks
GRN	Gene Regulation Networks
IN	Intercellular Networks
LN	Landscape Networks
MMO	Man-made Organic Reaction Networks
RNLO	Reaction Networks inside Living Organism
SN	Social Networks
VN	Vascular Networks

Graph Size

15 nodes to 250 nodes



Dynamical Functions

	Dim.	# Edge types
Netsims	1	1
Springs	4	1
Charged particles	4	2

- Comprehensive benchmarking
    - 13 structural inference methods
    - Measures accuracy, scalability, robustness and sensitivity
    - Over 706,800 CPU hours and 263,400 GPU hours

Results - Accuracy

- Only deep learning can cater **multi-dim.** feature
- Classical statistical models have **high ranks** consistently

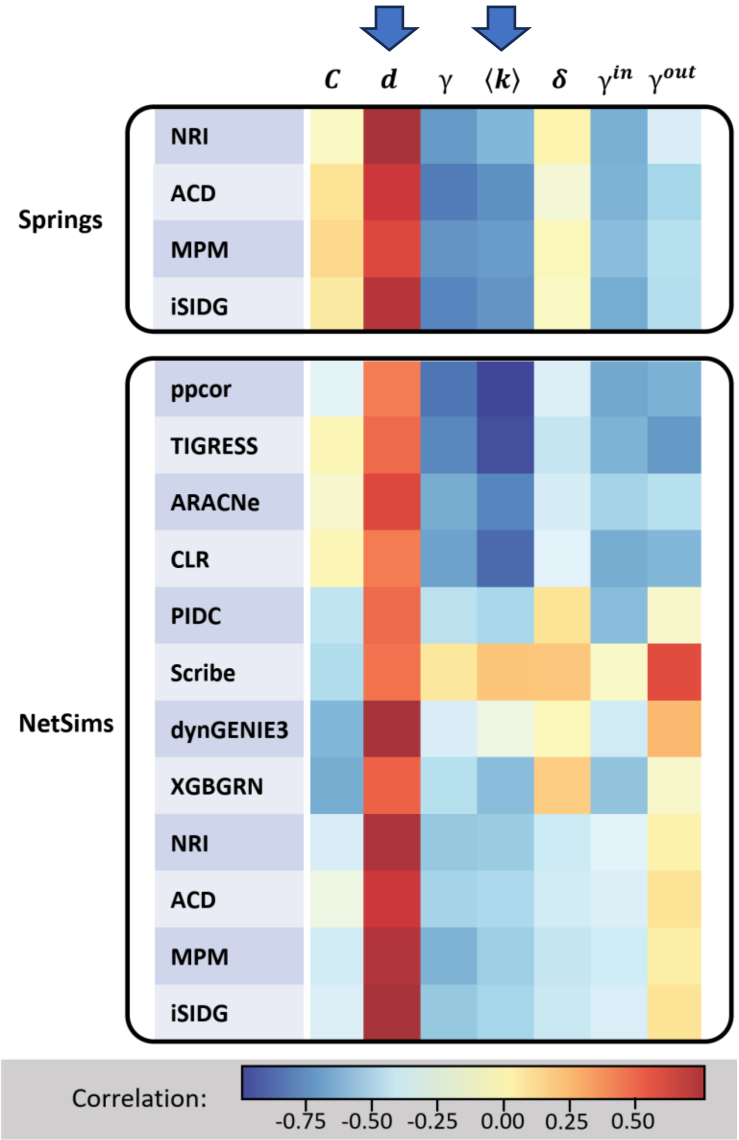
		BN	CRNA	FW	GCN	GRN	IN	LN	MMO	RNLO	SN	VN
Springs	NRI	98.99	73.19	76.07	91.03	90.15	88.56	90.46	85.07	78.96	81.36	93.37
	ACD	99.46	73.95	75.72	92.81	89.04	87.88	91.07	91.14	86.15	80.76	91.52
	MPM	99.64	73.15	75.74	90.57	89.29	88.77	91.15	90.48	84.73	79.29	88.97
	iSIDG	99.69	74.57	76.31	92.30	90.26	89.47	90.66	90.63	84.16	81.40	93.42
	RCSI	99.45	75.06	76.08	92.07	91.79	90.51	91.00	91.49	84.96	82.86	94.38
NetSims	ppcor	98.11	90.28	74.80	97.99	88.57	96.38	90.15	98.29	98.21	94.26	98.38
	TIGRESS	96.50	72.20	58.51	84.55	84.38	87.68	89.43	99.96	99.95	79.80	99.54
	ARACNe	96.79	77.33	63.26	93.30	70.18	85.69	76.67	95.39	96.05	80.37	98.03
	CLR	97.17	84.50	68.08	96.43	75.88	90.51	95.00	98.12	97.99	87.71	98.38
	PIDC	93.01	78.66	60.89	92.73	62.70	85.31	90.58	66.76	68.79	86.17	87.25
	Scribe	62.32	52.28	52.49	49.39	46.08	51.63	53.76	38.12	38.10	52.23	55.36
	dynGENIE3	97.61	51.93	49.63	48.65	59.21	61.66	54.81	27.40	30.34	54.60	96.33
	XGBGRN	100.00	87.01	64.83	95.42	82.96	99.63	97.26	69.34	78.43	99.56	98.83
	NRI	87.46	49.80	49.03	49.40	62.29	58.16	54.02	62.12	65.02	52.39	75.89
	ACD	89.92	49.57	50.31	46.46	66.64	57.60	56.77	63.38	59.55	54.56	70.85
	MPM	93.50	50.38	51.99	58.83	66.71	59.35	54.58	63.58	63.00	55.37	76.44
	iSIDG	93.63	50.85	51.41	53.05	61.66	58.59	55.85	63.60	63.10	56.63	77.94
	RCSI	94.44	50.77	52.35	54.32	65.83	57.66	57.87	64.08	62.64	57.93	79.54

Rank: Low High

Results - Correlation with Graph Properties

Model performances correlate:

- Positively with **average shortest path distance d**
- Negatively with **average degree $\langle k \rangle$**



Results - Scalability

- All models **deteriorate as system size increases** except PIDC and dynGENIE3.
- **Deep learning** methods are most **sensitive to graph size**.
- **Deep learning** methods can **infer multiple edge types**, but the performance drops
- Classical statistical methods are **highly scalable** without a significant drop in accuracy

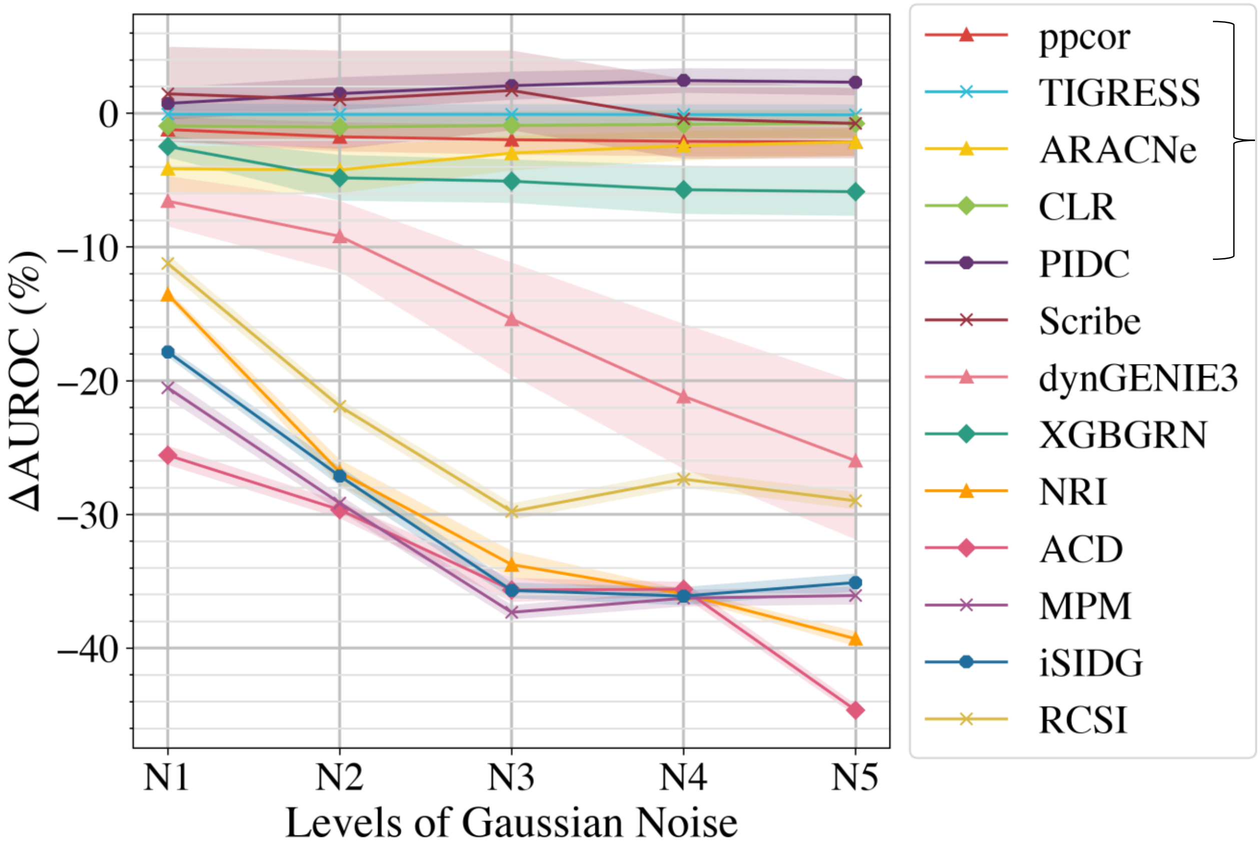
		n15	n30	n50	n100
Springs	NRI	93.42	86.39	85.37	80.52
	ACD	92.07	88.66	83.91	81.36
	MPM	94.26	87.63	82.93	81.18
	iSIDG	94.62	88.36	85.68	81.37
	RCSI	94.80	89.25	85.73	82.82
NetSims	ppcor	93.22	93.59	93.30	92.51
	TIGRESS	89.11	87.61	86.15	83.49
	ARACNe	85.71	85.27	84.95	83.36
	CLR	90.27	91.19	90.54	87.91
	PIDC	76.73	77.63	79.50	83.54
	Scribe	52.47	51.10	49.30	47.76
	dynGENIE3	54.92	56.90	56.51	61.55
	XGBGRN	90.47	91.89	89.15	82.40
	NRI	65.73	61.06	57.46	56.85
	ACD	65.21	58.81	58.60	57.27
	MPM	70.70	67.06	61.83	58.69
	iSIDG	68.18	61.88	61.06	58.43
	RCSI	69.12	64.48	61.52	58.48

		n15	n30	n50	n100
Charged Particles	NRI	72.14	71.66	68.98	64.35
	ACD	74.36	73.42	71.20	67.45
	MPM	75.10	74.89	72.04	67.82
	iSIDG	75.67	75.02	73.12	69.37
	RCSI	75.80	74.11	72.04	66.75

Rank: Low High

Results - Robustness to Additive Noise

- Most methods based on **classical statistics and information theory** are **resistant** to various levels of **Gaussian noise**.



Summary

https://structinfer.github.io/

- Structural inference: finding interaction graph behind dynamic systems
- We provide the DoSI dataset with 11 types of graphs, 3 dynamic functions and 213,445 trajectories
- We present a unified, systematic benchmark across 13 models from different fields
- We found that:
    - Only current deep learning methods can tackle multi-dimension features
    - Classical statistical methods remain strong on accuracy, robustness and scalability
    - Model performances are correlated positively with average shortest path distance and negatively with negatively with average degree