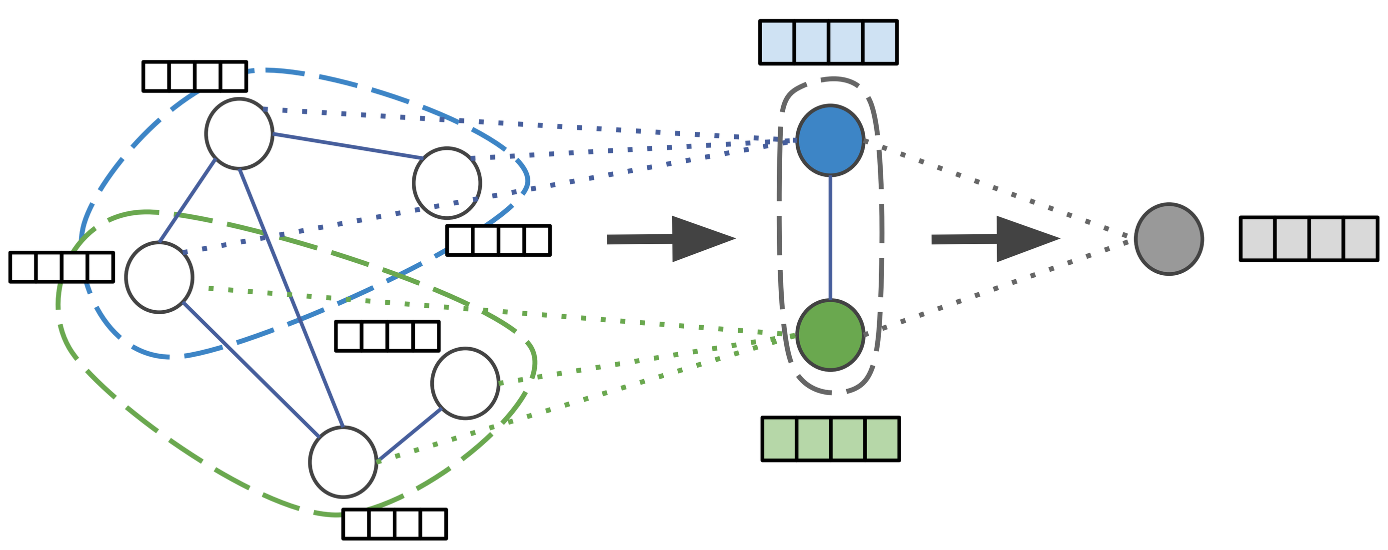
ASAP: Adaptive Structure Aware Pooling for Learning Hierarchical Graph Representations

Ekagra Ranjan[1], Soumya Sanyal[2], Partha Talukdar[2]

[1]Indian Institute of Technology Guwahati, [2]Indian Institute of Science Bangalore

Graph Pooling



Graph Pooling

- summarizing a graph to generate graph embedding.
- **Hierarchical pooling** - pooling recursively to form smaller graph. This enables processing of a larger area of the graph with the same compute.

Contributions

- **ASAP**, a sparse pooling operator capable of capturing local subgraph information hierarchically to learn global features with better edge connectivity in pooled graph.
- **Master2Token (M2T)**, a new self-attention framework, which is better suited for pooling.
- **LEConv**, a new convolution operator that can learn functions of local extremas in a graph substructure.

Master2Token (M2T) self-attention

Goal: Learn an overall representation of a cluster of nodes by attending to the relevant members using self-attention.

Approach

- Given a cluster, create a master-query $\mathbf{m}_i$ which is representative of the constituent nodes within cluster.
- Attend to all constituent nodes using additive-attention to learn the cluster representation $\mathbf{x}^c$

$$x_i^c = \sum_{j \in \mathcal{N}_i} \alpha_{i,j} x_j$$

where $\alpha_{i,j} = softmax(\vec{w}^T \sigma(W m_i \parallel x_j))$ and $m_i = \max\{x_{i1}, ..., x_{i|\mathcal{N}_i|}\}$

Advantage

- Attention computed w.r.t. a global query instead of a specific member is more intuitive for set aggregations.

Local Extrema Convolution (LEConv)

Goal: Select clusters corresponding to information local extremas to enable sampling of representative clusters from all parts of the graph.

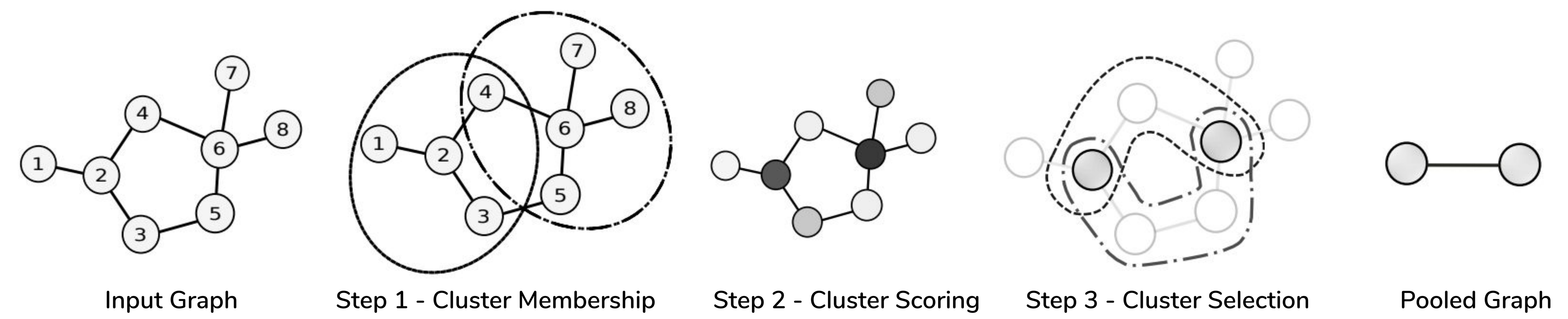
Formulation

$$\phi_i = \sigma(x_i W_1 + \sum_{j \in \mathcal{N}_i} A_{i,j}(x_i W_2 - x_j W_3))$$

Advantage

- Unlike GCNs, LEConv can learn difference functions.
- Helpful in scoring clusters by considering both its global and local importance through the use of self-loops.

ASAP Overview



Details of ASAP

- **Input:** Graph with N nodes having feature representation $X \in \mathbb{R}^{N \times d}$ and adjacency matrix $A \in \mathbb{R}^{N \times N}$.
- **Cluster Membership:** Consider clusters of 1-hop neighborhood with each node as a medoid. Use M2T attention to compute node membership in a cluster. Construct the membership matrix $S_{i,j} = \alpha_{i,j}$.
- **Cluster Scoring:** Score the clusters using LEConv. Darker shade denotes higher score in Figure (c) above.
- **Cluster Selection:** Select a fraction of top scoring clusters as part of the pooled graph. Recompute the adjacency matrix using the cluster membership matrix of nodes in the selected clusters using $A_{new} = S^T A S$.

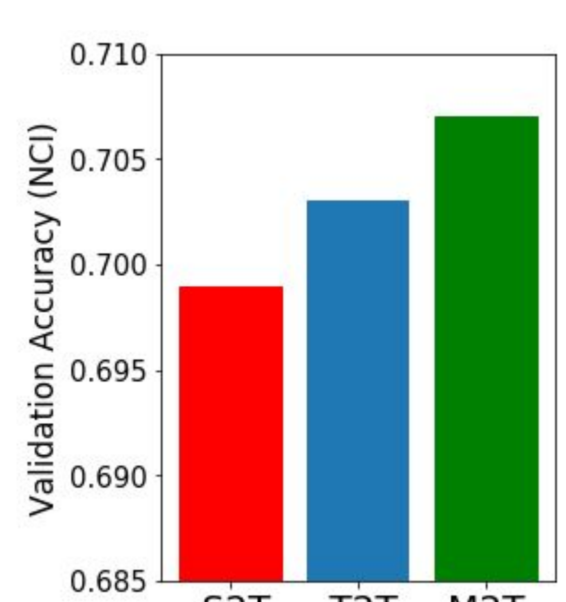
Results

Performance on Graph Classification: We report average accuracy and standard deviation for 20 random seeds. We observe that ASAP consistently outperforms all the baselines on all the datasets.

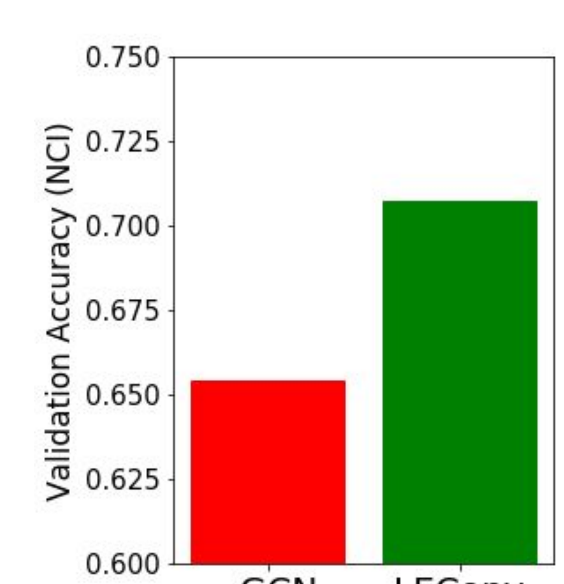
Method	D&D	PROTEINS	NCI1	NCI109	FRANKENSTEIN
SET2SET	71.60 ± 0.87	72.16 ± 0.43	66.97 ± 0.74	61.04 ± 2.69	61.46 ± 0.47
GLOBAL-ATTENTION	71.38 ± 0.78	71.87 ± 0.60	69.00 ± 0.49	67.87 ± 0.40	61.31 ± 0.41
SORTPOOL	71.87 ± 0.96	73.91 ± 0.72	68.74 ± 1.07	68.59 ± 0.67	63.44 ± 0.65
DIFFPOOL	66.95 ± 2.41	68.20 ± 2.02	62.32 ± 1.90	61.98 ± 1.98	60.60 ± 1.62
TOPK	75.01 ± 0.86	71.10 ± 0.90	67.02 ± 2.25	66.12 ± 1.60	61.46 ± 0.84
SAGPOOL	76.45 ± 0.97	71.86 ± 0.97	67.45 ± 1.11	67.86 ± 1.41	61.73 ± 0.76
ASAP (Ours)	$\mathbf{76.87 \pm 0.7}$	$\mathbf{74.19 \pm 0.79}$	$\mathbf{71.48 \pm 0.42}$	$\mathbf{70.07 \pm 0.55}$	$\mathbf{66.26 \pm 0.47}$

Ablation Study

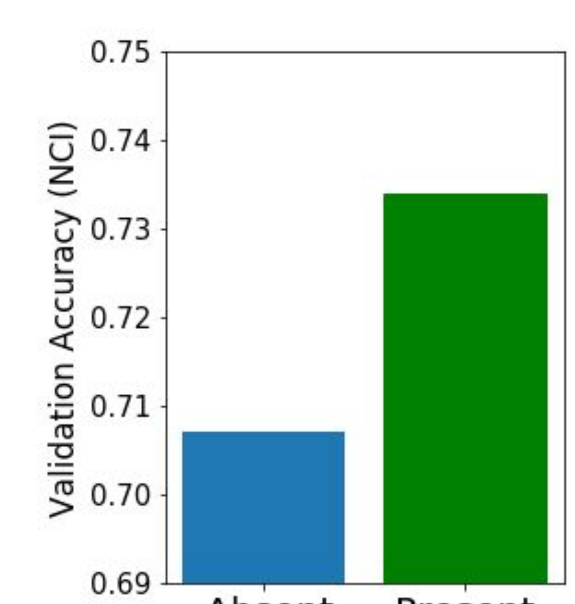
- **Effect of M2T Attention**: Master query is essential for cluster summarization.



- **Effect of LEConv scoring**: Difference improves performance as it can sample contrastive nodes.



- **Effect of Soft Edge Weights**: Membership information is crucial for graph pooling.



Desired Properties in Pooling

Property	DiffPool	TopK	SAGPool	ASAP
Sparse		✓	✓	✓
Node Aggregation	✓			✓
Soft Edge Weights	✓			✓
Variable number of clusters		✓	✓	✓

- Sparse pooling enables scalability to large graphs
- Node aggregation ensures better summarization of graph neighborhood
- Soft edge weights denote strengths/similarity between clusters
- Variable number of clusters can effectively handle graphs of any size

Source Code

github.com/malllabiisc/ASAP

Contact

ekagra.ranjan@gmail.com
soumyasanyal@iisc.ac.in