GIST: Distributed Training for Large-Scale GCNs

Cameron R. Wolfe* Jingkang Yang*, Fangshuo Liao*,
Arindam Chowdhury, Chen Dun, Artun Bayer, Santiago Segarra,
Anastasios Kyrillidis

*Equal Contribution

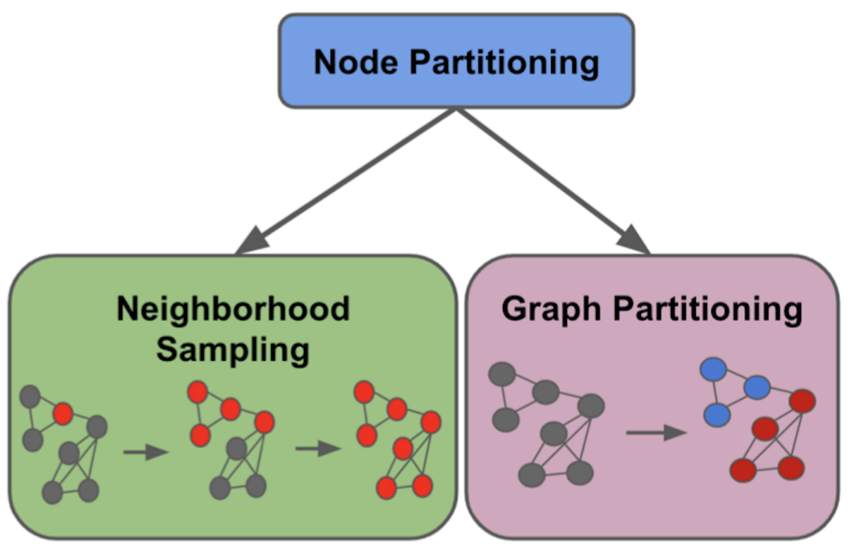
RICE

Central Question

GCN on Graph Structured Data:

$$H_{\ell+1} = \sigma\left(\bar{A} H_\ell \Theta_\ell\right); \quad H_0 = X$$

Current Efficient GCN Training



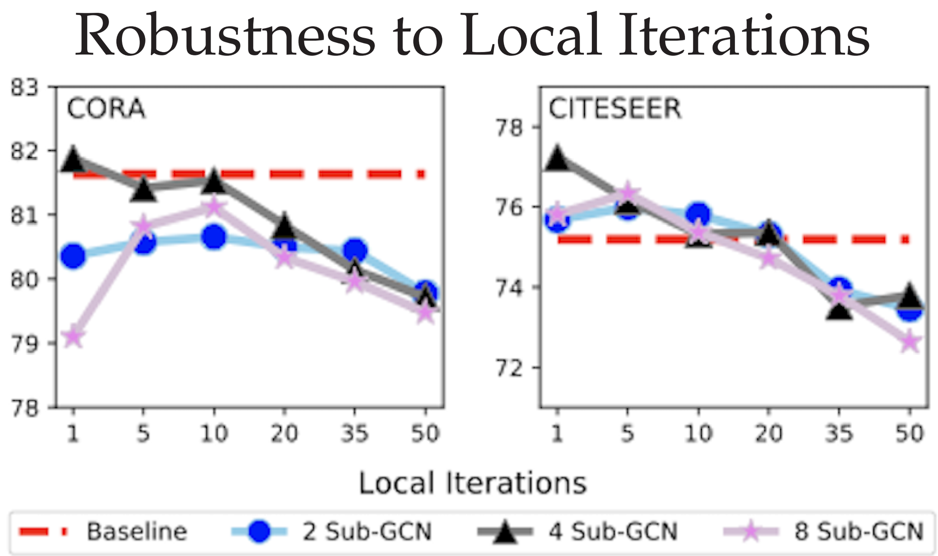
Our proposal: a novel distributed training framework for large-scale GCNs.

Design Ablation

Selective Partitioning of Hidden Layers

# Sub-GCNs	d_0	d_1	d_2	Cora	Citeseer
Baseline				81.52 ± 0.005	75.02 ± 0.018
2	✓	✓	✓	80.00 ± 0.010	$\mathbf{75.95} \pm 0.007$
	✓	✓		78.30 ± 0.011	69.34 ± 0.018
		✓	✓	$\mathbf{80.82} \pm 0.010$	75.82 ± 0.008
4	✓	✓	✓	76.78 ± 0.017	70.66 ± 0.011
	✓	✓		66.56 ± 0.061	68.38 ± 0.018
		✓	✓	$\mathbf{81.18} \pm 0.007$	$\mathbf{76.21} \pm 0.017$
8	✓	✓	✓	48.32 ± 0.087	45.42 ± 0.092
	✓	✓		53.60 ± 0.020	54.68 ± 0.030
		✓	✓	$\mathbf{79.58} \pm 0.006$	$\mathbf{75.39} \pm 0.016$

Robustness to Local Iterations



Combining GIST with LADIES

L	# Sub-GCNs	GIST + LADIES F1 Score	Time	Speedup
2	Baseline	89.73	3359.91s	1.00×
	2	89.29	1834.59s	1.83×
	4	88.42	1158.51s	2.90×
3	Baseline	89.57	4803.88s	1.00×
	2	86.52	2635.18s	1.82×
	4	86.72	1605.32s	3.00×

Methodology

- Disjointly partition model parameters into several sub-GCNs
- Train sub-GCNs independently in parallel
- Aggregate parameters into global model after several training iterations

Algorithm 1 `GIST` Algorithm

Parameters: T synchronization iterations, m sub-GCNs ζ local iterations, c clusters, $\mathcal{G}$ training graph.

$\Psi_{\mathcal{G}}(\cdot\,;\boldsymbol{\Theta}) \leftarrow$ randomly initialize GCN
$\{\mathcal{G}_{(j)}\}_{j=1}^{c} \leftarrow \texttt{Cluster}(\mathcal{G}, c)$
for $t = 0, \ldots, T-1$ **do**
  $\{\Psi_{\mathcal{G}}(\cdot\,;\boldsymbol{\Theta}^{(i)})\}_{i=1}^{m} \leftarrow \texttt{subGCNs}(\Psi_{\mathcal{G}}(\cdot\,;\boldsymbol{\Theta}), m)$
  Distribute each $\Psi_{\mathcal{G}}(\cdot\,;\boldsymbol{\Theta}^{(i)})$ to a different worker
  for $i = 1, \ldots, m$ **do**
    for $z = 1, \ldots, \zeta$ **do**
      $\Psi_{\mathcal{G}}(\cdot\,;\boldsymbol{\Theta}^{(i)}) \leftarrow \texttt{subTrain}(\boldsymbol{\Theta}^{(i)}, \{\mathcal{G}_{(j)}\}_{j=1}^{c})$
    end for
  end for
  $\Psi_{\mathcal{G}}(\cdot\,;\boldsymbol{\Theta}) \leftarrow \texttt{subAgg}(\{\boldsymbol{\Theta}^{(i)}\}_{i=1}^{m})$
end for

`Cluster`$(\cdot,\cdot)$: perform graph partitioning (METIS)
`subGCNs`$(\cdot,\cdot)$: divide GCN into disjoint Sub-GCNs
`subTrain`$(\cdot,\cdot)$: perform independent training
`subAgg`$(\cdot,\cdot)$: copy sub-GCN parameters into model

Communication Complexity

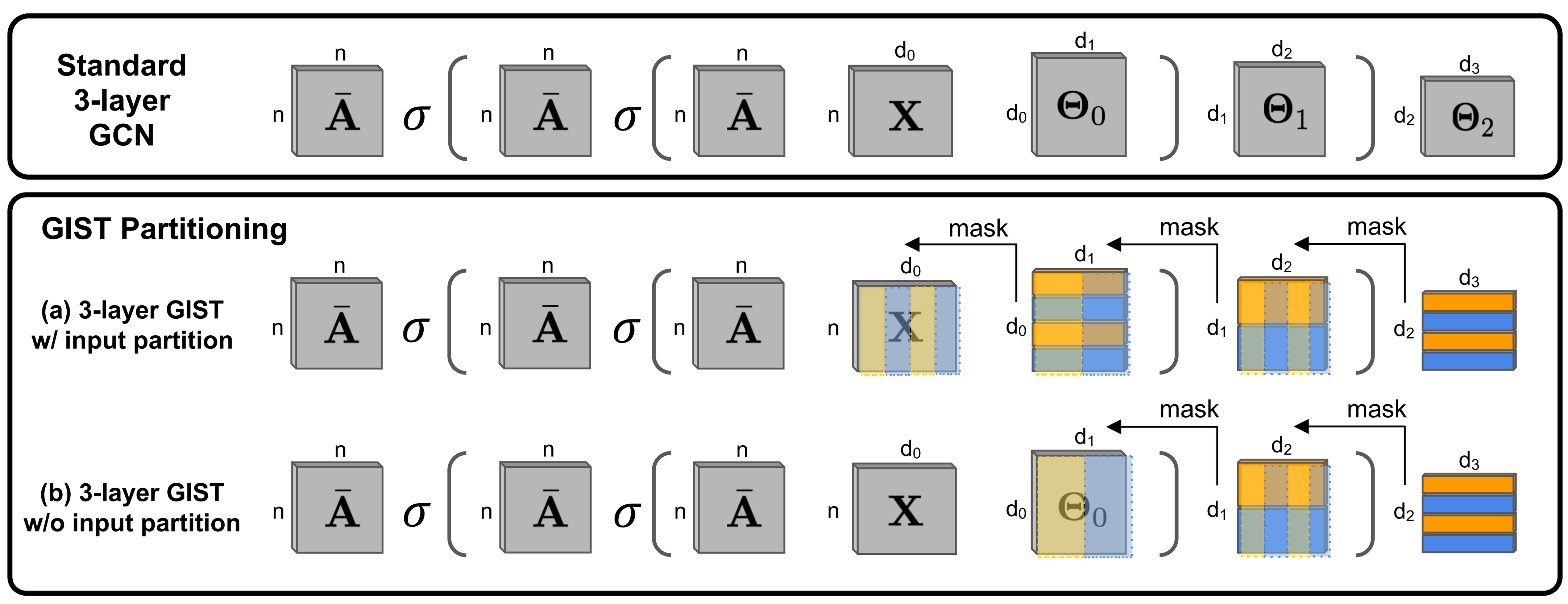
`Vanilla`: $\mathcal{O}(M d_i d_{i-1})$

`GIST`: $\mathcal{O}\left(\frac{1}{M} d_i d_{i-1}\right)$

Computational Complexity

`Vanilla`: $\mathcal{O}(N^2 d_i + N d_i d_{i-1})$

`GIST`: $\mathcal{O}\left(\frac{1}{M} N^2 d_i + \frac{1}{M^2} N d_i d_{i-1}\right)$



Ultra Wide GCN: GraphSAGE+Amazon2M

L	# Sub-GCNs	F1 Score (Time) $d_i = 400$	$d_i = 4096$	$d_i = 8192$	$d_i = 16384$	$d_i = 32768$
2	Baseline	89.38 (1.81hr)	90.58 (5.17hr)	OOM	OOM	OOM
	2	87.48 (1.25hr)	90.09 (1.70hr)	90.87 (2.76hr)	90.94 (9.31hr)	90.91 (32.31hr)
	4	84.82 (1.11hr)	88.79 (1.13hr)	89.76 (1.49hr)	90.10 (2.24hr)	90.17 (5.16hr)
	8	82.56 (1.13hr)	87.16 (1.11hr)	88.31 (1.20hr)	88.89 (1.39hr)	89.46 (1.76hr)
3	Baseline	89.73 (2.32hr)	90.99 (9.52hr)	OOM	OOM	OOM
	2	87.79 (1.56hr)	90.40 (2.12hr)	90.91 (4.87hr)	91.05 (17.7hr)	OOM
	4	85.30 (1.37hr)	88.51 (1.42hr)	89.75 (2.07hr)	90.15 (3.44hr)	OOM
	8	82.84 (1.37hr)	86.12 (1.34hr)	88.38 (1.37hr)	88.67 (1.88hr)	88.66 (2.56hr)
4	Baseline	89.77 (3.00hr)	91.02 (14.20hr)	OOM	OOM	OOM
	2	87.75 (1.79hr)	90.36 (2.77hr)	91.08 (6.92hr)	91.09 (26.44hr)	OOM
	4	85.32 (1.58hr)	88.50 (1.65hr)	89.76 (2.36hr)	90.05 (4.93hr)	OOM
	8	83.45 (1.56hr)	86.60 (1.55hr)	88.13 (1.61hr)	88.44 (2.30hr)	OOM

Large Scale Experiments

L	m	Reddit Dataset GraphSAGE F1	Time	Speedup	GAT F1	Time	Speedup	Amazon2M Dataset GraphSAGE ($d_i = 400$) F1	Time	Speedup	GraphSAGE ($d_i = 4096$) F1	Time	Speedup
2	-	96.09	105.78s	1.00×	89.57	1.19hr	1.00×	89.90	1.81hr	1.00×	91.25	5.17hr	1.00×
	2	96.40	70.29s	1.50×	90.28	0.58hr	2.05×	88.36	1.25hr	(1.45×)	90.70	1.70hr	3.05×
	4	96.16	68.88s	1.54×	90.02	0.31hr	3.86×	86.33	1.11hr	(1.63×)	89.49	1.13hr	(4.57×)
	8	95.46	76.68s	1.38×	89.01	0.18hr	6.70×	84.73	1.13hr	(1.61×)	88.86	1.11hr	(4.65×)
3	-	96.32	118.37s	1.00×	89.25	2.01hr	1.00×	90.36	2.32hr	1.00×	91.51	9.52hr	1.00×
	2	96.36	80.46s	1.47×	89.63	0.95hr	2.11×	88.59	1.56hr	(1.49×)	91.12	2.12hr	4.49×
	4	95.76	78.74s	1.50×	88.82	0.48hr	4.19×	86.46	1.37hr	(1.70×)	89.21	1.42hr	(6.72×)
	8	94.39	88.54s	(1.34×)	70.38	0.26hr	(7.67×)	84.76	1.37hr	(1.69×)	86.97	1.34hr	(7.12×)
4	-	96.32	120.74s	1.00×	88.36	2.77hr	1.00×	90.40	3.00hr	1.00×	91.61	14.20hr	1.00×
	2	96.01	91.75s	1.32×	87.97	1.31hr	2.11×	88.56	1.79hr	(1.68×)	91.02	2.77hr	5.13×
	4	95.21	78.74s	(1.53×)	78.42	0.66hr	(4.21×)	87.53	1.58hr	(1.90×)	89.07	1.65hr	(8.58×)
	8	92.75	88.71s	(1.36×)	66.30	0.35hr	(7.90×)	85.32	1.56hr	(1.93×)	87.53	1.55hr	(9.13×)