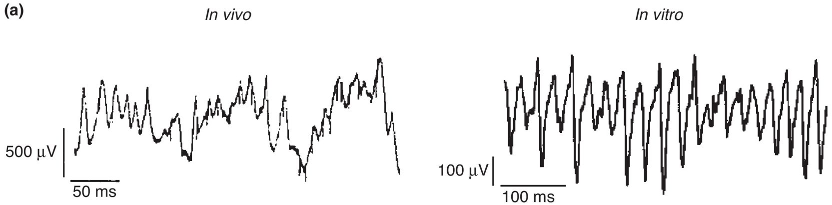
Graph-Coupled Oscillator Networks

T. Konstantin Rusch, Benjamin P. Chamberlain, James Rowbottom, Siddhartha Mishra, Michael M. Bronstein

Oscillators (for GNNs?)

Oscillators are ubiquitous in nature and engineering systems



Why oscillators for GNNs:

- Neurobiological motivation for networks of oscillatory neurons
- Expressivity of oscillators (Fourier series approximation)
- Well-behaved gradients of oscillators $\rightarrow$ Exploding/vanishing gradients problem mitigated?
- Desirable stability properties: A solution for the oversmoothing problem?

Oscillatory inductive bias for GNNs

Set-up:

- Undirected Graph $\mathcal{G} = (\mathcal{V}, \mathcal{E} \subseteq \mathcal{V} \times \mathcal{V})$
- Edges $\mathcal{E}$: pairs of nodes $\{i, j\}$ (denoted $i \sim j$)
- Node features $\mathbf{X}$

GraphCON based on graph dynamical system:

$$\mathbf{X}'' = \sigma(\mathbf{F}_\theta(\mathbf{X}, t)) - \gamma \mathbf{X} - \alpha \mathbf{X}' \iff \begin{cases} \mathbf{Y}' = \sigma(\mathbf{F}_\theta(\mathbf{X}, t)) - \gamma \mathbf{X} - \alpha \mathbf{Y}, \\ \mathbf{X}' = \mathbf{Y} \end{cases}$$

- General learnable 1-neighborhood coupling $\mathbf{F}_\theta$ (e.g. GCN, GAT,...):

$$\left(\mathbf{F}_\theta(\mathbf{X}, t)\right)_{ij} = \mathbf{F}_\theta\left(\mathbf{X}_i(t), \mathbf{X}_j(t), t\right) \quad \forall i \sim j,$$

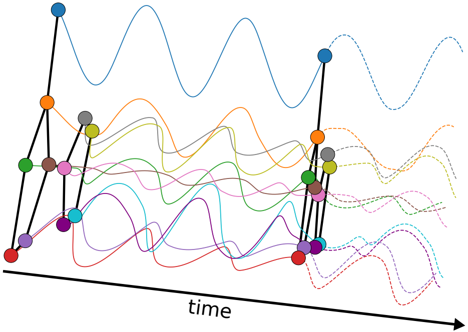
- Activation function σ
- Control parameters $\gamma, \alpha > 0$

IMEX discretizaton yields GraphCON:

$$\mathbf{Y}^n = \mathbf{Y}^{n-1} + \Delta t[\sigma(\mathbf{F}_\theta(\mathbf{X}^{n-1}, t^{n-1})) - \gamma \mathbf{X}^{n-1} - \alpha \mathbf{Y}^{n-1}],$$
$$\mathbf{X}^n = \mathbf{X}^{n-1} + \Delta t \mathbf{Y}^n,$$

for $n = 1, \ldots, N$, and

- $\Delta t > 0$ time-step
- $\mathbf{X}^n, \mathbf{Y}^n$ hidden node features at time $t^n = n\Delta t$

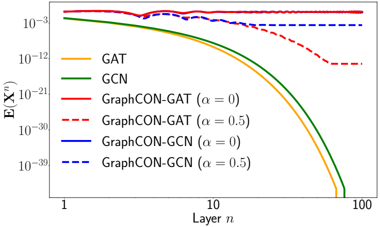


Properties of GraphCON

We provide mathematical definition of oversmoothing: exponential convergence to zero of layer-wise Dirichlet energy

Main result:

GraphCON mitigates the oversmoothing problem



We further show:

- GraphCON mitigates the exploding gradients problem
- GraphCON mitigates the vanishing gradients problem
- Naive multi-layer stacking of GNNs corresponds to fixed-point iteration of GraphCON $\rightarrow$ higher expressive power of GraphCON

Table: Transductive node classification test accuracy (MAP in %) on heterophilic datasets.

Homophily level	**Texas** **0.11**	**Wisconsin** **0.21**	**Cornell** **0.30**
GPRGNN	78.4 ± 4.4	82.9 ± 4.2	80.3 ± 8.1
H2GCN	$\mathbf{84.9 \pm 7.2}$	$\mathbf{87.7 \pm 5.0}$	$\mathbf{82.7 \pm 5.3}$
GCNII	77.6 ± 3.8	80.4 ± 3.4	77.9 ± 3.8
Geom-GCN	66.8 ± 2.7	64.5 ± 3.7	60.5 ± 3.7
PairNorm	60.3 ± 4.3	48.4 ± 6.1	58.9 ± 3.2
GraphSAGE	$\mathbf{82.4 \pm 6.1}$	81.2 ± 5.6	76.0 ± 5.0
MLP	80.8 ± 4.8	85.3 ± 3.3	81.9 ± 6.4
GAT	52.2 ± 6.6	49.4 ± 4.1	61.9 ± 5.1
GraphCON-GAT	82.2 ± 4.7	$\mathbf{85.7 \pm 3.6}$	$\mathbf{83.2 \pm 7.0}$
GCN	55.1 ± 5.2	51.8 ± 3.1	60.5 ± 5.3
GraphCON-GCN	$\mathbf{85.4 \pm 4.2}$	$\mathbf{87.8 \pm 3.3}$	$\mathbf{84.3 \pm 4.8}$

Table: Test accuracy in % on MNIST Superpixel 75.

Model	Test accuracy
ChebNet	75.62
MoNet	91.11
PNCNN	**98.76**
SplineCNN	95.22
GIN	97.23
GraphCON-GIN	98.53
GatedGCN	97.95
GraphCON-GatedGCN	98.27
GCN	88.89
GraphCON-GCN	**98.68**
GAT	96.19
GraphCON-GAT	**98.91**

Table: Test mean absolute error on ZINC (**without edge features, small 12k version**) restricted to small network sizes of $\sim 100k$ parameters.

Model	Test MAE
GIN	0.41 ± 0.008
GatedGCN	0.42 ± 0.006
GraphSAGE	0.41 ± 0.005
MoNet	0.41 ± 0.007
PNA	$\mathbf{0.32 \pm 0.032}$
DGN	$\mathbf{0.22 \pm 0.010}$
GCN	0.47 ± 0.002
GraphCON-GCN	$\mathbf{0.22 \pm 0.004}$
GAT	0.46 ± 0.002
GraphCON-GAT	$\mathbf{0.23 \pm 0.004}$

Conclusion / Outlook

- GraphCON: physics-inspired framework to construct very deep GNNs
- GraphCON provably overcomes the oversmoothing problem as well as exploding/vanishing gradient problem
- GraphCON reaches SOTA on a variety of different graph learning tasks

Main message: "Don't stack GNNs naively – use GraphCON!"

Future projects:

- Physics-inspired methods work – GraphCON is only the start

trusch@ethz.ch | github.com/tk-rusch | @tk_rusch