



Long-range Propagation through the Graph Wave Equation

Trenta Alessandro, Gravina Alessio, Hoang Tai, Ceni Andrea, Bacciu Davide
Department of Computer Science, University of Pisa

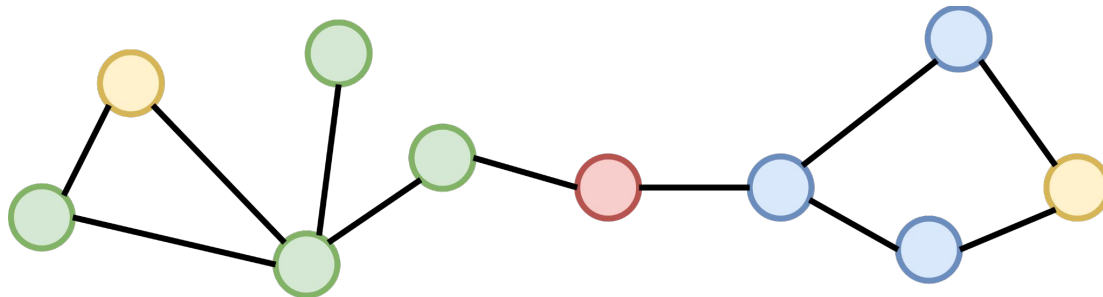
Learning on Graphs Italian Meetup - 10 June 2026



UNIVERSITÀ
DI PISA

The quest for Long-Range Propagation

- Making **distant nodes** communicate effectively is often hard.
- **Computational** and **topological** bottlenecks¹ hinder the performance of GNNs.
- Need for efficient solutions that do not require rewiring or global attentions.



1. Errica et al, "Oversmoothing, "Oversquashing", Heterophily, Long-Range, and more: Demystifying Common Beliefs in Graph Machine Learning", ICLR 2026

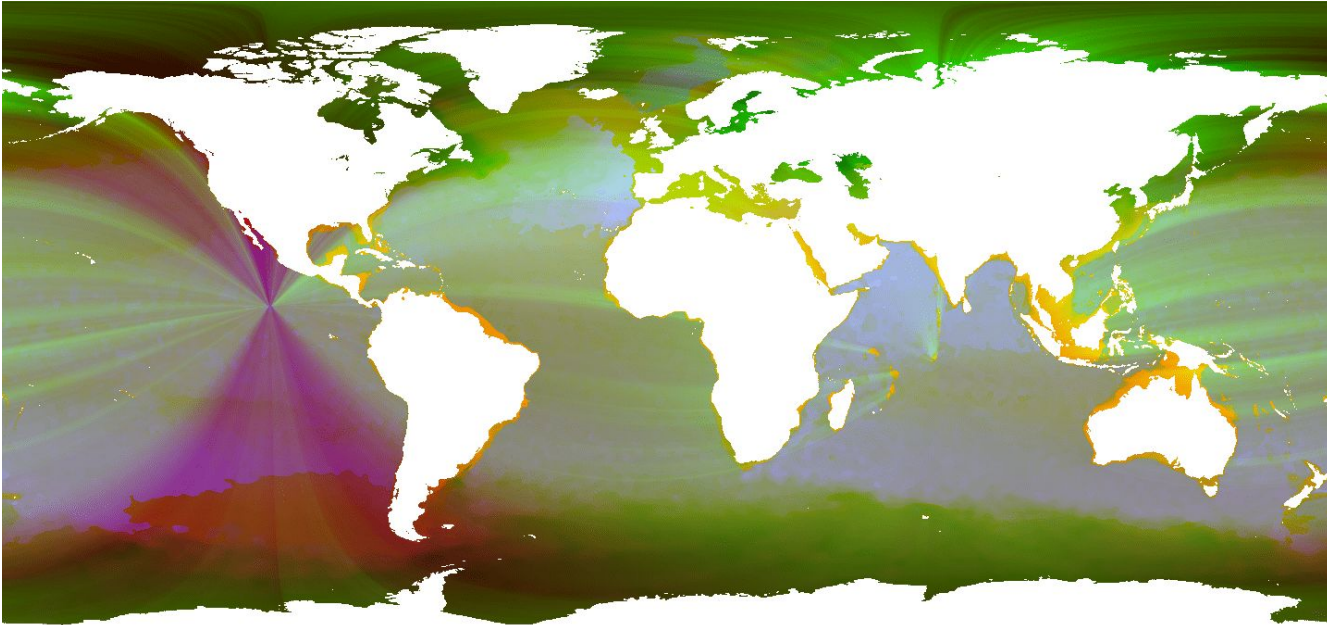
Differential Equation GNNs

- Model node states as a dynamical system through an ODE. $\frac{d\mathbf{X}(t)}{dt} = f(\mathbf{X}(t), \mathcal{G})$
- Input features are encoded in the initial condition. $\mathbf{X}(0) = \bar{\mathbf{X}}$
- Discretize the process and iterate via message passing

$$\mathbf{x}_u^{\ell+1} = \mathbf{x}_u^{\ell} + \Delta t \cdot f_u(\mathbf{x}_u, \cup_{v \in \mathcal{N}_u} \mathbf{x}_v)$$

- We can encode physical priors, e.g. energy conservation, in node dynamics.

The SOFAR Channel



Some nice properties of Waves



In Physics and Mathematics, they naturally describe propagation of information



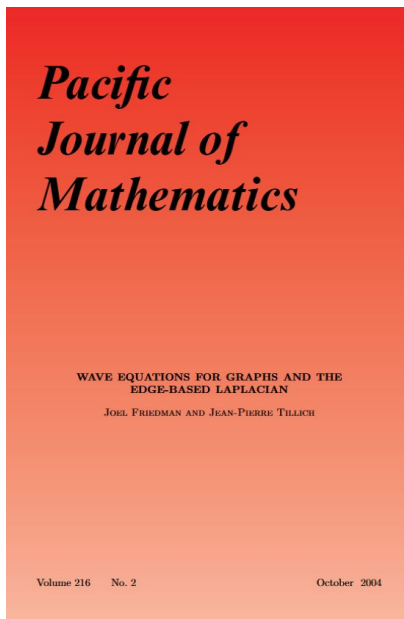
They can travel at constant speed indefinitely, without losing amplitude or energy



They can be linearly (de-)composed, allowing for the superposition of waves



A Mathematically Grounded Formulation



Joel Friedman and Jean-Pierre Tillich. Wave equations for graphs and the edge-based laplacian.

Pacific J. Math., 216(2):229–266, October 2004



UNIVERSITÀ
DI PISA

A Wave equation for Graphs

- We use a mathematically grounded formulation of the wave equation on Graphs

$$\frac{\partial^2 u}{\partial t^2} = c^2 \nabla^2 u$$

$$\ddot{\mathbf{X}}(t) = -\mathbf{L}^a \mathbf{X}(t)$$

A Wave equation for Graphs

We use a mathematically grounded formulation of the wave equation on Graphs.

$$\ddot{\mathbf{X}}(t) = -\mathbf{L}^a \mathbf{X}(t) \mathbf{W} \quad \mathbf{X}(0) = \bar{\mathbf{X}}$$

Here, we use the weighted Laplacian: weights modulate propagation on edges

$$a_{uv} = \text{MLP}(\mathbf{X}_u, \mathbf{X}_v)$$

On a local level:

$$(\mathbf{L}^a \mathbf{X})_v = \sum_{u \in \mathcal{N}(v)} a_{uv} (\mathbf{X}_v - \mathbf{X}_u)$$

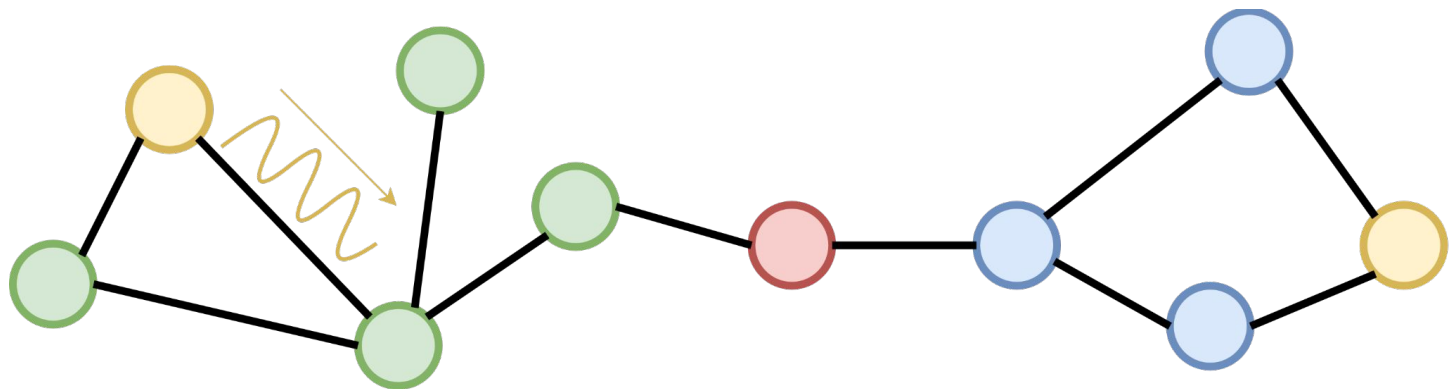
This is an MP operation!

Channel mixing

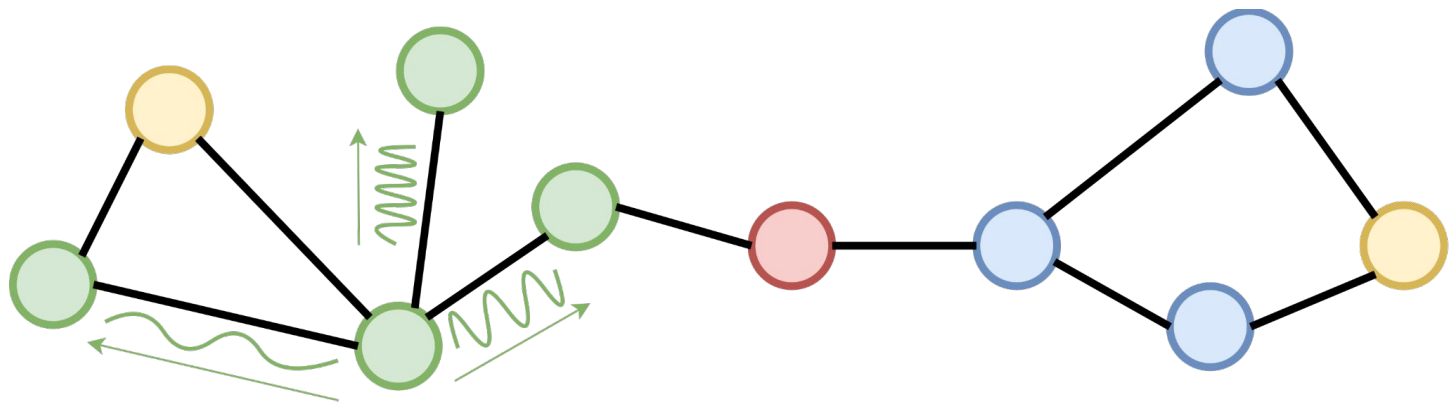
Input encoding



Interpretation

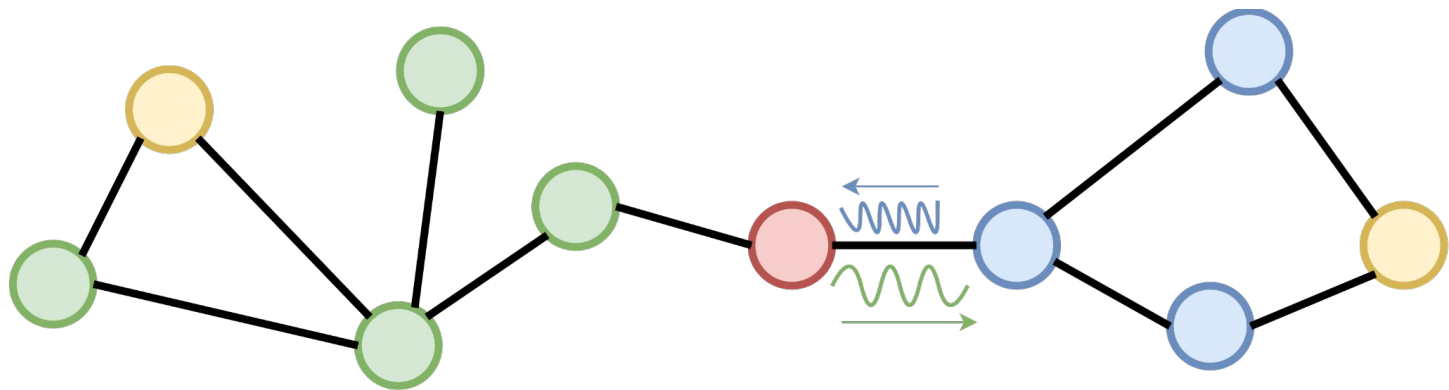


Interpretation



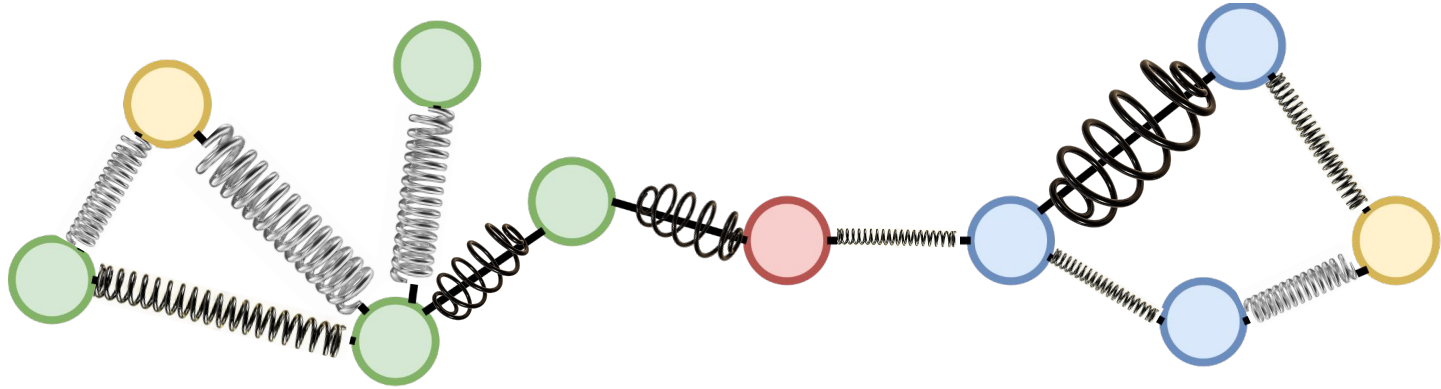
The adaptive resistance modulates how information travels on edges.
It can create paths, eliminate others, filter signals and alleviate computational bottlenecks.

Interpretation



Superposition of waves allows for multiple signals moving on edges, alleviating topological bottlenecks.

Interpretation



The equation $\ddot{\mathbf{X}}(t) = -\mathbf{L}^a \mathbf{X}(t)$ closely resembles the one of oscillator systems

SONAR - full model and Implementation

For additional flexibility, we use a dissipative and external forcing term

$$\ddot{\mathbf{X}}(t) = -\mathbf{L}^a \mathbf{X}(t) \mathbf{W} - D(\mathbf{X}(t)) \odot \dot{\mathbf{X}}(t) + F(\mathbf{X}(t)),$$

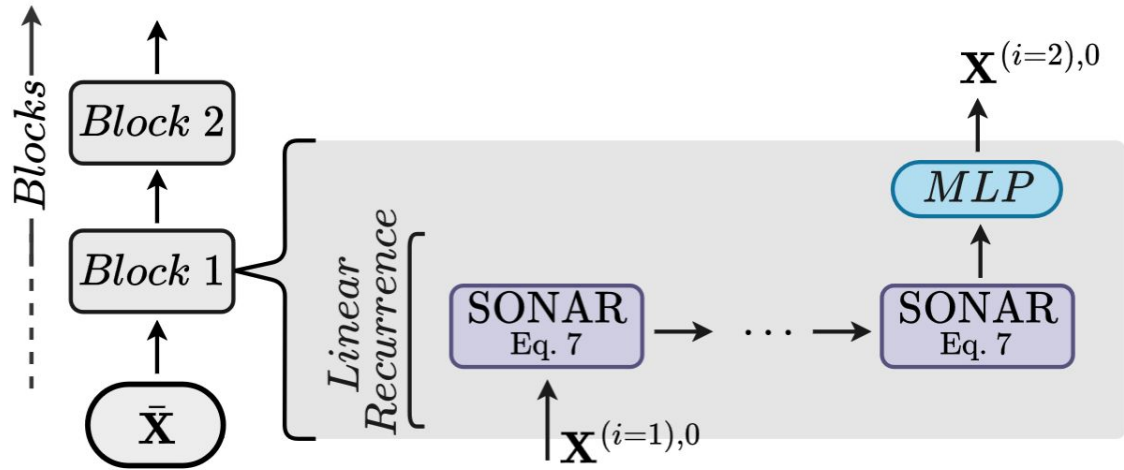
SONAR can be implemented with a simple message passing framework

$$\begin{cases} \mathbf{X}^{\ell+1} = \mathbf{X}^{\ell} + h \mathbf{V}^{\ell+1}, \\ \mathbf{V}^{\ell+1} = \mathbf{V}^{\ell} - h (\mathbf{L}^a \mathbf{X}^{\ell} \mathbf{W} + D(\mathbf{X}^{\ell}) \odot \mathbf{V}^{\ell} - F(\mathbf{X}^{\ell})), \end{cases}$$

Linear recurrence!

We can also stack multiple layers with MLPs in between to add non-linearities.

Deep SONAR architecture



Theoretical properties

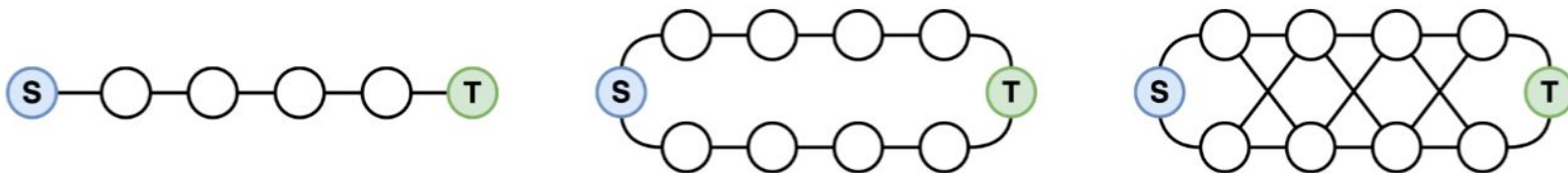
In its conservative form and on a single channel graph:

- Energy is conserved

$$E(t) = \underbrace{\sum_{v \in V} \frac{1}{2} \|\nabla \mathbf{x}_v(t)\|^2}_{\text{potential energy}} + \underbrace{\frac{1}{2} \left\| \frac{\partial \mathbf{X}(t)}{\partial t} \right\|^2}_{\text{kinetic energy}} .$$
- Cross influence oscillates but never vanishes

$$\frac{\partial \mathbf{x}_v(t)}{\partial \mathbf{x}_u(0)} = \cos \left(t \sqrt{\mathbf{L}} \right)_{vu} .$$

Results on Graph Transfer

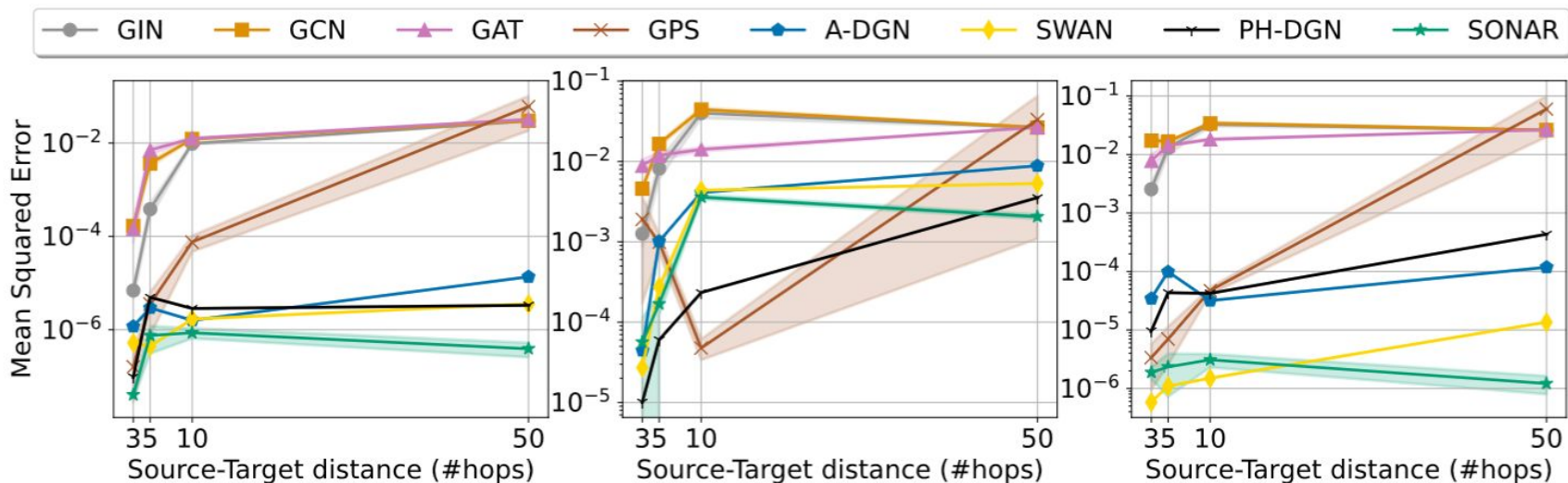


Transport label from source (S) to target (T), the rest of labels is random noise to keep untouched.

Di Giovanni et al. “On over-squashing in message passing neural networks: the impact of width, depth, and topology”, ICML 2023, with the setting of

Gravina et al. “On Oversquashing in Graph Neural Networks Through the Lens of Dynamical Systems”, AAAI 2025

Results on Graph Transfer



Results on Graph Transfer

Table 4: Sensitivity across different distances and recurrences in the Line-50 graph from Section 4.1.

SONAR					GCN				
N. Recurrences→ Distance ↓	25	50	75	100	N. Recurrences→ Distance ↓	25 $\times 10^{-5}$	50 $\times 10^{-5}$	75 $\times 10^{-5}$	100 $\times 10^{-5}$
10	0.0115	0.0413	0.573	1.955	10	0.7078	0.0	1.1470	0.3562
20	0.0037	0.0334	0.703	5.084	20	0.0	0.0	1.0490	0.2980
30	0.0273	0.0616	1.269	4.341	30	0.3725	0.9779	0.5259	0.0745
40	0.0131	0.0253	0.584	1.061	40	0.0	0.0	0.0	0.0
50	0.0002	0.0324	0.143	1.004	50	0.0	0.0	0.0	0.0

Results on Graph Property Prediction

Model	Diameter	SSSP	Eccentricity
MPNNs			
GCN	0.7424 \pm 0.0466	0.9499 \pm 0.0001	0.8468 \pm 0.0028
GAT	0.8221 \pm 0.0752	0.6951 \pm 0.1499	0.7909 \pm 0.0222
GraphSAGE	0.8645 \pm 0.0401	0.2863 \pm 0.1843	0.7863 \pm 0.0207
GIN	0.6131 \pm 0.0990	-0.5408 \pm 0.4193	0.9504 \pm 0.0007
GCNII	0.5287 \pm 0.0570	-1.1329 \pm 0.0135	0.7640 \pm 0.0355
AMP	-0.5891 \pm 0.0720	-3.9579 \pm 0.0769	0.0515 \pm 0.1819
DE-GNNs			
DGC	0.6028 \pm 0.0050	-0.1483 \pm 0.0231	0.8261 \pm 0.0032
GRAND	0.6715 \pm 0.0490	-0.0942 \pm 0.3897	0.6602 \pm 0.1393
GraphCON	0.0964 \pm 0.0620	-1.3836 \pm 0.0092	0.6833 \pm 0.0074
A-DGN	-0.5188 \pm 0.1812	-3.2417 \pm 0.0751	0.4296 \pm 0.1003
SWAN	-0.5981 \pm 0.1145	-3.5425 \pm 0.0830	-0.0739 \pm 0.2190
PH-DGN	-0.5473 \pm 0.1074	-4.2993\pm0.0721	-0.9348 \pm 0.2097
Graph Transformers			
GPS	-0.5121 \pm 0.0426	-3.5990\pm0.1949	0.6077 \pm 0.0282
Multi-hop GNNs			
DRew-GCN	-2.3692\pm0.1054	-1.5905 \pm 0.0034	-2.1004\pm0.0256
+ delay	-2.4018\pm0.1097	-1.6023 \pm 0.0078	-2.0291\pm0.0240
Our			
SONAR	-3.2906\pm0.0706	-6.7517\pm0.0590	-3.1187\pm0.0192

Predict topological properties of a Graph.

Gravina et al, "Anti-Symmetric DGN: a stable architecture for Deep Graph Networks", ICLR 2023

Improvement by orders of magnitude

Results on LRGB

Model	Peptides-func AP \uparrow	Peptides-struct MAE \downarrow	Pascal VOC-SP F1 \uparrow
MPNNs			
GatedGCN	58.64 \pm 0.77	0.3420 \pm 0.0013	0.2873 \pm 0.0219
GCN	59.30 \pm 0.23	0.3496 \pm 0.0013	0.1268 \pm 0.0060
GCNII	55.43 \pm 0.78	0.3471 \pm 0.0010	0.1698 \pm 0.0080
GINE	54.98 \pm 0.79	0.3547 \pm 0.0045	0.1265 \pm 0.0076
Multi-hop GNNs			
DIGL+MPNN+LapPE	68.30 \pm 0.26	0.2616 \pm 0.0018	0.2921 \pm 0.0038
DRew-GCN	69.96 \pm 0.76	0.2781 \pm 0.0028	0.1848 \pm 0.0107
DRew-GCN+LapPE	71.50\pm0.44	0.2536 \pm 0.0015	0.1851 \pm 0.0092
MixHop-GCN	65.92 \pm 0.36	0.2921 \pm 0.0023	0.2506 \pm 0.0133
MixHop-GCN+LapPE	68.43 \pm 0.49	0.2614 \pm 0.0023	0.2218 \pm 0.0174
Transformers			
GraphGPS+LapPE	65.35 \pm 0.41	0.2500 \pm 0.0005	0.3748\pm0.0109
Graph ViT	69.42 \pm 0.75	0.2449\pm0.0016	-
GRIT	69.88 \pm 0.82	0.2460\pm0.0012	-
SAN+LapPE	63.84 \pm 1.21	0.2683 \pm 0.0043	0.3230 \pm 0.0039
Transformer+LapPE	63.26 \pm 1.26	0.2529 \pm 0.0016	0.2694 \pm 0.0098
DE-GNNs			
GRAND	57.89 \pm 0.62	0.3418 \pm 0.0015	0.1918 \pm 0.0097
GraphCON	60.22 \pm 0.68	0.2778 \pm 0.0018	0.2108 \pm 0.0091
A-DGN	59.75 \pm 0.44	0.2874 \pm 0.0021	0.2349 \pm 0.0054
SWAN	67.51 \pm 0.39	0.2485 \pm 0.0009	0.3192 \pm 0.0250
PH-DGN	70.12\pm0.45	0.2465\pm0.0020	-
Ours			
SONAR	68.42 \pm 0.11	0.2525 \pm 0.0038	0.4058\pm0.0039
SONAR+LapPE	70.47\pm0.41	0.2486 \pm 0.0006	0.4082\pm0.0037

Actually, SONAR only needs 4 MP steps for this

On a "true" long-range task, we improve by 8%

Works well on Heterophilic Tasks as well

Model	Roman-empire Acc \uparrow	Amazon-ratings Acc \uparrow	Minesweeper AUC \uparrow	Tolokers AUC \uparrow	Questions AUC \uparrow
MPNNs					
GAT	80.87 \pm 0.30	49.09 \pm 0.63	92.01 \pm 0.68	83.70 \pm 0.47	77.43 \pm 1.20
Gated-GCN	74.46 \pm 0.54	43.00 \pm 0.32	87.54 \pm 1.22	77.31 \pm 1.14	76.61 \pm 1.13
GCN	73.69 \pm 0.74	48.70 \pm 0.63	89.75 \pm 0.52	83.64 \pm 0.67	76.09 \pm 1.27
SAGE	85.74 \pm 0.67	53.63\pm0.39	93.51\pm0.57	82.43 \pm 0.44	76.44 \pm 0.62
Graph Transformers					
Expformer	89.03\pm0.37	53.51\pm0.46	90.74 \pm 0.53	83.77\pm0.78	73.94 \pm 1.06
NAGphormer	74.34 \pm 0.77	51.26 \pm 0.72	84.19 \pm 0.66	78.32 \pm 0.95	68.17 \pm 1.53
GOAT	71.59 \pm 1.25	44.61 \pm 0.50	81.09 \pm 1.02	83.11 \pm 1.04	75.76 \pm 1.66
GPS	82.00 \pm 0.61	53.10\pm0.42	90.63 \pm 0.67	83.71\pm0.48	71.73 \pm 1.47
GPS _{GAT} +Performer (RWSE)	87.04 \pm 0.58	49.92 \pm 0.68	91.08 \pm 0.58	84.38\pm0.91	77.14 \pm 1.49
GT	86.51 \pm 0.73	51.17 \pm 0.66	91.85 \pm 0.76	83.23 \pm 0.64	77.95\pm0.68
GT-sep	87.32\pm0.39	52.18 \pm 0.80	92.29\pm0.47	82.52 \pm 0.92	78.05\pm0.93
Heterophily-Designated GNNs					
FAGCN	65.22 \pm 0.56	44.12 \pm 0.30	88.17 \pm 0.73	77.75 \pm 1.05	77.24 \pm 1.26
FSGNN	79.92 \pm 0.56	52.74 \pm 0.83	90.08 \pm 0.70	82.76 \pm 0.61	78.86\pm0.92
GBK-GNN	74.57 \pm 0.47	45.98 \pm 0.71	90.85 \pm 0.58	81.01 \pm 0.67	74.47 \pm 0.86
GPR-GNN	64.85 \pm 0.27	44.88 \pm 0.34	86.24 \pm 0.61	72.94 \pm 0.97	55.48 \pm 0.91
JacobiConv	71.14 \pm 0.42	43.55 \pm 0.48	89.66 \pm 0.40	68.66 \pm 0.65	73.88 \pm 1.16
Our					
SONAR	89.82\pm0.57	52.22 \pm 0.14	96.29\pm0.73	83.57 \pm 1.44	74.96 \pm 1.10

Working with spatiotemporal signals - STORM

Can we exploit wave-like dynamics for modeling spatiotemporal signals?

- The wave equation naturally provides signal propagation.
- The input signal drives the evolution of the system: **Forcing term!**

$$\ddot{\mathbf{H}}(t) = -\hat{\mathbf{L}}\mathbf{H}(t)\mathbf{W} - D(\mathbf{H}(t)) \odot \dot{\mathbf{H}}(t) + \boxed{F(\mathbf{X}(t), \mathbf{H}(t))}$$

Features of STORM

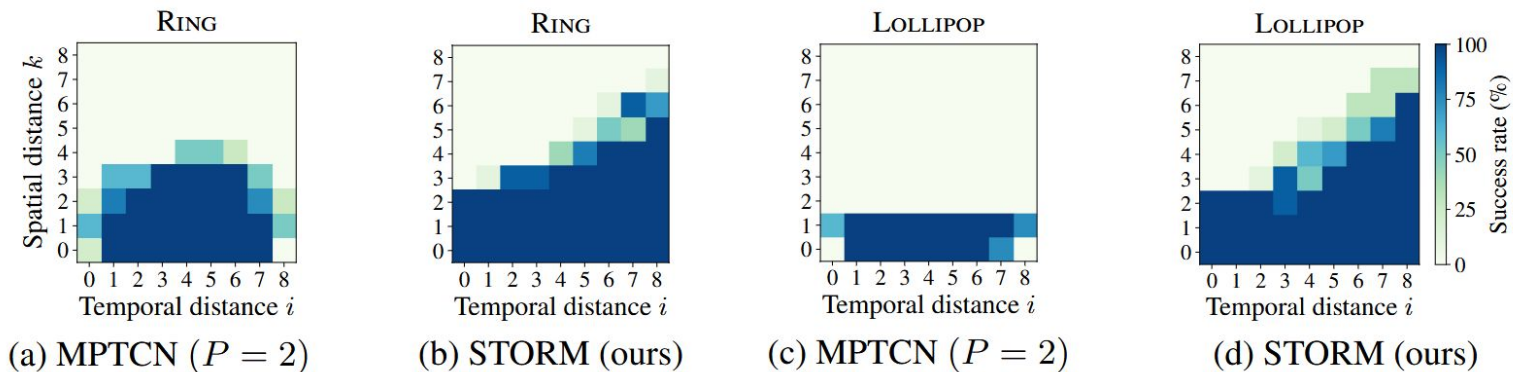
- Efficiency: no product graphs or spectral methods required + MP complexity.
- Feature Filtering: based on the adaptive resistance and dissipation, spectral analysis reveals that STORM can filter irrelevant features.
- STORM is a universal model.

Theorem 3.4 (Universality). *Let $\Phi : C([0, T]; \mathbb{R}^{n \times c}) \rightarrow C([0, T]; \mathbb{R}^{n \times c'})$ be a causal and continuous operator. Let Ψ_θ be the operator that maps the input $\mathbf{X}(t)$ to the solution of Equation (2) with initial conditions given by $\mathbf{H}_0 = \mathbf{X}(0)\mathbf{W}_H$ and $\mathbf{V}_0 = \mathbf{H}_0\mathbf{W}_V$. In other words, Ψ_θ represents STORM. Let $K \subset C([0, T]; \mathbb{R}^{n \times c})$ be compact. Then, for any $\epsilon > 0$, there exist parameters θ such that*

$$\sup_{t \in [0, T]} |\Phi(\mathbf{X})(t) - \Psi_\theta(\mathbf{X})(t)| \leq \epsilon \quad \forall \mathbf{X} \in K. \quad (11)$$

RocketMan Task

Introduced by Marisca et al.: Retrieve the average value at time $t-i$ of nodes distant k hops



Results - Forecasting and Changing Topologies

Model	Metr-LA			PeMS-Bay		
	MAE ↓	RMSE ↓	MAPE ↓	MAE ↓	RMSE ↓	MAPE ↓
HA	6.99	13.89	17.54%	3.31	7.54	7.65%
FC-LSTM	4.37	8.69	14.00%	2.37	4.96	5.70%
SVR	6.72	13.76	16.70%	3.28	7.08	8.00%
VAR	6.52	10.11	15.80%	2.93	5.44	6.50%
AdpSTGCN	3.40	7.21	9.45%	1.92	4.49	4.62%
ASTGCN	6.51	12.52	11.64%	2.61	5.42	6.00%
DCRNN	3.60	7.60	10.50%	2.07	4.74	4.90%
GMAN	3.44	7.35	10.07%	1.86	4.32	4.37%
Graph WaveNet	3.53	7.37	10.01%	1.95	4.52	4.63%
GTS	3.46	7.31	9.98%	1.95	4.43	4.58%
MTGNN	3.49	7.23	9.87%	1.94	4.49	4.53%
RGDAN	3.26	7.02	9.73%	1.82	4.20	4.28%
STAEformer	3.34	7.02	9.70%	1.88	4.34	4.41%
STD-MAE	3.40	7.07	9.59%	1.77	4.20	4.17%
STEP	3.37	6.99	9.61%	1.79	4.20	4.18%
STGCN	4.59	9.40	12.70%	2.49	5.69	5.79%
STSGCN	5.06	11.66	12.91%	2.26	5.21	5.40%
STORM (ours)	3.05	6.30	8.63%	1.59	3.69	3.59%

Twitter Tennis task	
Model	MSE ↓
A3T-GCN	0.477 ± 0.005
DCRNN	0.478 ± 0.004
DyGrAE	0.480 ± 0.005
DynGESN	0.300 ± 0.003
EvolveGCN	0.481 ± 0.003
GCRNN	0.477 ± 0.007
GC-LSTM	0.475 ± 0.010
HMM4G	0.333 ± 0.004
MPNN LSTM	0.482 ± 0.001
T-GCN	0.478 ± 0.004
STORM (ours)	0.302 ± 0.016





Key takeaways

- Oscillatory models are very powerful!
- Linear recurrence and energy conservation (in the right places) help!
- Combine components and let the model decide!
- Use physical and mathematical priors! Nature is efficient and powerful :)

Lanthaler et al. “Neural oscillators are universal”, NeurIPS 2023



Thank you for your attention!



UNIVERSITÀ
DI PISA