# Learning the right layers: a data-driven layer-aggregation strategy for semi-supervised learning on multilayer graphs

**Sara Venturini**

PhD Student, Computational Mathematics, University of Padova, Italy

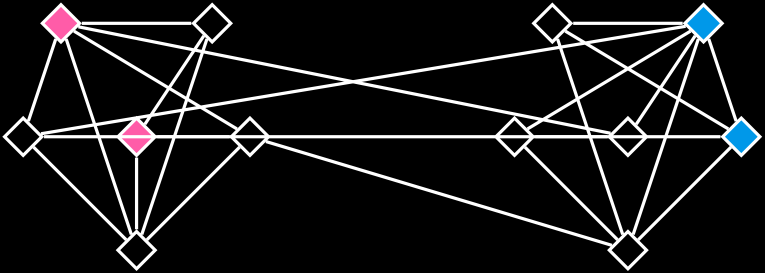
sara.venturini@math.unipd.it

Andrea Cristofari (University of Rome “Tor Vergata”)

Francesco Rinaldi (University of Padova)

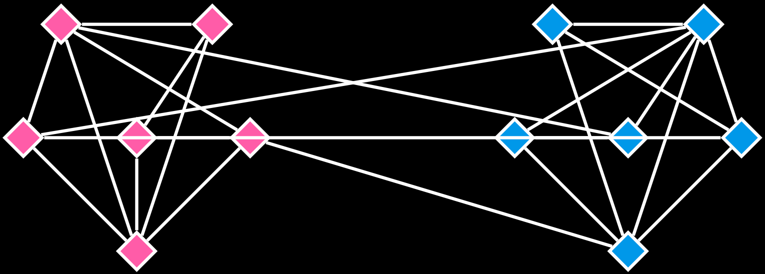
Francesco Tudisco (Gran Sasso Science Institute)

# Graph Semi-supervised Learning Problem



- $G = (V, E)$ graph (featureless)
- $C = \{C_1, \ldots, C_m\}$ set of classes of $G$
- set of input known labels for class : $Y_{ij} = 1$ if $i \in C_j$, and $Y_{ij} = 0$ otherwise

# Graph Semi-supervised Learning Problem



Aim: label the remaining vertices.

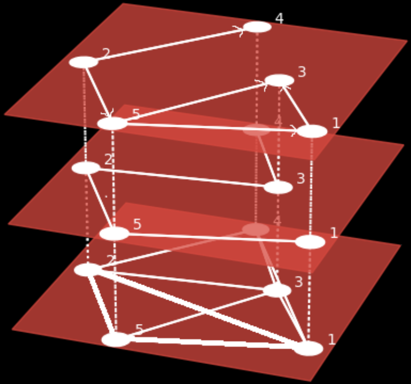
# Optimization problem

$$\min_{X} \varphi(X) := \|X - Y\|_F^2 + \frac{\lambda}{2} \sum_{i,j} A_{ij}^G \|X_{:i} - X_{:j}\|^2$$

where:

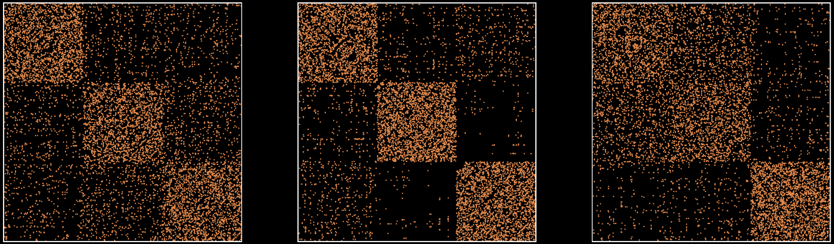
- $Y$ columns one-hot encoder vectors each class
- $A^G$ adjacency matrix of the graph
- $\lambda \geq 0$ regularization parameter

# Extension to Multilayer Networks



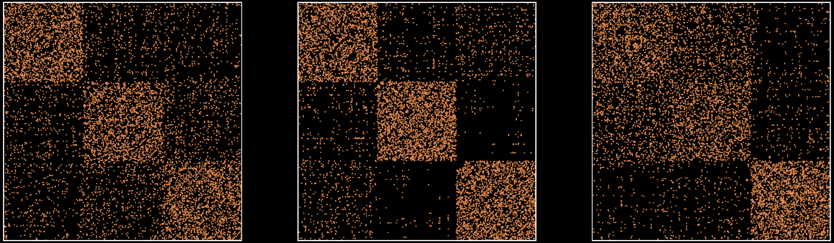

$A = \{A^{G^1}, \ldots, A^{G^K}\}$ with $A^{G^k}$ adjacency matrix layer $G^k$

Informative case



# Extension to Multilayer Networks

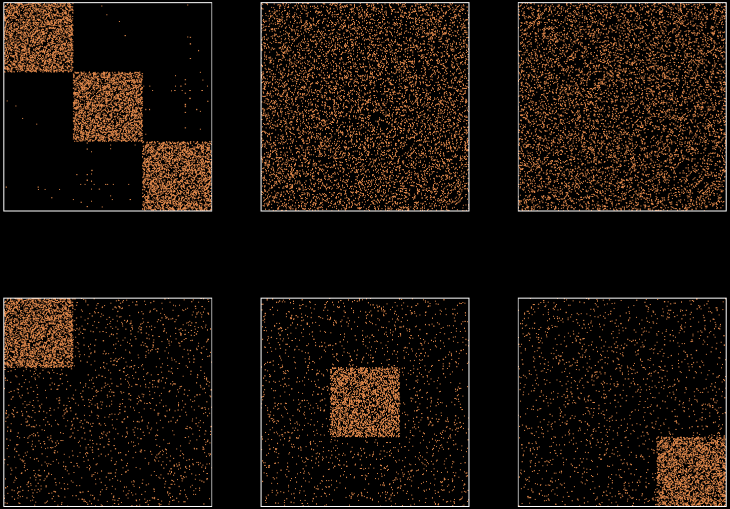
Informative case



Standard approaches:

- AGGR: $\frac{1}{K}\sum_{k=1}^{K} A_{ij}^{(k)}$
- UNION: $\max_{k=1,\dots,K} A_{ij}^{(k)}$

# Extension to Multilayer Networks



# Generalized mean adjacency model

The generalized mean adjacency matrix

$$A(\alpha, \beta)_{ij} = \left( \sum_{k=1}^{K} \beta_k \left(A_{ij}^{(k)}\right)^{\alpha} \right)^{1/\alpha} \quad \text{with } \alpha \in \mathbb{R}, \beta \geq 0, e^T \beta = 1,$$

| $\alpha \to -\infty$ | $\alpha = -1,\ \beta_k = 1/K$ | $\alpha \to 0,\ \beta_k = 1/K$ | $\alpha = 1,\ \beta_k = 1/K$ | $\alpha \to +\infty$ |
|---|---|---|---|---|
| $\min_{k=1,\ldots,K} A_{ij}^{(k)}$ | $\left( \frac{1}{K} \sum_{k=1}^{K} \frac{1}{A_{ij}^{(k)}} \right)^{-1}$ | $\left( \prod_{k=1}^{K} A_{ij}^{(k)} \right)^{1/K}$ | $\frac{1}{K} \sum_{k=1}^{K} A_{ij}^{(k)}$ | $\max_{k=1,\ldots,K} A_{ij}^{(k)}$ |
| Minimum | Harmonic | Geometric | Arithmetic | Maximum |

In order to learn the parameters $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}, \lambda)$, we split the available input labels into training and test sets: $Y^{tr}$ and $Y^{te}$, and consider the bilevel optimization model

$$\begin{aligned} \min_{\boldsymbol{\theta}} \quad & H(Y^{te}, X_{Y^{tr},\boldsymbol{\theta}}) \\ \text{s.t.} \quad & X_{Y^{tr},\boldsymbol{\theta}} = \arg\min_X \varphi(X, Y^{tr}, \boldsymbol{\theta}) \\ & \boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}, \lambda),\ \alpha \in \mathbb{R},\ \boldsymbol{\beta} \geq 0,\ \textstyle\sum_k \boldsymbol{\beta}_k = 1,\ \lambda \in \mathbb{R} \end{aligned}$$

with

- $H$ cross-entropy loss function
- $\varphi(X, Y, \theta) = \|X - Y\|_F^2 + \frac{\lambda}{2} \sum_{i,j} A(\alpha, \beta)_{ij} \|X_{:i} - X_{:j}\|^2$

# Lower level problem

The lower level problem

$$min_X \varphi(X, Y^{tr}, \boldsymbol{\theta})$$

is solved explicitly using Label Propagation,
over the graph induced by the generalized mean adjacency matrix

$$A(\alpha, \beta)_{ij} = \left( \sum_{k=1}^{K} \beta_k \left(A_{ij}^{(k)}\right)^{\alpha} \right)^{1/\alpha}$$

# Upper level problem

Feasible region:

$$S = \begin{cases} \alpha \in [-a, a] \\ \beta \geq 0, e^T\beta = 1 \\ \lambda \in [l_0, l_1] \end{cases}$$

We solve it using the Frank Wolfe algorithm with inexact gradient.
In each iteration we solve the linearized problem:

$$\hat{\theta} = \min_{\theta \in S} \widetilde{\nabla} H(\theta_n)^T(\theta - \theta_n)$$

which can be solved separately in the the variables $\theta = (\alpha, \beta, \lambda)$.

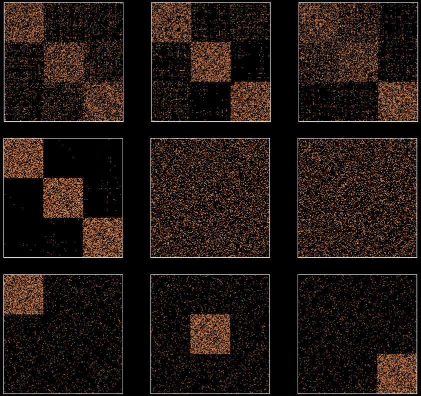
**Theorem (Informal)**

$\nabla H$ Lipschitz continuous, $S$ compact with finite diameter.
Let $\{\boldsymbol{\theta}_n\}$ a sequence generated by the Algorithm, where $\widetilde{\nabla} H$ and the step size satisfy some assumptions.
Then, we have a sublinear convergence rate of the duality gap:

$$g_n^* \leq \max(c_1 n^{-\frac{1}{2}}, c_2 n^{-1})$$

with appropriate constants $c_1$ and $c_2$,
$g_n^* = \min_{0 \leq i \leq n-1} -\nabla H(\boldsymbol{\theta}_i)^\top d_i^{FW}$, $d_i^{FW}$ direction obtained by the Frank-Wolfe algorithm with exact gradient.

# Synthetic Datasets



Performance Ratio

$$r_{a,d} = \frac{\mathcal{A}_{a,d}}{\max\{\mathcal{A}_{a,d} \text{ over all } a\}}$$

| | AGGR | UNION | BINOM | MULTI | SGMI | AGML | SMACD | GMM |
|---|---|---|---|---|---|---|---|---|
| **APR** | 0.87 | 0.89 | **0.98** | 0.97 | 0.78 | 0.61 | 0.50 | 0.86 |

# Real World Datasets

| | | 3sources | BBC | BBCSport | Wikipedia | UCI | cora | citeseer | dkpol | aucs | APR |
|---|---|---|---|---|---|---|---|---|---|---|---|
| AGGR | | 0.79 | 0.91 | 0.92 | 0.51 | 0.95 | 0.69 | 0.65 | 0.73 | 0.85 | 0.90 |
| | (+1) | 0.74 | 0.89 | 0.86 | 0.42 | 0.96 | 0.57 | 0.53 | 0.62 | 0.81 | |
| | (+2) | 0.07 | 0.88 | 0.80 | 0.39 | 0.96 | 0.49 | 0.47 | 0.58 | 0.77 | |
| UNION | | 0.75 | 0.90 | 0.92 | 0.51 | 0.92 | 0.69 | 0.65 | 0.69 | 0.85 | 0.86 |
| | (+1) | 0.66 | 0.87 | 0.85 | 0.42 | 0.92 | 0.57 | 0.53 | 0.60 | 0.65 | |
| | (+2) | 0.60 | 0.84 | 0.77 | 0.38 | 0.92 | 0.48 | 0.46 | 0.55 | 0.52 | |
| BINOM | | 0.75 | 0.88 | 0.92 | 0.62 | 0.97 | 0.74 | 0.66 | 0.62 | 0.85 | 0.94 |
| | (+1) | 0.76 | 0.87 | 0.91 | 0.57 | 0.97 | 0.63 | 0.59 | 0.54 | 0.81 | |
| | (+2) | 0.72 | 0.87 | 0.90 | 0.56 | 0.97 | 0.64 | 0.61 | 0.45 | 0.77 | |
| MULTI | | 0.74 | 0.86 | 0.88 | 0.64 | 0.96 | 0.76 | 0.65 | 0.76 | 0.85 | **0.98** |
| | (+1) | 0.73 | 0.87 | 0.87 | 0.62 | 0.96 | 0.76 | 0.63 | 0.72 | 0.81 | |
| | (+2) | 0.75 | 0.83 | 0.87 | 0.59 | 0.96 | 0.74 | 0.63 | 0.69 | 0.77 | |
| SGMI | | 0.75 | 0.76 | 0.84 | 0.61 | 0.94 | 0.72 | 0.51 | 0.31 | 0.75 | 0.84 |
| | (+1) | 0.58 | 0.76 | 0.83 | 0.59 | 0.94 | 0.72 | 0.52 | 0.31 | 0.76 | |
| | (+2) | 0.57 | 0.76 | 0.64 | 0.59 | 0.94 | 0.72 | 0.52 | 0.31 | 0.76 | |
| SMACD | | 0.62 | 0.69 | 0.73 | 0.24 | 0.33 | 0.34 | 0.36 | 0.26 | 0.58 | 0.56 |
| | (+1) | 0.60 | 0.66 | 0.60 | 0.25 | 0.30 | 0.28 | 0.32 | 0.24 | 0.56 | |
| | (+2) | 0.61 | 0.65 | 0.78 | 0.23 | 0.33 | 0.37 | 0.26 | 0.20 | 0.60 | |
| GMM | | 0.80 | 0.87 | 0.88 | 0.57 | 0.93 | 0.69 | 0.58 | 0.63 | 0.81 | 0.85 |
| | (+1) | 0.76 | 0.84 | 0.77 | 0.47 | 0.93 | 0.58 | 0.49 | 0.34 | 0.77 | |
| | (+2) | 0.71 | 0.81 | 0.73 | 0.43 | 0.93 | 0.55 | 0.45 | 0.27 | 0.76 | |

Learning the Right Layers: a Data-Driven Layer-Aggregation Strategy for Semi-Supervised Learning on Multilayer Graphs,
Venturini, S., Cristofari, A., Rinaldi, F., and Tudisco, F.

Sara Venturini
sara.venturini@math.unipd.it
https://saraventurini.github.io