Total Variation Graph Neural Networks

Jonas Berg Hansen and Filippo Maria Bianchi

Balanced K-cut Problem

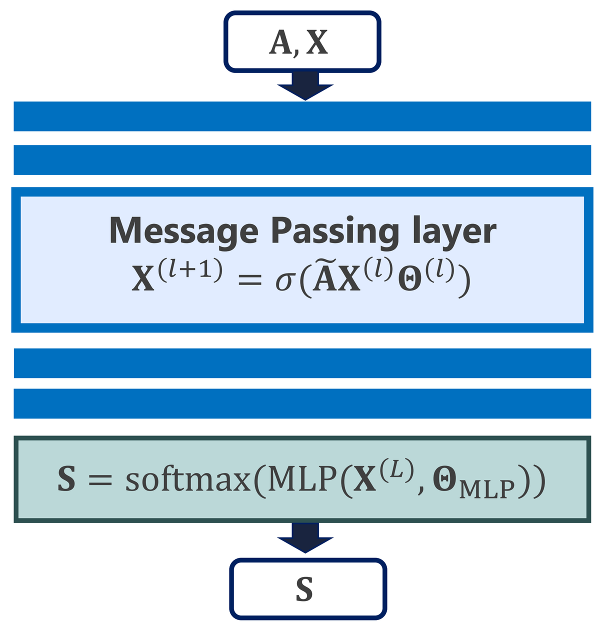
- The optimal cluster assignment is the solution of the following (combinatorial) problem:

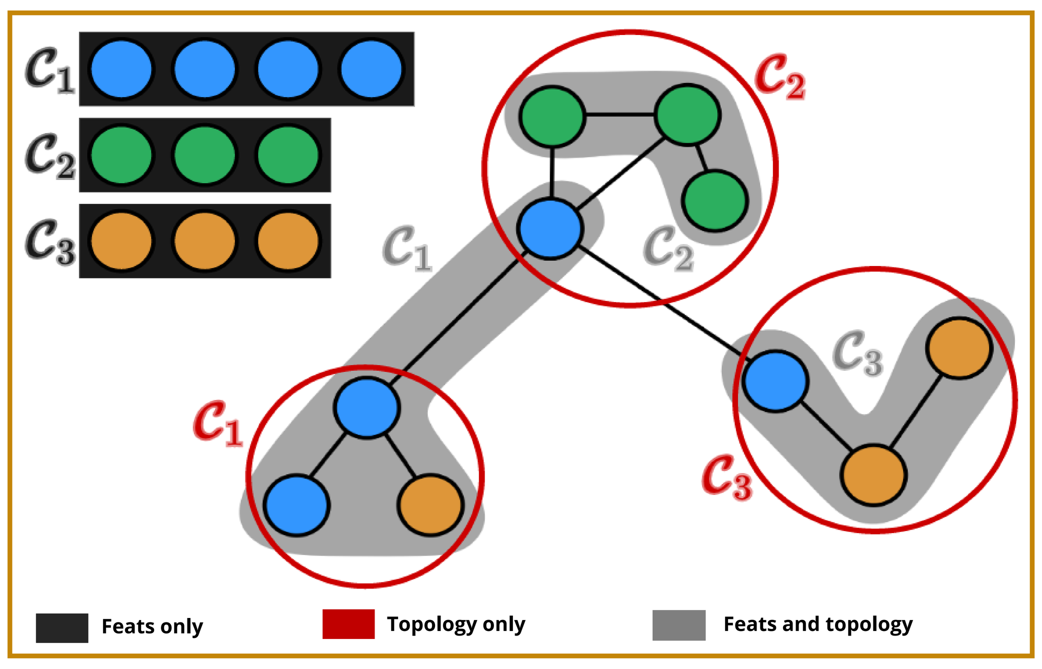
$$\min_{\mathcal{C}_1,\dots,\mathcal{C}_K} \sum_{k=1}^{K} \frac{cut(\mathcal{C}_k, \bar{\mathcal{C}}_k)}{\hat{S}(\mathcal{C}_k)}$$

- The numerator minimizes the volume of edges between clusters
- The denominator ensures balanced partitions and avoids degenerate solutions
- Spectral clustering adopts a continuous relaxation that replaces the numerator with a local quadratic variation (**LQV**) term: $\|\boldsymbol{s}\|_{LQV} = \sum_{(i,j)\in\mathcal{E}} a_{ij}(s_i - s_j)^2$
- Tighter relaxations adopt a graph total variation (**GTV**) term: $\|\boldsymbol{s}\|_{GTV} = \sum_{(i,j)\in\mathcal{E}} a_{ij}|s_i - s_j|$, but do not admit closed form solutions [1]

Notation
$\mathbf{X} \in R^{N\times F}$: node features
$\mathbf{A} \in R^{N\times N}$: adjacency matrix
$\mathbf{D} \in R^{N\times N}$: degree matrix
$\mathbf{S} \in R^{N\times K}$: cluster assignments

Clustering with GNNs

1 

 2

Example: MinCutPool[2]

Loss

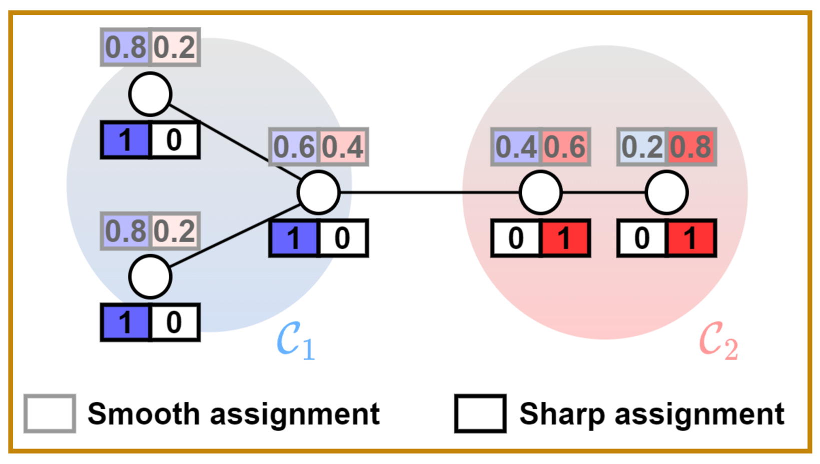
$$\underbrace{-\frac{\mathrm{Tr}(\mathbf{S}^T\tilde{\mathbf{A}}\mathbf{S})}{\mathrm{Tr}(\mathbf{S}^T\tilde{\mathbf{D}}\mathbf{S})}}_{\text{LQV term}} + \underbrace{\left\|\frac{\mathbf{S}^T\mathbf{S}}{\|\mathbf{S}^T\mathbf{S}\|_F} - \frac{\mathbf{I}_K}{\sqrt{K}}\right\|_F}_{\text{Balance term}}$$

MP layer

$$\mathbf{X}^{(l+1)} = \sigma\left[\underbrace{\tilde{\mathbf{A}}\mathbf{X}^{(l)}\boldsymbol{\Theta}_{\mathrm{MP}}^{(l)}}_{\text{LQV term}} + \mathbf{X}^{(l)}\boldsymbol{\Theta}_{\mathrm{SC}}^{(l)}\right]$$

3

1. The GNN Computes *soft assignments* **S** using MP layers and an MLP
2. GNN's clusters account both for *topology* and *node features*
3. Current approaches are based on loose graph cut relaxations with **LQV terms**
4. This results in *smooth* cluster assignments

 4

Total Variation GNN (TVGNN)

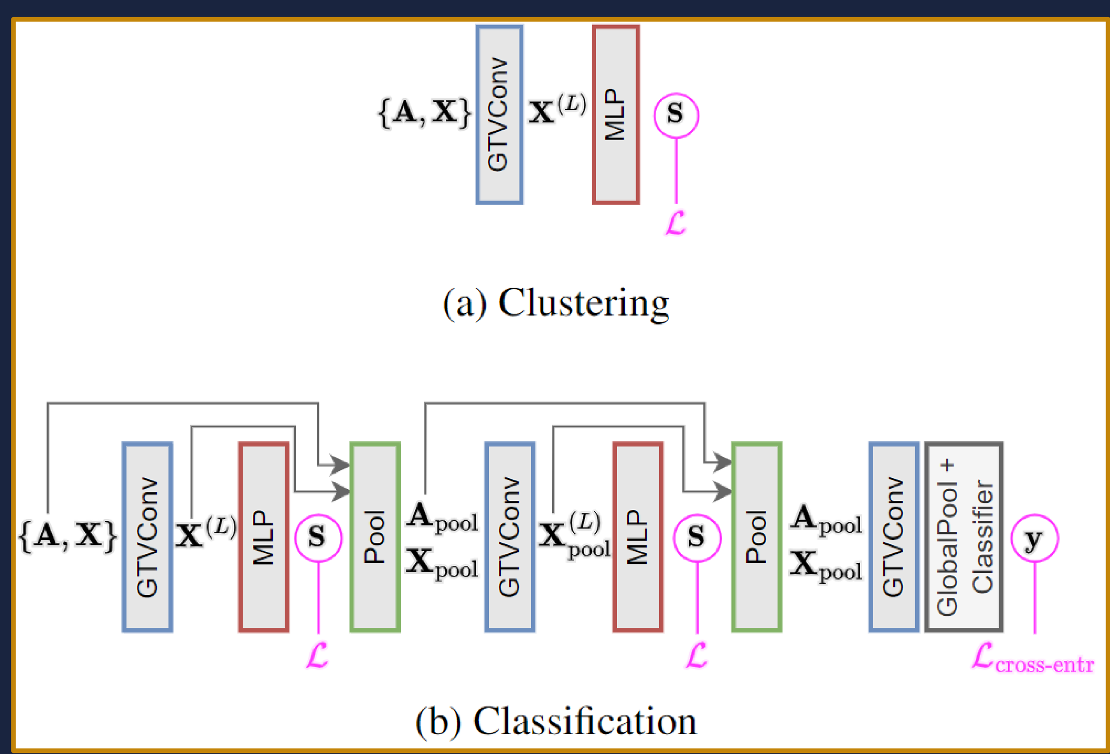
Loss based on a tighter relaxation of the minimum K-cut

$$\underbrace{\|\mathbf{S}\|_{\mathrm{GTV}}}_{\text{GTV term}} - \underbrace{\sum_{k=1}^{K}\|\mathbf{s}_{:,k} - \mathrm{quantile}_{K-1}(\mathbf{s}_{:,k})\|_{1,K-1}}_{\text{Balance term}}$$

Message passing layer derived from the gradient descent step of GTV

$$\mathbf{X}^{(l+1)} = \sigma\left[\underbrace{\left(\mathbf{I} - 2\delta\mathbf{L}_{\Gamma}^{(l)}\right)}_{\text{GTV term}}\mathbf{X}^{(l)}\boldsymbol{\Theta}^{(l)}\right]$$

Architecture

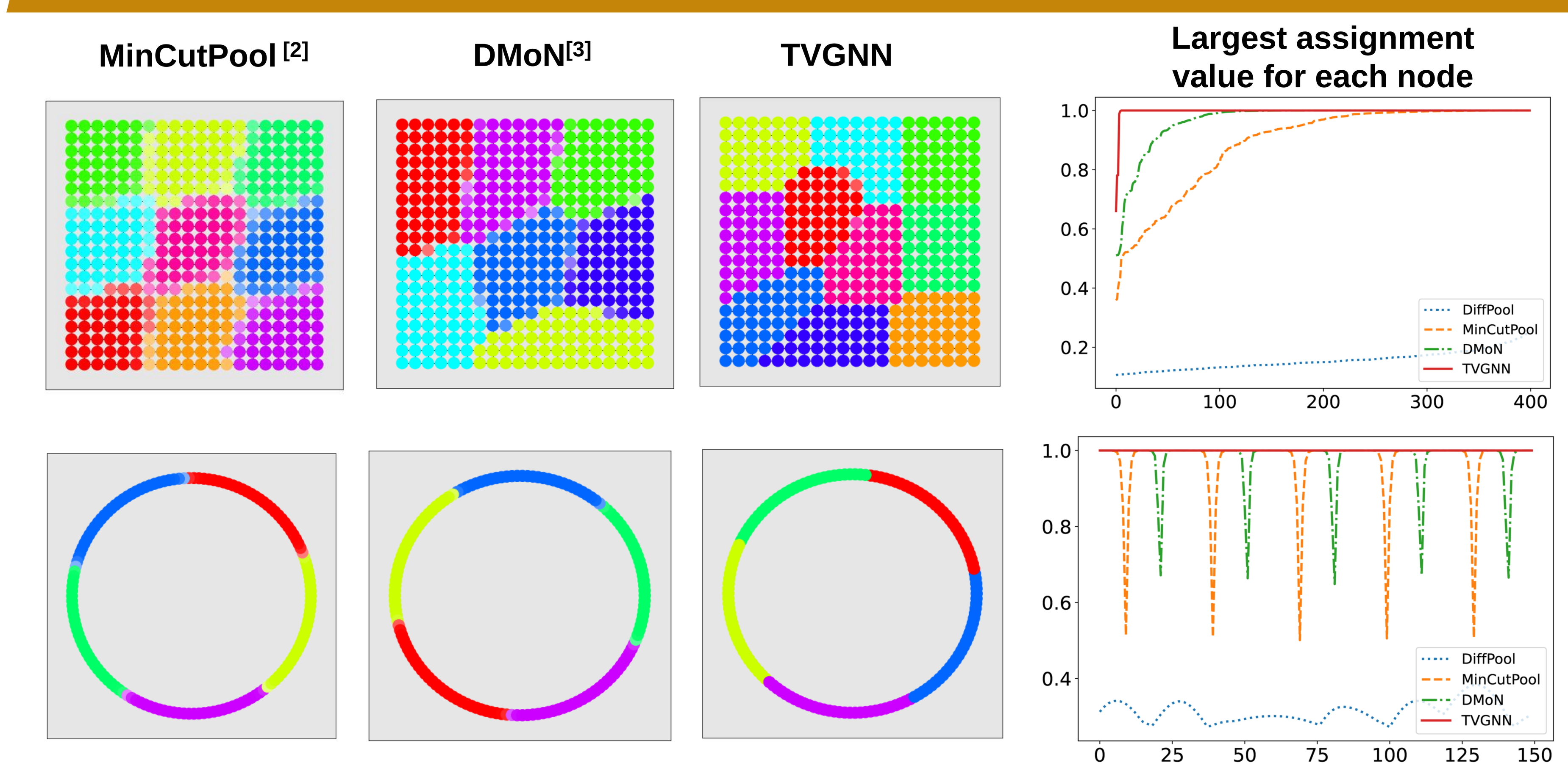


Properties

- Yields **sharper assignments** compared to methods that minimize LQV
- Implicitly **learns connectivity weights** given by $\mathbf{L}_{\Gamma} = \mathbf{D}_{\Gamma} - \boldsymbol{\Gamma}$ where

$$[\boldsymbol{\Gamma}]_{ij} = \frac{[\mathbf{A}]_{ij}}{\max\{\|\mathbf{x}_i - \mathbf{x}_j\|_1, \epsilon\}}$$

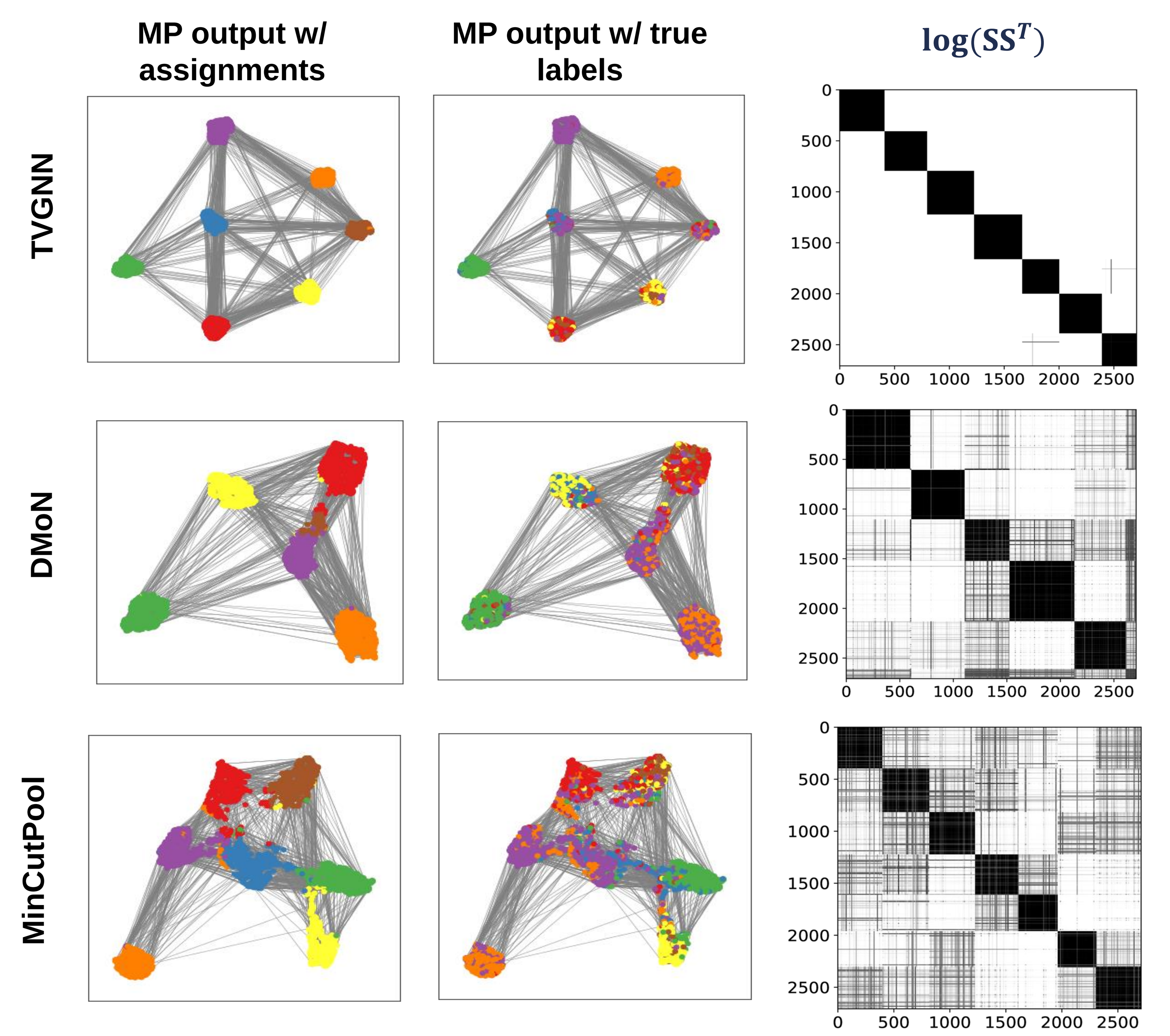
Results



Cluster assignments for simple 2D graphs

Dataset	Top-K	SAGPool	DiffPool	MinCutPool	DMoN	TVGNN	p-value
Bench-easy	53.8±31.8	53.8±31.8	99.0±0.3	99.0±0.3	98.8±0.5	**99.6**±0.6	.011*
Bench-hard	30.5±0.7	29.5±0.0	72.8±0.2	70.9±1.7	71.8±1.9	**75.3**±0.8	.001**
MUTAG	77.5±8.4	76.8±9.7	86.4±7.6	85.2±7.2	86.7±7.0	**88.4**±7.5	.606
Mutagenicity	68.4±8.4	68.2±7.8	78.5±1.5	78.4±1.4	77.1±1.3	**80.0**±1.3	.028*
NCI1	54.0±4.1	59.2±7.7	74.1±1.8	75.2±1.8	74.3±1.3	**77.3**±1.8	.018*
Proteins	69.6±2.7	70.4±2.5	74.6±4.2	75.7±3.0	75.2±3.3	**77.1**±2.9	.302
D&D	62.0±5.6	64.2±7.0	77.7±3.0	78.2±3.4	78.0±3.3	**79.5**±2.2	.323
COLLAB	73.4±6.9	75.6±2.5	78.4±1.6	79.3±1.1	79.5±0.7	**79.8**±1.1	.476
REDDIT-BINARY	54.0±10.0	50.0±0.1	80.9±2.7	82.3±3.2	82.6±2.9	**86.5**±2.8	.007**

Graph classification performance



Latent features and cluster assignments for Cora

References

1. Bresson, X., Laurent, T., Uminsky, D. and Von Brecht, J. *Multiclass total variation clustering*. Advances in Neural Information Processing Systems, 2013.
2. F. M. Bianchi, D. Grattarola, and C. Alippi. *Spectral clustering with graph neural networks for graph pooling*. International Conference on Machine Learning, 2020.
3. A. Tsitsulin, J. Palowitch, B. Perozzi, and E. Muller. *Graph clustering with graph neural networks*. arXiv preprint arXiv:2006.16904, 2020.

Contacts

filippo.m.bianchi@uit.no
@FilippoMariaBi1

Scan the QR code to access the official repository