Do Graph Neural Networks Follow Power Laws?

Vasil Georgeiv

Delft University of Technology, EEMCS Faculty | CSE3000 Research Project

Motivation

- **Labels are expensive** in many graph-learning applications.
- **Learning curves estimate returns** from collecting more labels.
- **Early prediction matters:** can small-budget results forecast larger-budget performance?
- **Open question:** do GNN learning curves follow power-law behavior, or do they saturate exponentially?

Research Question

Which parametric model better describes GNN learning curves and better extrapolates from small labeled-node budgets to larger budgets?

This project compares:

- a **power-law model with offset**,
- an **exponential model with offset**.

The comparison is made for semi-supervised node classification on two homophilic and two heterophilic datasets.

Learning-Curve Models

We report learning curves using test accuracy. For curve fitting, accuracy is converted to error so that both candidate models can be written as decreasing functions. Let x be the number of labeled training nodes and let $y(x)$ be the fitted error value.

Power law with offset

$$y(x) = ax^{-b} + c$$

This model captures diminishing returns: performance improves quickly at first, then more slowly as the label budget grows.

Exponential with offset

$$y(x) = ae^{-bx} + c$$

This model captures fast initial improvement followed by stronger saturation.

In both models, a controls the scale, b controls the decay rate, and c is the estimated asymptotic error level.

Datasets

Dataset	Type	Classes	Nodes	Homophily
Cora	Homophilic	7	2708	0.81
PubMed	Homophilic	3	19717	0.80
Chameleon	Heterophilic	5	890	0.24
Squirrel	Heterophilic	5	2223	0.21

- Homophilic and heterophilic graphs are both included.
- Budgets are limited by the available training candidate pool.

Experimental Setup

- **Task:** semi-supervised node classification.
- **GNN architecture:** ChebNet.
- **Budgets:** labeled-node budgets follow a doubling schedule: $5, 10, 20, 40, 80, 160, 320, 640, 1280$, depending on dataset size.
- **Metric:** mean test accuracy across repeated runs.
- **Repeated runs:** each budget is evaluated over multiple seeds to reduce variance from random label sampling and model initialization.

Full-Curve Fit

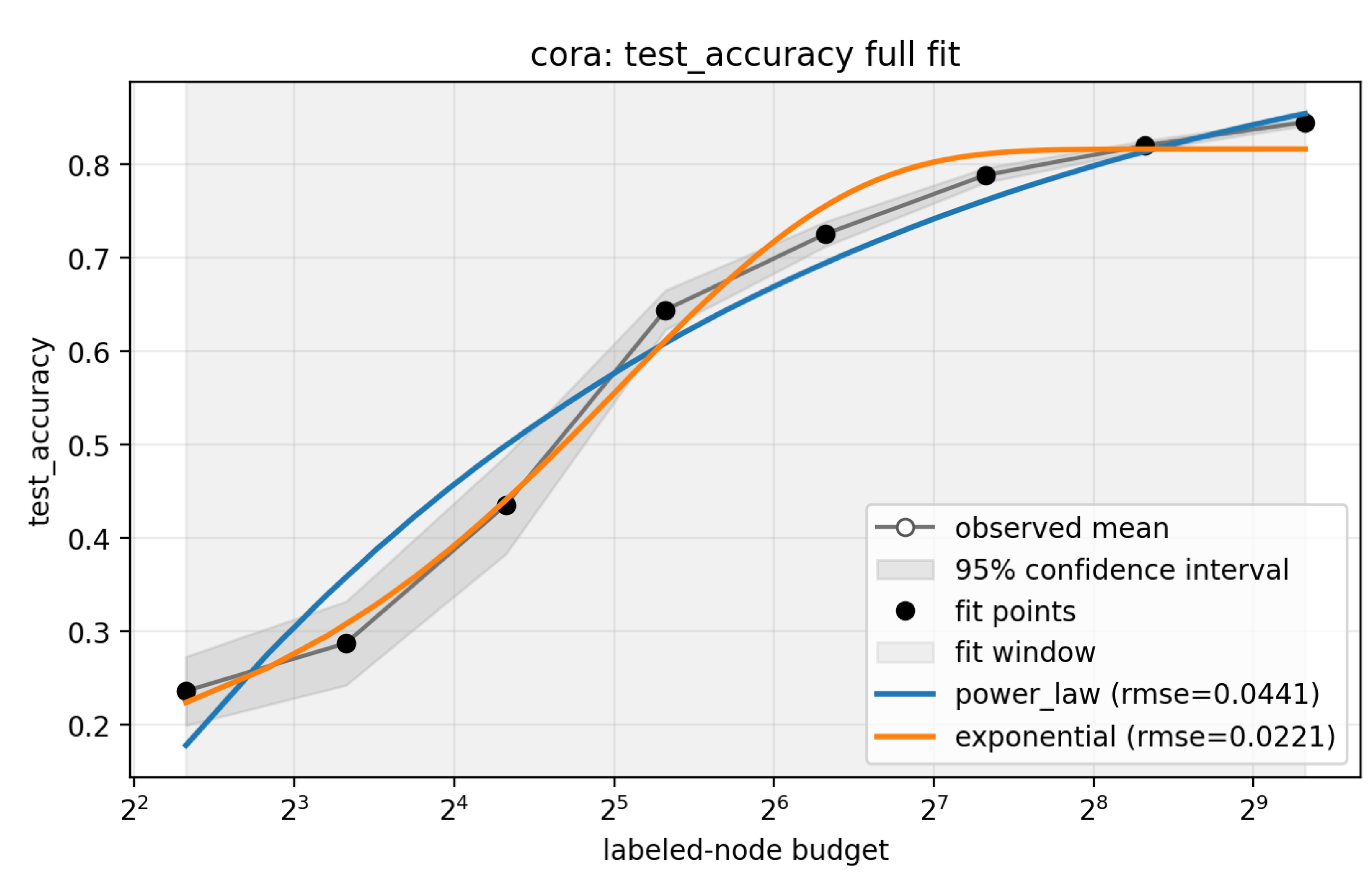
This experiment asks whether a curve model can describe the full observed learning curve.

1. Train the GNN at all feasible labeled-node budgets.
2. Average test accuracy across repeated runs for each budget.
3. Fit both parametric models to the full averaged curve.
4. Compare descriptive fit using RMSE.

Lower RMSE means that the fitted curve follows the observed learning curve more closely.

Dataset	Power RMSE	Exp. RMSE	Better
Cora	0.044	0.022	exponential
PubMed	0.007	0.027	power-law
Chameleon	0.014	0.010	exponential
Squirrel	0.004	0.008	power-law

Table 1. Full-curve descriptive fit on averaged test-accuracy learning curves.



Takeaways from Full-Curve Fit

- There is **no universal descriptive winner**.
- Exponential fits better on Cora and Chameleon.
- Power law fits better on PubMed and Squirrel.
- Homophily alone does not explain the preferred curve family: both homophilic and heterophilic datasets show mixed behavior.

Extrapolation

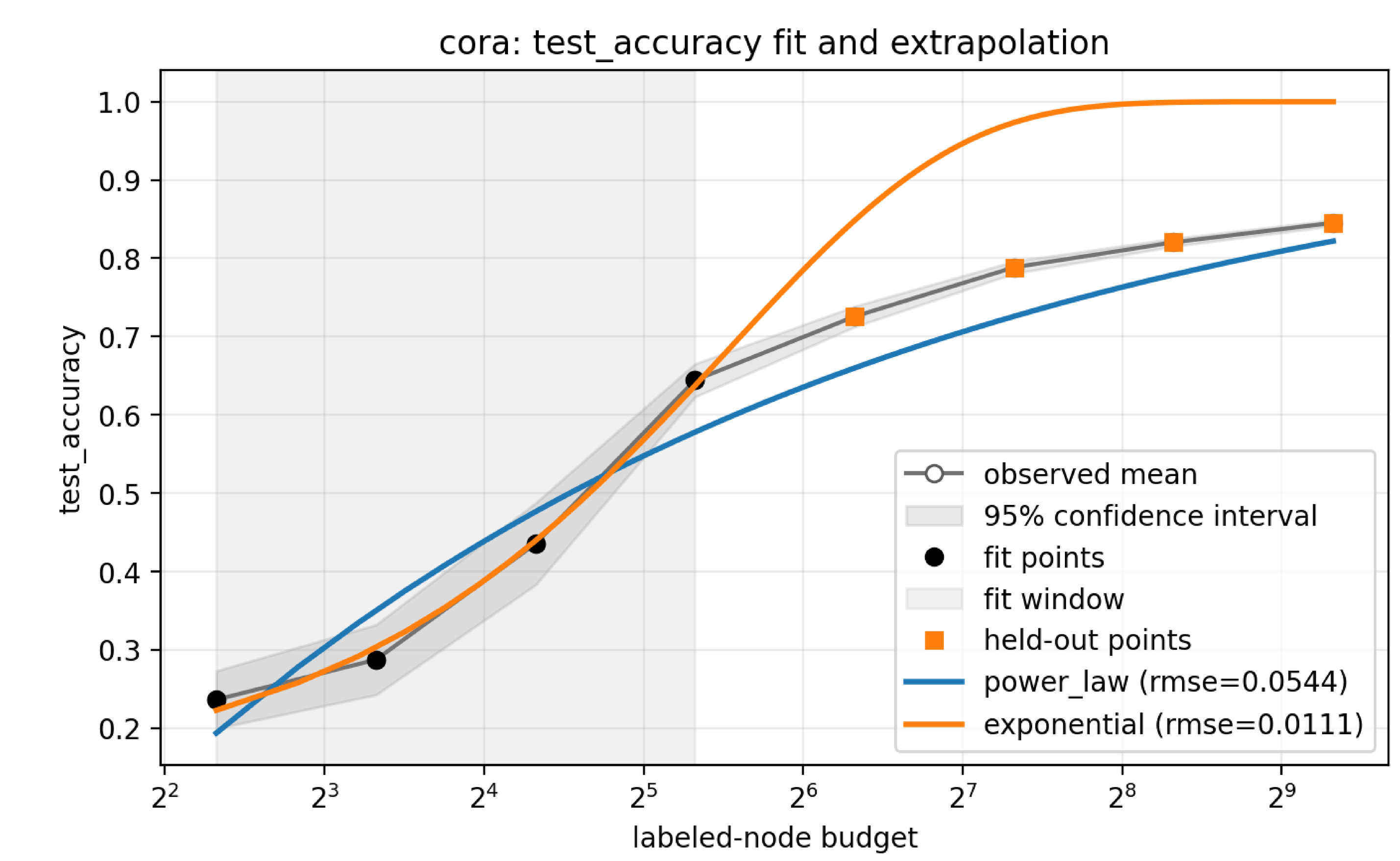
This experiment asks whether early-budget measurements can predict later performance.

1. Fit each model using only lower labeled-node budgets.
2. Predict performance at the larger budgets.
3. Compare predictions with the measured averaged test accuracy.

This is the main predictive test: a model may fit the observed part of the curve well but still extrapolate poorly.

Dataset	Fitting cutoffs	Extrapolation result
Cora	40, 80, 160	power-law better at all cutoffs
PubMed	40, 80, 160, 320	power-law better at all cutoffs
Chameleon	40, 80	mixed result
Squirrel	40, 80, 160	power-law better at all cutoffs

Table 2. Summary of held-out extrapolation results using RMSE.



Main Finding

Descriptive fit and extrapolation accuracy are different objectives.

- Exponential models sometimes fit the observed curve better.
- Power-law models more often extrapolate better to larger budgets.
- Therefore, choosing a curve only by full-curve fit can be misleading.

Overall, power-law behavior appears useful for extrapolating ChebNet learning curves, but it should not be assumed to hold universally.

Future Work

- Repeat the study with more GNN architectures.
- Evaluate more homophilic and heterophilic graph datasets.
- Study whether fitted parameters are stable across random splits and label samplings.
- Test additional curve families or non-parametric extrapolation methods.