Adapting Precomputed Features for Efficient Graph Condensation

Yuan Li[1], Jun Hu[1], Zemin Liu[2], Bryan Hooi[1],
Jia Chen[3], Bingsheng He[1]

[1]National University of Singapore
[2]Zhejiang University
[3]GrabTaxi Holdings Pte. Ltd.

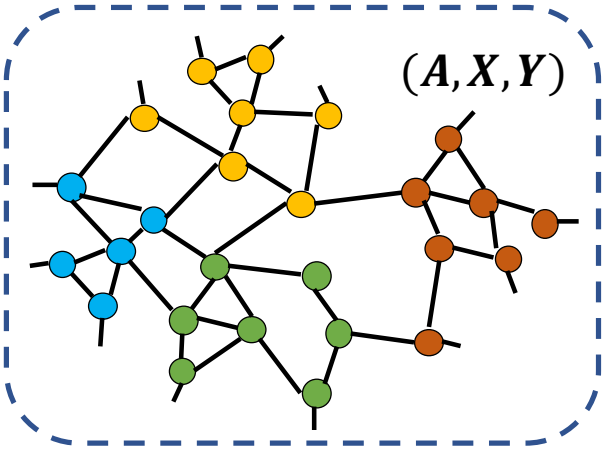
Graph Condensation (GC)

- GNNs face scalability issues on large-scale graphs.
- Graph Condensation[1] compresses a graph into a smaller, synthetic one, while preserving critical information.

[1] Jin et al. Graph Condensation for Graph Neural Networks, ICLR 2022

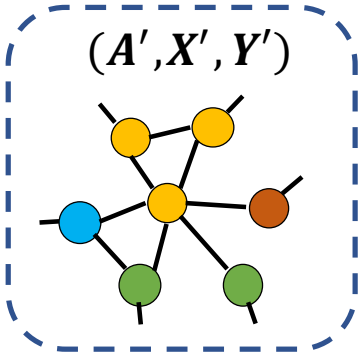
Example of Graph Condensation

Test accuracies
GCN: 93.9%
SGC: 93.5%
APPNP: 94.3%
GraphSAGE: 93.0%



153,932 training nodes

Condense



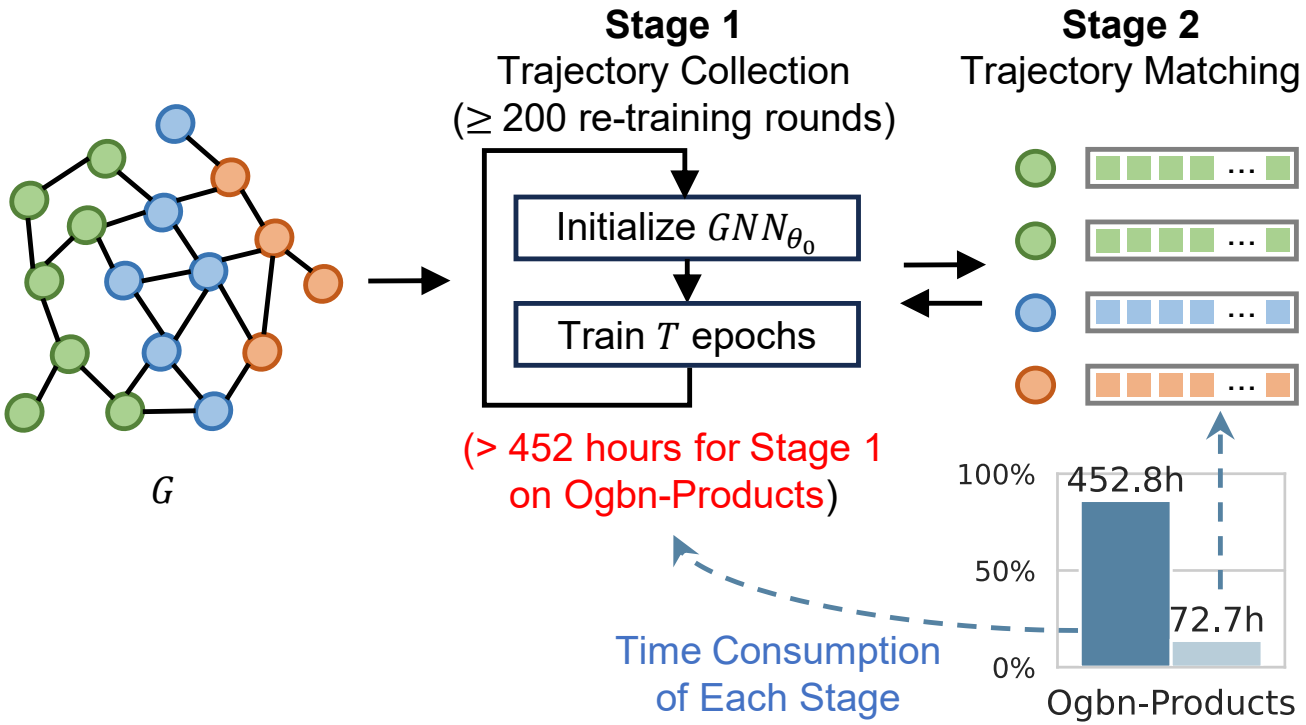
154 training nodes

Test accuracies
GCN: 89.4%
SGC: 89.6%
APPNP: 87.8%
GraphSAGE: 89.1%

Related Work

- **Gradient Matching:** Match GNN gradients between graphs.
- **Distribution Matching:** Align embedding distributions.
- **Trajectory Matching:** Match training dynamics over multiple steps.

Challenge: Inefficient GNN Re-training[2]



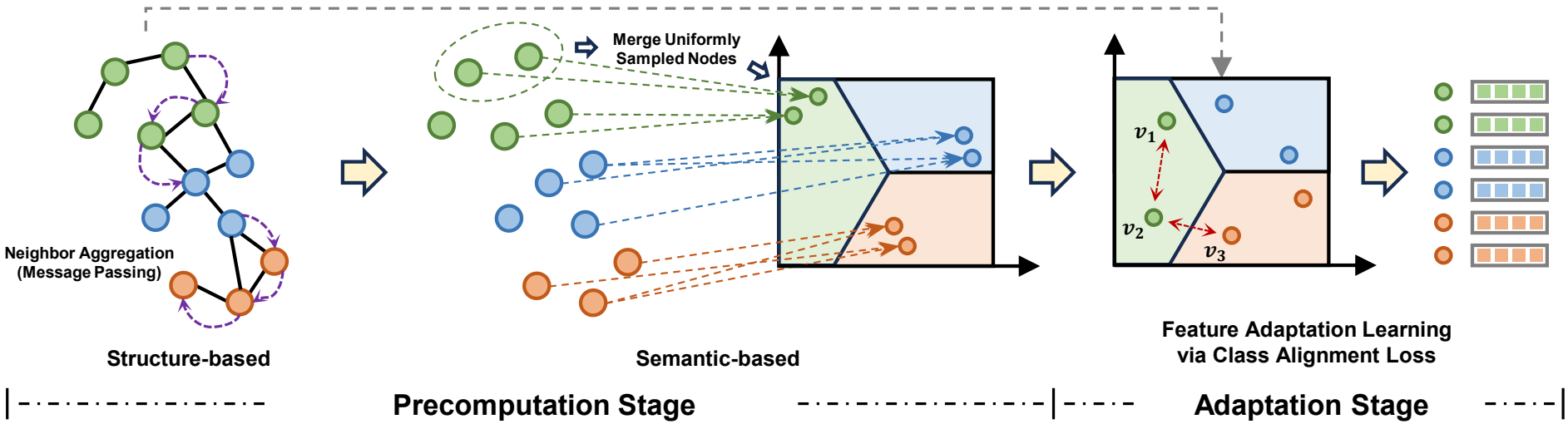
Research Question

Can we achieve efficient and effective GC without GNN re-training?

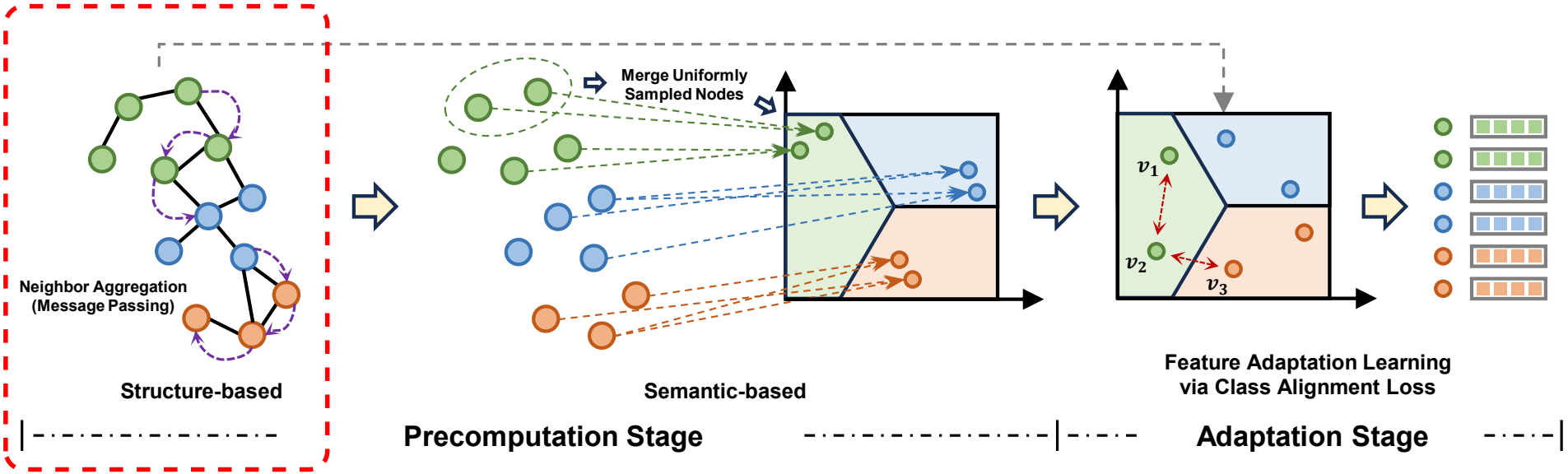
[2] Zhang et al. Navigating Complexity: Toward Lossless Graph Condensation via Expanding Window Matching, ICML 2024

Our Proposal: GC via Precompute-then-Adapt (GCPA)

1. **Precomputation:** Extract structural & semantic information.
2. **Adaptation:** Refine features via class-wise alignment & diversity.

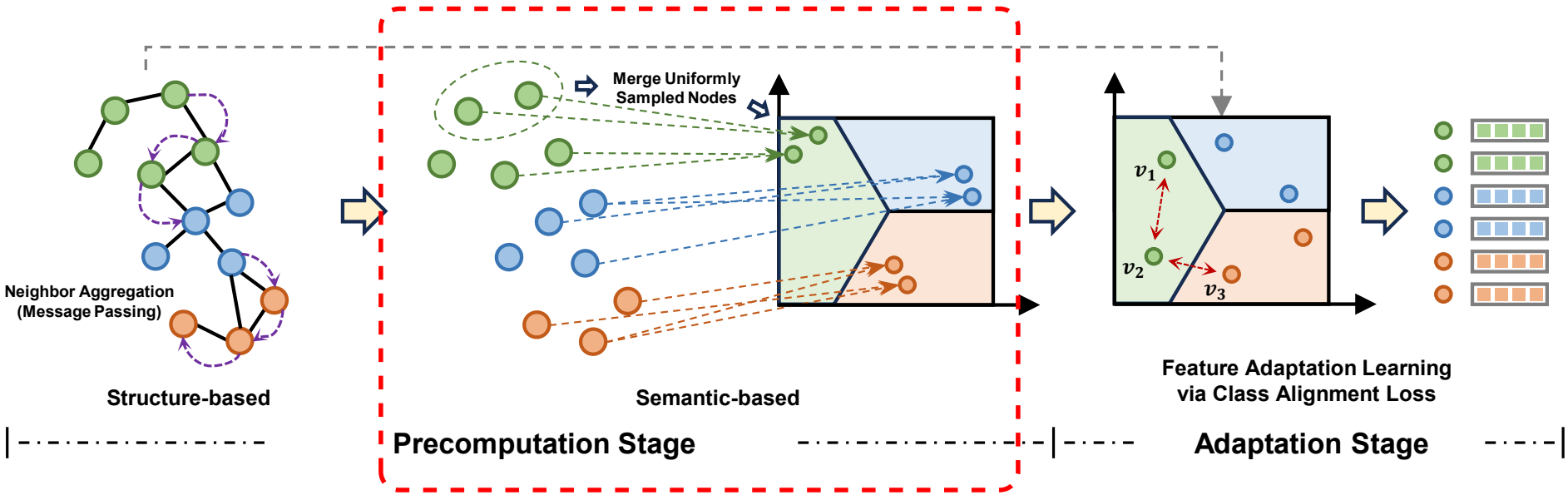


Structure-based Precomputation



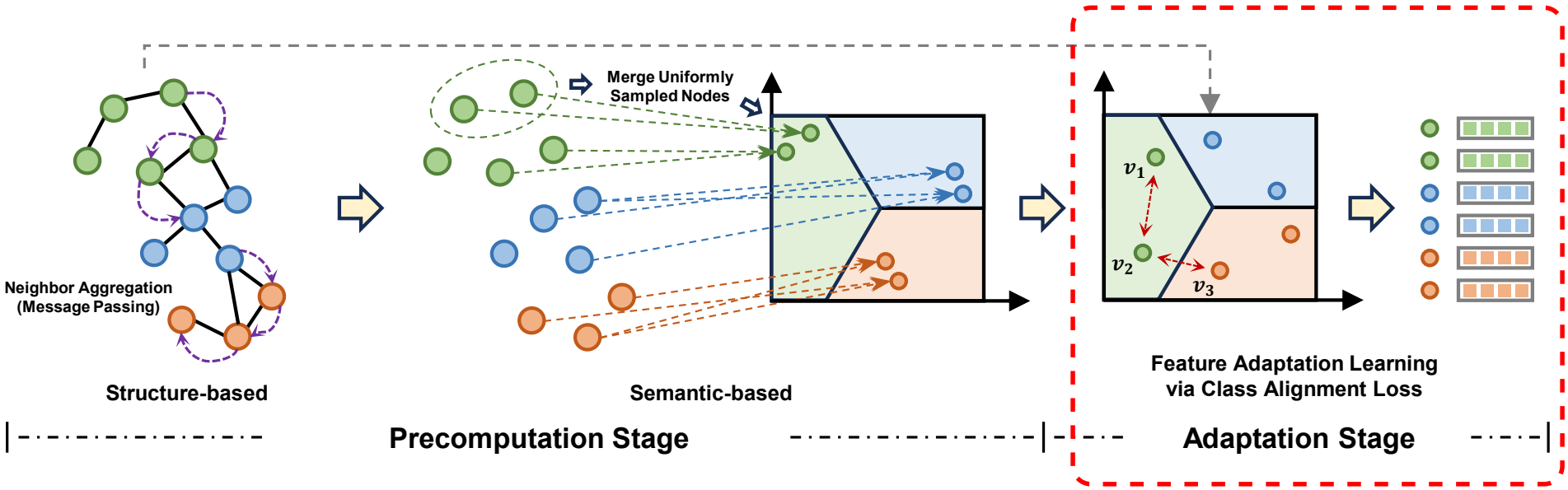
Aggregate multi-hop neighbors.

Semantic-based Precomputation



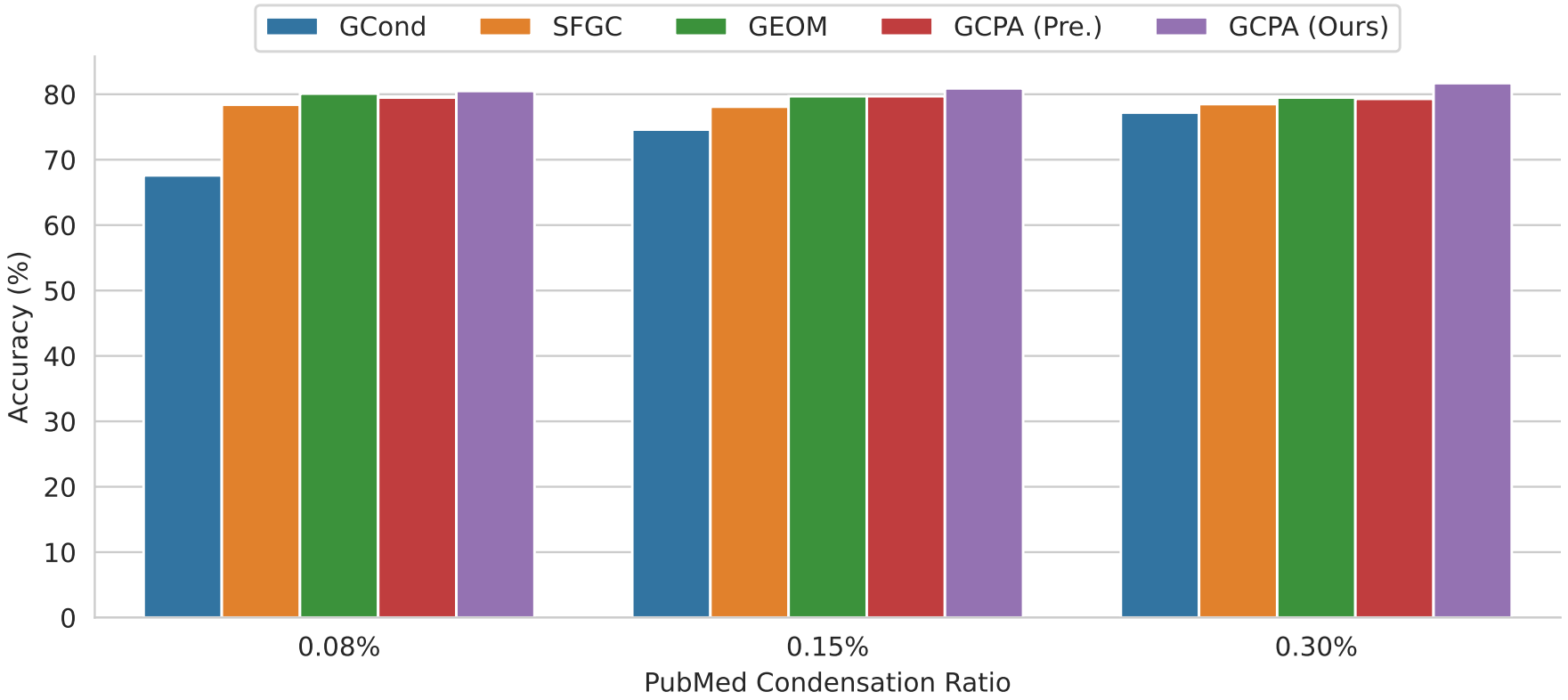
Mean-pool over uniformly sampled nodes per class.

Adaptation Learning



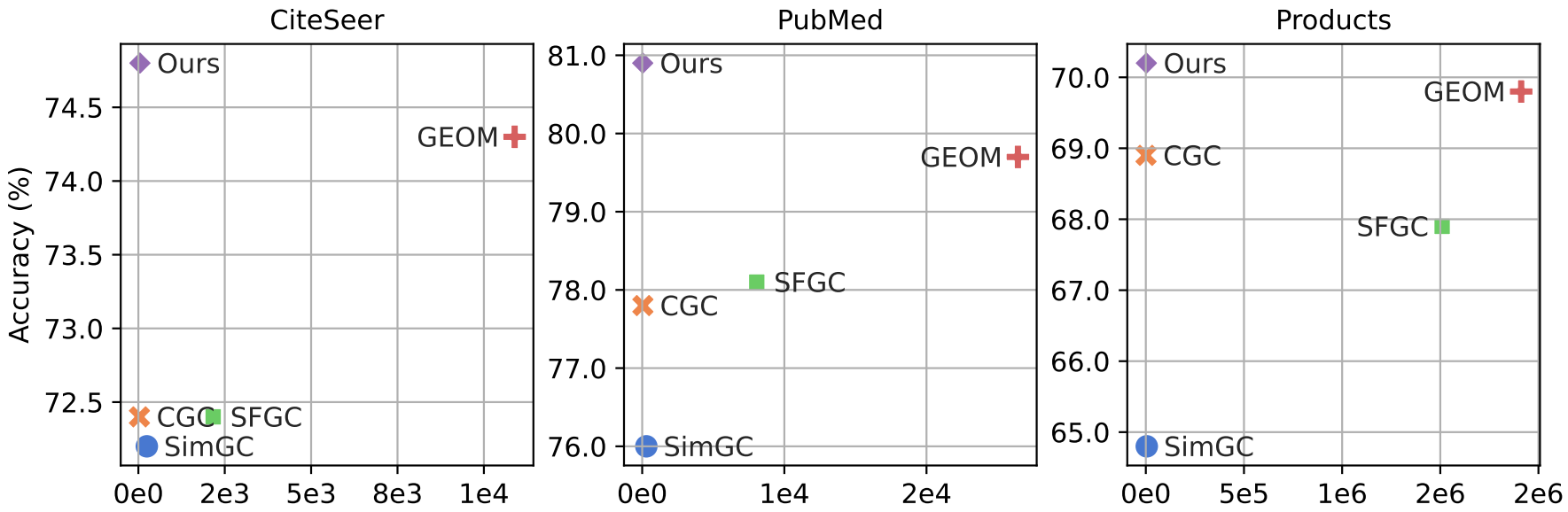
Align same-class features; Promote diversity to avoid collapse.

Performance Comparison



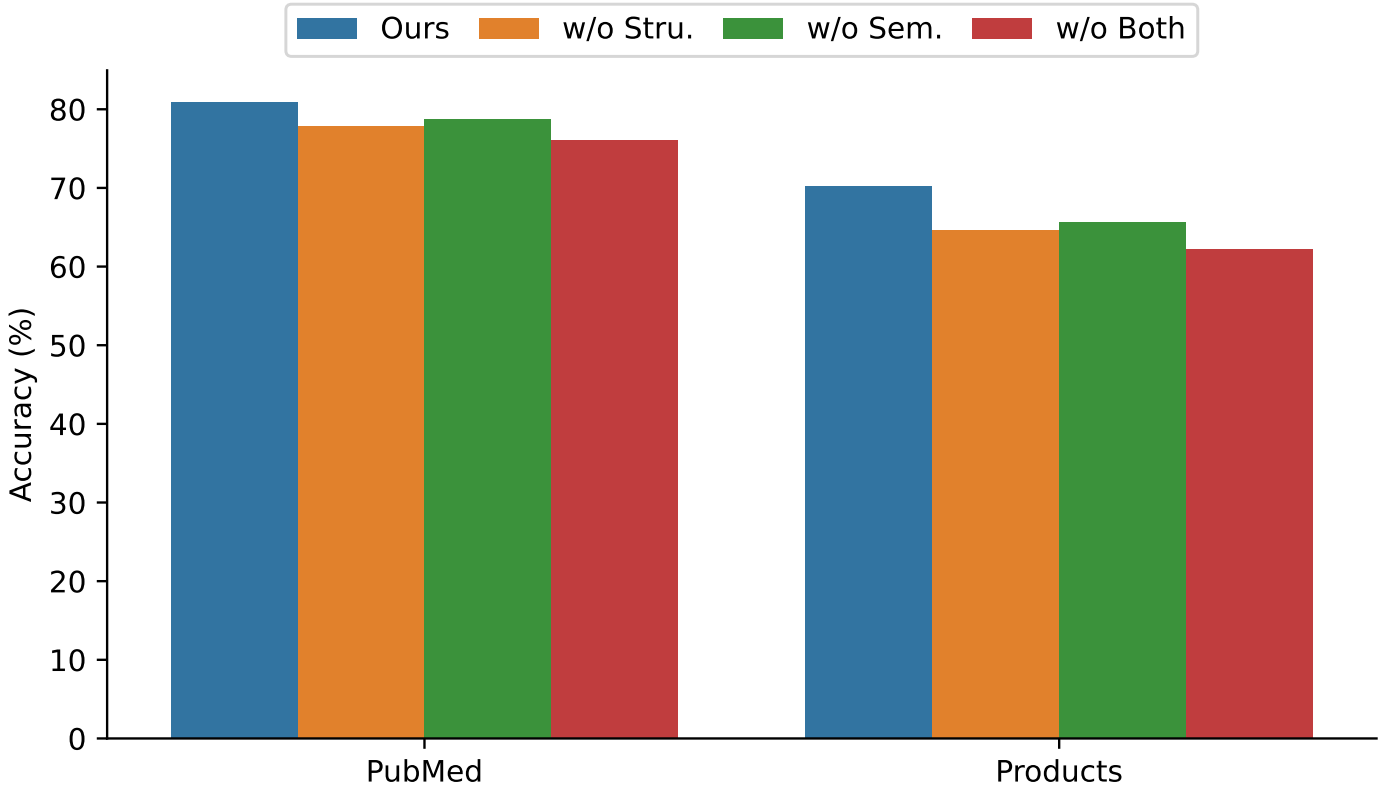
Precomputation + adaptation achieve SOTA.

Performance vs. Efficiency



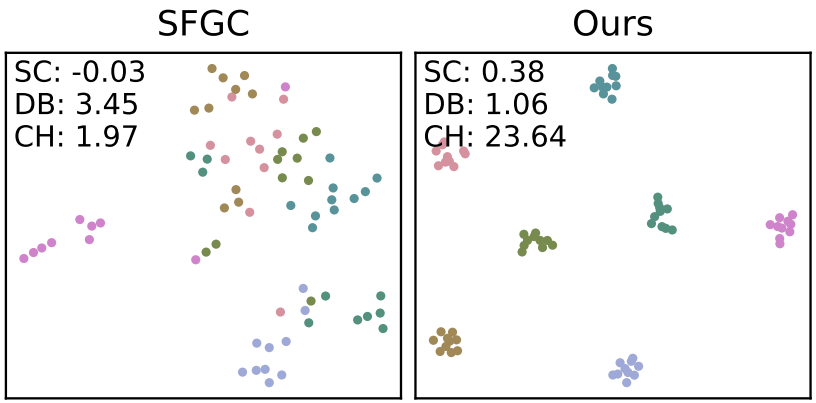
Strong tradeoff between performance and efficiency.

Ablation on Precomputation Stage.

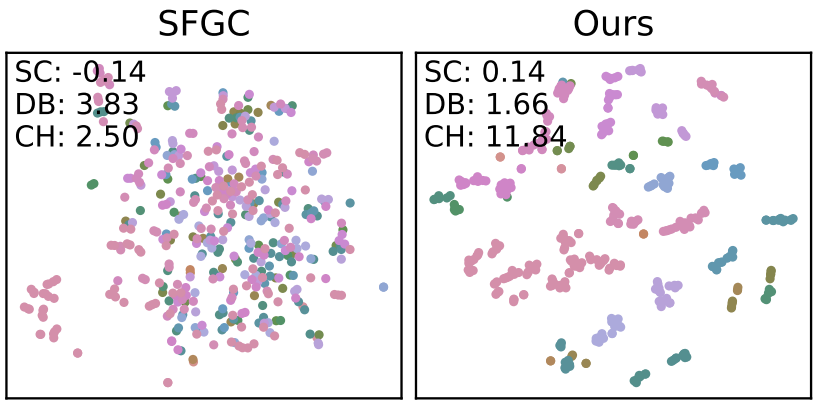


Both structure and semantics contribute to performance.

Feature Visualization



(a) Cora



(b) Arxiv

Clear class boundaries and superior clustering metrics.

Conclusion and Future Work

- Existing graph condensation methods require costly re-training.
- We propose GCPA for efficient GC via precomputation & adaptation.
- GCPA achieves SOTA performance with up to 2,455× speedup.

Future Work:

- Explore more precomputation approaches.
- Extend to more graph types.

Acknowledgements

- Supported by NRF Singapore and IMDA under Trust Tech Funding Initiative.

- Code: `https://github.com/Xtra-Computing/GCPA`