A Unified Framework for
Spectral Clustering in Sparse Graphs

Lorenzo Dall'Amico[†]   LORENZO.DALL-AMICO@GIPSA-LAB.FR
Romain Couillet[†] [*]   ROMAIN.COUILLET@UNIV-GRENOBLE-ALPES.FR
Nicolas Tremblay[†]   NICOLAS.TREMBLAY@GIPSA-LAB.FR

[†] *GIPSA-lab Université Grenoble Alpes, CNRS, Grenoble INP*
11 Rue des Mathématiques, 38400 Saint-Martin-d'Hères, France

[*] *Laboratoire d'Informatique de Grenoble (LIG), Université Grenoble Alpes*
700 Av. Centrale, 38401 Saint-Martin-d'Hères, France

Editor: Tina Eliassi-Rad

Abstract

This article considers spectral community detection in the regime of sparse networks with heterogeneous degree distributions, for which we devise an algorithm to efficiently retrieve communities. Specifically, we demonstrate that a well parametrized form of *regularized Laplacian matrices* can be used to perform spectral clustering in sparse networks without suffering from its degree heterogeneity. Besides, we exhibit important connections between this proposed matrix and the now popular *non-backtracking* matrix, the *Bethe-Hessian* matrix, as well as the standard Laplacian matrix. Interestingly, as opposed to competitive methods, our proposed improved parametrization inherently accounts for the hardness of the classification problem. These findings are summarized under the form of an algorithm capable of both estimating the number of communities and achieving high-quality community reconstruction.

Keywords: community detection, sparsity, heterogeneous degree distribution, spectral clustering, unsupervised learning

1. Introduction

Graph theory has found many applications in a variety of domains that span from modern biology, to technology and social sciences (Barabási et al., 2016). Although the underlying represented systems may be fundamentally different, some common features emerge in complex networks. Some of these are (Newman, 2003):

1. Their *heterogeneous degree distribution*: the number of connections of each node (their *degree*) is far from following a homogeneous distribution and, instead, is typically a broad distribution (Barabási and Albert, 1999) in which few nodes (called *hubs*) have a very large degree while the majority has only few connections.

2. *Sparsity*: the average degree is typically much smaller than the size of the graph and a small fraction of all the possible connections are present (Barabási, 2013); consequently, the average degree can be considered as roughly independent of the size of the graph.

3. the *clustering effect*: as a consequence of the nodes affinity encoded by edges (Borgatti and Halgin, 2011), in real graphs typically emerge small groups of densely connected nodes, called *clusters* or *communities*.

The remainder of the article focuses on *community detection*, that is in determining a good community assignment of the nodes of a given graph. While point 3. of the list above is of central concern, the (nuisance) properties 1. and 2. need to be accounted for to guarantee high performance community assignments on real-world graphs.

1.1 Community Detection

One of the most natural tasks in graph theory is community detection, i.e., the identification of similarity groups on a given graph. Practically, for an unweighted and undirected graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ with $|\mathcal{V}| = n$ nodes and $|\mathcal{E}|$ edges, community detection consists in finding a non-overlapping partition of the nodes that identifies underlying communities in a completely unsupervised manner. There is no unique definition of a community, but a general criterion is to impose that nodes in the same community have more inter-connections than nodes in different communities, as a consequence of the stronger affinity among members of the same community (Fortunato, 2010).

A possible way of formalizing this intuition consists in defining the class labels as the solution of an optimization problem such as *MinCut*, *RatioCut*, and *NormalizedCut* (Von Luxburg, 2007). These definitions do not make any assumption on the underlying graph but only on what a satisfactory community assignment should be like. The aforementioned optimization problems are, however, NP-hard and only approximate solutions can be obtained. Another possible approach (adopted in the following) is instead to consider community detection as a statistical inference problem. The graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ is seen as a realization of a random process in which the class-label assignment is encoded by some hidden parameters of the generative model that have to be inferred. To simultaneously account for sparsity and heterogeneity that typically characterize real graphs, in this article we consider graphs generated according to a k-class sparse degree-corrected stochastic block model (DC-SBM) (Karrer and Newman, 2011), that is

$$\mathbb{P}(A_{ij} = 1 | \ell_i, \ell_j, \theta_i, \theta_j) = \theta_i \theta_j \frac{C_{\ell_i, \ell_j}}{n} \quad 1 \leq i < j \leq n \tag{1}$$

in which $\boldsymbol{\ell} = \{1, \cdots, k\}^n$ denotes the node-wise labelling vector ($\ell_i = p$ if node i is in class p), $\boldsymbol{\theta}$ is the vector of intrinsic connection "probabilities" which are used to produce an arbitrary degree distribution and are independent of the label vector $\boldsymbol{\ell}$. The entries of $\boldsymbol{\theta}$ are positive independent random variables $\theta_i \in [\theta_{\min}, \theta_{\max}]$, with $\theta_{\min} > 0$, having first moment $\mathbb{E}[\theta_i] = 1$ and second moment $\mathbb{E}[\theta_i^2] = \Phi$. We denote with $C \in \mathbb{R}^{k \times k}$ the symmetric class-affinity matrix, with entries independent of n. The term $1/n$ in (1) bounds the average degree $\bar{d} = \frac{1}{n} \mathbf{1}_n^T A \mathbf{1}_n = O_n(1)$ to an n-independent value, setting the problem in the sparse regime.

A major advantage of a well defined generative model results in statistically tractable conclusions, like the existence in asymptotically large graphs ($n \to \infty$) generated from the (degree-corrected) stochastic block model of a limiting *detectability threshold*. Specifically, for $k = 2$ classes, one can in general identify a parameter α, function of the community

statistics and such that, beyond a threshold ($\alpha > \alpha_c$), partial label reconstruction can be theoretically achieved, whereas below the threshold ($\alpha < \alpha_c$) no algorithm can perform better than random guess. For $k > 2$ classes, the situation is more involved: it is conjectured (Decelle et al., 2011) that there exists an *easy* detection zone in which a non-trivial clustering can be found in polynomial time, a *hard* detection zone in which non-trivial clustering can be found but only in exponential time and an *impossible* detection zone in which no algorithm can find a non-trivial partition.

In the very specific setting where $k = 2$ and both classes are of equal size ($n/2$), the detectability threshold assumes a simple expression. Letting $C_{\ell_i,\ell_j} = c_{\text{in}}$ if $\ell_i = \ell_j$ and c_{out} otherwise, in the *stochastic block model* (for which $\boldsymbol{\theta} = \mathbf{1}_n$) it was initially conjectured (Decelle et al., 2011) and later proved (Mossel et al., 2012; Massoulié, 2014) that the condition to non-trivial clustering is given by

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

where $c = (c_{\text{in}} + c_{\text{out}})/2$. Equation (2) was later generalized in (Gulikers et al., 2018, 2017a) to the DC-SBM but now with $\alpha_c = 2/\sqrt{\Phi}$ (we recall that $\Phi = \mathbb{E}[\theta_i^2]$).

If on one side random generative models like the DC-SBM give a solid theoretical understanding, the applicability to real graphs of a DC-SBM inspired clustering algorithm is a serious concern to be taken into account and which will be addressed here.

1.2 Spectral Clustering: Related Work

Regardless of the definition of a community considered (as a solution of an optimization problem or of an inference problem), different classes of algorithms can be considered to find exact or approximate solutions for community detection. Spectral clustering is a class of algorithms according to which each node of the graph is mapped to a low dimensional space. Spectral clustering hence provides a small dimensional graph embedding from which nodes are partitioned using common clustering algorithms, like *k-means*. The name *spectral* comes from the fact that the embedding is obtained using some eigenvectors (later referred to as *informative*) of a suitable graph representation matrix. Spectral clustering allows one to obtain an approximate fast solution of the optimal class assignment both in the case in which clusters are defined as the solution of an NP-hard optimization problem (Von Luxburg, 2007; Zhang and Rohe, 2018) or of an inference problem (Krzakala et al., 2013).

Algorithm 0 A typical spectral clustering algorithm on a graph $\mathcal{G}$ with k classes

---

1: **Input** : $\mathcal{G}$, k
2: Choose the matrix $M \in \mathbb{R}^{n \times n}$, a suited representation of the graph
3: Find the k largest (or smallest) eigenvalues of M and stack the corresponding eigenvectors in the columns of a matrix $\boldsymbol{X} \in \mathbb{R}^{n \times k}$
4: (*Optional*) Normalize the rows of $\boldsymbol{X}$
5: Estimate community labels as the output of some small dimensional clustering method (such as *k-means* or *expectation maximization*) performed on the points in $\mathbb{R}^k$ defined by the rows of $\boldsymbol{X}$
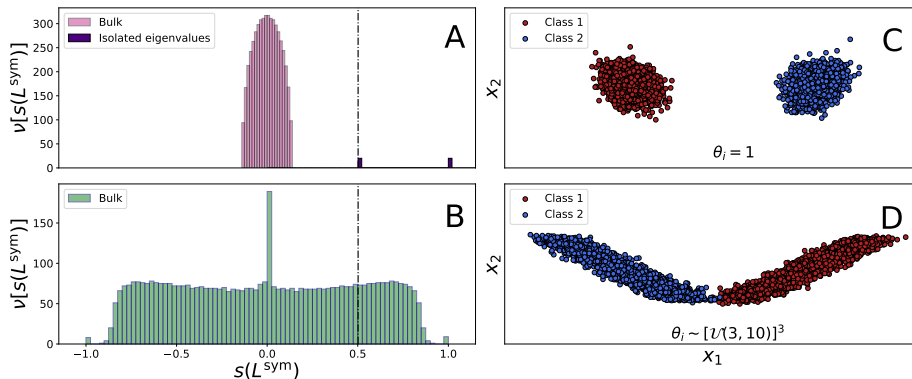6: **return** Estimated label community vector.

Figure 1: For all the four plots $n = 5000$, $k = 2$, $\boldsymbol{\theta} = \mathbf{1}_n$ if not otherwise specified. (Left) spectrum of the matrix L^{sym} in the: A) dense regime ($c_{\mathrm{in}}/n = 0.06$, $c_{\mathrm{out}}/n = 0.02$); B) sparse regime ($c_{\mathrm{in}} = 6$, $c_{\mathrm{out}} = 2$). (Right) first eigenvectors $\boldsymbol{x}_1$ vs second eigenvector $\boldsymbol{x}_2$ of L^{sym} in the dense regime for: C) $\boldsymbol{\theta} = \mathbf{1}_n$; D) $\theta_i \sim [\mathcal{U}(3, 10)]^3$.

Algorithm 0 provides the general structure of a k-class spectral clustering algorithm based on an "affinity" matrix M. Algorithm 0 relies on the fact that some of the eigenvectors of these appropriate affinity matrices contain the class structure of the graph. Step 3 of Algorithm 0 underlines that the informative eigenvalues are generally found to be the largest or smallest eigenvalues, so that there is an algorithmically clear definition of which eigenvectors should be used for clustering.

In order to review the most relevant spectral clustering algorithms for community detection, let us define the adjacency matrix $A \in \{0, 1\}^{n \times n}$ of the graph $\mathcal{G}$ by $A_{ij} = \mathbb{1}_{(ij) \in \mathcal{E}}$ where $\mathbb{1}_x$ equals 1 if the condition x is verified and zero otherwise, and the diagonal degree matrix $D = \mathrm{diag}(A\mathbf{1}_n)$, where $\mathbf{1}_n$ is the vector with all entries equal to one. In (Fiedler, 1973), and in the case of $k = 2$ communities, it was proposed to reconstruct communities using the eigenvector corresponding to the second smallest eigenvalue of the *combinatorial graph Laplacian* matrix $L = D - A$. It was then shown (see *e.g.* Von Luxburg, 2007) that this eigenvector provides a relaxed solution of the *RatioCut* problem. Based on this result, one can build a spectral algorithm that, referring to Algorithm 0, sets $M = L$; in this case, the informative eigenvectors correspond to the smallest eigenvalues and step 4 is not performed. Similarly, the *normalized graph Laplacian* matrices ($L^{\mathrm{rw}} = D^{-1}A$ and $L^{\mathrm{sym}} = D^{-1/2}AD^{-1/2}$) can be used to solve a relaxed form of the *NCut* problem (Von Luxburg, 2007). In (Shi and Malik, 2000), $M = L^{\mathrm{rw}}$, the informative eigenvectors correspond to the k largest eigenvalues and step 4 is again not performed, while in (Ng et al., 2002), $M = L^{\mathrm{sym}}$, the informative eigenvectors correspond to the k largest eigenvalues and step 4 is performed. Another classical choice of matrix M is in (Newman, 2006), where the authors propose an alternative method to define communities through a NP-hard optimization problem by maximizing the so-called modularity of the graph. For $k = 2$ classes, a relaxed solution can be obtained by considering the largest eigenvector of the modularity matrix $A - \frac{\boldsymbol{d}\boldsymbol{d}^T}{2|\mathcal{E}|}$, where $\boldsymbol{d} = A\mathbf{1}_n$ is the degree vector. For more than two communities, the algorithm is slightly more involved and deviates from the general structure of Algorithm 0.

Many results justifying the earlier methods exist in the asymptotic limit where $n \gg 1$, in the so-called *dense* regime, in which the average degree (number of connections) grows with the size (n) of the network. However, as we anticipated in the introduction, real networks tend instead to be *sparse* as well as *heterogeneous* in their degree distribution, following broad distributions such as power laws. These two characteristics (sparsity and heterogeneity) make it difficult to theoretically predict the behavior and performances of Algorithm 0 and to choose a proper matrix M to work with. From a linear algebra viewpoint, sparsity in the node degrees has been shown to spread the eigenvalues of the Laplacian matrix, with the deleterious effect to "swallow" the isolated (smallest or largest) informative eigenvalues within the so-called "bulk" of uninformative eigenvalues: as a result, the corresponding informative eigenvectors are no longer found to be associated with dominant (largest or smallest) eigenvalues, as shown in Figure 1B, and as opposed to Figure 1A. Besides, by losing the isolation of informative eigenvalues, the associated eigenvectors tend to merge with the eigenvectors associated to close-by (non-informative) eigenvalues. Heterogeneity in the degrees also induces a severe negative effect: it modifies the amplitude of the i-th entry (for all $1 \leq i \leq n$) of the informative eigenvectors by a non-trivial function of the degree of node i. This further compromises the efficiency of the last classification step (usually performed through k-*means*), as shown in Figure 1D, to be opposed to Figure 1C. Several contributions have thus tackled the challenging problem of devising efficient spectral clustering in the sparse regime and for an arbitrary degree distribution.

One prominent line of research was devoted to defining spectral algorithms capable of retrieving communities as soon as theoretically possible ($\alpha > \alpha_c$, Equation (2)) on sparse graphs generated from the SBM, for which we recall that there is a sharp transition between the "detectable" and "undetectable" situations. The authors of (Krzakala et al., 2013) proposed an algorithm based on the (non-symmetric) non-backtracking matrix $B_{(ij)(lm)} = \delta_{jl}(1 - \delta_{im})$, with $B \in \{0,1\}^{2|\mathcal{E}| \times 2|\mathcal{E}|}$. In (Massoulié, 2014) it was indeed shown that the eigenvectors associated to the largest (in modulus) eigenvalues of B have a non-trivial correlation with the actual underlying communities, as soon as $\alpha > \alpha_c$. A closely related algorithm is the one proposed in (Saade et al., 2014) which instead uses the eigenvectors attached to the smallest eigenvalues of the Bethe-Hessian matrix $H_r = (r^2 - 1)I_n + D - rA \in \mathbb{R}^{n \times n}$, for $r = \sqrt{\rho(B)}$ (B being the non-backtracking matrix just defined and $\rho(\cdot)$ the spectral radius). These algorithms have the benefit of achieving non-trivial classification down to the detectability threshold, but they have only been investigated under a sparse SBM setting; hence they do not cope with degree heterogeneity. The extension to the sparse DC-SBM case was treated in (Gulikers et al., 2017a), in which the authors show that spectral clustering on B also works down to the (generalized) threshold.

All these results are powerful as they propose algorithms capable of reaching the information-theoretic threshold, but they also have inherent weaknesses as they only guarantee a positive correlation of their output classification with the underlying true structure. Specifically, these positive correlations do not imply that the classification performance is maximal. In particular, even in (Gulikers et al., 2017a) where spectral clustering on the DC-SBM is studied, the problem of eigenvector pollution due to degree heterogeneity (which is one of our main focuses in this article) is not discussed and *a fortiori* not corrected.

Another line of research which studies the reconstruction of sparse DC-SBM networks suggests to exploit regularization (i.e., using $D_\tau = D + \tau I_n$ and $A_\tau = A + \tau \mathbf{1}_n \mathbf{1}_n^T$ in place of D and A) as a solution to maintain the location of the informative eigenvalues in their dominant (smallest or largest) positions (Amini et al., 2013; Joseph and Yu, 2016; Lei and Rinaldo, 2015; Le et al., 2018). A particularly interesting method in this direction is proposed in (Qin and Rohe, 2013) which uses the *regularized symmetric reduced Laplacian* matrix $L_\tau^{\mathrm{sym}} = D_\tau^{-1/2} A D_\tau^{-1/2}$, for $\tau = \bar{d}$, the average degree. In terms of Algorithm 0, the algorithm proposed in (Qin and Rohe, 2013) sets $M = L_\tau^{\mathrm{sym}}$, searches for the k eigenvectors corresponding to the k largest eigenvalues of M, and then performs the normalization step 4 on the rows of the resulting matrix $\boldsymbol{X}$. While (Qin and Rohe, 2013) deals with graphs with arbitrary degree distribution (generated from the DC-SBM), the authors do not discuss whether their algorithm can achieve non-trivial clustering down to the threshold, i.e. as soon as $\alpha > \alpha_c$. We will bring new conclusions in this direction in Section 3.

1.3 Contributions

In the state-of-the-art methods we presented, while sparsity is not properly accounted for in (Shi and Malik, 2000; Ng et al., 2002), the line of research based on B (and consequently H_r) (Krzakala et al., 2013; Saade et al., 2014; Bordenave et al., 2015; Gulikers et al., 2017a) does account for sparsity and provides methods to reach the detectability threshold; yet, all they guarantee is the possibility to obtain a positive, possibly suboptimal, correlation between the algorithm output and the underlying community structure. As for the works on regularization (Qin and Rohe, 2013; Le and Levina, 2015), they only establish theoretical results of perfect community recovery far from the (most interesting) detection threshold.

The present work solves both the issues of sparsity and heterogeneity at once by proposing a simple spectral clustering scheme which, under a sparse DC-SBM setting, provably performs non-trivial clustering as soon as $\alpha > \alpha_c$ and is robust to degree heterogeneity, as it retrieves eigenvectors not infected by the node degrees. This claim is supported by three parallel arguments having the side advantage to unify in a joint framework the ideas of (Krzakala et al., 2013; Saade et al., 2014; Shi and Malik, 2000; Fiedler, 1973; Qin and Rohe, 2013).

We further discuss how our algorithm, analyzed under the DC-SBM assumption, can be extended to real graphs. In particular, the algorithm provides an accurate estimate of the number of communities when unknown, and is observed to systematically achieve good scores (both in terms of modularity or log-likelihood of the DC-SBM posterior probability) on all real-world graphs which we experimented on.

In summary, our main contributions are as follows:

1. We devise a practical spectral algorithm exploiting the eigenvectors corresponding to the smallest eigenvalues of the Bethe-Hessian matrix H_r for a set of properly chosen values of r. We further provide strong arguments and extensive numerical simulations to claim that the proposed method achieves non-trivial reconstruction down to the DC-SBM detectability threshold.
2. The proposed algorithm is extensively tested on synthetic DC-SBM graphs (for which the performance is measured in terms of overlap between the estimated and ground-truth labels) as well as on real networks (for which the performance is measured in terms

of modularity and DC-SBM log-likelihood). In both cases our algorithm outperforms or is on par with the standard competing spectral techniques for sparse graphs.

3. In passing, we provide a new vision to spectral clustering, and in particular a compelling new connection between our proposed approach and all aforementioned standard spectral methods, so far treated independently. In particular, we show that the *regularized reduced random walk Laplacian* matrix $L_\tau^{\text{rw}} = D_\tau^{-1}A$ (that shares the same eigenvalues as L_τ^{sym}) can be used to perform non-trivial clustering down to the detectability threshold for a choice of τ strongly related to our proposed algorithm.

A Python implementation of the codes needed to reproduce the results of this article can be found at lorenzodallamico.github.io/codes/: all data, experiments and running times refer to this implementation. On top of the Python codes, we developed a more efficient Julia implementation of the proposed Algorithm 2, available in the CoDeBetHe.jl package which gains a factor x5 ∼ x10 in terms of computational time over the Python implementation.

The remainder of the article is organized as follows: Section 2 formally introduces the problem at hand and provides three complementary supporting arguments to the claim that the eigenvectors corresponding to the smallest eigenvalues of a set of Bethe-Hessian matrices H_r, with different values of r, can be used to reconstruct classes. Section 3 draws the connection between the above H_r and the regularized Laplacian and proves that, under a correct choice of τ related to r above, the matrix L_τ^{rw} is also a suitable candidate to reconstruct communities. Section 4 introduces the final form of the proposed algorithm and provides simulation outputs on real networks. Section 5 finally closes the article.

1.4 Notation

- Matrices are indicated with standard font capital letters (M). The only exception is $\boldsymbol{X}$, the matrix concatenating the eigenvectors of the spectral clustering algorithm.
- $M_{\bullet,p}$ indicates the p-th column of M and $M_{p,\bullet}$ its p-th row.
- Vectors are denoted in bold $\boldsymbol{v}_p$, while scalar and vector elements in standard font $v_{p,i}$.
- We denote by $s_i^\uparrow(M)$ the i-th smallest eigenvalue of a Hermitian matrix M, i.e., $s_1^\uparrow(M) \le s_2^\uparrow(M) \le \cdots \le s_n^\uparrow(M)$, and $s_i^\downarrow(M)$ the i-th largest. For a non-Hermitian matrix M, we denote by $s_{R,i}^{\uparrow/\downarrow}(M)$ the i-th smallest/largest real eigenvalue, while $s_i^{\uparrow/\downarrow_{|\cdot|}}(M)$ indicates the i-th smallest/largest *in modulus*. When using the notation $s_i(M)$ we consider a generic eigenvalue of M.
- $S(M) = \{s_i(M), \ 1 \le i \le n\}$ denotes the set of eigenvalues of $M \in \mathbb{R}^{n \times n}$.
- The notation $\mathbf{1}_n$ indicates the vector of size n containing all ones: $\mathbf{1}_n^T = (1, 1, \cdots, 1)$.
- We denote the neighborhood of the node i of a graph with adjacency matrix A by $\partial i = \{j \in \mathcal{V} \ : \ A_{ij} = 1\}$.

2. Informative Eigenvectors of B and H_r in the Sparse Regime

Consider an undirected and unweighted graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ with $|\mathcal{V}| = n$ nodes and $|\mathcal{E}|$ edges. Our objective is to devise a community detection algorithm on $\mathcal{G}$ which is resilient to the typically heterogeneous and sparse nature of real graphs, and also accounts for the fact that the actual grouping of nodes in communities is not clear-cut in a real graph (i.e., there usually exists

no "optimal" number of communities and no "ground-truth" for the allocation of nodes into communities).

While our proposed algorithm (Algorithm 2) is applicable to any arbitrary graph (as discussed in Section 4), for analytic purposes, we start the article by assuming that $\mathcal{G}$ is constituted of exactly k communities and generated as a sparse DC-SBM graph; that is, the edges of $\mathcal{G}$ are drawn independently according to Equation (1). Before delving into the technical details though, we first provide a short digest of our main result.

2.1 Informal Statement of the Main Result

Let us present our central result (Claim 1) and the resulting new algorithm for community detection in simple terms. The result revolves around the shape of the eigenvectors attached to the smallest (algebraic) eigenvalues of the Bethe-Hessian matrix $H_r \in \mathbb{R}^{n \times n}$, defined as

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

for carefully chosen values of $r \geq 1$; we recall that A and D are the adjacency and degree matrices of $\mathcal{G}(\mathcal{V}, \mathcal{E})$, respectively, while I_n is the identity matrix of size n.

To each k-class graph, community detection is performed by first associating a set of k Bethe-Hessian matrices $\{H_{\zeta_p}\}_{p=1,\dots,k}$, with ζ_p defined so that the p-th smallest eigenvalue of H_{ζ_p} is equal to zero. We then extract from each matrix the eigenvector $\boldsymbol{x}_p$ attached to the zero eigenvalue (so that $H_{\zeta_p}\boldsymbol{x}_p = 0$). The eigenvectors $\{\boldsymbol{x}_p\}_{p=1,\dots,k}$ are stacked in the columns of the matrix $\boldsymbol{X}$ defined in Algorithm 0 and used to produce the small dimensional node embedding on which the *k-means* algorithm is applied.

Claim 1 justifies the relevance of the above procedure as a highly performing community detection method. Specifically, for a k-class DC-SBM graph, Claim 1 states that:

- The largest value of ζ_p that can be defined is ζ_k. We show that, as a consequence, the $(k+1)$-th smallest eigenvalue of H_r is always positive for all r and the maximal number of negative eigenvalues of H_r as function of r is precisely equal to k. This allows one to build an estimator for the number of classes.
- The eigenvectors $\{\boldsymbol{x}_p\}_{p=1,\dots,k}$ and the k dominant eigenvectors of the non-backtracking matrix B are correlated to the class structure under the same hypothesis. This, for $k = 2$ classes of equal size, means that $\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$ can be used to reconstruct communities as soon as theoretically possible, i.e. when $\alpha > \alpha_c$, (Equation (2)).
- The entries of $\{\boldsymbol{x}_p\}_{p=1,\dots,k}$ are not polluted by the degree heterogeneity; hence do not suffer the problem shown in Figure 1D that hinders the performance of *k-means*.

The properties of $\{\boldsymbol{x}_p\}$ allow us to define an efficient algorithm for community detection in sparse graphs with a heterogeneous degree distribution, solving at once the two main challenges discussed in Section 1.

2.2 Main Result

In this section, we proceed with the formal statement of Claim 1, together with a clear definition of the hypotheses under which it is formulated.

2.2.1 Model and Setting

Let $\mathcal{G}$ be a k class DC-SBM generated graph. We consider graphs $\mathcal{G}$ whose classes are not necessarily of equal size and denote by $0 < \pi_p < 1$, for every class $1 \leq p \leq k$, the fraction of nodes belonging to class p. We correspondingly define the diagonal matrix $\Pi = \operatorname{diag}(\pi_1, \ldots, \pi_k)$ (note that in particular $\operatorname{Tr}(\Pi) = 1$). We next denote by $(\nu_p, \boldsymbol{v}_p)$ the eigenvalue-eigenvector pairs of $C\Pi$, i.e., $C\Pi \boldsymbol{v}_p = s_p^{\downarrow}(C\Pi)\boldsymbol{v}_p \equiv \nu_p \boldsymbol{v}_p$, where we recall $s_p^{\downarrow}(C\Pi)$ denotes the p-th largest eigenvalue of $C\Pi$, hence the eigenvalues are sorted as $\nu_1 \geq \ldots \geq \nu_k$. Note that the eigenvalues of $C\Pi$ are all real as $C\Pi$ has the same spectrum as $\Pi^{1/2}C\Pi^{1/2}$, which is a symmetric and real matrix (thus diagonalizable in $\mathbb{R}$). In order to set ourselves under an asymptotically non-trivial community detection setting, the following assumptions on matrix $C\Pi$ need to be posed:

Assumption 1 *The symmetric matrix C and the diagonal matrix Π with $\operatorname{Tr}(\Pi) = 1$ have entries independent of n and satisfy the following hypotheses:*

1. *$(C\Pi)_{pq} > 0, \ \forall \ 1 \leq p, q \leq k$*

2. *$C\Pi\boldsymbol{1}_k = c\boldsymbol{1}_k$, for some positive c*

3. *For $1 \leq p \leq k$, $\nu_p > \sqrt{\frac{c}{\Phi}}$.*

Assumption 1 is fundamental to our analysis. Its three key requests deserve a detailed motivation and explanation:

1. $(C\Pi)_{pq} > 0$: this implies that there is a non-null probability of connection between any two classes. As a consequence, one may apply the Perron-Frobenius theorem on $C\Pi$: the eigenvalue of $C\Pi$ of largest modulus, denoted with $c > 0$, is positive, simple and its corresponding eigenvector is the only one that can be chosen with all positive entries. By definition of the eigenvalues $\{\nu_p\}_{p=1}^k$, we have in particular $c = \nu_1$.

2. $C\Pi\boldsymbol{1}_k = c\boldsymbol{1}_k$: this assumption imposes that the Perron-Frobenius eigenvector of $C\Pi$ is the vector of all ones, $\boldsymbol{1}_k$. From Equation (1) and the independence of the entries of A, one can show that this assumption implies that $\mathbb{E}[\bar{d}]$, the expected value of the average degree, is equal to c. In the two-class symmetric case, $c = (c_{\text{in}} + c_{\text{out}})/2$ as introduced in Equation (2). Assumption 1 also importantly ensures that the average degree does not depend on the class; in fact, the expected degree of each node i belonging to class p equals $\theta_i(C\Pi\boldsymbol{1}_k)_p = \theta_i c$, which is independent of p. this is a standard assumption in the literature and sets the problem under a scenario for which the degree of each node, being independent of its label, does not bring any information on the class structure (Krzakala et al., 2013; Decelle et al., 2011; Bordenave et al., 2015).

3. $\nu_p > \sqrt{c/\Phi}$: this assumption is made for consistency between the actual number of classes k and the number of classes $\hat{k}$ that can be detected with the spectral methods described in the following and consists in imposing $k = \hat{k}$. Recalling that $\nu_1 = c$, this assumption notably implies that $c\Phi > 1$, which is a necessary and sufficient condition to have a giant component in $\mathcal{G}$, as stated in Theorem 1 in Appendix A. In the case of two classes of equal sizes, the condition $\nu_p > \sqrt{c/\Phi}$ is equivalent to

setting the problem above the detectability threshold $\alpha > \alpha_c$. As a consequence of this assumption, for all large n with high probability, the spectrum of the affinity matrices under study (B, H_r, L_τ^{rw}, etc.) can be decomposed as the union of a set of contiguous (non-informative) eigenvalues, collectively referred to as the *bulk*, and of a set of k (informative) *isolated* eigenvalues $s_1, \ldots, s_k$ found at non-vanishing distance from the bulk (that is, up to multiplicity, s_i is isolated if for some $\epsilon > 0$, we have for all n and $j \neq i$, $|s_i - s_j| > \epsilon$). As opposed to bulk eigenvalues, the isolated eigenvalues are referred to as *informative* as their corresponding eigenvectors are non-trivially correlated to the community structure.

It must be noted that, from Assumption 1, we necessarily have $\nu_p > 0$ for all $1 \leq p \leq k$. However, the results of this paper remain valid if one replaces the third point of Assumption 1 by $|\nu_p| > \sqrt{\frac{c}{\Phi}}$ (that is, ν_p for $p \geq 2$ may be of arbitrary sign and only its modulus may be lower-bounded; $\nu_1 = c$, being the Perron-Frobenius eigenvalue, is necessarily positive in any case). As the notations to describe this more general case are more cumbersome, we prefer here to focus on the simpler case where $\nu_p > \sqrt{\frac{c}{\Phi}}$ in the core of the article and defer the discussion of the general case to Section 3.2.2.

2.2.2 CHARACTERIZATION OF THE INFORMATIVE EIGENVECTORS OF H_r AND B

Before stating the main claim of the article, we first need to introduce the values $r = \zeta_p$ which must be carefully selected to best operate on the Bethe Hessian H_r.

Definition 1 (ζ_p) *Consider an arbitrary graph $\mathcal{G}$ composed of n_{CC} connected components. Let $\mathcal{G}^{(j)}$ be the subgraph reduced to the j-th connected component, and $H_r^{(j)}$ its associated Bethe-Hessian matrix, for $r \in \mathbb{R}$ (Equation (3)). Let p be an integer between 1 and n. We define, if it exists, $\zeta_p^{(j)}$ as:*

$$\zeta_p^{(j)} = \min_{r \geq 1} \{r : s_p^\uparrow(H_r^{(j)}) = 0\}, \tag{4}$$

where $s_p^\uparrow(\cdot)$ is the p-th smallest eigenvalue. In words, $\zeta_p^{(j)}$ is, if it exists, the smallest value of the parameter $r \geq 1$ such that the p-th smallest eigenvalue of $H_r^{(j)}$ is null.

Here we list a few important properties of these values:

1. At $r = 1$, $H_1^{(j)} = D^{(j)} - A^{(j)}$ is the combinatorial Laplacian of the subgraph associated to the j-th connected component. As this subgraph is by definition connected, it is well known (Chung, 1997) that its smallest eigenvalue is null and that $s_{p \geq 2}^\uparrow\left(H_1^{(j)}\right) > 0$. Consequently, $\zeta_1^{(j)} = 1$ always exists and if $\zeta_{p \geq 2}^{(j)}$ exists, it is strictly superior to 1.

2. If $\zeta_{p \geq 2}^{(j)}$ exists, then $\zeta_2^{(j)}, \ldots, \zeta_{p-1}^{(j)}$ necessarily exist and are smaller or equal to $\zeta_p^{(j)}$. In fact, if $\zeta_p^{(j)}$ exists, it means that at $r = \zeta_p^{(j)}$, 0 is the p-th smallest eigenvalue of $H_r^{(j)}$: the $p-1$ smallest are thus ≤ 0. Given that these $p-1$ smallest are ≥ 0 at $r = 1$ and by continuity of $s_p^\uparrow(H_r^{(j)})$, they necessarily cross zero before $\zeta_p^{(j)}$. Similarly, if $\zeta_{p \geq 2}^{(j)}$ does not exist, then all $\zeta_q^{(j)}$ for $q > p$ do not exist.

3. Empirically, on as many connected graphs as we could think of, we have observed that the function $s_p^\uparrow(H_r)$ for $r \geq 1$ either never crosses zero (in which case ζ_p does not exist), crosses zero exactly twice and is convex between these two crossings (in which case ζ_p is the lowest of the two values), or, in very symmetric cases, touches zero exactly once without crossing it (in which case ζ_p is that value).

We are now in position to state our main result.

Claim 1 *Let $\mathcal{G}(\mathcal{V}, \mathcal{E})$ be generated according to Equation (1), that is, a DC-SBM with k communities. Let D and A be its degree and adjacency matrices, and $H_r = (r^2-1)I_n + D - rA$, for $r \in \mathbb{R}$ its Bethe-Hessian matrix. Provided that Assumption 1 is satisfied, we have with high probability for large n that:*

- *There is only one connected component j for which $\zeta_{p \geq 2}^{(j)}$ exists: it is the giant component. In the following, abusing notation, we simply write ζ_p instead of $\zeta_p^{(j)}$ to refer to the ζ's associated to this giant component. One has $\zeta_1 = 1$ and, if it exists, $\zeta_{p \geq 2}$ verifies[1]*

$$\zeta_p = \min_{r>1} \{r : s_p^\uparrow(H_r) = 0\}. \tag{5}$$

- *The largest p for which ζ_p exists is equal to k, the number of underlying communities of the DC-SBM.[2] One has $1 = \zeta_1 < \zeta_2 \leq \ldots \leq \zeta_k \leq \sqrt{c\Phi}$. More precisely:*

$$\forall\, p \text{ s.t. } 1 \leq p \leq k, \quad \zeta_p = \frac{c}{\nu_p} + o_n(1). \tag{6}$$

- *For $2 \leq p \leq k$ and $\zeta_p \leq r \leq \sqrt{c\Phi}$, the p smallest eigenvalues of H_r are isolated. In particular, zero is an isolated eigenvalue of H_{ζ_p}. Its corresponding eigenvector $\boldsymbol{x}_p$ is correlated to the community structure and the entries of $\boldsymbol{x}_p$ are in expectation (over realizations of A) independent of the degrees of the graph.*

In simple terms, Claim 1 predicts that, in a graph of k communities, spectral clustering can be successfully performed by successively retrieving the eigenvector associated to the null eigenvalue of each of the matrices $H_{\zeta_2}, \ldots, H_{\zeta_k}$. Note that these eigenvectors have null entries for nodes that are not connected to the giant component. Thus, all the nodes not connected to the giant component will be arbitrarily classified: this is a structural consequence of sparsity, for which complete recovery is not feasible (Mossel et al., 2014).

In the specific case of $k = 2$ classes of equal size with $C_{\ell_i, \ell_j} = c_{\text{in}}$ if $\ell_i = \ell_j$ and c_{out} if not, Claim 1 states that the ζ's exist only up to $p = k = 2$, and that[3] $\zeta_2 = (c_{\text{in}} + c_{\text{out}})/(c_{\text{in}} - c_{\text{out}}) + o_n(1)$. Also, the relevant community information is found in the

---

1. This statement allows to define ζ_p with respect to the Bethe-Hessian matrix of the whole graph, instead of the Bethe-Hessian matrix of its giant component. The fact that Eq. (5) is verified with high probability is not evident. In fact, considering all the disconnected components at once might change the ordering of the eigenvalues. More details are to be found in Appendix B.

2. Precisely, we will see that $s_{k+1}^\uparrow(H_r) > 0$ for all $r > 1$ with high probability, so that ζ_{k+1} is not defined; in fact it was interestingly shown in (Saade et al., 2014) and we will verify that $s_{k+1}^\uparrow(H_{\sqrt{c\Phi}}) \downarrow 0^+$ as $n \to \infty$, but the limit is never reached.

3. One can indeed easily show that $\nu_2 = \frac{c_{\text{in}} - c_{\text{out}}}{2}$ in this case of two equal-sized classes

eigenvector $\boldsymbol{x}_2$ associated with the zero eigenvalue (the second smallest) of H_r for $r = \zeta_2$. As we will see next, this value of r differs from the choice $r = \sqrt{\rho(B)}$ advocated by (Saade et al., 2014), unless $\alpha = \frac{c_{\text{in}} - c_{\text{out}}}{\sqrt{c}} = \alpha_c = \frac{2}{\sqrt{\Phi}}$, i.e., exactly at the phase transition point.

Figure 3 provides a visual representation of the typical spectrum of H_r for r not too far from a ζ_p: the eigenvalue of H_r closest to zero is clearly isolated and the eigenvector associated to this eigenvalue is strongly aligned to the community structure. In Figure 4, for a typical realization of a DC-SBM, $s_p^{\uparrow}(H_r)$ is represented as a function of r for $p = 1, 2, 3$ and 4 in solid lines and $p = 5$ in dotted line. In this instance, ζ_5 does not exist (and all subsequent ζ_p's do not exist either), $\zeta_1 = 1$, and ζ_2, ζ_3 and ζ_4 exist and lie between 1 and 2.

The non-obvious parts of the claim are (i) of course that such ζ_p's do exist up to $p = k$, (ii) that they are concentrated around a deterministic value depending on the statistics of the model, (iii) importantly, that zero is indeed an isolated eigenvalue in the spectrum of H_{ζ_p}, (iv) that the associated eigenvector is informative for the underlying community structure and that is not infected by the degrees of the graph.

The statement of Claim 1 is formulated as a claim in the sense that, while efforts have been made to rigorously prove parts of the result (see, *e.g.*, Coste and Zhu, 2021), the mathematical tools required to fully prove Claim 1, to the best of our knowledge, do not exist yet. Instead, the remainder of this section will propose three complementary supporting elements, arising from non-rigorous but convincing approximations, in particular borrowing arguments from the field of statistical physics.

2.3 Supporting Arguments to Claim 1

This section details our heuristic arguments in support of Claim 1. Specifically, we will successively show

- under Section 2.3.1 that the vectors $\boldsymbol{x}_p$, solution of $H_{\zeta_p}\boldsymbol{x}_p = 0$, are informative in the sense that they are correlated with the community labels.

- under Section 2.3.2, that the informative null eigenvalue of H_{ζ_p} (associated to the informative eigenvector $\boldsymbol{x}_p$) is located in p-th smallest position and is isolated. This result is algorithmically crucial to determine ζ_p itself.

- under Section 2.3.3, that the entry $x_{p,i}$ is essentially independent of d_i, the degree of node i; precisely, it is strictly independent of d_i on average (over random allocation of the labels) and only depends on d_i through a noise term of order $1/\sqrt{d_i}$ otherwise.

2.3.1 LINEARIZATION OF THE BELIEF PROPAGATION EQUATIONS

We first proceed to our agenda by arguing that the eigenvectors $\boldsymbol{x}_p$ of H_{ζ_p} are correlated with the community labels.

A first approach to the question of community detection in sparse graphs consists in estimating the community allocation probabilities $\mathbb{P}(\ell_i | A)$ using belief propagation (BP) approximation, based on the local-tree approximation of sparse graphs. The BP node

marginal is given by (Decelle et al., 2011, Equation 28), which we recall here:

$$\mathbb{P}(\ell_i|A) \approx \frac{\pi_{\ell_i}}{Z_i} e^{-h_{\ell_i}} \prod_{j \in \partial i} \sum_{\ell_j} C_{\ell_i,\ell_j} \eta_{ij}(\ell_j), \tag{7}$$

in which Z_i is the probability normalization constant, and the $\eta_{ji}(\ell_i)$ (the "messages") and $h(\ell_i)$ are the solutions to (see Decelle et al., 2011, Equations 26,27):

$$\eta_{ji}(\ell_i) = \frac{1}{Z_{ji}} \pi_{\ell_i} e^{-h_{\ell_i}} \prod_{m \in \partial i \setminus j} \sum_{\ell_m} C_{\ell_m,\ell_i} \eta_{im}(\ell_m) \tag{8}$$

$$h_{\ell_i} = \frac{1}{n} \sum_{k \in \mathcal{V}} \sum_{\ell_k} C_{\ell_i,\ell_k} \mathbb{P}(\ell_k|A). \tag{9}$$

Equation (8) admits the trivial $\eta_{ji}(\ell_i) = \pi_{\ell_i}$ which corresponds to the so-called *paramagnetic fixed point* (Zhang, 2015). Let us perform a linear expansion of Equation (8) around this trivial fixed point by denoting $\eta_{ji}(\ell_i) = \pi_{\ell_i} + b_{ji}(\ell_i)$. Letting $\boldsymbol{b}^T = [\boldsymbol{b}_1^T, \ldots, \boldsymbol{b}_k^T] \in \mathbb{R}^{2|\mathcal{E}|k}$, with $\boldsymbol{b}_p \in \mathbb{R}^{2|\mathcal{E}|}$ containing the entries of $b_{ij}(p)$, we obtain as in (Krzakala et al., 2013):

$$(T \otimes B)\boldsymbol{b} = \boldsymbol{b} + o_n(1) \tag{10}$$

where $\otimes$ is the Kronecker product, $T = \frac{\Pi C}{c} - \Pi \mathbf{1}_k \mathbf{1}_k^T$ and $B \in \{0,1\}^{2|\mathcal{E}| \times 2|\mathcal{E}|}$ is the non-backtracking matrix, which we recall is defined on the set of directed edges $\mathcal{E}^d$ of $\mathcal{G}$ as:

$$B_{(ij)(kl)} = \delta_{jk}(1 - \delta_{il}), \ \forall \ (ij), (kl) \ \in \ \mathcal{E}^d, \tag{11}$$

in which δ is the Kronecker symbol. Before proceeding, we recall an important property of the spectrum of B rigorously established in (Bordenave et al., 2015; Gulikers et al., 2017a) and which we will extensively use in the following. Under Assumption 1, with high probability, the matrix B has k isolated real eigenvalues equal to[4] $\nu_p \Phi + o_n(1)$, $1 \leq p \leq k$, while all other (complex) eigenvalues are contained within a disk of the complex plane of radius $\sqrt{\rho(B)} = \sqrt{c\Phi} + o_n(1)$.

From Equation (10), our interest is in the eigenvalues of $(T \otimes B)$ equal (or close) to one. From the properties of the Kronecker product, the set of eigenvalues of $(T \otimes B)$ is $S(T \otimes B) = \{s_i(B)s_j(T) \ : \ 1 \leq i \leq 2|\mathcal{E}|, 1 \leq j \leq k\}$. This induces a one-to-one relation between the $k-1$ non-zero eigenvalues[5] of T and the $k-1$ eigenvalues of B, that need to satisfy the relation $s_q(T)s_t(B) = 1$ for some t, q. From the expression of T, it comes that the p-th largest eigenvalue of T satisfies $s_p^{\downarrow}(T) = \frac{\nu_{p+1}}{c}$ for $1 \leq p \leq k-1$ (see Appendix C for details) so that, for $2 \leq p \leq k$, there must exist $\zeta_p \in S(B)$ such that:

$$\zeta_p = \frac{c}{\nu_p} + o_n(1). \tag{12}$$

To these eigenvalues correspond $k-1$ *exact* eigenvectors $\boldsymbol{g}_p \in \mathbb{R}^{2|\mathcal{E}|}$ of B satisfying $B\boldsymbol{g}_p = \zeta_p \boldsymbol{g}_p$ (with $\boldsymbol{g}_p = \boldsymbol{b}_p + o_n(1)$) that are naturally "informative" as they are associated to structural

---

4. Recall that ν_p is the p-th largest eigenvalue of $C\Pi$.
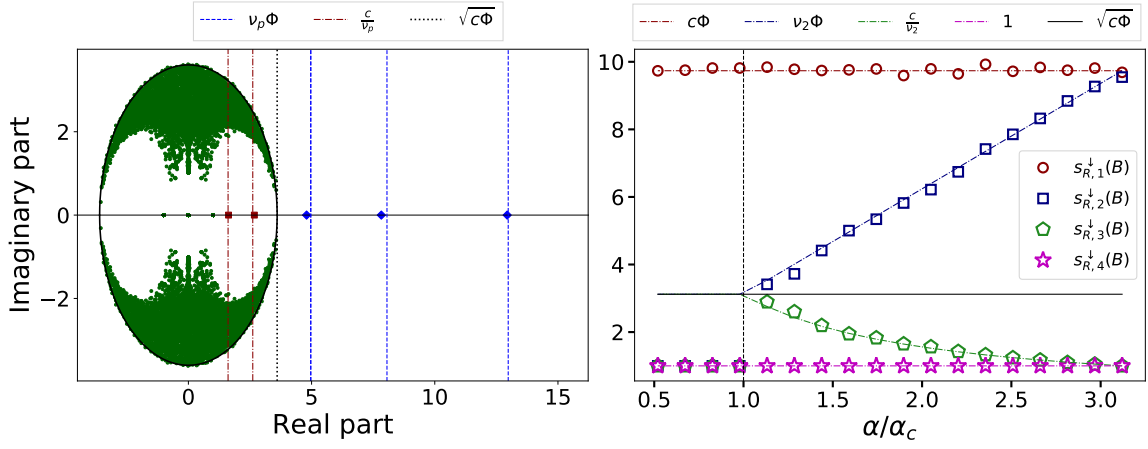5. Note that $s_1^{\uparrow}(T) = 0$.

Figure 2: (Left): spectrum of the matrix B on the complex plane. The green dots represent the uninformative eigenvalues, the blue diamonds the real eigenvalues outside the bulk and the red squares the real eigenvalues inside the bulk. The lines indicate the theoretical values. For this network $n = 5000$, $k = 3$, $\theta_i \sim [\mathcal{U}(3,10)]^3$, $c = 8$. The off-diagonal elements of C are distributed according to $C_{p>q} \sim \mathcal{N}(c_{\text{out}}, c_{\text{out}}/k)$ with $c_{\text{out}} = 4$ and the diagonal elements are then fixed to have $C\Pi\mathbf{1}_k = c\mathbf{1}_k$. The element of $\boldsymbol{\pi}$ are distributed as $\pi_i \sim \mathcal{N}(1/k, 1/2k)$. The entries are then rescaled sot that $\text{Tr}(\Pi) = 1$. (Right): position of the first four real eigenvalues of B as a function of α. For this simulation $n = 5000$, $k = 2$, $\theta_i = [\mathcal{U}(3,10)]^3$, $c_{\text{in}} : 12 \to 7$, $c_{\text{out}} : 0 \to 5$ with $c = 6$.

eigenvalues of T. Also, from a belief propagation standpoint, these eigenvectors are small deviations from the uninformative fixed point, so must point towards informative directions.

Note importantly here that Equation (12) *defines* ζ_p as a real eigenvalue of the matrix B. So far, this definition needs not correspond to the claimed definition of ζ_p as introduced in Claim 1: Section 2.3.2 will show that these two definitions are indeed equivalent. In addition, by Assumption 1 and the Perron-Frobenius theorem, $\nu_1 = c$ has unit multiplicity, so that $\zeta_p > 1 + o_n(1)$ for all $p > 1$. Along with Equation (12), we thus have $1 + o_n(1) < \zeta_p \leq \sqrt{c\Phi} + o_n(1)$, and thus ζ_p is asymptotically confined inside the bulk of radius $\sqrt{\rho(B)} = \sqrt{c\Phi} + o_n(1)$ of B. Figure 2 confirms, in agreement with the formal result of (Coste and Zhu, 2021) obtained in the slightly non sparse regime $\log(n)/c = o_n(1)$, that inside the disk of radius $\sqrt{\rho(B)}$, these are the only informative eigenvalues of B. Specifically, the real eigenvalues of B inside the bulk are divided between (i) the (non-informative) eigenvalues $-1, 0, 1$, and (ii) the k eigenvalues $\gamma = \zeta_p \approx \frac{c}{\nu_p} \leq \sqrt{c\Phi}$ just described.

Having identified the $k-1$ informative eigenvectors $\boldsymbol{g}_{2\leq p \leq k} \in \mathbb{R}^{2|\mathcal{E}|}$ of B, we now consider how to process $\boldsymbol{g}_p$ in order to obtain a vector of size n whose entries are in one to one mapping with the nodes of the graph. From the linearization of the marginal probability distribution of the labels (Equation (7)), the term $\sum_{j\in\partial i} b_{ij}(\ell_i) \approx \sum_{j\in\partial i} g_{\ell_i, ij} \equiv x_{\ell_i, i}$ brings

information on the label of node i, hence on the class structure. Now, let $\boldsymbol{g}$ be the solution of $B\boldsymbol{g} = \gamma\boldsymbol{g}$ for some γ and $g_i^{\mathrm{in}} = \sum_{j\in\partial i} g_{ij}$. Then, thanks to the Ihara-Bass formula (Krzakala et al., 2013; Terras, 2010), we find that

$$[(\gamma^2 - 1)I_n + D - \gamma A]\boldsymbol{g}^{\mathrm{in}} = 0. \tag{13}$$

The matrix in brackets appears to be the Bethe-Hessian matrix $H_r = (r^2 - 1)I_n + D - rA$, here for $r = \gamma$. This provides an explicit link between B and H_r. According to our earlier discussion, the informative eigenvectors of the matrix B have corresponding eigenvalues $\gamma = \zeta_p$ and there are consequently $k-1$ informative eigenvectors $\boldsymbol{x}_p$ for $2 \le p \le k$ such that

$$H_{\zeta_p}\boldsymbol{x}_p = 0. \tag{14}$$

Note that, for $\zeta_1 = 1$ we have $H_1 = D - A$ and $\boldsymbol{x}_1$ is the vector $\boldsymbol{1}_n$, which is irrelevant to reconstruct communities.

Let us specifically focus on the particular case of two classes of equal size and consider the question of the detectability threshold. In this case the matrix $C\Pi$ has a unique non-trivial eigenvalue equal to $\nu_2 = (c_{\mathrm{in}} - c_{\mathrm{out}})/2$. In this setting, B has two informative eigenvalues $\approx \nu_2\Phi$ and $\approx \frac{c}{\nu_2}$ on either side of the disk (bulk) of radius $\sqrt{c\Phi}$. As the detection problem becomes harder, the two eigenvalues get closer together until they (asymptotically) meet right at the detectability transition where $\nu_2\Phi = \sqrt{c\Phi}$ (and $\alpha = \alpha_c$). Further increasing the detection difficulty (that is, for $\alpha < \alpha_c$), the two eigenvalues now become complex, each being the complex conjugate of the other. This behavior is shown in Figure 2 (right panel). Summarizing, we have

$$\nu_2\Phi = \frac{c}{\nu_2} = \sqrt{c\Phi}, \quad \text{when } \alpha = \alpha_c$$
$$\nu_2\Phi > \sqrt{c\Phi} > \frac{c}{\nu_2} > 1, \quad \text{when } \alpha > \alpha_c.$$

The fact that the second eigenvalue of largest amplitude $\approx \nu_2\Phi$ of B remains isolated down to the *detectability threshold* is the strongest argument in favor of the algorithm proposed by (Krzakala et al., 2013). The authors in (Krzakala et al., 2013) however ignored the effect on the corresponding eigenvalue $\approx \frac{c}{\nu_2}$ which, from our present discussion, is similar. This behavior can be extended to more than two classes, obtaining $\zeta_p \approx \frac{c}{\nu_p} \le \nu_p\Phi$ with equality at the transition point where $\nu_p\Phi = \sqrt{c\Phi}$. This can be summarized as follows:

Element 1 *Under Assumption 1 for all large n with high probability, the non-backtracking matrix B of a graph $\mathcal{G}(\mathcal{V},\mathcal{E})$ generated from a k-class DC-SBM has $k-1$ isolated real eigenvalues inside the disk of radius $\sqrt{c\Phi}$ (bulk). These eigenvalues are found at positions $1 < \zeta_p = \frac{c}{\nu_p} + o_n(1)$ for $2 \le p \le k$. Besides, the eigenvectors $\boldsymbol{g}_p$ of B associated with these eigenvalues ζ_p bring information about the community structure, which can be extracted through the vectors $\boldsymbol{x}_p \in \mathbb{R}^n$, defined as:*

$$x_{p,i} = \sum_{j\in\partial i} g_{p,ij}, \quad B\boldsymbol{g}_p = \zeta_p\boldsymbol{g}_p;$$

or equivalently satisfying

$$H_{\zeta_p}\boldsymbol{x}_p = 0.$$

Element 1 provides a first statement of Claim 1, according to which the class information should be retrieved from the vector $\boldsymbol{x}_p$ solution to $H_{\zeta_p}\boldsymbol{x}_p = 0$, and that $\zeta_p = c/\nu_p + o_n(1)$. This argument however does not specify the location (in the ordered list of the eigenvalues) of the null eigenvalue of H_{ζ_p} to which the eigenvector $\boldsymbol{x}_p$ corresponds nor the structure of the eigenvector $\boldsymbol{x}_p$, and in particular its dependence on the degrees of the graph.

The subsequent sections will cover these aspects.

2.3.2 THE EXCITED STATES OF THE ISING HAMILTONIAN ON $\mathcal{G}$

In this section, through a statistical physics mapping between the graph $\mathcal{G}$ and a system of interacting *spins*, we aim to informally justify why the smallest eigenvalues of the matrix H_r correspond to "informative states" of the system and why the zero eigenvalue of H_{ζ_p} (associated by definition with the informative eigenvectors $\boldsymbol{x}_p$) should be isolated and in the p-th smallest position of the spectrum of H_{ζ_p}.

2.3.2.1 The Smallest Eigenvalues of H_r are Informative

Consider the Ising model of interacting *spins* $\sigma_i \in \{\pm 1\}$ on the graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, in which the intensity of spin coupling is controlled by a parameter r. Its Hamiltonian $\mathcal{H}(\boldsymbol{\sigma})$ expresses the energy of a given configuration $\boldsymbol{\sigma}$ as (see *e.g.* Mezard et al., 2009):

$$\mathcal{H}(\boldsymbol{\sigma}) = - \sum_{(ij) \in \mathcal{E}} \operatorname{ath}\left(\frac{1}{r}\right) \sigma_i \sigma_j. \tag{15}$$

The parameter $\operatorname{ath}(1/r)$ is related to the temperature of the system and it appears through an inverse hyperbolic tangent for computational ease. The vector $\boldsymbol{\sigma}$ is a random variable, distributed according to the Boltzmann distribution $\mu(\boldsymbol{\sigma})$:

$$\mu(\boldsymbol{\sigma}) = \frac{1}{Z} e^{-\mathcal{H}(\boldsymbol{\sigma})}. \tag{16}$$

Equation (16) implies that the configurations with a low energetic cost are realized with high probability. Since $\boldsymbol{\sigma}$ is a random variable, to properly characterize its *stable average configuration*, we will focus our study on its statistical mean. The mean $\boldsymbol{m}^* = \langle \boldsymbol{\sigma} \rangle$, as well as the covariance $\boldsymbol{\chi}^* = \{\langle \sigma_i \sigma_j \rangle\}_{i,j}$ and all subsequent moments of μ (here $\langle \cdot \rangle$ denotes an expectation taken over the distribution (16)), can be obtained from the explicit expression of Z which, seen as a function of the temperature, is a moment generating function. But an explicit expression of Z cannot be obtained in general. A common approximation, adapted to sparse systems, is the *Bethe* approximation (Bethe, 1935) $Z_{\text{Bethe}}(\boldsymbol{m}, \boldsymbol{\chi}) \approx Z$, which is asymptotically exact as $n \to \infty$ (see Appendix D for details). The function $F_{\text{Bethe}}(\boldsymbol{m}, \boldsymbol{\chi}) = - \log Z_{\text{Bethe}}(\boldsymbol{m}, \boldsymbol{\chi})$ is called the *Bethe free energy* and is a variational approximation of the actual free energy $F = - \log Z = \langle \mathcal{H}(\boldsymbol{\sigma}) \rangle - S[\mu(\boldsymbol{\sigma})]$, where $S[\cdot]$ denotes the entropy of a distribution. One can show (see Appendix D) that $F_{\text{Bethe}}(\boldsymbol{m}, \boldsymbol{\chi}) \geq F$, so the best estimate of the *stable configuration* $(\boldsymbol{m}^*, \boldsymbol{\chi}^*)$ is

$$(\boldsymbol{m}^*, \boldsymbol{\chi}^*) \approx \arg \min_{\boldsymbol{m}, \boldsymbol{\chi}} \big(- \log Z_{\text{Bethe}}(\boldsymbol{m}, \boldsymbol{\chi}) \big) \equiv \arg \min_{\boldsymbol{m}, \boldsymbol{\chi}} F_{\text{Bethe}}(\boldsymbol{m}, \boldsymbol{\chi}). \tag{17}$$

For r sufficiently large (that is, at high temperature) the free energy favors disordered configurations with high entropy. In this case, the function $F_{\text{Bethe}}(\boldsymbol{m}, \boldsymbol{\chi})$ has a unique
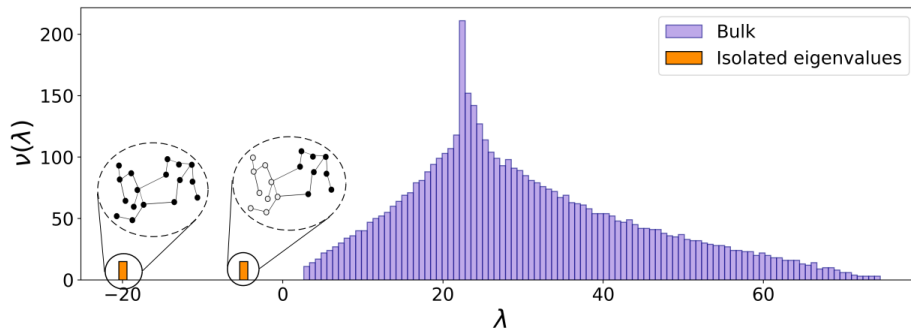
Figure 3: Spectrum of the matrix H_r for $r = 3/2\sqrt{c\Phi}$. On the x axis the eigenvalues, on the y axis the respective histogram. In purple the bulk of uninformative eigenvalues and in orange the two isolated eigenvalues—zoomed to simplify the readability—. The two drawings give a sketchy representation of the *ferromagnetic* ground state (all black spins) and of the informative state (half black and half white). For this network, $n = 5000$, $k = 2$, $\boldsymbol{\pi} = \mathbf{1}_k/2$, $c_{\text{in}} = 8$, $c_{\text{out}} = 3$, $r = 6.4$, $\theta_i \sim [\mathcal{U}(3,5)]^4$.

minimum at $\boldsymbol{m}^* = 0$ and the system is said to be in the *paramagnetic phase*. For smaller (more interesting) values of r, the free energy tends to favor configurations with a small energetic cost (minimizing Equation (15)). In this case the minima satisfy $\boldsymbol{m}^* \neq 0$ and $\boldsymbol{m} = 0$ becomes a saddle point (Leone et al., 2002): the stable average configuration of the spins has a non trivial ordering. In Appendix D we show that, on a graph $\mathcal{G}$ with k communities, there exist exactly k directions from the point $\boldsymbol{m} = 0$ along which the free energy *can* exhibit a local minimum. The limitation to k directions is a direct consequence of the fact that $\text{rank}(\mathbb{E}[A]) = k$. These directions are those along which a non-trivial organization of the spins becomes stable at sufficiently low temperature: they are naturally correlated to the underlying community structure of $\mathcal{G}$.

In order to formally carry out the stability analysis (i.e., to identify the above non-trivial directions of free energy descent), one needs to study the eigenvalues and eigenvectors of the Hessian matrix $\nabla_{\boldsymbol{m}}^2 F_{\text{Bethe}}$ of the function F_{Bethe} at the paramagnetic point $\boldsymbol{m} = 0$. In (Watanabe and Fukumizu, 2009; Saade et al., 2014), the authors show that this Hessian matrix at $\boldsymbol{m} = 0$ is strictly proportional to the Bethe-Hessian, H_r. This induces a natural link to H_r, which in the previous section was defined from the non-backtracking matrix B (arising from a linearization of the belief propagation equations). The eigenvectors associated to the negative eigenvalues of the Hessian H_r precisely correspond to the sought-for directions towards the local (or global) minima of the Bethe free energy. According to our earlier discussion, only k such directions may exist, so that only up to k eigenvalues (the k smallest) of H_r can be negative (i.e., in the limit of large n, $s_{k+1}^{\uparrow}(H_r) \geq 0, \forall \ r > 1$), and their corresponding eigenvectors are *all* correlated to the class labels.

We are thus left to showing why specifically the p-th smallest eigenvalue of H_r for the particular choice of $r = \zeta_p$ is of utmost importance, and why it corresponds to the null isolated eigenvalue of H_{ζ_p}.
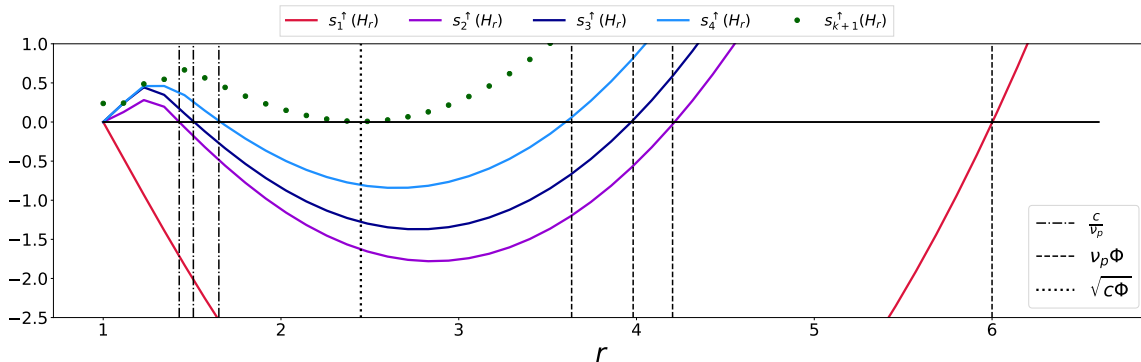
17

Figure 4: Behavior of $s_p^\uparrow(H_r)$ as a function of r for $1 \le p \le k$ in solid lines, and for $p = k+1$ in dotted line. The vertical lines are the theoretical values of c/ν_p (dashed dotted line), of $\sqrt{\rho(B)}$ (dotted line) and of $\nu_p \Phi$ (dashed lines). For this simulation, $n = 50.000$, $c = 6$, $c_{\mathrm{out}} = 2$, $k = 4$, $C_{p>q} \sim \mathcal{N}(c_{\mathrm{out}}, 2c_{\mathrm{out}}/k)$ and $\pi \propto \mathbf{1}_k$.

2.3.2.2 Zero is an Isolated Eigenvalue of H_{ζ_p}

This result follows from the observation, reported in Figure 4, that the k-th smallest eigenvalues $s_1^\uparrow(H_r), \ldots, s_k^\uparrow(H_r)$ of H_r successively equal zero in this order as r increases. Starting from $r = 1$, $H_r = H_1 = D - A$ for which we know that $s_j^\uparrow(H_1) = 0$ for $1 \le j \le n_{\mathrm{CC}}$ with n_{CC} the number of connected components and $s_i^\uparrow(H_1) \ge 0$, $i > n_{\mathrm{CC}}$. Increasing r beyond $r = 1$, the successive smallest eigenvalues of H_r first all increase and remain equal to the left edge of the bulk before escaping, each in turn (and in the order $s_2^\uparrow(H_r), \ldots, s_k^\uparrow(H_r)$), the bulk of non-informative eigenvalues. At their point of escape, they successively shift until they cross zero: this is where the successive values $\zeta_2 \le \ldots \le \zeta_k$ are defined. This in particular implies that the p-th smallest eigenvalue of H_{ζ_p} coincides with (i) the null eigenvalue of H_{ζ_p}, as well as with (ii) its largest isolated (so the last informative) eigenvalue. This allows us to redefine ζ_p as in Equation (5) of Claim 1.

Note in passing that, letting r further increase beyond ζ_k, the left edge of the non-informative bulk of H_r progressively shifts back (from positive values) towards zero until it reaches asymptotically zero in the limit where $r = \sqrt{\rho(B)}$ (Saade et al., 2014), before increasing again away for $r > \sqrt{\rho(B)}$. These last findings may be summarized as follows.

Element 2 *Given a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ with k classes, generated from the DC-SBM, the p smallest eigenvalues of the matrix H_r are isolated for $\zeta_p \le r \le \sqrt{c\Phi}$, for all $2 \le p \le k$. The entries of the p smallest eigenvectors are correlated with the class labels. Besides, in the specific case where $r = \zeta_p$, the p-th smallest eigenvalue of H_{ζ_p} is equal to zero.*

Element 2 answers the problem of locating the informative eigenvalue-eigenvector pair $(0, \boldsymbol{x}_p)$ of H_{ζ_p} introduced in Element 1, and of justifying why it is isolated. The last step, developed in the next section, consists in better understanding the structural content of the informative eigenvector $\boldsymbol{x}_p$.

2.3.3 Parametrization to Provide Resilience to Degree Heterogeneity

From a purely algebraic standpoint, the Bethe-Hessian matrix H_r may be seen as a regularized combinatorial Laplacian. In (Dall'Amico et al., 2019) we studied the problem of a good parametrization r of H_r for which the informative eigenvectors are resilient to degree heterogeneity. We briefly report here the main conclusions, based on a local approximation of the neighborhood of each node. See (Dall'Amico et al., 2019, Section 2) for further details.

The argument goes as follows: exploiting sparsity, the graph $\mathcal{G}$ can be shown to locally converge to a Galton-Watson tree in which the offsprings are statistically independent (Dembo et al., 2010; Mossel et al., 2015; Salez, 2011; Mossel et al., 2014; Gulikers et al., 2018).

Fixing A (and thus the degrees d_i), we now perform a Bayesian analysis *on a random allocation of the class labels.* Considering node j, labeled as ℓ_j, as the *root* of the tree, the probability for offspring i to have label ℓ_i reads

$$
\begin{aligned}
\mathbb{P}(\ell_i|\ell_j, A_{ij}=1) &= \frac{\mathbb{P}(\ell_i, \ell_j|A_{ij}=1)}{\mathbb{P}(\ell_j|A_{ij}=1)} = \frac{\iint d\theta_i d\theta_j \mathbb{P}(\ell_i, \ell_j, \theta_i \theta_j|A_{ij}=1)}{\mathbb{P}(\ell_j)} \\
&= \frac{\iint d\theta_i d\theta_j \mathbb{P}(A_{ij}=1|\theta_i, \theta_j, \ell_i, \ell_j)\mathbb{P}(\ell_i)\mathbb{P}(\ell_j)\mathbb{P}(\theta_i)\mathbb{P}(\theta_j)}{Z\pi_{\ell_j}} \\
&= \frac{\pi_{\ell_i} C_{\ell_i, \ell_j}}{c} = \frac{(\Pi C)_{\ell_i, \ell_j}}{c} = \frac{(C\Pi)_{\ell_j, \ell_i}}{c}.
\end{aligned}
\tag{18}
$$

Recalling the definition of the eigenvalue-eigenvector pairs $(\nu_p, \boldsymbol{v}_p)$ of $C\Pi$ (i.e., $C\Pi \boldsymbol{v}_p = \nu_p \boldsymbol{v}_p$), let us define $\boldsymbol{u}_p \in \mathbb{R}^n$ with $u_{p,i} = v_{p,\ell_i}$ the n-dimensional class-wise expansion of $\boldsymbol{v}_p$: this vector $\boldsymbol{u}_p$ is inherently random as the class allocations $\ell_1, \ldots, \ell_n$ are here considered random. Then, for n large, under the limiting tree approximation with conditionally independent offsprings, we take an expectation over the random allocation of the labels different from ℓ_i ($\boldsymbol{\ell}_{\backslash i}$), with A and ℓ_i known:

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{\ell}_{\backslash i}}[(A\boldsymbol{u}_p)_i|A, \ell_i] &= \sum_{j \in \partial i} \mathbb{E}_{\boldsymbol{\ell}_{\backslash i}}[u_{p,j}|A, \ell_i] = \sum_{j \in \partial i} \mathbb{E}_{\boldsymbol{\ell}_{\backslash i}}[v_{p,\ell_j}|A, \ell_i] \approx \sum_{j \in \partial i} \sum_{\ell_j} \mathbb{P}(\ell_j|\ell_i, A_{ij}=1) v_{p,\ell_j} \\
&= \frac{1}{c} \sum_{j \in \partial i} \sum_{\ell_j} (C\Pi)_{\ell_i, \ell_j} v_{p,\ell_j} = \frac{1}{c} \sum_{j \in \partial i} (C\Pi v_p)_{\ell_i} = \frac{d_i}{c} \nu_p v_{p,\ell_i} = d_i \frac{\nu_p}{c} u_{p,i}
\end{aligned}
\tag{19}
$$

where the approximation follows from the fact that conditional independence of the neighbors of a given node only holds asymptotically; in particular, the approximation becomes an equality on a tree rooted at node i. As a consequence of Equation (19), using $H_r = (r^2 - 1) + D - rA$, we find that

$$
\mathbb{E}_{\boldsymbol{\ell}_{\backslash i}}[H_r \boldsymbol{u}_p|A, \ell_i] \approx \left[(r^2 - 1)I_n + D\left(1 - r\frac{\nu_p}{c}\right)\right] \boldsymbol{u}_p.
\tag{20}
$$

In order for this equation to be an approximate eigenvector equation for arbitrary degrees in D, the right hand-side term proportional to D must vanish. That is, one must select

$$
r = \frac{c}{\nu_p} \approx \zeta_p
$$

(this last approximation having been introduced and discussed in Section 2.3.1). This result implies that the eigenvectors $\boldsymbol{x}_p$ defined as $H_{\zeta_p} \boldsymbol{x}_p = 0$, for $1 \leq p \leq k$, correspond to a

noisy version of $\boldsymbol{u}_p$ which is not affected *on average over the class allocation* by the degree distribution. In (Dall'Amico et al., 2019), we confirmed —so long that the average node degree is not too small— that the approximation holds, beyond the average, for every typical realization of the class allocation. In a nutshell, this behavior unfolds from the following remark: denoting $\boldsymbol{\xi}_p$ the "noise" vector satisfying $\boldsymbol{x}_p = \boldsymbol{u}_p + \boldsymbol{\xi}_p$, and thus $H_{\zeta_p}(\boldsymbol{u}_p + \boldsymbol{\xi}_p) = 0$,

$$
\begin{aligned}
\mathrm{Var}_{\boldsymbol{\ell}_{\backslash i}}[(H_{\zeta_p}\boldsymbol{u}_p)_i|A, \ell_i] &= \mathrm{Var}_{\boldsymbol{\ell}_{\backslash i}}[(H_{\zeta_p}\boldsymbol{\xi}_p)_i|A, \ell_i] \\
&= \mathrm{Var}_{\boldsymbol{\ell}_{\backslash i}}[(\zeta_p^2 - 1 + d_i)\xi_{p,i} - \zeta_p(A\boldsymbol{\xi}_{p,i})|A, \ell_i] \\
&\approx O_{d_i}(d_i^2)\mathrm{Var}_{\boldsymbol{\ell}_{\backslash i}}[\xi_{p,i}|A, \ell_i] + O_{d_i}(d_i)\mathrm{Var}_{\boldsymbol{\ell}_{\backslash i}}[\xi_{p,i}|A, \ell_i] \\
&\approx O_{d_i}(d_i^2)\mathrm{Var}_{\boldsymbol{\ell}_{\backslash i}}[\xi_{p,i}|A, \ell_i]
\end{aligned}
\tag{21}
$$

where we recall that the underlying random variable is the random class allocation. There we used the fact that the random variables $\xi_{p,i}$ and $[A\boldsymbol{\xi}_p]_i$ are essentially independent and that the variance of the sum of the d_i asymptotically independent variables $\xi_{p,1}, \ldots, \xi_{p,n}$ grows linearly with d_i (and thus becomes essentially negligible). Now, proceeding as in Equation (19), we can compute $\mathrm{Var}_{\boldsymbol{\ell}_{\backslash i}}[(H_{\zeta_p}\boldsymbol{u}_p)_i|A, \ell_i]$ from a direct calculation of $\mathbb{E}_{\boldsymbol{\ell}_{\backslash i}}[(H_{\zeta_p}\boldsymbol{u}_p)_i^2|A, \ell_i]$ and $\mathbb{E}_{\boldsymbol{\ell}_{\backslash i}}^2[(H_{\zeta_p}\boldsymbol{u}_p)_i|A, \ell_i]$, and we obtain $\mathrm{Var}_{\boldsymbol{\ell}_{\backslash i}}[(H_{\zeta_p}\boldsymbol{u}_p)_i|A, \ell_i] = O_{d_i}(d_i)$.

Combining with Equation (21), we get that, for d_i sufficiently large, $\mathrm{Var}_{\boldsymbol{\ell}_{\backslash i}}[\xi_{p,i}|A, \ell_i] = O_{d_i}(d_i^{-1})$. The vector $\boldsymbol{x}_p$ can therefore be written as the sum of the deterministic information and a noise with amplitude inversely proportional to the square root of the degree, consistently predicting that nodes with higher degrees are easier to classify.

These results are summarized under the form of our last argument.

Element 3 *The eigenvector $\boldsymbol{x}_p$ ($1 \leq p \leq k$), solution to $H_{\zeta_p}\boldsymbol{x}_p = 0$, is a noisy version of the vector $\boldsymbol{u}_p$, defined as $u_{p,i} = v_{p,\ell_i}$, where $C\Pi\boldsymbol{v}_p = \nu_p\boldsymbol{v}_p$, where the noise for entry i scales as $1/\sqrt{d_i}$ for d_i sufficiently large and is zero on average. Consequently, the entries of $\boldsymbol{x}_p$ do not, to first order, depend on the degree distribution but only on the labels.*

Element 3 combined with Elements 1, 2 completes our justification of Claim 1.

2.4 Comments and Performance Comparison

Let us now discuss how Claim 1 can be exploited in practice to obtain an efficient spectral clustering algorithm for sparse graphs with a heterogeneous degree distribution.

From Claim 1 and our detailed analysis in Sections 2.3.1–2.3.3, for all large n with high probability, under Assumption 1, the $k - 1$ Bethe-Hessian matrices $H_{\zeta_2}, \ldots, H_{\zeta_k}$ have an eigenvalue equal to zero, which is isolated. The corresponding eigenvectors $\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ (which are *not* necessarily orthogonal as they correspond to distinct matrices) are all informative in the sense that they are noisy realizations of piece-wise constant versions of the eigenvectors $\boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ of the matrix $C\Pi$ (each "piece" identifying each class). Besides, $\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ are, to first order, resilient to the degree distribution in the network.

This so far assumes that the number of classes k is known, and that all classes satisfy the separability condition of Assumption 1. Yet, in practice, k is generally unknown. The following remark provides a consistent estimator for k, as proposed in (Krzakala et al., 2013; Saade et al., 2014).

Algorithm 1 Community Detection with the Bethe-Hessian

1: **Input** : adjacency matrix of undirected graph $\mathcal{G}$
2: Estimate $c = \frac{1}{n} \sum_i d_i$ and $\Phi = \frac{1}{nc} \sum_i d_i^2$
3: Detect number of classes $\hat{k}$ according to Remark 1
4: **for** $2 \leq p \leq \hat{k}$ **do**
5: $\quad \zeta_p \leftarrow r$ such that $s_p^{\uparrow}(H_r) = 0$ on $1 < r \leq \sqrt{c\Phi}$
6: $\quad X_{\bullet,p} \leftarrow \boldsymbol{x}_p$ such that $H_{\zeta_p} \boldsymbol{x}_p = 0$
7: Estimate community labels $\hat{\boldsymbol{\ell}}$ from the node embedding $\boldsymbol{X} = [\boldsymbol{X}_{\bullet,1}, \ldots, \boldsymbol{X}_{\bullet,\hat{k}}]$.
8: **return** Estimated number $\hat{k}$ of communities and label vector $\hat{\boldsymbol{\ell}}$.

Remark 1 (Estimation of the number of classes) *Under Assumption 1, for all large n with high probability, $\hat{k}_{\mathrm{B}} = \hat{k}_{\mathrm{H}} = k$ where*

$$\hat{k}_{\mathrm{B}} = \left| \left\{ i, \ \mathscr{R}[s_i(B)] > \sqrt{c\Phi} \right\} \right|$$

$$\hat{k}_{\mathrm{H}} = \left| \left\{ i, \ s_i(H_r)|_{r=\sqrt{c\Phi}} < 0 \right\} \right|,$$

where $\mathscr{R}$ indicates the real part.

The remark is justified by the fact that $r = \sqrt{c\Phi}$ is the (limiting) right edge of the bulk of the non-backtracking matrix B and that the eigenvalues of B whose real part exceeds $\sqrt{c\Phi}$ (in the limit $n \to \infty$) are the isolated largest real eigenvalues of B. These are mapped, according to Figure 4 and the discussion of Section 2.3.2, to the number of communities. Likewise, from the connection between H_r and B, for $r = \sqrt{c\Phi}$, the negative eigenvalues of H_r are one-to-one mapped to the largest real eigenvalues of B.

The arguments above naturally lead to Algorithm 1 for community detection using the Bethe-Hessian matrix. Algorithm 1 is a meta-algorithm, written for the user readability. Section 4 will provide an efficient implementation (Algorithm 2) of Algorithm 1 accounting for deeper algorithmic considerations.

Applying Algorithm 1, Figure 5 compares the overlap performance (defined in Equation (22)), achieved by Algorithm 1 versus competing methods for synthetic DC-SBM graphs as a function of the hardness of the problem α.

$$\text{Overlap} = \max_{\hat{\boldsymbol{\ell}} \in \mathcal{P}} \frac{1}{1 - \frac{1}{k}} \left(\sum_{i=1}^{n} \delta(\hat{\ell}_i, \ell_i) - \frac{1}{k} \right), \tag{22}$$

where $\hat{\boldsymbol{\ell}}$ is the estimate of the label eigenvector and $\mathcal{P}$ the set of the permutations of $\hat{\boldsymbol{\ell}}$.

On the left hand-side of Figure 5 is depicted a symmetric two-class scenario, for which α is defined as per Equation (2) and community reconstruction is asymptotically feasible if and only if $\alpha > \alpha_c$. On the right hand-side a more involved multiple class scenario is devised: the matrix C is obtained here by drawing random Gaussian numbers with mean c_{out} and variance $f = c_{\text{out}}/k$, before being symmetrized and tuning the diagonal elements to guarantee $C\Pi\boldsymbol{1}_k = c\boldsymbol{1}_k$. In this case, we still define $\alpha = 2(c - c_{\text{out}})/\sqrt{c}$ but α_c now only represents the transition conjectured in (Decelle et al., 2011) of the planted partition model
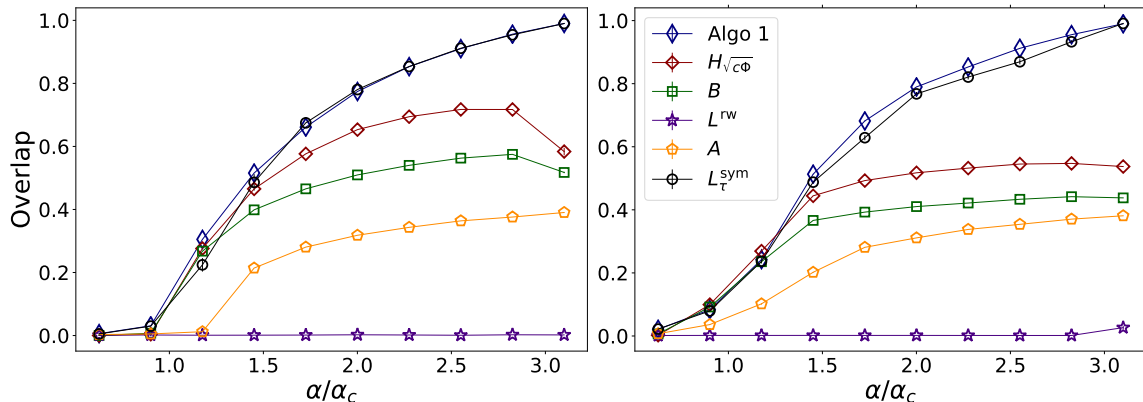
Figure 5: Overlap comparison as a function of the hardness α of the detection problem. Blue sharp diamonds are the result given by Algorithm 1, red diamonds the algorithm using the Bethe-Hessian of (Saade et al., 2014), green squares are for the non-backtracking of (Krzakala et al., 2013), yellow pentagons the adjacency matrix, purple stars the random walk Laplacian, black dots are the algorithm of (Qin and Rohe, 2013). For both graphs $n = 50.000$, $\theta_i \sim [\mathcal{U}(3,10)]^4$, $\boldsymbol{\pi} \propto \mathbf{1}_k$, $c = 5$. (Left): $k = 2$, $c_{\text{out}} : 0.05 \to 4$. (Right): $k = 5$, $\pi_p \sim \mathcal{N}(k^{-1}, (4k)^{-1})$, $C_{a>b} \sim \mathcal{N}(c_{\text{out}}, c_{\text{out}}k^{-1})$, $c_{\text{out}} : 0.05 \to 4$. $\alpha := 2(c - c_{\text{out}})/\sqrt{c}$ and $\alpha_c = 2/\sqrt{\Phi}$. Averages over 10 samples.

for which $f = 0$. In practice, in our simulation with $f \neq 0$, there are multiple transitions that are close (but in general different) to α_c.

In both cases, Algorithm 1 outperforms the competing algorithms. For $k = 2$, as $\alpha \to \alpha_c$, Algorithm 1, the algorithm of (Saade et al., 2014) based on $H_{\sqrt{c\Phi}}$ and the one of (Krzakala et al., 2013) based on B give essentially the same result, which confirms that they are indeed equivalent at the phase transition. For easier detection problems though, except for the algorithm of (Qin and Rohe, 2013) which is only slightly less accurate, the performance of all methods is largely improved by Algorithm 1. Interestingly, it can be shown here that $|\sqrt{c\Phi} - \zeta_2| < |\nu_2\Phi - \zeta_2|$, giving an intuition on why the Bethe-Hessian method of (Saade et al., 2014) performs better than the non-backtracking approach of (Krzakala et al., 2013). As for the standard spectral clustering algorithm which exploits the dominant eigenvectors of A, it can only perform non trivial community reconstruction for α far beyond the threshold $\alpha = \alpha_c$. The algorithm of (Shi and Malik, 2000), based on the popular random walk Laplacian L^{rw}, is here incapable of making any non-trivial reconstruction for the considered set of parameters, suggesting that its dominant eigenvectors are not informative.

This last point is quite interesting and raises the possibility that the informative eigenvectors of L^{rw} are possibly lost in the bulk. The next section will confirm this affirmation and make the statement more precise.

3. Relation Between H_r and the Regularized Random Walk Laplacian L_τ^{rw}

In this section we show that there is a strong connection between the Bethe-Hessian matrix and the regularized random walk Laplacian matrix $L_\tau^{\mathrm{rw}} = D_\tau^{-1}A$, where $D_\tau = D + \tau I_n$. We will specifically prove that, when τ is carefully chosen, the matrix L_τ^{rw} has isolated informative eigenvalues in its largest positions. Remarking that L_τ^{rw} has the same eigenvalues as the matrix $L_\tau^{\mathrm{sym}} = D_\tau^{-1/2}AD_\tau^{-1/2}$, we also find a strong connection between L_τ^{rw} and the community detection method of (Qin and Rohe, 2013).

The mapping between H_r and L_τ^{rw} unfolds from the following basic remark: for $\zeta_p > 1$ (as defined in Equation (5)),

$$[(\zeta_p^2 - 1)I_n + D - \zeta_p A]\boldsymbol{x}_p = 0 \tag{23}$$

$$\Leftrightarrow [D + (\zeta_p^2 - 1)I_n]^{-1}A\boldsymbol{x}_p = \frac{1}{\zeta_p}\boldsymbol{x}_p \tag{24}$$

inducing a natural mapping between the Bethe-Hessian matrix at $r = \zeta_p$ to the regularized random walk Laplacian at $\tau = \zeta_p^2 - 1$. In particular, the vector $\boldsymbol{x}_p$ solution of Equation (23) coincides with the eigenvector of $L_{\zeta_p^2-1}^{\mathrm{rw}}$ associated with the eigenvalue $1/\zeta_p$.

In the following this relation is further analyzed and will be used to provide an alternative method to detect communities down to the detectability threshold.

3.1 Main Result

The spectral methods of (Shi and Malik, 2000; Ng et al., 2002) do not lead to good partitions in the sparse regime, as evidenced in Figure 5. The reason, which we briefly commented in Section 1, is that the bulk of uninformative eigenvalues of both L^{rw} and L^{sym} undergo a certain spreading in the sparse regime (see Figure 1B) which "swallows" the informative eigenvalues within the bulk. This is not to say that informative eigenvectors vanish: rather, these eigenvectors are associated to *non-isolated* eigenvalues; this was already observed in (Joseph and Yu, 2016) for the benchmark network *Political blogs* (Adamic and Glance, 2005). This has two major negative consequences: (i) it may be practically infeasible (especially as the clustering task is more difficult) to identify the correct informative eigenvalue, and (ii) since the eigenvectors associated to close-by eigenvalues (in the bulk, the typical distance between consecutive eigenvalues is $O_n(1/n)$) tend to "spread" across the eigenvectors of these neighboring eigenvalues, not a single eigenvector but a collection of neighboring eigenvectors need to be considered. This is clearly deleterious to spectral clustering.

This problem is partially solved through regularization through a parameter τ (added to D as $D_\tau = D + \tau I_n$): it was indeed shown in a series of related works that regularization helps clustering in sparse networks (Lei and Rinaldo, 2015; Joseph and Yu, 2016; Qin and Rohe, 2013). In all these works, the results obtained (or inferred) are not straightforwardly applicable to the sparse regime. Furthermore, according to the analysis performed in these works (however only far from the transition point), large values of τ would seem preferable; however, in practice, it is rather observed that small values entail better performances.

We answer here to the question of why small parametrizations τ should be preferred to large τ, as well as which is the smallest value of τ for which we can guarantee the existence of isolated informative eigenvectors of L_τ^{rw}, down to the detectability threshold. Besides,

in accordance to the previous sections, we determine a value for τ which is optimal in its (i) ensuring the existence of these informative eigenvectors all the way to the detectability threshold, and most importantly (ii) which is resilient to degree heterogeneity in the graph. Formally, the result is formulated as follows.

Claim 2 *Consider the graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, built on a sparse DC-SBM as per Equation (1) with k communities. Let ζ_p for $2 \leq p \leq k$ be defined as per Equation (5) (imposing $\zeta_1 = 1$) and $\tau \in \mathbb{R}$ be such that $\zeta_p^2 - 1 \leq \tau \leq c\Phi - 1$. Then, under Assumption 1, for all large n with high probability, the p largest eigenvalues of the matrix L_τ^{rw} are isolated. In particular,*

$$s_p^\downarrow(L_{\zeta_p^2-1}^{\mathrm{rw}}) = \frac{1}{\zeta_p}.$$

Note that the proposition is not an obvious consequence of the equivalence between Equation (23) and Equation (24) as it is not clear that the eigenvalue $1/\zeta_p$ of $L_{\zeta_p^2-1}^{\mathrm{rw}}$ corresponds to the p-th largest and that it remains isolated (as is zero in the spectrum of H_{ζ_p}).

Since Claim 2 asserts that the eigenvalue $1/\zeta_p$ of $L_{\zeta_p^2-1}^{\mathrm{rw}}$ is the p-th largest and is isolated, it can be clearly identified for all finite but large n: its corresponding informative eigenvector $\boldsymbol{x}_p$ (from Equation 24) can thus not be confused with other eigenvectors. This eigenvector $\boldsymbol{x}_p$ is also the solution to $H_{\zeta_p}\boldsymbol{x}_p = 0$ (from Equation (23)); its properties have been extensively discussed in Section 2: it is in particular asymptotically insensitive to the degrees of the graph. Intuitively, by picking τ away from $\zeta_p^2 - 1$, the entries of the p-th eigenvector are likely to be more polluted by the degrees of the network. This suggests that large values of τ (as studied by the authors of Joseph and Yu, 2016) are likely to lead to sub-optimal partitions.

Claim 2 further asserts that, for any τ in the interval $[\zeta_p^2 - 1, c\Phi - 1]$, the p dominant eigenvalues of L_τ^{rw} are isolated. Since in addition $\zeta_p \leq \zeta_k \leq \sqrt{c\Phi}$, regardless of the hardness of the detection problem, the k largest eigenvalues of $L_{c\Phi-1}^{\mathrm{rw}}$ (i.e., $L_{r^2-1}^{\mathrm{rw}}$ for $r = \sqrt{c\Phi}$) must be isolated. Note in passing that, in (Qin and Rohe, 2013), the authors propose the regularization $\tau = \bar{d} = \frac{1}{n}\sum_i d_i$ from a heuristic intuition. While suboptimal according to the claim, this choice is somewhat meaningful as τ must indeed essentially grow with the average degree $\bar{d} \approx c$.

Besides, as a corollary, this provides a convenient alternative method to estimate the number of communities, based on the regularized Laplacian matrix.

Remark 2 (Second estimation of the number of classes) *As a direct consequence of Claim 2, we have that, for all large n with high probability $k = \hat{k}_{\mathrm{B}} = \hat{k}_{\mathrm{H}} = \hat{k}_{\mathrm{L}}$, where*

$$\hat{k}_{\mathrm{L}} = \left|\left\{i \ : \ s_i(L_\tau^{\mathrm{rw}})|_{\tau=c\Phi-1} \geq \frac{1}{\sqrt{c\Phi}}\right\}\right|.$$

Note that, since $S(L_\tau^{\mathrm{rw}}) = S(L_\tau^{\mathrm{sym}})$, this calculation can be performed on a symmetric matrix, bringing gain in computational efficiency.

The technical supporting elements of Claim 2 can be found in Appendix E. In a nutshell, letting $\tau(r) = -s_p^\uparrow(D - rA)$, we show that there is a one-to-one map between the isolated eigenvalues of $L_{\tau(r)}^{\mathrm{rw}}$ and of H_r, only reversed in order (the smallest of H_r are mapped to the

largest of $L^{\text{rw}}_{\tau(r)}$). This is particularly valid for $r = \zeta_p$ for which $-s^{\uparrow}_p(D - \zeta_p A) = \zeta_p^2 - 1$. We further show that the function $\tau(r)$ is bijective, thereby extending the results of Claim 1 for $r \in [\zeta_p, \sqrt{c\Phi}]$ into $\tau \in [\zeta_p^2 - 1, c\Phi - 1]$.

3.2 Side Comments

This section provides further interpretations and insights of the results introduced in the previous chapters. Firstly, we relate Algorithm 1 to other commonly adopted methods for spectral clustering and secondly we discuss how our result can be extended in the presence of disassortative communities.

3.2.1 Connection Among the Spectral Algorithms Based on $B, H_r, L^{\text{rw}}_\tau, L^{\text{rw}}, L$

We showed in the previous sections the deep connection between the belief propagation equations and the regularized Laplacian matrix L^{rw}_τ, by successively passing through the non-backtracking B and the Bethe-Hessian matrices H_r, so far treated in parallel (and with different tools) in the literature. From a practical perspective, we notably pointed that adding the regularization τI_n to the degree matrix D (i) favors efficient clustering in sparse networks but (ii) is optimally tuned for τ taken as a function of the hardness of the detection task (explicitly, for $k = 2$, a function of $\alpha = (c_{\text{in}} - c_{\text{out}})/\sqrt{c}$). In particular, for more challenging clustering tasks, larger values of r in H_r and τ in L^{rw}_τ should be employed. On the opposite, in the limit of trivially simple clustering, $r \to 1$ (so that $H_r \to D - A$, the standard Laplacian) and $\tau \to 0$ (so $L^{\text{rw}}_\tau \to L^{\text{rw}}$, the standard random walk Laplacian). This suggests that the classical Laplacian matrices are still "optimal" in the sparse regime, yet only for easy clustering tasks, i.e., possibly far beyond the detectability threshold.

Most importantly, while the aforementioned "difficulty" of a clustering task is of course not accessible to the practitioner, we showed that the optimal tuning of the hyperparameters r and τ can be practically obtained by retrieving isolated eigenvalues in the spectra of H_r, B, or L^{rw}_τ (through the fundamental quantities $\zeta_2, \ldots, \zeta_k$). In passing, the relative distance of the ζ_p's to the bulk of non-informative eigenvalues is a further clue for the practitioner of the level of confidence of the ultimate clustering result.

3.2.2 Extension to Disassortative Communities

Another remark concerns the possible existence of disassortative communities in the graph, i.e., groups of nodes which are identified as a class because they repel (rather than attract) each other. As a concrete example one may think of the vertices of a graph as the words contained in a text with edges if two words appear next to each other: on this graph, adjectives and nouns represent two disassortative communities (Newman and Girvan, 2004). Under a two-class DC-SBM model, a disassortative network can be easily generated by imposing $c_{\text{out}} > c_{\text{in}}$. In this case, the second largest eigenvalue of $C\Pi$ will therefore be negative and equal to $\nu_2 = (c_{\text{in}} - c_{\text{out}})/2 < 0$.

The problem of detection of two disassortative communities can be addressed both with the non-backtracking and the Bethe Hessian matrices (as shown in Krzakala et al., 2013; Saade et al., 2014). In particular, in the presence of two disassortative communities, $s_2^{\downarrow|\cdot|}(B) < 0$, where we recall $s_p^{\downarrow|\cdot|}(M)$ indicates the p-th largest eigenvalues in modulus. Similarly, the
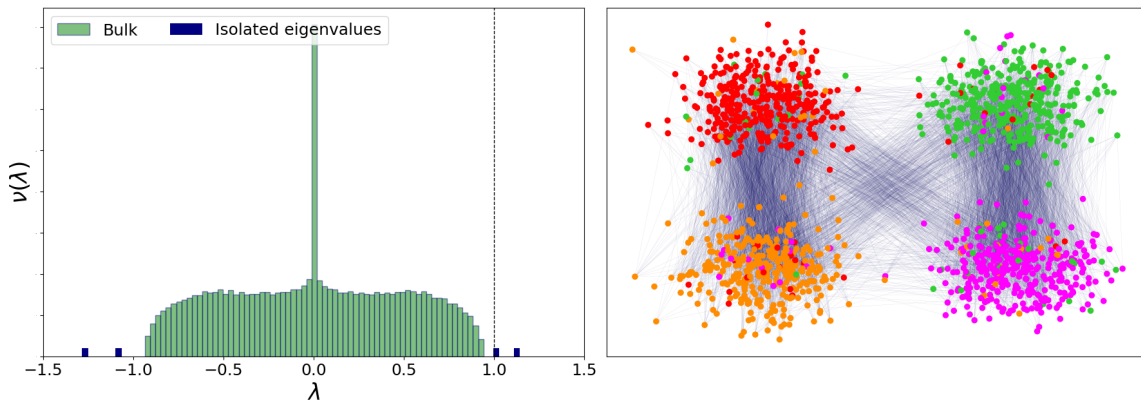
Figure 6: (Left) Spectrum of $\zeta_4 L^{\mathrm{rw}}_{\zeta_4^2-1}$, $n = 5000$, with $c_{\mathrm{in}} = 4$, $c_{\mathrm{out}} = 0.5$, $\theta_i \sim [\mathcal{U}(3,5)]^4$ with C generated according Equation (26). (Right) Realization of $\mathcal{G}$ from the SBM with C defined as per Equation (26), with $c_{\mathrm{in}} = 5$ and $c_{\mathrm{out}} = 1$, $n = 1500$. The position of the nodes is assigned according to the true labels as $\mathcal{N}(\mu_{\ell_i}, I_2)$ and the colors are the output of Algorithm 1 on $\mathcal{G}$.

matrix H_r can detect two disassortative communities for $r < 0$, becoming a deformed version (equal in the limit of easy clustering) of the signless Laplacian $D + A$, frequently used to study bipartite graphs (Cvetković et al., 2007).

While defining communities when both assortative and disassortative elements are present is debatable, at least the result of Claim 2 can be extended to the case where the matrix $C\Pi$ has possibly both positive and negative eigenvalues, as follows.

Remark 3 (Eigenvalues of $C\Pi$ of arbitrary sign) *Assume, unlike stipulated in Assumption 1, that the sign of the eigenvalues ν_p of $C\Pi$ is arbitrary (except for $\nu_1 = c$), but that $|\nu_p| > \sqrt{\frac{c}{\Phi}}$ in absolute value. We sort the eigenvalues ν_p according to their modulus as $\nu_1 > |\nu_2| \geq \ldots \geq |\nu_k|$. In this setting, Claim 2 generalizes by stating that the p eigenvalues of L^{rw}_τ with largest modulus are isolated and that*

$$s_p^{\downarrow|\cdot|}(L^{\mathrm{rw}}_{\zeta_p^2-1}) = \frac{1}{\zeta_p} = \frac{\nu_p}{c} + o_n(1). \tag{25}$$

As suggested above, while the mathematical result is well defined for $\nu_p < 0$, the interpretation of the negative eigenvalues of $C\Pi$ is not straightforward for $k > 2$ and the very definition of a *community* might not be evident. As an example, the right frame of Figure 6 displays a network designed from the SBM with $k = 4$, $\Pi \propto I_k$ and C defined as:

$$C = \begin{pmatrix} c_{\mathrm{in}} & c_{\mathrm{out}} \\ c_{\mathrm{out}} & c_{\mathrm{in}} \end{pmatrix} \otimes \begin{pmatrix} c_{\mathrm{out}} & c_{\mathrm{in}} \\ c_{\mathrm{in}} & c_{\mathrm{out}} \end{pmatrix} \tag{26}$$

with '$\otimes$' the Kronecker product. This graph has a *hierarchical* structure: it can be divided into two assortative communities, each of them composed of two disassortative communities.

In Figure 6 (right) the position of each node is assigned according to its true label, while the color is assigned according to the output of k-means on the vectors $\boldsymbol{x}_p$ of $L^{\mathrm{rw}}_{\zeta_p^2-1}$ with largest modulus. We can observe that our algorithm performs well also in this particular case. For the sake of clarity, we close here the discussion on the generalization to cases where ν_p may be negative and, unless otherwise stated, we are thus once again under Assumption 1.

4. Implementation of the Algorithm

Based on the various discussions and results from the previous section, we now formally introduce our proposed algorithm along with pragmatic discussions on implementation cost, optimization and robustness to real-world network configurations.

4.1 Applicability to Real Networks

As opposed to the sparse stochastic block model, the DC-SBM model of Equation (1) accounts both for the sparsity and heterogeneity of real networks. Yet, it does not capture other fundamental aspects of real graphs. For instance, the presence of many triangle in social networks (Holland and Leinhardt, 1971) and more generally of short loops may invalidate the local tree-like approximation that underlies our theoretical findings. Nevertheless, its usability on real data sets suggest that some properties of B are valid for various graph topologies. In particular, the following properties of B seem to hold in general:

- All complex eigenvalues come in pairs of complex conjugates and most of them are bounded by a circle on the complex plane of radius $r_{\mathrm{radius}} \approx \sqrt{\rho(B)}$. This statement originates both from empirical observations and from heuristic arguments on the asymptotic density of the eigenvalues of B dicussed in (Krzakala et al., 2013).
- The number of real eigenvalues of B, different from $\{-1, 0, 1, \rho(B)\}$, is even. Half of the eigenvalues are larger in modulus than r_{radius}, while the other half lies between 1 and r_{radius}. All of them are isolated.

It is easy to verify that the two earlier properties hold for *any* graph topology if $D = cI_n$ and can be extended to the case in which $D \approx cI_n$ as suggested in (Coste and Zhu, 2021). Given the connection between B and the Bethe Hessian matrix H_r, from these two points, we can claim that the steps of Algorithm 1 are all well defined on arbitrary graphs.

Some properties of B are however likely model-dependent and thus not resilient to arbitrary graphs, such as

- $\rho(B)$ may be far from $\frac{\sum_i d_i^2}{\sum_i d_i}$, which is a natural estimator for $c\Phi$ in the DC-SBM model.
- For an arbitrary network, ζ_p (as defined in Claim 1) may be far from $\frac{\rho(B)}{s^{\downarrow}_{R,p}(B)}$ in general.

These observations impose that, when devising a spectral clustering algorithm adapted to real networks, these purely DC-SBM considerations should not be exploited.

A further specificity of real graphs is their possibility to be inherently made of disjoint components. We have instead so far worked with the assumption that $\mathcal{G}$ has a giant component and that the communities are contained within the giant component. In practice, communities may live in disconnected subgraphs (in a two-class DC-SBM, this would correspond to setting

$c_{\text{out}} = 0$). In this case, one can perform a two-stage clustering. First the connected components are detected, looking into the eigenvectors with zero eigenvalue of $H_{\zeta_1} = D - A$. Afterwards, each connected component is treated independently. We will thus consider, without loss of generality, that the real graphs which we will consider are connected.

4.2 Estimation of k

The problem of estimating the number of communities in an unsupervised manner is in general non-trivial. Methods based on the non-backtracking rather than Bethe-Hessian matrix have been studied and exploited to efficiently recover communities regardless of the generative model (Le and Levina, 2015). In the course of our argumentation, we suggested different ways to estimate the number of communities, which are asymptotically equivalent (recall Remark 1 and Remark 2):

1. $\hat{k}_{\text{B}} = |\{s_{R,i}(B) > \sqrt{\rho(B)}\}|$

2. $\hat{k}_{\text{H}} = |\{s_i\left(H_{\sqrt{\rho(B)}}\right) < 0\}|$

3. $\hat{k}_{\text{L}} = \left|\left\{s_i\left(L^{\text{sym}}_{\rho(B)-1}\right) > \frac{1}{\sqrt{\rho(B)}}\right\}\right|$.

From a computational standpoint, note that the eigenvalues of the matrix B different from ± 1 are the same as the eigenvalues of the matrix $B' \in \mathbb{R}^{2n \times 2n}$ (Krzakala et al., 2013; Coste and Zhu, 2021), defined as

$$B' = \begin{pmatrix} A & I_n - D \\ I_n & 0 \end{pmatrix} \tag{27}$$

so all computation involving the eigenvalues of B can be performed in an efficient manner on the matrix B'. All three estimators above require the value of $\rho(B)$ that can be obtained efficiently by direct computation. But not all estimates perform the same in practice. Estimator 1 is in particular not efficient as not only real but also complex eigenvalues of B need be evaluated, which sensibly slows down the algorithm. Estimators 2 and 3 do not suffer this limitation. Estimator 3 is nonetheless preferred as estimating eigenvalues with largest, rather than smallest, algebraic value can in general be performed more efficiently. Furthermore, when the matrix $C\Pi$ has at the same time both positive and negative eigenvalues, all communities can be detected using the same matrix $L^{\text{sym}}_{\rho(B)-1}$, as opposed to the Bethe-Hessian for which one needs to consider both $H_{\sqrt{\rho(B)}}$ and $H_{-\sqrt{\rho(B)}}$. Consequently, the estimator $\hat{k}_{\text{L}}$ is the one that will be adopted in our final Algorithm 2. The complexity of estimating $\hat{k}$ scales as $O(n\hat{k}^3)$, as computing the p largest eigenvalues of a sparse matrix of size $n \times n$ (that is, containing $O(n)$ non-null elements) costs $O(np^2)$ with state-of-the-art methods such as restarted Arnoldi methods (see for instance Saad, 2011).

Subroutine 1 details how to estimate[6] $\hat{k}$ and it will be referenced in Algorithm 2.

---

6. Subroutine 1 is actually a naïve implementation to estimate the number of communities, as it requires to compute several times the same eigenvectors. In the CoDeBetHe.jl package a more efficient implementation is proposed, based random projections, that allows to efficiently estimate $\hat{k}$.

Subroutine 1 `estimate_number_of_classes`

1: **Input** : adjacency matrix of a connected, undirected graph $\mathcal{G}$
2: Compute $\rho(B) \leftarrow s_1^{\downarrow|\cdot|}(B')$, with B' defined in Equation (27)
3: $j \leftarrow 1$
4: **while** $v > 0$ **do**
5: $\quad v \leftarrow s_{j+1}^{\downarrow}(L_{\rho(B)-1}^{\mathrm{sym}}) - \frac{1}{\sqrt{\rho(B)}}$
6: $\quad j \leftarrow j+1$
7: $\hat{k} \leftarrow j$
8: **return** Estimated number $\hat{k}$ of communities.

Of course, as per Assumption 1 in the DC-SBM setting, since for all $1 \leq p \leq k$, $\nu_p > \sqrt{\frac{c}{\Phi}}$, all communities were claimed "visible". In practice though, this becomes a stringent condition. If, in particular, only $\kappa < k$ eigenvalues of $C\Pi$ exceed the threshold, fewer eigenvectors will be exploited and fewer classes will be looked for by the algorithm (as is also the case of the algorithms in Krzakala et al., 2013; Saade et al., 2014). In particular, $k - \kappa$ communities might be merged to close-by communities or spread across the κ detected communities.

4.3 Estimation of $\{\zeta_p\}$

From Section 2.3.1, the values of the ζ_p for $1 \leq p \leq k$ can be estimated as

$$\zeta_p = \frac{c}{\nu_p} + o_n(1) = \frac{\rho(B)}{s_p^{\downarrow}(B)} + o_n(1). \tag{28}$$

This estimation of ζ_p (via B') is computationally efficient but may be quite inaccurate for graphs not generated from the DC-SBM model. Conversely, a naive line search for $\zeta_p \in (1, \sqrt{\rho(B)})$ satisfying $s_p^{\uparrow}(H_{\zeta_p}) = 0$ is computationally inefficient.

We propose here a faster method, motivated by a Courant-Fischer theorem argument. The details and proof of convergence of the proposed method are given in Appendix F.

In a few words, starting from an initial guess r_0, we devise an iterative sequence $r_0, r_1, \ldots$, such that $r_t \to \zeta_p$. In Appendix F we show that the convergence is guaranteed when setting $r_0 = \zeta_{p+1}$, under the convention $\zeta_{k+1} = \sqrt{\rho(B)}$. The values of ζ_p are then estimated from the largest, ζ_k, to the smallest, ζ_2.

The algorithm builds on two parts: (i) the computation of a $p \times p$ matrix obtained from the eigenvectors of H_{r_t} with computational cost of $O(np^2)$, (ii) a subsequent line-search using this matrix, with computational cost $O(p^3)$. The advantage of this method is that the line search (which requires many iterations) is computationally cheap, while the most expensive part of the algorithm needs to be performed much fewer times with respect to the greedy line-search to obtain the same accuracy. The total complexity of the algorithm needed to compute the vector $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_k)^T$, scales as $O(nk^3)$.

The proposed algorithm is described in the following subroutine. We indicate with $X_r \in \mathbb{R}^{n \times p}$ the matrix containing in its columns the p smallest eigenvectors of H_r and $S_r \in \mathbb{R}^{p \times p}$ the diagonal matrix with the corresponding eigenvalues. For further details, the reader is referred to Appendix F.

Subroutine 2 `compute_ζ`

1: **Input** : adjacency matrix of a connected, undirected graph $\mathcal{G}$, number of classes k

2: $r \leftarrow \sqrt{s_1^{\downarrow,|\cdot|}(B')}$, where B' is defined in Equation (27)

3: $p \leftarrow k$

4: **while** $p > 1$ **do**

5:     **repeat**

6:        Compute $S_r = \mathrm{diag}\left(s_1^{\uparrow}(H_r), \ldots, s_p^{\uparrow}(H_r)\right)$ and $X_r \in \mathbb{R}^{n \times p}$, with $H_r X_r = X_r S_r$

7:        $\Lambda_r \leftarrow X_r^T D X_r \in \mathbb{R}^{p \times p}$ ($D \in \mathbb{N}^{n \times n}$, the diagonal degree matrix)

8:        $r^* \leftarrow r' : s_1^{\downarrow}(r' S_r + (r - r')\Lambda_r) = (r - r')(1 + r r')$, with line-search on $r' \in (1, r)$

9:        $r \leftarrow r^*$

10:     **until** convergence

11:     $\delta \leftarrow$ multiplicity of $s_p^{\uparrow}(H_r) = 0$

12:     Set $\zeta_p, \zeta_{p-1}, \cdots, \zeta_{p-\delta+1}$ to r

13:     $p \leftarrow p - \delta$

    return Vector $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, \ldots, \zeta_k)^T$.

Note that, although the subroutine `compute_ζ` only outputs the vector $\boldsymbol{\zeta}$, it can also be used to directly compute the informative eigenvectors $\{\boldsymbol{x}_p\}$.

Figure 7 provides a typical output of the computation of $\boldsymbol{\zeta}$ and confirms the accuracy of the proposed algorithm. On the top line the algorithm is tested on a network created from the DC-SBM model, for which $\rho(B)/s_{R,p}^{\downarrow}(B)$ is a valid estimator for ζ_p. The horizontal line indicates $\sqrt{\rho(B)} = \sqrt{c\Phi}$ which is the upper bound of ζ_p. In our simulations, C is generated randomly (see the caption of Figure 4), for the first two points we obtained $\nu_3 < \sqrt{c\Phi}$, invalidating Assumption 1. In this case we see that $c/\nu_p > \sqrt{c\Phi}$ and the corresponding estimated value of ζ_p saturates at $\sqrt{c\Phi}$. On the contrary, whenever Assumption 1 is verified, the estimate of ζ_p is correct.

In the bottom line we compare the two methods to estimate ζ_p on the three real networks that clearly show that these two methods are different on graphs not generated from the DC-SBM. Given that Subroutine 2 provides a direct computation of the eigenvalues of B, it should be preferred.

For illustrative purposes, we tested the execution time of Subroutine 2 for a SBM network with $k = 4$ classes of equal size, $n = 10^5$, $c_{\text{in}} = 26.20$ on the diagonal elements of C and $c_{\text{out}} = 4.60$ for the off diagonal elements. The values of ζ_p were computed to machine precision in approximately 40 seconds on a standard laptop using our Python implementation, while in approximately 5 seconds with the CoDeBetHe.jl Julia implementation. The complexity of this subroutine scales linearly with n, and thus can be applied to large networks but cubically with respect to k (and not quadratically as usual in the spectral clustering context) decreasing the computational efficiency when a rather large number of classes is present.

4.4 Projection of the Embedded Points on a Hyper-Sphere

The study performed so far identifies the presence of k informative eigenvalues and describes the content of the associated k eigenvectors $\boldsymbol{X} \equiv [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k] \in \mathbb{R}^{n \times k}$ with $\boldsymbol{x}_p$ the eigenvector
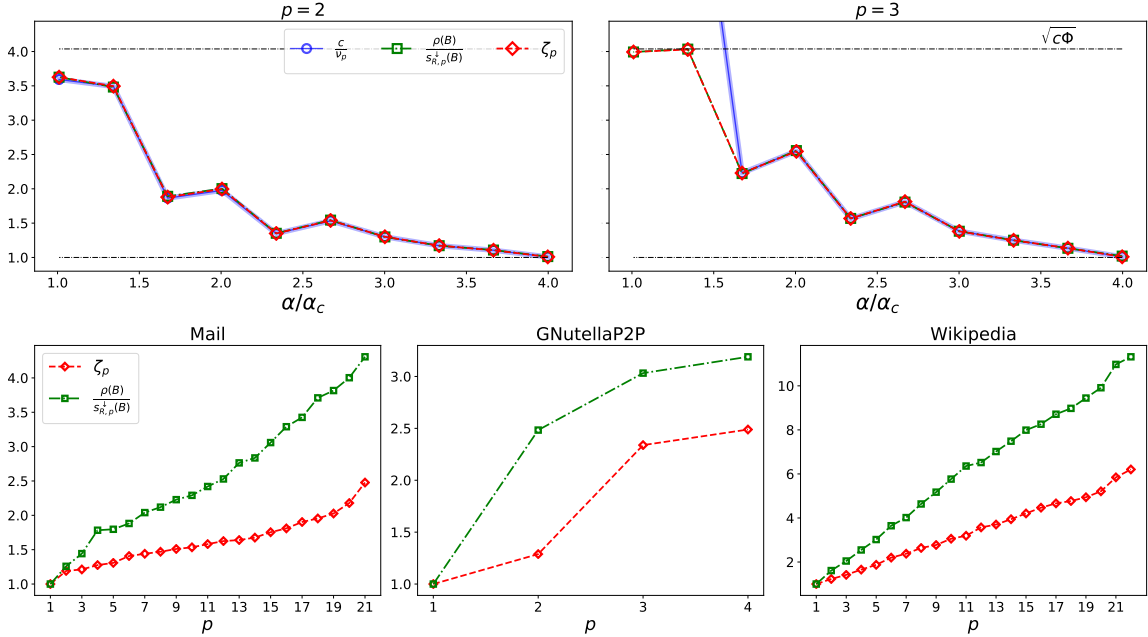
Figure 7: (Top: DC-SBM setting) Exact value of c/ν_p ($p = 2$ on the left and $p = 3$ on the right) in blue, compared to the estimate of ζ_p obtained using the dominant eigenvalues of B (green) and its direct computation using Subroutine 2 (red), for problems of different hardness. For this simulation, $n = 50.000$, $k = 3$, $\theta_i \sim [\mathcal{U}(3, 10)]^3$, $c = 10$. For each point $\pi_i \sim \mathcal{N}(1/k, 1/2k)$ and the off-diagonal elements of C are distributed as $\sim \mathcal{N}(c_{\text{out}}, c_{\text{out}}/k)$ and $c_{\text{out}} : 0.1 \to 7.5$. (Bottom: real networks) Estimate of the values of ζ_p as function of p, computed as $\rho(B)/s_p^{\downarrow}(B)$ (green squares) vs the direct computation of ζ_p as the output of Subroutine 2 (red diamonds) on the three real networks taken from (Leskovec and Krevl, 2014). The value of $\hat{k}$ is estimated according to Subroutine 1.

associated to the p-th smallest eigenvalue of H_{ζ_p} (in particular, the vectors $\boldsymbol{x}_p$'s need not be orthogonal). The rows $\boldsymbol{X}_{1,\bullet}, \dots, \boldsymbol{X}_{n,\bullet} \in \mathbb{R}^k$ of the matrix $\boldsymbol{X}$ form a k-dimensional feature for every node, which are used in a last small-dimensional clustering step, usually employing the *k-means* algorithm. The fact that *k-means* is particularly efficient when the low-dimensional clusters are quite "isotropic" strongly motivates the need for the entries of the vectors $\boldsymbol{x}_p$ not to be affected by the node degrees (which would otherwise spread the clusters unevenly).

Yet, to further tackle residual degree dependence, a classical method, prior to *k-means*, consists in normalizing all vectors $\boldsymbol{X}_{i,\bullet}$ to $\|\boldsymbol{X}_{i,\bullet}\| = 1$ (this is Step 4 of Algorithm 0) . This method is motivated by the assumption that the degree dependence in each X_{ij} is separable from the label dependence, a fact that is verified in sufficiently dense DC-SBM networks (Jin et al., 2015) and to some extent also in sparser graphs (Qin and Rohe, 2013). Besides, under this normalization, the *k-means* algorithm is restricted to the unitary hypersphere, improving its convergence to the genuine solution, especially when in presence of many communities.

Algorithm 2 Community Detection in sparse and heterogeneous graphs

---

1: **Input** : adjacency matrix of a connected, undirected graph $\mathcal{G}$
2: $\hat{k} \leftarrow$ `estimate_number_of_classes` (Subroutine 1)
3: $\boldsymbol{\zeta} \leftarrow$ `compute_ζ` (Subroutine 2)
4: Initialize $\boldsymbol{X} \in \mathbb{R}^{n \times \hat{k}}$
5: **for** $p = 1 : \hat{k}$ **do**
6: $\quad$ $\boldsymbol{X}_{\bullet,p} \leftarrow \boldsymbol{x}_p \in \mathbb{R}^n$, where $\boldsymbol{x}_p$ is the eigenvector with eigenvalue $s_p^{\uparrow}(H_{\zeta_p}) = 0$.
7: Normalize the rows of $\boldsymbol{X}_{i,\bullet} \leftarrow \boldsymbol{X}_{i,\bullet}/\|\boldsymbol{X}_{i,\bullet}\|$
8: Estimate community labels $\hat{\boldsymbol{\ell}}$ as output of $\hat{k}$-class *k-means* on the rows of $\boldsymbol{X}$.
9: **return** Estimated number $\hat{k}$ of communities and label vector $\hat{\boldsymbol{\ell}}$.

As such, while our proposed algorithm naturally discards degree dependence in the entries of $\boldsymbol{X}$ under the DC-SBM setting, the reality of practical networks may disrupt this expected behavior and the projection of the vectors $\boldsymbol{X}_{i,\bullet}$ on the unit hyper-sphere both alleviates this deleterious effect and further improves the convergence of *k-means*. We thus adopt this normalization step in our final Algorithm 2 and will confirm its practical gains when clustering real graphs.

The total theoretical complexity of Alg. 2 is dominated by subroutines 1 and 2, that both run in $O(n\hat{k}^3)$, as the *k-means* step costs only $O(n\hat{k}^2)$. This is to compare with the usual complexity of spectral clustering algorithms on sparse graphs that are in $O(n\hat{k}^2)$ (see for instance Tremblay and Loukas, 2020). This additional cost comes with a better classification performance in many real-world graphs, as discussed in the next section. In practice, to give an order of magnitude of computation times[7], running Algorithm 2 on a SBM[8] with $k = 4$ classes of equal size, $n = 10^5$ (resp. 10^6), $c_{\text{in}} = 26.20$ on the diagonal elements of C and $c_{\text{out}} = 4.60$ for the off diagonal elements, takes approximately 85 (resp., 2200) seconds on a laptop. If k is known in advance, this times drop to approximately 45 (resp., 700) seconds.

4.5 Algorithm and Performance on Real Networks

In this section we compare the performance of Algorithm 2 versus competing spectral methods on real-world networks that do not have ground-truth label assignment as well as LFR synthetic benchmark networks.

Measuring the quality of an inferred partition $\hat{\boldsymbol{\ell}}$ is in general not straightforward since communities are not uniquely defined. Two different scores are then adopted adopted. One is the modularity[9] (Newman and Girvan, 2004; Newman, 2006), $\mathcal{M}$:

$$\mathcal{M} = \frac{1}{2|\mathcal{E}|} \sum_{i,j=1}^{n} \left(A_{ij} - \frac{d_i d_j}{2|\mathcal{E}|} \right) \delta(\hat{\ell}_i, \hat{\ell}_j). \tag{29}$$

---

7. The laptop's RAM is 7.7Gb with Intel Core i7-6600U CPU @ 2.6GHz x 4
8. These times refer to our Python implementation, while the CoDeBetHe.jl implementation runs for $n = 10^5$ (resp., 10^6) in approximately 15 (resp., 450) seconds if k is not known and in approximately 6 (resp., 100) seconds if k is known a priori.
9. Note that the measure of the modularity is meaningful on *assortative* or *disassortative* networks but not on "hybrid" networks for which a more involved description would be needed.
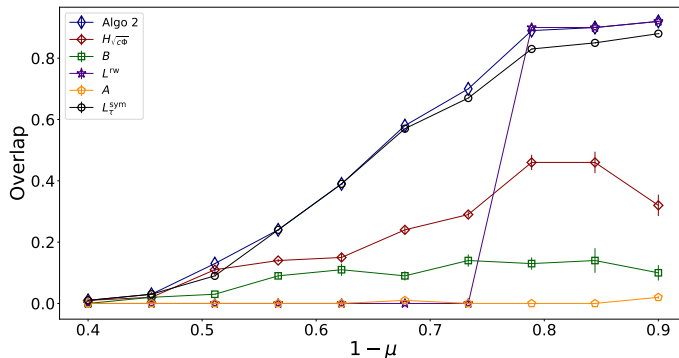
Figure 8: Overlap comparison on different LFR (Lancichinetti et al., 2008) benchmark graphs as a function of the mixing parameter μ. Blue sharp diamonds are the result given by Algorithm 2, red diamonds the algorithm using the Bethe-Hessian of (Saade et al., 2014), green squares are for the non-backtracking of (Krzakala et al., 2013), yellow pentagons the adjacency matrix, purple stars the random walk Laplacian, black dots are the algorithm of (Qin and Rohe, 2013). The parameters of the LFR are $n = 50.000$, $d_{\max} = 15\log(n)$, $\hat{d} = 5$, $\tau_1 = -3$, $\tau_2 = 1$, `min_class_size` $= n/8$, `max_class_size` $= n/8$. Variances are taken over 8 runs of *k-means*.

High values of $\mathcal{M}$ correspond to good quality partitions. Alternatively, the partition quality is evaluated in terms of (normalized) posterior negative log-likelihood of the DC-SBM, $\mathcal{L}$:

$$\mathcal{L} = -\frac{1}{2|\mathcal{E}|}\left[\sum_{(ij)\in\mathcal{E}} \log\left(\hat{\theta}_i\hat{\theta}_j \frac{\hat{C}_{\hat{\ell}_i,\hat{\ell}_j}}{n} \right) + \sum_{(ij)\notin\mathcal{E}} \log\left(1 - \hat{\theta}_i\hat{\theta}_j \frac{\hat{C}_{\hat{\ell}_i,\hat{\ell}_j}}{n} \right) \right], \tag{30}$$

where $\hat{\theta}_i = d_i/\bar{d}$ and $\hat{C}_{ab} = \left(\sum_{i\,:\,\hat{\ell}_i=a}\sum_{j\,:\,\hat{\ell}_j=b} A_{ij} \right) / \left(\sum_{i\,:\,\hat{\ell}_i=a}\sum_{j\,:\,\hat{\ell}_j=b} \theta_i\theta_j \right)$. Good quality clustering correspond, in this case, to low values $\mathcal{L}$.

Table 1 compares the results of different clustering algorithms[10] of 15 real-world networks of increasing size. For all networks, the number of communities, when not available, is estimated through $\hat{k}_{\mathrm{L}}$ (see Subroutine 1) and then the same value is used for all competing techniques (which in general do not provide their own dedicated estimator of k). The underlined numbers in the k column indicate instead that k is known. Furthermore, for all networks, community detection is performed only on the largest connected component of the graph and n, c, Φ, k refer to the characteristics of this dominant connected component.

Given the embedding $\boldsymbol{X}$, 5 iterations of *k-means* are run and the best partition $\hat{\boldsymbol{\ell}}$ (in terms of *k-means* is kept) and the scores $\mathcal{M}, \mathcal{L}$ are computed. To keep stochastic of *k-means* into account, this step is repeated for 8 times and Tables 1 reports the results of $\mathcal{M}, \mathcal{L}$ in terms of `mean ± standard deviation`.

---

10. The choice of the spectral algorithm considered for comparison is based on two criterion: i) $H_{\sqrt{c\Phi}}$, B and L_τ^{sym} are the state-of-the-art spectral methods for community detection in *sparse* graphs, while A and L^{rw} are algorithms of great relevance in the literature; ii) these are the methods that are well explained by our unified framework.

Data set	n	c	Φ	k	Alg 1_B	Alg 2_{wp}	Alg 2	A	$H_{\sqrt{c\Phi}}$	B	L^{rw}	L_r^{sym}
Karate	34	4.6	1.7	2	**0.86**	**0.86**	**0.86**	**0.86**	**0.86**	**0.86**	0.98	**0.86**
					0.37	**0.37**	**0.37**	**0.37**	**0.37**	**0.37**	0.36	**0.37**
Dolphins	62	5	1.3	2	**1.25**	**1.25**	**1.25**	1.38 ±0.01	1.38	1.38	**1.25**	**1.25**
					0.38	**0.38**	**0.38**	0.22 ±0.01	0.22	0.22	**0.38**	**0.38**
Polbooks	105	8.4	1.4	3	**1.17**	1.16	**1.17**	1.26 ±0.01	1.21	1.28 ±0.01	1.16	1.22
					0.51	0.50	**0.51**	0.46	0.50	0.47 ±0.01	0.50	**0.51**
Football	115	10.7	1	12	**0.83**	**0.83**	**0.83**	**0.83**	**0.83**	**0.83**	**0.83**	**0.83**
					0.60	**0.60**	**0.60**	**0.60**	**0.60**	**0.60**	**0.60**	**0.60**
Mail	1133	9.6	1.9	21	1.95 ±0.01	**1.89 ±0.02**	**1.88 ±0.01**	2.13 ±0.01	2.03 ±0.01	2.05 ±0.01	1.99 ±0.03	**1.87 ±0.01**
					0.45 ±0.01	0.49 ±0.01	**0.52 ±0.01**	0.31	0.40	0.36 ±0.01	0.50 ±0.02	0.51
Polblogs	1222	27.4	3	2	**1.50**	**1.50**	**1.50**	1.62	1.63	1.65	1.73	**1.50**
					0.43	**0.43**	**0.43**	0.25	0.27	0.24	0.00	**0.43**
Tv	3892	8.9	3	41	2.01 ±0.03	1.89 ±0.09	**1.50 ±0.01**	1.99 ±0.09	2.02 ±0.05	2.05 ±0.05	2.06 ±0.15	**1.47 ±0.01**
					0.55 ±0.02	0.70 ±0.06	**0.84**	0.53 ±0.04	0.53 ±0.02	0.53 ±0.02	0.59 ±0.12	0.79 ±0.01
Facebook	4039	43.7	2.4	55	1.17 ±0.04	1.05 ±0.06	**0.88 ±0.01**	1.13 ±0.03	1.21 ±0.03	1.21 ±0.03	1.07 ±0.14	**0.87 ±0.01**
					0.52 ±0.01	**0.76 ±0.02**	**0.76**	0.44 ±0.03	0.47 ±0.01	0.47 ±0.02	**0.75 ±0.05**	0.59 ±0.01
GrQc	4158	6.5	2.8	29	2.40 ±0.04	2.22 ±0.04	**1.89 ±0.01**	2.39 ±0.06	2.40 ±0.04	2.39 ±0.01	2.44 ±0.09	**1.90 ±0.02**
					0.50 ±0.02	0.68 ±0.09	**0.80**	0.49 ±0.04	0.50 ±0.01	0.50 ±0.01	0.58 ±0.10	0.78
Power grid	4941	2.7	1.5	25	3.63 ±0.02	2.61 ±0.01	**2.57 ±0.01**	3.84 ±0.01	3.63 ±0.03	3.71 ±0.01	2.61 ±0.01	2.80 ±0.02
					0.32 ±0.01	0.92	**0.93**	0.17 ±0.01	0.35 ±0.03	0.28 ±0.01	0.92	0.86 ±0.01
Politicians	5908	14.1	3	62	1.78 ±0.04	1.60 ±0.05	1.43 ±0.01	1.94 ±0.06	1.90 ±0.02	1.92 ±0.02	1.77 ±0.11	**1.40 ±0.01**
					0.52 ±0.02	**0.83 ±0.01**	**0.84 ±0.01**	0.54 ±0.02	0.50 ±0.02	0.47 ±0.01	0.78 ±0.07	0.73 ±0.01
GNutella P2P	6299	6.6	2.7	4	3.32 ±0.01	3.28	**3.21**	3.33	3.37	3.38 ±0.01	3.41	**1.88**
					0.20 ±0.01	0.26	**0.40**	0.19 ±0.01	0.14	0.13 ±0.01	0.00	0.35
Wikipedia	7066	28.3	5.1	22	1.93	**1.88 ±0.01**	**1.88**	1.99 ±0.01	1.97	1.96	2.12 ±0.07	**1.44 ±0.01**
					0.20	0.23 ±0.01	0.26	0.15	0.17	0.14 ±0.01	**0.38 ±0.02**	0.27
HepPh	11204	21.0	6.2	60	1.75 ±0.06	1.98 ±0.05	**1.52 ±0.01**	1.74 ±0.08	1.79 ±0.04	1.68 ±0.05	2.12 ±0.07	**1.44 ±0.01**
					0.39 ±0.02	0.37 ±0.06	**0.57**	0.42 ±0.04	0.39 ±0.02	0.44 ±0.03	0.25 ±0.09	0.51 ±0.01
Vip	11565	11.6	4.4	53	2.60 ±0.03	2.48 ±0.13	**2.22 ±0.01**	2.72 ±0.02	2.64 ±0.03	2.67 ±0.03	2.65 ±0.01	**2.18 ±0.01**
					0.34 ±0.01	**0.58 ±0.15**	**0.60 ±0.01**	0.27 ±0.01	0.32 ±0.01	0.30 ±0.01	0.50 ±0.12	0.54 ±0.01

Table 1: Modularity and log-likelihood comparison on real networks (Leskovec and Krevl, 2014; Zachary, 1977; Lusseau et al., 2003; Girvan and Newman, 2002; Adamic and Glance, 2005). The value of k is estimated according to Algorithm 2 only if it is not known. When k is known in advance, it appears in the table as an underlined number. All algorithms are performed on the largest connected component of the network. For each data set, the first row in the score $\mathcal{L}$ Equation (30), while the second is $\mathcal{M}$, Equation (29). When the standard deviation is not indicated it means that it is below 0.01.
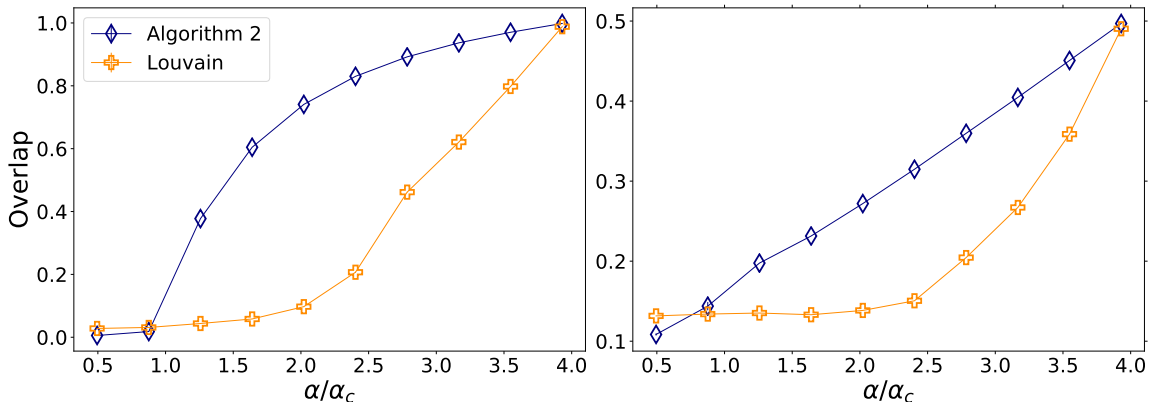
34

Figure 9: Comparison of the overlap (left) and modularity (right) resulting from the label estimation of Algorithm 2 (blue diamonds) and Louvain algorithm (orange crosses) applied on the giant component of $\mathcal{G}$, as function of the detection hardness. For this simulation, $n = 50.000$, $k = 2$, $\pi \propto \mathbf{1}_2$, $c = 8$, $c_{\text{out}} = 0.05 \to 7$. Averages are taken over 10 samples.

The first score column of Table 1 indicates the output of the meta-algorithm Algorithm 1 for which ζ_p is estimated from $\rho(B)/s^{\downarrow}_{R,p}(B)$. The second score column provides the output of Algorithm 2 in which the last step of projection on the hypersphere (Step 4 of Algorithm 0) is not performed. Comparing these two columns, it is clear that Algorithm 2 generally provides much better partitions both in terms of $\mathcal{M}$ and $\mathcal{L}$, as a consequence of $\rho(B)/s^{\downarrow}_{R,p}(B)$ being an inappropriate estimator for ζ_p in general. The following columns display the results scores obtained by Algorithm 2 with the projection step (highlighted in cyan), by clustering based on the leading eigenvectors of A, by the Bethe-Hessian as per (Saade et al., 2014), by the non-backtracking matrix as per (Krzakala et al., 2013), by the random walk Laplacian as per (Shi and Malik, 2000) and by the symmetric normalized Laplacian of (Qin and Rohe, 2013).

The algorithm of (Shi and Malik, 2000) based on L^{rw} provides in certain cases very competitive partitions (*e.g.*, in the Wikipedia data set) but is quite unreliable as it may dramatically fail in others (GNutella P2P and Polblogs). Algorithm 2 without the normalization step provides systematically good partitions, all comparable to those of L^{sym}_{τ}. This is an evidence that Algorithm 2 effectively produces a node embedding which is significantly resilient to degree heterogeneity. Finally, the improved version of Algorithm 2 including the projection step further improves the quality of the partition on most data sets, providing on all data sets but Wikipedia the highest reported modularity and lowest, or second lowest measured value of $\mathcal{L}$.

To test Algorithm 2 we computed the vector $\boldsymbol{\zeta}$ up to machine error precision, that is, the convergence stopping criterion in Subroutine 2 is met when r's update is below machine precision. Notably, due to the non-linearity of the *k-means* step, larger errors in the estimate of the values of $\boldsymbol{\zeta}$ lead, within a certain range, to the same partition. The same classification precision can therefore be reached in fewer iterations, thus faster, if needed.

To further test the efficiency of Algorithm 2, we compared the clustering performances on some large size Lancichinetti-Fortunato-Radicchi networks (Lancichinetti et al., 2008)

that notoriously constitute a benchmark to evaluate clustering performance on synthetic graphs with ground truth labels that resemble real-world networks. In this case we have a clear definition of communities (provided by the generative model), hence the performance is evaluated in terms of overlap. Figure 8 shows the performance of the different spectral algorithms considered earlier as a function of the mixing parameters $1 - \mu$, the fraction of links connecting nodes in the same community. The result is in complete agreement with the observations in Table 1 and Figure 5 which all evidence: i) the importance of choosing a proper parametrization for H_r; ii) that Algorithm 2 is robust on real-world networks.

Let us mention that, comparing the columns "Alg 2_{wp}" and "$H_{\sqrt{c\Phi}}$", it appears evident that Algorithm 2 indeed achieves good quality embeddings on real-world graphs. In addition, it appeared from our simulations that the projection step on the unitary hypersphere of the rows of X significantly helps (both in terms of quality of the partitioning and of the variance of the score) the convergence of *k-means* at a low computational cost.

As a side comment, although formally not strictly comparable on even grounds, we evaluated the performances reached by Algorithm 2 against the popular greedy Louvain method (Blondel et al., 2008) using its *scikit-network* implementation (Pedregosa et al., 2011). The Louvain method comes with an estimate of the number of communities and relies on a different notion of communities than the one we used. Specifically, it is a hierarchical algorithm which looks for a partition maximizing the modularity, while Algorithm 2 relies on the DC-SBM assumption. Figure 9 shows that Algorithm 2 largely outperforms the Louvain method on a DC-SBM graphs, both in terms of overlap (left) and modularity (right), except below α_c. There, the Louvain method reaches higher modularity, but this is likely an incidental artifact of the modularity optimization constraint (Guimera et al., 2004).

The situation of real networks is less straightforward. We indeed observed that when the number of communities estimated by both algorithms is similar, they both produce node partitions with a similar modularity. This is however no longer the case when the estimated k are more distinct (*e.g.*, on GNutella P2P). While we recommend our estimation method of k due to its interpretability (that we discussed all along the article), we acknowledge that the Louvain algorithm often provides very competitive outcomes, at a smaller computational cost but at the expense of any theoretical guarantee. Similar considerations can be made in relation to the more recently proposed Leiden method (Traag et al., 2019).

5. Comments and Conclusions

This article provides a critical analysis of the current state-of-the-art spectral algorithms for community detection in sparse graphs. While these methods were mostly developed in parallel and from differing lines of arguments, we showed that they all relate and can be jointly optimized to tackle community detection in heterogeneous and sparse graphs. This in passing bridges ideas from statistics and physics. Notably, we discovered that the hyperparameters involved in these algorithms, so far chosen statistically, must instead be smartly selected, this selection being a function of the difficulty of the clustering problem. Besides this fundamental observation, we also showed that there is an efficient (with computational complexity linear in n) routine capable of estimating these optimal parameters, which comes from a deep relation between the non-backtracking matrix and the Bayes optimal solution.

The necessity in all these methods to adapt the choice of the graph representation matrix to the hardness of the problem at hand is a profound observation, likely not restricted to community detection on sparse graphs. Indeed, the classical spectral algorithms based on Laplacian matrices may all be interpreted as a continuous relaxation of an NP-hard discrete optimization problem, providing good (and sometimes provably optimal) solutions on dense but generally not on sparse networks or data. In light of this observation, it may appear that some common optimization problems, beyond clustering, have similarly been devised in the past to perform well in easy scenarios, but would require a task-hardness related update to perform better in harder settings.

Practically, examples of how the results of this article can be generalized to more involved scenarios include the dynamical community detection studied in (Dall'Amico et al., 2020) on time-evolving graphs and the i.i.d. sparsification (improving cost efficiency) of kernel matrices for spectral clustering (Dall'Amico et al., 2021). The latter work shows that the common kernel spectral clustering algorithms can be interpreted as a dense limit of a weighted Bethe-Hessian, that can be conveniently adopted in both the sparse and dense regimes. Beyond spectral clustering, we envision that a carefully parametrized Bethe-Hessian matrix can be exploited as a powerful alternative to the combinatorial graph Laplacian.

Acknowledgments

Appendix A. Existence of a Giant Component in the Sparse DC-SBM

In this appendix we state and prove Theorem 1 which provides a necessary and sufficient condition for the existence of a giant component in a multi-class (with unequal sizes) sparse DC-SBM. This theorem is an application of (Bollobás et al., 2007, Theorem 3.1) to the present DC-SBM setting. Note in particular that the number of nodes in each class p is deterministic and equal to $n\pi_p$. In Bollobas' setting, however, the class of each node i is drawn independently and is equal to p with probability π_p: $n\pi_p$ is thus only the expected number of nodes in class p. Asymptotically, we will see that this of course does not affect the result. To be precise, the generative procedure under analysis in the theorem is the following:

- letting k be the number of classes, and n the number of nodes, we consider a symmetric matrix $C \in \mathbb{R}^{k \times k}$ where $C_{\ell\ell'} > 0$ encodes the affinity between classes ℓ and ℓ'; a diagonal matrix $\Pi = \text{diag}(\pi_1, \ldots, \pi_k)$, where $0 < \pi_\ell < 1$ is the expected proportion of nodes in class ℓ and $\text{Tr}(\Pi) = 1$.

- we generate n nodes and, for each node, draw independently the class p to which it belongs with probability π_p. Also, for each node i, we draw independently its intrinsic connectivity variable $\theta_i \in [\theta_{\min}, \theta_{\max}]$, according to an underlying distribution $\nu(\theta)$ verifying $\mathbb{E}(\theta) = 1$ and $\mathbb{E}(\theta^2) = \Phi$.

- For each pair of nodes (i, j), we generate an edge with probability $\min\left(\theta_i \theta_j \frac{C_{\ell_i, \ell_j}}{n}, 1\right)$ where ℓ_i is the class of node i.

Theorem 1 (Percolation threshold in the DC-SBM) *Consider a graph $\mathcal{G}$ generated according to the above DC-SBM procedure. Suppose that the constant vector is the Perron-Frobenius eigenvector of $C\Pi$ associated to eigenvalue $c > 0$. Then, for all large n with high probability, the graph $\mathcal{G}$ has a giant component if and only if $c\Phi > 1$.*

Proof In our proof we will use (Bollobás et al., 2007, Theorem 3.1) which requires to write the graph $\mathcal{G}$ as a graphon. Let us define the "node" variable $\boldsymbol{x} = \theta \boldsymbol{v} \in \mathcal{S}$, with probability distribution $\mu(\boldsymbol{x})$, where

- $\theta \in \Theta = [\theta_{\min}, \theta_{\max}]$ is a scalar that encodes the degree heterogeneity, distributed according to $\nu(\cdot)$ and satisfying $\int_\Theta d\nu(\theta) = 1$ (normalization), $\mathbb{E}[\theta] = \int_\Theta \theta \, d\nu(\theta) = 1$ and $\mathbb{E}[\theta^2] = \int_\Theta \theta^2 \, d\nu(\theta) = \Phi$.

- $\boldsymbol{v} \in \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k\}$ with $\boldsymbol{e}_p \in \mathbb{R}^k$ satisfying $e_{p,\ell} = \delta_{\ell p}$; $\boldsymbol{v}$ is a k-dimensional random vector with law $\mathbb{P}(\boldsymbol{v} = \boldsymbol{e}_p) = \pi_p$; that is, $\boldsymbol{v}$ encodes the class to which node $\boldsymbol{x}$ belongs.

- The probability density of $\boldsymbol{x}$ is denoted by $\mu(\boldsymbol{x})$ and is defined on $\mathcal{S} = \cup_{\ell=1}^k \mathcal{S}_\ell$ where $\mathcal{S}_\ell = [\theta_{\min} \boldsymbol{e}_\ell, \theta_{\max} \boldsymbol{e}_\ell]$. The measure $\mu(\boldsymbol{x})$ is equal to $\pi_\ell \nu(\theta)$ if $\boldsymbol{x} \in \mathcal{S}_\ell$ and it is indeed normalized:

$$\int_\mathcal{S} d\mu(\boldsymbol{x}) = \sum_{\ell=1}^k \pi_\ell \int_\Theta d\nu(\theta) = \sum_{\ell=1}^k \pi_\ell = 1.$$

Given two such nodes, we next define the kernel $\kappa(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^T C \boldsymbol{y}$. The kernel κ satisfies the three conditions of (Bollobás et al., 2007, Definition 2.7) and is a so-called graphical kernel.

With these notations at hand, the generative procedure of the DC-SBM under consideration is equivalent to: drawing n independent values $\{\boldsymbol{x}_i\}_{i=1,\ldots,n}$ from $\mu(\cdot)$ and then generating the edges independently according to $p_{ij} = \min\left(\frac{\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)}{n}, 1\right)$; we thus fall precisely under Bollobas' framework.

In order to use the core argument of (Bollobás et al., 2007, Theorem 3.1), we still need to define the linear operator T_κ, an operator on $f : \mathbb{R}^k \to \mathbb{R}$, endowed with the norm $\|f\|_2^2 = \int_\mathcal{S} f^2(\boldsymbol{x}) \, d\mu(\boldsymbol{x})$, as:

$$\forall \, \boldsymbol{x} \in \mathcal{S}, \quad (T_k f)(\boldsymbol{x}) = \int_\mathcal{S} \kappa(\boldsymbol{x}, \boldsymbol{y}) f(\boldsymbol{y}) \, d\mu(\boldsymbol{y})$$

The square 2-norm of this operator, $\|T_\kappa\|_2$, reads

$$\|T_\kappa\|_2^2 = \sup_{\|f\|_2 \leq 1} \int_{\mathcal{S}} (T_\kappa f)^2(\boldsymbol{x}) \, d\mu(\boldsymbol{x}). \tag{31}$$

According to (Bollobás et al., 2007, Theorem 3.1), with high probability, a giant component appears in $\mathcal{G}$ if and only if $\|T_\kappa\| > 1$. We thus are left to evaluating $\|T_\kappa\|$ for the kernel $\boldsymbol{x}^T C \boldsymbol{y}$: we will show that it equals $c\Phi$, first by showing that $\|T_\kappa\| \leq c\Phi$ and then finding a particular function f for which the bound is attained.

Let f be a function from $\mathbb{R}^k$ to $\mathbb{R}$. We introduce its associated vector $\boldsymbol{\omega}_f \in \mathbb{R}^k$ as follows.

$$(T_\kappa f)(\boldsymbol{x}) = \int_{\mathcal{S}} \kappa(\boldsymbol{x}, \boldsymbol{y}) f(\boldsymbol{y}) \, d\mu(\boldsymbol{y}) = \int_{\mathcal{S}} \boldsymbol{x}^T C \boldsymbol{y} f(\boldsymbol{y}) \, d\mu(\boldsymbol{y})$$

$$= \boldsymbol{x}^T C \sum_{\ell=1}^{k} \pi_\ell \boldsymbol{e}_\ell \int_{\Theta} \theta \, f(\theta \boldsymbol{e}_\ell) \, d\nu(\theta) \equiv \boldsymbol{x}^T C \Pi \boldsymbol{\omega}_f, \tag{32}$$

where $\forall \ell$, $\omega_{f,\ell} = \int_{\Theta} \theta \, f(\theta \boldsymbol{e}_\ell) \, d\nu(\theta)$. From Equation (32), one thus has:

$$\int_{\mathcal{S}} (T_\kappa f)^2(\boldsymbol{x}) d\mu(\boldsymbol{x}) = \int_{\mathcal{S}} \boldsymbol{\omega}_f^T \Pi C \boldsymbol{x} \boldsymbol{x}^T C \Pi \boldsymbol{\omega}_f \, d\mu(\boldsymbol{x})$$

$$= \Phi \, \boldsymbol{\omega}_f^T \Pi C \left(\sum_{\ell=1}^{k} \pi_\ell \boldsymbol{e}_\ell \boldsymbol{e}_\ell^T \right) C \Pi \boldsymbol{\omega}_f$$

$$= \Phi \, \boldsymbol{\omega}_f^T \Pi C \Pi C \Pi \boldsymbol{\omega}_f. \tag{33}$$

Injecting the result of Equation (33) into Equation (31) we obtain

$$\|T_\kappa\|_2^2 = \Phi \sup_{\|f\|_2 \leq 1} \boldsymbol{\omega}_f^T \Pi C \Pi C \Pi \boldsymbol{\omega}_f = \Phi \sup_{\|f\|_2 \leq 1} \frac{\boldsymbol{\omega}_f^T \Pi C \Pi C \Pi \boldsymbol{\omega}_f}{\boldsymbol{\omega}_f^T \Pi \boldsymbol{\omega}_f} \cdot \boldsymbol{\omega}_f^T \Pi \boldsymbol{\omega}_f$$

$$\leq \Phi \sup_{\|f\|_2 \leq 1} \frac{\boldsymbol{\omega}_f^T \Pi C \Pi C \Pi \boldsymbol{\omega}_f}{\boldsymbol{\omega}_f^T \Pi \boldsymbol{\omega}_f} \cdot \sup_{\|f\|_2 \leq 1} \boldsymbol{\omega}_f^T \Pi \boldsymbol{\omega}_f \tag{34}$$

Analyzing the first element, we can write for $\boldsymbol{v}_f = \Pi^{1/2} \boldsymbol{\omega}_f$:

$$\sup_{\|f\|_2 \leq 1} \frac{\boldsymbol{\omega}_f^T \Pi C \Pi C \Pi \boldsymbol{\omega}_f}{\boldsymbol{\omega}_f^T \Pi \boldsymbol{\omega}_f} = \sup_{\|f\|_2 \leq 1} \frac{\boldsymbol{v}_f^T \Pi^{1/2} C \Pi C \Pi^{1/2} \boldsymbol{v}_f}{\boldsymbol{v}_f^T \boldsymbol{v}_f} \leq \sup_{\boldsymbol{v} \in \mathbb{R}^k} \frac{\boldsymbol{v}^T \Pi^{1/2} C \Pi C \Pi^{1/2} \boldsymbol{v}}{\boldsymbol{v}^T \boldsymbol{v}}$$

i.e.:

$$\sup_{\|f\|_2 \leq 1} \frac{\boldsymbol{\omega}_f^T \Pi C \Pi C \Pi \boldsymbol{\omega}_f}{\boldsymbol{\omega}_f^T \Pi \boldsymbol{\omega}_f} \leq s_1^{\downarrow}(\Pi^{1/2} C \Pi C \Pi^{1/2}) = s_1^{\downarrow}((C\Pi)^2) = c^2,$$

as, by hypothesis, c is the Perron eigenvalue of $C\Pi$, and as such is larger than the modulus of any other eigenvalue: $s_1^{\downarrow}((C\Pi)^2)$ is indeed c^2. We can therefore rewrite Equation (34) as

$$\|T_\kappa\|_2^2 \leq c^2 \Phi \sup_{\|f\|_2 \leq 1} \boldsymbol{\omega}_f^T \Pi \boldsymbol{\omega}_f.$$

Analyzing the right-hand side term, we have

$$\boldsymbol{\omega}_f^T \Pi \boldsymbol{\omega}_f = \sum_{\ell=1}^{k} \pi_\ell \omega_{f,\ell}^2 = \sum_{\ell=1}^{k} \pi_\ell \left(\int_\Theta \theta f(\theta \boldsymbol{e}_\ell) \, d\nu(\theta) \right)^2$$

which, by Cauchy-Schwartz's inequality is bounded as

$$\left(\int_\Theta \theta f(\theta \boldsymbol{e}_\ell) \, d\nu(\theta) \right)^2 \leq \left(\int_\Theta \theta^2 d\nu(\theta) \right) \left(\int_\Theta f^2(\theta \boldsymbol{e}_\ell) \, d\nu(\theta) \right) \leq \Phi,$$

where in the last step we used the fact that the norm of f is less than or equal to one. We thus obtain that $\sup_{\|f\|_2 \leq 1} \boldsymbol{\omega}_f^T \Pi \boldsymbol{\omega}_f \leq \Phi$ and conclude that

$$\|T_\kappa\|_2^2 \leq (c\Phi)^2. \tag{35}$$

We are thus left to showing that there exists a function f for which the bound is reached. Let us consider $\bar{f}(\boldsymbol{x}) = \|\boldsymbol{x}\|/\sqrt{\Phi}$. It is easy to check that this function has unit norm

$$\|\bar{f}\|_2^2 = \int_\mathcal{S} \left(\frac{\|\boldsymbol{x}\|}{\sqrt{\Phi}} \right)^2 \, d\mu(\boldsymbol{x}) = \frac{1}{\Phi} \int_\mathcal{S} \theta^2 \|\boldsymbol{e}_\ell\|^2 \, d\mu(\boldsymbol{x}) = \frac{1}{\Phi} \int_\mathcal{S} \theta^2 \, d\mu(\boldsymbol{x}) = 1. \tag{36}$$

Furthermore, observe that $\boldsymbol{\omega}_{\bar{f}} = \sqrt{\Phi} \mathbf{1}_k$. Then, we have

$$\|T_\kappa\|_2^2 = \Phi \sup_{\|f\|_2 \leq 1} \boldsymbol{\omega}_f^T \Pi C \Pi C \Pi \boldsymbol{\omega}_f \geq \Phi \boldsymbol{\omega}_{\bar{f}}^T \Pi C \Pi C \Pi \boldsymbol{\omega}_{\bar{f}} = \Phi^2 \mathbf{1}_k^T \Pi (C\Pi)^2 \mathbf{1}_k = (c\Phi)^2 \tag{37}$$

as, by hypothesis, we suppose that $C\Pi \mathbf{1}_k = c \mathbf{1}_k$.

Combining Equations (35) and (37), we find that $\|T_\kappa\| = c\Phi$ so that, from (Bollobás et al., 2007, Theorem 3.1), we conclude that the percolation transition is at $c\Phi = 1$ and a giant component exists if and only if $c\Phi > 1$.

∎

Appendix B. Definition of ζ on the Sparse DC-SBM

In this Appendix we show that for a graph generated from the sparse DC-SBM model, defined in Equation (1), the two following definitions of $\zeta_p^{(j)}$ (where j is the giant component) are indeed equivalent with high probability for $p \geq 2$ if $\zeta_p^{(j)}$ exists:

$$\zeta_p^{(j)} = \min_{r \geq 1} \{r : s_p^\uparrow(H_r^{(j)}) = 0\}, \tag{38}$$

$$\zeta_p^{(j)} = \min_{r > 1} \{r : s_p^\uparrow(H_r) = 0\}. \tag{39}$$

We will further discuss that, with high probability, $\zeta_p^{(j)} > 1$ exists only for the giant component.

40

This appendix is structured as follows. First we enunciate two lemmas and one corollary to state that when an arbitrary connected graph has at most one loop, the spectrum of its associated non-backtracking matrix B does not have real eigenvalues that are larger than one in modulus.

We then proceed by arguing that with high probability, for large DC-SBM graphs, the small connected components do not have more than one loop, implying that all the real eigenvalues of B that are larger than one in modulus come from the giant component. Finally we show that this last statement, implies that the two definitions of $\zeta_p^{(j)}$ given in Equations (38,39) are indeed equivalent with high probability.

Lemma 1 (Spectrum of B on a tree) *Let $\mathcal{T}(\mathcal{V}, \mathcal{E})$ be a tree. Then, all the eigenvalues of its associated non-backtracking matrix $B(\mathcal{T})$, as defined in Equation (11), are equal to zero.*

Lemma 2 (Spectrum of $\mathcal{G}$ plus a node) *Let $\mathcal{G}'$ be a graph obtained by adding one node and one edge to the graph $\mathcal{G}$, i.e. for $i \notin \mathcal{V}$ and an arbitrary $j \in \mathcal{V}$, $\mathcal{V}' = \mathcal{V} \cup \{i\}$ and $\mathcal{E}' = \mathcal{E} \cup (ij)$. Then, all the non-zero eigenvalues of $B(\mathcal{G}')$ are equal to the non-zero eigenvalues of $B(\mathcal{G})$, where $B(\mathcal{G})$ is the non-backtracking matrix defined in Equation (11) of the graph $\mathcal{G}$.*

Corollary 1 (Spectral radius of a connected graph with $|\mathcal{E}| = |\mathcal{V}|$) *Let $\mathcal{G}(\mathcal{V}, \mathcal{E})$ be a connected graph with $|\mathcal{E}| = |\mathcal{V}|$. Then $\rho(B) = 1$, where B is defined in Equation (11) and $\rho(\cdot)$ is the spectral radius.*

Proof [Lemma 1] Consider an arbitrary vector $\boldsymbol{g} \in \mathbb{R}^{2|\mathcal{E}|}$ and define $\boldsymbol{g}^{(m)} \in \mathbb{R}^{2|\mathcal{E}|}$ as

$$g_{ij}^{(m)} = \sum_{(k\ell) \ : \ d(ij, k\ell) = m} g_{k\ell}. \tag{40}$$

The notation $d(ij, k\ell) = m$ indicates that there exists a non-backtracking path from the edge (ij) to the edge $(k\ell)$ of length m. From a straightforward calculation one obtains

$$(B\boldsymbol{g}^{(m)})_{ij} = \sum_{\ell \in \partial j \backslash i} \sum_{(kq):d(j\ell,kq)=m} g_{kq} = g_{ij}^{(m+1)}. \tag{41}$$

For all trees, for any two edges we have that $d(ij, k\ell) \le n-1$, so there is a value of $m_\mathrm{c} \le n-1$, which represents the maximal distance between any two directed edges, such that, for all vectors $\boldsymbol{g}$,

$$B\boldsymbol{g}^{(m_\mathrm{c})} = 0. \tag{42}$$

This relation comes from the fact that, by definition of m_c, no two edges are at a distance equal to $m_\mathrm{c} + 1$, so $\{(k\ell : d(ij, k\ell) = m_\mathrm{c} + 1\} = \emptyset$ for any edge (ij). Now, let us consider $\boldsymbol{g}$ to be an eigenvector of B, such that $B\boldsymbol{g} = \gamma\boldsymbol{g}$. Then we can write

$$0 = B\boldsymbol{g}^{(m_\mathrm{c})} = B^2\boldsymbol{g}^{(m_\mathrm{c}-1)} = \cdots = B^{m_\mathrm{c}-1}\boldsymbol{g} = \gamma^{m_\mathrm{c}-1}\boldsymbol{g}. \tag{43}$$

Thus concluding that any eigenvector $\boldsymbol{g}$ of B is associated to eigenvalue zero. Note that in other words, this means that B is nilpotent. ∎

Proof [Lemma 2] Let i be the newly added node and j the node in $\mathcal{V}$ to which i is attached. The matrix $B(\mathcal{G}')$ can be written by adding to the matrix $B(\mathcal{G})$ two rows and two columns corresponding to the directed edges (ij) and (ji). We introduce the notation $\mathbb{1}^{(\bullet j)} \in \mathbb{R}^{2|\mathcal{E}|}$. Its element-wise definition reads for all $(yx) \in \mathcal{E}_d$ (the set of directed edges of $\mathcal{G}$) as $\mathbb{1}_{yx}^{(\bullet j)} = \delta_{xj}$. Similarly, we define $\mathbb{1}^{(j\bullet)} \in \mathbb{R}^{2|\mathcal{E}|}$ with $\mathbb{1}_{yx}^{(j\bullet)} = \delta_{yj}$. Denote with $M = (\mathbb{1}^{(j\bullet)}, \boldsymbol{0}_{2|\mathcal{E}|}) \in \{0,1\}^{2|\mathcal{E}|\times 2}$ and $M' = (\boldsymbol{0}_{2|\mathcal{E}|}, \mathbb{1}^{(\bullet j)}) \in \{0,1\}^{2|\mathcal{E}|\times 2}$, where $\boldsymbol{0}_{2|\mathcal{E}|}$ is the null vector of size $2|\mathcal{E}|$. The matrix $B(\mathcal{G}')$ can be written as:

$$B(\mathcal{G}') = \begin{pmatrix} B(\mathcal{G}) & M' \\ M^T & 0_{2\times 2} \end{pmatrix}$$

We now look for the non-zero eigenvalues of γ of $B(\mathcal{G}')$. Recall that I_n is the $n \times n$ identity matrix. Consider $\gamma \in \mathbb{C}^*$. Using a block matrix determinantal formula one has:

$$\det\left(B(\mathcal{G}') - \gamma I_{2|\mathcal{E}'|}\right) = \det(-\gamma I_2)\det\left(B(\mathcal{G}) - \gamma I_{2|\mathcal{E}|} + \frac{1}{\gamma}M'M^T\right).$$

It is straightforward to check that $M'M^T = 0_{2|\mathcal{E}|\times 2|\mathcal{E}|}$, thus we have $\det\left(B(\mathcal{G}') - \gamma I_{2|\mathcal{E}'|}\right) = \gamma^2\det\left(B(\mathcal{G}) - \gamma I_{2|\mathcal{E}|}\right)$. A non-zero eigenvalue of $B(\mathcal{G}')$ (cancelling the determinant) is thus necessarily also an eigenvalue of $B(\mathcal{G})$, ending the proof. ∎

Proof [Corollary 1] A connected graph with $|\mathcal{E}| = |\mathcal{V}|$ can be obtained by adding an edge between any two nodes of a particular tree defined on $\mathcal{V}$. The graph $\mathcal{G}$ thus contains only one loop. We now apply Lemma 2 "backwards", i.e., removing leaves from $\mathcal{G}$ without affecting the non-zero eigenvalues of B.

By iteratively removing all leaves, the graph $\mathcal{G}$ reduces to a loop. We are thus left to prove that $\rho(B) = 1$ on a loop. It is straightforward to check that on any d-regular graph (all nodes having d neighbours), the vector $\boldsymbol{1}_{2|\mathcal{E}|}$ is the Perron-Frobenius eigenvector of B with eigenvalue equal to $d-1$. A loop is a d-regular graph with $d = 2$, hence the result. ∎

With the results of the former two Lemmas, we proceed to argue that all the eigenvalues of B that are larger than one in modulus come from the giant component. Since the graph is disconnected, $S(B) = \bigcup_{j=1}^{n_{CC}} S\left(B^{(j)}\right)$, i.e., each connected component contributes independently to the eigenvalues of B. The expected size of the small connected components grows as[11] $\log(n)$. We claim that this implies that the probability that a small connected component contains two or more loops is $o_n(1)$, thus tending to zero as n grows to infinity. Applying Lemmas 1 (for zero loops) and Corollary 1 (for one loop), we conclude that, with high probability all the real eigenvalues of B larger than one in modulus will come from the

---

11. Indeed, the kernel κ defined in the proof of Theorem 1 is irreducible in the sense of (Bollobás et al., 2007, Def. 2.10), as the nodes cannot be split into two parts that have no chance of being connected (all the entries of C and all π_i are supposed strictly positive). One can thus apply (Bollobás et al., 2007, Thm. 3.12) stating that the small components are with high probability of order $\log n$.

giant component. From the Ihara-Bass formula (Terras, 2010), for all the eigenvalues $\gamma \neq \pm 1$, $\gamma \in S(B) \iff \det[H_\gamma] = 0$. From the definition of $\zeta_{p \geq 2}^{(j)}$ we gave in Equation (4), we have $\det[H_{\zeta_{p \geq 2}^{(j)}}] = 0$, thus all the $\zeta_p^{(j)} > 1$ are in the spectrum of B. We may conclude that, if $\exists j$ such that $\zeta_{p \geq 2}^{(j)} > 1$, then with high probability j is the giant component. We now conclude showing that this statement implies the equivalence between Equations (38) and (39).

Thanks to the properties we discussed in Section 2, for $p > 1$, then certainly $\zeta_p^{(j)} > 1$, if it exists. We denote with j the giant component and we want to prove that for $r, p > 1$

$$s_p^\uparrow(H_r^{(j)}) = s_p^\uparrow(H_r) \tag{44}$$

Consider $\zeta_p^{(j)} > 1$ such that $s_p^\uparrow(H_{\zeta_p^{(j)}}^{(j)}) = 0$, then certainly there exists[12] $q \geq p$ such that $s_q^\uparrow(H_{\zeta_p^{(j)}}) = 0$ and so it is enough to show that $p = q$. We proceed with a proof by contradiction. Suppose that $p \neq q$, then there exists $j' \neq j$ such that $s_1^\uparrow\left(H_{\zeta_p^{(j)}}^{(j')}\right) \leq 0$. Applying Gershgorin circle theorem, it is easy to show that for $r > d_{\max} - 1$, the matrix H_r is positive definite. Consequently, there exists $r \geq \zeta_p^{(j)} > 1$ such that $s_1^\uparrow(H_r^{(j')}) = 0$. From the Ihara-Bass formula, r is thus in $S(B)$, so there is a real eigenvalue of B larger than one not coming from the giant component. This is in contradiction with what stated above, thus concluding our argument.

Appendix C. The Eigenvalues of T

We here want to show that $s_p^\downarrow(T) = \frac{\nu_{p+1}}{c}$ for $1 \leq p < k$. Consider the following equivalent identities for the matrix $T = \frac{\Pi C}{c} - \Pi \mathbf{1}_k \mathbf{1}_k^T$:

$$\left(\frac{\Pi C}{c} - \Pi \mathbf{1}_k \mathbf{1}_k^T\right) \boldsymbol{a}_p = s_p^\downarrow(T) \boldsymbol{a}_p$$

$$\left(\frac{C\Pi}{c} - \mathbf{1}_k \mathbf{1}_k^T \Pi\right) \boldsymbol{b}_p = s_p^\downarrow(T) \boldsymbol{b}_p; \qquad \boldsymbol{b}_p = \Pi^{-1} \boldsymbol{a}_p$$

Since $C\Pi \mathbf{1} = c\mathbf{1}_k$ and $\mathbf{1}_k^T \Pi \mathbf{1}_k = 1$, then $\boldsymbol{b}_k = \mathbf{1}_k$. By introducing $\boldsymbol{d}_p = \Pi^{1/2} \boldsymbol{b}_p$, we can write:

$$\left(\frac{\Pi^{1/2} C \Pi^{1/2}}{c} - \Pi^{1/2} \boldsymbol{b}_k \boldsymbol{b}_k^T \Pi^{1/2}\right) \boldsymbol{d}_p = s_p^\downarrow(T) \boldsymbol{d}_p$$

$$\left(\frac{\Pi^{1/2} C \Pi^{1/2}}{c} - \boldsymbol{d}_k \boldsymbol{d}_k^T\right) \boldsymbol{d}_p = s_p^\downarrow(T) \boldsymbol{d}_p$$

---

12. As $S(H_r) = \cup_{j=1}^{n_{\mathrm{CC}}} S\left(H_r^{(j)}\right)$.

Given that the $\boldsymbol{d}_k$ are eigenvectors of a symmetric matrix, then $\boldsymbol{d}_p^T \boldsymbol{d}_q = \delta_{pq}$, so, for $p < k$

$$\frac{\Pi^{1/2} C \Pi^{1/2}}{c} \boldsymbol{d}_p = s_p^{\downarrow}(T) \boldsymbol{d}_p$$

$$\frac{C \Pi}{c} \boldsymbol{b}_p = s_p^{\downarrow}(T) \boldsymbol{b}_p$$

From this last equation the result follows directly.

Appendix D. The Excited States of the Ising Hamiltonian on $\mathcal{G}$

In this section we give an explicit connection between the problem described in Section 2.3.2 and its statistical physics analogue. Consider the graph $\mathcal{G}$ obtained from the k-class DC-SBM model as per Equation (1). To study the stability of the configurations of $\langle \boldsymbol{\sigma} \rangle$ as a function of r, one needs to compute the free energy F, defined in Equation (46) (which we recall is a function of the moments of the Boltzmann distribution $\mu(\boldsymbol{\sigma})$):

$$\mu(\boldsymbol{\sigma}) = \frac{1}{Z} e^{-\mathcal{H}(\boldsymbol{\sigma})} \tag{45}$$

$$F = -\log Z = \sum_{\boldsymbol{\sigma}} \mu(\boldsymbol{\sigma})[\mathcal{H}(\boldsymbol{\sigma}) + \log \mu(\boldsymbol{\sigma})]. \tag{46}$$

To evaluate F, a common way to proceed is to exploit the *Bethe* approximation that is exact on a tree and is thus relevant on sparse graphs. By denoting with $\mu_{ij}(\cdot)$ and $\mu_i(\cdot)$ the edge and the node marginals of $\mu(\cdot)$, the Bethe free energy is defined as:

$$\mu^{\mathrm{B}}(\boldsymbol{\sigma}; \boldsymbol{m}, \boldsymbol{\chi}) = \frac{\prod_{(ij) \in \mathcal{E}} \mu_{ij}^{\mathrm{B}}(\sigma_i, \sigma_j; \boldsymbol{m}, \boldsymbol{\chi})}{\prod_{i \in \mathcal{V}} \left[\mu_i^{\mathrm{B}}(\sigma_i; \boldsymbol{m}, \boldsymbol{\chi}) \right]^{d_i - 1}} \tag{47}$$

$$F_{\mathrm{Bethe}}(\boldsymbol{m}, \boldsymbol{\chi}) = \sum_{\boldsymbol{\sigma}} \mu^{\mathrm{B}}(\boldsymbol{\sigma}; \boldsymbol{m}, \boldsymbol{\chi}) \left[\mathcal{H}(\boldsymbol{\sigma}) + \log \mu^{\mathrm{B}}(\boldsymbol{\sigma}; \boldsymbol{m}, \boldsymbol{\chi}) \right]. \tag{48}$$

By a direct computation, one can verify that $F_{\mathrm{Bethe}}(\boldsymbol{m}, \boldsymbol{\chi}) = F + D_{\mathrm{KL}}\left(\mu^{\mathrm{B}}(\boldsymbol{\sigma}; \boldsymbol{m}, \boldsymbol{\chi}) \| \mu(\boldsymbol{\sigma}) \right)$, where $D_{\mathrm{KL}}(\cdot \| \cdot)$ is the Kullback-Leibler divergence which is always greater or equal to zero. In order to find the best estimate of the exact free energy, one needs to find the minimum of F_{Bethe} with respect to $\boldsymbol{m}, \boldsymbol{\chi}$. At large r one sees that F_{Bethe} has a unique minimum in $\boldsymbol{m} = 0$, $\boldsymbol{\chi} = \frac{1}{r} \mathbf{1}_{2|\mathcal{E}|}$, whereas for r small enough, there appear multiple minima and $\boldsymbol{m} = 0$ is a saddle point. To quantitatively study this behavior, one needs to consider the Hessian matrix of the Bethe free energy at the paramagnetic point, i.e.:

$$\left. \frac{\partial^2 F_{\mathrm{Bethe}}(\boldsymbol{m}, \boldsymbol{\chi})}{\partial m_i \partial m_j} \right|_{\boldsymbol{m}=0; \boldsymbol{\chi}=\frac{1}{r}\mathbf{1}_{2|\mathcal{E}|}} = \frac{(H_r)_{ij}}{r^2 - 1}. \tag{49}$$

The eigenvectors corresponding to the negative eigenvalues of the Bethe-Hessian matrix H_r represent the directions along which the paramagnetic phase becomes unstable and they will have a non-trivial structure.

We now detail an approach inspired from (Suchecki, 2006) but adapted to the Bethe approximation to study the stability of the paramagnetic phase. The following equations

hold:

$$\mu_i^{\mathrm{B}}(\sigma_i) = \frac{1}{z_i} \prod_{k \in \partial i} \xi_{ki}(\sigma_i) \tag{50}$$

$$\xi_{ji}(\sigma_i) \propto \sum_{\boldsymbol{\sigma}_j} e^{\mathrm{ath}\left(\frac{1}{r}\right)\sigma_i \sigma_j} \prod_{k \in \partial j \backslash i} \xi_{kj}(\sigma_j). \tag{51}$$

It is possible to verify that by plugging the ansatz of Equations (50,51), one recovers the factorization of Equation (47). We define the following change of variable $\xi_{ji}(\sigma_i) = e^{h_{ji}\sigma_i}$, resulting in $h_{ji} = \frac{1}{2}\log\frac{\xi_{ji}(+1)}{\xi_{ji}(-1)}$. This leads to the following expressions:

$$h_{ji} = \mathrm{ath}\left[\frac{1}{r}\mathrm{th}\left(B^T\boldsymbol{h}\right)_{ji}\right] \tag{52}$$

$$m_i = \mathrm{th}\left[\sum_{j \in \partial i} h_{ji}\right]. \tag{53}$$

We next linearize Equations (52) and (53) at the paramagnetic point, i.e., for $\boldsymbol{h} \to 0$ in the high temperature regime (large r). For mathematical convenience we here assume the regime where $c \gg 1$ and r of the order of c or larger. We will comment in the end on how the results we obtain still hold in the sparse regime in which $c = O_n(1)$.

$$m_i = \sum_{j \in \partial i} h_{ji} = \frac{1}{r}\sum_{j \in \mathcal{V}}\left(B^T\boldsymbol{h}\right)_{ji} = \frac{1}{r}\sum_{j,k,m \in \mathcal{V}} A_{ij}A_{km}\delta_{jm}(1 - \delta_{ik})h_{km}$$

$$= \frac{1}{r}\left[\sum_{j,k \in \mathcal{V}} A_{ij}A_{jk}h_{kj} - \sum_{j \in \mathcal{V}} A_{ij}h_{ij}\right] = \frac{1}{r}\sum_{j \in \mathcal{V}} A_{ij}(m_j - h_{ij}). \tag{54}$$

Further developing the second term of the summation

$$\sum_{j \in \mathcal{V}} A_{ij}h_{ij} = \sum_{j \in \partial i} h_{ij} = \frac{1}{r}\sum_{j \in \partial i}\left(B^T\boldsymbol{h}\right)_{ij} = \frac{(d_i - 1)m_i}{r}, \tag{55}$$

we obtain

$$m_i = \frac{1}{r}\left[\sum_{j \in \mathcal{V}} A_{ij}m_j - \frac{(d_i - 1)m_i}{r}\right] = \frac{1}{r}\sum_{j \in \mathcal{V}} A_{ij}m_j + O_n\left(\frac{1}{r^2}\right). \tag{56}$$

Equation (56) gives a self consistent relation between the magnetizations of the stable (or metastable) configurations. We study the magnetizations taking an average over the realizations of A as proposed in (Suchecki, 2006)

$$\mathbb{E}[m_i] \approx \frac{1}{r}\sum_{j \in \mathcal{V}} \theta_i \theta_j \frac{C_{\ell_i,\ell_j}}{n} \mathbb{E}[m_j]. \tag{57}$$

Defining $s_a = \frac{1}{n\pi_a} \sum_{i \in \mathcal{V}_a} \theta_i \mathbb{E}[m_i]$, where $\mathcal{V}_a = \{i \in \mathcal{V} : \ell_i = a\}$, gives

$$
\begin{aligned}
s_a &= \frac{1}{n\pi_a} \sum_{i \in \mathcal{V}_a} \theta_i \mathbb{E}[m_i] \approx \frac{1}{rn^2\pi_a} \sum_{b=1}^{k} \sum_{i \in \mathcal{V}_a} \sum_{j \in \mathcal{V}_b} \theta_i^2 \theta_j C_{ab} \mathbb{E}[m_j] \\
&= \frac{1}{rn\pi_a} \sum_{b=1}^{k} \sum_{i \in \mathcal{V}_a} \theta_i^2 C_{ab} s_b \pi_b = \frac{\Phi}{r} \sum_{b=1}^{k} (C\Pi)_{ab} s_b.
\end{aligned} \tag{58}
$$

We thus obtain the relation controlling the stability of the solutions:

$$
C\Pi s = \frac{r}{\Phi} s = \nu_p s, \tag{59}
$$

The transition temperatures are hence given by $r = \nu_p \Phi$, as expected. For $r > c\Phi$ Equation (59) only admits the trivial solution $s = 0$ which represents the paramagnetic configuration. At $r = c\Phi$ there is a paramagnetic-ferromagnetic transition, consistent with the result of (Leone et al., 2002). For smaller values of r there appear other directions (aligned to the eigenvectors of $C\Pi$) along which the Bethe free energy has local minima.

The main result of this appendix is thus to argue that (i) there are only k directions along which the paramagnetic phase can get unstable, hence H_r can have at most k negative eigenvalues, so that $s_{k+1}^{\uparrow}(H_r) \geq 0$ for $r > 1$; (ii) the directions of instability precisely correspond to the eigenvectors of the matrix $C\Pi$.

To conclude, we comment the case $c = O_n(1)$. We just detailed an approach to understand and interpret the problem at hand with tools taken from statistical physics. For mathematical convenience we had to assume the average degree c to be rather large and we were able to predict the exact positions of the transition temperatures, which coincide with the largest eigenvalues of the non-backtracking matrix. Our approach in determining the positions of these eigenvalues is non-rigorous and does not intend to substitute the existing proofs, but rather to give a physical intuition of the problem at hand. In particular, the spectrum of B has been rigorously studied in (Gulikers et al., 2017b; Bordenave et al., 2015) in the regime $c = O_n(1)$. The conclusion above is not altered: the matrix B has k isolated eigenvalues, so only the k smallest eigenvalues of the Bethe-Hessian matrix can become negative, in correspondence to the emergence of new local minima in the Bethe free energy profile. According to our earlier argument, we expect these minima (represented by the eigenvectors of H_r) to be correlated with the class structure of A.

Appendix E. Supporting Arguments for Claim 2

To support Claim 2, we will rely on the following three intermediary results that will be proved subsequently.

Lemma 3 *Let D and A be the degree and adjacency matrices of any graph of size n. Let $r > 1$ and p an integer between 2 and n. If $s_p^{\uparrow}(D - rA) < 0$ then*

$$
s_p^{\downarrow}\left(L_{-s_p^{\uparrow}(D-rA)}^{\mathrm{rw}}\right) = \frac{1}{r}
$$

and both eigenvalues $s_p^{\uparrow}(D - rA)$ and $1/r$ share the same eigenvector.

The fact that the matrix $L^{\mathrm{rw}}_{-s_p^{\uparrow}(D-rA)}$ has an eigenvalue equal to $1/r$ comes easily by construction of $L^{\mathrm{rw}}_{-s_p^{\uparrow}(D-rA)}$. The main result of Lemma 3 is to state that this eigenvalue is the p-th largest. Note that the condition $s_p^{\uparrow}(D-rA) < 0$ could be loosened, but we don't need a stronger result in the following.

Lemma 4 *Let D and A be the degree and adjacency matrices of any arbitrary graph. Let p be an integer ≥ 2. Suppose that there exists $r_p > 1$ such that i/ $\forall\ r \geq r_p$ the eigenvalue $s_p^{\uparrow}(D-rA)$ is simple, and ii/ $s_p^{\uparrow}(D-r_p A) < 0$. Then:*

$$\forall\ r \geq r_p, \quad \partial_r s_p^{\uparrow}(D-rA) < 0.$$

In words, once the function $s_p^{\uparrow}(D-rA)$ becomes negative, it is strictly decreasing.

The eigenvalue simplicity assumption of this lemma is technical and enables us to properly define and manipulate the derivative of the p-th smallest eigenvalue. In case of multiplicity, the tools involved are more complicated (see for instance Greenbaum et al., 2020) and not included here. Note that according to Claim 1 this assumption is verified for a network generated from the DC-SBM because on a large random graph with independent entries all eigenvalues are simple with probability one.

Lemma 5 *Let $\mathcal{G}(\mathcal{V},\mathcal{E})$ be a DC-SBM graph generated according to Equation (1), and D and A its degree and adjacency matrices respectively. Consider $2 \leq p \leq k$ and let ζ_p be defined as per Equation (5). Let $r \geq \zeta_p$. If $s_p^{\uparrow}(D-rA) < 0$ is isolated with associated eigenvector $\boldsymbol{x}_p$, then $\boldsymbol{x}_p$ is an eigenvector of $L^{\mathrm{rw}}_{-s_p^{\uparrow}(D-rA)}$, whose corresponding eigenvalue is also isolated. Otherwise, for $1 \leq p \leq n$ and $r > 1$, if $s_p^{\uparrow}(D-rA)$ belongs to the bulk of non informative eigenvalues of $D-rA$, then the corresponding eigenvalue of $L^{\mathrm{rw}}_{-s_p^{\uparrow}(D-rA)}$ also belongs to the bulk of $L^{\mathrm{rw}}_{-s_p^{\uparrow}(D-rA)}$.*

Based on Lemmas 3-5 and Claim 1 we now proceed to the justification of Claim 2, that we recall here for convenience.

Claim 2 *Consider the graph $\mathcal{G}(\mathcal{V},\mathcal{E})$, built on a sparse DC-SBM as per Equation (1) with k communities. Let ζ_p for $2 \leq p \leq k$ be defined as per Equation (5) (imposing $\zeta_1 = 1$) and $\tau \in \mathbb{R}$ be such that $\zeta_p^2 - 1 \leq \tau \leq c\Phi - 1$. Then, under Assumption 1, for all large n with high probability, the p largest eigenvalues of the matrix L^{rw}_τ are isolated. In particular,*

$$s_p^{\downarrow}(L^{\mathrm{rw}}_{\zeta_p^2-1}) = \frac{1}{\zeta_p}.$$

For simplicity, will use the notation $s_p^{\uparrow}(D-rA) = s_p(r)$. The first part of Claim 2 asserts that for $\zeta_p^2 - 1 \leq \tau \leq c\Phi - 1$ the p largest eigenvalues of the matrix L^{rw}_τ are isolated. According to Claim 1, for $\zeta_p \leq r \leq \sqrt{c\Phi}$ the eigenvalue $s_p(r)$ is isolated. Thanks to Lemma 4, since $s_p(\zeta_p) < 0$ we have: $\partial_r s_p(r) < 0$ for all r in the interval $\zeta_p \leq r \leq \sqrt{c\Phi}$, implying $s_p^{\uparrow}(D-rA) < 0$ for all r in the interval $\zeta_p \leq r \leq \sqrt{c\Phi}$. The hypotheses of Lemma 3 and Lemma 5 are therefore verified. We can then assert that the eigenvalue $1/r$ is the p-th largest of the matrix $L^{\mathrm{rw}}_{-s_p^{\uparrow}(D-rA)}$ and it is isolated.

Further exploiting Lemma 4, letting $\tau(r) = -s_p^{\uparrow}(D - rA)$, we have $\partial_r \tau(r) > 0$ for all r in the interval $\zeta_p \leq r \leq \sqrt{c\Phi}$. The function $\tau(r)$ is thus bijective and increasing on this interval. Since the eigenvalue $s_p^{\uparrow}(D - rA)$ is isolated for all $\zeta_p \leq r \leq \sqrt{c\Phi}$, the corresponding eigenvalue of $L_{\tau(r)}^{\mathrm{rw}}$ equal to $1/r$ is isolated for

$$\zeta_p^2 - 1 = -s_p(\zeta_p) \leq \tau \leq -s_p(\sqrt{c\Phi}).$$

Note now that, for $1 \leq p \leq k$, $s_p^{\uparrow}(D - \sqrt{c\Phi}A) \leq s_{k+1}^{\uparrow}(D - \sqrt{c\Phi}A) = c\Phi - 1$. This implies that for $\zeta_p^2 - 1 \leq \tau \leq c\Phi - 1$, the top p eigenvalues of L_{τ}^{rw} are certainly isolated.

For the particular case where $r = \zeta_p$, by definition of ζ_p, $-s_p^{\uparrow}(D - \zeta_p A) = \zeta_p^2 - 1 > 0$ and the result is straightforwardly obtained by applying Lemma 3, concluding our argument.

We now proceed to the proof of the three lemmas. For the sake of clarity we here enunciate the Courant-Fischer theorem that will be of fundamental use in the sequel.

Theorem 2 (Courant-Fischer, see for instance Bhatia, 2013) *Let $M \in \mathbb{C}^{n \times n}$ be a Hermitian matrix and U a vector subspace of $\mathbb{C}^n$. Then,*

$$s_p^{\uparrow}(M) = \min_{U:\dim(U)=p} \max_{\boldsymbol{z} \in U, \boldsymbol{z} \neq 0} \frac{\boldsymbol{z}^T M \boldsymbol{z}}{\boldsymbol{z}^T \boldsymbol{z}}$$

$$s_p^{\uparrow}(M) = \max_{U:\dim(U)=n-p+1} \min_{\boldsymbol{z} \in U, \boldsymbol{z} \neq 0} \frac{\boldsymbol{z}^T M \boldsymbol{z}}{\boldsymbol{z}^T \boldsymbol{z}}.$$

Proof [Lemma 3] Let $r > 1$. For simplicity of notation, we write $s_p(r) \equiv s_p^{\uparrow}(D - rA)$. Define the matrix M_r as

$$M_r = -s_p(r)I_n + D - rA = D_{-s_p(r)} - rA$$

where we recall the notation $D_{\tau} = D + \tau I_n$. Note that one has $s_p^{\uparrow}(M_r) = 0$. Letting $\tilde{M}_r \equiv D_{-s_p(r)}^{-1/2} M_r D_{-s_p(r)}^{-1/2}$, we then have

$$L_{-s_p(r)}^{\mathrm{sym}} = \frac{1}{r}\left[I_n - \tilde{M}_r\right].$$

Recalling that L_{τ}^{rw} and L_{τ}^{sym} share the same spectrum, one has in particular:

$$s_p^{\downarrow}(L_{-s_p(r)}^{\mathrm{rw}}) = s_p^{\downarrow}(L_{-s_p(r)}^{\mathrm{sym}}) = \frac{1}{r}s_p^{\downarrow}\left(I_n - \tilde{M}_r\right) = \frac{1}{r}\left(1 - s_p^{\uparrow}(\tilde{M}_r)\right).$$

Thus, proving the lemma amounts to proving that $s_p^{\uparrow}(\tilde{M}_r) = 0$, which we do now by first proving that $s_p^{\uparrow}(\tilde{M}_r) \leq 0$ and then that $s_p^{\uparrow}(\tilde{M}_r) \geq 0$.

Denote by $X = (\boldsymbol{x}_1 | \dots | \boldsymbol{x}_p) \in \mathbb{R}^{n \times p}$ the matrix concatenating the eigenvectors associated to the p smallest eigenvalues of M_r. One has:

$$\forall\, 1 \leq q \leq p, \qquad M_r \boldsymbol{x}_q = s_q^{\uparrow}(M_r)\boldsymbol{x}_q \quad \text{and} \quad s_q^{\uparrow}(M_r) \leq s_p^{\uparrow}(M_r) = 0.$$

Now define the vector space V as $V = \mathrm{span}\left(D_{-s_p(r)}^{1/2} X\right)$. Since $s_p(r)$ is strictly negative by hypothesis, $D_{-s_p(r)} \succ 0$ (is positive definite), which in turn implies that $\dim(V) = p$. By

definition of V, $\forall\ \boldsymbol{z} \in V$, $\exists\ \boldsymbol{u} \in \mathbb{R}^p : \boldsymbol{z} = D^{1/2}_{-s_p(r)} X \boldsymbol{u}$. From the Courant-Fischer theorem, we can write:

$$
s_p^\uparrow(\tilde{M}_r) = \min_{U:\dim(U)=p}\ \max_{\boldsymbol{z}\in U, \boldsymbol{z}\neq 0} \frac{\boldsymbol{z}^T \tilde{M}_r \boldsymbol{z}}{\boldsymbol{z}^T \boldsymbol{z}} \leq \max_{\boldsymbol{z}\in V, \boldsymbol{z}\neq 0} \frac{\boldsymbol{z}^T \tilde{M}_r \boldsymbol{z}}{\boldsymbol{z}^T \boldsymbol{z}} = \max_{\boldsymbol{u}\in\mathbb{R}^p, \boldsymbol{u}\neq 0} \frac{\boldsymbol{u}^T X^T D^{1/2}_{-s_p(r)} \tilde{M}_r D^{1/2}_{-s_p(r)} X \boldsymbol{u}}{\boldsymbol{u}^T X^T D_{-s_p(r)} X \boldsymbol{u}}
$$

i.e.:

$$
s_p^\uparrow(\tilde{M}_r) \leq \max_{\boldsymbol{u}\in\mathbb{R}^p, \boldsymbol{u}\neq 0} \frac{\boldsymbol{u}^T X^T M_r X \boldsymbol{u}}{\boldsymbol{u}^T X^T D_{-s_p(r)} X \boldsymbol{u}} = \max_{\boldsymbol{u}\in\mathbb{R}^p, \boldsymbol{u}\neq 0} \frac{\sum_{q=1}^{p}(\boldsymbol{x}_q^T \boldsymbol{u})^2 s_q^\uparrow(M_r)}{\boldsymbol{u}^T X^T D_{-s_p(r)} X \boldsymbol{u}} \leq 0
$$

where for the last step we exploited $s_q^\uparrow(M_r) \leq 0$ and $D_{-s_p(r)} \succ 0$. We thus conclude that $s_p^\uparrow(\tilde{M}_r) \leq 0$.

To prove that the equality holds, we exploit the second relation of the Courant-Fischer theorem. We define $\bar{X} = (\boldsymbol{x}_p|\ldots|\boldsymbol{x}_n) \in \mathbb{R}^{n\times(n-p+1)}$ the matrix concatenating the eigenvectors associated to the $n - p + 1$ largest eigenvalues of M_r. For $q \geq p$, $s_q^\uparrow(M_r) \geq 0$. We further define $W = \text{span}\left(D^{1/2}_{-s_p(r)} \bar{X}\right)$, satisfying $\dim(W) = n - p + 1$. We can write:

$$
s_p^\uparrow(\tilde{M}_r) = \max_{U:\ \dim(U)=n-p+1}\ \min_{\boldsymbol{z}\in U, \boldsymbol{z}\neq 0} \frac{\boldsymbol{z}^T \tilde{M}_r \boldsymbol{z}}{\boldsymbol{z}^T \boldsymbol{z}} \geq \min_{\boldsymbol{z}\in W, \boldsymbol{z}\neq 0} \frac{\boldsymbol{z}^T \tilde{M}_r \boldsymbol{z}}{\boldsymbol{z}^T \boldsymbol{z}}
$$

i.e.:

$$
s_p^\uparrow(\tilde{M}_r) \geq \min_{\boldsymbol{u}\in\mathbb{R}^{n-p+1}, \boldsymbol{u}\neq 0} \frac{\boldsymbol{u}^T \bar{X}^T M_r \bar{X} \boldsymbol{u}}{\boldsymbol{u}^T \bar{X}^T D_{-s_p(r)} \bar{X} \boldsymbol{u}} = \min_{\boldsymbol{u}\in\mathbb{R}^p, \boldsymbol{u}\neq 0} \frac{\sum_{q=p}^{n}(\boldsymbol{x}_q^T \boldsymbol{u})^2 s_q^\uparrow(M_r)}{\boldsymbol{u}^T \bar{X}^T D_{-s_p(r)} \bar{X} \boldsymbol{u}} \geq 0.
$$

As a consequence, $s_p^\uparrow(\tilde{M}_r) \geq 0$. Combining both inequalities, we obtain that $s_p^\uparrow(\tilde{M}_r) = 0$.

The fact that the eigenvectors are shared comes from the following. Let $\boldsymbol{x}_p(r)$ be the eigenvector of $D - rA$ associated to $s_p(r)$, *i.e.*, $[D - rA]\boldsymbol{x}_p(r) = s_p(r)\boldsymbol{x}_p(r)$. This can be re-written as $L^{\text{rw}}_{-s_p(r)}\boldsymbol{x}_p(r) = \frac{1}{r}\boldsymbol{x}_p(r)$.

$\blacksquare$

Proof [Lemma 4] Once again, let us write $s_p(r) = s_p^\uparrow(D - rA)$ to lighten notations. Let p, r_p be an integer ≥ 2 and a scalar > 1 (if they exist) such that $s_p(r_p) < 0$. Let $r \geq r_p$. As $s_p(r)$ is an eigenvalue, $\exists\ \boldsymbol{x}_p(r)$ with $||\boldsymbol{x}_p(r)||^2 = 1$ such that:

$$
[D - rA]\boldsymbol{x}_p(r) = s_p(r)\boldsymbol{x}_p(r),
$$

which implies in particular that $\boldsymbol{x}_p(r)^T A \boldsymbol{x}_p(r) = \frac{1}{r}\boldsymbol{x}_p(r)^T D_{-s_p(r)}\boldsymbol{x}_p(r)$. As we suppose $s_p(r)$ to be simple, we can apply the eigenvalue perturbation theorem (see, for instance Greenbaum et al., 2020):

$$
\partial_r s_p(r) = -\boldsymbol{x}_p^T(r) A \boldsymbol{x}_p(r) = -\frac{1}{r}\boldsymbol{x}_p(r)^T D_{-s_p(r)}\boldsymbol{x}_p(r). \tag{60}
$$

In Equation (60), $D_{-s_p(r_p)} \succ 0$ (as $s_p(r_p) < 0$ by hypothesis), consequently $\partial_r s_p(r)|_{r=r_p} < 0$. We now want to show that for all $r > r_p$, $\partial_r s_p(r) < 0$. We proceed with a proof by contradiction.

Suppose that there exists a value $r' > r_p$ such that $\partial_r s_p(r)|_{r=r'} \geq 0$. From Equation (60) it follows that a necessary (but not sufficient) condition to be verified is that $s_p(r') \geq 0$. From a continuity argument on the function $s_p(r)$, and as $s_p(r_p) < 0$ and $s_p(r') \geq 0$ with $r' > r_p$, there exists $r'' \in (r_p, r')$ such that $\partial_r s_p(r)|_{r=r''} > 0$ and $s_p(r'') < 0$. Invoking once again Equation (60), no such r'' can exist, invalidating the hypothesis we made by absurd. We thus conclude that

$$\forall\ r \geq r_p, \quad \partial_r s_p(r) < 0, \tag{61}$$

finishing the proof.

∎

Proof [Lemma 5] Consider the values $\{\zeta_p\}_{2 \leq p \leq k}$ as defined in Eq.(5). As $\mathcal{G}$ is generated from a DC-SBM we know from Claim 1 that these $k - 1$ values of ζ_p exist with high probability. Define $\zeta_{k+1} = \sqrt{c\Phi}$ and fix p an integer such that $2 \leq p \leq k$. Let us write once more $s_p(r) = s_p^{\uparrow}(D - rA)$ to lighten notations. Note that $s_p^{\uparrow}(H_{\zeta_p}) = 0$, $\forall\ 2 \leq p \leq k$ and $s_{k+1}^{\uparrow}(H_{\zeta_{k+1}}) = o_n(1)$, as shown in (Saade et al., 2014). We thus have $s_p(\zeta_p) < 0$, $\forall\ 2 \leq p \leq k + 1$ and it is isolated by hypothesis. We can apply Lemma 4, for $r_p = \zeta_p$: for all $r \geq \zeta_p$, $\partial_r s_p(r) < 0$ and $s_p(r) < 0$. Consequently, there exists a unique value of $r' \geq r$, satisfying $s_{p+1}(r') = s_p(r)$. From lemma 3, we know that:

- as $s_p(r) < 0$: $\boldsymbol{x}_p(r)$, the eigenvector of $D - rA$ associated to $s_p(r)$, is also the eigenvector of $L^{\text{rw}}_{-s_p(r)}$ associated to its p-th largest eigenvalue, $1/r$.

- as $s_{p+1}(r') < 0$: $\boldsymbol{x}_{p+1}(r')$, the eigenvector of $D - r'A$ associated to $s_{p+1}(r')$, is also the eigenvector of $L^{\text{rw}}_{-s_{p+1}(r')} = L^{\text{rw}}_{-s_p(r)}$ associated to its $(p+1)$-th largest eigenvalue, $1/r'$.

Thus, the p-th (resp. $(p+1)$-th) largest eigenvalue of $L^{\text{rw}}_{-s_p(r)}$ is $1/r$ (resp. $1/r'$). Consider $r \geq \zeta_p$. By hypothesis, $s_p(r)$ is an isolated eigenvalue, that is, we can write

$$O_n(1) = s_{p+1}(r) - s_p(r) = s_{p+1}(r) - s_{p+1}(r') = \int_{r'}^{r} dx\ \partial_x s_p(x) = \kappa(r - r')$$

for some constant $\kappa = O_n(1)$ independent of n, representing the average value of $\partial_x s_p(x)$ on the integration interval. The constant $\kappa = O_n(1)$ because for any $r, r' = O_n(1)$, we have that $s_p(r), s_p(r') = O_n(1)$. The eigengap for $L^{\text{rw}}_{-s_p(r)}$ is $1/r - 1/r' = (r' - r)/(rr')$, thus

$$O_n(1) = O_n(s_{p+1}(r) - s_p(r)) = O_n(r - r') = O_n\left(\frac{1}{r} - \frac{1}{r'}\right) = O_n(1).$$

So, if $s_p(r)$ is isolated, the eigenvalue $1/r$ of L^{rw}_{τ} is isolated as well.

On the other hand, if $s_p(r)$ is in the bulk, then $\exists\ q\ :\ |s_p(r) - s_q(r)| = o_n(1)$. By an argument of continuity on $s_p(r)$, one can analogously define an r' satisfying $s_p(r) = s_q(r)$, concluding that in this case $|r - r'| = o_n(1)$ and so eigenvalues in the bulk are mapped into eigenvalues in the bulk.

∎

50

Appendix F. Fast Estimate of the ζ_p

We here describe a fast algorithm to estimate the values of the ζ_p using the Courant-Fisher theorem.

Let us denote by $X_{r_t}^T = (\boldsymbol{x}_1^T(r_t), \dots, \boldsymbol{x}_p^T(r_t))^T \in \mathbb{R}^{n \times p}$ where $\boldsymbol{x}_p(r_t)$ is the eigenvector of H_{r_t} corresponding to the p-th smallest eigenvalue, $s_p^{\uparrow}(H_{r_t})$, while $S_{r_t} = \text{diag}\left(s_1^{\uparrow}(H_{r_t}), \dots, s_p^{\uparrow}(H_{r_t})\right)$. For another value $r' \neq r_t$, and applying the Courant-Fischer theorem, we can write

$$s_p^{\uparrow}(H_{r'}) = \min_{U:\dim(U)=p} \max_{\boldsymbol{z} \in U, \boldsymbol{z} \neq 0} \frac{\boldsymbol{z}^T H_{r'} \boldsymbol{z}}{\boldsymbol{z}^T \boldsymbol{z}} \leq \max_{\boldsymbol{u} \in \mathbb{R}^p} \frac{\boldsymbol{u}^T X_{r_t}^T H_{r'} X_{r_t} \boldsymbol{u}}{\boldsymbol{u}^T \boldsymbol{u}} = s_1^{\downarrow}(X_{r_t}^T H_{r'} X_{r_t})$$

i.e.:

$$s_p^{\uparrow}(H_{r'}) \leq s_1^{\downarrow}(X_{r_t}^T[(r'^2 - r_t^2)I_n - (r' - r_t)A + H_{r_t}]X_{r_t}) = (r'^2 - r_t^2) + s_1^{\downarrow}(S_{r_t} - (r' - r_t)X_{r_t}^T A X_{r_t}).$$

We can further simplify the earlier expression by exploiting the identity:

$$S_{r_t} = X_{r_t}^T H_{r_t} X_{r_t} = (r_t^2 - 1)I_p + X_{r_t}^T D X_{r_t} - r_t X_{r_t}^T A X_{r_t}.$$

We thus obtain

$$s_p^{\uparrow}(H_{r'}) \leq \frac{1}{r_t}\left[(r' - r_t)(1 + r'r_t) + s_1^{\downarrow}\big((r_t - r')X_{r_t}^T D X_{r_t} + r'S_{r_t}\big)\right] \equiv \frac{f_{r_t}(r')}{r_t}. \tag{62}$$

We now study the function $f_{r_t}(r')$ for $r_t \in (\zeta_p, \sqrt{\rho(B)})$ and define $r_{t+1} \in (\zeta_p, r_t)$ as the solution (if it exists) to

$$f_{r_t}(r_{t+1}) = 0. \tag{63}$$

The idea is to iteratively approach ζ_p from the right and substitute $r_t \leftarrow r_{t+1}$. If $r_{t+1} \in (\zeta_p, r_t)$ as defined above exists, then $\{r_t\}_{t \geq 0}$ is a lower-bounded decreasing sequence: it thus converges to a limit r_∞ (potentially different from ζ_p). Exploiting (Rellich and Berkowitz, 1969, Theorem in I.4), denoting with A, B two Hermitian matrices and λ a scalar, $s_1^{\downarrow}(A + \lambda B)$ is a convex function of λ. As a consequence, the function $f_{r_t}(r')$ is convex and has either no root or two roots. Since $\lim_{|r'| \to \infty} f_{r_t}(r') = +\infty$, it is enough to find a value of r' for which $f_{r_t}(r') < 0$ to prove that this function has two roots. With a straightforward computation, one can verify that $f_{r_t}(r_t) < 0$, so $f_{r_t}(r')$ has two roots, satisfying $r_{t+1} < r_t$ and $r^+ > r_t$. By construction, since $s_p^{\uparrow}(H_{r'})$ is negative in the considered interval and $s_p^{\uparrow}(H_{r'}) \leq f_{r_t}(r')/r_t$, if $f_{r_t}(r_{t+1}) = 0$, then $\zeta_p \leq r_{t+1}$. Consequently, $f_{r_t}(r_{t+1}) = 0$ has a unique solution satisfying

$$\zeta_p \leq r_{t+1} < r_t$$

and the algorithm converges to $r_\infty = \lim_{t \to \infty} r_t \geq \zeta_p$. We are left to prove that $r_\infty = \zeta_p$. By convergence of r_t, we have $r_{t+1} - r_t = o_t(1)$. Plugging this relation solution into Equations (62,63), we obtain

$$s_1^{\downarrow}(r_{t+1}S_{r_{t+1}}) = r_{t+1}s_p^{\uparrow}(H_{r_{t+1}}) = o_t(1).$$
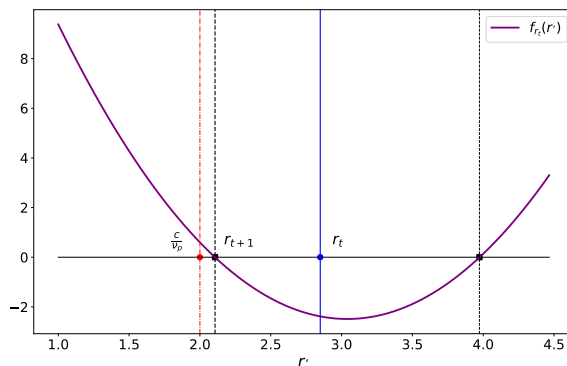
Figure 10: Plot of the $f_{r_t}(r')$ (in purple). The blue continuous line indicates the value of r_t, the red dashed dotted line the theoretical value of $\zeta_p = c/\nu_p$, the two dashed black lines are the two roots of $f_{r_t}(r')$, the smaller of which is r_{t+1}. For this simulation $n = 50\,000$, $k = 2$, $c = 5$, $c_{\text{out}} = 2.5$, $r = \sqrt{c\Phi}$, $\theta_i \sim [\mathcal{U}(3,10)]^3$, $\boldsymbol{\pi} \propto I_k$.

Since $s_p^\uparrow(H_r) = 0$ has a unique solution $(r = \zeta_p)$ in the interval $r \in (1, \sqrt{\rho(B)})$, we obtain $r_t = \zeta_p + o_t(1)$ and so

$$r_\infty = \zeta_p.$$

The initial value of r can be chosen as $r_0 = \zeta_{p+1}$ (setting $\zeta_{k+1} = \sqrt{\rho(B)}$ that certainly falls in the right interval for all the ζ_p.

References

Lada A Adamic and Natalie Glance. The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pages 36–43. ACM, 2005.

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*, 41 (4):2097–2122, 2013.

Albert-László Barabási. Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1987):20120375, 2013.

Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

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

Hans A Bethe. Statistical theory of superlattices. *Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences*, 150(871):552–575, 1935.

Rajendra Bhatia. *Matrix analysis*, volume 169. Springer Science & Business Media, 2013.

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

Béla Bollobás, Svante Janson, and Oliver Riordan. The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31(1):3–122, 2007.

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*, pages 1347–1357. IEEE, 2015.

Stephen P Borgatti and Daniel S Halgin. On network theory. *Organization science*, 22(5): 1168–1181, 2011.

F.R.K. Chung. *Spectral graph theory*. Amer Mathematical Society, 1997.

Simon Coste and Yizhe Zhu. Eigenvalues of the non-backtracking operator detached from the bulk. *Random Matrices: Theory and Applications*, 10(03):2150028, 2021.

Dragoš Cvetković, Peter Rowlinson, and Slobodan K Simić. Signless laplacians of finite graphs. *Linear Algebra and its applications*, 423(1):155–171, 2007.

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*, pages 4039–4049, 2019.

Lorenzo Dall'Amico, Romain Couillet, and Nicolas Tremblay. Community detection in sparse time-evolving graphs with a dynamical bethe-hessian. *Advances in Neural Information Processing Systems*, 33, 2020.

Lorenzo Dall'Amico, Romain Couillet, and Nicolas Tremblay. Nishimori meets bethe: a spectral method for node classification in sparse weighted graphs. *arXiv preprint arXiv:2103.03561*, 2021.

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*, 84(6):066106, 2011.

Amir Dembo, Andrea Montanari, et al. Gibbs measures and phase transitions on sparse random graphs. *Brazilian Journal of Probability and Statistics*, 24(2):137–211, 2010.

Miroslav Fiedler. Algebraic connectivity of graphs. *Czechoslovak mathematical journal*, 23 (2):298–305, 1973.

Santo Fortunato. Community detection in graphs. *Physics reports*, 486(3-5):75–174, 2010.

Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

Anne Greenbaum, Ren-cang Li, and Michael L Overton. First-order perturbation theory for eigenvalues and eigenvectors. *SIAM review*, 62(2):463–482, 2020.

Roger Guimera, Marta Sales-Pardo, and Luís A Nunes Amaral. Modularity from fluctuations in random graphs and complex networks. *Physical Review E*, 70(2):025101, 2004.

Lennart Gulikers, Marc Lelarge, and Laurent Massoulié. Non-Backtracking Spectrum of Degree-Corrected Stochastic Block Models. In *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*, volume 67 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 44:1–44:27, 2017a. ISBN 978-3-95977-029-3. doi: 10.4230/LIPIcs.ITCS.2017.44.

Lennart Gulikers, Marc Lelarge, and Laurent Massoulié. A spectral method for community detection in moderately sparse degree-corrected stochastic block models. *Advances in Applied Probability*, 49(3):686–721, 2017b.

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*, 28(5):3002–3027, 2018.

Paul W Holland and Samuel Leinhardt. Transitivity in structural models of small groups. *Comparative group studies*, 2(2):107–124, 1971.

Jiashun Jin et al. Fast community detection by score. *The Annals of Statistics*, 43(1):57–89, 2015.

Antony Joseph and Bin Yu. Impact of regularization on spectral clustering. *The Annals of Statistics*, 44(4):1765–1791, 2016.

Brian Karrer and Mark EJ Newman. Stochastic blockmodels and community structure in networks. *Physical review E*, 83(1):016107, 2011.

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*, 110(52):20935–20940, 2013.

Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi. Benchmark graphs for testing community detection algorithms. *Physical review E*, 78(4):046110, 2008.

Can M Le and Elizaveta Levina. Estimating the number of communities in networks by spectral methods. *arXiv preprint arXiv:1507.00827*, 2015.

Can M Le, Elizaveta Levina, and Roman Vershynin. Concentration of random graphs and application to community detection. In *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018*, pages 2925–2943. World Scientific, 2018.

J. Lei and A. Rinaldo. Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43(1):215–237, 2015.

M Leone, A Vázquez, A Vespignani, and Riccardo Zecchina. Ferromagnetic ordering in graphs with arbitrary degree distribution. *The European Physical Journal B-Condensed Matter and Complex Systems*, 28(2):191–197, 2002.

Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford large network dataset collection. `http://snap.stanford.edu/data`, June 2014.

David Lusseau, Karsten Schneider, Oliver J Boisseau, Patti Haase, Elisabeth Slooten, and Steve M Dawson. The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology*, 54(4):396–405, 2003.

Laurent Massoulié. Community detection thresholds and the weak ramanujan property. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 694–703. ACM, 2014.

Marc Mezard, Marc Mezard, and Andrea Montanari. *Information, physics, and computation*. Oxford University Press, 2009.

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

Elchanan Mossel, Joe Neeman, and Allan Sly. Belief propagation, robust reconstruction and optimal recovery of block models. In *Conference on Learning Theory*, pages 356–370, 2014.

Elchanan Mossel, Joe Neeman, and Allan Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162(3-4):431–461, 2015.

M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103:8577–8582, 2006. ISSN 0027-8424,1091-6490. doi: 10.1073/pnas.0601602103.

Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2): 167–256, 2003.

Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.

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

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

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

Franz Rellich and Joan Berkowitz. *Perturbation theory of eigenvalue problems*. CRC Press, 1969.

Yousef Saad. *Numerical Methods for Large Eigenvalue Problems*. Classics in Applied Mathematics 66. SIAM, 2nd. edition, 2011. ISBN 1-61197-072-5 978-1-61197-072-2.

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

Justin Salez. *Some implications of local weak convergence for sparse random graphs*. PhD thesis, Université Pierre et Marie Curie-Paris VI; Ecole Normale Supérieure de Paris ..., 2011.

Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *Departmental Papers (CIS)*, page 107, 2000.

Janusz A. Suchecki, Krzysztof; Hołyst. Ising model on two connected barabasi-albert networks. *Physical Review E*, 74:011122, 2006. ISSN 1539-3755,1550-2376.

Audrey Terras. *Zeta functions of graphs: a stroll through the garden*, volume 128. Cambridge University Press, 2010.

Vincent A Traag, Ludo Waltman, and Nees Jan van Eck. From louvain to leiden: guaranteeing well-connected communities. *Scientific reports*, 9(1):1–12, 2019.

Nicolas Tremblay and Andreas Loukas. Approximating Spectral Clustering via Sampling: A Review. In Frédéric Ros and Serge Guillaume, editors, *Sampling Techniques for Supervised or Unsupervised Tasks*, pages 129–183. Springer International Publishing, 2020.

Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4): 395–416, 2007.

Yusuke Watanabe and Kenji Fukumizu. Graph zeta function in the bethe free energy and loopy belief propagation. In *Advances in Neural Information Processing Systems*, pages 2017–2025, 2009.

Wayne W Zachary. An information flow model for conflict and fission in small groups. *Journal of anthropological research*, 33(4):452–473, 1977.

Pan Zhang. Nonbacktracking operator for the ising model and its applications in systems with multiple states. *Physical Review E*, 91(4):042120, 2015.

Yilin Zhang and Karl Rohe. Understanding regularized spectral clustering via graph conductance. In *Advances in Neural Information Processing Systems*, pages 10631–10640, 2018.