MS-IMAP - A Multi-Scale Graph Embedding Approach for Interpretable Manifold Learning

Shay Deutsch [1] Lionel Yelibi [1] Alex Lin [1] Arjun Ravi Kannan [1]

shaydeu@gmail.com ylblio001@myuct.ac.za atlin271@gmail.com arjun.kannan@gmail.com

Motivations

Challenges in unsupervised learning:

Developing robust methods for unsupervised learning using graph embeddings methods is still a challenging problem due to several reasons. We name a few key challenges:

- Noisy data
- Computational complexity
- Topology: Lack of "good" trade-off between local and global methods
- Feature importance and interpretation.

To address these limitations, this project proposes three contributions.

1. Multiscale graph representation with Spectral Graph Wavelet (SGW)[1]: preprocess data by capturing localized, multiscale features in graph-structured data. Enhances signal filtering, denoising, and smoothness, improves feature extraction, and aid graph-based learning.
2. Tensor-Based Constrastive Learning improves Separability and Interpretation.
3. The Laplacian score selects features that preserve local structure in data, identifying those most relevant for clustering and classification.

Clustering a synthetic two-moon dataset: From left to right the original dataset, three modern methods, and ours are shown:

1. UMAP[3]: Fuzzy Cross-Entropy Optimization.
2. t-SNE[4]: KL-divergence optimization.
3. HeatGeo[2]: Heat-geodesic Optimization.
4. MS-IMAP: Tensor-Based SGW Fuzzy Cross-Entropy Optimization improves significantly the separation of clusters for clustering.

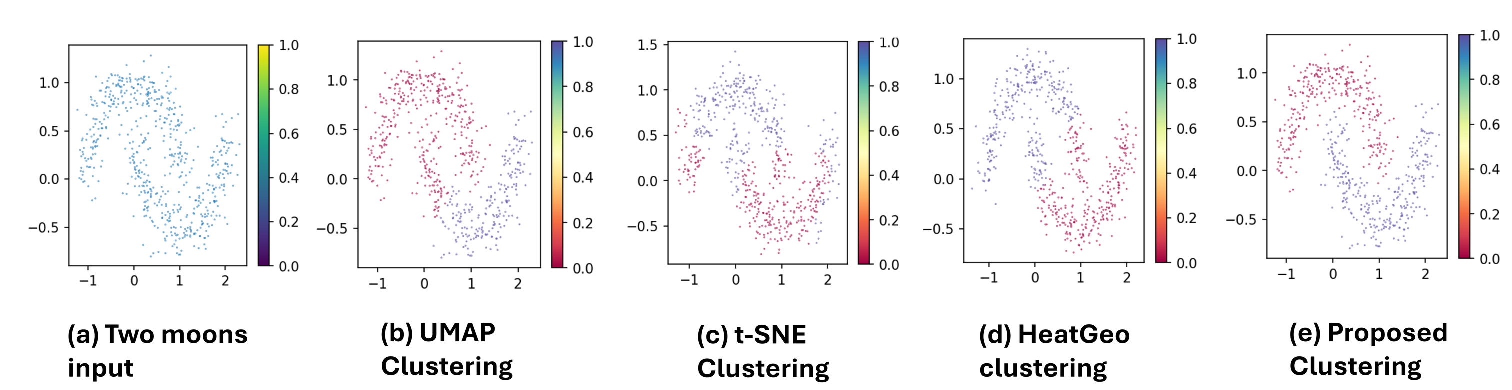

Figure 1:Clustering results using two moons

Multiscale representations using SGW

Given a bandpass filter in the spectral domain $g(\lambda)$ and a wavelet ψ centered on node i at scale s. Wavelets $\psi(s,i)$ are calculated by applying them to the delta function on a single vertex i:

$$\psi(s,i) = \Phi g(s\lambda)\Phi^T \delta_i \tag{1}$$

The value of $\psi(s,i)$ with respect to a vertex m can be written as:

$$\psi(s,i)(m) = \sum_{l=1}^{N} g(s\lambda_l)\phi_l^*(i)\phi_l(m)$$

Given a graph signal $f \in \mathbb{R}^N$, the ***SGW coefficient*** at node i and scale s can be expressed as follows:

$$\psi_f(s,i) = \sum_{l=1}^{N} g(s\lambda_l)\hat{f}(\lambda_l)\phi_l(i) \tag{2}$$

3D-Tensor Encoding

The raw features are transformed into their multiscale representations through spectral graph wavelet feature extraction. These representations are then concatenated into a 3D tensor $\psi_x(i,j,k)$ of shape $K \times D \times N$, where K denotes the number of scales s_j, D is the feature dimensionality, and N represents the number of nodes (or data points). This 3D tensor captures the multiscale characteristics of the data across different resolutions. The resulting encoded tensor will serve as an initialization for subsequent optimization, which is performed using stochastic gradient descent (SGD) to fine-tune the representations for the specific task.

3D-Tensor Based Optimization

The stochastic gradient descent (SGD) update rule applied at iteration t to the 3D tensor is given by:

$$(\tilde{\psi}_x^{(t+1)})_{i,j,k} = (\tilde{\psi}_x^{(t)})_{i,j,k} - \alpha \frac{\partial \mathcal{L}}{(\partial \psi_x)_{i,j,k}} \tag{3}$$

Here, α is the learning rate, and the gradient of the loss function $\mathcal{L}$ with respect to the tensor ψ_x is computed to update the representation at each scale s_k. The fuzzy cross entropy loss is defined as:

$$\mathcal{L}(\tilde{\psi}_x|W) = \sum_{i,j,k}\left(w_{ij}\log\frac{w_{ij}}{v_{ij}^{\psi_x(s_k,:,:)}} + (1-w_{ij})\log\frac{1-w_{ij}}{1-v_{ij}^{\psi_x(s_k,:,:)}}\right) \tag{4}$$

In this equation, w_{ij} denotes the dependency strength between nodes in the graph W, and $v_{ij}^{\psi_x}$ is the predicted dependency for the given scale s_k.

Optimization is performed across all scales s_k using SGD to minimize the loss for each scale. Batches are formed via positive and negative sampling based on node dependencies in the graph.

The final optimized embeddings across scales are summed as:

$$\tilde{\psi}_x = \sum_k \tilde{\psi}_x(s_k,:,:) \tag{5}$$

The resulting embedding $\tilde{\psi}_x$ retains the same dimensionality as the raw features, ensuring a one-to-one correspondence between the two.

Laplacian Score

Embedded features $(\tilde{\psi}_x)_l$ are evaluated using the Laplacian Score, which projects them onto the graph Laplacian L and Degree matrix D to quantify how well the features preserve local structure in the data.

The Laplacian Score $L_s(\tilde{\psi}_x)_l$ of a single feature f_r, which is used to construct coordinate l in the embedding space $(\tilde{\psi}_x)_l$, is defined as:

$$L_s(\tilde{\psi}_x)_l = \frac{(\tilde{\psi}_x)_l^T L (\tilde{\psi}_x)_l}{(\tilde{\psi}_x)_l^T D (\tilde{\psi}_x)_l} \tag{6}$$

In this expression, the numerator $(\tilde{\psi}_x)_l^T L(\tilde{\psi}_x)_l$ measures the smoothness of feature f_r on the graph, while the denominator $(\tilde{\psi}_x)_l^T D(\tilde{\psi}_x)_l$ measures the importance of the feature in maintaining local relationships. Features with lower Laplacian Scores are more effective at preserving the local structure of the data, making them highly useful for feature selection, interpretability, and importance ranking in downstream tasks.

Experimental Results

Method / Accuracy	ARI	AMI
UMAP	0.54	0.51
t-SNE	0.42	0.35
ISOMAP	0.36	0.3
Diffusion Maps	0.25	0.19
HeatGeo	0.54	0.52
MS-IMAP Method 1	0.75	0.73
MS-IMAP Method 2	**0.89**	**0.87**

Table 1:Comparison of clustering performance on the Two Moons datasets.

Dataset	Census		Zilionis		AWA	
Method / Accuracy	ARI	AMI	ARI	AMI	ARI	AMI
UMAP	0.05	0.08	0.45	0.71	0.73	0.80
t-SNE	0.01	0.02	0.64	0.34	0.68	0.76
HeatGeo	0.15	0.10	N/A	N/A	0.65	0.74
MS-IMAP Method 1	0.15	0.14	0.69	**0.76**	0.72	0.80
MS-IMAP Method 2	**0.22**	**0.15**	**0.70**	**0.76**	**0.74**	**0.81**

Table 2:Clustering results comparison using ARI and AMI the Census, Zilionis, and AWA datasets. Note HeatGeo had code execution issues on the Zilionis dataset.

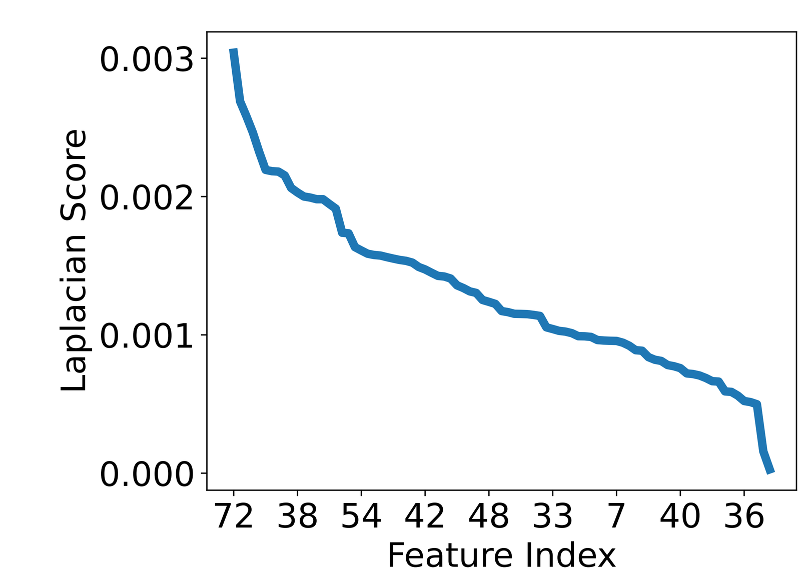

Figure 2:AWA Dataset Laplacian Score plotline showing in order from high to low, the importance of the original features with respect to the embeddings.

References

[1] David K. Hammond, Pierre Vandergheynst, and Rémi Gribonval. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 2011.

[2] Guillaume Huguet, Alexander Tong, Edward De Brouwer, Yanlei Zhang, Guy Wolf, Ian Adelstein, and Smita Krishnaswamy. A heat diffusion perspective on geodesic preserving dimensionality reduction. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

[3] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction, 2018.

[4] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, pages 2579–2605, 2008.