COMMUNITY DETECTION IN SPARSE REALISTIC GRAPHS: IMPROVING THE BETHE HESSIAN

Lorenzo Dall'Amico[1], *Romain Couillet*[1,2]

[1]GIPSA-lab, University of Grenoble–Alpes, [2]CentraleSupélec, University of Paris–Saclay

ABSTRACT

This article improves over the recently proposed Bethe Hessian matrix for community detection on sparse graphs, assuming here a more realistic setting where node degrees are inhomogeneous. We notably show that the parametrization proposed in the seminal work on the Bethe Hessian clustering can be ameliorated with positive consequences on correct classification rates. Extensive simulations support our claims.

Index Terms— community detection; Bethe Hessian; spectral clustering; statistical physics.

1. INTRODUCTION

Community detection on graphs [1] is a cornerstone topic in machine learning, much related to unsupervised classification (or clustering) [2], and consists in grouping nodes of strong affinity in distinct classes. Theoretically speaking, given a statistical generative model for a graph $\mathcal{G}$ with classes, the first question to consider is their detectability and the capability to associate each node to its genuine class.

The most popular and versatile approach to perform community detection on graphs is the belief propagation algorithm; however, the latter is computationally expensive, offers no convergence guarantee and is theoretically hard to analyze. Most convincing (since well performing, theoretically analyzable and computationally appealing) among the proposed alternative approaches to community retrieval are *spectral methods* that consist in reading the community classes directly off the dominant eigenvectors of a matrix representation of $\mathcal{G}$, thereby reminiscent of spectral clustering [2]. Assuming a two-class *stochastic block model* (SBM) for the generative graph model with n nodes – where the probability for node i to connect to node j equals $p_{\rm in} \in [0,1]$ if they belong to the same class or $p_{\rm out} \in [0,1]$ otherwise, and every edge is drawn independently – a natural spectral community detection method consists in extracting the class information from the dominant eigenvectors of the adjacency matrix $A \in \{0,1\}^{n\times n}$, where $A_{ij} = 1$ if nodes i and j are connected, and $A_{ij} = 0$ otherwise.

Couillet's work is supported by the GSTATS UGA IDEX DataScience chair and the ANR RMT4GRAPH Project (ANR-14-CE28-0006).

It was indeed shown that, as $n \to \infty$ and $p_{\rm in}, p_{\rm out}$ are independent of n, which is referred to as a *dense graph* community detection problem, spectral clustering on A is "optimal" in the sense that:

(a) there exists a minimal value for $(p_{\rm in}-p_{\rm out})/\sqrt{p_{\rm in}+p_{\rm out}}$ below which community detection is infeasible;

(b) spectral clustering on A returns non-trivial classification (that is on average better than random guess) as soon as this threshold is exceeded.

In statistical physics terms, this asymptotic decidability thresholding effect is referred to as a *phase transition phenomenon*.

Yet, the conditions under which spectral clustering on A is optimal rely on two key ingredients:

(i) the statistical block model for $\mathcal{G}$ is quite elementary;

(ii) the graph is dense (the node degrees scale with n).

Both conditions are deemed *unrealistic* as not representative of real world graphs. To address issue (i), a line of works was developed [3, 4] in a K-class *degree corrected* stochastic block model (DC-SBM), where

$$P(A_{ij} = 1) = q_i q_j C(x_i, x_j)$$

with $q_i > 0$ some *intrinsic* connectivity amplitude for node i, $x_i \in \{1,\dots,K\}$ the label of the class of node i, and $C(x_i, x_j)$ some class-wise affinity parameter. In [4], it is shown that spectral clustering on A is no longer optimal in that the phase transition phenomenon in general arises well below the detectability power of A; an improvement is then proposed in [4] which shows that there exists $\alpha > 0$ depending on the law of the q_i's such that performing spectral clustering on $D^{-\alpha}AD^{-\alpha}$ rather than A drastically pushes the phase transition to smaller discriminative values of $C(a,b)$. Yet, to date, no theoretical optimal phase transition in the dense (as well as sparse) DC-SBM setting is known.

Addressing limitation (ii) is theoretically much harder. Assuming that the probability for $A_{ij} = 1$ scales like $1/n$, i.e., the average nodal degree is of order $O(1)$ with respect to n, it has long been seen in simulations that spectral clustering on A is largely suboptimal and, here again, even under the SBM setting, the optimal phase transition position is unknown. Yet, via the impulse of *statistical physics* tricks,

mostly consisting in either approximating (linearizing) belief propagation or mapping the community detection problem into an Ising model analog, new spectral clustering algorithms were proposed that are shown in practice (and sometimes in theory [5]) to dramatically improve over spectral clustering on A; this is notably the case of spectral clustering on the *non-backtracking operator* B [6] and on the (closely related) *Bethe Hessian* [7].

The work we propose in this article precisely revolves around the Bethe Hessian matrix that we aim at *analyzing and improving* in a *sparse DC-SBM* setting. The Bethe Hessian matrix H_r is defined as

$$H_r = (r^2 - 1)I_n + D - rA \tag{1}$$

for a certain parameter r, with $D = \mathrm{diag}(d_1, \ldots, d_n)$ ($d_i = [A1_n]_i$) the degree matrix. Assuming a two-class SBM model with label $x_i \in \{-1, 1\}$ for node i, H_r corresponds to the Hessian matrix of the Bethe free energy of an Ising model on the graph, the Bethe approximation consisting of using belief propagation to express the probability distribution of the system and r relates to the physical *temperature* in the model (see details in Section 2).

In [7], the authors propose various intuitive arguments for setting r to some presumably optimal value for spectral clustering on H_r. Precisely, they advise $r = \sqrt{\rho(B)}$ (i.e., the spectral radius of the non-backtracking operator B from [6]) that in a homogeneous case reduces to:

$$r = \sqrt{c}, \quad c \equiv \frac{1}{2}(c_{\mathrm{in}} + c_{\mathrm{out}}) \tag{2}$$

where $c_{\mathrm{in}} \equiv np_{\mathrm{in}}$ and $c_{\mathrm{out}} \equiv np_{\mathrm{out}}$ are constant in the sparse model. Besides, it is claimed in [7] that the optimal phase transition for H_r appears as soon as

$$c_{\mathrm{in}} - c_{\mathrm{out}} > 2\sqrt{c} \tag{3}$$

which is consistent with the *proved* phase transition of the non-backtracking operator [5].

In [4], it is however shown in simulations that H_r (for the above choice of r) often performs far from optimaly in the DC-SBM model and for realistic graphs, even for rather sparse scenarios, where spectral clustering on the (supposedly suboptimal) matrices $D^{-\alpha}AD^{-\alpha}$ (even for $\alpha = 0$) often outperforms H_r. This is mostly due to the fact that the dominant eigenvectors of H_r are strongly affected by the varying degrees. This suggests that either a post-processing on the eigenvectors of H_r is needed or that, in the first place, the SBM-adapted Ising model should be revised to accommodate for the more diverse degrees observed in real graphs.

In this article, rather than updating the Ising model, we make the following observation: there exists a value $r = r_c$ different from that recommended in (2) for which the Bethe Hessian is *insensitive to the node heterogeneity*, precisely

$$r_c = \frac{c_{\mathrm{in}} + c_{\mathrm{out}}}{c_{\mathrm{in}} - c_{\mathrm{out}}}$$

(and $r_c = \sqrt{c}$ right at the phase transition $c_{\mathrm{in}} - c_{\mathrm{out}} = 2\sqrt{c}$).

Thus, our main finding is that the Bethe Hessian itself, although developed under an SBM assumption, with properly set temperature, is *resilient* to degree heterogeneity and thus likely well-performing on realistic graphs, at least for $r = r_c$.

2. MODEL AND MAIN RESULTS

2.1. Preliminaries

We consider a 2-class symmetric n-node graph $\mathcal{G}$ generated from a sparse DC-SBM model, that is with adjacency matrix $A \in \{0, 1\}^{n \times n}$ such that

$$P(A_{ij} = 1) = q_i q_j \frac{C(x_i, x_j)}{n} \tag{4}$$

for $q_1, \ldots, q_n > 0$ random and independently drawn with $\mathrm{E}[q_i] = 1$, $x_i \in \{-1, 1\}$ the class label of node i, and $C(x_i, x_j) = c_{\mathrm{in}} > 0$ if $x_i = x_j$ or $C(x_i, x_j) = c_{\mathrm{out}} < c_{\mathrm{in}}$ if $x_i \neq x_j$. For a given random realization of A, we then define the Bethe Hessian matrix H_r as per (1).

The authors in [7] claim that, in the class detectability regime (3), the eigenvector corresponding to the second smallest eigenvalue of H_r is very much aligned to the vector $x \in \{-1, 1\}^n$ containing the labels x_i, and therefore contains the structural information about the classes that can be retrieved by spectral clustering.

Selecting the optimal value of r is however a delicate matter. Indeed, H_r derives from the analog Ising model of interaction between the "spins" x_i defined from the dimensionless Hamiltonian with temperature-related parameter r:

$$\beta E(\{x\}; r) = - \sum_{\substack{i,j \in \{1,\ldots,n\} \\ A_{ij}=1}} \mathrm{atanh}(r^{-1}) x_i x_j$$

At high temperature ($r \to \infty$), the free energy is dominated by the entropy contribution and the x_i's become independent, thereby not raising any clustering. On the opposite, at low temperature ($r \downarrow 1$), spontaneous magnetization occurs, leading to a favorite orientation of the spins. In [7], the authors argue that, under an SBM setting, the optimal value of r should be found over the range

$$\sqrt{c} \leq r \leq c \tag{5}$$

since, for r *somewhat larger* to the right-edge, the eigenvalues of H_r no longer isolate. On this range lies $r = \sqrt{\rho(B)}$, which perfectly maps the Bethe Hessian to the related non-backtracking operator B and is thus claimed in [7] to be the

likely optimal parameter for H_r. As previously pointed out, for the very specific case of the SBM setting, this value is found to be $r = \sqrt{c}$.

We argue in the following, and support in simulations, that an appropriate choice of r, better adapted to the DC-SBM model, can be found outside the range (5).

2.2. Main Result

Let us start by considering $H_r x$:

$$(H_r x)_i = (r^2 - 1)x_i + d_i x_i - r \sum_{k \in \mathcal{N}(i)} x_k$$

where $k \in \mathcal{N}(i) \Leftrightarrow A_{ik} = 1$ (that is $\mathcal{N}(i)$ is the set of neighbors of i). Denoting $\partial_i^S \equiv \{j,\ A_{ij} = 1 \text{ and } x_i = x_j\}$ the set of neighbors of node i belonging to the *same* class and, similarly, $\partial_i^O \equiv \{j,\ A_{ij} = 1 \text{ and } x_i \neq x_j\}$ the set of neighbors of i from the *opposite* class, this simply reads

$$(H_r x)_i = x_i \left[(r^2 - 1) + d_i - r \left(|\partial_i^S| - |\partial_i^O| \right) \right].$$

We now make the assumption (or rather the heuristic approximation) that, although the average node degree is of order $O(1)$ with respect to n, one can approximately claim that

$$\frac{|\partial_i^S|}{d_i} \simeq \frac{c_{\text{in}}}{c_{\text{in}} + c_{\text{out}}}, \quad \frac{|\partial_i^O|}{d_i} \simeq \frac{c_{\text{out}}}{c_{\text{in}} + c_{\text{out}}}$$

at least for those nodes i having many neighbors (close to the decidability threshold (3), the approximation is mostly adequate to scenarios where c_{out} is rather large). Then,

$$(H_r x)_i \simeq x_i \left[(r^2 - 1) + d_i \left(1 - r \frac{c_{\text{in}} - c_{\text{out}}}{c_{\text{in}} + c_{\text{out}}} \right) \right] \tag{6}$$

where, under the DC-SBM model (4), the d_i's may in general be quite different. Thus, in order to retrieve an approximate eigenvector equation for x, one must set

$$r \equiv r_c = \frac{c_{\text{in}} + c_{\text{out}}}{c_{\text{in}} - c_{\text{out}}} \tag{7}$$

in which case $H_{r_c} x \simeq (r_c^2 - 1)x$. As such, for $r = r_c$, one expects to see one dominant eigenvector of H_{r_c} *not tainted by the degrees* d_i, as opposed to what was observed in [4] for $r = \sqrt{c}$ or $r = \rho(B)$.

Remark 1 (Homogeneous case)**.** *Note that in the homogeneous case where the q_i's are all equal and thus the d_i's are expected to be approximately the same,* (6) *is an approximate eigenvector equation for all r's. And thus r_c is not a particularly preferred candidate.*

It is now interesting to see that, as pointed out previously, $r = r_c$ necessarily falls *away* from the conservative range (5) of utmost interest to the authors in [7]. Indeed, for c_{in} and c_{out} compliant with (3), we find

$$r_c \leq \sqrt{c}$$

with equality right *at* the transition of (3).

This said, it is still important to note that the authors in [7] have identified (mostly through simulations) the eigenvector carrying the class information as the one associated to the second smallest eigenvalue of H_r for all positive r's inducing an asymptotic phase transition. This observation seems also to hold in the DC-SBM case.

As such, our final claim may then be formulated as:

Claim 1 (Spectral Clustering on H_{r_c})**.** *Assume a sparse DC-SBM model for a graph $\mathcal{G}$. Then, community detection on $\mathcal{G}$ is efficiently performed, irrespective of the heterogeneity of the degrees, by performing spectral clustering on the eigenvector attached to the second smallest eigenvalue of H_{r_c} with r_c given in Equation* (7)*.*

2.3. Estimation of r_c

A subsequent difficulty for practical application is that $c_{\text{in}} - c_{\text{out}}$, and thus r_c, is not directly accessible. Several solutions here exist to retrieve a good approximation for r_c. One may for instance iteratively perform spectral clustering on H_r starting with, say, $r = \sqrt{\rho(B)}$ that can be estimated as $\frac{\sum_i d_i^2}{\sum_i d_i} - 1$ (according to [7]), obtain a first estimate of the class components, from which c_{in} and c_{out} are further estimated, and so on. Another initialization option follows from

$$(D^{-1}Ax)_i = \sum_{k \in \mathcal{N}(i)} \frac{x_k}{d_i} = \frac{|\partial_i^S| - |\partial_i^O|}{d_i} x_i \simeq \frac{c_{\text{in}} - c_{\text{out}}}{c_{\text{in}} + c_{\text{out}}} x_i. \tag{8}$$

As such, r_c can be retrieved, with the same approximation made above on $H_{r_c} x$, as a corresponding isolated (inverse) eigenvalue of $D^{-1}A$.[1]

3. NUMERICAL RESULTS

This section provides numerical support for our claimed results. We start first by considering synthetic DC-SBM graphs with various laws for the q_i's. For comparison fairness and adaptability to uneven class cardinalities, spectral clustering is systematically performed using the k-means algorithm rather than on a sign-based method (as opposed to [7])

We first focus on the case of two even size classes. While the coming observations have been verified to be equally valid

[1] This, in passing, raises the question as to why $D^{-1}A$ would not be an equally valid matrix for spectral clustering as H_{r_c}. The answer however lies in the badly understood reasons why spectral clustering on $D^{-1}A$ as well as A, $D - A$, $D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ does not perform well in the sparse case.

for various heterogeneous settings, we will here depict the most interesting and visible case where $q_i \in \{0.4, 1.6\}$ with $P(q_i = 0.4) = P(q_i = 1.6) = \frac{1}{2}$. In this case, both H_{r_c} and $H_{\sqrt{\rho(B)}}$ essentially have the same performance in terms of overlap (which measures the distance to random guess on a $[0, 1]$ scale), both overtaking that of $D^{-1}A$. However, a careful control of the second smallest eigenvectors of H_{r_c}, $H_{\sqrt{\rho(B)}}$ and second largest of $D^{-1}A$ (Figure 1) reveals that the second suffers from the presence of two distinct values for the q_i's by exhibiting four 'plateaus' rather than two. This is not the case of either H_{r_c} or $D^{-1}A$. Yet, when asked to retrieve exactly two classes, k-means usually performs a correct partitioning, hence the equal overlap performance. Drawing on this observation, Figure 2 compares the performance of the same three methods and for the same choice of q_i but now for two classes of uneven sizes $\frac{n}{3}$ and $\frac{2n}{3}$, respectively. In this more asymmetric situation, the performance of $H_{\sqrt{\rho(B)}}$ is strongly affected by the q_i's that k-means wrongly confuses for the genuine class divisions. Spectral clustering on H_r does not suffer this limitation.

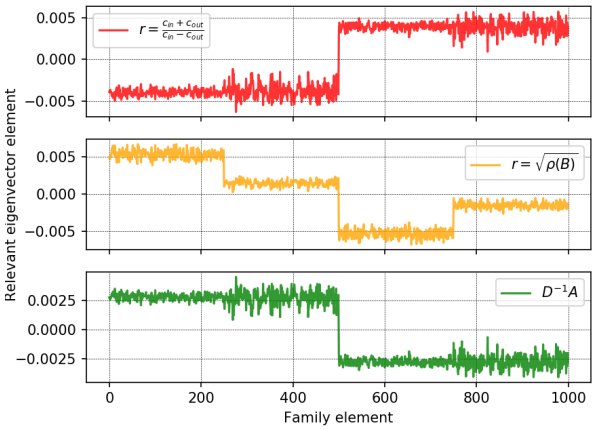

Fig. 1. Second dominant eigenvector of H_{r_c}, $H_{\sqrt{\rho(B)}}$, and $D^{-1}A$ for q_i distributed as $\frac{1}{2}\delta_{0.4} + \frac{1}{2}\delta_{1.6}$. In this case the q_i's, $i = 1, \ldots, n$, are sorted in four $\frac{n}{4}$-sized consecutive blocks as $(1.6, .4, 1.6, .4)$.

Table 1 provides a comparative overlap performance, on the same real graphs as in [7], of $H_{\sqrt{\rho(B)}}$, for the iterated method discussed in Subsection 2.3 with initialization at $r = \sqrt{\rho(B)}$ (indicated as $\sqrt{\rho(B)}^+$) or $r = 1/\lambda_2(D^{-1}A)$ (indicated as $\lambda_2^+(D^{-1}A)$), and for the oracle optimal $r = r_{\text{opt}} \in \mathbb{R}$. Consistently with our intuitive findings, it is observed that r_{opt} is systematically rather far from $\sqrt{\rho(B)}$ while $r = 1/\lambda_2(D^{-1}A)$ and further iterates are close. In terms of overlap, the proposed methods outperform the $H_{\sqrt{\rho(B)}}$ approach, although some overlaps remain much lower than optimal, despite the proximity of r to r_{opt}; this is likely due both to a finite-dimensional effect as well as to the specificities of the possibly far-from-DCSBM looking graphs.

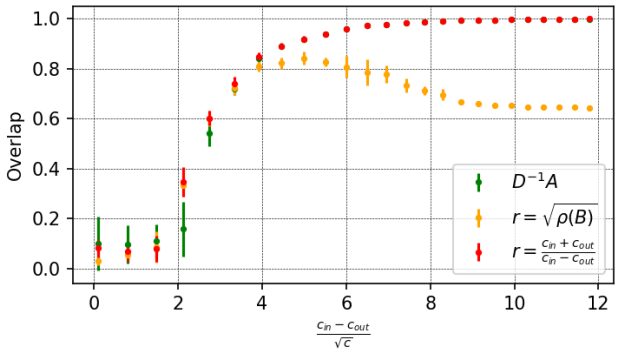

Fig. 2. Overlap performance of the three methods for an uneven population $|\mathcal{C}_1| = 2|\mathcal{C}_2|$ of classes $\mathcal{C}_1$ and $\mathcal{C}_2$, with q_i's distributed as $\frac{1}{2}\delta_{0.4} + \frac{1}{2}\delta_{1.6}$.

Graph / r	$\sqrt{\rho(B)}$	$\sqrt{\rho(B)}^+$	$\lambda_2^+(D^{-1}A)$	r_{opt}
Polblogs	0.32	0.59	0.03	0.90
	(*8.01*)	(*8.01*)	(*1.09*)	(*1.15*)
		(*1.94*)	(*1.00*)	
Karate	1	0.94	0.94	1
	(*1.78*)	(*2.14*)	(*1.15*)	(*1.78*)
		(1.19)	(1.19)	
Dolphins	0.93	0.97	0.97	0.97
	(*1.61*)	(*1.61*)	(*1.04*)	(*1.08*)
		(*0.97*)	(*1.08*)	

Table 1. Overlap performance on benchmark graphs and, in parentheses, starting and final values of r for the iterated estimates (r^+).

4. CONCLUDING REMARKS

This article proposes an improvement over the recently developed Bethe Hessian approach to community detection on sparse graphs. We showed that the proposed new parametrization, while performing similarly on homogeneous graphs, brings significant gains on more realistic heterogeneous graphs, as confirmed by simulations on real networks.

The cornerstone of our approach however lies in benefiting from a fortunate cancelling of the heterogeneity effect on the matrix second smallest eigenvector for a precise parameter setting. Estimating the latter satisfactorily , a point of crucial importance as demonstrated in our simulations, requires a more thorough analysis. A line of further improvement lies in updating the energy potential formulation from which the Bethe Hessian derives to already accommodate for graph heterogeneity, without resorting to a fine posterior parameter tuning. These are the objects of future investigations.

5. REFERENCES

[1] Mark EJ Newman, "Detecting community structure in networks," *The European Physical Journal B*, vol. 38, no. 2, pp. 321–330, 2004.

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

[3] Amin Coja-Oghlan and André Lanka, "Finding planted partitions in random graphs with general degree distributions," *SIAM Journal on Discrete Mathematics*, vol. 23, no. 4, pp. 1682–1714, 2009.

[4] Hafiz Tiomoko Ali and Romain Couillet, "Improved spectral community detection in large heterogeneous networks," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 8344–8392, 2017.

[5] 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.

[6] 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.

[7] 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.