Spectral methods for graph clustering

Lorenzo Dall'Amico, Romain Couillet, Nicolas Tremblay
Email: `lorenzo.dallamico@isi.it`
Website: `lorenzodallamico.github.io`

My PhD manuscript

My publications

ABSTRACT

The "what"
- Graph spectral clustering
- Application to static, dynamical and weighted graphs

The "how"
- Physics-inspired arguments
- A constructive pipeline for efficient spectral clustering

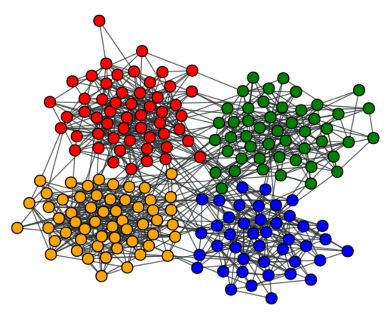
The results
- Practical and highly performing algorithms for clustering
- A unified interpretation of different methods

Graph clustering (I)

Definition

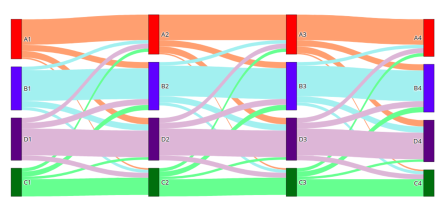
- **Partition** the nodes of a graph $\mathcal{G}(\mathcal{V},\mathcal{E})$ in k groups according to the edge configuration
- **Homophily**: "similar" nodes are strongly connected
- **Unsupervised** learning task

Three applications



- **Community detection**
INPUT: unweighted and undirected graph (*e.g.* a social network)
OUTPUT: affinity classes sharing, for instance, common interests

- **Dynamical community detection**
INPUT: a temporal graph sequence
OUTPUT: time-dependent evolution of the affinity classes

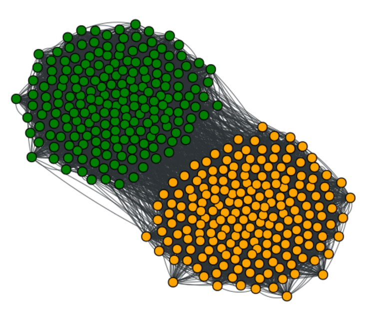


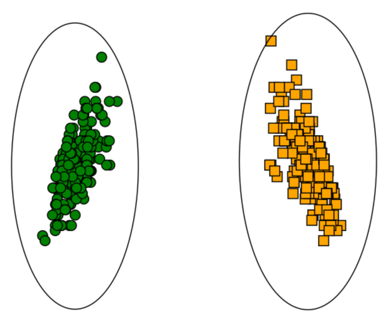
- **High dimensional vectors clustering**
INPUT: a set of feature vectors
REPRESENTATION: a weighted graph with edge weights measuring the proximity between the feature vectors
OUTPUT: clustering of the input data

Spectral clustering

Basic idea
- Define a **graph matrix representation** M $(A, D-A, D^{-1/2}AD^{-1/2})$
- Compute $X \in \mathbb{R}^{n\times k}$ with $X_{:,i}$ the eigenvector associated with the i-th smallest or largest eigenvalue of M
- The row $X_{i,:}$ maps node i to a $\mathbb{R}^k$

Spectral clustering in sparse graphs
- **Sparsity** is a *necessary* problem to deal with
- For community detection **unrelated contributions** suggest that *regularization* helps spectral clustering in sparse graphs
$H_r = (r^2-1)I_n + D - rA$, $L_\tau = D_\tau^{-1/2}AD_\tau^{-1/2}$ with $D_\tau = D + \tau I_n$





- What is the **optimal** regularization?
- Are the **dense** and **sparse** worlds to be treated **separately**?
- Can we design a **constructive pipeline** for spectral clustering?

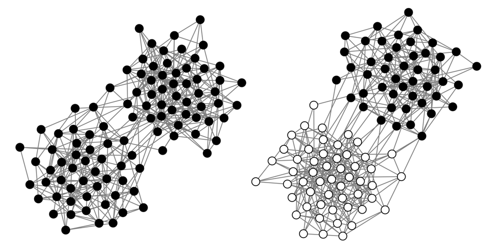
A constructive method for clustering (II)

The Bethe-Hessian matrix

- The Ising Hamiltonian: $\mathcal{H}(\boldsymbol{s}) = -\sum_{(ij)\in\mathcal{E}} J_{ij}s_is_j$, with $\boldsymbol{s}\in\{-1,1\}^n$
- The Boltzmann distribution: $\mu(\boldsymbol{s}) = \frac{e^{-\beta\mathcal{H}(\boldsymbol{s})}}{Z}$, with β the inverse temperature
- The free energy: $F_\beta(\boldsymbol{m}) = \beta\langle\mathcal{H}(\boldsymbol{s})\rangle_\mu - S_\mu$, with $\boldsymbol{m} = \langle\boldsymbol{s}\rangle_\mu$

The temperature balances the *energy* and the *entropy*

If $\mathcal{G}(\mathcal{V},\mathcal{E})$ has clusters, $F_\beta(\boldsymbol{m})$ has **local minima** associated to them



- F_β cannot be computed analytically → **Bethe approximation** $F_\beta^{\text{Bethe}}(\boldsymbol{m})$
- Hessian matrix of F_β^{Bethe} at $\boldsymbol{m} = \boldsymbol{0}_n$: the Bethe-Hessian matrix

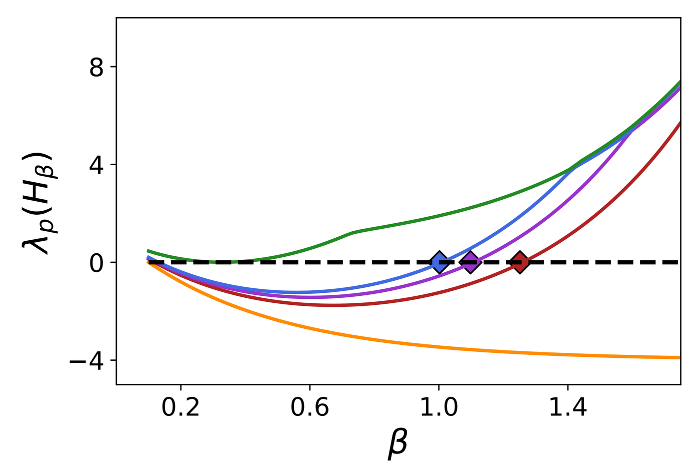
$$(H_\beta)_{ij} = \delta_{ij}\left(1 + \sum_{k\in\partial i}\frac{\text{th}^2(\beta J_{ik})}{1-\text{th}^2(\beta J_{ik})}\right) - \frac{\text{th}(\beta J_{ij})}{1-\text{th}^2(\beta J_{ij})}$$

Eigenvectors with negative eigenvalues of H_β to approximate the local minima of $F_\beta(\boldsymbol{m})$

Optimal temperature

- On synthetic static, dynamic and weighted graphs we established the **optimal temperatures** β
- They can be computed with an **unsupervised** and fast algorithm
- The value of k can be estimated using the Bethe-Hessian matrix

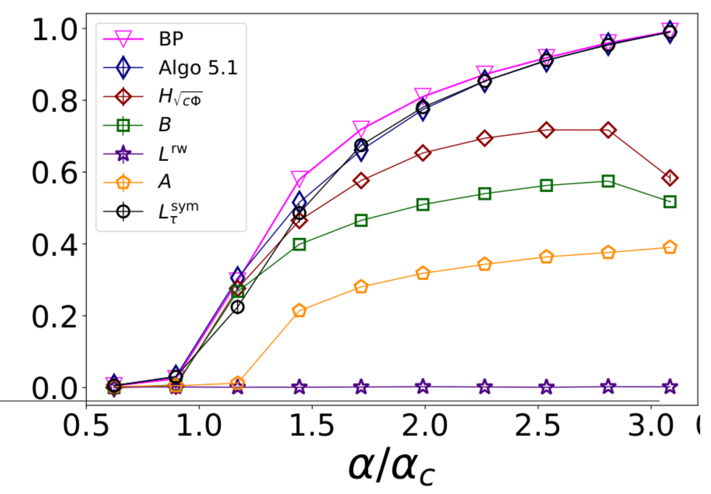
- $\hat{\beta}_p = \max_\beta\{\beta\ :\ \lambda_p(H_\beta) = 0\}$
- The largest p for which $\hat{\beta}_p$ is defined provides an **estimate of k**
- The embedding is obtained $X_{:p} = \boldsymbol{x}_p$ s.t. $H_{\hat{\beta}_p}\boldsymbol{x}_p = 0$



Main contributions (III)

New algorithms for spectral clustering

- The **optimal Bethe-Hessian** based clustering leads to **SOA** performances in spectral clustering in **all three applications** considered.
- On synthetic datasets the performance is close to **Bayes optimal**
- We released a **Julia** package called **CoDeBetHe.jl** with an efficient implementation of our algorithms



A unified framework for spectral clustering

Our constructive method (II) reduces to **commonly adopted** spectral methods for **sub-optimal** approximations or values of β:

- For $J = A - \frac{\boldsymbol{dd}^T}{2|\mathcal{E}|}$ and the **naive mean field** approximation: modularity matrix [Newman2006]
- For J the feature covariance matrix and the **naive mean field** approximation: PCA
- For $\beta\to\infty$: $D-A$ [Fiedler1973]

A **smooth transition** between the dense and the sparse worlds

$D-A$	$H_{\beta_{\text{opt}}}$		$H_{\beta_{\text{Saade}}}$
Trivial	Optimal		Worst case
$D^{-1/2}AD^{-1/2}$	$L_{f(\beta_{\text{opt}})}$	L_c	$L_{f(\beta_{\text{Saade}})}$
Trivial	Optimal	Qin13	Worst case

Conclusion

- **Self-adapting** algorithms
- **Bridge** between several theoretical results
- Definition of a **constructive pipeline** for spectral clustering either

Possibility to extend the **temperature-aware** pipeline to more involved settings?

ACKNOWLEDGMENTS

The work has been conducted at GIPSA lab in Grenoble during my PhD under the supervision of Romain Couillet and of Nicolas Tremblay. I further acknowledge the support from all the permanent staff and PhD students of the GAIA team.