NEURAL INFORMATION PROCESSING SYSTEMS

NeurIPS 2024

December 10, 2024

Vancouver
Vancouver Convention Center

State Space Models on Temporal Graphs: A First-Principles Study

Jintang Li, Xinzhou Jin, Liang Chen, Zibin Zheng, Sun Yat-sen University
Ruofan Wu, Boqun Ma, Coupang

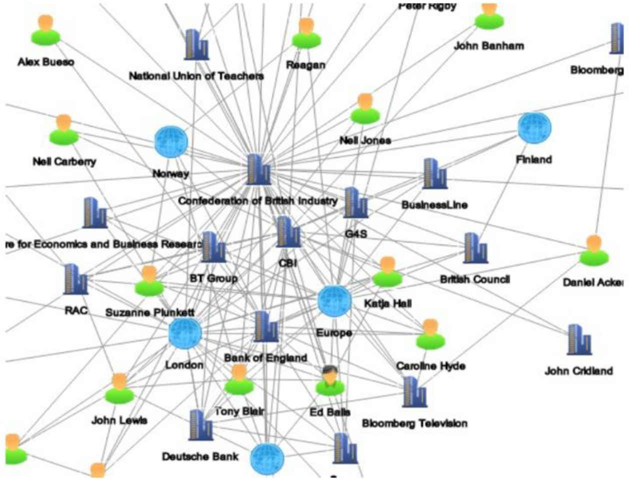
中山大學 SUN YAT-SEN UNIVERSITY

coupang

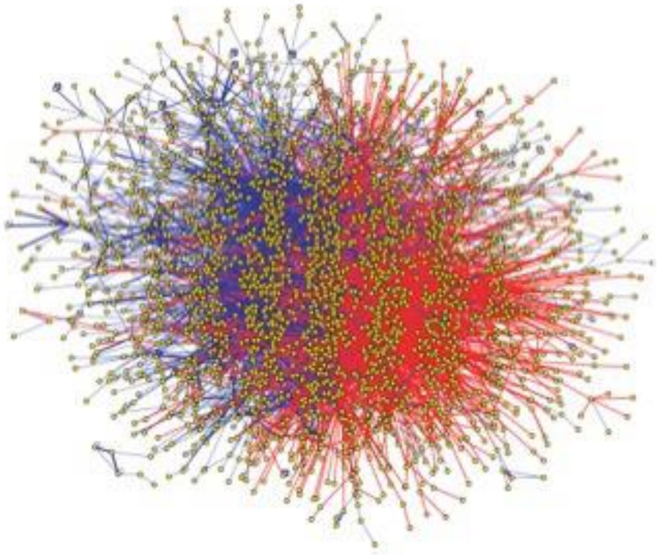
BACKGROUND GRAPHS & GRAPH NEURAL NETWORKS

Networks are everywhere - Graphs are natural way to model such networks

Social Networks



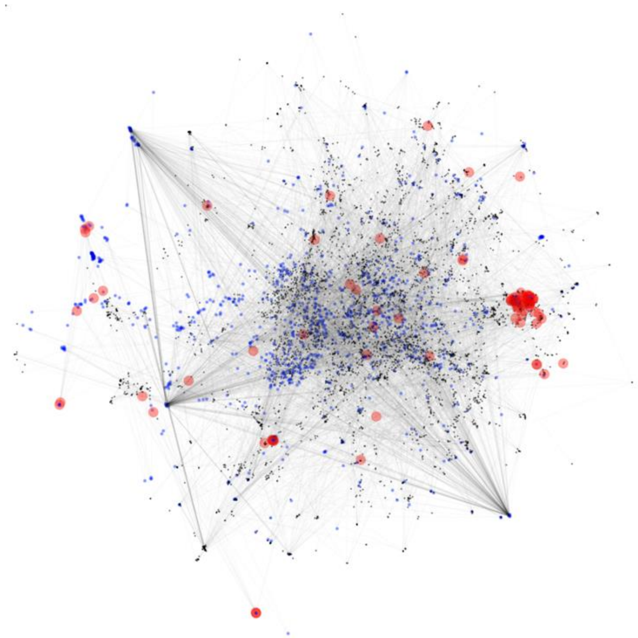
Financial transactions

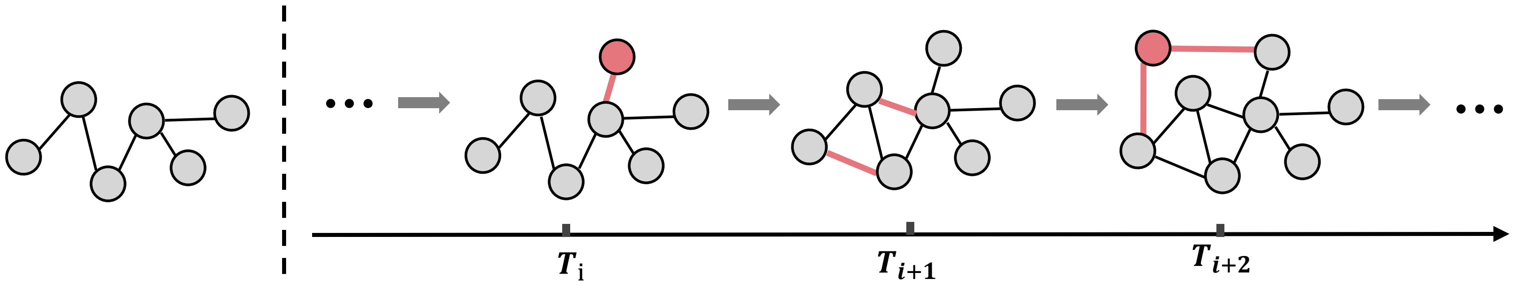


Protein-Protein
Interaction Networks

MOTIVATION TEMPORAL GRAPH REPRESENTATION LEARNING

Graphs are often dynamic - nodes, edges, and attributes may evolve constantly over time



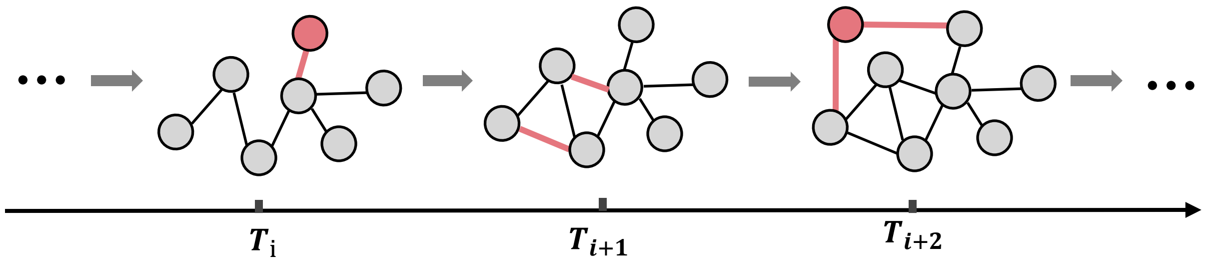


(a) Static Graph

(b) Temporal Graph

MOTIVATION TEMPORAL GRAPH REPRESENTATION LEARNING

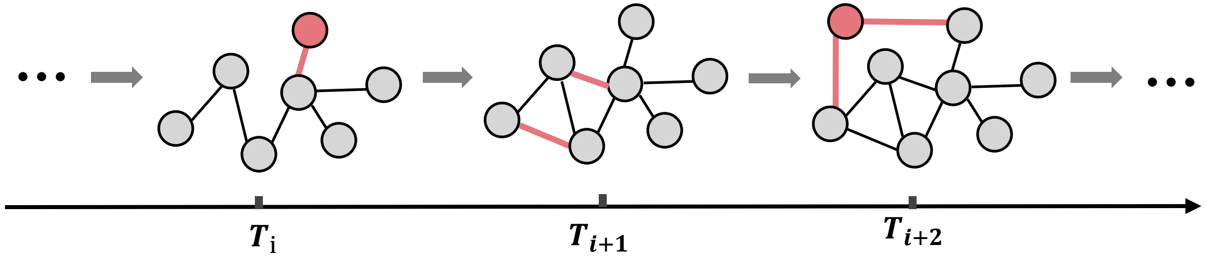
GNNs developed for static graphs are not directly applicable for temporal graphs



GNN?

MOTIVATION TEMPORAL GRAPH REPRESENTATION LEARNING

Generalize GNNs to temporal graphs by additionally considering the time dimension.



GNN

+

Time

NEURAL INFORMATION PROCESSING SYSTEMS

Generalize GNNs to temporal graphs by additionally considering the time dimension.

From RNNs, SNNs, Transformers, to SSMs

	RNNs [46, 4, 18]	SNNs [38, 8]	Transformers [42]	SSMs (S6 [10])
Training	Slow	Slow	Fast	Fast
Inference	Fast	Fast	Slow	Fast
Para. Size	Medium	Extremely small	Large	Small
Performance	☆☆☆	☆☆☆	☆☆☆☆☆	☆☆☆☆
Limitations	Forgetting	Vanishing gradients	Mem. & Time: $O(n^2)$	?

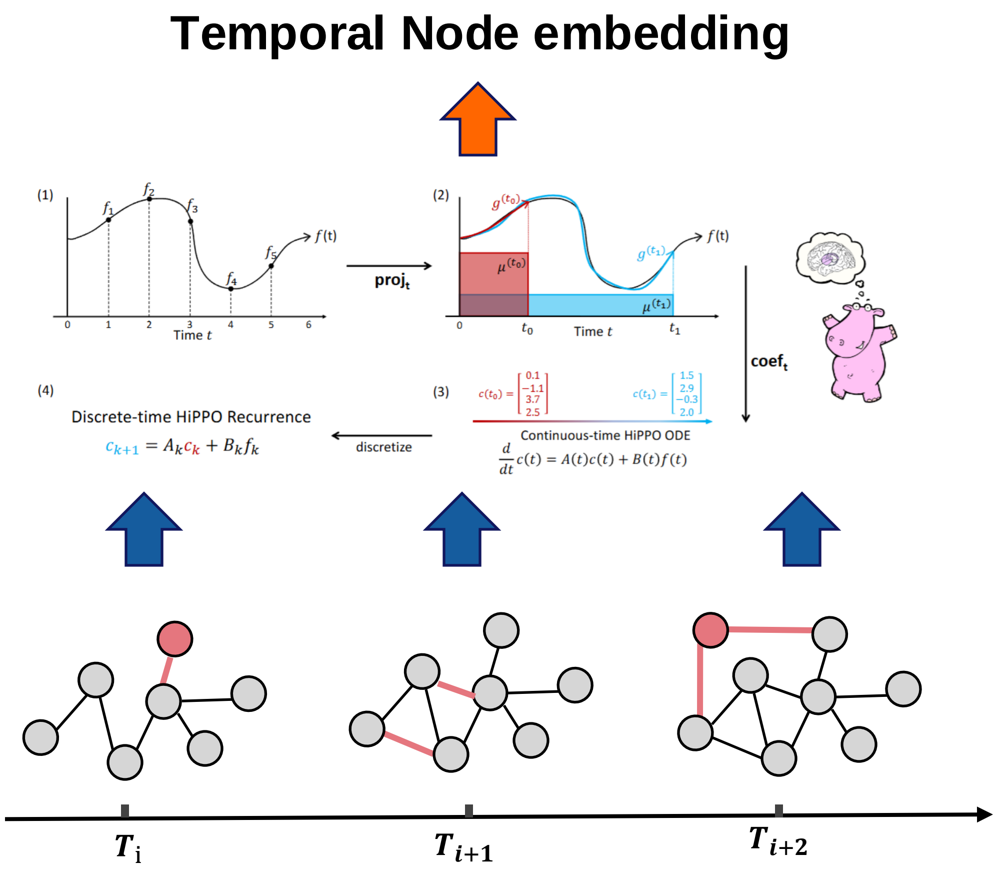
GNN

+

Time

NEURAL INFORMATION PROCESSING SYSTEMS

GraphSSM: Modeling temporal graph dynamics with SSMs



GHiPPO: HiPPO on Temporal Graphs

Laplacian-regularized online approximation

$$\mathcal{L}_t(Z;G,X,\mu) = \overbrace{\int_0^t \|X(s) - Z(s)\|_2^2 d\mu_t(s)}^{\text{HiPPO term}} + \alpha \overbrace{\int_0^t Z(s)^\top L(s) Z(s) d\mu_t s.}^{\text{Regularization term}}$$

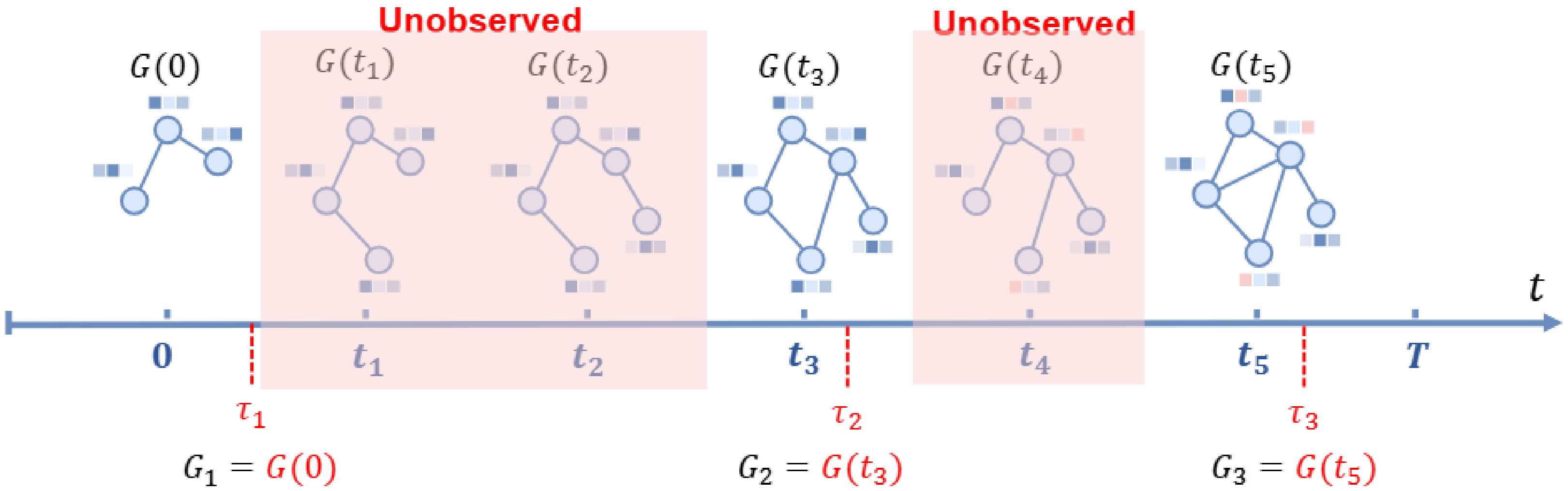
Graph memory projection operator

$$\mathrm{GPROJ}_t(G,X) = \underset{Z: z_v \in \mathcal{P}_N \forall v \in V}{\arg\min} \ \mathcal{L}_t(Z;G,X,\mu).$$

Extending HiPPO to GHiPPO

$$\mathrm{GHIPPO}(G,X) = \overbrace{\boxed{\mathrm{CoEF}_t}}^{\text{Coefficient projection}}(\mathrm{GPRoJ}_t(G,X)).$$

Mixed Discretization for Unobserved Graph Mutations



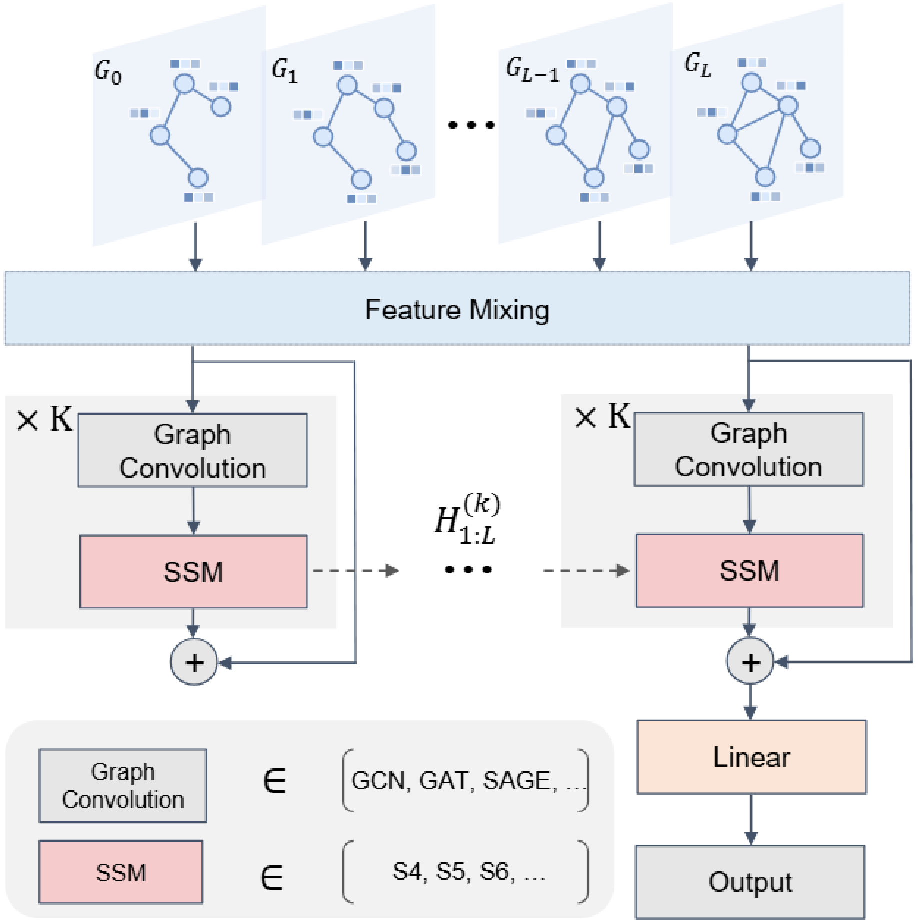
- Ordinary ZOH: $\widehat{X}_l^{(\mathrm{O})} = \mathrm{GNN}_\theta\left(X_l, G_l\right)$
- Feature Mixing: $\widehat{X}_l^{(\mathrm{F})} = \mathrm{GNN}_\theta\left(\mathrm{Mix}_\phi\left(X_{l-1}, X_l\right), G_l\right)$
- Representation Mixing: $\widehat{X}_l^{(\mathrm{R})} = \mathrm{Mix}_\phi\left(\mathrm{GNN}_\theta\left(X_{l-1}, G_{l-1}\right), \mathrm{GNN}_\theta\left(X_l, G_l\right)\right)$

GraphSSM Framework

$$H_{1:L}^{(k)} = \sigma\left(\text{SSMLayer}\left(H_{1:L}^{(k-1)}, G_{1:L}\right)\right) + \textbf{Linear}\left(H_{1:L}^{(k-1)}\right), 1 \leq k \leq K,$$

- ❑GraphSSM-S4: Single-Input, Single-Output (SISO) configuration
- ❑GraphSSM-S5: Multiple-Input, Multiple-Output (MIMO) configuration
- ❑GraphSSM-S6: input-controlled time intervals and state matrices (Δ, B, C)

GraphSSM: Overall framework



- ❑Backbone: Graph Neural Networks
  - ❑ GCN, GraphSAGE, GAT
- ❑Sequence Model: State Space Models
  - ❑ S4, S5, S6 (Mamba)
- ❑Mixed Discretization
  - ❑ Unobserved graph mutations
  - ❑ Feature mixing
  - ❑ Representation mixing

EXPERIMENTS SETTINGS

4k → 577k

- Datasets: DBLP3, Brain, Reddit, DBLP-10, arXiv, Tmall
- Dataset splits: 8/2 for training/test
- Comparison methods:
  - Static methods: DeepWalk & Node2Vec
  - Dynamic methods: HTNE, MMDNE, DynamicTriad, MPNN, STAR, EvolveGCN, SpikeNet & ROLAND
- Metrics: Micro-F1 & Macro-F1
- Task: Temporal Node Classification

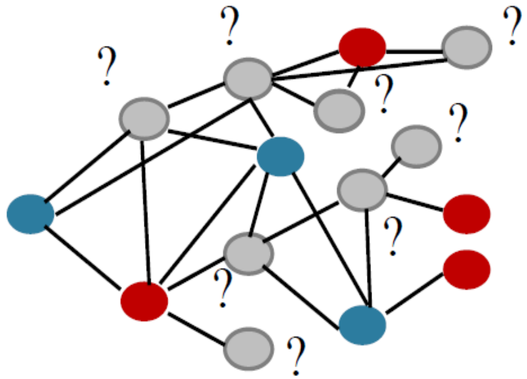
	DBLP-3	Brain	Reddit	DBLP-10	arXiv	Tmall
#Nodes	4,257	5,000	8,291	28,085	169,343	577,314
#Edges	23,540	1,955,488	264,050	236,894	2,315,598	4,807,545
#Features	100	20	20	128	128	128
#Classes	3	10	4	10	40	5
#Time Steps	10	12	10	27	35	186
Category	Citation	Biology	Society	Citation	Citation	E-commerce
$TC_{structure}$	0.139	0.024	0.030	0.823	0.580	0.811
$TC_{feature}$	0.468	0.070	0.556	0.823	1.000	0.712

Table: The statistics of datasets.

$$\mathrm{TC}_{\text{structure}} = \frac{1}{L-1}\sum_{l}^{L-1}\frac{\mathcal{E}_l \cap \mathcal{E}_{l+1}}{\mathcal{E}_l \cup \mathcal{E}_{l+1}},$$

$$\mathrm{TC}_{\text{feature}} = \frac{1}{L-1}\sum_{l}^{L-1}\mathrm{Sim}(X_l, X_{l+1}),$$

$$\text{where} \quad \mathrm{Sim}(X_l, X_{l+1}) = \frac{1}{N_V}\sum_{v\in V}\frac{\langle x_{l,v}, x_{l+1,v}\rangle}{\|x_{l,v}\|\,\|x_{l+1,v}\|}.$$

t_0 →  → t_n

EXPERIMENTS TEMPORAL NODE CLASSIFICATION

	DBLP-3		Brain		Reddit		DBLP-10	
	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1
DeepWalk [36]	$47.53_{\pm 0.4}$	$47.21_{\pm 0.2}$	$51.45_{\pm 0.6}$	$51.03_{\pm 0.8}$	$26.82_{\pm 0.6}$	$26.75_{\pm 0.4}$	66.38	67.12
Node2Vec [9]	$48.79_{\pm 0.3}$	$48.42_{\pm 0.4}$	$53.51_{\pm 0.5}$	$52.95_{\pm 0.6}$	$25.47_{\pm 0.6}$	$25.44_{\pm 0.5}$	67.31	66.93
HTNE [54]	$48.98_{\pm 0.2}$	$48.74_{\pm 0.3}$	$22.31_{\pm 0.8}$	$22.12_{\pm 0.5}$	$26.96_{\pm 0.5}$	$26.80_{\pm 0.7}$	68.79	68.36
M^2DNE [28]	$49.12_{\pm 0.5}$	$48.87_{\pm 0.4}$	$23.79_{\pm 0.5}$	$23.54_{\pm 0.6}$	$25.79_{\pm 0.6}$	$25.61_{\pm 0.4}$	69.71	69.75
DynamicTriad [50]	$48.78_{\pm 0.5}$	$48.63_{\pm 0.6}$	$21.71_{\pm 0.7}$	$21.94_{\pm 0.7}$	$28.76_{\pm 0.5}$	$28.51_{\pm 0.5}$	66.95	66.42
MPNN [32]	$81.78_{\pm 0.6}$	$81.46_{\pm 1.2}$	$90.97_{\pm 1.4}$	$91.01_{\pm 1.5}$	$40.85_{\pm 1.3}$	$40.64_{\pm 1.2}$	$67.74_{\pm 0.3}$	$65.05_{\pm 0.5}$
STAR [47]	**$84.74_{\pm 1.0}$**	**$84.20_{\pm 1.2}$**	$92.08_{\pm 1.3}$	$92.23_{\pm 1.3}$	$43.42_{\pm 2.3}$	$43.43_{\pm 2.4}$	$72.98_{\pm 1.5}$	$72.03_{\pm 1.2}$
tNodeEmbed [39]	$84.51_{\pm 1.2}$	$83.57_{\pm 1.1}$	**$92.35_{\pm 0.8}$**	**$92.30_{\pm 1.0}$**	$42.11_{\pm 1.8}$	$42.06_{\pm 1.3}$	$74.19_{\pm 1.8}$	$74.23_{\pm 2.2}$
EvolveGCN [34]	$84.01_{\pm 1.5}$	$83.12_{\pm 1.5}$	$92.20_{\pm 1.3}$	$92.00_{\pm 1.0}$	$41.24_{\pm 1.3}$	$41.11_{\pm 1.5}$	$71.32_{\pm 0.5}$	$71.20_{\pm 0.7}$
SpikeNet [25]	$83.92_{\pm 1.5}$	$83.04_{\pm 1.1}$	$92.00_{\pm 1.2}$	$91.97_{\pm 1.2}$	$40.42_{\pm 2.0}$	$40.20_{\pm 2.1}$	$74.86_{\pm 0.5}$	$74.65_{\pm 0.5}$
ROLAND [48]	$84.21_{\pm 1.4}$	$84.06_{\pm 1.5}$	$92.14_{\pm 1.2}$	$91.85_{\pm 1.1}$	**$44.22_{\pm 2.2}$**	**$44.25_{\pm 1.9}$**	**$75.01_{\pm 1.1}$**	**$74.98_{\pm 1.0}$**
GRAPHSSM	**$85.26_{\pm 0.9}$**	**$85.00_{\pm 1.3}$**	**$93.52_{\pm 1.0}$**	**$93.54_{\pm 0.9}$**	**$49.21_{\pm 0.5}$**	**$49.05_{\pm 0.7}$**	**$76.80_{\pm 0.3}$**	**$76.00_{\pm 0.4}$**

Table: Temporal node classification performance.

GraphSSM achieves superior performance in the node classification task

EXPERIMENTS TEMPORAL NODE CLASSIFICATION

NEURAL INFORMATION PROCESSING SYSTEMS

	arXiv		Tmall	
	Micro-F1	Macro-F1	Micro-F1	Macro-F1
DeepWalk [36]	$66.54_{\pm 0.3}$	$43.01_{\pm 0.3}$	57.88	49.53
Node2Vec [9]	$67.71_{\pm 0.5}$	$43.69_{\pm 0.4}$	60.66	54.58
HTNE [54]	$65.48_{\pm 0.3}$	$42.25_{\pm 0.3}$	62.64	54.93
M^2DNE [28]	$66.91_{\pm 0.5}$	$43.52_{\pm 0.6}$	64.65	58.47
DynamicTriad [50]	$61.10_{\pm 0.2}$	$38.25_{\pm 0.3}$	60.72	51.16
MPNN [32]	$64.68_{\pm 1.7}$	$41.22_{\pm 1.5}$	$58.07_{\pm 0.6}$	$50.27_{\pm 0.5}$
STAR [47]	OOM	OOM	OOM	OOM
tNodeEmbed [39]	OOM	OOM	OOM	OOM
EvolveGCN [34]	$65.17_{\pm 1.4}$	$43.01_{\pm 1.3}$	$61.77_{\pm 0.6}$	$55.78_{\pm 0.6}$
SpikeNet [25]	$66.69_{\pm 0.9}$	$43.96_{\pm 1.0}$	**$66.10_{\pm 0.3}$**	**$62.40_{\pm 0.6}$**
ROLAND [48]	**$68.27_{\pm 1.2}$**	**$48.01_{\pm 1.3}$**	OOM	OOM
GRAPHSSM	**$70.67_{\pm 0.7}$**	**$49.97_{\pm 0.5}$**	**$66.29_{\pm 0.1}$**	**$62.57_{\pm 0.1}$**

Table: Temporal node classification performance on large graphs.

GraphSSM scales to large and long-range temporal graphs

EXPERIMENTS ABLATION STUDY

NEURAL INFORMATION PROCESSING SYSTEMS

	DBLP-3		Brain		Reddit		DBLP-10	
	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1
GraphSSM-S4	**85.26**$_{\pm 0.9}$	**85.00**$_{\pm 1.3}$	**93.52**$_{\pm 1.0}$	**93.54**$_{\pm 0.9}$	**49.21**$_{\pm 0.5}$	**49.05**$_{\pm 0.7}$	**76.80**$_{\pm 0.3}$	**76.00**$_{\pm 0.4}$
GraphSSM-S5	84.32$_{\pm 1.5}$	83.57$_{\pm 1.9}$	92.20$_{\pm 0.6}$	92.05$_{\pm 0.7}$	**44.75**$_{\pm 0.4}$	**44.79**$_{\pm 0.4}$	73.19$_{\pm 0.6}$	72.95$_{\pm 0.4}$
GraphSSM-S6	**85.74**$_{\pm 0.5}$	**85.23**$_{\pm 0.6}$	**93.80**$_{\pm 0.3}$	**94.47**$_{\pm 0.6}$	42.52$_{\pm 0.9}$	41.73$_{\pm 1.1}$	**75.26**$_{\pm 0.3}$	**74.81**$_{\pm 0.2}$

S4 is the best architecture for learning over temporal graph sequences

EXPERIMENTS ABLATION STUDY

- ❑Feature Mixing: $\widehat{X}_l^{(\mathrm{F})} = \mathrm{GNN}_\theta\left(\mathrm{Mix}_\phi\left(X_{l-1}, X_l\right), G_l\right)$
- ❑Representation Mixing: $\widehat{X}_l^{(\mathrm{R})} = \mathrm{Mix}_\phi\left(\mathrm{GNN}_\theta\left(X_{l-1}, G_{l-1}\right), \mathrm{GNN}_\theta\left(X_l, G_l\right)\right)$

	DBLP-3		Brain		Reddit		DBLP-10	
	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1
$\widehat{X}_1^{(\mathrm{O})} + \widehat{X}_2^{(\mathrm{O})}$	$84.51_{\pm 0.9}$	$84.28_{\pm 0.9}$	$91.56_{\pm 1.1}$	$91.99_{\pm 0.7}$	$48.05_{\pm 2.8}$	$47.99_{\pm 3.0}$	$75.62_{\pm 0.5}$	$74.65_{\pm 0.6}$
$\widehat{X}_1^{(\mathrm{F})} + \widehat{X}_2^{(\mathrm{O})}$	$\mathbf{85.12}_{\pm 0.5}$	$\mathbf{84.82}_{\pm 0.3}$	$\mathbf{92.36}_{\pm 0.8}$	$92.54_{\pm 0.9}$	$\mathbf{49.06}_{\pm 1.9}$	$\mathbf{49.06}_{\pm 1.8}$	$\mathbf{76.67}_{\pm 0.6}$	$\mathbf{75.95}_{\pm 0.7}$
$\widehat{X}_1^{(\mathrm{R})} + \widehat{X}_2^{(\mathrm{O})}$	$84.98_{\pm 1.1}$	$84.79_{\pm 1.0}$	$\mathbf{93.52}_{\pm 1.0}$	$\mathbf{93.54}_{\pm 0.9}$	$\mathbf{49.21}_{\pm 0.5}$	$\mathbf{49.05}_{\pm 0.7}$	$\mathbf{77.76}_{\pm 0.5}$	$\mathbf{77.54}_{\pm 0.6}$
$\widehat{X}_1^{(\mathrm{O})} + \widehat{X}_2^{(\mathrm{R})}$	$\mathbf{85.26}_{\pm 0.9}$	$\mathbf{85.00}_{\pm 1.3}$	$91.84_{\pm 1.9}$	$91.88_{\pm 1.7}$	$47.88_{\pm 1.8}$	$47.94_{\pm 1.8}$	$75.41_{\pm 0.7}$	$74.89_{\pm 1.0}$

Feature mixing generally leads to enhanced model performance

CONCLUSION

1. GHiPPO abstraction, a novel construct predicated on the objective of Laplacian regularized online function approximation
2. GraphSSM, a flexible state space framework designed for temporal graphs, addressing the key algorithmic challenge of unobserved graph mutations
3. GraphSSM exhibits advanced performance on temporal graph learning and scales to large temporal graphs with long-range snapshots

Thanks & QA?

Jintang Li

Sun Yat-sen University

https://arxiv.org/abs/2406.00943

https://github.com/EdisonLeeeee/GraphSSM

https://edisonleeeee.github.io/

lijt55@mail2.sysu.edu.cn