Adaptive Message Passing: a General Framework to Mitigate Oversmoothing, “Oversquashing”, and Underreaching

Federico Errica, Henrik Christiansen, Viktor Zaverkin,
Takashi Maruyama, Mathias Niepert, Francesco Alesiani

NEC Laboratories Europe & University of Stuttgart

ICML International Conference On Machine Learning

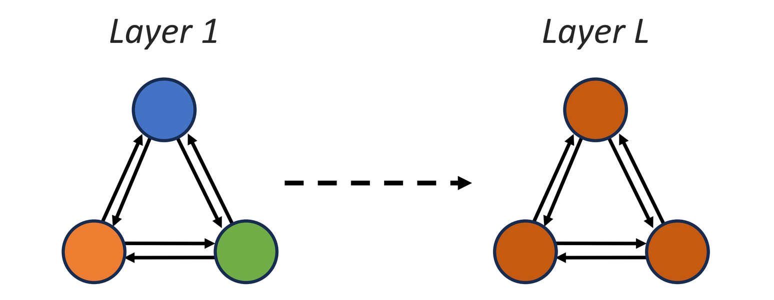
Paper *Code*

NEC

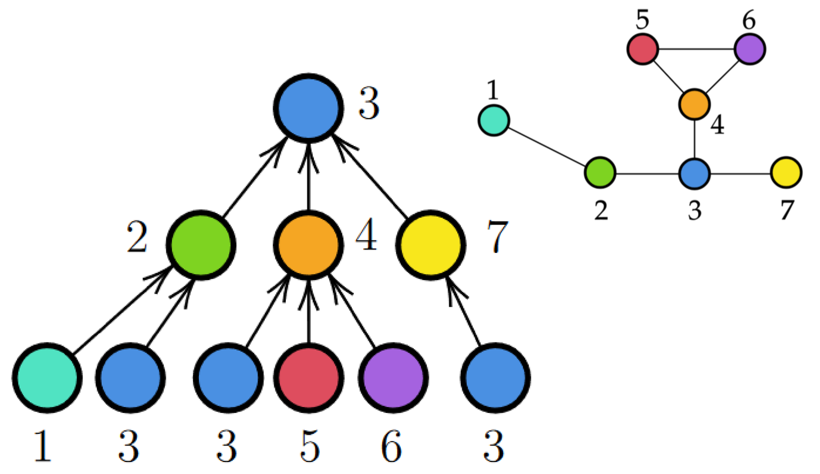
Universität Stuttgart

Technical Issues of Message Passing

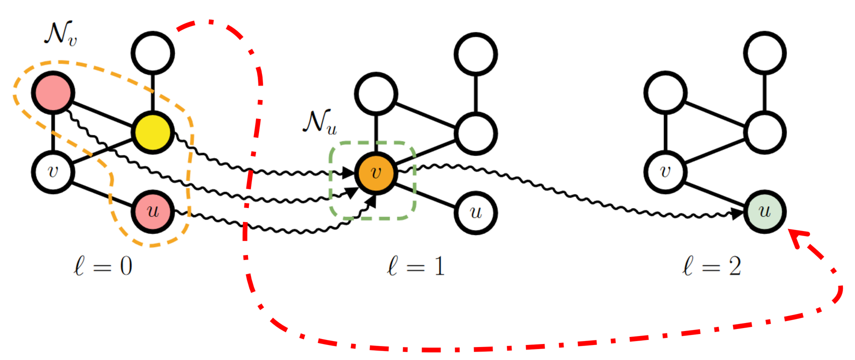
Oversmoothing: node representations converge to the same value



Computational Bottlenecks: exponential amount of info needs to be compressed into a node representation

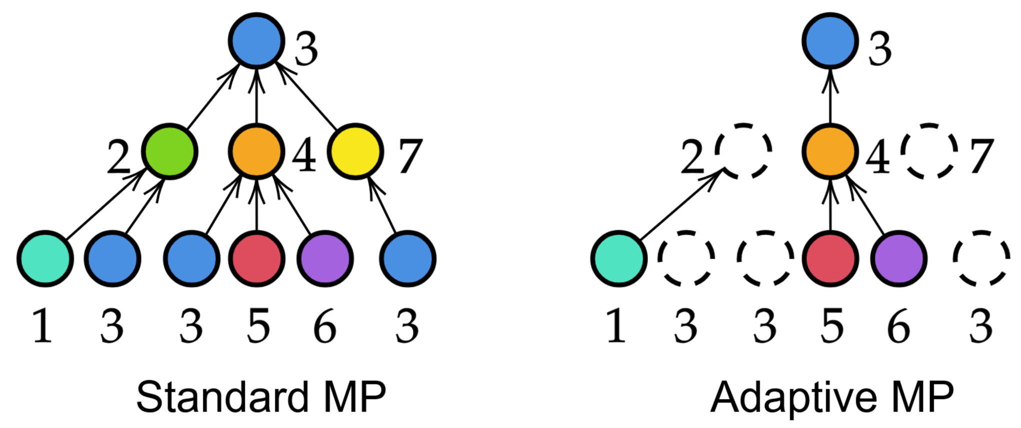


Underreaching: we use less layers than needed (more layers increase the 2 other problems)



Our Solutions

Learn to **Filter Messages**



$$\boldsymbol{h}_v^{\ell} = \phi^{\ell}\left(\boldsymbol{h}_v^{\ell-1},\ \Psi(\{\boldsymbol{F}_i(u,\ell-1)\odot\psi^{\ell}(\boldsymbol{h}_u^{\ell-1},\boldsymbol{a}_{uv}^{\ell})\mid u\in\mathcal{N}_v\})\right)$$

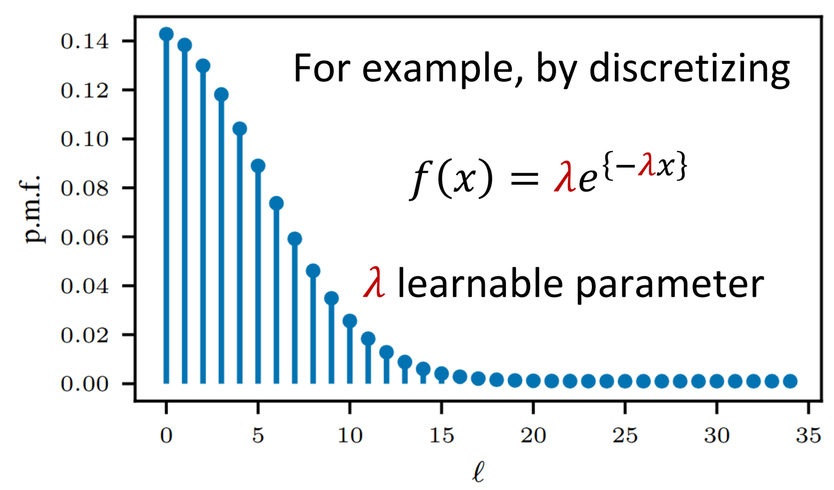
Node Embedding *Permutation Invariant Function* *Learnable Functions* *Filtering Function*

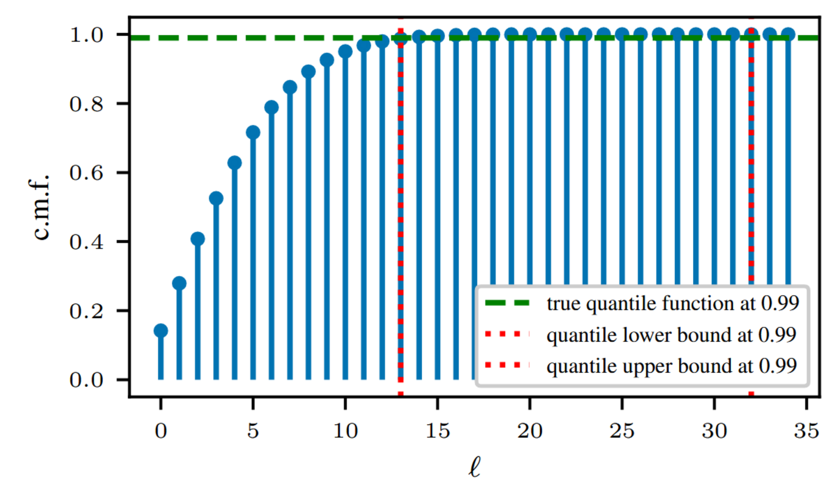
Theorem 3.1. *For AMP with m layers and $u, v \in \mathcal{V}$,*

$$\left\|\frac{\partial h_v^{(m)}}{\partial h_u^{(0)}}\right\|_{L^1} \le d\left(\left(c_{\mathrm{up}}\left(c_{\mathrm{rs}} I + c_{\mathrm{mp}}\left(c_F k_h + k_F\right) A\right)\right)^m\right)_{vu}$$

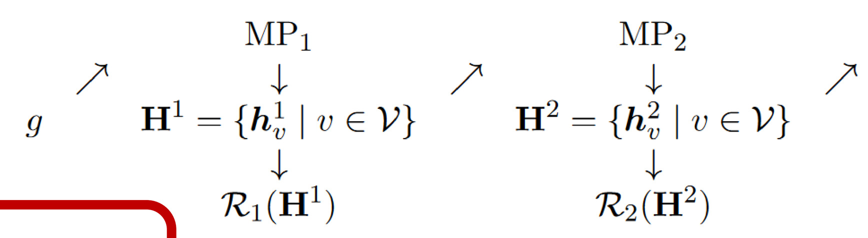
Learn the Depth via simple Backpropagation

1. **Define importance distribution** over (infinite) layers
2. **Dynamically Truncate** the distribution to a finite value based on λ (using the quantile function)
3. **Every** layer's **output** prediction **weighted by importance** – needs 1 readout per layer





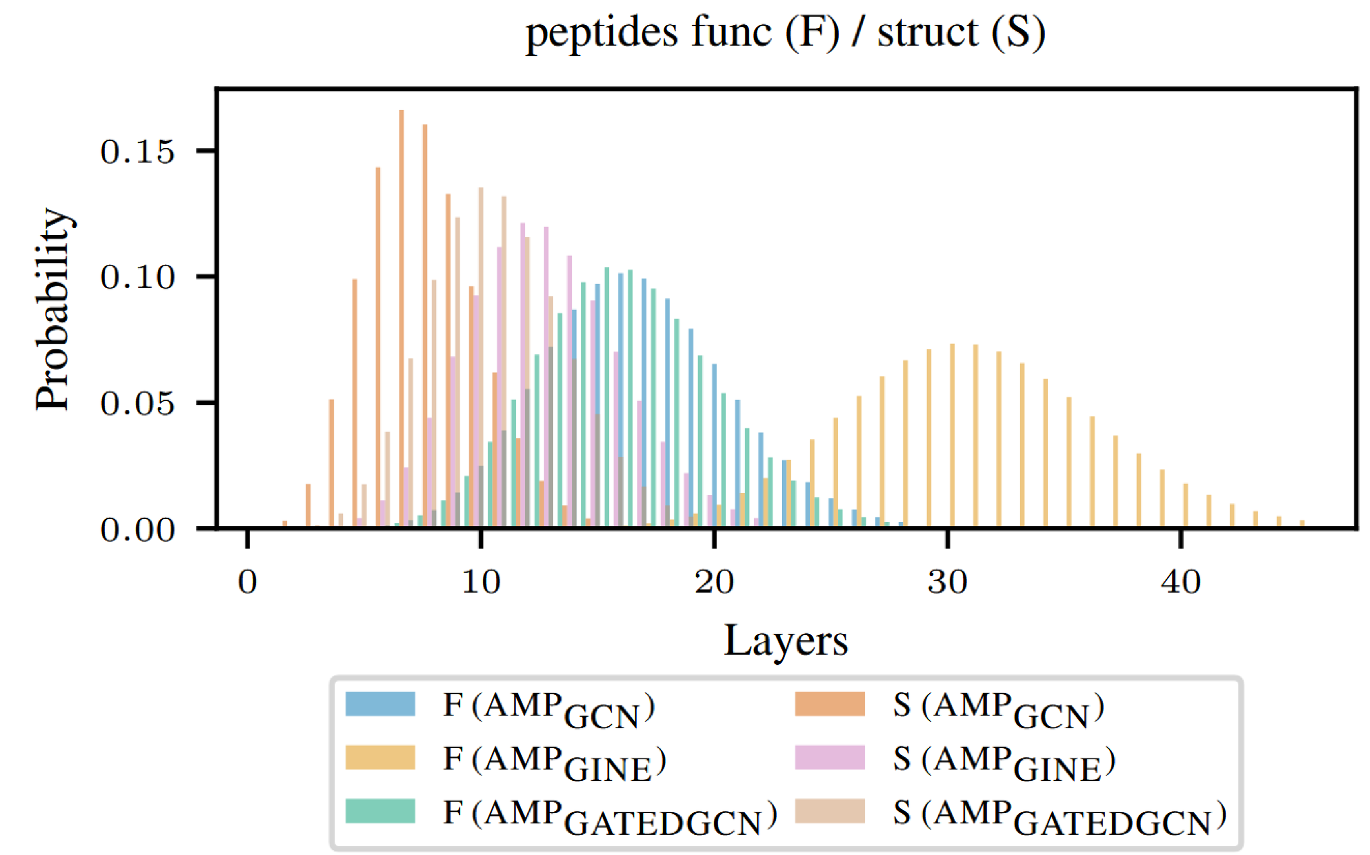
Deep Graph Network architecture:



Advantages

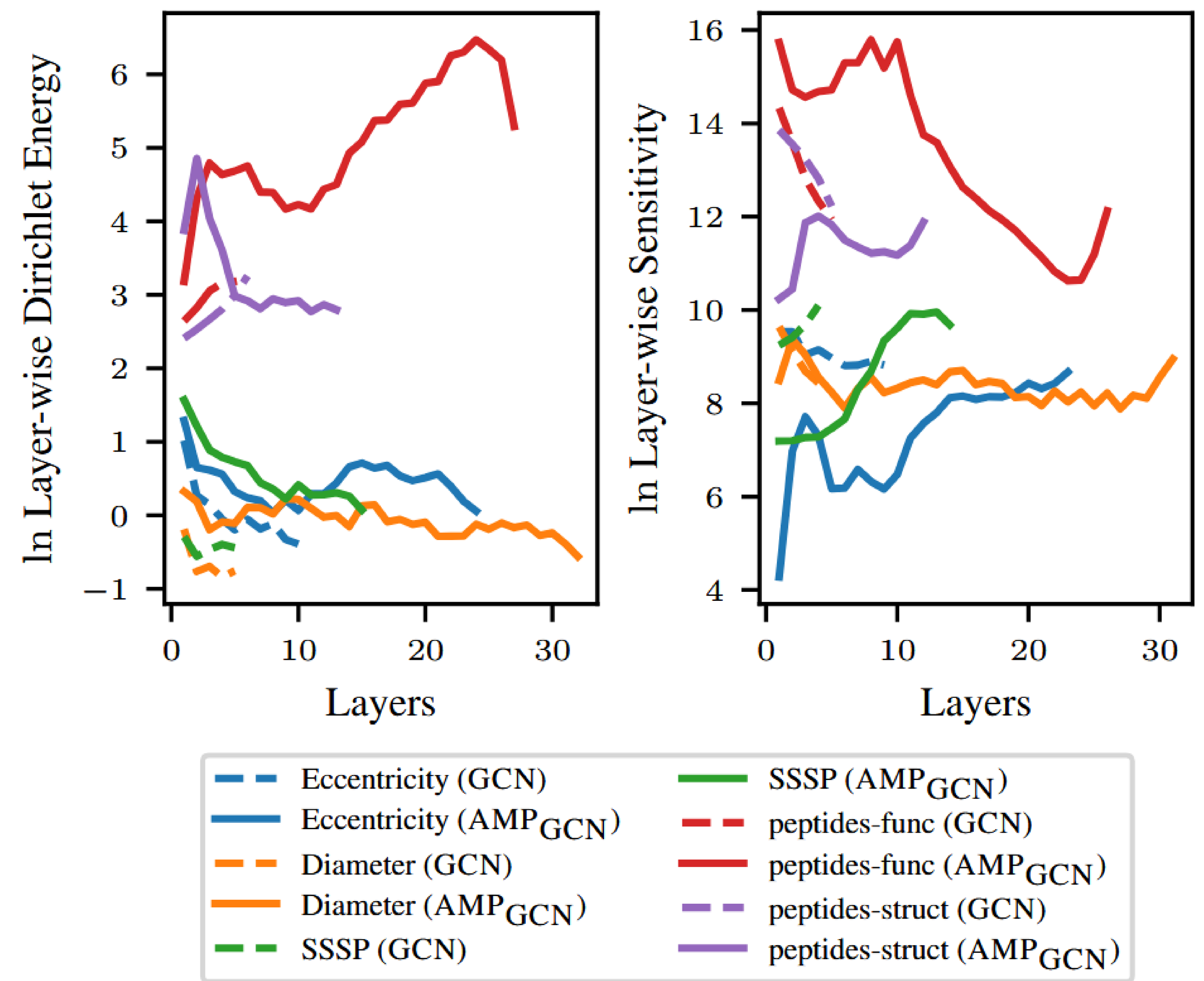
1. The model's **depth grows/shrinks during training** according to the task!

We can encourage deeper/shallower models using a principled prior distribution



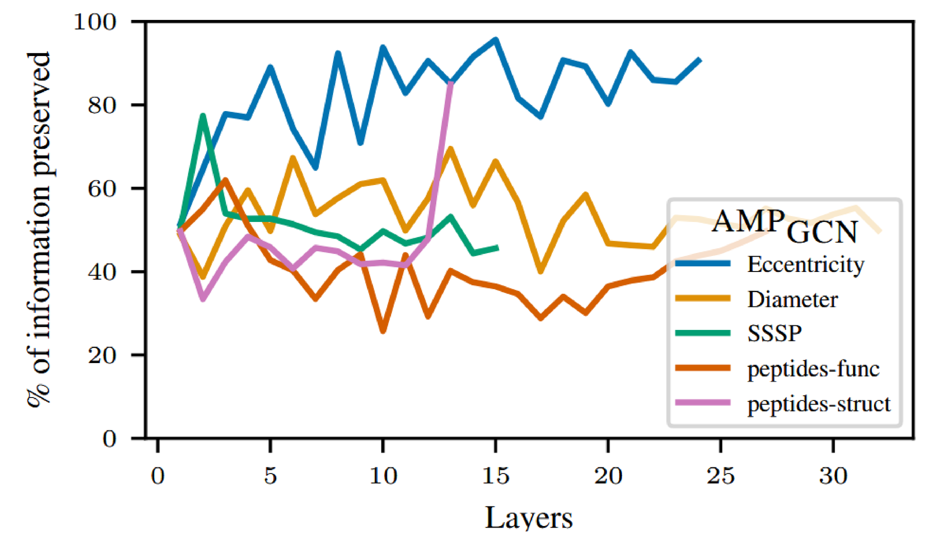
2. The model learns to **control oversmoothing and “oversquashing”**

The amount of information to propagate is learned based on the task



3. **Competitive performance with a classical GCN**

	Method	peptides-func Test AP ↑	peptides-struct Test MAE ↓
LRGB	GCN	0.5930 ± 0.0023	0.3496 ± 0.0013
	GINE	0.5498 ± 0.0079	0.3547 ± 0.0045
	GatedGCN	0.6069 ± 0.0035	0.3357 ± 0.0006
	Transformer	0.6326 ± 0.0126	0.2529 ± 0.0016
	SAN	0.6439 ± 0.0075	0.2545 ± 0.0012
	GPS	0.6535 ± 0.0041	0.2500 ± 0.0005
Re-Eval	GCN	0.6860 ± 0.0050	0.2460 ± 0.0007
	GINE	0.6621 ± 0.0067	0.2473 ± 0.0017
	GatedGCN	0.6765 ± 0.0047	0.2477 ± 0.0009
	GPS	0.6534 ± 0.0091	0.2509 ± 0.0014
Others	CRaWl	0.7074 ± 0.0032	0.2506 ± 0.0022
	$\mathrm{DRew}_{\mathrm{GCN}}$	0.7150 ± 0.0044	0.2536 ± 0.0015
	$\mathrm{DRew}_{\mathrm{GatedGCN}}$	0.6977 ± 0.0026	0.2539 ± 0.0007
	Exphormer	0.6527 ± 0.0043	0.2481 ± 0.0007
	GRIT	0.6988 ± 0.0082	0.2460 ± 0.0012
	Graph ViT	0.6942 ± 0.0075	0.2449 ± 0.0016
	G-MLPMixer	0.6921 ± 0.0054	0.2475 ± 0.0015
	LASER	0.6440 ± 0.0010	0.3043 ± 0.0019
	CO-GNN	0.6990 ± 0.0093	-
	$\mathrm{NBA}_{\mathrm{GCN}}$	0.7207 ± 0.0028	0.2472 ± 0.0008
	$\mathrm{NBA}_{\mathrm{GatedGCN}}$	0.6982 ± 0.0014	0.2466 ± 0.0012
	PH-DGN	0.7012 ± 0.0045	0.2465 ± 0.0020
	GRED	0.7041 ± 0.0049	0.2584 ± 0.0015
	PR-MPNN	0.6825 ± 0.0086	0.2477 ± 0.0005
	IPR-MPNN	0.7210 ± 0.0039	0.2422 ± 0.0007
	$\mathrm{AMP}_{\mathrm{GCN}}$	$0.7161^{\dagger} \pm 0.0047$	$0.2446^{\dagger} \pm 0.0026$
	$\mathrm{AMP}_{\mathrm{GINE}}$	$0.7065^{\dagger} \pm 0.0105$	$0.2468^{\dagger} \pm 0.0026$
	$\mathrm{AMP}_{\mathrm{GatedGCN}}$	$0.6943^{\dagger} \pm 0.0046$	$0.2480^{\dagger} \pm 0.0012$



	Diameter	Rel Imp	SSSP	Rel Imp	Eccentricity	Rel Imp
GCN	0.6146 ± 0.0375		0.9132 ± 0.0051		0.7398 ± 0.0705	
GAT	1.4367 ± 0.3558		0.6070 ± 0.0375		1.0714 ± 0.0616	
GraphSAGE	0.6146 ± 0.0744		-1.0139 ± 0.0120		1.0859 ± 0.0001	
GIN	0.2408 ± 0.0154		-0.2648 ± 0.4437		0.9229 ± 0.0002	
GCNII	0.5057 ± 0.0309		-0.9172 ± 0.4396		0.7112 ± 0.0255	
DGC	0.5601 ± 0.0220		-0.0254 ± 0.0077		0.8051 ± 0.0017	
GRAND	0.9477 ± 0.2160		0.1909 ± 0.3103		0.7450 ± 0.1369	
ADGN	-0.4530 ± 0.0883		-3.5448 ± 0.2749		0.0547 ± 0.0732	
$\mathrm{AMP}_{\mathrm{GCN}}$	$-0.1072^{\dagger} \pm 0.0791$	-81%	$0.5440^{\dagger} \pm 0.0108$	-57%	$0.6054^{\dagger} \pm 0.0919$	-26%
$\mathrm{AMP}_{\mathrm{GIN}}$	$-0.4874^{\dagger} \pm 0.1111$	-81%	$-3.0628^{\dagger} \pm 0.3159$	-99%	$0.4093^{\dagger} \pm 0.0546$	-69%
$\mathrm{AMP}_{\mathrm{ADGN}}$	$-0.5891^{\dagger} \pm 0.0720$	-27%	$-3.9579^{\dagger} \pm 0.0769$	-61%	$0.0515^{\dagger} \pm 0.1819$	-1%
Avg Rel Imp		-63%		-72%		-32%

Oversmoothing, "Oversquashing", Heterophily, Long-Range, and more: **Demystifying Common Beliefs in Graph Machine Learning**

HOT OFF THE PRESS