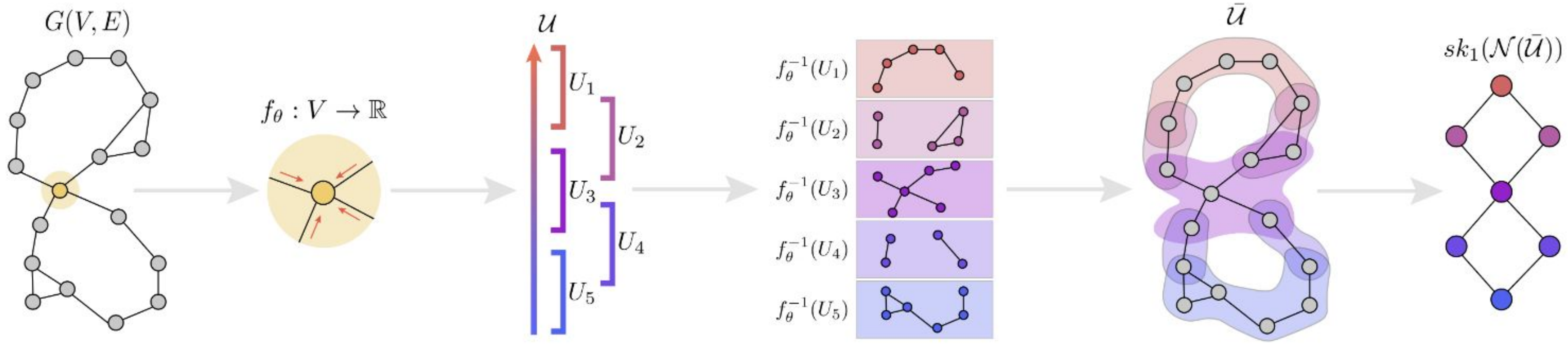
Deep Graph Mapper

Seeing Graphs through the Neural Lens

Cristian Bodnar*, Cătălina Cangea*, Pietro Liò
{cb2015, ccc53, pl219}@cam.ac.uk

University of Cambridge

Topological Data Analysis and Beyond 2020
NeurIPS workshop



(a) (b) (c) (d) (e) (f)

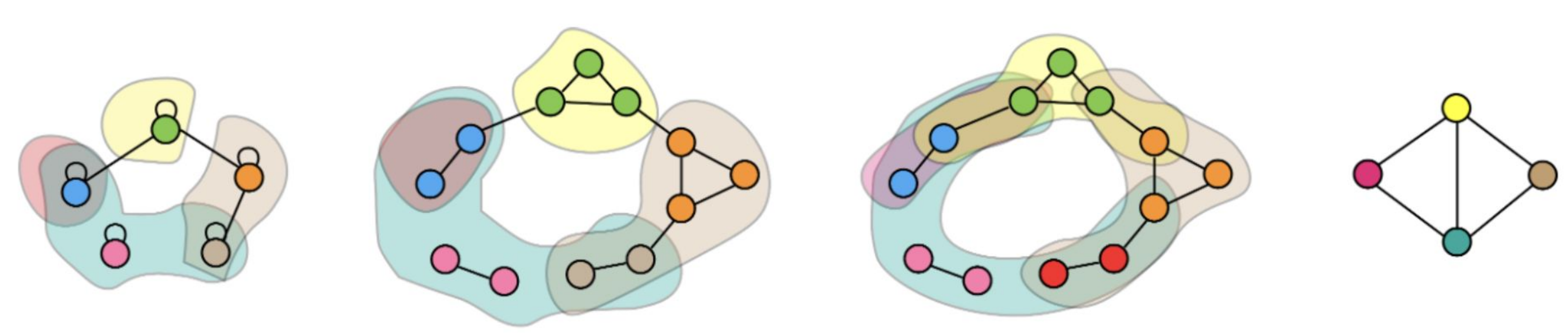
Contributions

We adopt a **topological perspective over graphs** and bring the following **contributions**:

- **Deep Graph Mapper (DGM)**: A (differentiable) Mapper-based algorithm for graph pooling / coarsening.
- Prove that **DGM generalises other popular pooling algorithms** based on soft-cluster assignments such as DiffPool or min-cut pooling.
- Demonstrate that DGM is **competitive with other state-of-the-art pooling algorithms**.
- Show how the algorithm can be used for **improved graph visualisations.**

Relationship to soft-cluster assignment methods

We offer a visual proof for the **relationship to soft cluster assignments**.



(a) (b) (c) (d)

Differentiable Mapper Pooling (DMP)

We first propose a **differentiable** pooling algorithm based on a 'lens function' **parameterised by GNNs**. The soft cluster assignment matrix is computed using a **kernel density estimation** approach.

$$\phi(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{\delta}\right)$$

$$S_{ij} = \frac{\phi(\sigma(f_\theta(\mathbf{X}_l))_i, x_j)}{\sum_{j=1}^{n} \phi(\sigma(f_\theta(\mathbf{X}_l))_i, x_j)}$$

PageRank-based Mapper Pooling (MPR)

We also design a **non-differentiable** pooling method leveraging a **PageRank-based** 'lens function', which exploits the **power-law distributions** often present in graph datasets. The soft cluster assignment matrix is computed from the **pull back cover**.

$$f(\mathbf{X})_i \triangleq \mathbf{PR}_i = \sum_{j \in N(i)} \frac{\mathbf{PR}_j}{|N(i)|}$$

$$S_{ij} = \frac{\mathbb{I}_{i \in f^{-1}(U_j)}}{|\{U_k | i \in f^{-1}(U_k)\}|}$$

Pooling results

Model	D&D	Mutag	NCI1	Proteins
DMP (Ours)	77.3 ± 3.6	84.0 ± 8.6	70.4 ± 4.2	$\mathbf{75.3 \pm 3.3}$
MPR (Ours)	$\mathbf{78.2 \pm 3.4}$	80.3 ± 6.0	69.8 ± 1.8	75.2 ± 2.2
Top-k	75.1 ± 2.2	82.5 ± 6.8	67.9 ± 2.3	74.8 ± 3.0
minCUT	77.6 ± 3.1	82.9 ± 6.0	68.8 ± 2.1	73.5 ± 2.9
DiffPool	77.9 ± 2.4	$\mathbf{94.7 \pm 7.1}$	68.1 ± 2.1	74.2 ± 0.3
WL	77.4 ± 2.6	74.5 ± 6.5	$\mathbf{76.4 \pm 2.7}$	74.7 ± 3.2
Flat	69.9 ± 2.2	71.8 ± 4.3	65.5 ± 1.7	70.2 ± 2.6
avg-MLP	63.7 ± 1.4	69.1 ± 5.8	55.7 ± 2.8	61.8 ± 1.7

Model	Collab	IMDB-B	IMDB-M	Reddit-B	Reddit-5k
DMP (Ours)	81.4 ± 1.2	$\mathbf{73.8 \pm 4.5}$	$\mathbf{50.9 \pm 2.5}$	86.2 ± 6.8	51.9 ± 2.1
MPR (Ours)	$\mathbf{81.5 \pm 1.0}$	73.4 ± 2.7	50.6 ± 2.0	86.3 ± 4.8	52.3 ± 1.6
Top-k	75.0 ± 1.1	69.6 ± 3.8	45.0 ± 2.8	79.4 ± 7.4	48.5 ± 1.1
minCUT	79.9 ± 0.8	70.7 ± 3.5	50.6 ± 2.1	$\mathbf{87.2 \pm 5.0}$	$\mathbf{52.9 \pm 1.3}$
DiffPool	81.3 ± 0.1	72.4 ± 3.1	50.3 ± 1.8	79.0 ± 1.1	50.4 ± 1.7
WL	78.5 ± 1.1	72.1 ± 3.1	50.7 ± 2.9	66.7 ± 10.4	49.2 ± 1.4
Flat	80.9 ± 1.4	73.6 ± 4.2	48.5 ± 2.4	70.0 ± 10.8	49.5 ± 1.7
avg-MLP	74.8 ± 1.3	71.5 ± 2.9	49.5 ± 2.2	53.6 ± 6.2	45.9 ± 1.6

Visualisations