Website link

Read paper

UGC: Universal Graph Coarsening

Mohit Kataria[1], Sandeep Kumar[2,1], Jayadeva[2,1]

[1]Yardi School of Artificial Intelligence, [2]Department of Electrical Engineering

Indian Institute of Technology, Delhi

Mohit Kataria
Mohit.Kataria@scai.iitd.ac.in

Sandeep Kumar
ksandeep@ee.iitd.ac.in

Jayadeva
jayadeva@ee.iitd.ac.in

Outline

- **Background**
- Proposed framework UGC
- Quality checks for the coarsened graph
- Experiments
- Conclusion

Graph Coarsening

- The objective is to reduce an input graph $\mathcal{G}$(V, A, X) with *N*-nodes into a new graph $\mathcal{G}_c$(V′, A′, X′) with *n*-nodes.
- The **G**raph **C**oarsening (GC) problem requires learning of a coarsening matrix **C**, which defines the linear mapping from V → V′.

Original Graph

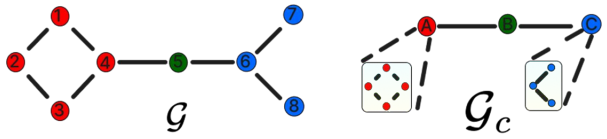
- Feature size $X \in \mathbb{R}^{N \times d}$
- Vertices are of order $\mathcal{O}(N)$
- Edges are of order $\mathcal{O}(N^2)$
- Features are of order $\mathcal{O}(Nd)$

Coarsened Graph

- Feature size $X' \in \mathbb{R}^{n \times d}$
- Vertices are of order $\mathcal{O}(n)$
- Edges are of order $\mathcal{O}(n^2)$
- Features are of order $\mathcal{O}(nd)$

Toy Example

$$\mathcal{C}^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$



- Every non-zero entry C_{ij} denotes the mapping of the i^{th} node of $\mathcal{G}$ to the j^{th} super-node $\mathcal{G}_c$.
- A valid C matrix must belong to set S defined as

$$\mathcal{S} = \left\{ \mathcal{C} \in \mathbb{R}^{N \times n}, \mathcal{C}_{ij} \in \{0,1\}, \|\mathcal{C}_i\| = 1, \langle \mathcal{C}_i, \mathcal{C}_j \rangle = 0, \forall i \neq j, \langle \mathcal{C}_l, \mathcal{C}_l \rangle = d_i, \|\mathcal{C}_i^T\|_0 \geq 1 \right\} \quad (1)$$

- $\|\mathcal{C}_i^T\|_0 \geq 1$ makes sure that no supernode is empty and $\langle \mathcal{C}_i, \mathcal{C}_j \rangle = 0$ ensures that each node of $\mathcal{G}$ is mapped to a unique supernode.

What has been done on graph coarsening?

- **Optimization and Heuristics**
  - Loukas 2018: Two variants, edge-based (LVE) and neighborhood-based (LVN)
  - Kumar 2023: FGC
  - Dorfler 2013: Kron reduction
  - Chen 2011: Algebraic Distance
  - Livne 2011: Affinity GS
  - Dhillon 2007: Heavy Edge
- **GNN based graph condensation**
  - Jin 2021: GCond
  - Zheng 2023: SFGC
- **Scaling GNN using coarsening methods**
  - Huang 2021: SCAL
  - Cai 2021: GOREN

Research gaps

- Existing optimization and heuristic based graph coarsening methods are computationally demanding.
- Existing graph condensation methods require full graph training to get a condensed graph, due to which these methods are not suitable for the scalability of GNN models.
- Lack of graph coarsening methods for heterophilic graphs.
- How to employ graph coarsening methods for scalibility of graph neural networks.

Research gaps

- Existing graph coarsening methods are computationally demanding.

  UGC uses a hashing-based method, which is super fast.

- Existing graph condensation methods require full graph training to get a condensed graph.

  UGC doesn't require full graph training.

- Lack of graph coarsening methods for heterophilic graphs.

  UGC uses both feature level and structure level information to handle heterohily graphs.

- How to employ graph coarsening methods for scalibility of graph neural networks.

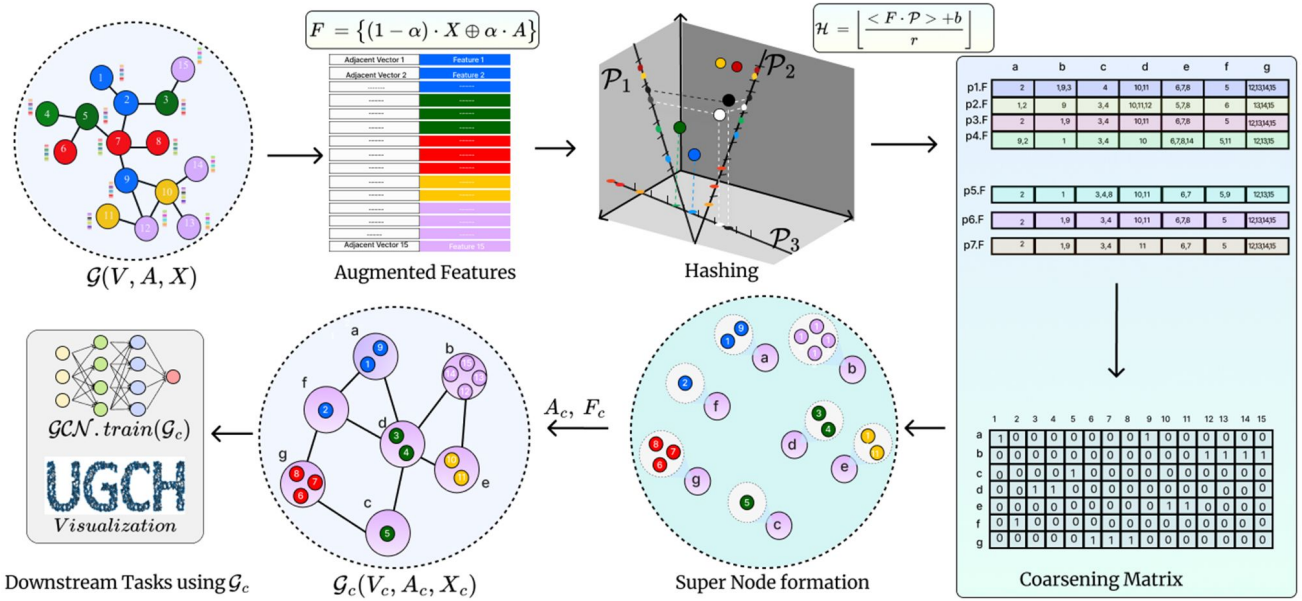
  A coarsened graph can be used to scale GNN based methods.

Outline

- Background
- **Proposed framework UGC**
- Quality checks for the coarsened graph
- Experiments
- Conclusion

Proposed framework: UGC

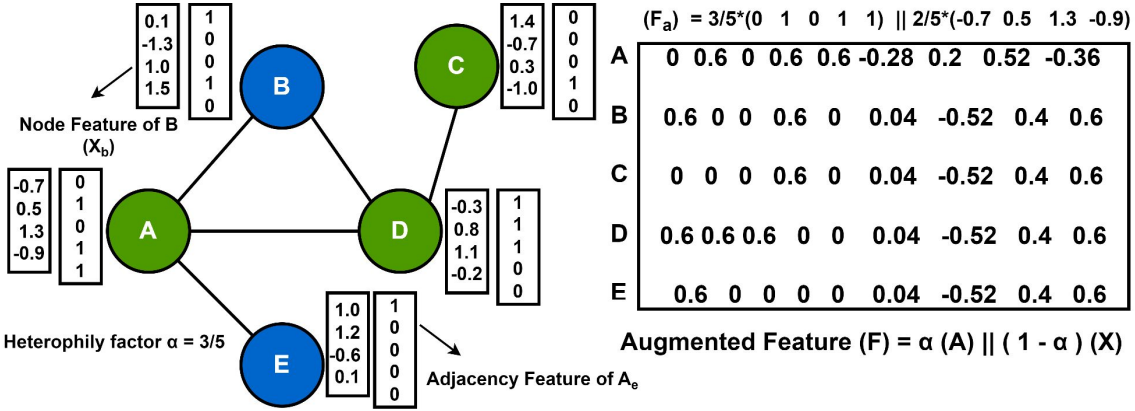
The UGC framework comprises three main components: (a) construction of an *augmented feature matrix*; (b) construction of a *coarsening matrix;* and (c) construction of a *coarsened graph*.



Augmented feature matrix

- In order to create a universal GC framework suitable for both homophilic as well as heterophilic datasets, it is important to consider features at both i) the node level, i.e., features, and ii) the structure-level, i.e., adjacency matrix, together.

- A heterophily factor $0 \leq \alpha \leq 1$ may be used to denote the degree of heterophily. α is calculated as the fraction of edges between nodes of different classes to the total number of edges.



Coarsening matrix

- UGC uses the Locality Sensitive Hashing (LSH) technique to create coarsening matrix. LSH is defined as

> **Definition**: Let d be a distance measure, and let $d_1 < d_2$ be two distances. A family of functions F is said to be $(d_1, d_2, p_1, p_2)-$sensitive if for every $f \in F$ the following two conditions hold:
>
> 1. If $d(x, y) \leq d_1$ then probability $[f(x) = f(y)] \geq p_1$
> 2. If $d(x, y) \geq d_2$ then probability $[f(x) = f(y)] \leq p_2$

- $F_i \in \mathbb{R}^d$ represent the augmented feature vector of node $\boldsymbol{v_i}$.

- Let $\mathcal{W} \in \mathbb{R}^{d \times l}$ and $b \in \mathbb{R}^l$ be the random hashing matrices with *l* hash functions. The hash indices generated by the k^{th} projector for the i^{th} node is given as $h_i^k = \lfloor \frac{1}{r} * (\mathcal{W}_k \cdot F_i + b_k) \rfloor$

- The hash value assigned to the i^{th} node is given by $h_i = maxOccured\{h_i^1, h_i^2 .... h_i^l\}$ which defines the linear mapping $\pi : V \rightarrow V'$ and construction of the coarsening matrix.

Coarsened graph

- A pair of super-nodes, say $\widetilde{v_i}$ and $\widetilde{v_j}$, in $\mathcal{G}_c$ are connected; if any of the nodes $u \in \pi^{-1}(\widetilde{v_i})$ has an edge to any of the nodes, say $v \in \pi^{-1}(\widetilde{v_j})$ in $\mathcal{G}$, i.e., $\exists\, u \in \pi^{-1}(\widetilde{v_i}), v \in \pi^{-1}(\widetilde{v_j})$ such that A_{uv} != 0.

- The coarsened graph ($\mathcal{G}_c$) is weighted, and the weight assigned to the edge between nodes $\widetilde{v_i}$ and $\widetilde{v_j}$ is given by, $\widetilde{A}_{ij} = \sum_{(u \in \pi^{-1}(\widetilde{v_i}), v \in \pi^{-1}(\widetilde{v_j}))} A_{uv}$ and the adjacency matrix of $\mathcal{G}_c$ is defined as $\widetilde{A} = \mathcal{C}^T A \mathcal{C}$.

- Supernode features are calculated as $\widetilde{F} = \mathcal{C}^T F$

Outline

- Background
- Proposed framework UGC
- **Quality checks for the coarsened graph**
- Experiments
- Conclusion

Quality checks for the coarsened graph

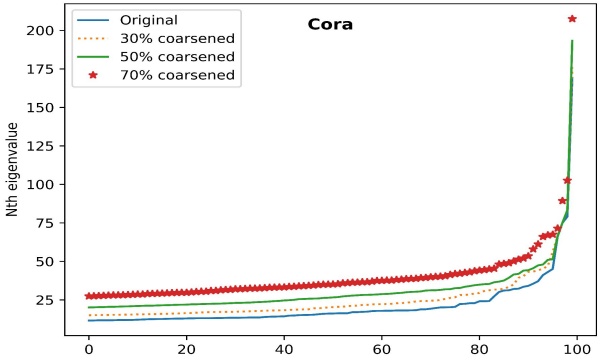
UGC employs different matrices to quantify the quality of the coarsened graph.

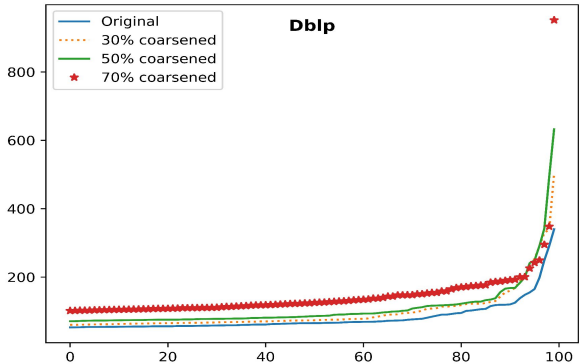
- Spectral Similarity
- ε- similarity
- LSH similarity
- Using node classification accuracy when trained on the coarsened graph.

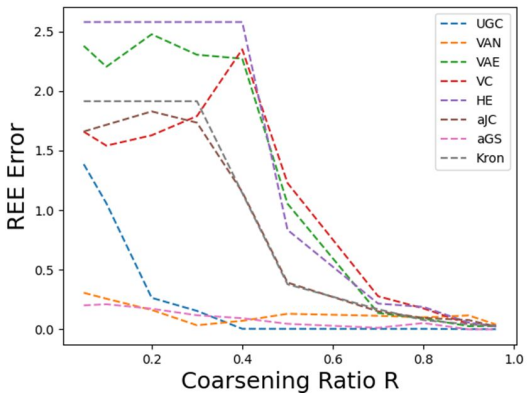
Spectral Similarity

- Relative Eigen Error (REE) gives the means to quantify the measure of the eigen properties of the original graph $\mathcal{G}$ that are preserved in the coarsened graph $\mathcal{G}_c$.

- REE is defined as follows: $REE(L, L_c, k) = \frac{1}{k} \sum_{i=1}^{k} \frac{|\tilde{\lambda_i} - \lambda_i|}{\lambda_i}$ where λ_i and $\tilde{\lambda_i}$ are the top *k* eigenvalues of the original graph Laplacian L and the coarsened graph Laplacian L_c.







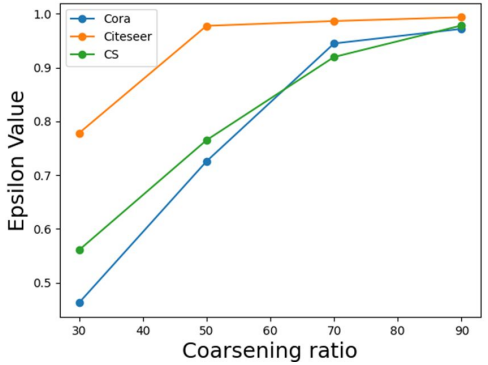
ε-similarity

- UGC gives a coarsened graph, which satisfies the ε-similarity theorem, which is stated as:

Theorem: The input graph $\mathcal{G}(L, F)$ and the coarsened graph $\mathcal{G}_c(L_c, \widetilde{F})$ obtained using the proposed UGC algorithm are ϵ-similar with $\epsilon \geq 0$, i.e.,

$$(1-\epsilon)\|F\|_L \leq \|\widetilde{F}\|_{L_c} \leq (1+\epsilon)\|F\|_L$$

where L and L_c are the laplacian matrices of $\mathcal{G}$ and $\mathcal{G}_c$ respectively.



- To give a strict bound on the ε(< 1) we updated

$\widetilde{F}$ to $\widehat{F}$ by minimizing the term

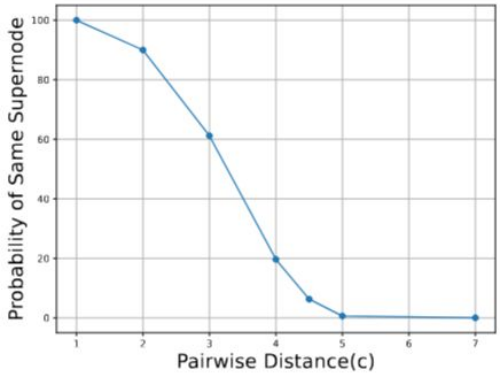
$\min_{\widehat{F}} f(\widehat{F}) = \mathrm{tr}(\widehat{F}^T \mathcal{C}^T L \mathcal{C} \widehat{F}) + \frac{\alpha}{2}\|\mathcal{C}\widehat{F} - F\|_F^2$

which aim to enforce the Dirichlet smoothness condition in super-node features

LSH similarity

- The LSH family used in our framework ensures that the probability of two nodes going to the same supernode is directly related to the distance between their features.

Theorem: The probability that two nodes v and u will collide and go to a super-node under a hash function drawn uniformly at random from a 2-stable distribution is inversely proportional to $c = ||v - u||_2$ and it is represented by $p(c) = Pr_{w,b}\left[h_{w,b}(v) = h_{w,b}(u)\right] = \int_0^r \frac{1}{c} f_p\left(\frac{t}{c}\right)\left(1 - \frac{t}{r}\right) dt.$



Outline

- Background
- Proposed framework UGC
- Quality checks for the coarsened graph
- **Experiments**
- Conclusion

Experiments

The conducted experiments establish the performance of UGC concerning

- Computational efficiency,
- Preservation of spectral properties,
- Potential extensions of the coarsened graph $\mathcal{G}_c$ into real-world applications. We have used node classification tasks on real-world datasets.
- Model agnostic behaviour of UGC.

Run time

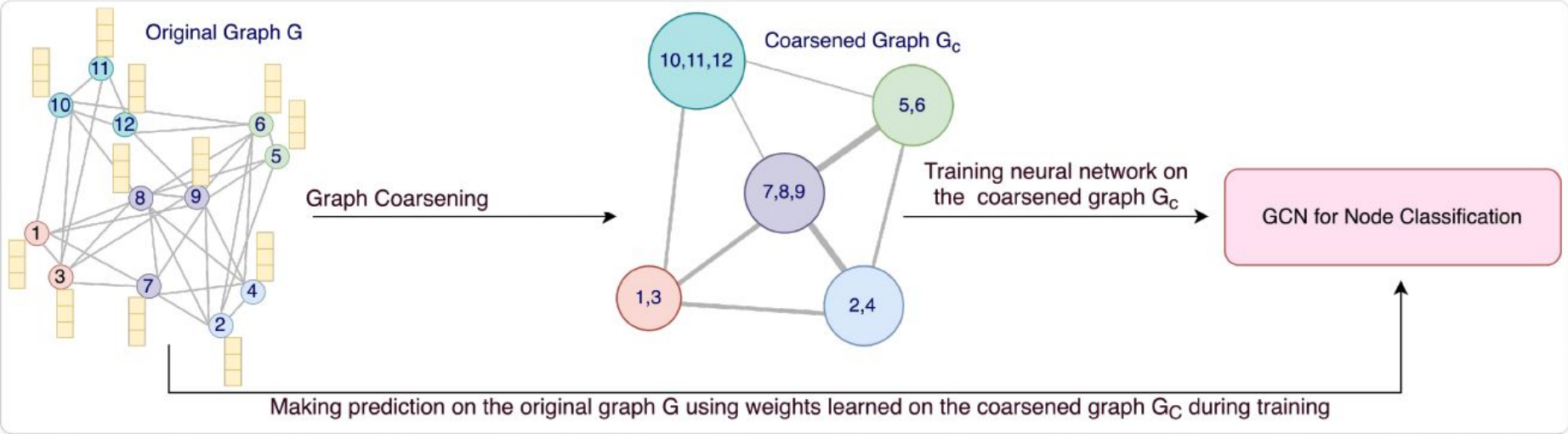
- UGC's main contribution lies in its computational efficiency. The time required to compute the coarsening matrix $\boldsymbol{C}$ is summarized in below Table.

Data/Method	Cora	Cite.	CS	PubMed	DBLP	Physics	Flickr	Reddit	Yelp	Squirrel	Cham.	Cor.	Texas	Film
Var. Neigh.	6.64	8.72	23.43	24.38	22.79	58.0	OOM	OOM	OOM	33.26	12.2	1.34	0.63	27.67
Var. Edges	5.34	7.37	16.72	18.69	20.59	67.16	OOM	OOM	OOM	46.45	12.65	1.31	0.76	26.6
Var. Cliq.	7.29	9.8	24.59	61.85	38.31	69.80	OOM	OOM	OOM	28.91	10.55	1.56	1.14	33.04
Heavy Edge	0.7	1.41	7.50	12.03	8.39	39.77	OOM	OOM	OOM	18.08	5.41	1.62	1.17	11.79
Alg. Dist	0.93	1.55	9.63	10.48	9.67	46.42	OOM	OOM	OOM	18.03	5.24	1.58	0.81	12.65
Affinity GS	2.36	2.53	169.05	168.3	110.9	924.7	OOM	OOM	OOM	20.00	5.83	1.81	1.24	20.65
Kron	0.63	1.37	8.72	5.81	7.09	34.53	OOM	OOM	OOM	20.62	7.25	1.73	0.97	12.29
UGC	**0.41**	**0.71**	**3.1**	**1.62**	**1.86**	**6.4**	**8.9**	**16.17**	**170.91**	**2.14**	**0.49**	**0.04**	**0.03**	**1.38**

- UGC is able to coarsen down massive datasets like *Yelp (716.8k nodes)*, which was previously not possible.
- UGC is the fastest graph coarsening method.

Scaling GNN via graph coarsening

To scale the training process, we used coarsened graph $\mathscr{G}_c$ to train a GNN model; all the predictions are made on test data from the original graph.



Node classification accuracy

- UGC demonstrated superior performance compared to existing methods in 7 out of the 9 datasets. Reported are the accuracy of the GNN models when trained with 50% coarsen graph.

Data/Method	Cora	DBLP	PubMed	Physics	Squirrel	Cham.	Cor.	Texas	Film
Var.Neigh.	79.75	77.05	77.87	93.74	19.67	20.03	52.49	34.51	15.67
Var.Edges	81.57	**79.93**	78.34	93.86	20.22	29.95	55.32	30.59	21.8
Var.Clique	80.92	79.15	73.32	92.94	19.54	31.92	58.8	33.92	20.35
Heavy Edge	79.90	77.46	74.66	93.03	20.36	33.3	54.67	29.18	19.16
Alg. Dis.	79.83	74.51	74.59	93.94	19.96	28.81	**59.91**	18.61	19.23
Aff. GS	80.20	78.15	80.53	93.06	20.00	27.58	54.06	21.18	20.34
Kron	80.71	77.79	74.89	92.26	18.03	29.1	55.02	31.14	17.41
UGC(fea.)	83.92	75.50	**85.65**	94.70	20.71	29.9	55.6	52.4	22.6
UGC(fea+Ad)	**86.30**	75.50	84.77	**96.12**	**31.62**	**48.7**	54.7	**57.1**	**25.4**

- Results from four diverse models, namely GCN, GraphSage, GIN, and GAT, have been incorporated to demonstrate the robustness and model-agnostic nature of UGC.

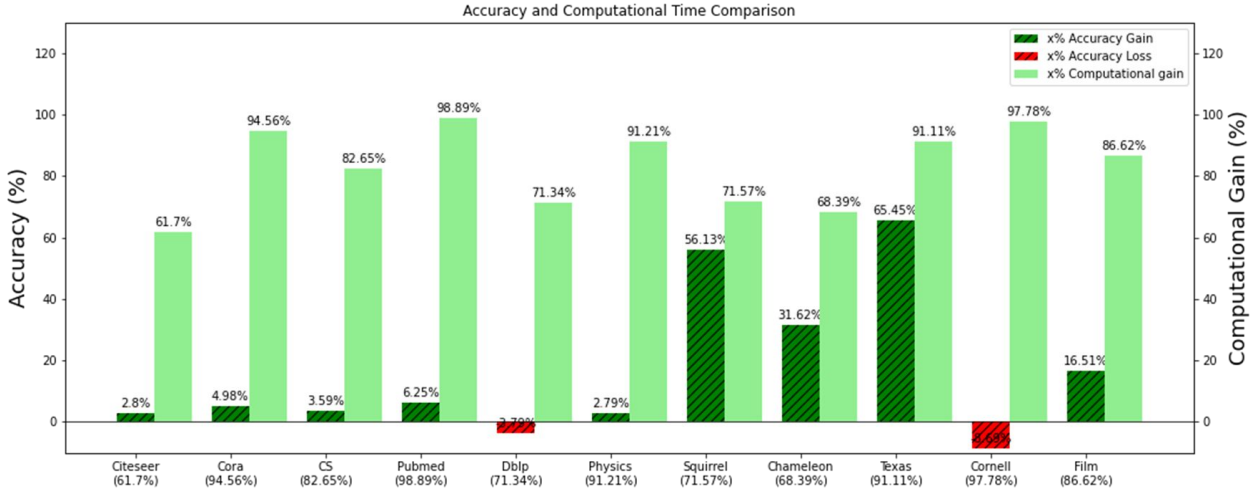
Model/Data	Cora	Pubmed	Physics	Squirrel
GCN	86.30	84.77	96.12	31.62
GraphSage	69.39	85.72	94.49	61.23
GIN	67.23	84.12	85.15	44.72
GAT	74.21	84.37	92.60	48.75

Outline

- Background
- Proposed framework UGC
- Quality checks for the coarsened graph
- Experiments
- **Conclusion**

Conclusion

- UGC is the fastest graph coarsening method.
- UGC preserves spectral properties.
- UGC satisfy ε-similarity and LSH similarity.
- UGC scales training of GNN models.



Thank you