# Projet Geometric Data Analysis (J. Feydy) - MVA 2022

**Clustering with Multi-Layer Graphs : a Spectral Perspective,**
**Dong, X., Frossard, P., Vandergheynst, P., and Nefedov, N., 2011.**

Capucine GARÇON - Sonia MAZELET

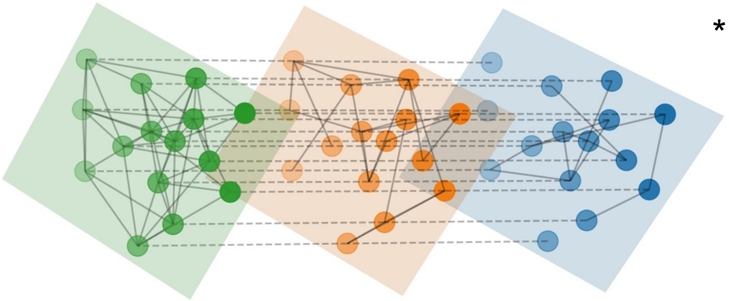
# Sommaire

*Clustering with Multi-Layer Graphs : a Spectral Perspective*

# 1° Context

- **Multi-layer graphs:** group of graphs that share the same nodes but have different vertices

  ⇒ represent multimodal data

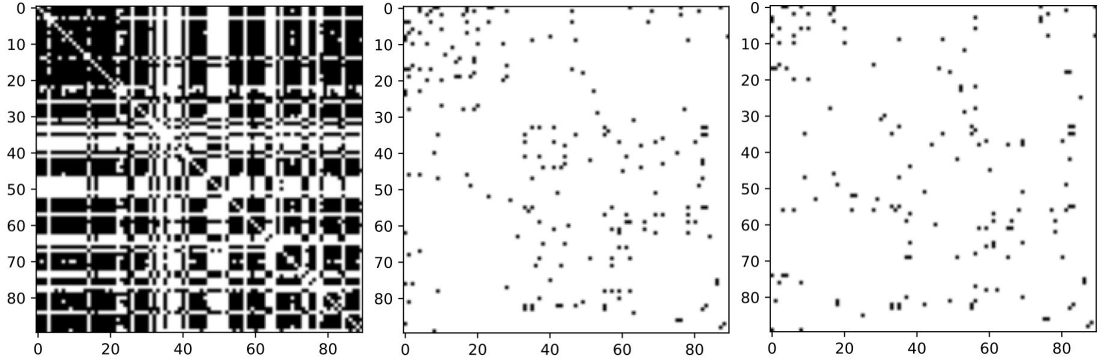


*

**multi-graph generated by networkx and matplotlib*

# 1° Context

- **Goal:** Perform clustering on multi layer graphs using a joint spectrum of the Laplacian matrices of the different layers.




**Laplacian matrices of a 3-layer graph constructed with the MIT Reality Mining dataset.**
**From left to right: Proximity, Calls and Friendship.**

# 2°1 Spectral Clustering

- Method for finding $k$ clusters in simple (one layer) graphs.
- Look for the $k$ smallest eigenvalues, compute the matrix having the $k$ associated eigenvectors for columns, and perform K-means clustering on its lines.

  $\Rightarrow$ projecting the vertices into the low-dimensional spectral domain.
- **Goal of the article:** Generalize to multi-layer graphs.

# 2°2 Clustering with generalized eigen-decomposition (SC-GED)

- **Idea:** Make a joint-spectrum, as close as possible to the spectrums of the layers, construct a joint Laplacian with this spectrum and perform K-means clustering on its lines.

- For $\mathrm{M}$ the number of layers $L_{rw}^{(i)}$ the normalized Laplacian of each layer and $\Lambda^{(i)}$ its diagonal matrix of eigenvalues:

$$\arg\min_{P,Q} S = \frac{1}{2}\sum_{i=1}^{M} \underbrace{||L_{rw}^{(i)} - P\Lambda^{i}Q||_F^2}_{\substack{\text{joint spectrum is close to the} \\ \text{spectrum of the Laplacian each layer}}} + \underbrace{\frac{\alpha}{2}(||P||_F^2 + ||Q||_F^2)}_{\text{control of } P \text{ and } Q} + \underbrace{\frac{\beta}{2}||PQ - I_n||_F^2}_{Q \text{ is close to the inverse of } P}$$

# 2°3 Clustering with spectral regularization (SC-SR)

- **Idea**: Use the relative importance of each layer, and construct a function $f : V \to \mathbb{R}$ ( $V$ number of vertices) 'smooth' on each layer, on which to perform clustering.
- **Smooth function:** function that preserves the proximity between vertices.
- **For one layer:**

$$\arg\min_{f} f^T L f \quad \text{st} \quad ||f|| = 1, f \perp 1$$

- **Observation:** If the eigenvalues are small, the eigenvectors of $L$ are good candidates.

# 2°3 Clustering with spectral regularization

- **For two layers**: Take the first layer to be the most informative.

$$\arg\min_{f} \underbrace{\frac{1}{2}||f_i - u_i||^2}_{\substack{f \text{ is close to the} \\ \text{eigenvectors of the first layer}}} + \underbrace{\lambda f_i^T L_{rw}^{(2)} f_i}_{\substack{f \text{ is smooth on} \\ \text{the second layer}}}$$

- **Generalization to $\mathrm{M}$ layers:** After finding f for the first two layers, iterate on all the other layers, sorted in order of importance.

- The final function plays the role of a joint smooth spectrum, on which to perform clustering

# 3° Nos expérimentations

- Implementation of single layer clustering algorithm : Spectral Clustering
- Implementation of multi-layer clustering algorithms :
  - **Clustering with generalized eigen-decomposition**.
  - **Clustering with spectral regularization**.
  - **Kernel K-means**, as a baseline for our experiments.
- Datasets used :
  - Cora Dataset (two layers)
  - MIT Reality Mining Dataset (three layers).

**One issue :** Finding the most informative layer ?

Experiments available at : https://github.com/CapucineGARCON/MVA2022-Geometric-Data-Analysis

# 3°1 Evaluation criteria

- **Purity** : $$Purity(\Omega, C) = \frac{1}{N}\sum_k \max_j \underbrace{|\omega_k \cap c_j|}_{\substack{\text{number of common} \\ \text{vertices in } \omega_k \text{ and } c_j}}$$

- **Normalized Mutual Information (NMI)** : For $I$ the mutual information and $H$ the entropy

$$NMI(\Omega, C) = \frac{I(\Omega, C)}{(H(\Omega) + H(C))/2}$$

# 3°2 How to find the most informative layer ?

- The **most informative layer is necessary** for :
  - SC-GED : the initialization of the P and Q matrices.
  - SC-SR : to build the first low-dimensional embedding U.

- Implementation of a function to order the layer from the most informative to the least :
  1. Performing **Spectral clustering on each layer** independently;
  2. **NMI** between predictions of each layer;
  3. The most layer is **the one with the highest NMI with each of the other layers**;
  4. Sorting layers by NMI with the first one.

# 3°3 Our experiments on Cora Dataset

- Nodes : 40 different papers.
- Layer Citations : more information on relations due to the scarcity of the matrix → best results.
- Layer **Words : more informative** → getting better results when using this layer as the most informative. *
- Multi-layer clustering not better that single layer clustering.

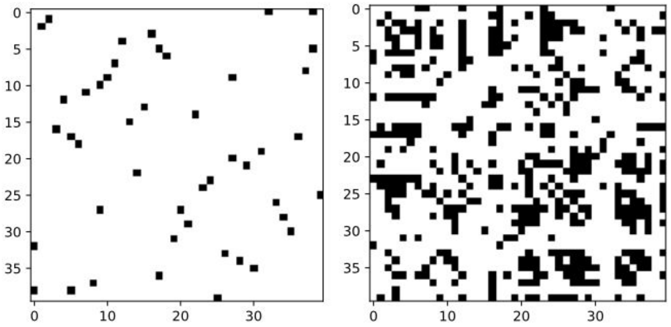


**Figure 1 : Laplacian matrices of a 2-layer graph constructed with the Cora dataset. From left to right: Citations, Words.**

| | Cora Dataset | | | | |
|---|---|---|---|---|---|
| | Words | Citations | SC-GED* | SC-SR* | K-Kmeans* |
| **Purity** | 0.70 | **0.75** | 0.70 | 0.70 | 0.60 |
| **NMI** | 0.23 | **0.30** | 0.27 | 0.23 | 0.05 |

# 3°3 Our experiments on MIT Reality Mining Dataset

- Nodes : 90 different people.
- Result from our method : Calls, Friendship, Proximity (from the most informative to the least).
- Better results obtained with : **Proximity, Calls, Friendship**. *
  → Proximity contains more informations.
- **Multi-layer clustering performs better** than single layer clustering.

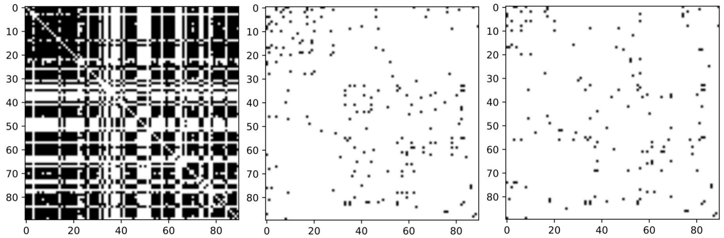


**Figure 2 : Laplacian matrices of a 3-layer graph constructed with the MIT Reality Mining dataset. From left to right: Proximity, Calls and Friendship.**

| | MIT Reality Mining Dataset | | | | | |
|---|---|---|---|---|---|---|
| | Friendship | Proximity | Calls | SC-GED* | SC-SR* | K-Kmeans* |
| **Purity** | 0.42 | 0.41 | 0.37 | 0.43 | 0.51 | **0.61** |
| **NMI** | 0.20 | 0.08 | 0.11 | 0.13 | 0.22 | **0.37** |

# Limitations

Limitations :

- **Results highly depends on the choice of the most informative layer** which is sometimes difficult to determine.
- Have to keep in mind that we do not use the same dataset.

Extensions:

- Study the impact of the trade-off parameters $\lambda$ , $\beta$ , $\alpha$ .
- Study how the number of layers impacts the performance.

# Conclusion

- Generally, **multi-layer clustering performs better** than single-layer clustering.
- **SC-SR is better than SC-GED** : performs better and **computationally efficient**.
- SC-SR enables to capture specificity about each layer.

# Références

[1] Argyriou, A., Herbster, M., Pontil, M. (2005). Combining graph Laplacians for semi-supervised learning. *Advances in Neural Information Processing Systems* 18 (2005).

[2] Dong, X., Frossard, P., Vandergheynst, P., and Nefedov, N. (2012). *Clustering with Multi-Layer Graphs : a Spectral Perspective. In Proceedings of the IEEE Transactions on Signal Processing* 60, 11, 5820-5831.

[3] Shi, J., and Malik, J. 2000. Normalized cuts and image segmentation. IEEE *Transactions on pattern analysis and machine intelligence* 22, 8 (2000), 888–905.

[4] Zhou, D., and JC Burges, C. 2007. Spectral clustering and transductive learning with multiple views. In *Proceedings of the 24th international conference on Machine learning.* 1159–1166.

**Thank you for your attention !**