Spectral Graph Pruning Against Over-Squashing and Over-Smoothing

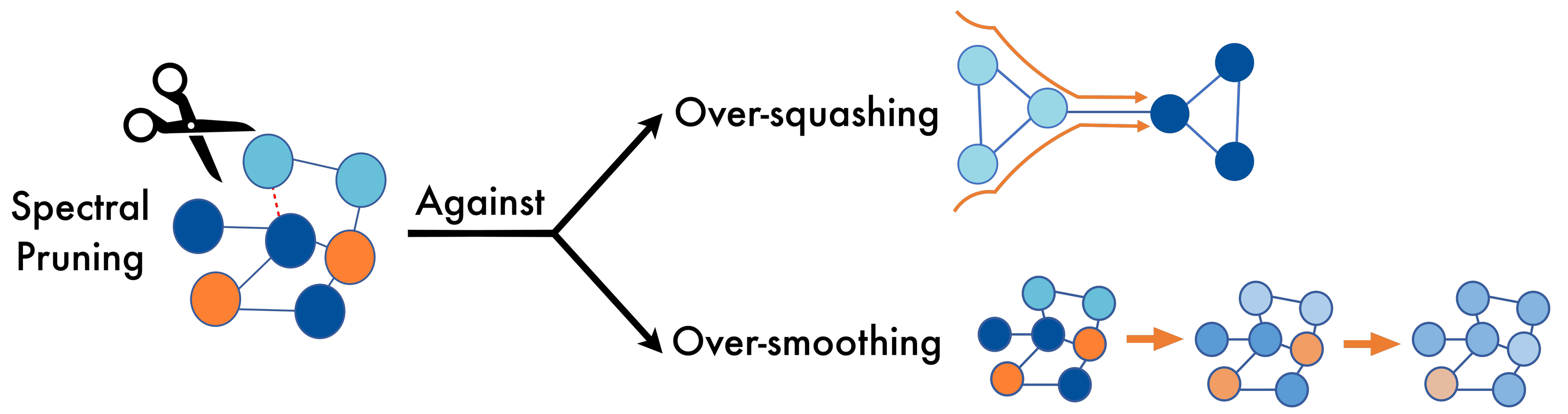
Adarsh Jamadandi* [1,2] Celia Rubio-Madrigal* [2] Rebekka Burkholz[2]

*Equal contribution [1]Universität des Saarlandes [2]CISPA Helmholtz Center for Information Security

Background

GNNs may suffer from two issues: **over-smoothing** (node features become indistinguishable with more layers) and **over-squashing** (restricted information flow via bottlenecks).

Common approach: Rewiring the graph by different criteria, like maximizing the spectral gap by adding edges. However, this can **worsen over-smoothing**, so over-squashing and over-smoothing are usually treated as opposites [1, 2].



Main contribution

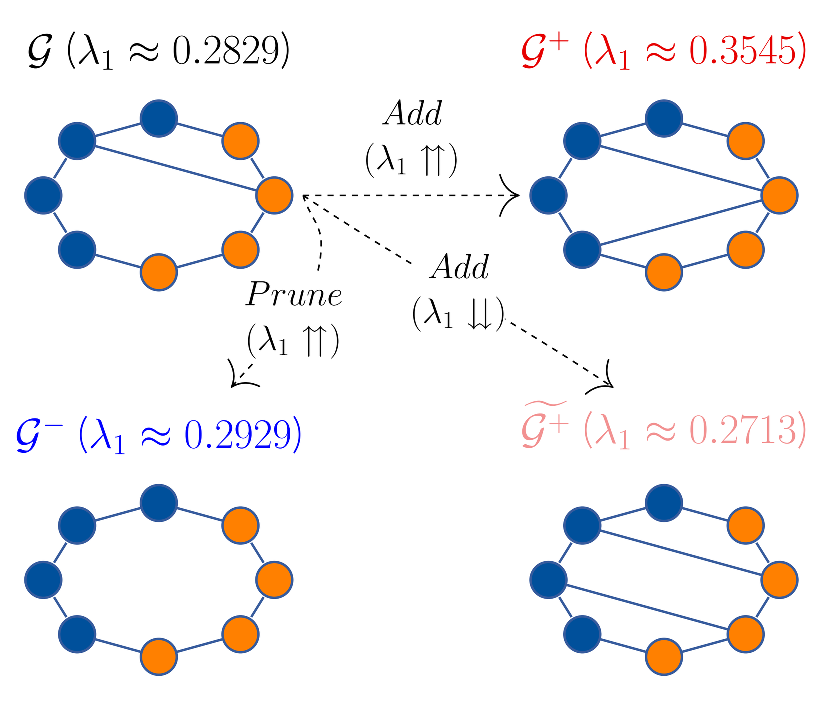
Introducing the Braess paradox: Adding extra capacity to a network can, in some cases, lead to a reduction in overall flow (and viceversa) [3, 4].

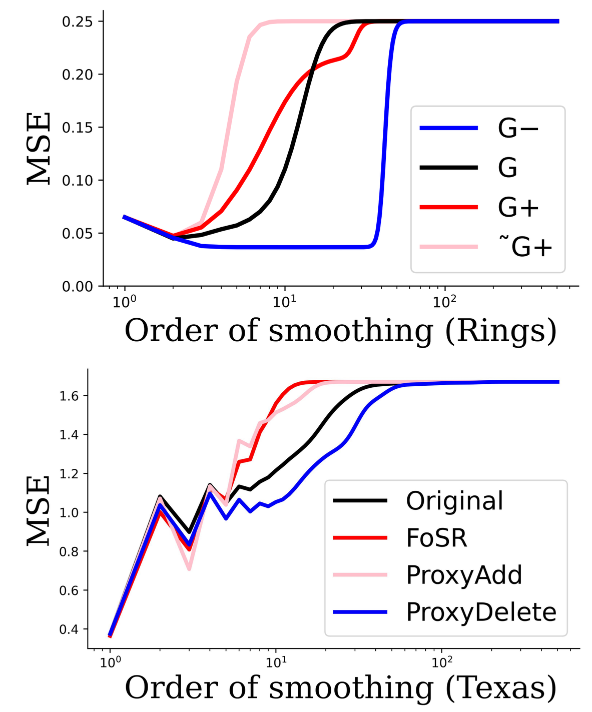
Therefore, we can find **edge deletions** that maximize the spectral gap.

Key idea: Over-smoothing and over-squashing are **not a trade-off**, because maximizing the spectral gap by edge deletions:

1. Helps reduce over-squashing, both theoretically and empirically.
2. Helps reduce over-smoothing, as defined in the testbed by [5] which considers node features in addition to the graph structure.

Trade-off counterexample: $\mathcal{G}^-$ has one fewer edge than $\mathcal{G}$ but a higher spectral gap and a lower rate of smoothing (in **black** vs. blue ↓).





Proposed rewiring methods

1. **EldanAdd/EldanDelete**: Based on a lemma by [4] that states a sufficient condition for the Braess paradox to occur.
2. **ProxyAdd/ProxyDelete**: Better and constant-time approximation of λ using matrix perturbation theory [6, 7]:

$$\acute{\lambda} \approx \lambda + \Delta w_{u,v}((f_u - f_v)^2 - \lambda(f_u^2 + f_v^2))$$

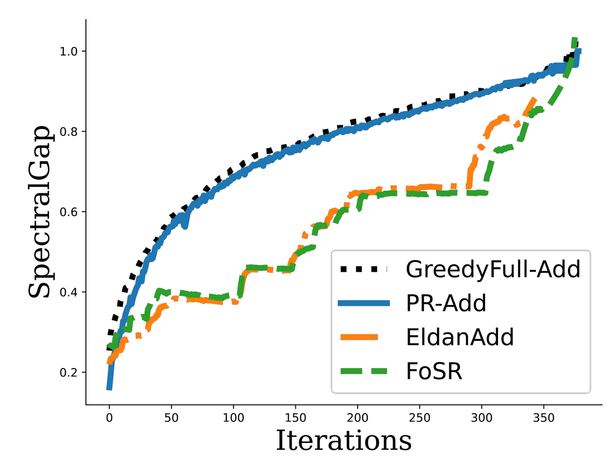


Table 1. Runtimes (in seconds) for 50 edge modifications.

Method	Cora	Citeseer	Chameleon	Squirrel
FoSR [1]	4.69	5.33	5.04	19.48
SDRF [8]	19.63	173.92	17.93	155.95
ProxyAdd	4.30	3.13	1.15	9.12
ProxyDelete	1.18	0.86	1.46	7.26

Experimental evidence

GNN benchmarks: Improvements on the Long Range Graph Benchmark [9] (T2) and on large heterophilic datasets [10] (T3).

Lottery tickets: We can use our methods to find Graph Lottery Tickets. We compare them with UGS [11] (T4). Our methods can provide a stopping criterion, and can be used to perform Pruning at Initialization.

Table 2. Amazon-Ratings.

Method	#EdgesAdded	Accuracy	#EdgesDeleted	Accuracy	Layers
GCN	-	47.20±0.33	-	47.20±0.33	10
GCN+FoSR	25	49.68±0.73	-	-	10
GCN+Eldan	25	48.71±0.99	100	**50.15±0.50**	10
GCN+Proxy	10	49.72±0.41	50	**49.75±0.46**	10
GAT	-	47.43±0.44	-	47.43±0.44	10
GAT+FoSR	25	51.36±0.62	-	-	10
GAT+Eldan	25	51.68±0.60	50	**51.80±0.27**	10
GAT+Proxy	20	49.06±0.92	100	**51.72±0.30**	10
GCN	-	47.32±0.59	-	47.32±0.59	20
GCN+FoSR	100	49.57±0.39	-	-	20
GCN+Eldan	50	**49.66±0.31**	20	48.32±0.76	20
GCN+Proxy	50	49.48±0.59	500	**49.58±0.59**	20
GAT	-	47.31±0.46	-	47.31±0.46	20
GAT+FoSR	100	51.31±0.44	-	-	20
GAT+Eldan	20	51.40±0.36	20	**51.64±0.44**	20
GAT+Proxy	50	47.53±0.90	20	**51.69±0.46**	20

Table 3. Long Range Graph Benchmark.

Method	PascalVOC-SP (Test F1 ↑)	Peptides-Func (Test AP ↑)	Peptides-Struct (Test MAE ↓)
Baseline-GCN	0.1268±0.0060	0.5930±0.0023	0.3496±0.0013
DRew+GCN	0.1848±0.0107	**0.6996±0.0076**	0.2781±0.0028
FoSR+GCN	0.2157±0.0057	0.6526±0.0014	0.2499±0.0006
ProxyAdd+GCN	**0.2213±0.0011**	0.6789±0.0002	**0.2465±0.0004**
ProxyDelete+GCN	0.2170±0.0015	0.6908±0.0007	**0.2470±0.0080**

Table 4. Pruning for lottery tickets.

Method	Cora			Citeseer			Pubmed		
Metrics	GS	WS	Acc	GS	WS	Acc	GS	WS	Acc
UGS	79.85%	97.86%	68.46±1.89	78.10%	97.50%	**66.50±0.60**	68.67%	94.52%	76.90±1.83
EldanDelete+UGS	79.70%	97.31%	68.73±0.01	77.84%	96.78%	64.60±0.00	70.11%	93.17%	**78.00±0.42**
ProxyDelete+UGS	78.81%	97.24%	**69.26±0.63**	77.50%	95.83%	65.43±0.60	78.81%	97.24%	75.25±0.25

Conclusions

1. Over-smoothing and over-squashing are **not necessarily diametrically opposed**: both can be mitigated by spectral based edge deletions.
2. We propose a greedy **graph pruning algorithm** that maximizes the spectral gap in a computationally efficient way. It can also be utilized to add edges.
3. We **connect literature** on three seemingly disconnected topics: over-smoothing, over-squashing, and graph lottery tickets.

References: [1] Karhadkar, K. et al. in *ICLR* (2023). [2] Giraldo, J. H. et al. in *ACM CIKM* (2023). [3] Braess, D. *Unternehmensforschung* **12**, 258–268 (1968). [4] Eldan, R. et al. *Random Structures & Algorithms* **50** (2017). [5] Keriven, N. in *LoG* (2022). [6] Stewart, G. et al. (1990). [7] Bojchevski, A. et al. in *ICML* (2019). [8] Topping, J. et al. in *ICLR* (2022). [9] Dwivedi, V. P. et al. in *NeurIPS* (2022). [10] Platonov, O. et al. in *ICLR* (2023). [11] Chen, T. et al. in *ICML* (2021).

NEURAL INFORMATION PROCESSING SYSTEMS