Learning Curves of GNNs vs. MLP vs. Tikhonov

Calin Radoi

Responsible Professor: Elvin Isufi | Supervisors: Chengen Liu, Mohamed Jebali

EEMCS, Delft University of Technology, The Netherlands

1 Why this matters

- GNNs fuse **node features** + **graph topology** for semi-supervised node classification [1].
- But a higher-capacity model is *not* automatically better: with few labels, topology can **help**, **hurt**, or be **redundant**.
- We trace **learning curves**, test macro-F1 vs. labelled nodes per class n_l [2], not a single benchmark number.

▶ **One score hides data-efficiency; the curve reveals it.**

2 Research question

When do graph structure and node features. . .

- **synergise** implicitly,
- need **explicit** topological priors,
- or become **redundant** / **harmful**?

▶ **Core lens: *implicit* topology (forward-pass message passing) vs. *explicit* topology (a backward-pass smoothness penalty).**

3 Five information regimes

Model	Feat.	Topo.	Expl. prior
MLP (feature-only)	✓		
Tikhonov (topology-only)		✓	✓
ChebNet GNN [3]	✓	✓ impl.	
MLP + Laplacian	✓		✓
GNN + Laplacian	✓	✓ impl.	✓

Topology-only closed form & explicit Laplacian penalty:

$$\hat{\mathbf{Y}} = (\mathbf{I} + \lambda\mathbf{L})^{-1}\mathbf{Y}_0, \quad \mathcal{L}_{reg} = \frac{\lambda}{|E|}\sum_{(u,v)\in E}\|\mathbf{p}_u - \mathbf{p}_v\|_2^2$$

Diagnostic – **Dirichlet energy** of labels (smoothness):

$$\mathcal{E} = \frac{1}{|E|}\sum_{(u,v)\in E}\|\mathbf{Y}_u - \mathbf{Y}_v\|_2^2$$

▶ **One task, five regimes ⇒ isolate what topology contributes.**

4 Datasets & setup

Dataset	Dirichlet $\mathcal{E}$	Avg. deg.	Topology
Cora [4]	0.38	∼3.9	Homophilic / sparse
PubMed [4]	0.40	∼4.5	Homophilic / sparse
Chameleon [5]	1.53	∼19.9	Heterophilic / med.
Squirrel [5]	1.59	∼42.3	Heterophilic / dense

- 2 layers, 64 hidden; ChebNet $K{=}3$; $\lambda{=}1.0$; Adam (lr 0.01); n_l : $5 \to 1280$, ∼20 runs/budget; filtered splits.
- **Fair baseline:** tuned on MLP then frozen; λ swept to guarantee optimal prior.

5 Homophily: structure wins — until features do

Q: Does graph structure help on smooth ($\mathcal{E} < 0.4$) graphs?

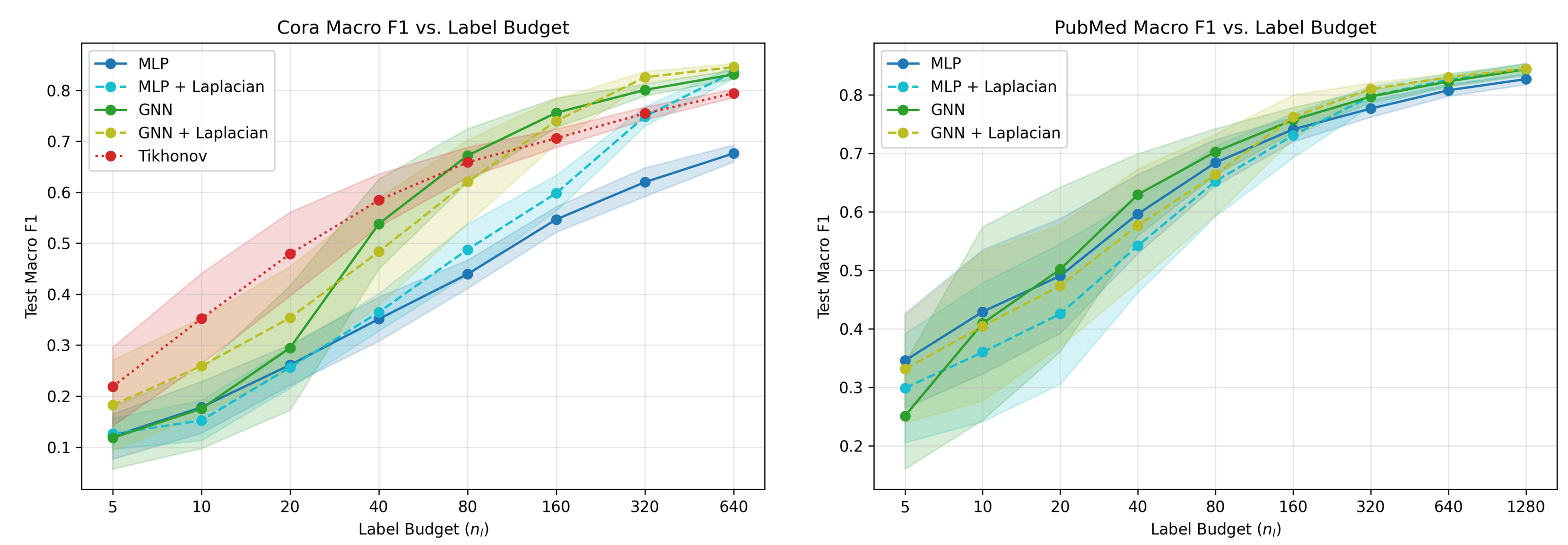

Fig. 1: Test macro-F1 vs. label budget. **Cora** (left): a large, persistent *structural benefit gap*. **PubMed** (right): the MLP catches the GNN at higher budgets.

A:

- **Cora** (weak bag-of-words): GNN ≫ MLP at almost every budget (Welch $t{=}34.6$, $p{<}10^{-4}$ [6]); Tikhonov already *matches* the neural nets at $n_l{=}5$.
- **PubMed** (rich TF-IDF features): MLP ≈ GNN (≈ 0.84 macro-F1) — the *feature-dominance effect*: when features are descriptive, structure adds little.

▶ **Topology helps most when features are weak *and* labels are scarce.**

6 Heterophily: the topological trap

Q: What if neighbours disagree, i.e. the prior is violated ($\mathcal{E} > 1.5$)?

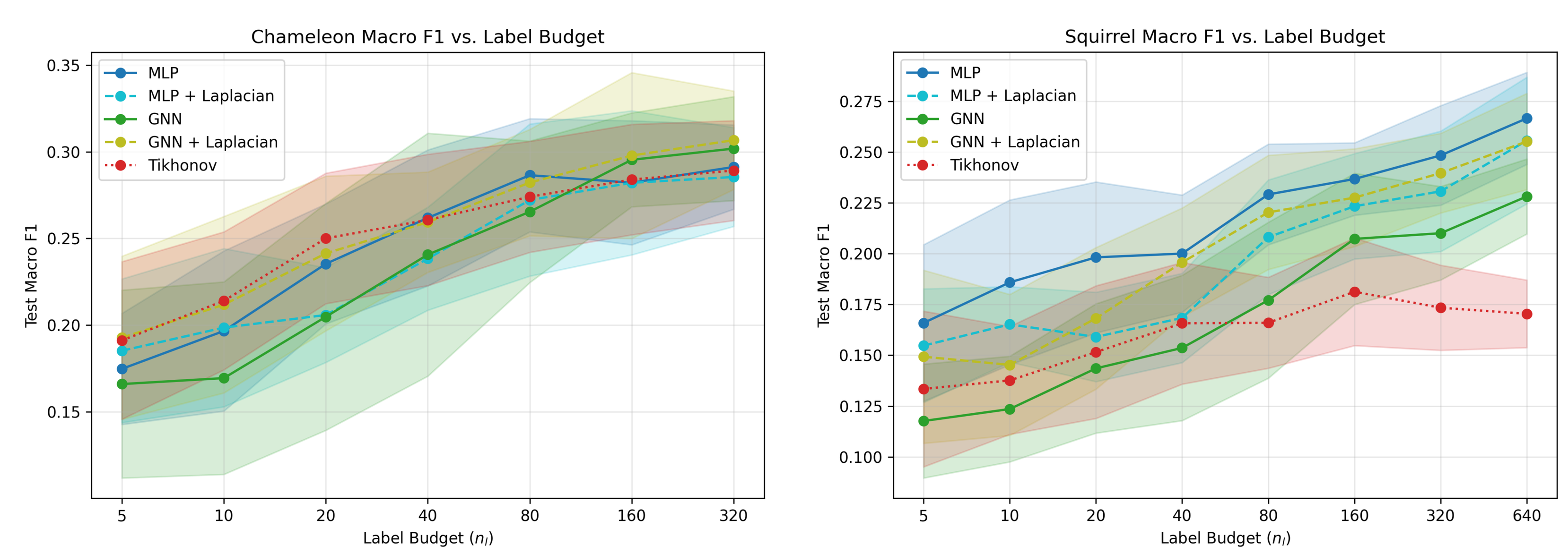

Fig. 2: **Chameleon** (left) and dense **Squirrel** (right): all regimes collapse to a feature floor; on Squirrel the analytic Tikhonov solver drops *below* the feature-only MLP.

A:

- All five models flatline at a **feature floor** (≈ 0.28 on Chameleon); GNN vs. MLP *not* significant ($t{=}0.95$, $p > 0.05$).
- Dense **Squirrel** (deg ∼42): Tikhonov over-smooths across ∼42 conflicting neighbours → $0.17 <$ MLP 0.26.
- Regularised neural nets do *not* crash below the MLP: Adam adaptively down-weights the toxic prior — *optimisation resilience*.

▶ **Heterophily *plus* density breaks classical solvers; weak features cap neural nets.**

7 Is an explicit prior redundant on a GNN?

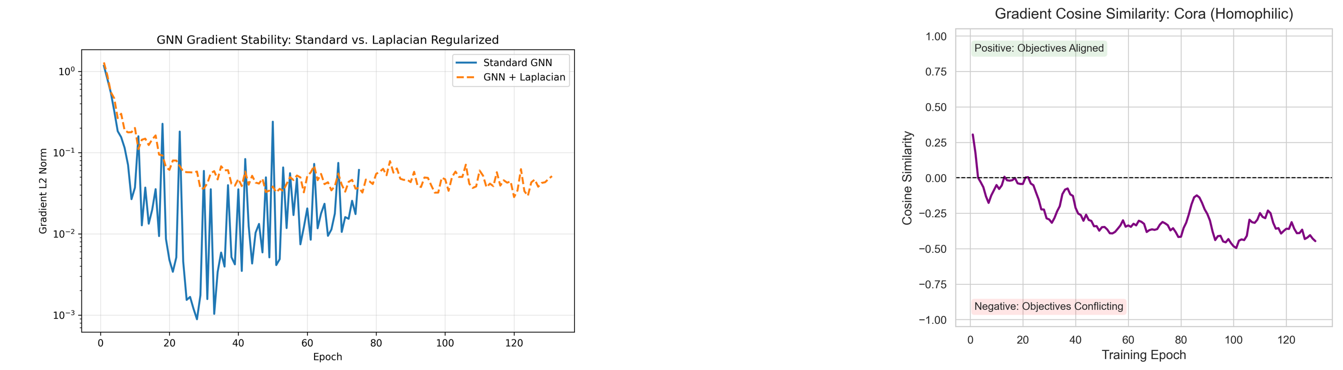

Fig. 3: Cora ($n_l{=}20$): the penalty damps gradient jitter (left); CE vs. Laplacian gradients stay *negatively* aligned (right).

- The penalty steadies optimisation but gives **no accuracy gain**; the gradient cosine similarity stays **negative** (near -1 under heterophily) — the prior *fights* the labels.

▶ **A GNN already smooths in the forward pass ⇒ an explicit prior is algorithmically redundant.**

8 Take-aways for practitioners

1. **Scarce labels + smooth graph** → cheap Tikhonov / Laplacian-MLP as a structural anchor.
2. **Rich node features** → a plain MLP; skip message passing.
3. **Dense heterophily** → avoid analytic solvers; a well-tuned MLP is the safest baseline.

9 Limitations & future work

- Four static datasets, fixed capacity; no temporal / scale-free graphs.
- **Next:** adaptive spectral filters (e.g. GPR-GNN [7]), signed regularisers for heterophily, energy-aware constraint toggling.

References

[1] T. N. Kipf et al. "Semi-Supervised Classification with Graph Convolutional Networks". *CoRR* abs/1609.02907 (2016).
[2] T. Viering et al. "The Shape of Learning Curves: A Review". English. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45.6 (2023), pp. 7799–7819.
[3] M. Defferrard et al. *Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering*. 2017.
[4] P. Sen et al. "Collective classification in network data". *AI magazine* 29.3 (2008), pp. 93–93.
[5] H. Pei et al. *Geom-GCN: Geometric Graph Convolutional Networks*. 2020.
[6] B. L. Welch. "The Generalization of 'Student's' Problem when Several Different Population Variances are Involved". *Biometrika* 34.1/2 (1947), pp. 28–35.
[7] E. Chien et al. *Adaptive Universal Generalized PageRank Graph Neural Network*. 2021.