TEXAS A&M UNIVERSITY
Engineering

GraphFM: Improving Large-Scale GNN Training via Feature Momentum

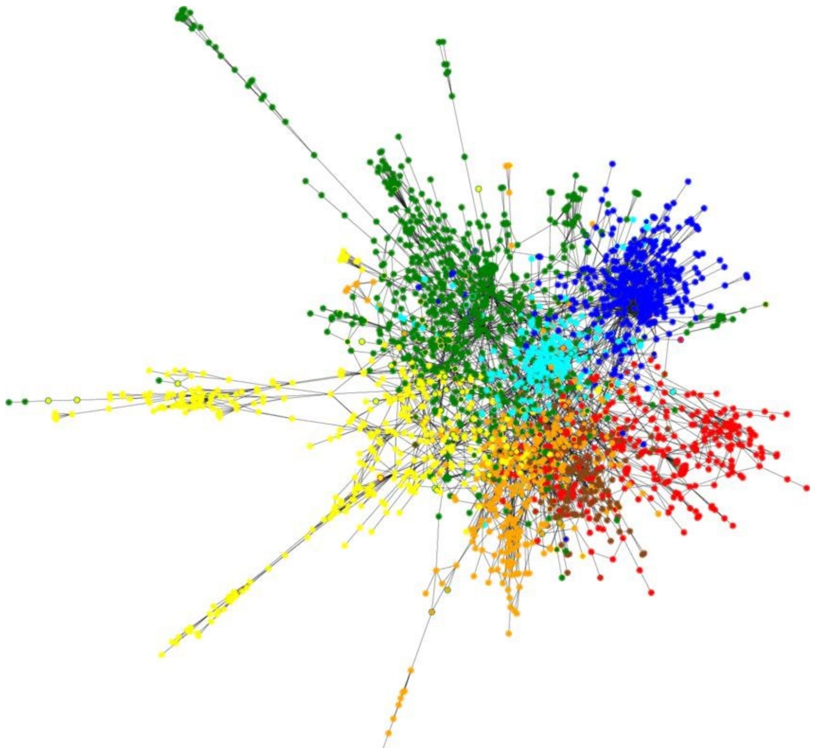
Haiyang Yu*, Limei Wang*, Bokun Wang*, Meng Liu, Tianbao Yang, and Shuiwang Ji

TEXAS A&M UNIVERSITY
Engineering

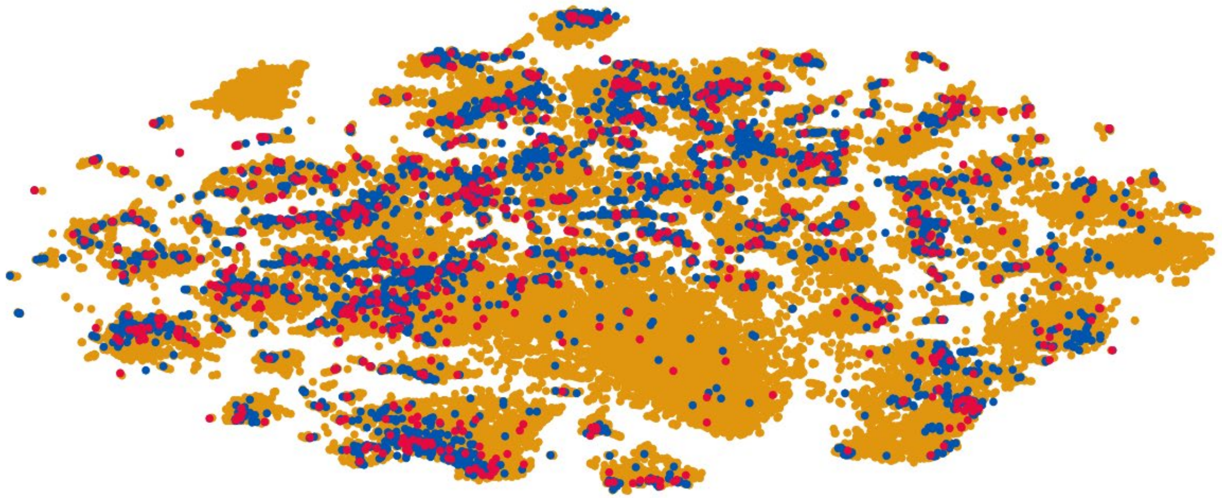
Background

large-scale graphs

TEXAS A&M UNIVERSITY Engineering



Small-scale
Cora
2,708 nodes
5,278 edges



Large-scale
Ogbn-products:
2,449,029 nodes
61,859,140 edges

Reference: Geometric deep learning on graphs and manifolds using mixture model CNNs, CVPR 2017
Open Graph Benchmark: Datasets for Machine Learning on Graphs, NeurIPS 2020

Problem: biased gradient estimation

TEXAS A&M UNIVERSITY Engineering

- Two layer model with parameters $\boldsymbol{w}$.

$$F(\mathbf{w}) = f_1 \circ f_2(\mathbf{w})$$

- Introduce the random variable ξ, ζ to denote the sampling procedure,
  - For the unbiased features and gradients

$$\mathbb{E}[f_1(\cdot;\xi)] = f_1(\cdot), \quad \mathbb{E}[\nabla f_1(\cdot;\xi)] = \nabla f_1(\cdot)$$
$$\mathbb{E}[f_2(\cdot;\zeta)] = f_2(\cdot), \quad \mathbb{E}[\nabla f_2(\cdot;\zeta)] = \nabla f_2(\cdot)$$

  - Unbiased gradients

$$\nabla \widehat{F}(\mathbf{w}) = \nabla f_2(\mathbf{w};\zeta)^\top \nabla f_1(f_2(\mathbf{w});\xi)$$

  - The true gradients during the sampling procedure in training steps - biased

$$\nabla \widehat{F}(\mathbf{w}) = \nabla f_2(\mathbf{w};\zeta)^\top \nabla f_1(\boxed{f_2(\mathbf{w};\zeta)};\xi)$$

TEXAS A&M UNIVERSITY
Engineering

Methods

Feature Momentum

$$\hat{f}_{2,t} = (1-\beta)\hat{f}_{2,t-1} + \beta f_2(\mathbf{w}_t;\zeta)$$

where t denotes the training step.

Contribution:

- GraphFM-IB: apply FM to node-wise sampling method GraphSAGE
  - Rigorous convergence analysis
  - Less GPU memory consumption

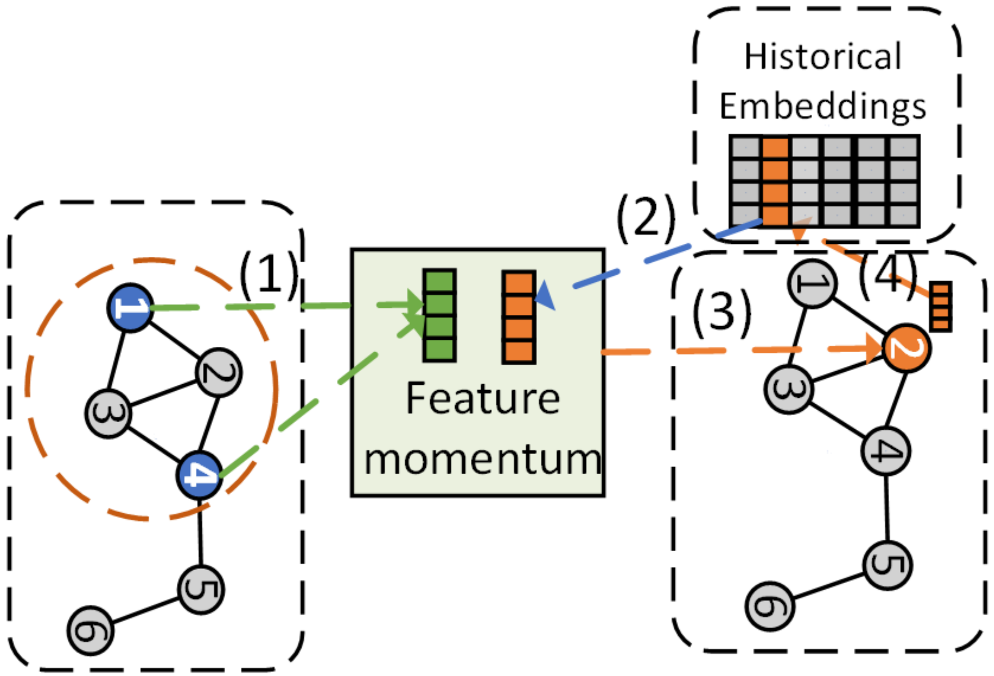
- GraphFM-OB: apply FM to subgraph sampling method GNNAutoScale
  - Provide theoretical insight to alleviate the staleness problem of historical embeddings

- Consistently performance improvement

GraphFM-IB

TEXAS A&M UNIVERSITY
Engineering

Sampling region
Target nodes
Sampled nodes
Other nodes

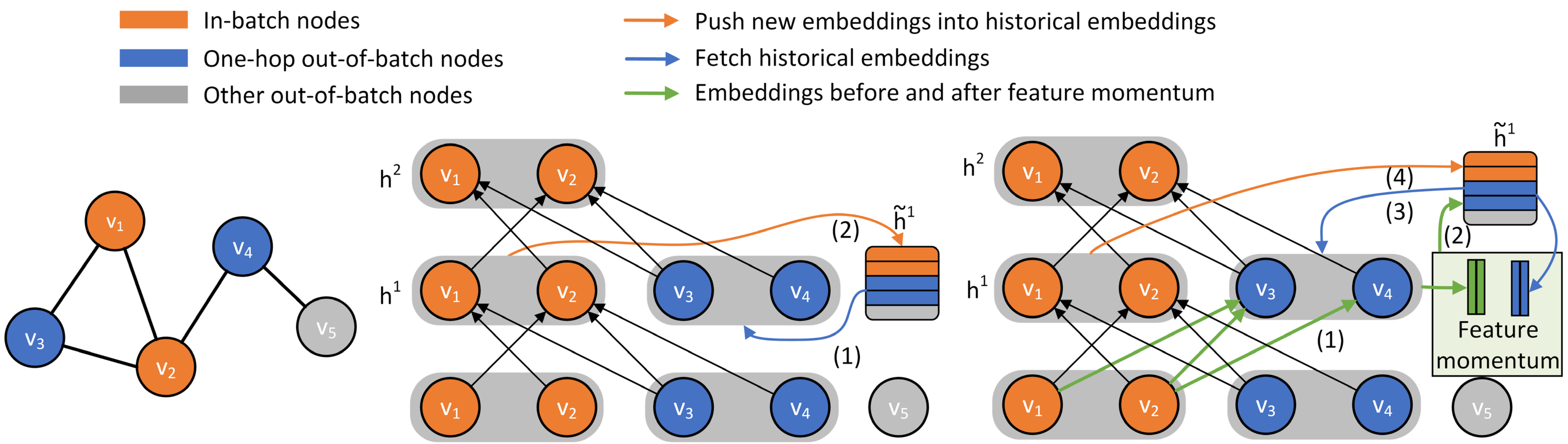


Push new embeddings into historical embeddings

Fetch historical embeddings

Embeddings before and after feature momentum

GraphFM-OB



(a) Original graph (b) Forward propagation in GNNAutoscale (c) Forward propagation with feature momentum

TEXAS A&M UNIVERSITY
Engineering

Experiments

Overall Performance

Backbones	Methods	Flickr	Reddit	Yelp	ogbn-arxiv	ogbn-products
	VR-GCN	0.482 ± 0.003	0.964 ± 0.001	0.640 ± 0.002	–	–
	FastGCN	0.504 ± 0.001	0.924 ± 0.001	0.265 ± 0.053	–	–
	GraphSAINT	0.511 ± 0.001	0.966 ± 0.001	0.653 ± 0.003	–	0.791 ± 0.002
	Cluster-GCN	0.481 ± 0.005	0.954 ± 0.001	0.609 ± 0.005	–	0.790 ± 0.003
	SIGN	0.514 ± 0.001	0.968 ± 0.000	0.631 ± 0.003	0.720 ± 0.001	0.776 ± 0.001
SAGE	GraphSAGE	0.501 ± 0.013	0.953 ± 0.001	0.634 ± 0.006	0.715 ± 0.003	0.783 ± 0.002
	GraphFM-IB	<u>0.513 ± 0.009</u>	<u>0.963 ± 0.005</u>	<u>0.641 ± 0.001</u>	0.713 ± 0.002	<u>0.792 ± 0.003</u>
GCN	GNNAutoScale	0.5400	0.9545	0.6294	0.7168	0.7666
	GraphFM-OB	<u>0.5446</u>	0.9540	–	<u>0.7181</u>	<u>0.7688</u>
GCNII	GNNAutoScale	0.5620	0.9677	0.6514	0.7300	0.7724
	GraphFM-OB	<u>0.5631</u>	<u>0.9680</u>	<u>**0.6529**</u>	<u>**0.7310**</u>	<u>0.7742</u>
PNA	GNNAutoScale	0.5667	**0.9717**	0.6440	0.7250	0.7991
	GraphFM-OB	<u>**0.5710**</u>	0.9712	<u>0.6450</u>	<u>0.7290</u>	<u>**0.8047**</u>

GraphFM-IB consumption

Methods	Neighbor sizes	Reddit	Flickr
GraphSAGE	2 layer full-batch	OOM	0.513/4,860M/1.7s
GraphSAGE	[25,10]	0.957/3,080M/6.5s	0.512/1,740M/1.6s
GraphSAGE	[1,1]	0.931/2,250M/3.3s	0.490/1,310M/1.2s
GraphFM-IB + SAGE	[1,1]	0.957/2,300M/3.9s	0.503/1,480M/1.4s
GraphSAGE	[4,4]	0.955/2,320M/4.0s	0.507/1,390M/1.3s
GraphFM-IB + SAGE	[4,4]	0.958/2,450M/4.2s	0.511/1,540M/1.5s
GraphSAGE	4 layer full-batch	OOM	0.514/11,000M/5.2s
GraphSAGE	[25,10,10,10]	0.962/10,110M/53s	0.514/6,480M/3.6s
GraphSAGE	[1,1,1,1]	0.951/2,700M/5.2s	0.502/1,360M/1.7s
GraphFM-IB + SAGE	[1,1,1,1]	0.962/2,860M/6.2s	0.513/1,700M/2.0s
GraphSAGE	[2,2,2,2]	0.958/2,870M/5.8s	0.509/1,470M/1.8s
GraphFM-IB + SAGE	[2,2,2,2]	0.963/3,130M/7.5s	0.513/1,900M/2.4s

TEXAS A&M UNIVERSITY Engineering

Thank you!