# Optimal Laplacian Regularization for Sparse Spectral Community Detection

Lorenzo Dall'Amico, Romain Couillet, Nicolas Tremblay

# OPTIMAL LAPLACIAN REGULARIZATION FOR SPARSE SPECTRAL COMMUNITY DETECTION

*Lorenzo Dall'Amico*[†] *Romain Couillet*[†⋆] *Nicolas Tremblay*[†]

[†] GIPSA-lab, Université Grenoble Alpes, CNRS, Grenoble INP
[⋆]L2S, CentraleSupélec, University of Paris Saclay

**ABSTRACT**

Regularization of the classical Laplacian matrices was empirically shown to improve spectral clustering in sparse networks. It was observed that small regularizations are preferable, but this point was left as a heuristic argument. In this paper we formally determine a proper regularization which is intimately related to alternative state-of-the-art spectral techniques for sparse graphs.

***Index Terms*—** Regularized Laplacian, Bethe-Hessian, spectral clustering, sparse networks, community detection

## 1. INTRODUCTION

Community detection [1] is one of the central unsupervised learning tasks on graphs. Given a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes ($|\mathcal{V}| = n$) and $\mathcal{E}$ the set of the edges, it consists in finding a label assignment – $\hat{\ell}_i$ – for each node $i$, in order to reconstruct the underlying community structure of the network. The community detection problem has vast applications in different fields of science [2] and can be seen as the simplest form of clustering, *i.e.* the problem of dividing objects into similarity classes. We focus on unweighted and undirected graphs that can be represented by their adjacency matrices $A \in \{0,1\}^{n\times n}$, defined as $A_{i,j} = \mathbb{1}_{(ij)\in\mathcal{E}}$, where $\mathbb{1}_x = 1$ if the condition $x$ is verified and is zero otherwise.

A satisfactory label partition $\hat{\boldsymbol{\ell}}$ can be formulated as the solution of an optimization problem over various cost functions, such as *Min-Cut* or *Ratio-Cut* [3]. These optimization problems are NP-hard and a common way to find an approximate – but fast – solution, is by allowing a continuous relaxation of these problems. Defining the degree matrix $D \in \mathbb{N}^{n\times n}$ as $D = \mathrm{diag}(A\mathbf{1}) = \mathrm{diag}(\boldsymbol{d})$, the eigenvectors associated to the smallest eigenvalues of the combinatorial Laplacian matrix $L = D - A$ and to the largest eigenvalues of the matrix $D^{-1}A$ (that we refer to as $L^{\mathrm{rw}}$) provide an approximate solution to the *Ratio-Cut* and the *Min-Cut* problems, respectively (Section 5 of [3]). Retrieving communities from matrix eigenvectors is the core of *spectral clustering* [4, 5].

Although spectral algorithms are well understood in the *dense* regime in which $|\mathcal{E}|$ grows faster than $n$ (*e.g.* [6, 7]), much less is known in the much more challenging *sparse* regime, in which $|\mathcal{E}| = O(n)$. This regime is particularly interesting from a practical point of view, since most real networks are very sparse, and the degrees $d_i$ of each node $i$ scales as $d_i = (A\mathbf{1})_i \ll n$. Rigorous mathematical tools are still being developed, but some important results have already tackled the problem of community detection in sparse graphs.

We focus here on how regularization helps spectral clustering in sparse networks. In the dense regime (Figure 1A), for a graph $\mathcal{G}$ with $k$ communities, the $k$ largest eigenvalues of $L^{\mathrm{rw}}$ are isolated and the respective eigenvectors carry the information of the low rank structure of the graph. In the sparse regime instead, (Figure 1.B) the isolated informative eigenvalues are lost in the *bulk* of uninformative eigenvalues. Regularization avoids the spreading of the uninformative bulk and enables the recovery of the low rank structure of the matrix, as depicted in Figure 1.C. Among the many contributions proposing different types of regularization [8, 9, 10, 11, 12], we focus on the likely most promising one proposed by [8] that recovers communities from the eigenvectors corresponding to the largest eigenvalues of the matrix $L_\tau^{\mathrm{sym}} = D_\tau^{-1/2} A D_\tau^{-1/2}$, where $D_\tau = D + \tau I_n$. The authors build a line of argument under the assumption that the graph is generated from the degree-corrected stochastic block model [13] (DC-SBM). In the literature, the characterization of the parameter $\tau$ was never properly addressed and its assignment was left to a heuristic choice. More specifically, both in [8] and [10] the results provided by the authors seem to suggest a large value of $\tau$, but it is observed experimentally that smaller values of $\tau$ give better partitions. In the end, the authors in [8] settle on the choice of $\tau = \frac{1}{n}\mathbf{1}^T D\mathbf{1}$, *i.e.*, the average degree.

A fundamental aspect of community detection on sparse graphs generated from the DC-SBM, defined in Equation (1), is the existence of an information-theoretic threshold for community recovery [14, 15, 16, 17]: if the parameters of the generative model do not meet certain conditions, then no algorithm can assign the labels better than random guess.

In this article we study the problem of community detection on a network generated from the sparse DC-SBM and show why a small value of $\tau$ is preferable, drawing a connection to other existing algorithms based on the Bethe-Hessian matrix [18, 19], coming from statistical physics intuitions. We further show for which value of $\tau$ the leading eigenvectors of $L_\tau^{\mathrm{rw}} = D_\tau^{-1}A$ (and equivalently $L_\tau^{\mathrm{sym}}$) allow for non-trivial community reconstruction as soon as theoretically possible, addressing a question not answered in [8, 9, 11]. The correct parametrization of $\tau$ depends on the hardness of the detection problem and, for our proposed choice of $\tau$, the matrix $L_\tau^{\mathrm{rw}}$ has an eigenvector, corresponding to an isolated eigenvalue whose entry $i$ only depends on the class label of node $i$ and can be used to retrieve the community labels.

The remainder of the article is organized as follows: in Section 2 we present the generative model of the graph and the theoretical results about the detectability transition; in Section 3 we give the main result together with its algorithmic implementation; Section 4 closes the article.

**Notations.** Matrices are indicated with capital ($M$), vectors with bold ($\boldsymbol{v}_p$), scalar and vector elements with standard ($v_{p,i}$) letters. We denote by $s_i^{\uparrow}(M)$ the $i$-th smallest eigenvalue of a Hermitian

matrix $M$ and by $s_i^{\downarrow}(M)$ the $i$-th largest. $s_i(M)$ indicates a generic eigenvalue of $M$. The notation $M\boldsymbol{x} \approx 0$ indicates that, for all large $n$ with high probability, $\boldsymbol{x}$ is an approximate eigenvector of $M$ with eigenvalue $o_n(1)$.

## 2. MODEL

Consider a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ with $|\mathcal{V}| = n \gg 1$. We consider the DC-SBM [13] as a generative model for the $k$-class graph $\mathcal{G}$. Letting $\boldsymbol{\ell} \in \{1, \cdots, k\}^n$ be the vector of the true labels of a $k$-class network, the DC-SBM generates edges independently according to

$$\mathbb{P}(A_{ij} = 1 | \ell_i, \ell_j, \theta_i, \theta_j) = \theta_i \theta_j \frac{C_{\ell_i, \ell_j}}{n}. \tag{1}$$

The vector $\boldsymbol{\theta}$ allows for any degree distribution on the graph and satisfies $\frac{1}{n}\sum_{i\in\mathcal{V}} \theta_i = 1$ and $\frac{1}{n}\sum_{i\in\mathcal{V}} \theta_i^2 = \Phi = O_n(1)$. The matrix $C \in \mathbb{R}^{k\times k}$ is the class affinity matrix. Letting $\Pi = \mathrm{diag}(\boldsymbol{\pi}) \in \mathbb{R}^{k\times k}$, where $\pi_p$ is the fraction of nodes having label $p$, we assume that $C\Pi\mathbf{1} = c\mathbf{1}$, where it is straightforward to check that $c = \mathbb{E}[\frac{1}{n}\mathbf{1}^T D\mathbf{1}] = O_n(1)$ is the expected average degree, while, denoting with $d_i = D_{ii}$ the degree of node $i$, $\mathbb{E}[d_i] = c\theta_i$. This is a standard assumption [14, 20, 21, 22] that means that the expected degree of each node does not depend on its community, hence that the degree distribution does not contain any class information.

Considering the model of Equation (1) for $k = 2$ and $\boldsymbol{\pi} \propto \mathbf{1}_2$, we denote $C_{\ell_i,\ell_j} = c_{\text{in}}$ if $\ell_i = \ell_j$ and $C_{\ell_i,\ell_j} = c_{\text{out}}$ otherwise. As shown in [14, 15, 16], non-trivial community reconstruction is theoretically feasible if and only if

$$\alpha \equiv \frac{c_{\text{in}} - c_{\text{out}}}{\sqrt{c}} > \frac{2}{\sqrt{\Phi}} \equiv \alpha_c. \tag{2}$$

The parameter $\alpha$ regulates the hardness of the detection problem: for large $\alpha$ we have easy recovery, for $\alpha \leq \alpha_c$ the problem has asymptotically no solution. When $k > 2$ we distinguish two transitions: one from impossible to hard detection (the solution can be obtained in exponential time) and one from hard to easy detection (the solution can be obtained in polynomial time) [14].

## 3. MAIN RESULT

In this section we study the relation between the matrix $L_\tau^{\text{rw}} = D_\tau^{-1}A$ and the Bethe-Hessian matrix [18], defined as

$$H_r = (r^2 - 1)I_n + D - rA, \quad r \in \mathbb{R}. \tag{3}$$

We exploit some important results concerning the spectrum of $H_r$ to better understand why regularization helps in sparse networks.

### 3.1. Relation between $L_\tau^{\text{rw}}$ and the Bethe-Hessian matrix

In [18] it was shown that the Bethe-Hessian matrix can be efficiently used to reconstruct communities in sparse graphs. This matrix comes from the strong connection existing between the problem of community detection and statistical physics. The authors of [18] originally proposed to perform spectral clustering with the $k$ smallest eigenvectors of $H_r$ for $r = \sqrt{c\Phi}$. For this choice of $r$, if the problem is in the easy (polynomial) detectable regime, then only the $k$ smallest eigenvalues of $H_r$ are negative, while $s_{k+1}^{\uparrow}(H_r) \approx 0$. In [23] we refined this approach in a two-class setting, showing that there exists a parametrization – depending on the clustering difficulty – that leads to better partitions under a generic degree distribution and, at the same time, provides non-trivial clustering as soon as theoretically possible. In [19] we extended our reasoning for more than two classes and studied the shape of the informative eigenvectors. We here recall our main findings.

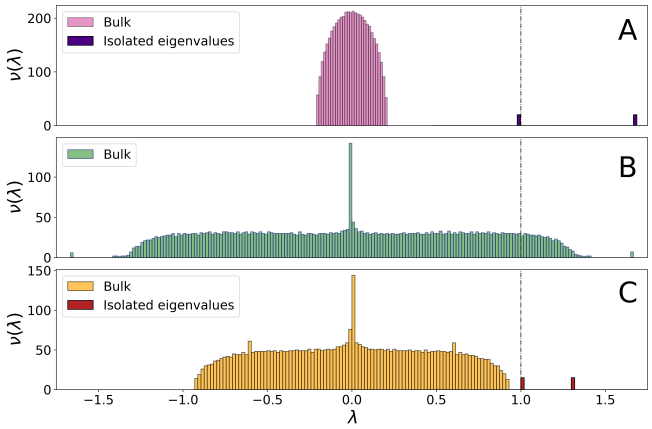


**Fig. 1**. In all three figures $r = \zeta_2 = \frac{c_{\text{in}}+c_{\text{out}}}{c_{\text{in}}-c_{\text{out}}}$, $n = 5000$, $\theta_i \sim [\mathcal{U}(3,7)]^3$, $k = 2$. **A:** Spectrum of $rD^{-1}A$ in the dense regime ($c_{\text{in}}/n = 0.08$, $c_{\text{out}}/n = 0.02$). **B:** Spectrum of $rD^{-1}A$ in the sparse regime ($c_{\text{in}} = 8$, $c_{\text{out}} = 2$). **C:** Spectrum of $rD_{r^2-1}^{-1}A$ in the sparse regime ($c_{\text{in}} = 8$, $c_{\text{out}} = 2$).

**Result 1** (from [19])**.** *Let $p$ be an integer between 1 and $k$. The equation $s_p^{\uparrow}(H_r) \approx 0$ is verified for*

$$r = \zeta_p = \frac{c}{s_p^{\downarrow}(C\Pi)} \quad \in (1, \sqrt{c\Phi}). \tag{4}$$

*For $p = k + 1$, the solution is $r = \sqrt{c\Phi}$. The eigenvector solution to $H_{\zeta_p}\boldsymbol{x}_p \approx 0$ has a non-trivial alignment to the community label vector and $\mathbb{E}[x_{p,i}] = v_{p,i}$, where $C\Pi\boldsymbol{v}_p = s_p^{\downarrow}(C\Pi)\boldsymbol{v}_p$.*

This result implies that, for $r = \zeta_p$, the $p$-th smallest eigenvalue of $H_{\zeta_p}$ is close to zero and the corresponding eigenvector $\boldsymbol{x}_p$ is a noisy version of the corresponding eigenvector $\boldsymbol{v}_p \in \mathbb{R}^k$ related to the $p$-th largest eigenvalue of $C\Pi$. Importantly, $\mathbb{E}[x_{p,i}]$ does not depend on $d_i$, hence it is suited to reconstruct communities regardless of the degree distribution. Since the eigenvectors of $C\Pi$ constitute a sufficient basis to identify the classes, the vectors $\boldsymbol{x}_p$, for $1 \leq p \leq k$, can be exploited to recover the community labels by stacking them in the columns of a matrix $X \in \mathbb{R}^{n\times k}$ and running the *k-means* algorithm on the rows of $X$.

**Remark 1.** *Having eigenvectors whose entries are, in expectation, independent of the degree distribution is of fundamental importance in the* k-means *step. If this were not the case, then the class information would be affected by the uninformative degree distribution, compromising the performance of the algorithm, as shown in [19].*

### 3.2. Improved regularization for the regularized random walk Laplacian $L_\tau^{\text{rw}}$

We here work on the strong connection between the Bethe-Hessian matrix for $r = \zeta_p$ and the regularized random walk Laplacian. The

following equivalent identities indeed hold:

$$[(\zeta_p^2-1)I_n + D - \zeta_p A]\boldsymbol{x}_p \approx 0$$
$$D_{\zeta_p^2-1}^{-1} A \boldsymbol{x}_p \approx \frac{1}{\zeta_p}\boldsymbol{x}_p,$$

where $D_{\zeta_p^2-1} = D + (\zeta_p^2-1)I_n$. This notably suggests that, for $\tau = \zeta_p^2 - 1$, the matrix $L_\tau^{\rm rw}$ has an eigenvector whose entries are not affected by the degree distribution, but depend only on the class labels, as depicted in Figure 2. With Figure 2, consistently with Remark 1, we further underline that for the *k-means* step it is fundamental to obtain two well separated density clouds in the $k$-dimensional space spanned by the rows of $X$, instead of a continuum of points, as evidenced by the histograms. Since there is a unique value of $r$ that allows the matrix $H_r$ to have a "clean" eigenvector [19], also the choice of $\tau$ is unique, as depicted in Figure 3. An "informative" eigenvector however does not imply that such eigenvector corresponds to a dominant isolated eigenvalue. This is of fundamental importance because, if the informative eigenvector corresponds to a non-isolated eigenvalue, then i) it is algorithmically challenging to locate it and ii) the information is likely to spread out on the neighboring eigenvectors. This is however not the case thanks to the following two propositions:

**Proposition 1.** *Consider the graph $\mathcal{G}(\mathcal{V},\mathcal{E})$ built on a sparse DC-SBM as per Equation* (1) *with $k$ communities. Further let $\zeta_p$, be defined as in Equation* (4)*, satisfying $\zeta_p \le \sqrt{c\Phi}$.*

*Then, for all large $n$ with high probability, $\zeta_p^{-1}$ is the $p$-th largest eigenvalue of $L_{\zeta_p^2-1}^{\rm rw}$ and it is isolated.*

**Proposition 2.** *The $p$ largest eigenvalues of $L_\tau^{\rm rw}$ are isolated, for $\zeta_p^2 - 1 \le \tau \le c\Phi - 1$ with high probability for all large $n$.*

Proposition 1 guarantees that, for the proposed parametrization, the informative eigenvector is isolated and can be found in the $p$-th largest position. Thanks to Proposition 2 instead, we know that for $\tau = c\Phi - 1 \ge \zeta_k^2 - 1$, all the $k$ informative eigenvalues will be isolated.

We now give a sketch of proof of Proposition 1 and 2.

*Sketch of Proof of Proposition 1.* Consider the eigenvector equation of $L_\tau^{\rm rw}$, for $0 \le \tau \le c\Phi - 1$,

$$L_\tau^{\rm rw}\boldsymbol{x}_p = s_p(L_\tau^{\rm rw})\boldsymbol{x}_p.$$

We define $r$ by $\tau = r^2 - 1$, so that $1 \le r \le \sqrt{c\Phi}$. The earlier equation can be rewritten in the following form:

$$rL_{r^2-1}^{\rm rw}\boldsymbol{x}_p = s_p(rL_{r^2-1}^{\rm rw})\boldsymbol{x}_p.$$

Define $\bar r_p$ such that $s_p(\bar r_p L_{\bar r_p^2-1}) = 1$ (we assume the existence of such an $\bar r_p$). Then, necessarily,

$$H_{\bar r_p}\boldsymbol{x}_p = 0.$$

From the properties of the Bethe-Hessian matrix, $\bar r_p$ can assume only $k+1$ discrete values: $\bar r_p \approx \zeta_p$ for $1 \le p \le k$ and $\bar r_p \approx \sqrt{c\Phi}$ for $p = k+1$. Letting $\epsilon \to 0$, then $s_p\left((\bar r_p+\epsilon)L_{(\bar r_p+\epsilon)^2-1}\right) \neq 1$, $\forall\, 1 \le p \le n$, meaning that there is no other eigenvalue in the neighborhood of $s_p(\bar r_p L_{\bar r_p^2-1}^{\rm rw})$ and hence we conclude it is isolated. As a consequence, the eigenvalues belonging to the *bulk* of $rL_{r^2-1}^{\rm rw}$ are necessarily smaller than 1 in modulus.

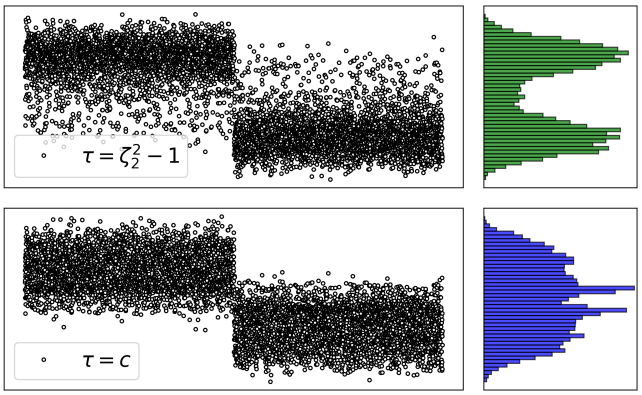


**Fig. 2**. Eigenvector of second largest eigenvalue of $L_{\zeta_2^2-1}^{\rm rw}$ (top) and $L_c^{\rm rw}$ (bottom) with histogram of the densities of the entries of the eigenvector (right). For this simulation, $n = 15000$, $k = 2$, $\theta_i \sim [\mathcal{U}(3,15)]^5$, $c_{\rm out} = 3$, $c_{\rm in} = 17$. Only nodes $i$ with $d_i > 0$ are considered.

Intuitively, looking at the symmetry of the spectrum of $L_\tau^{\rm rw}$ (Figure 1), the isolated eigenvalues are in largest positions. Formally, exploiting the Courant-Fischer theorem one can prove that 1 is indeed the $p$-th largest eigenvalue of $\zeta_p L_{\zeta_p^2-1}^{\rm rw}$. We can write:

$$s_q^{\downarrow}(rL_{r^2-1}^{\rm rw}) < 1, \quad \text{for } q \ge k+1, \quad r < \sqrt{c\Phi} \tag{5a}$$
$$s_p^{\downarrow}(\zeta_p L_{\zeta_p^2-1}^{\rm rw}) = 1, \quad \text{for } p \le k. \tag{5b}$$

Equation (5b) states that the informative eigenvector has an eigenvalue equal to 1, while Equation (5a) imposes that all the uninformative eigenvalues belonging to the bulk are smaller than 1, hence of the informative eigenvalue, so the result. □

*Sketch of Proof of Proposition 2.* This proposition is a direct consequence of $s_p^{\uparrow}(D - rA)$ being isolated for $\zeta_p \le r \le \sqrt{c\Phi}$ [19].

Define $\tilde r$ such that, $s_p^{\uparrow}(D - rA) = s_{p+1}^{\uparrow}(D - \tilde r A) = -\tau$. By construction, the matrix $L_\tau^{\rm rw}$ thus has two eigenvalues that are equal to $r^{-1}$ and $\tilde r^{-1}$. As a consequence of $s_p^{\uparrow}(D - rA)$ being isolated, $r^{-1}$ is away from $\tilde r^{-1}$.

The change of variable $\tau = -s_p^{\uparrow}(D - rA)$ provides a one-to-one mapping between the smallest isolated eigenvalues of $H_r$ and the largest isolated eigenvalues of $L_\tau^{\rm rw}$. It follows that, for $\zeta_p^2 - 1 \le \tau \le c\Phi - 1$, the top $p$ eigenvalues of $L_\tau^{\rm rw}$ are isolated.

□

### 3.3. Comments on the result and algorithm.

Figure 3 compares the performance of reconstruction in terms of the overlap

$$Ov = \left(\frac{1}{n}\sum_i \delta(\hat\ell_i, \ell_i) - \frac{1}{k}\right)\frac{1}{1-\frac{1}{k}} \tag{6}$$

and evidences that i) small regularizations produce better node partitions, ii) the proposed, $\alpha$-dependent, regularization surpasses all fixed values of $\tau$, iii) for $\tau < \zeta_2^2 - 1$ good partitions are achieved on easy problems, but the information does not correspond to isolated eigenvectors for hard detection problems, and iv) the performance using $H_{\zeta_2}$ and $L_{\zeta_2^2-1}^{\rm rw}$ are the same, since they are using the same informative eigenvectors.

We next list important messages of Proposition 1 and 2.

1. $L_\tau^{\rm sym}$ *vs.* $L_\tau^{\rm rw}$ : as opposed to [8], we studied the matrix $L_\tau^{\rm rw}$ instead of $L_\tau^{\rm sym}$. These two matrices have the same eigenvalues, but not the same eigenvectors. Our line of argument suggests – in good agreement with the observation of [3] – that it is more convenient to use the eigenvectors of the matrix $L_\tau^{\rm rw}$ in order to obtain eigenvectors whose entries are not affected by the degree distribution.
2. *The value of* $\tau$ *at the transition*: consider $k = 2$, then $\zeta_2 = (c_{\rm in}+c_{\rm out})/(c_{\rm in}-c_{\rm out})$. When $\alpha = \alpha_c$ (Equation (2)), $\zeta_2 = \sqrt{c\Phi}$ and therefore $\tau = c\Phi - 1 \approx c$. This observation allows us to understand why, in practice, the regularization $\tau = c$ appears to be a good choice. When $\tau = c\Phi - 1$ then certainly – regardless the hardness of the problem, as long as $\alpha > \alpha_c$ – the second largest eigenvalue of $L_\tau^{\rm rw}$ is isolated. As $c$ is in the order of magnitude of $c\Phi - 1$ in sparse graphs, $\tau = c$ will lead – in most cases – to a "satisfying" spectrum, in the sense that the informative eigenvalues are isolated. When $k > 2$, $\zeta_p = \sqrt{c\Phi}$ represents the transition from easy to hard detection and the argument can be generalized.
3. *The regularization is a function of the hardness of the detection problem*: once again consider $k = 2$. For easy detection problems ($c_{\rm out} \to 0$), we have $\zeta_2 = (c_{\rm in} + c_{\rm out})/(c_{\rm in} - c_{\rm out}) \to 1$, while, as already mentioned in point 2, $\zeta_2$ increases up to $\sqrt{c\Phi}$ in harder scenarios. This implies that harder problems need a larger regularization. Note also that, in the trivial case for which $\zeta_p \to 1$, $\forall\ p$, (when we have $k$ nearly disconnected clusters), the Bethe-Hessian falls into the combinatorial graph Laplacian $\lim_{r\to 1} H_r = D - A$ and the regularized random walk Laplacian into its non-regularized counterpart $\lim_{\tau\to 0} L_\tau^{\rm rw} = L^{\rm rw}$.
4. *Estimating the values of* $\zeta_p$: thanks to Equation (4) the values of $\zeta_p$ can be obtained by searching for the solution to $s_p^{\uparrow}(H_r) = 0$ on $r \in (1, \sqrt{c\Phi})$.
5. *Estimating the number of classes*: it was shown in [18] that all and only the informative eigenvalues of $H_r$ at $r = \sqrt{c\Phi}$ are negative, allowing an unsupervised method to estimate $k$. From this result, also $L_\tau^{\rm rw}$ allows to estimate $k$ as follows:

$$\hat{k} = \left| \{ i : s_i(L^{\rm rw}_{c\Phi-1}) > \frac{1}{\sqrt{c\Phi}} \} \right| \qquad (7)$$

6. *Disassortative networks*: even though we assumed for simplicity all the eigenvalues of the matrix $C\Pi$ to be positive – hence that there is a larger probability to get connected to nodes in the same community (*assortativity*) – the above results can be easily generalized to the case in which $C\Pi$ has negative eigenvalues and so in which there are *disassortative* communities.

The results of Section 3.2 naturally unfold in Algorithm 1 for community detection in sparse graphs.

**Algorithm 1** Community Detection with the regularized Laplacian

**Input** : adjacency matrix of undirected graph $\mathcal{G}$
Estimate $k$ : $\hat{k} \leftarrow \left| \left\{ i,\ s_i(D^{-1}_{c\Phi-1}A) > \frac{1}{\sqrt{c\Phi}} \right\} \right|$.
**for** $1 \le p \le \hat{k}$ **do**
  $\zeta_p \leftarrow \arg\underset{r}{} \left[ s_p^{\uparrow}(H_r) = 0 \right]$
  $X_p \leftarrow \boldsymbol{x}_p : \zeta_p D^{-1}_{\zeta_p^2-1} A \boldsymbol{x}_p = \boldsymbol{x}_p$
Estimate community labels $\hat{\boldsymbol{\ell}}$ as output of $\hat{k}$-class *k-means* on the rows of $X = [X_2, \ldots, X_{\hat{k}}]$.
**return** Estimated number $\hat{k}$ of communities and label vector $\hat{\boldsymbol{\ell}}$.

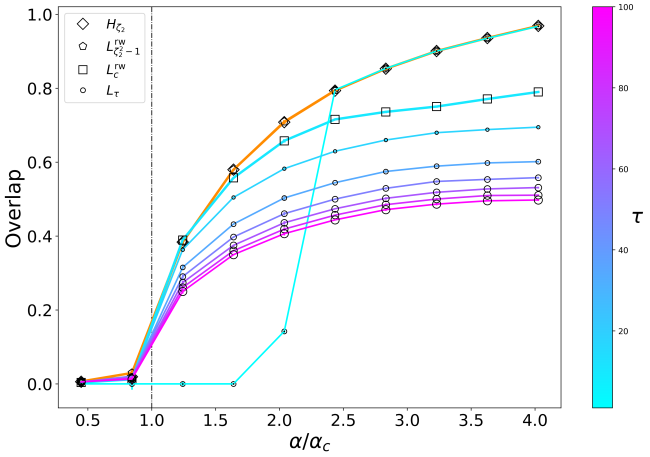


**Fig. 3**. Overlap comparison using the eigenvector with second largest eigenvalue of the matrix $L_\tau$ for different values of $\tau$. For these simulations $n = 50.000$ separated in two equal-size classes, $c_{\rm in} = 11 \to 19$, $c_{\rm out} = 9 \to 1$ with $c = 10$ constant, $\theta_i \sim [\mathcal{U}(3, 15)]^5$. Averages are taken over 5 realizations. The color code indicates the values of $\tau = 1 \to c^2$ also encoded by the size of the dots (small dots are for small $\tau$). The orange line with pentagons is obtained for $L^{\rm rw}_{\zeta_2^2-1}$, the line with squares is obtained for $L_c^{\rm rw}$. The orange line with diamonds is for $H_{\zeta_2}$.

## 4. CONCLUSION

In this article we discussed the regularization parameter $\tau$ of the matrix $L_\tau^{\rm rw} = (D + \tau I_n)^{-1}A$, used to reconstruct communities in sparse graphs. We explained why small and, most importantly, difficulty-adapted regularizations perform better than large (and difficulty-agnostic) ones.

Our findings notably shed light on the connection between the two benchmark approaches to community detection in sparse networks, provided for one by the statistics community and for the other by the physics community; these approaches have so far have been treated independently. We strongly suggest that bridging both sets of results has the capability to improve state-of-the-art knowledge of machine learning algorithms in sparse conditions (for which a direct application of standard algorithms is often inappropriate). Similar outcomes could arise for instance in KNN-based kernel learning or for any algorithm involving numerous data which, for computational reasons, imposes a sparsification of the information matrices.

In this view, we notably intend to generalize the algorithm proposed in this article (i) to richer graph and data clustering problems, (ii) to the often more realistic semi-supervised setting (where part of the nodes are known), while in passing (iii) enriching our understanding on existing algorithms.

## 5. ACKNOWLEDGEMENTS

Couillet's work is supported by the IDEX GSTATS DataScience Chair and the MIAI LargeDATA Chair at University Grenoble Alpes. Tremblay's work is supported by CNRS PEPS I3A (Project RW4SPEC).

## 6. REFERENCES


[1] Santo Fortunato, "Community detection in graphs," *Physics reports*, vol. 486, no. 3-5, pp. 75–174, 2010.

[2] Albert-László Barabási et al., *Network science*, Cambridge university press, 2016.

[3] Ulrike Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.

[4] Jianbo Shi and Jitendra Malik, "Normalized cuts and image segmentation," *Departmental Papers (CIS)*, p. 107, 2000.

[5] Andrew Y Ng, Michael I Jordan, and Yair Weiss, "On spectral clustering: Analysis and an algorithm," in *Advances in neural information processing systems*, 2002, pp. 849–856.

[6] M. E. J. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, vol. 103, pp. 8577–8582, 2006.

[7] Jiashun Jin et al., "Fast community detection by score," *The Annals of Statistics*, vol. 43, no. 1, pp. 57–89, 2015.

[8] Tai Qin and Karl Rohe, "Regularized spectral clustering under the degree-corrected stochastic blockmodel," in *Advances in Neural Information Processing Systems*, 2013, pp. 3120–3128.

[9] Arash A Amini, Aiyou Chen, Peter J Bickel, Elizaveta Levina, et al., "Pseudo-likelihood methods for community detection in large sparse networks," *The Annals of Statistics*, vol. 41, no. 4, pp. 2097–2122, 2013.

[10] Antony Joseph and Bin Yu, "Impact of regularization on spectral clustering," *arXiv preprint arXiv:1312.1733*, 2013.

[11] J. Lei and A. Rinaldo, "Consistency of spectral clustering in stochastic block models," *The Annals of Statistics*, vol. 43, no. 1, pp. 215–237, 2015.

[12] Can M Le, Elizaveta Levina, and Roman Vershynin, "Concentration of random graphs and application to community detection," *arXiv preprint arXiv:1801.08724*, 2018.

[13] Brian Karrer and Mark EJ Newman, "Stochastic blockmodels and community structure in networks," *Physical review E*, vol. 83, no. 1, pp. 016107, 2011.

[14] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová, "Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications," *Physical Review E*, vol. 84, no. 6, pp. 066106, 2011.

[15] Elchanan Mossel, Joe Neeman, and Allan Sly, "Stochastic block models and reconstruction," *arXiv preprint arXiv:1202.1499*, 2012.

[16] Laurent Massoulié, "Community detection thresholds and the weak ramanujan property," in *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*. ACM, 2014, pp. 694–703.

[17] Lennart Gulikers, Marc Lelarge, Laurent Massoulié, et al., "An impossibility result for reconstruction in the degree-corrected stochastic block model," *The Annals of Applied Probability*, vol. 28, no. 5, pp. 3002–3027, 2018.

[18] Alaa Saade, Florent Krzakala, and Lenka Zdeborová, "Spectral clustering of graphs with the bethe hessian," in *Advances in Neural Information Processing Systems*, 2014, pp. 406–414.

[19] Lorenzo Dall'Amico, Romain Couillet, and Nicolas Tremblay, "Revisiting the bethe-hessian: improved community detection in sparse heterogeneous graphs," in *Advances in Neural Information Processing Systems*, 2019, pp. 4039–4049.

[20] Florent Krzakala, Cristopher Moore, Elchanan Mossel, Joe Neeman, Allan Sly, Lenka Zdeborová, and Pan Zhang, "Spectral redemption in clustering sparse networks," *Proceedings of the National Academy of Sciences*, vol. 110, no. 52, pp. 20935–20940, 2013.

[21] Charles Bordenave, Marc Lelarge, and Laurent Massoulié, "Non-backtracking spectrum of random graphs: community detection and non-regular ramanujan graphs," in *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*. IEEE, 2015, pp. 1347–1357.

[22] Lenka Zdeborová and Florent Krzakala, "Statistical physics of inference: Thresholds and algorithms," *Advances in Physics*, vol. 65, no. 5, pp. 453–552, 2016.

[23] Lorenzo DallAmico and Romain Couillet, "Community detection in sparse realistic graphs: Improving the bethe hessian," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 2942–2946.