Fixed Aggregation Features Can Rival GNNs

Celia Rubio-Madrigal   Rebekka Burkholz

CISPA Helmholtz Center for Information Security

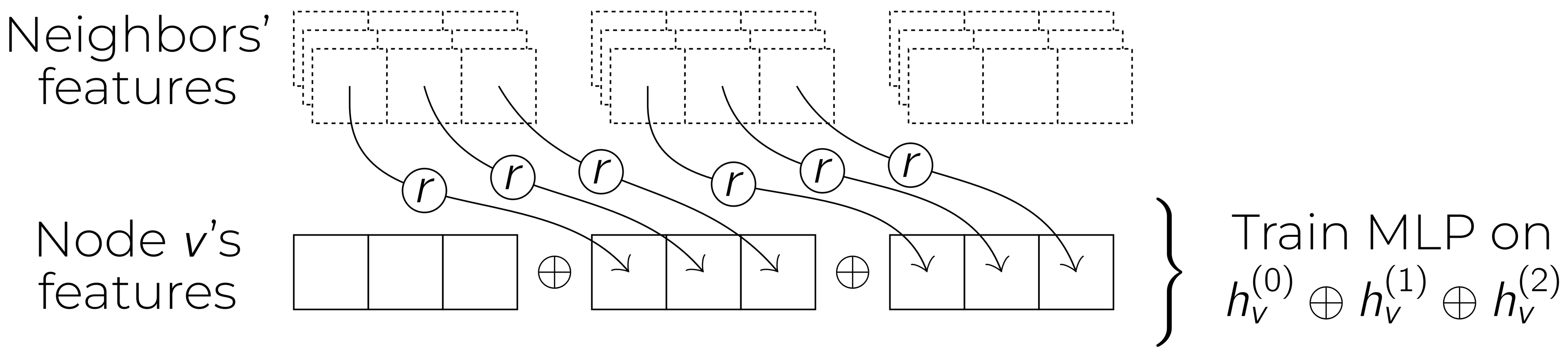

FIGURE 1: Fixed Aggregation Features (FAFs) for one reducer r.

Background and motivation

- Well-tuned classic GNNs can rival GTs [1].
  $\Longrightarrow$ Reducing complexity is possible. What must remain?
- Scalability inspires precomputing multi-hop features [2,3,4].
  $\Longrightarrow$ How necessary is learning the aggregation process?
- We replace trainable with fixed, concatenated aggregation.
  $\Longrightarrow$ Can it be maximally **expressive**? Can it be **performant**?

Method: Fixed Aggregation Features (FAFs)

FAFs are a preprocessing step (Fig. 1). For depth $k \in \{1, \dots, K\}$ and reducer $r \in \mathcal{R} \subset \{\Phi, \text{mean}, \text{sum}, \text{max}, \text{min}, \text{std}, \dots\}$:

$$h_v^{(0,r)} = x_v, \qquad h_v^{(k,r)} = r\Big(\{\, h_u^{(k-1,r)} \,\}_{u \in N(v)}\Big). \tag{1}$$

We then train an MLP on the concatenated representation:

$$z_v = x_v \oplus \Big(\bigoplus_{r \in \mathcal{R}} \bigoplus_{k \in \{1,\dots,K\}} h_v^{(k,r)} \Big) \tag{2}$$

If r is **injective**, the neighborhood information is preserved [5].

Theory: Existence of lossless aggregation

KOLMOGOROV–ARNOLD REPRESENTATION [6]: There exists a **fixed**, injective Φ (and ϕ) s.t. any multivariate f can be written as

$$f\big(\{\, h_u \,\}_{u \in N(v)}\big) = g\,(\Phi(\{\, h_u \,\}_{u \in N(v)}) = g\,(\Sigma_{u \in N(v)}\, \phi(h_u)) \tag{3}$$

for a univariate g (modeled by an MLP). While Φ is discontinuous, the learnable part g inherits continuity from f.

ICML International Conference On Machine Learning

CISPA HELMHOLTZ CENTER FOR INFORMATION SECURITY

Use FAFs instead of MLPs as tabular baselines for graph benchmarks

Many graph benchmarks do not require either *learned* or *information-preserving* aggregation.

We need to start using richer benchmarks, e.g., with long-range dependencies, where FAFs may struggle more.

We transform graph into tabular data; in theory, without information loss

Information-preserving aggregation is neither necessary nor sufficient.

Mean and sum have amenable inductive biases for better generalization.

References:
[1] Luo et al., NeurIPS 2024 [2] Frasca et al., ICML 2020 GRL+
[3] Zhang et al., KDD 2022 [4] Deng et al., DAC 2024
[5] Xu et al., ICLR 2019 [6] Schmidt-Hieber, NN 2021
[7] Wu et al., ICML 2019 [8] Gallicchio & Micheli, IJCNN 2010
[9] Rubio-Madrigal et al., ICLR 2025 [10] Roth et al., LoG 2024
[11] Eremeev et al. [12] Choi et al. [13] Hayler et al.

Experiments on node classification

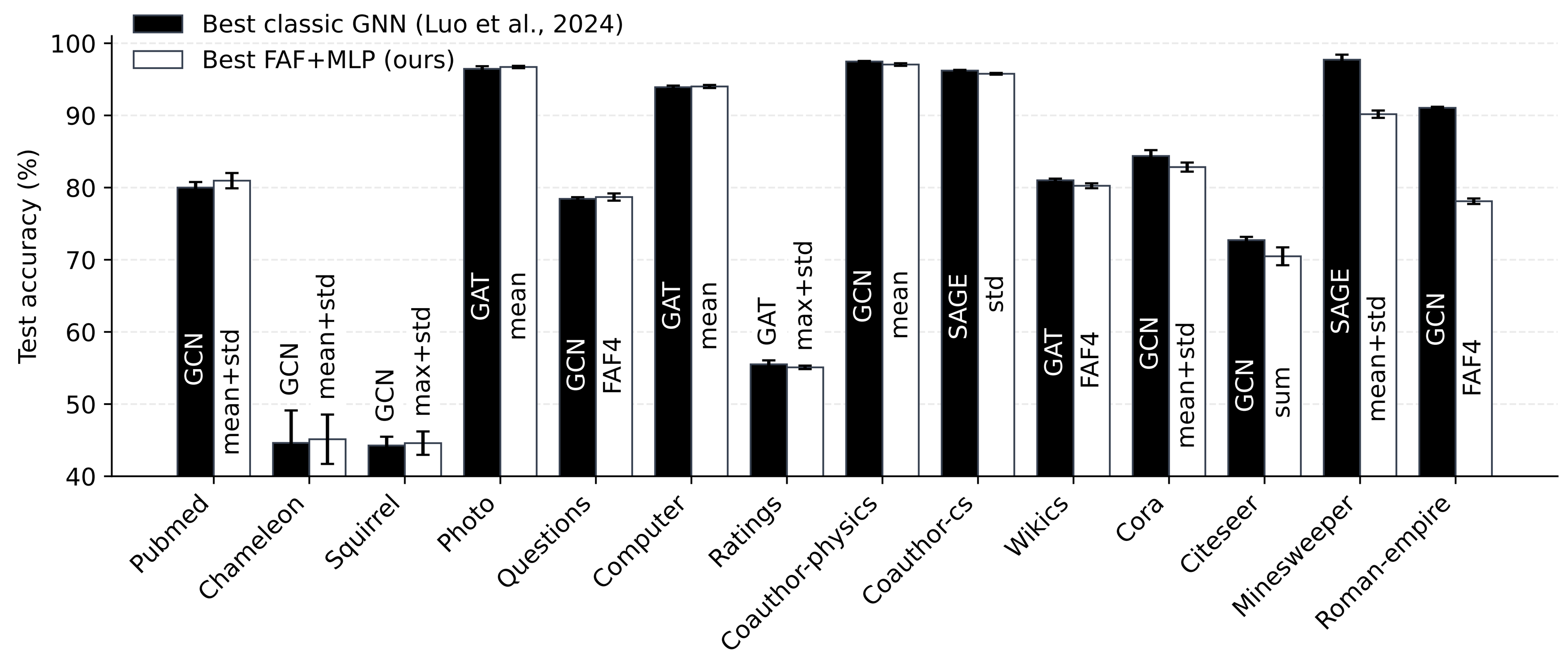

FIGURE 2: On 12/14 node classification datasets, we rival well-tuned GCN, GAT, and GraphSAGE with FAFs trained on MLP.

Ablations and derived insights

- HOPS: Accuracy peaks at $K = 2$-4, and 2 is often enough.
  $\Longrightarrow$ Most current benchmarks **don't require depth**.
  $\Longrightarrow$ Higher depths may overfit.
- (NON)-INJECTIVITY: Φ behaves badly. Mean/sum/etc. are not lossless, but yield more amenable features and are enough.
  $\Longrightarrow$ They **don't require injectivity** but better inductive bias.
- CLASSIFIER: A **well-tuned** MLP outperforms a linear layer. Concatenating **all hops** outperforms only using the last hop.
  $\Longrightarrow$ Different than SGC, freezing weights, or reservoirs [7,8].
- REDUCERS: Mean is often the best, and std complements it. Max/sum are slightly more helpful on specific datasets.
  $\Longrightarrow$ Concatenating **all reducers** is a robust default (FAF4).

Benefits of the tabular form

We turn the graph into a tabular matrix, unlocking many tools:

- INTERPRETABILITY: feature importance per k and r.
- EFFICIENCY: no backpropagation through convolutions.
- AUGMENTATIONS: like structural encodings or rewiring [9,10].
- FOUNDATION MODELS: graph data ready for TFMs [11,12,13].

Code: https://github.com/celrm/fixed-aggregation-features