Capacity Scaling and Learning Curves in Graph Neural Networks

Tudor Pipernea

Delft University of Technology, EEMCS Faculty | CSE3000 Research Project

Take-home message

ChebNet capacity should scale with graph *topology* and label budget, not with raw parameter count.

Motivation

Graph Neural Networks often break the usual deep-learning rule that "bigger is better": scaling width, depth, or filter reach can *reduce* accuracy. So how large should a model be for a given amount of labelled data?

Research Question

How should ChebNet capacity, through hidden width h, depth L, and Chebyshev order K, scale with the number of labelled training nodes?

Datasets

Dataset	Type	Nodes	Cls.	Homoph.
Cora	Citation	2,708	7	0.81
PubMed	Citation	19,717	3	0.80
Chameleon	Web	890	5	0.24
Squirrel	Web	2,223	5	0.21

Homophilic citation graphs vs. heterophilic web graphs (homophily $0.81 \rightarrow 0.21$).

Protocol

Each knob is swept against the label budget: $n \leq 1280$, $h \leq 256$, $L \leq 5$, $K \leq 10$. 20 runs per setting; metric is mean test cross-entropy with bootstrap bands.

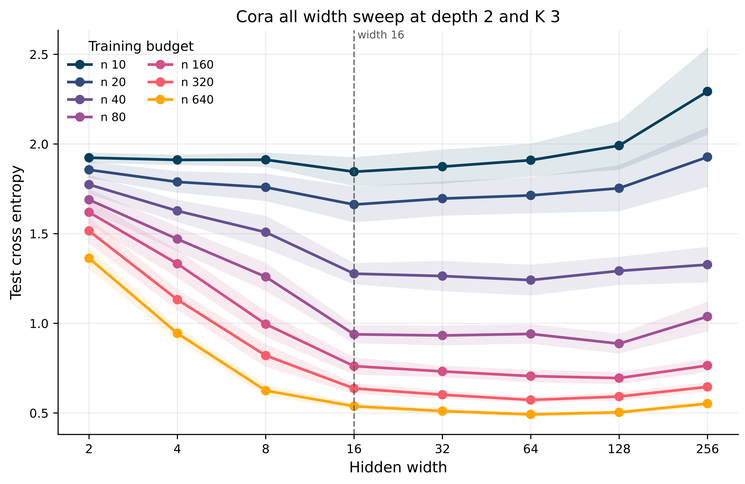
Three rules for sizing ChebNet

Size, don't oversize: width past $h \approx 16$ rarely helps — parameter count misleads.

Shape to the graph: shallow (L=2) for homophilic, deeper (L=5) only for heterophilic.

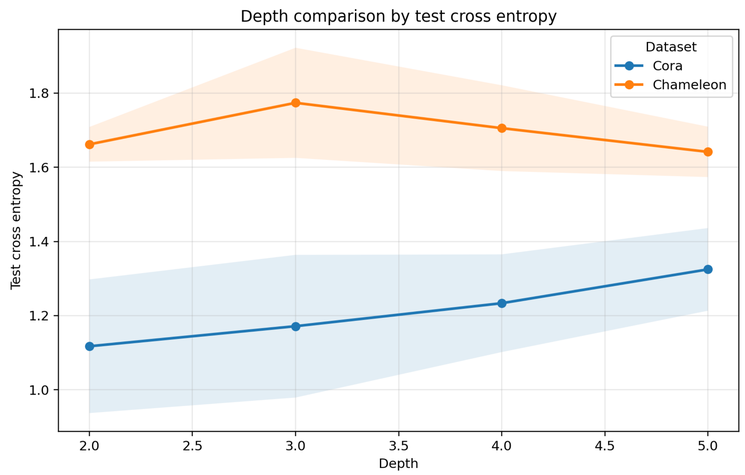
Reach before depth: tune K=2–3 first; add layers last.

Width: gains flatten near $h = 16$



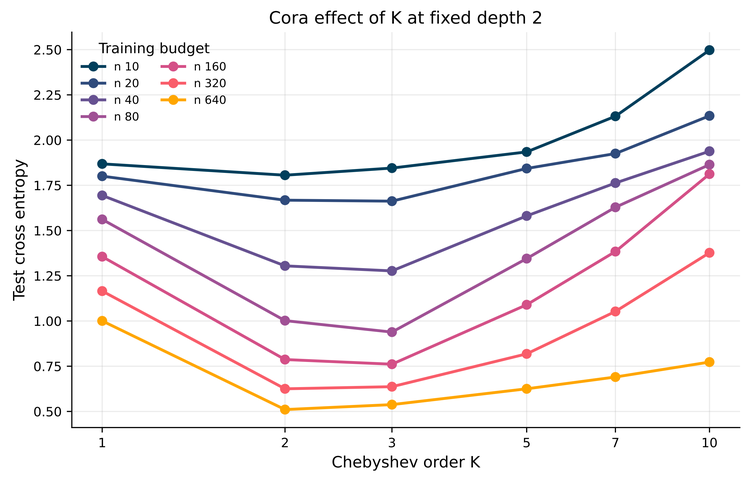
Inflection at $h = 16$. Beyond it, citation graphs gain little, while web graphs tend to overfit under small label budgets.

Depth: match the graph topology



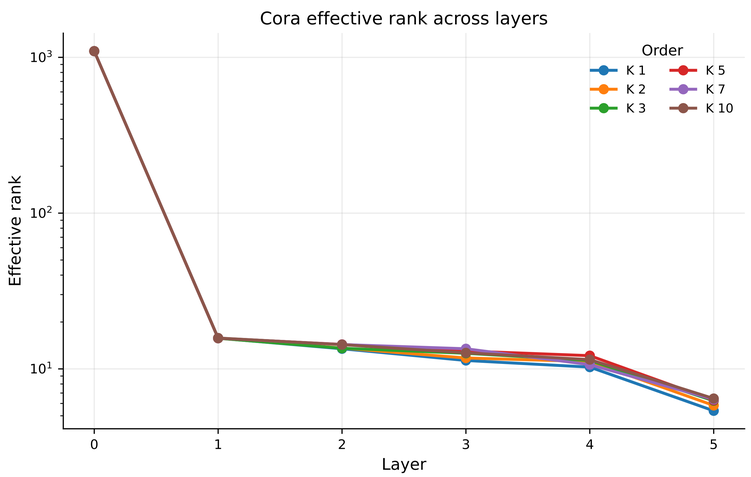
$L = 2$ is best for citation graphs; $L = 5$ for web graphs. Depth sets the receptive field, not just the parameter count.

Order K: a moderate optimum



Both extremes hurt: small K under-reaches, while large K (≥ 5) mixes in distant, unrelated nodes and dilutes the local signal. **$K = 2$–3 is safest** on both graph types.

Diagnostic: representation compression



As depth grows, **effective rank drops sharply** ($>10^3 \rightarrow <10$ over a few layers). The trend appears even in *untrained* models, indicating a partly structural smoothing tendency rather than a purely training effect.