LOG CONFERENCE

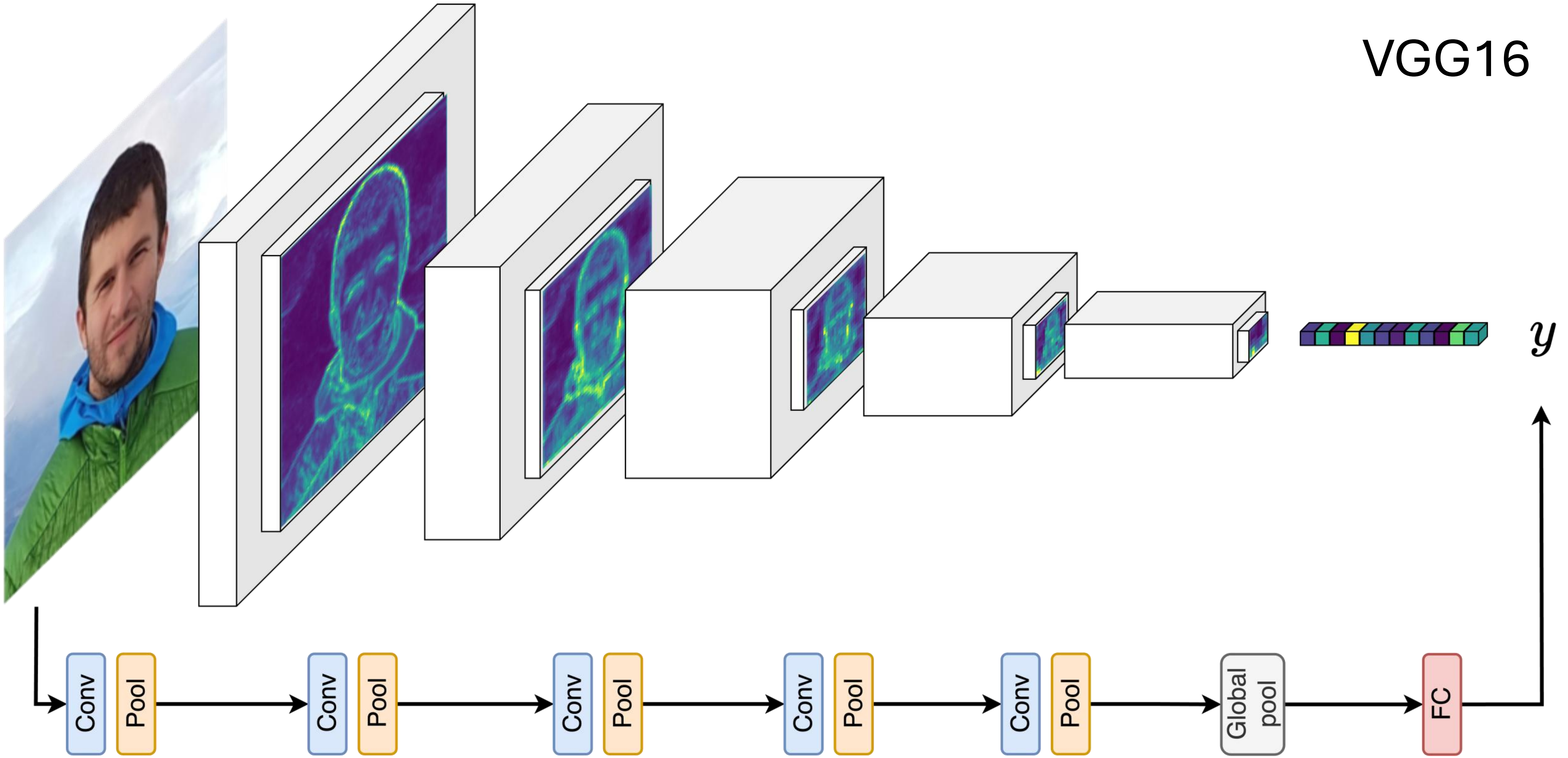
Siena – Italian Meetup

Hierarchical pooling in Graph Neural Networks

Filippo Maria Bianchi Carlo Abate

Introduction

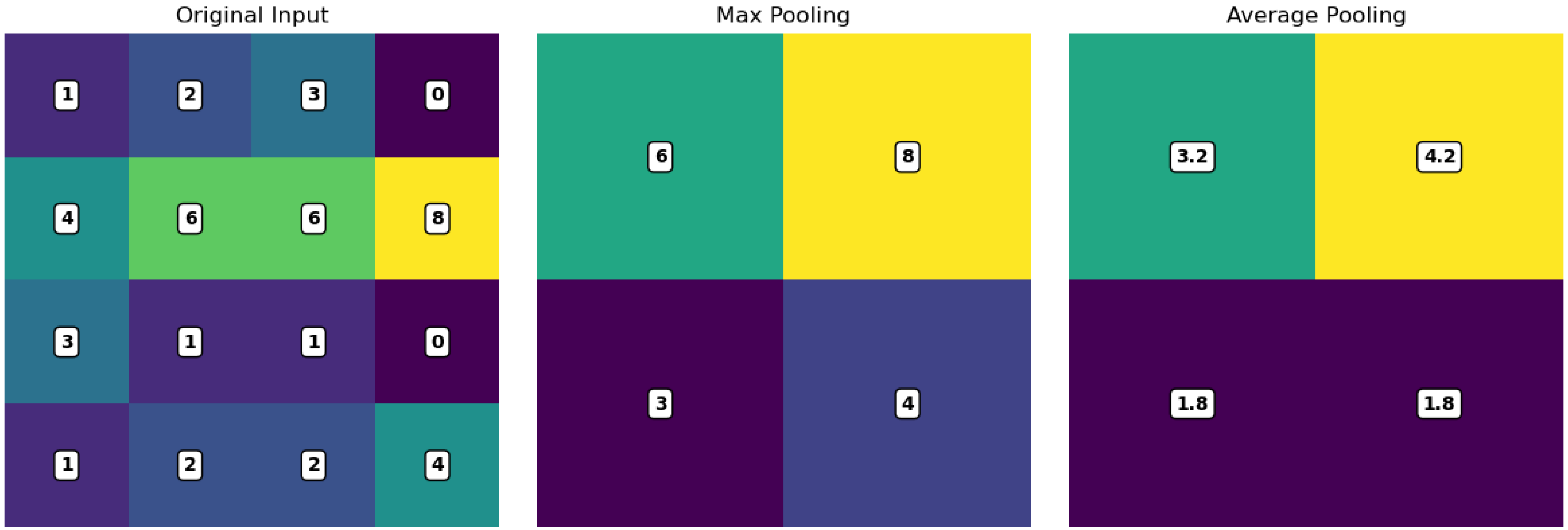
CNN architectures

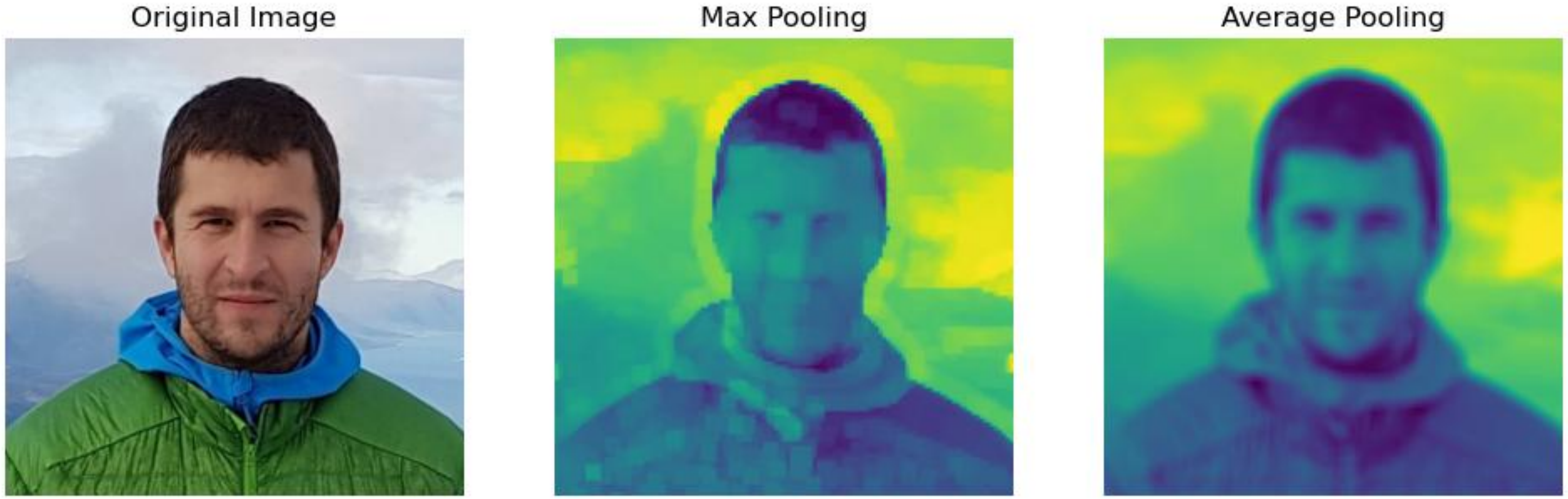


📖 Simonyan K. and Zisserman A., “Very Deep Convolutional Networks for Large-Scale Image Recognition”,

Pooling in CNNs

Extract local summaries from adjacent pixels

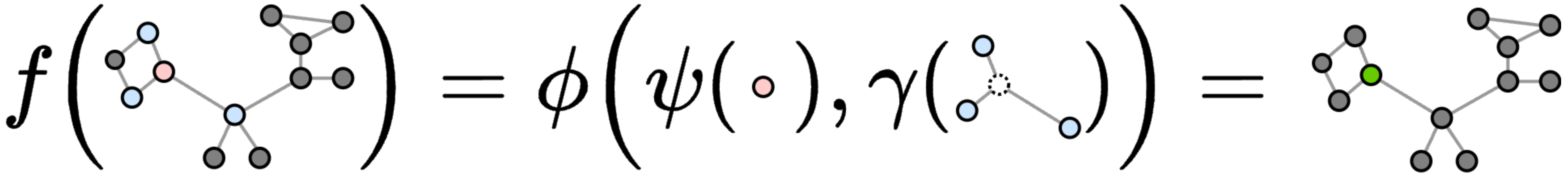




GNNs and MP

Graph Neural Networks

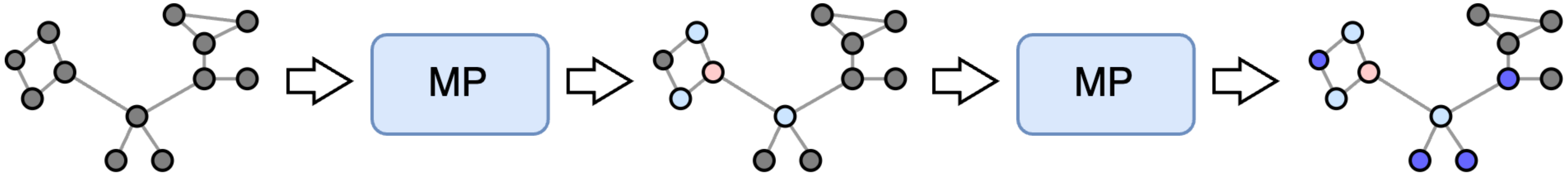
One or more Message-Passing (MP) layers learn a node representation for each node.



GNN architectures

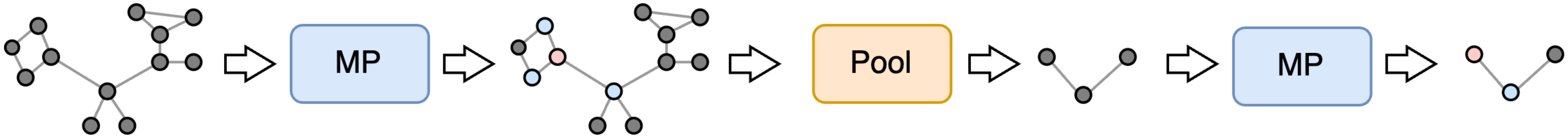
Flat (classic)

- Each MP progressively learns **more complex** representations and accounts for **longer** spatial relationships.



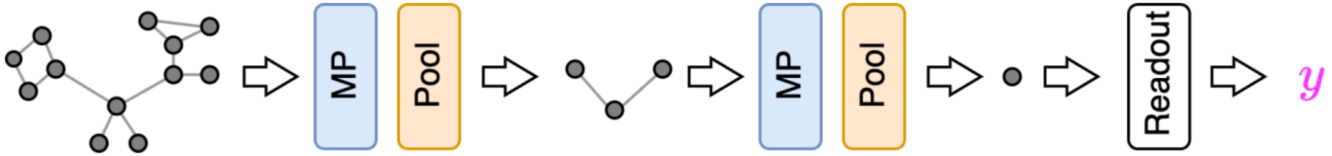
Hierarchical

- **Distill** global information from the graph
- Quickly **expand** the receptive field using fewer MP layers

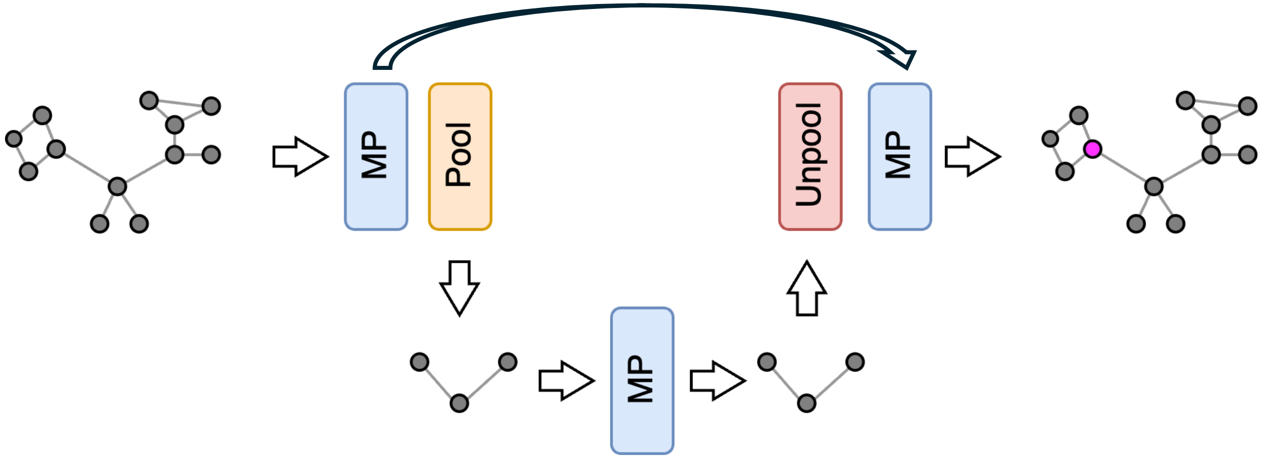


Hierarchical GNNs tasks

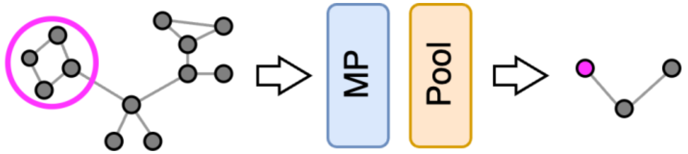
- Graph-level



- Node-level

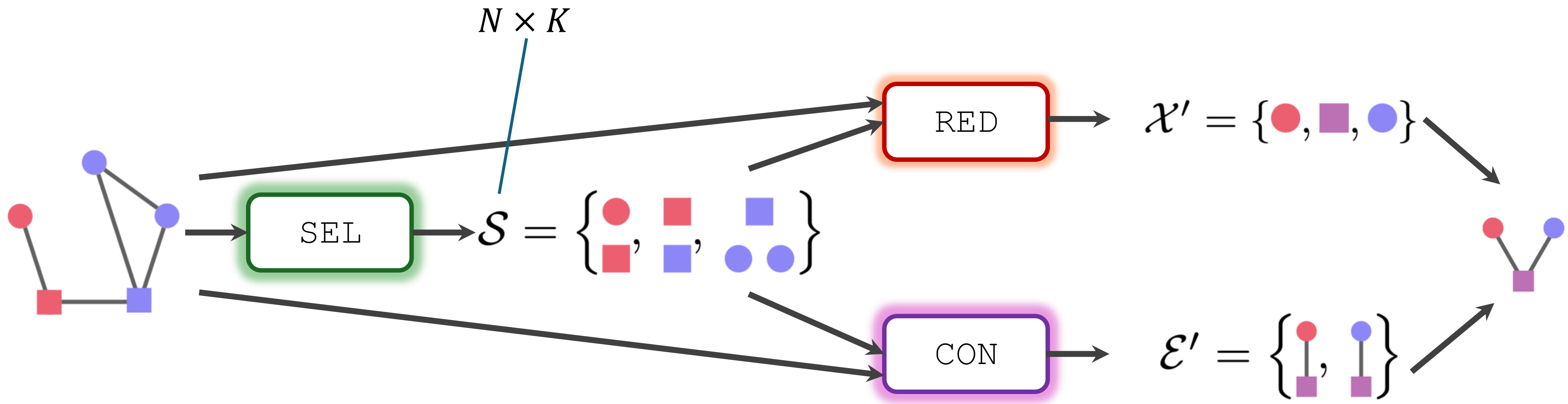


- Community-level



SRC – Select, Reduce, Connect

- Describe pooling operators as a composition of three functions.
- Different methods are obtained with different SEL, RED, CON.



📖 Grattarola, D., et al. "Understanding pooling in graph neural networks.", 2022.

Global pooling

- Combines only node features
- Does not account for topology

Global pooling

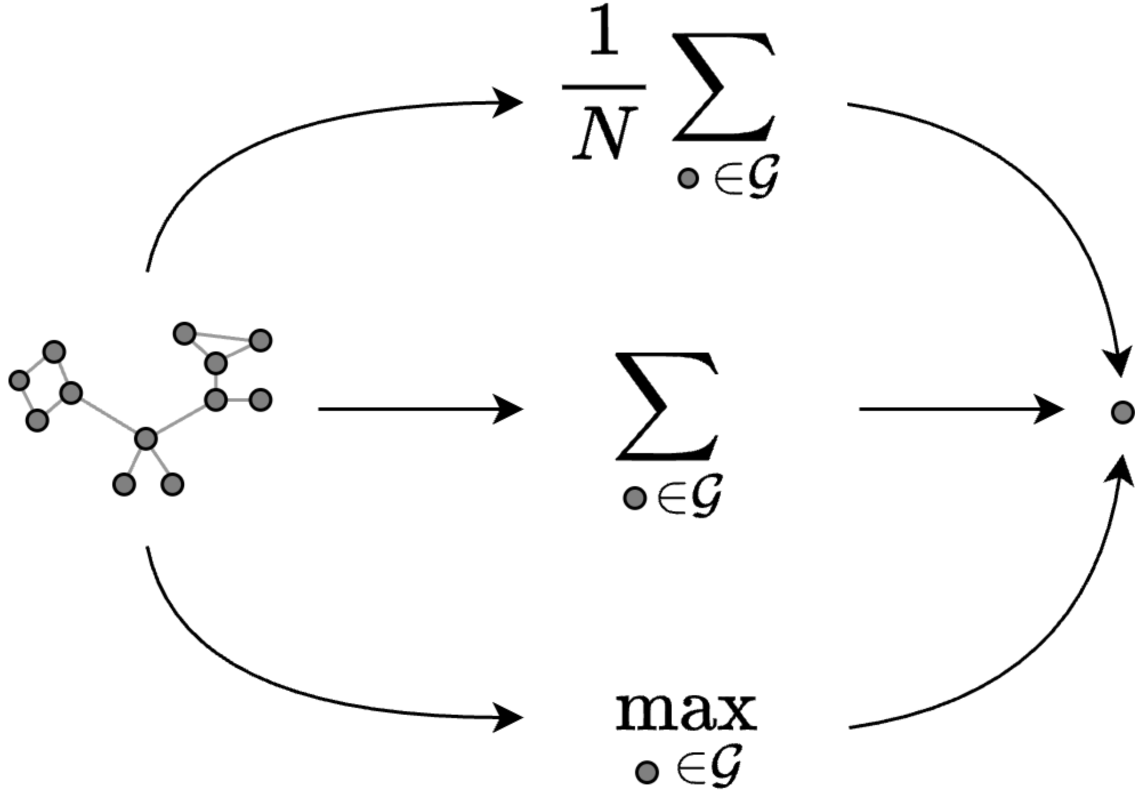
- Combines only node features
- Does not account for topology

$$\boldsymbol{s} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

Global pooling

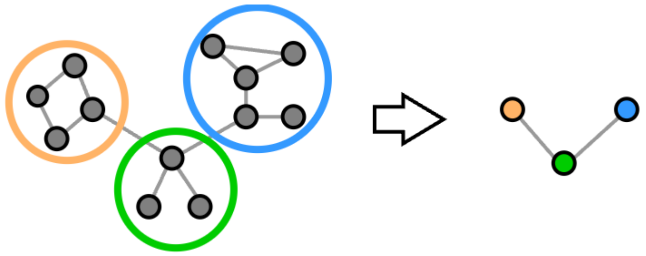
- Combines only node features
- Does not account for topology

$$s = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

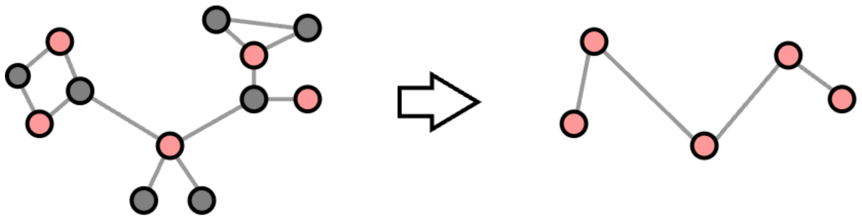


Hierarchical graph pooling families

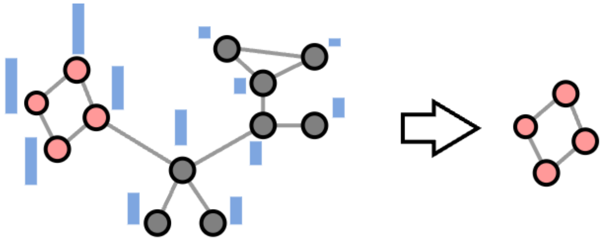
- Soft clustering



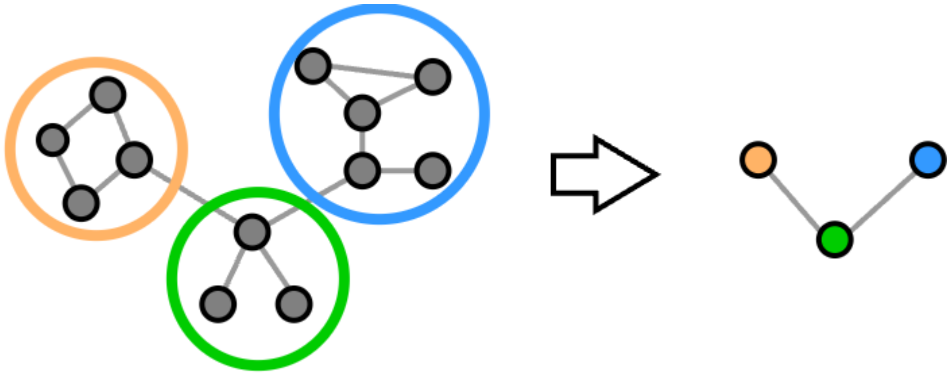
- 1-over-k



- Scoring-based

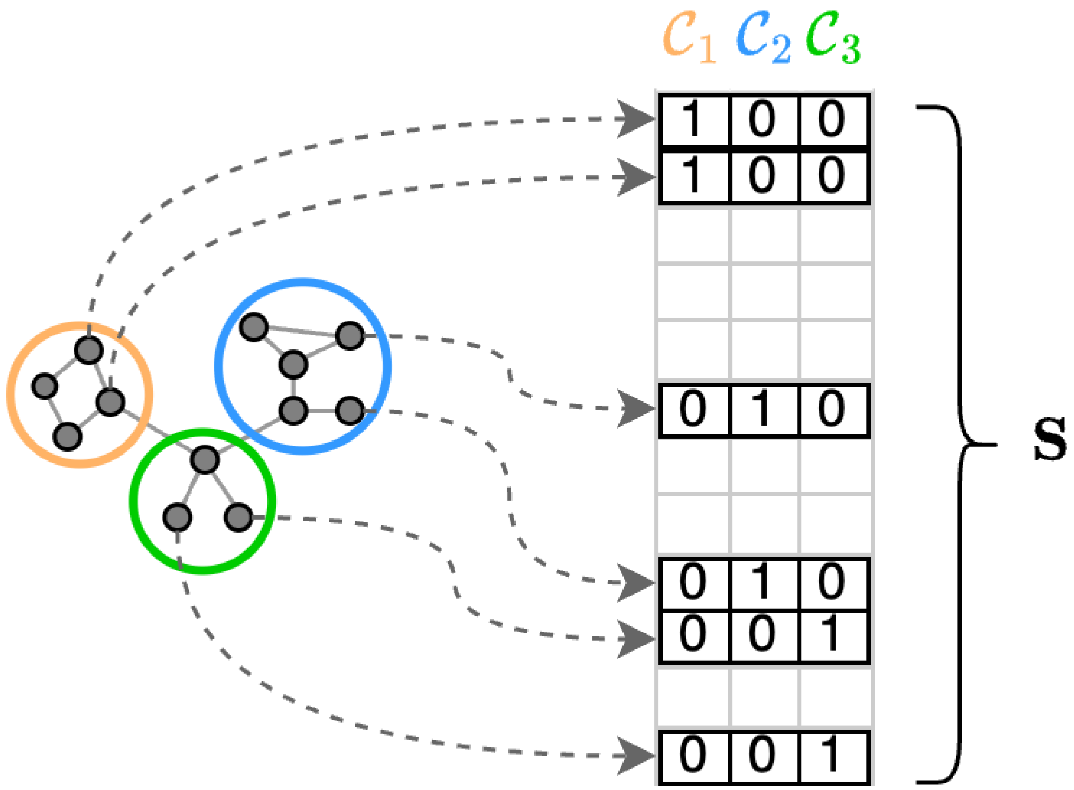


Soft Clustering

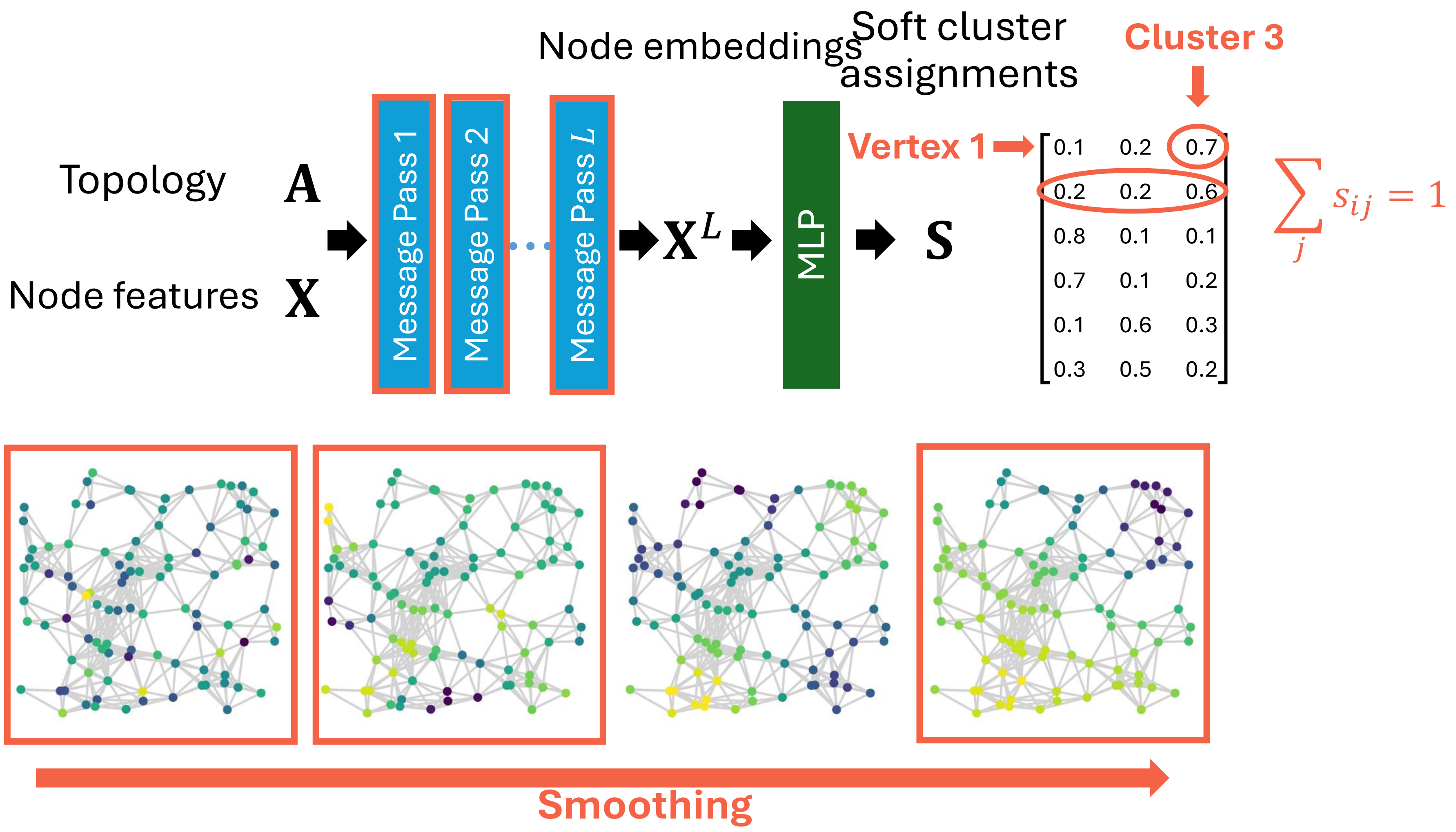


Cluster-based

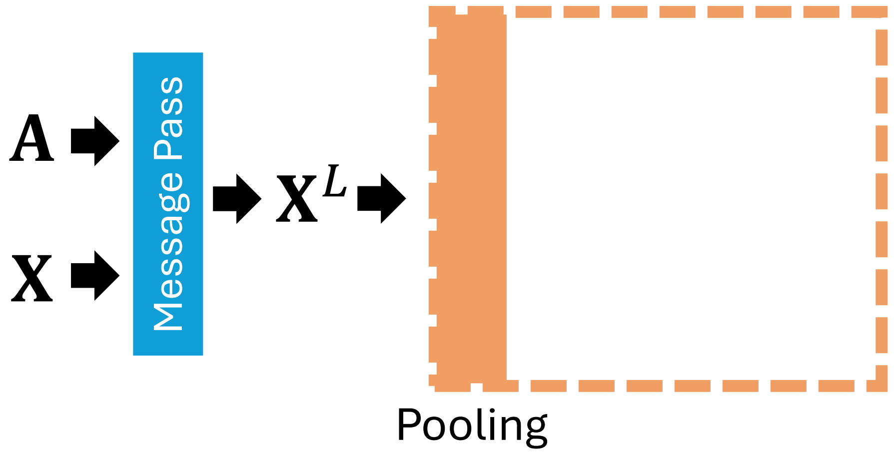
- Aggregates nodes that have **similar features** and are **strongly connected**
- The clustering partition is induced by an assignment matrix $\boldsymbol{S}$.



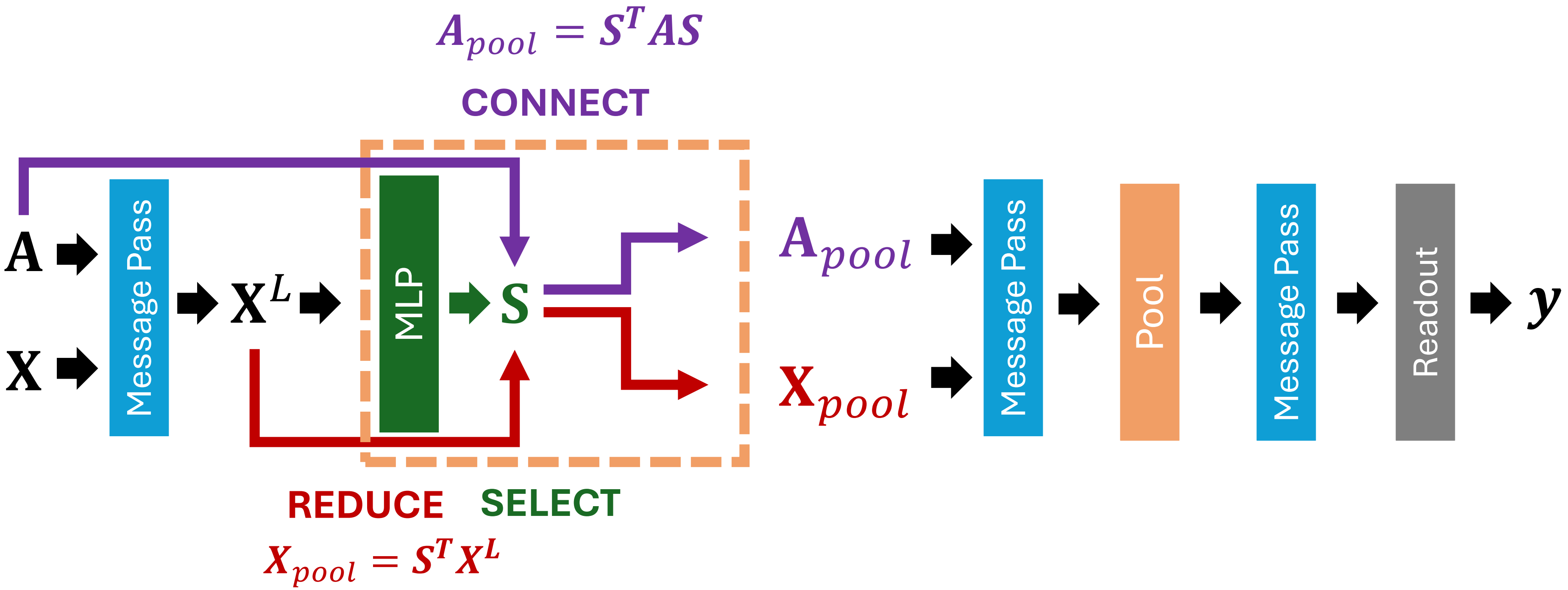
General framework



Workflow

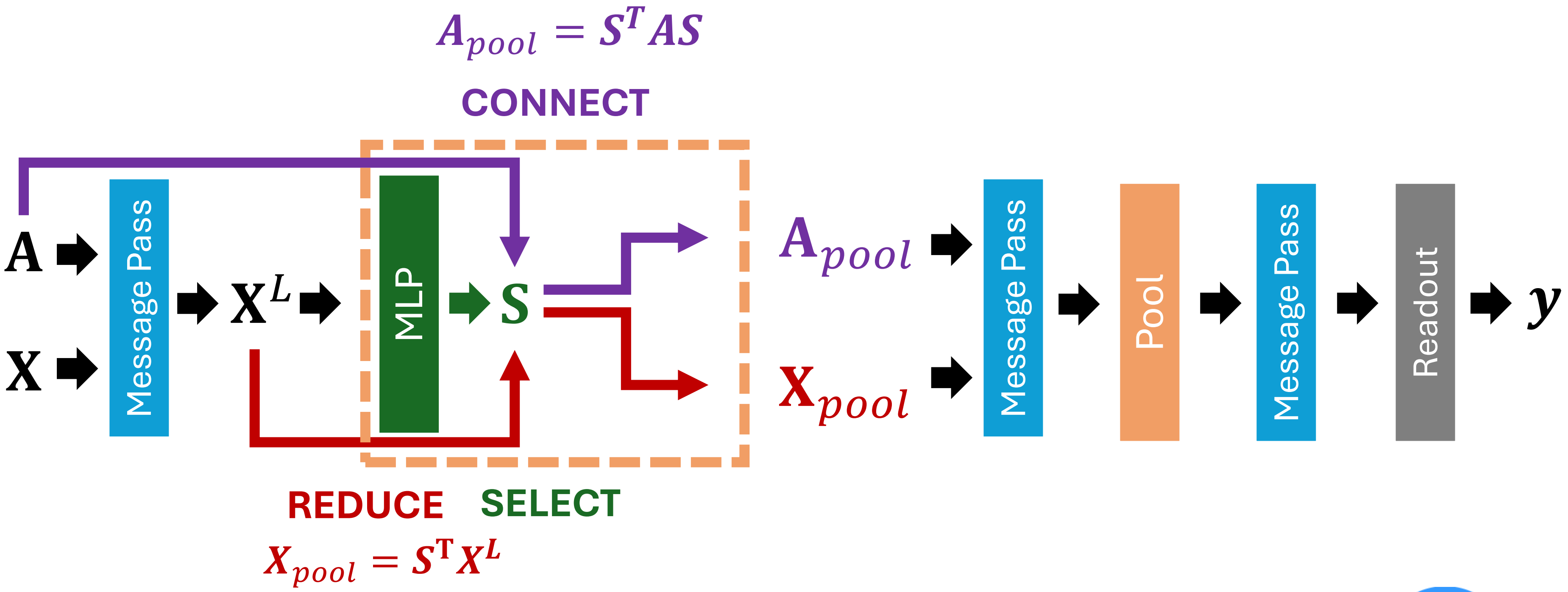


Workflow

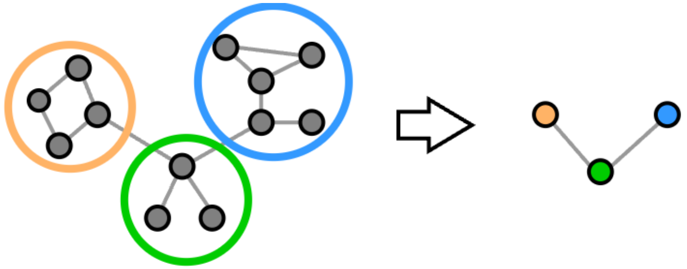


$A_{pool} = S^T AS$
CONNECT
A
X
Message Pass
X^L
MLP
S
A_{pool}
X_{pool}
Message Pass
Pool
Message Pass
Readout
y
REDUCE
SELECT
$X_{pool} = S^T X^L$

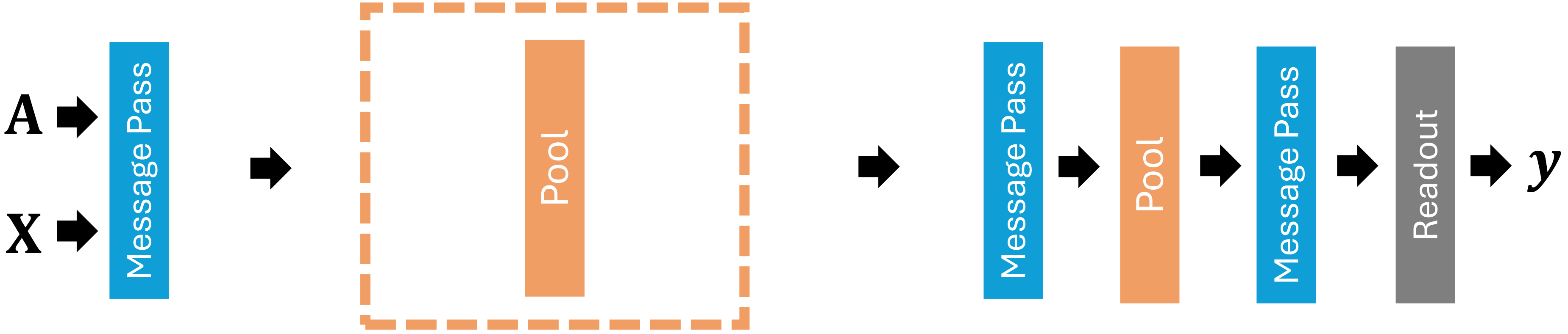
Workflow



$A_{pool} = S^T AS$
CONNECT
A
X
Message Pass
X^L
MLP
S
A_{pool}
X_{pool}
Message Pass
Pool
Message Pass
Readout
y
REDUCE
SELECT
$X_{pool} = S^T X^L$

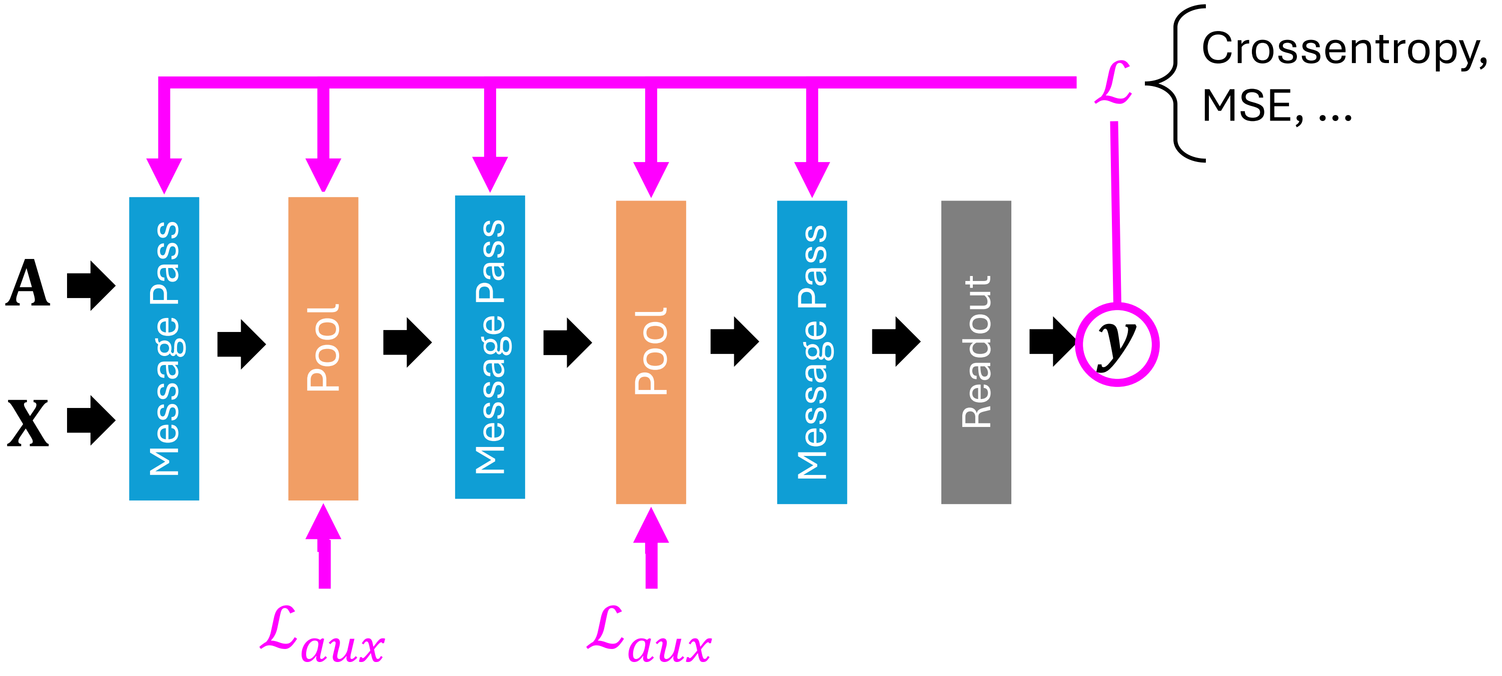


Training



A
X
Message Pass
Pool
Message Pass
Pool
Message Pass
Readout
y

Training



Degenerate solutions
- All nodes in one cluster
- Uniform membership

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0.3 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.3 \end{bmatrix}$$

Auxiliary losses

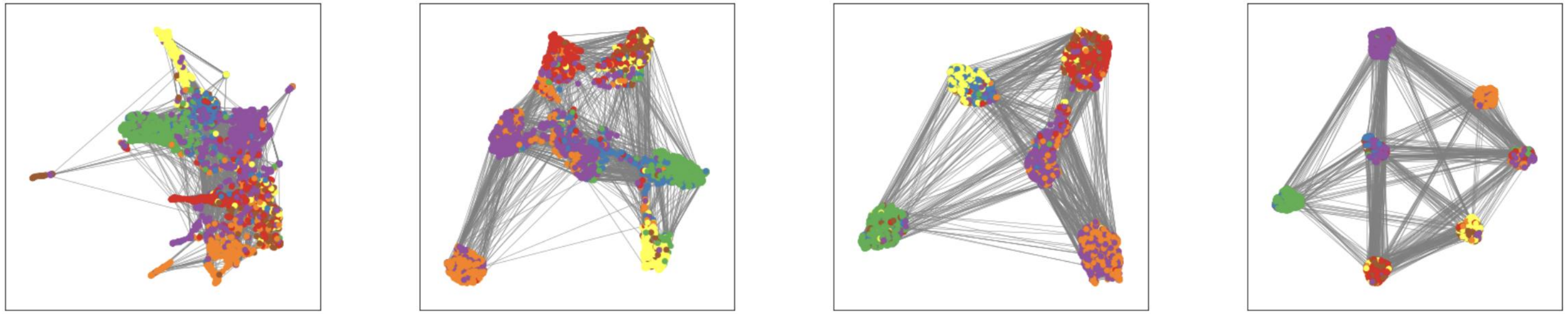
	Cluster assignments ↕ Topology	Well-formed clusters (no degenerate solutions)	Reference
Diffpool	Link reconstruction: $\left\lVert \mathbf{A} - \mathtt{softmax}(\mathbf{S})\mathtt{softmax}(\mathbf{S})^\top \right\rVert_F$	Entropy: $\frac{1}{N}\sum_{n=1}^{N} H(\mathbf{S})$	📖 Ying, Z., et al. "Hierarchical graph representation learning with differentiable pooling", 2018.
Mincut	mincut: $-\frac{\mathrm{Tr}(\mathbf{S}^\top\mathbf{A}\mathbf{S})}{\mathrm{Tr}(\mathbf{S}^\top\mathbf{D}\mathbf{S})}$	Orthogonality: $\left\lVert \frac{\mathbf{S}^\top\mathbf{S}}{\lVert\mathbf{S}^\top\mathbf{S}\rVert_F} - \frac{\mathbf{I}_C}{\sqrt{C}} \right\rVert_F$	📖 Bianchi, F. M. et al. “Spectral clustering with Graph Neural Networks for Graph Pooling”, 2020.
DMoN	Modularity: $-\frac{1}{2M}\mathrm{Tr}(\mathbf{S}^\top\mathbf{B}\mathbf{S})$	Collapse: $\frac{\sqrt{C}}{N}\left\lVert \sum_i \mathbf{C}_i^\top \right\rVert_F - 1$	📖 Tsitsulin, A. et al. “Graph clustering with graph neural networks”, 2023.
TVGNN	Asym Cheeger Cut: $\frac{1}{2M}\sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{j=i}^{N} a_{i,j}\lvert s_{i,k} - s_{j,k}\rvert$	Balance: $\sum_{k=1}^{K} \lvert\lvert \boldsymbol{s}_{:,k} - \mathrm{quant}_\rho(\boldsymbol{s}_{:,k})\rvert\rvert_{1,\rho}$	📖 Hansen, J. B. et al. “Total Variation Graph Neural Networks”, 2023.

Cluster assignments
↕
Topology

Well-formed clusters
(no degenerate solutions)

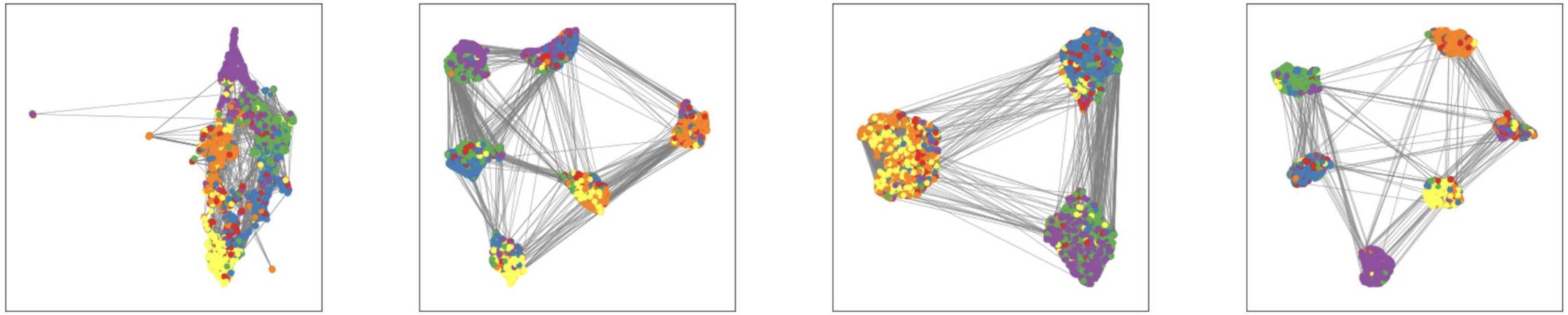
Comparison

Cora



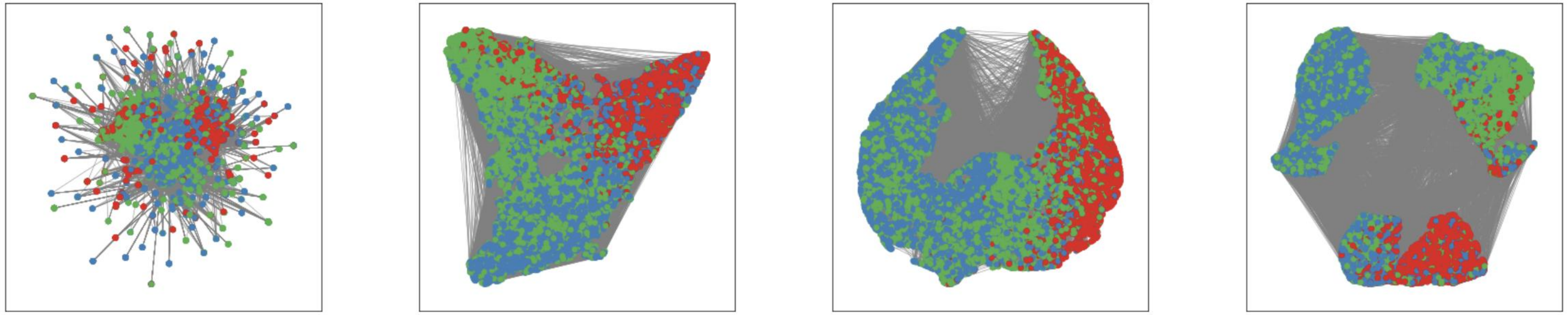
(e) DiffPool (f) MinCutPool (g) DMoN (h) TVGNN

Citeseer



(e) DiffPool (f) MinCutPool (g) DMoN (h) TVGNN

Pubmed



(e) DiffPool (f) MinCutPool (g) DMoN (h) TVGNN

Pros and cons

✅ Flexible, retains all graph information

✅ Good performance in downstream tasks

❌ High space complexity

❌ Fixed size in pooled graphs

Pros and cons

✅ Flexible, retains all graph information

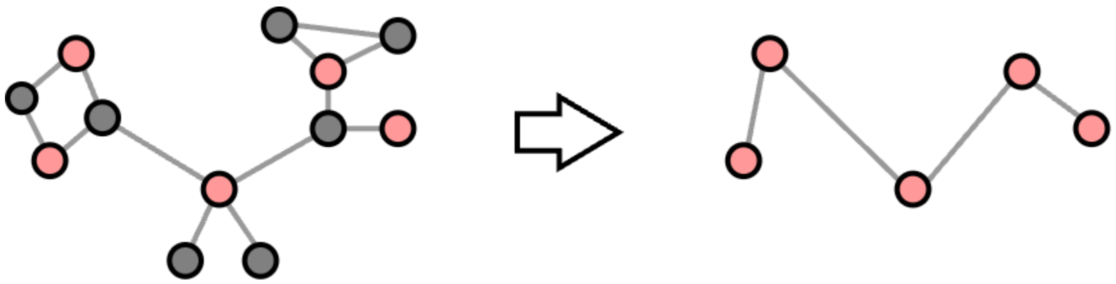
✅ Good performance in downstream tasks

❌ High space complexity

❌ Fixed size in pooled graphs

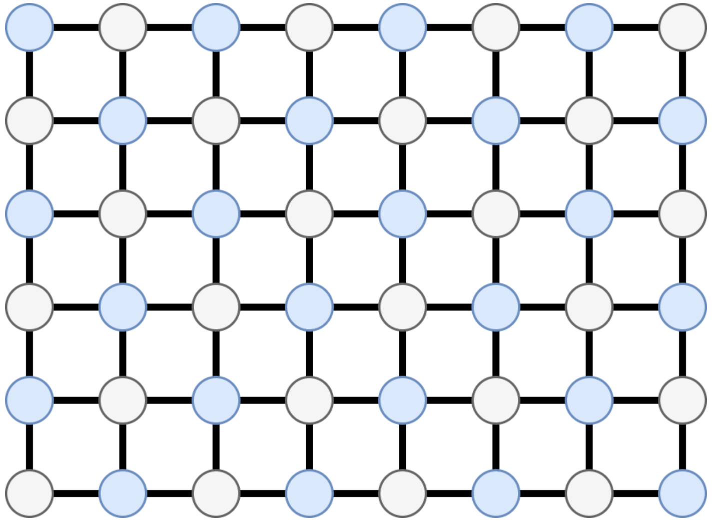
😎 BN-Pool: a Bayesian Nonparametric Approach for Graph Pooling
Open talk by **Daniele Castellana**

1-over-k



1-over-k

- Extends 1-over-k sampling from regular structures to graphs



Regular grid

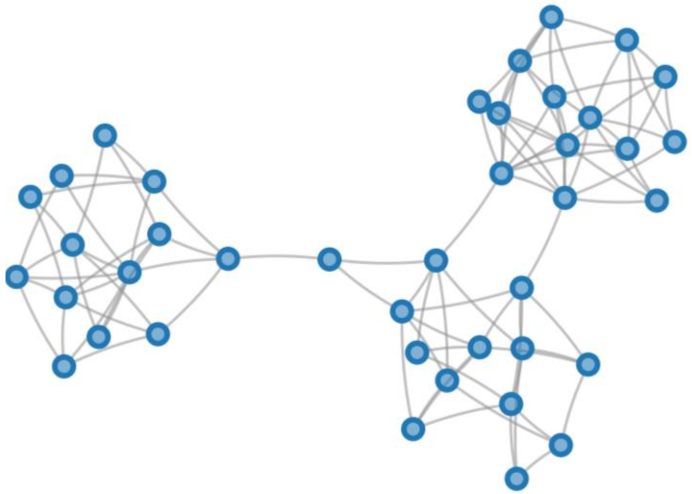
1-over-2

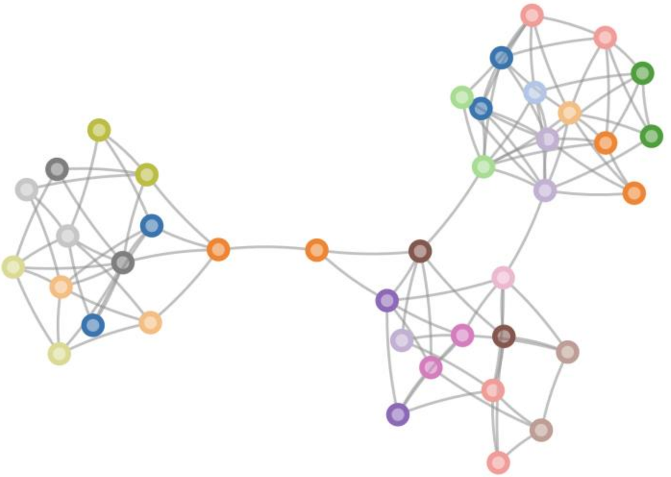
Sequence

1-over-4

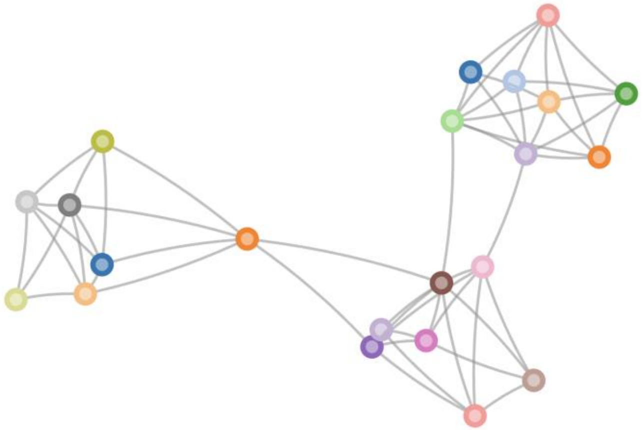
Graclus

No feats are considered!





SELECT

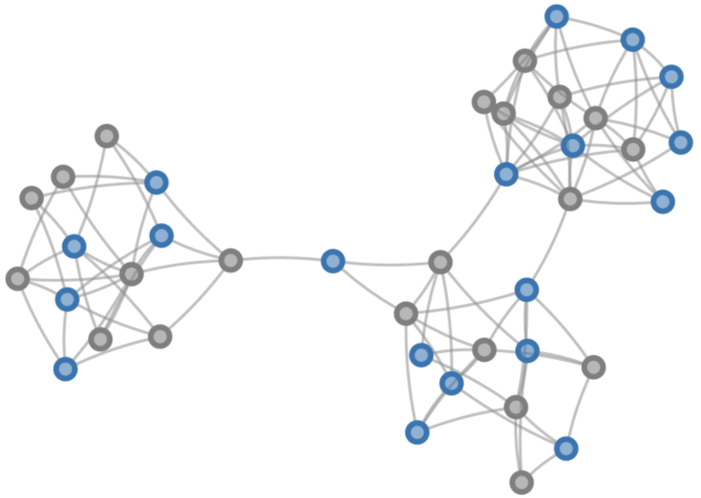


REDUCE & **CONNECT**

📖 Dhillon I. S., et al. "Weighted graph cuts without eigenvectors a multilevel approach", 2018.

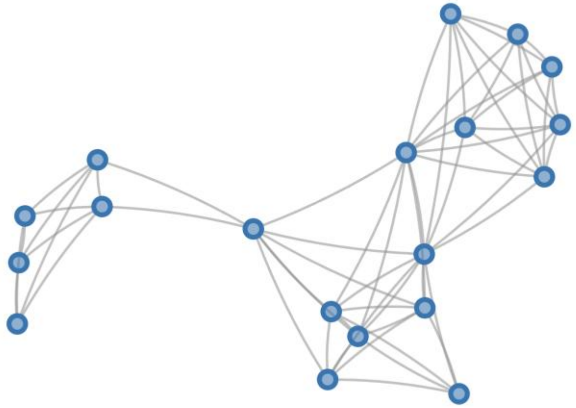
NDP – Node Decimation Pooling

Solution to MAXCUT and node coloring



SELECT

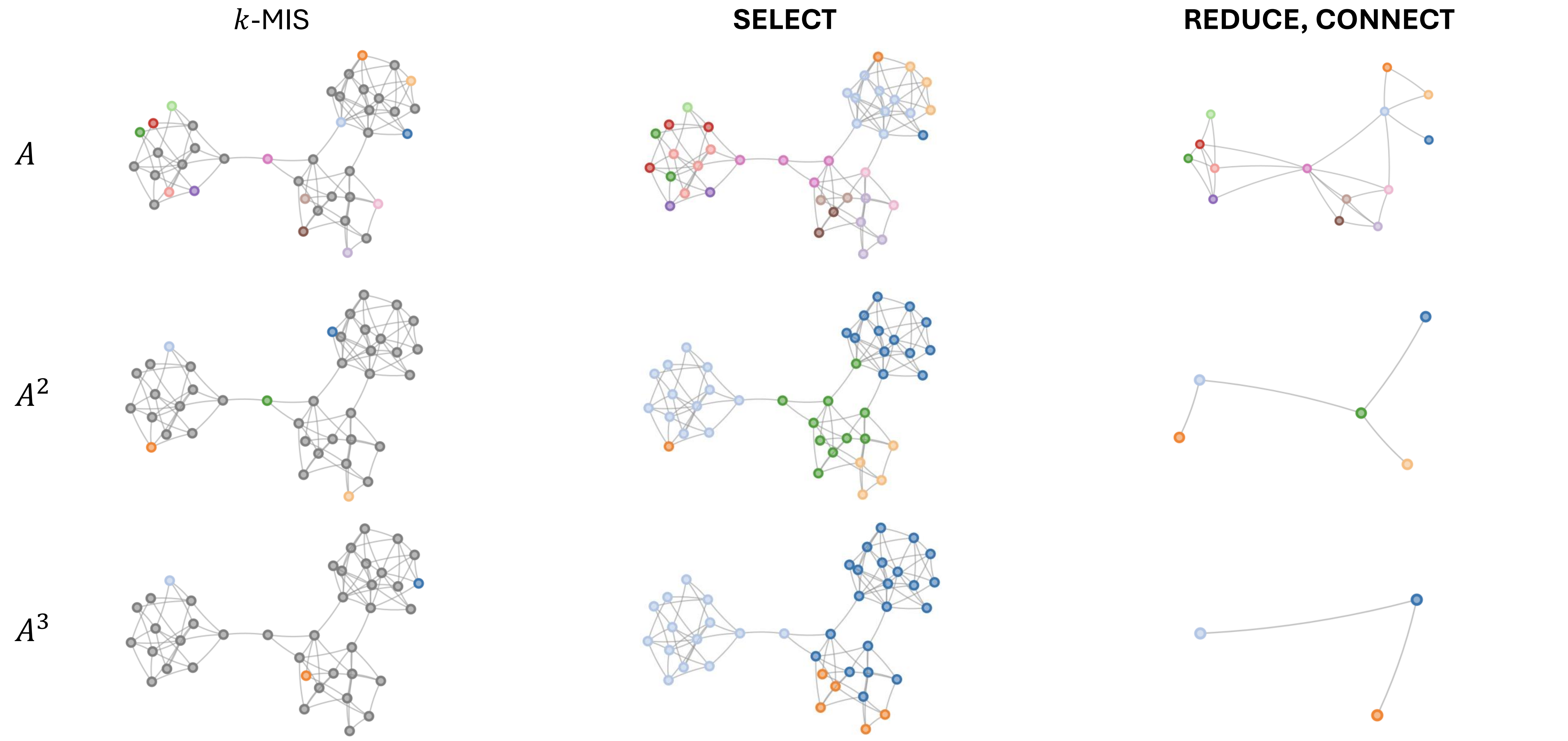
Kron reduction



REDUCE & CONNECT

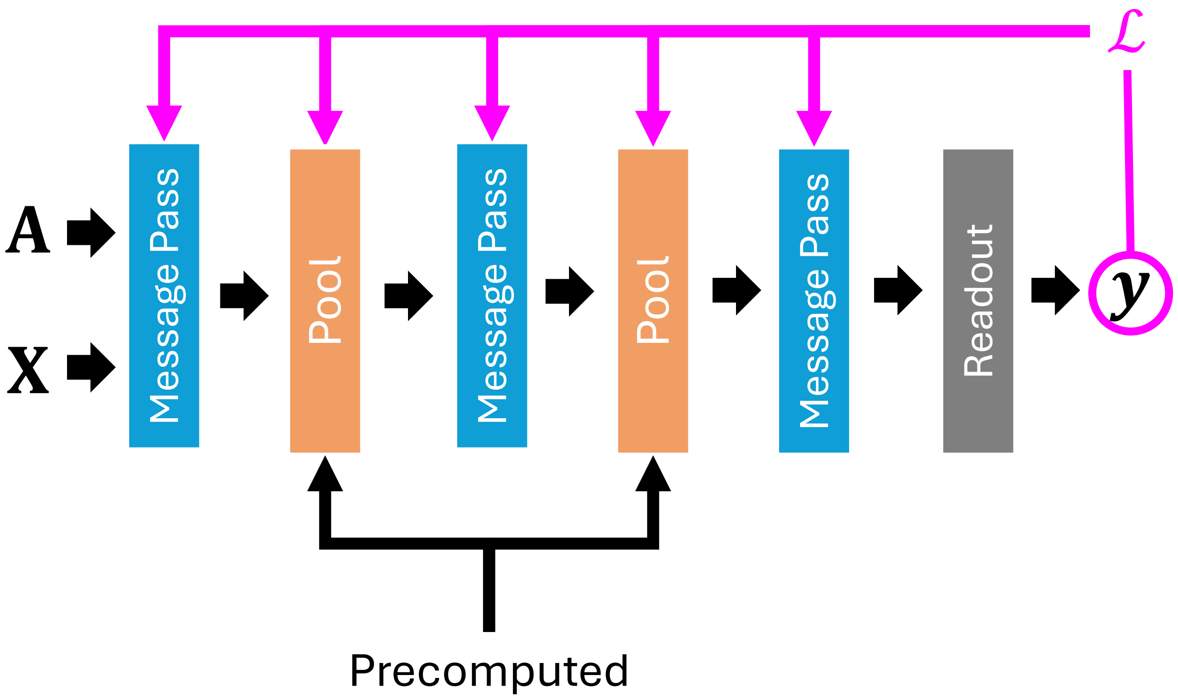
📖 Bianchi F. M., et al. "Hierarchical representation learning in graph neural networks with node decimation pooling", 2020.

k-Maximal Independent Sets (MIS)



📖 Bacciu D., et al, “Generalizing Downsampling from Regular Data to Graphs”, 2022.

Training



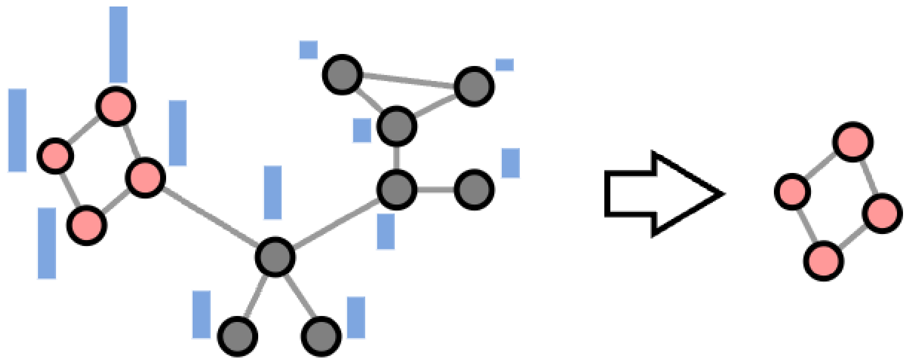
$\mathcal{L}$
A
X
Message Pass
Pool
Message Pass
Pool
Message Pass
Readout
y
Precomputed

Pros and cons

- ✅ Precomputed, very fast training.
- ✅ Lack of trainable parameters are good for small datasets.

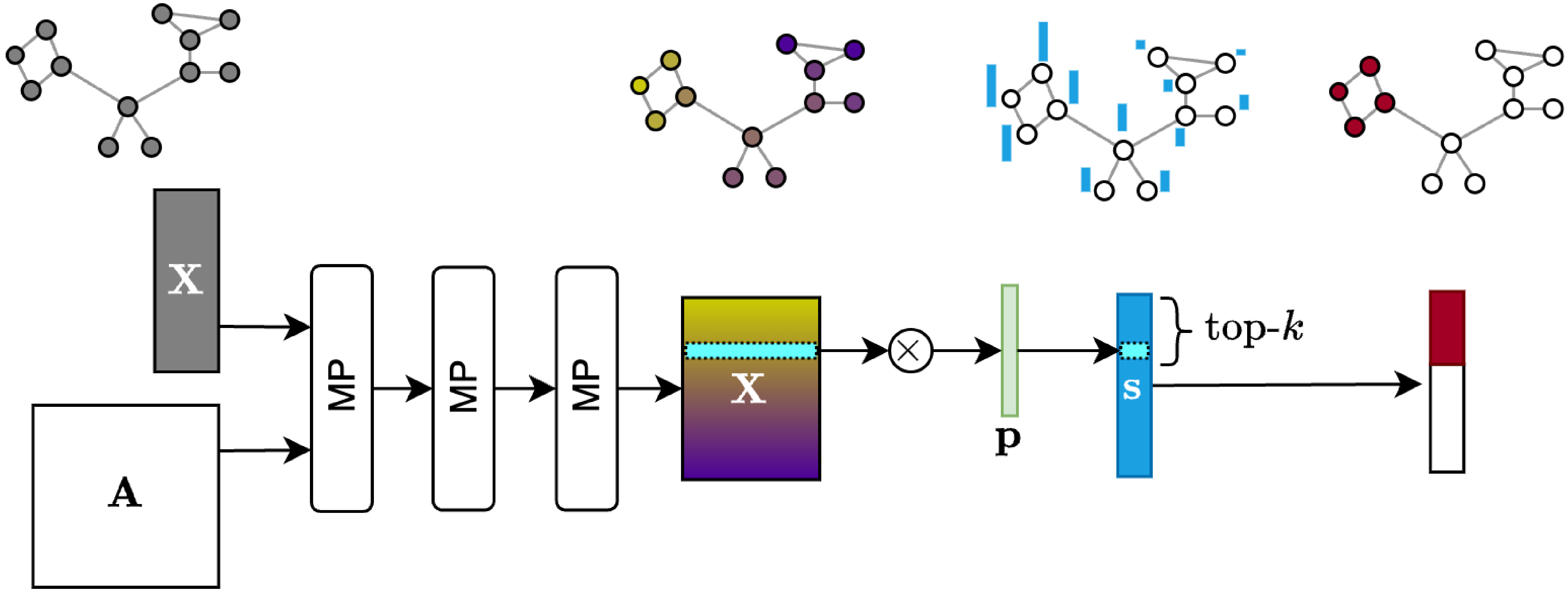
- ❌ Lack of flexibility
- ❌ Does not account for node features and downstream task

Scoring-based



Select

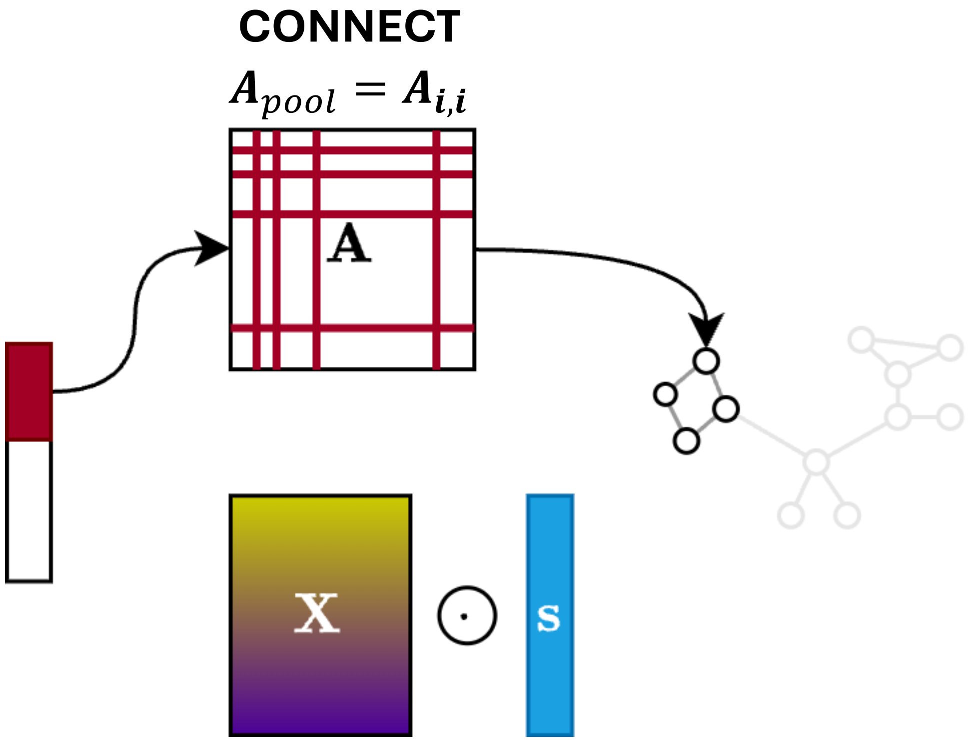
- Assigns a score to each node and keep nodes with highest score



📖 Gao H. & Jin S., "Graph U-Nets", 2019.

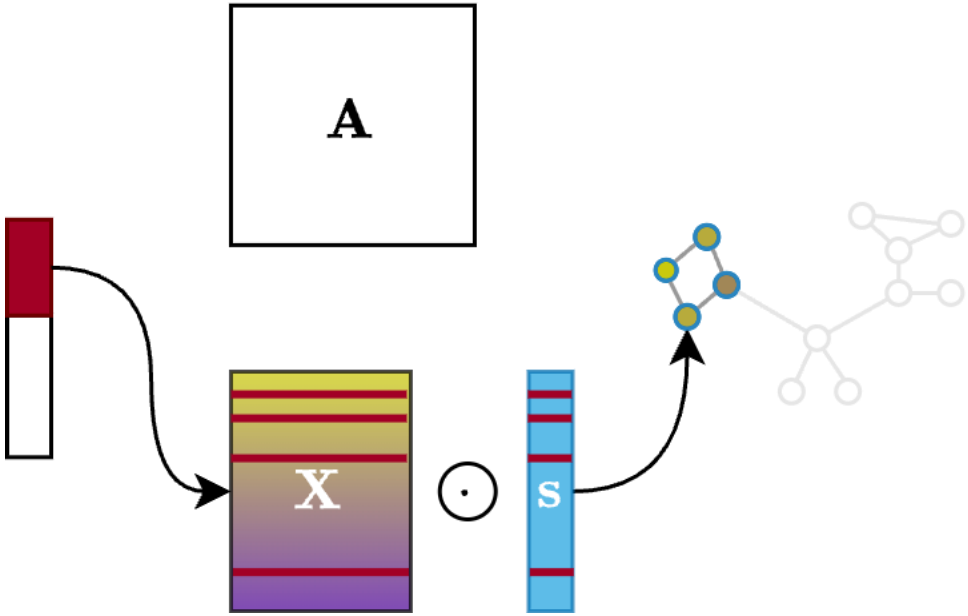
Connect, Reduce

$$\boldsymbol{i} = \mathrm{top}k(\boldsymbol{s})$$



Connect, Reduce

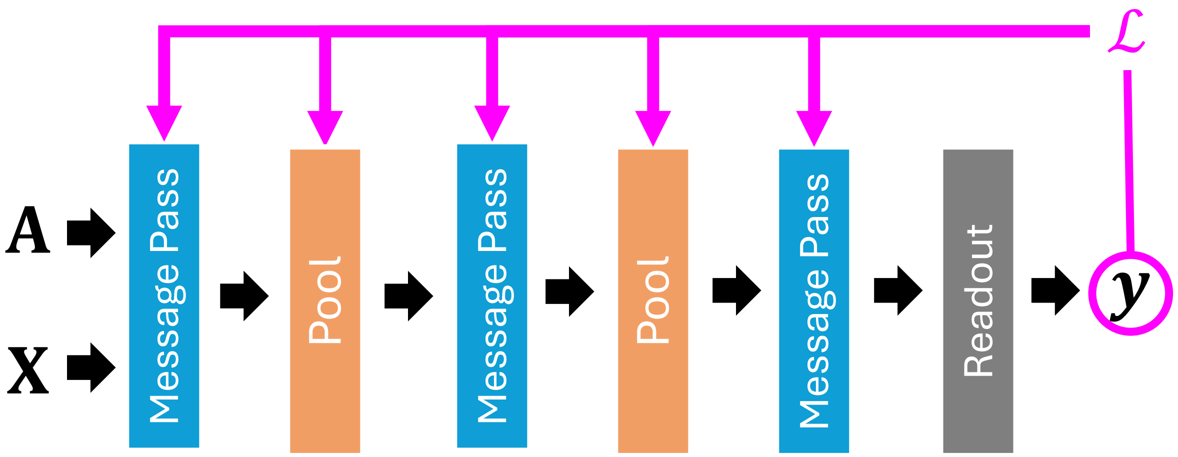
$\boldsymbol{i} = \mathrm{top}k(\boldsymbol{s})$



REDUCE

$$\boldsymbol{X}_{pool} = \boldsymbol{X}_{\boldsymbol{i}} \odot \boldsymbol{s}$$

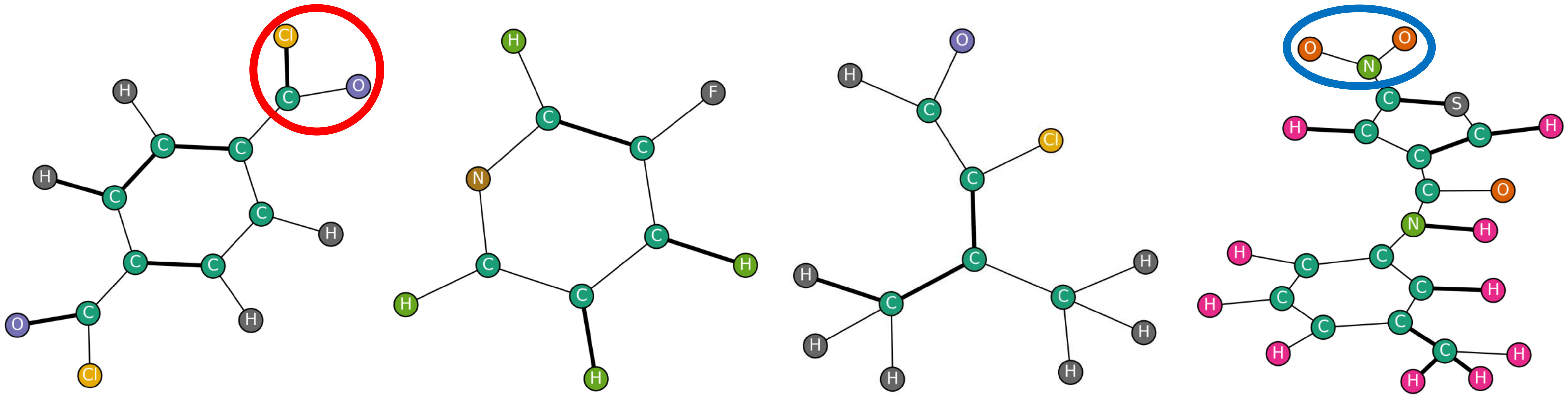
Training



$\mathcal{L}$
A
X
Message Pass
Pool
Message Pass
Pool
Message Pass
Readout
y

When does it work?

- Different scoring-based methods compute the scores differently.
- However, pooled graph are chunks of connected nodes.
- Works e.g., when the class is given by a specific substructure.



1 0 0 2

Pros and cons

- ✅ Few parameters, low computational complexity

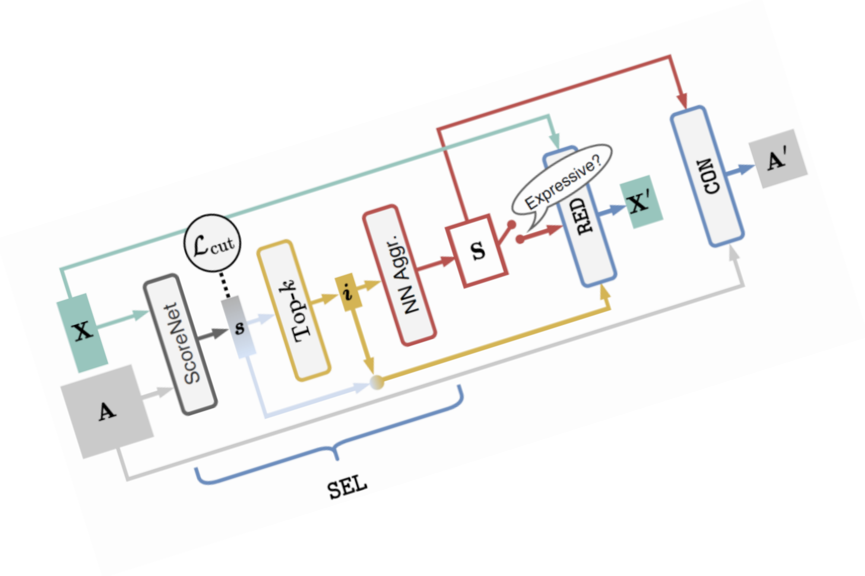
- ❌ Completely discard some graph information
- ❌ Fails in tasks where preserving the graph structure matters

Pros and cons

- ✅ Few parameters, low computational complexity

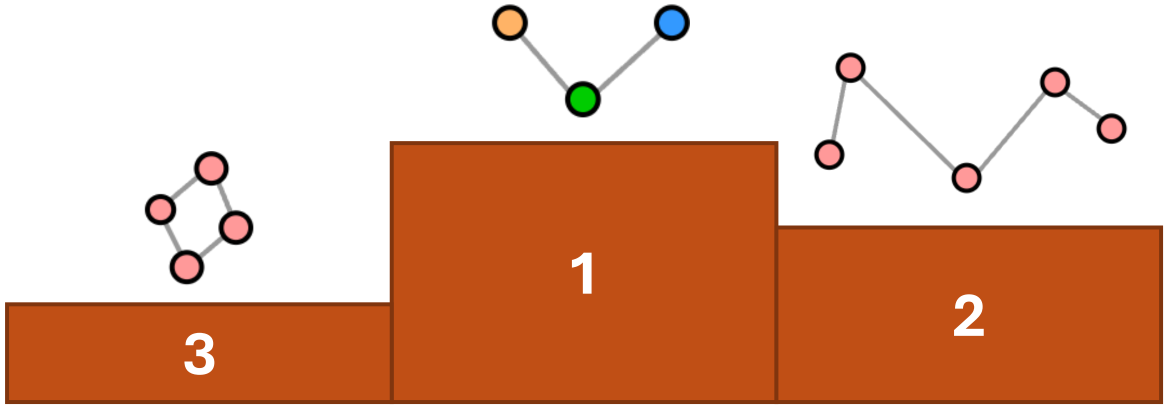
- ❌ Completely discard some graph information
- ❌ Fails in tasks where preserving the graph structure matters

😎 MaxCutPool



come to the poster! 😎

Evaluation Procedures



Evaluation procedures 1/4

🤔 How to measure the performance of a pooling method?

1. Performance of GNN with pooling layers on downstream tasks (graph classification, regression, etc…).

- ⚠️ Empirical and indirect evaluation.
- ⚠️ Difficult to separate the effect of pooling from other GNN components.

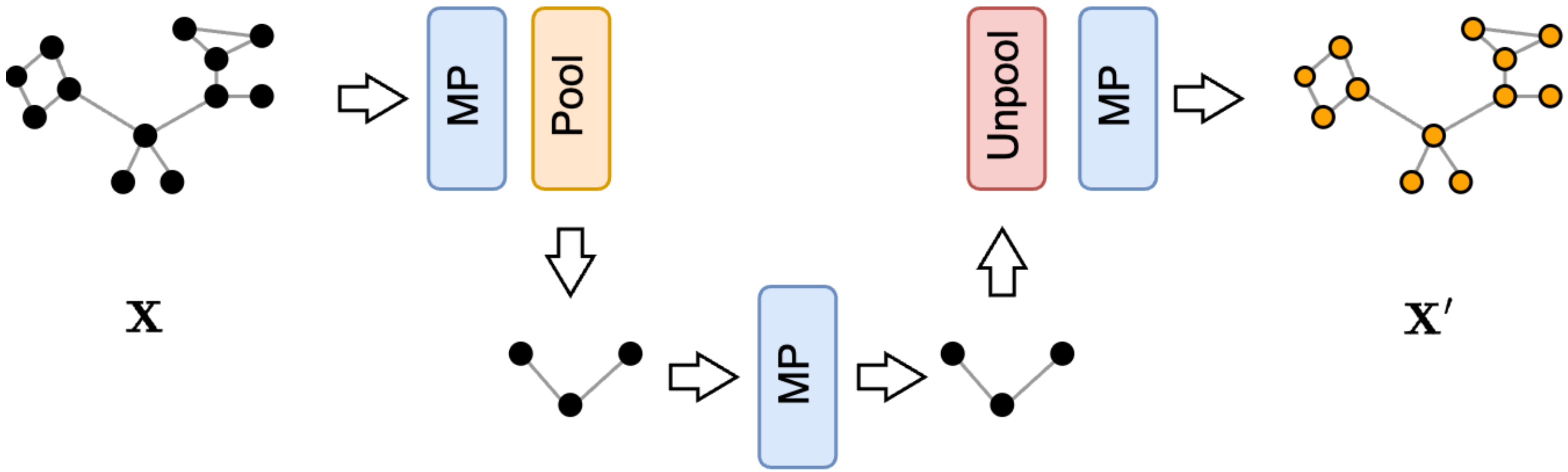
🏆 Soft clustering methods usually perform better

📖 Wang P., et al. “A Comprehensive Graph Pooling Benchmark: Effectiveness, Robustness and Generalizability”, 2024.

Evaluation procedures 2/4

2. Preserving node information

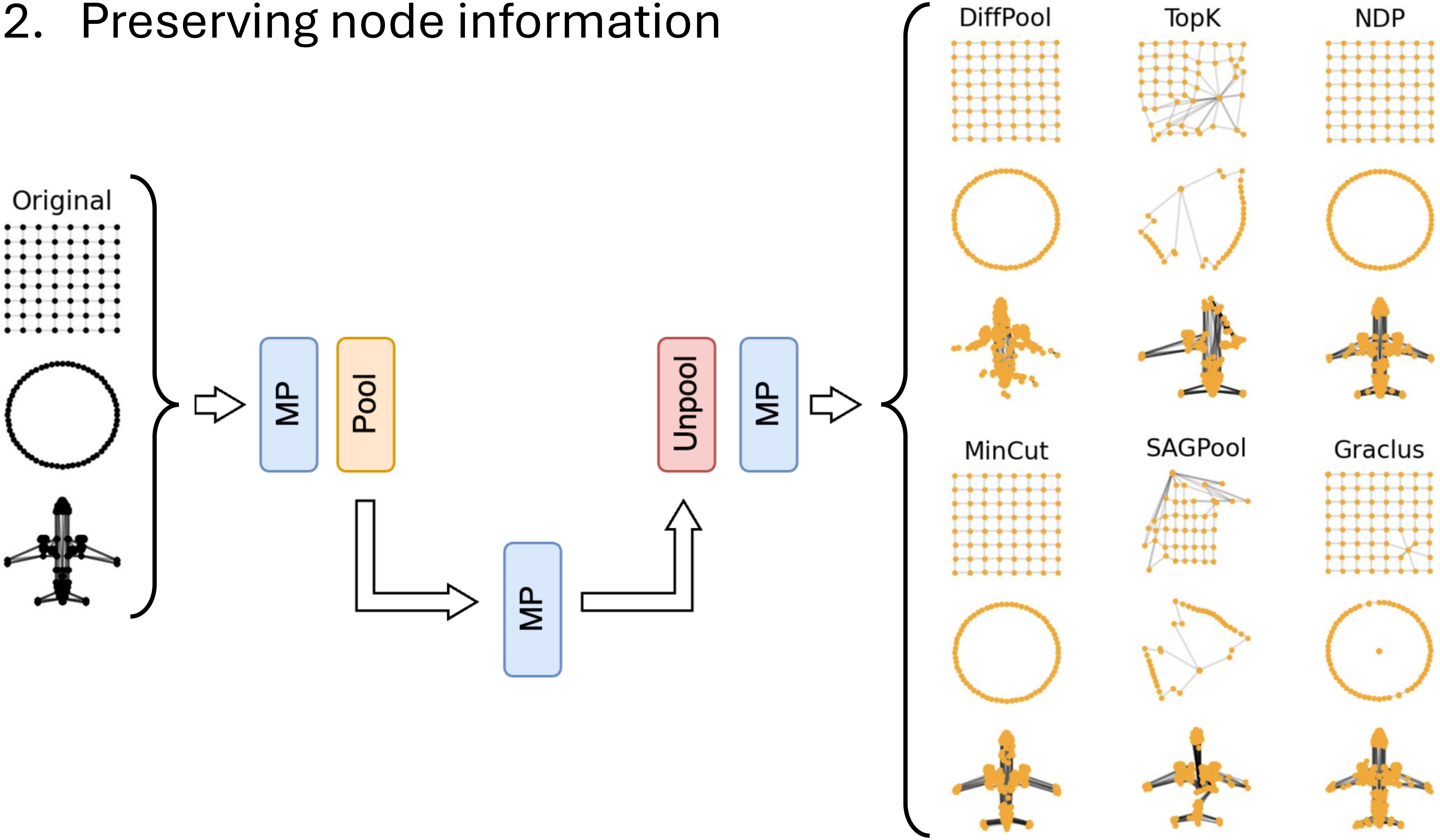
$$\mathcal{L}_{\mathrm{rec}} = \|\mathbf{X} - \mathbf{X}'\|^2$$



📖 Grattarola, D., et al. "Understanding pooling in graph neural networks.", 2022.

Evaluation procedures 2/4

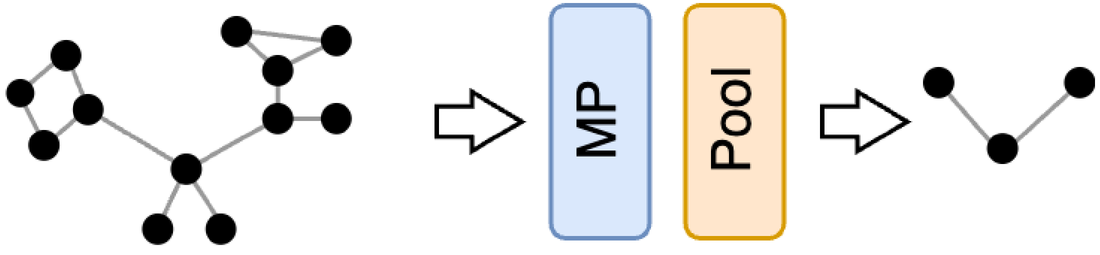
2. Preserving node information



Evaluation procedures 3/4

3. Preserving topology

$$\mathcal{L}_{\text{struct}} = \sum_{i=0}^{D} \|\mathbf{X}_{:,i}^\top \mathbf{L} \mathbf{X}_{:,i} - \mathbf{X}_{:,i}'^\top \mathbf{L}' \mathbf{X}_{:,i}'\|^2$$

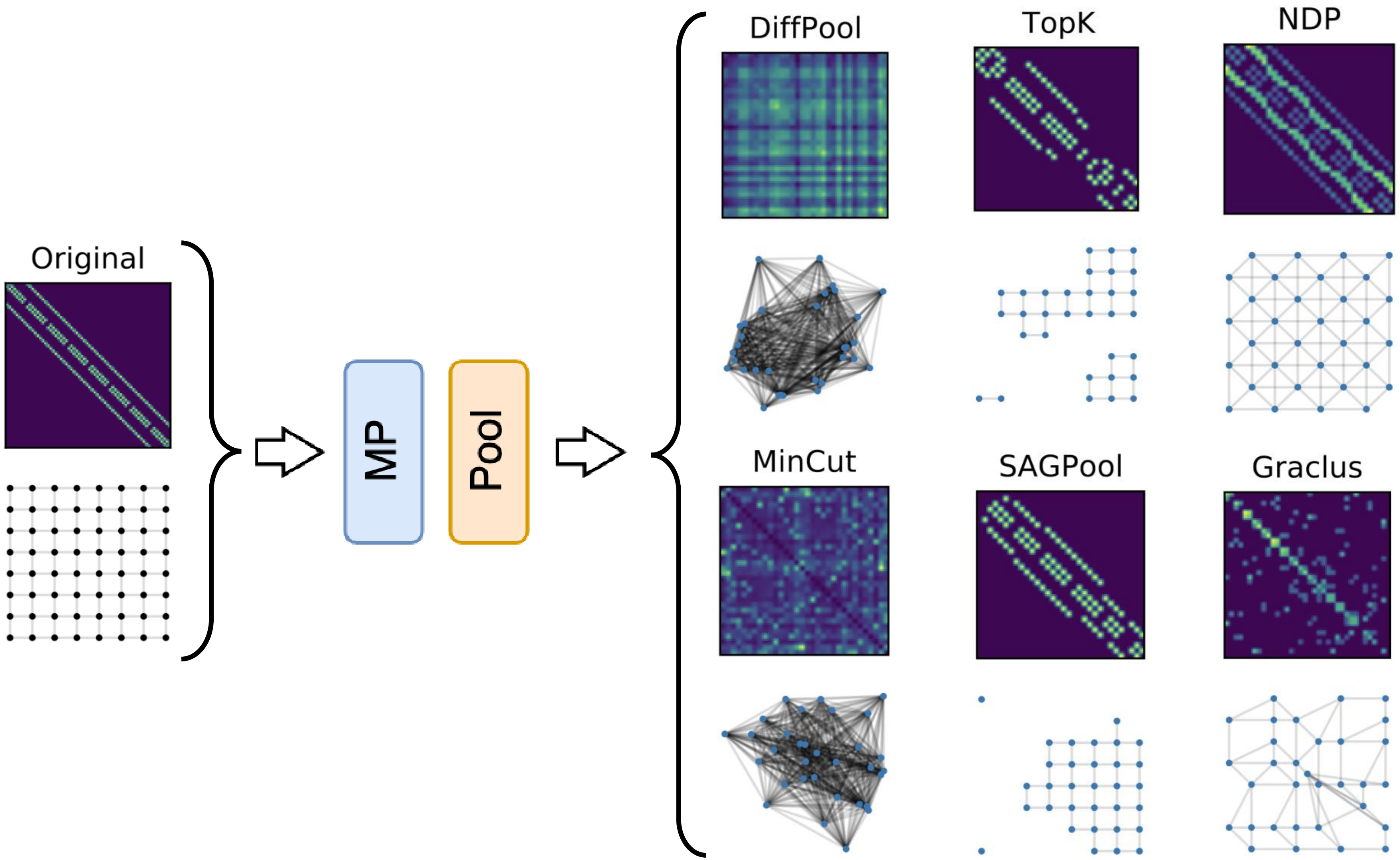


$\mathbf{L}$ $\mathbf{L}'$

📖 Grattarola, D., et al. "Understanding pooling in graph neural networks.", 2022.

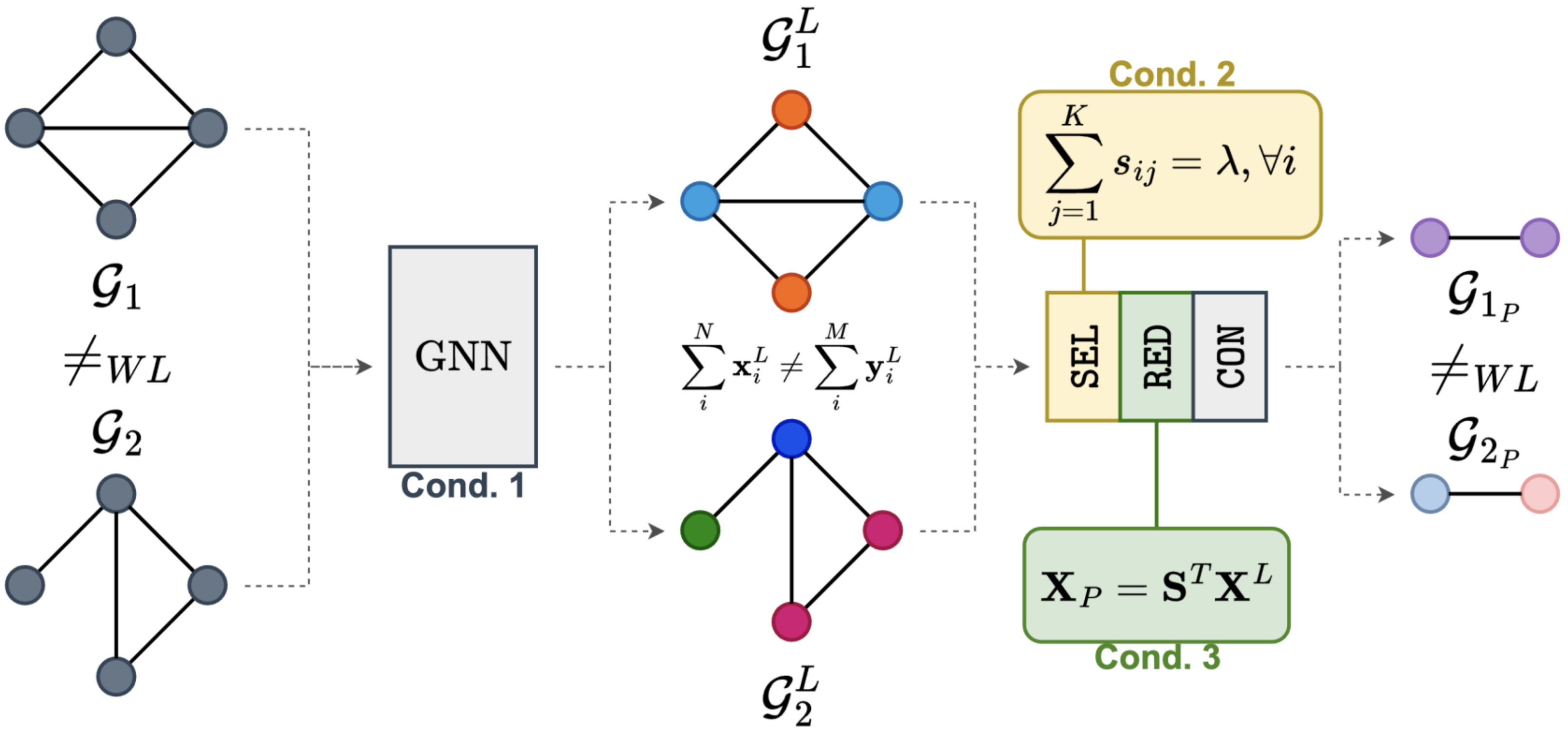
Evaluation procedures 3/4

3. Preserving topology



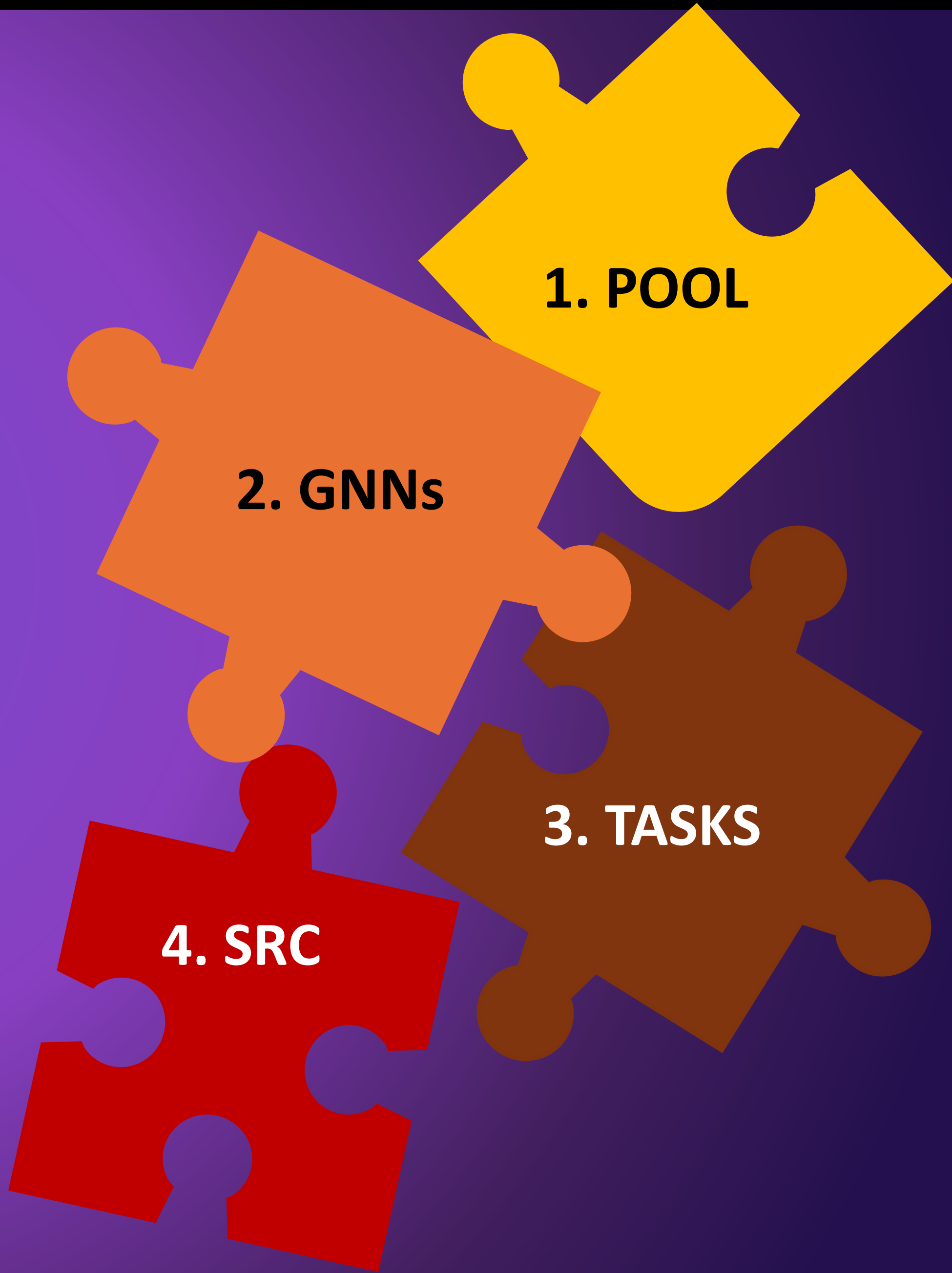
Evaluation procedures 4/4

4. Expressiveness

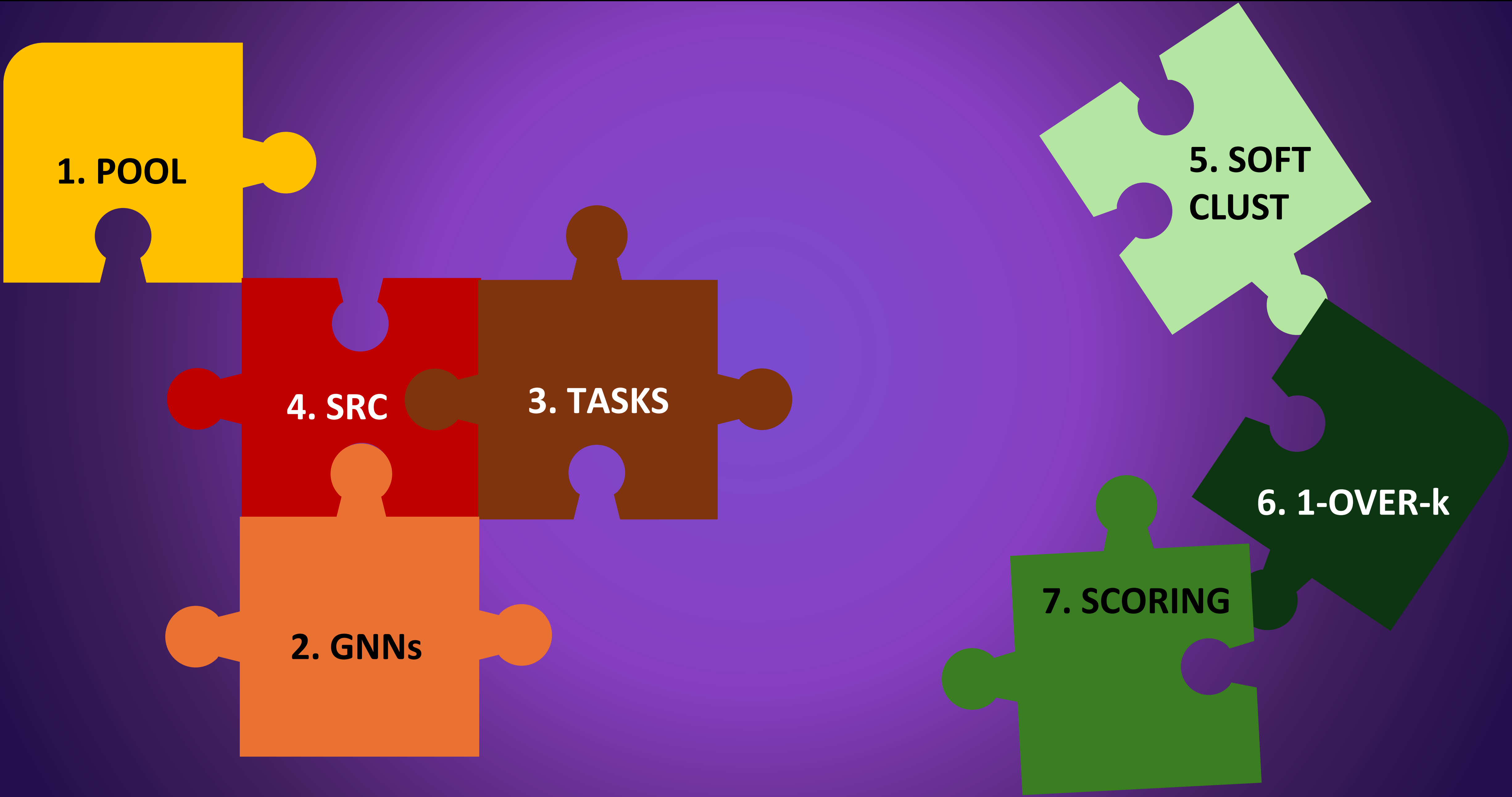


📖 Bianchi F. M. & Lachi V., “The expressive power of pooling in Graph Neural Networks”, 2023.

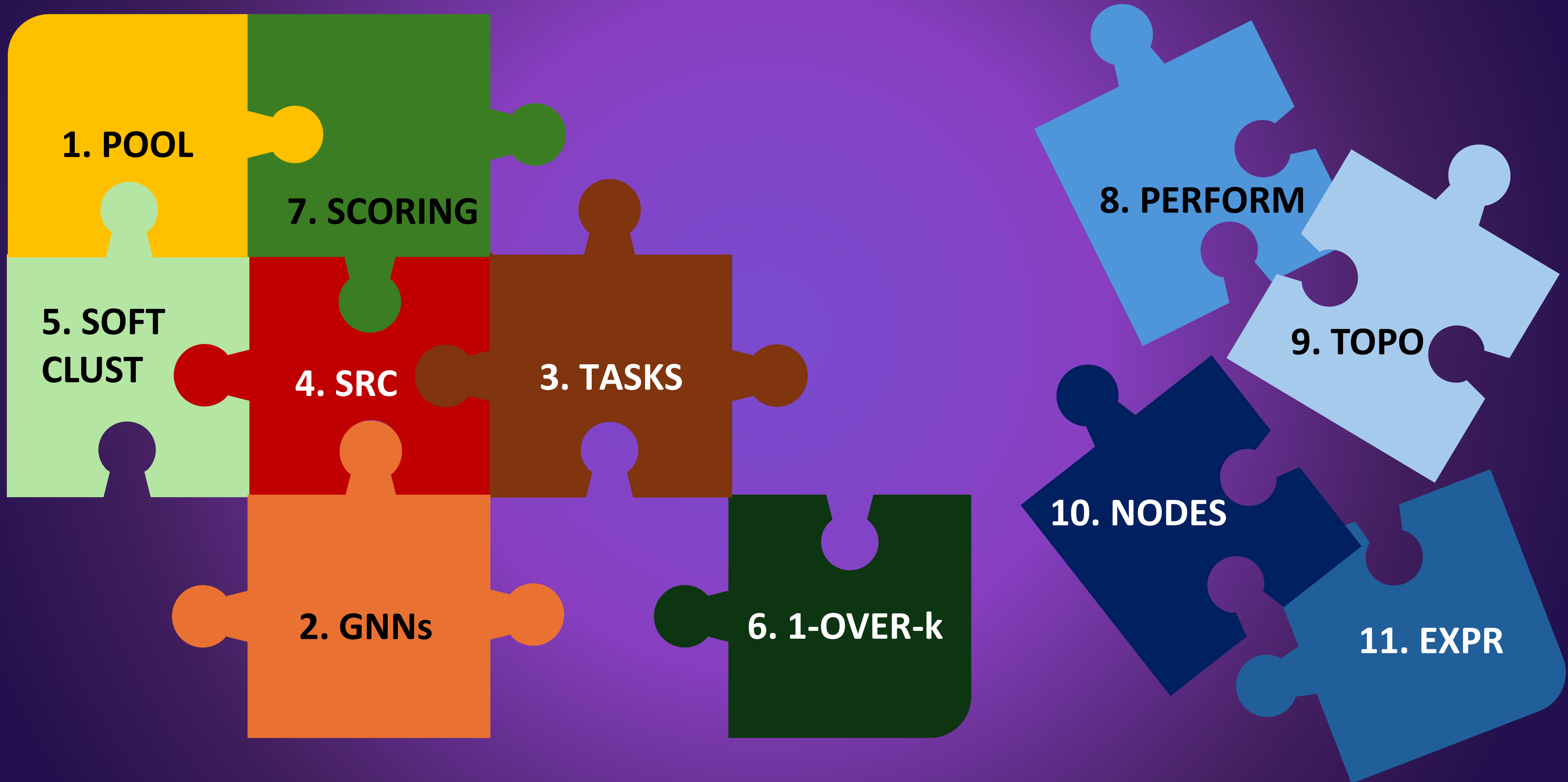
Summary



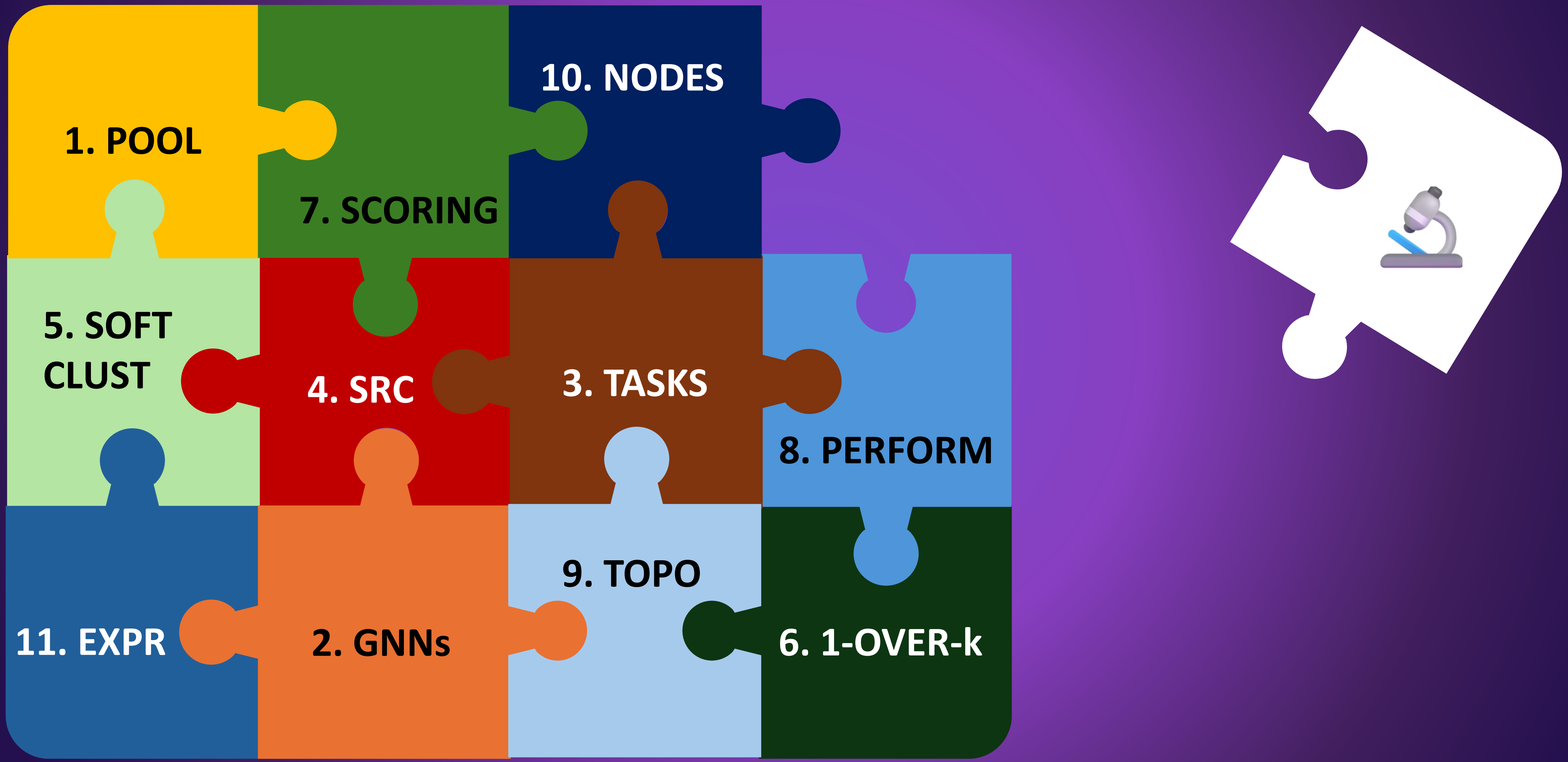
Summary



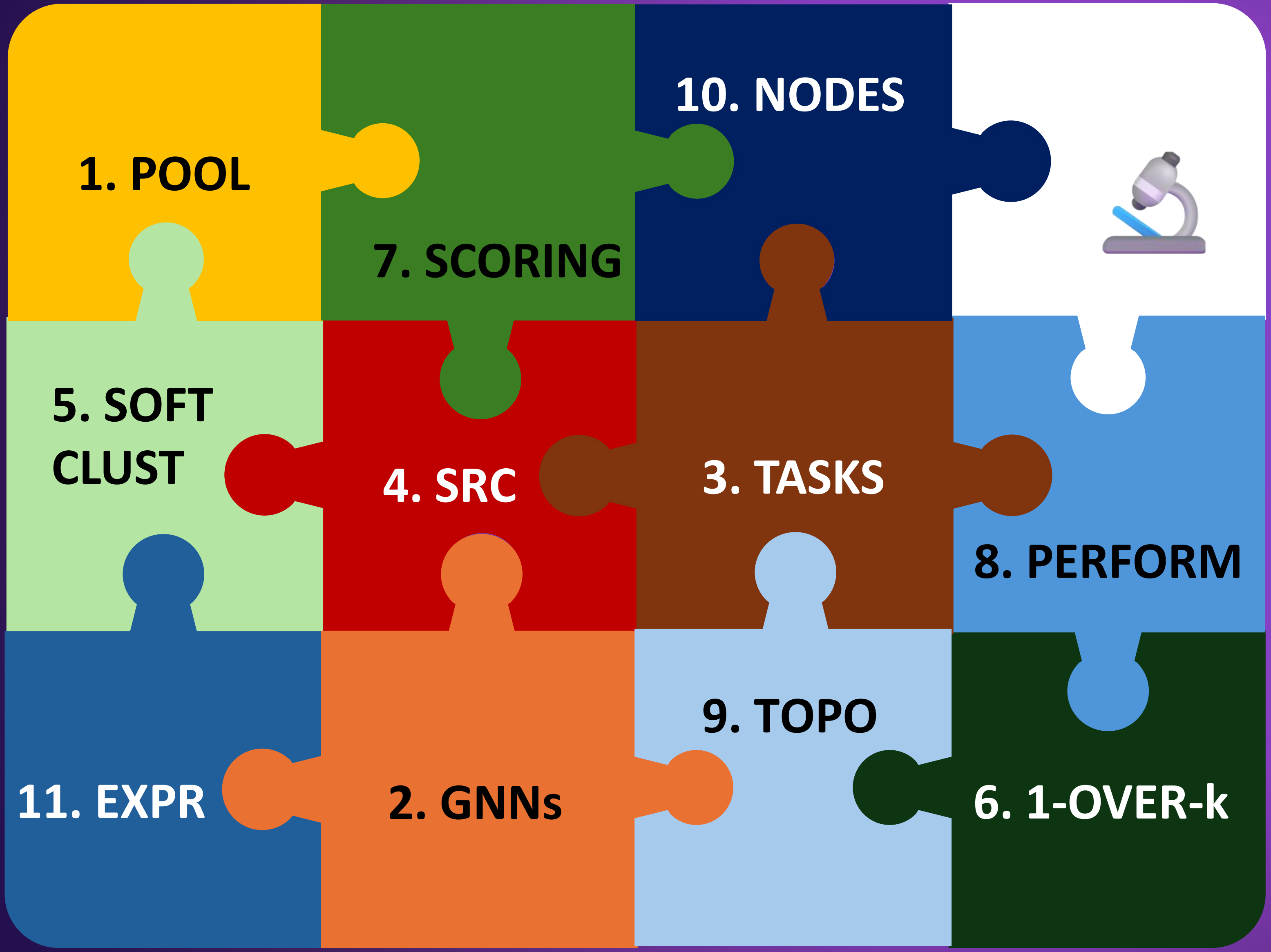
Summary


1. POOL
7. SCORING
5. SOFT CLUST
4. SRC
3. TASKS
2. GNNs
6. 1-OVER-k
8. PERFORM
9. TOPO
10. NODES
11. EXPR

Summary


1. POOL
7. SCORING
10. NODES
5. SOFT CLUST
4. SRC
3. TASKS
8. PERFORM
11. EXPR
2. GNNs
9. TOPO
6. 1-OVER-k

Summary



1. POOL
7. SCORING
10. NODES
5. SOFT CLUST
4. SRC
3. TASKS
8. PERFORM
11. EXPR
2. GNNs
9. TOPO
6. 1-OVER-k

The End

Blog

https://gnn-pooling.notion.site/

Northernmost Graph Machine Learning Group

https://ngmlgroup.github.io/