

Eigenvalues of the non-backtracking operator detached from the bulk

Simon Coste

INRIA Paris, DYOGENE Team, Office C330, Paris, France
simon.coste@inria.fr

Yizhe Zhu*

*Department of Mathematics, University of California
San Diego, La Jolla, CA 92093, USA*
yiz084@ucsd.edu

Received 14 February 2020

Revised 29 June 2020

Accepted 30 July 2020

Published 19 September 2020

We describe the non-backtracking spectrum of a stochastic block model with connection probabilities $p_{\text{in}}, p_{\text{out}} = \omega(\log n)/n$. In this regime we answer a question posed in [L. Dall’Amico, R. Couillet and N. Tremblay, Revisiting the Bethe–Hessian: Improved community detection in sparse heterogeneous graphs, in *Advances in Neural Information Processing Systems* (2019), pp. 4039–4049] regarding the existence of a real eigenvalue “inside” the bulk, close to the location $\frac{p_{\text{in}}+p_{\text{out}}}{p_{\text{in}}-p_{\text{out}}}$. We also introduce a variant of the Bauer–Fike theorem well suited for perturbations of quadratic eigenvalue problems, which could be of independent interest.

Keywords: Non-backtracking operator; stochastic block model; non-Hermitian perturbation; quadratic eigenvalue problem.

Mathematics Subject Classification 2020: 60C05, 05C80, 60B20

1. Introduction

For any real matrix A with size $n \times n$, its non-backtracking operator B is the real matrix indexed by the coordinates of the non-zero entries of A , and is defined by

$$B_{(i,j),(k,\ell)} = A_{k,\ell} \mathbf{1}_{j=k, i \neq \ell}. \quad (1.1)$$

The non-backtracking matrix of a graph is the non-backtracking matrix of its adjacency matrix, and it is closely related to the Zeta function of the graph [29]. Its

*Corresponding author.

spectrum was first studied in the case of finite graphs and their universal covers [5, 7, 19, 20, 27, 29]. Recently, the non-backtracking operator attracted a lot of attention from random graph theory as a very powerful tool. In the spectral theory of random graphs, it was a key element in a new proof of the Alon–Friedman theorem for random regular graphs [11]. In the same vein it has been used later to study the eigenvalues of random regular hypergraphs [16], random bipartite biregular graphs [13] and homogeneous or inhomogeneous Erdős–Rényi graphs [3, 8, 9, 12, 18, 21, 31]. Very recently, the real eigenvalues were used to prove estimates on the vector-coloring number of a graph [6].

Most of the results focus on the eigenvalues of large magnitude, those which lie outside the bulk of the spectrum. They are known to be the “most informative” eigenvalues, as they capture some essential features about the structure of the graph. For instance, in community detection, the appearance of certain outliers indicates when the community structure can be recovered [12, 21]. A cornerstone result was that even in the difficult *dilute* case, where the connection probabilities are of order $1/n$, reconstruction was feasible (under some condition) by looking at the eigenvalues of the non-backtracking matrix appearing outside the “bulk” of eigenvalues.

It was recently observed in [15, Sec. 3.2] that in fact, there is a real eigenvalue isolated *inside* the bulk that corresponds to the ratio of the two largest eigenvalues *outside* the bulk, as displayed in the right panel of Fig. 1. Recall that B is non-Hermitian with a complex spectrum, so that “inside” the bulk is understood

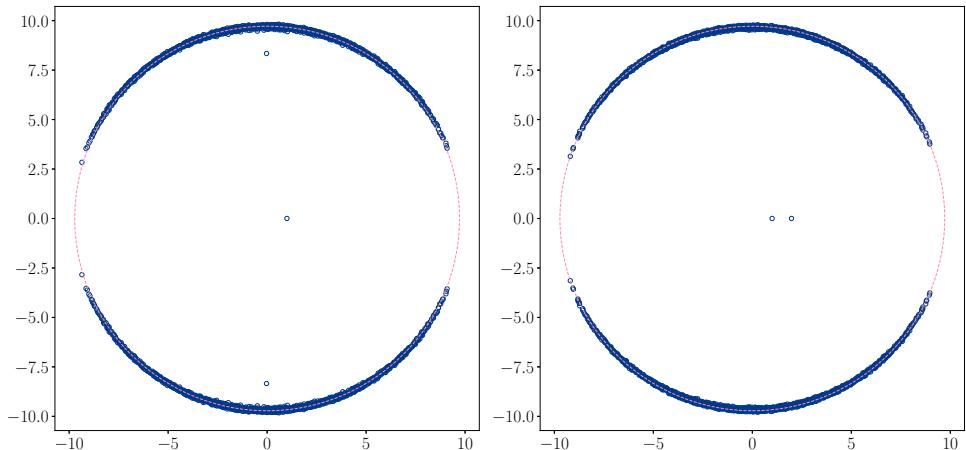


Fig. 1. (Color online) The non-backtracking spectrum of two SBM graphs. The circle of radius $\sqrt{\alpha - 1}$, where α is the mean degree, is drawn in pink. We have cropped the pictures, the outliers do not appear. On the left panel we took $p = q = (\log n)^2/n$, a classical Erdős–Rényi graph. On the right panel, we took $p = 3(\log n)^2/n$ and $q = (\log n)^2/n$, which corresponds to a two-block SBM. We clearly see the two bulk insiders around 1 and 2. In both cases the number of vertices is $n = 1000$.

as eigenvalues inside the circle of the spectrum. To the best of our knowledge, this phenomenon has not been rigorously studied yet. In this paper, we prove the existence of this real eigenvalue inside the bulk for the stochastic block model (SBM) in the regime where the mean degree goes to infinity faster than $\log n$.

Notations

Throughout the paper we will adopt the conventional notations $a_n = o(b_n)$ when $\lim_{n \rightarrow \infty} a_n/b_n = 0$, $a_n = \omega(b_n)$ when $\lim_{n \rightarrow \infty} a_n/b_n = \infty$ and $a_n = O(b_n)$ when $|a_n/b_n|$ is bounded. All the results depend on the parameter n , the size of the graph, which is seen as large through $n \rightarrow \infty$. For any matrix M , we denote the spectral norm of M as $\|M\|$.

1.1. Setting: The SBM in the logarithmic regime

Consider a stochastic block model $G(n, p, q)$ with an even number n of vertices, two blocks of equal size $n/2$, and two probability parameters p, q : if i, j are vertices in the same block then they are connected with probability p , and if they are in different blocks they are connected with probability q . We will place ourselves under the regime

$$p, q = \frac{\omega(\log n)}{n}, \quad p, q \rightarrow 0, \quad C_1 \leq \frac{|p - q|}{p + q} \leq C_2 \quad (\text{A})$$

for some constants $C_1, C_2 \in (0, 1)$. The last condition is technical: we will see that it is here to ensure that two separated outliers appear in the spectrum of the adjacency matrix, and are of the same order. This assumption is crucial in our perturbation analysis, see Remark 4.1 for further discussion.

It is known (see, for example, [10, Chap. 3]) that when $q = p = \omega(\log n/n)$, the $G(n, p, p)$ graph (the Erdős–Rényi model) is “almost regular” in the sense that all degrees are concentrated around $(n - 1)p$. In general, it can be shown that when $p, q = \omega(\log n)/n$ then $G(n, p, q)$ is almost regular with degrees concentrated around

$$\frac{np}{2} - p + \frac{nq}{2} = \frac{n(p + q)}{2} - p.$$

See Sec. 3.4 for a proof. For this reason, we will denote by α the mean degree and by β the mean difference degree:

$$\alpha := \mathbb{E}[d_i] = \frac{n(p + q)}{2} - p,$$

$$\beta := \mathbb{E}[d_i^{\text{in}}] - \mathbb{E}[d_i^{\text{out}}] = \left(\frac{n}{2} - 1\right)p - \frac{n}{2}q = \frac{n}{2}(p - q) - p,$$

where d_i is the number of neighbors of vertex i , and d_i^{out} (respectively d_i^{in}) is the number of neighbors of vertex i which do not have the same type as i (respectively which have the same type as i). Under assumption (A), β can be either positive or negative, and α, β are of the same order.

Our assumptions (A) imply that the mean degree α is $\omega(\log n)$ and the mean difference degree β has the same order as α . Finally, the adjacency matrix of the graph is the $n \times n$ matrix A defined by $A_{i,j} = \mathbf{1}_{i \sim j}$, where $i \sim j$ denotes the event that i and j are connected.

1.2. Main results

Let $G = (V, E)$ be any finite graph with adjacency matrix A . The Ihara–Bass formula gives a connection between the spectrum of B defined in (1.1) and a *quadratic eigenvalue problem*: for any complex z ,

$$\det(B - z\mathbf{I}) = (z^2 - 1)^{|E| - |V|} \det(z^2\mathbf{I} - zA + D - \mathbf{I}), \quad (1.2)$$

see [7, 20, 27]. The zeros of the polynomial $z \mapsto \det(z^2\mathbf{I} - zA + D - \mathbf{I})$ are usually called, with an abuse of language, the *non-backtracking spectrum* of B , the two additional eigenvalues ± 1 appearing with multiplicity $|E| - |V|$ in (1.2) being usually called *trivial*. The non-backtracking spectrum can be expressed as the eigenvalues of a smaller matrix H :

$$H = \begin{bmatrix} A & \mathbf{I} - D \\ \mathbf{I} & 0 \end{bmatrix}, \quad (1.3)$$

where $D = \text{diag}(A\mathbf{1})$ is the diagonal degree matrix. This representation of the spectrum of B in terms of H is extremely useful: to compute the spectrum of B , we do not have to construct the matrix B , and we can analyze the spectrum of B directly from H .

Using these facts, we answer the question posed in [15] regarding the existence of an isolated eigenvalue inside the bulk, at least under the assumptions (A). We also give a detailed description of the non-backtracking spectrum that is similar to the one given in [31] for Erdős–Rényi graphs.

Theorem 1.1. *Let B be the non-backtracking operator of a stochastic block model $G(n, p, q)$ satisfying assumption (A). We order the $2|E|$ eigenvalues of B by decreasing modulus: $|\lambda_1(B)| \geq \dots \geq |\lambda_{2|E|}(B)|$.*

With probability $1 - o(1)$, the spectrum of B can be described as follows. First, the smallest eigenvalues in modulus are the trivial eigenvalues -1 and 1 , each with multiplicity $|E| - |V|$.

Then, in the non-trivial eigenvalues $\lambda_1(B), \dots, \lambda_{2n}(B)$, there are four real eigenvalues which are isolated, two “outliers”

$$\lambda_1(B) = \alpha + O(\alpha^{3/4}), \quad \lambda_2(B) = \beta + O(\alpha^{3/4}), \quad (1.4)$$

and two “insiders”

$$\lambda_{2n-1}(B) = \frac{\alpha}{\beta} + o(1) \quad \text{and} \quad \lambda_{2n}(B) = 1. \quad (1.5)$$

All the other eigenvalues $\lambda_k(B)$ with $k \in \{3, \dots, 2n - 3\}$ are located within distance $o(\sqrt{\alpha})$ of a circle of radius $\sqrt{\alpha - 1}$. Moreover, the real parts of eigenvalues of $\frac{B}{\sqrt{\alpha}}$ are asymptotically distributed as the semi-circle distribution supported on $[-1, 1]$.

To present our approach in the most efficient and clear way, we state and prove the theorem in the simplest regime, when there are only two blocks and when the community structure appears in the spectrum through the presence of an extra outlier near β . It is straightforward to check that our proof can be extended to a diversity of settings, when the mean degree is $\omega(\log n)$. We state this as an informal result:

Assume that with high probability the degree of each vertex is $(1 + o(1))\alpha$ and the spectrum of the adjacency matrix has $k = O(1)$ outliers far outside $[-2\sqrt{\alpha}, 2\sqrt{\alpha}]$, say $\lambda_1 \approx \alpha$, and $\lambda_2, \dots, \lambda_k$. Assume $\lambda_1, \dots, \lambda_k$ are of the same order. Then with high probability its non-backtracking spectrum will have k eigenvalues near λ_i for $i \in [k]$, then k eigenvalues near α/λ_i , and all the other eigenvalues will be located within distance $o(\sqrt{\alpha})$ of the circle of radius $\sqrt{\alpha - 1}$.

Remark 1.1. The concentration results we have in Sec. 3 work for general inhomogeneous random graphs with outliers of the same order. Our modified Bauer–Fike theorem given in Theorem 2.1 works for general inhomogeneous random graphs as well, as long as each vertex has almost regular degree $(1 + o(1))\alpha$. Therefore all the analysis in Sec. 4 can be extended to the case we mentioned above for k outliers.

The presence of bulk insiders in Theorem 1.1 and in the preceding statement are illustrated in Fig. 1 for a realization of an Erdős–Rényi graph and a realization of an SBM graph. Note that the description of the spectrum of B in the preceding theorem is much more precise than [31, Theorem 1.5]. This comes from the fact that their perturbation parameter $R = c\sqrt{\log n/p}$ goes to infinity (see [31, Theorem 1.5] for the exact statement, where the scaling parameter is different from ours). Our method includes a tailored version of the Bauer–Fike theorem suited for perturbations of matrices like (1.3), which yields perturbation bounds that are better than the classical Bauer–Fike theorem in terms of the order of magnitude, and without which the existence of the two eigenvalues at 1 and near β/α would not follow. We think such variants of the Bauer–Fike theorem could be of independent interest.

1.3. Bulk insiders and community detection

The real eigenvalue of B inside the bulk is closely related to community detection problems for SBMs. An interesting heuristic spectral algorithm based on the Bethe Hessian matrix was proposed in [26, 32]. The Bethe Hessian matrix, sometimes called deformed Laplacian [15, 17], is defined as

$$H(r) := (r^2 - 1)I + D - rA,$$

where $r \in \mathbb{R}$ is a regularizer to be carefully tuned. It is conjectured in [26] that a spectral algorithm based on the eigenvectors associated with the negative

eigenvalues of $H(r)$ with $r = \sqrt{\rho(B)}$ is able to reach the information-theoretic threshold confirmed in [12, 23–25] for community detection in the dilute regime. In a subsequent work [15], the authors crafted a spectral algorithm based on $H(r)$ with $r = \alpha/\beta$ and empirically showed it outperforms already known spectral algorithms. Their choice of r was motivated by the conjectured value of the real eigenvalue inside the bulk. The gain in using $H(r)$ instead of B in spectral algorithms mainly comes from the fact that $H(r)$ has a smaller dimension than B , is Hermitian, and is easiest to build from A — nearly no preprocessing is needed, in contrast with non-backtracking matrices [12, 21], self-avoiding path matrices [23] or graph powering matrices [1].

The relation between the Bethe Hessian matrix and the non-backtracking operator is given by the Ihara–Bass formula (see [12] above). Therefore, a good understanding of the real eigenvalues of the non-backtracking operator is the first step towards understanding the theoretical guarantee of the heuristic algorithms proposed in [15, 26].

Unfortunately, our proof techniques do not work in the dilute regime. In the regime studied in this paper, community detection problems are now very well understood and clustering based on the second eigenvector of A has been proven to yield *exact reconstruction* (see [2]). Our result should instead be seen as a preliminary step in view of (1) proving the existence of bulk insiders in the dilute regime, (2) showing their usefulness in practical reconstruction. It will be helpful in practice to have a better understanding of the eigenvectors for the Bethe Hessian matrix, and we leave it as a future direction.

The key obstacle is the lack of concentration of degrees profiles, which tells us random graphs with bounded expected degrees are far away from being “roughly” regular (see also the discussion in Sec. 3.4). Without this property, our perturbation analysis does not apply.

Organization of the paper

In Sec. 2, we first state some classical facts on the non-backtracking spectrum of graphs then we state and prove a perturbation theorem which is well suited for quadratic eigenvalue problems and improves the classical Bauer–Fike results. In Sec. 3, we gather several facts on stochastic block models. Then we study the spectrum of H as in (1.3) and a suitably chosen perturbation of H (defined later in (3.2)). In Sec. 4 we prove the main theorem.

2. Perturbation of the Non-Backtracking Spectrum

2.1. The non-backtracking spectrum

When the graph is regular with degree d , the diagonal matrix satisfies $D = dI$, and we can relate the eigenvalues of B with the eigenvalues of A through exact algebraic relations as in the following elementary lemma.

Lemma 2.1. *Let \hat{A} be a Hermitian matrix with eigenvalues $\lambda_1, \dots, \lambda_n$ and let η be a non-zero complex number. Then, the characteristic polynomial of the matrix*

$$\hat{H}_0 = \begin{bmatrix} \hat{A} & \eta \mathbf{I} \\ \mathbf{I} & 0 \end{bmatrix}$$

is given by $\chi_{\hat{H}_0}(z) = \prod_{k=1}^n (z^2 - \lambda_k z - \eta)$, and the eigenvalues of \hat{H}_0 are the $2n$ complex numbers (counted with multiplicities) which are solutions of $z^2 - z\lambda_k - \eta = 0$ for $k \in [n]$.

Similar exact relations have also been used when the graph has a very specific structure, like bipartite biregular (see [13]). When the graph G does not exhibit such a simple structure, the relation between A and B becomes more involved. Several Ihara–Bass-like formulas are available (see, for instance, [4, 8, 32]), but they are usually hard to analyze.

As cleverly noted in [31], the spectrum of H as in (1.3) is hard to describe in terms of the spectrum of A , but the spectrum of \hat{H}_0 in the preceding lemma is completely explicit in terms of the spectrum of \hat{A} , even if \hat{A} has no specific structure. It is therefore quite natural to study the spectrum of \hat{H}_0 using the spectrum of \hat{A} , then use perturbation theorems to infer results on the “true” non-backtracking spectrum, the spectrum of H . This is done in [31] through a combination of the Bauer–Fike theorem and a refinement of the Tao–Vu replacement principle [28].

The celebrated Bauer–Fike theorem says that if a square matrix \hat{A} is diagonalizable, say $\hat{A} = P\Delta P^{-1}$ for a diagonal matrix Δ and a non-singular matrix P , then under a perturbation E , every eigenvalue of the matrix $\hat{A} + E$ is within distance ε of an eigenvalue of \hat{A} , where $\varepsilon = \kappa(P)\|E\|$, and $\kappa(P) = \|P\|\|P^{-1}\|$ is the condition number (see, for instance, [12]).

We observe that the Bauer–Fike theorem, while optimal in the worst case, is indeed extremely wasteful when applied to H and H_0 . Taking into account the specific structure of H and H_0 yields a better perturbation bound at virtually no cost, as shown in the next section.

2.2. Bauer–Fike theorems for quadratic eigenvalue problems

A quadratic eigenvalue problem (QEP) consists of finding the zeroes of the polynomial equation

$$0 = \det(z^2 M - z \hat{A} - X), \quad (2.1)$$

where M, \hat{A}, X are square matrices, and M is non-singular. Such problems appear in a variety of contexts and there exists an extensive literature on them, mainly from a numerical point of view (see the survey [30]).

The triplet (M, \hat{A}, X) can be replaced with by the triplet $(\mathbf{I}, \hat{A}M^{-1}, XM^{-1})$ without changing the problem, so we will be interested in the case where $M = \mathbf{I}$. In

this case, one can easily check that the solutions of (2.1) are the eigenvalues of the $2n \times 2n$ matrix

$$Q_{\hat{A},X} := \begin{bmatrix} \hat{A} & X \\ I & 0 \end{bmatrix},$$

which is called a *linearization* of the problem. In this section, we will present extensions of the Bauer–Fike theorem for linearizations of quadratic eigenvalue problems.

If both matrices \hat{A} and X are diagonal, say $\hat{A} = \text{diag}(\hat{a}_i)$ and $X = \text{diag}(x_i)$, then it is easily seen through elementary linear algebra operations that

$$\det(Q_{\hat{A},X} - zI) = \prod_{i=1}^n (z^2 - \hat{a}_i z - x_i) \quad (2.2)$$

and the eigenvalues of $Q_{\hat{A},X}$ are the $2n$ complex solutions of the collection of n quadratic equations $z^2 - \hat{a}_i z - x_i = 0$ for $1 \leq i \leq n$.

We say the matrices \hat{A} and X are *co-diagonalizable* if there is a common non-singular matrix P such that $P\hat{A}P^{-1}$ and PXP^{-1} are diagonal. If \hat{A}, X are co-diagonalizable, then the identity (2.2) still holds with \hat{a}_i being the eigenvalues of \hat{A} and x_i being those of X . As a consequence, we say that $Q_{\hat{A},X}$ is *QEP-diagonalizable* if \hat{A} and X are co-diagonalizable. This is equivalent to ask that the matrix $z^2I - z\hat{A} - X$ is diagonalizable for any $z \in \mathbb{C}$.

Our main tool for the perturbation analysis is the following theorem.

Theorem 2.1. *Let \hat{A}, \hat{B}, X, Y be $n \times n$ matrices. We define*

$$L_0 = \begin{bmatrix} \hat{A} & X \\ I & 0 \end{bmatrix}, \quad L = \begin{bmatrix} \hat{B} & Y \\ I & 0 \end{bmatrix}.$$

Suppose L_0 is QEP-diagonalizable, with \hat{A} and X being diagonalized by the common matrix P . Then, for any eigenvalue μ of L , there is an eigenvalue ν of L_0 such that

$$|\mu - \nu| \leq \sqrt{\kappa(P)} \sqrt{\|X - Y + \mu(\hat{A} - \hat{B})\|}. \quad (2.3)$$

Moreover, “multiplicities are preserved” in the following sense: Denote $\varepsilon(\mu)$ the RHS of (2.3) and $\varepsilon = \max_{\mu \in \text{Spec}(L)} \varepsilon(\mu)$. If ν_1, \dots, ν_n are the eigenvalues of L_0 and \mathcal{K} is a subset of $[n]$ such that

$$\left(\bigcup_{j \in \mathcal{K}} \mathcal{B}(\nu_k, \varepsilon) \right) \cap \left(\bigcup_{j \notin \mathcal{K}} \mathcal{B}(\nu_k, \varepsilon) \right) = \emptyset,$$

where $\mathcal{B}(\nu_k, \varepsilon) = \{z \in \mathbb{C} : |z - \nu_k| \leq \varepsilon\}$ for $1 \leq k \leq n$, then the number of eigenvalues of L in $\bigcup_{j \in \mathcal{K}} \mathcal{B}(\nu_k, \varepsilon)$ is exactly equal to $|\mathcal{K}|$.

Remark 2.1. Theorem 2.1 is stated for two general matrices L_0 and L . However, the inequality (2.3) will yield good perturbation bound only when we can

control the difference between \hat{A}, \hat{B} , the difference between X, Y , and the condition number $\kappa(P)$.

Proof. Assume μ is an eigenvalue of L . The matrix $R_\mu := \mu^2 I - \mu \hat{B} - Y$ is then singular. Assume, in addition, that μ is not an eigenvalue of L_0 . Then, the matrix $S_\mu := \mu^2 I - \mu \hat{A} - X$ is non-singular. We have

$$\begin{aligned} R_\mu &= \mu^2 I - \mu \hat{B} - Y = \mu^2 I - \mu \hat{A} - X + X + \mu \hat{A} - \mu \hat{B} - Y \\ &= S_\mu + X + \mu \hat{A} - \mu \hat{B} - Y \\ &= S_\mu (I + S_\mu^{-1} (X + \mu \hat{A} - \mu \hat{B} - Y)). \end{aligned}$$

As a consequence the matrix $I + S_\mu^{-1} (X + \mu \hat{A} - \mu \hat{B} - Y)$ is singular, which directly implies that -1 is an eigenvalue of $S_\mu^{-1} (X + \mu \hat{A} - \mu \hat{B} - Y)$. Therefore by the definition of spectral norm,

$$1 \leq \|S_\mu^{-1}\| \cdot \|X + \mu \hat{A} - \mu \hat{B} - Y\| = \|S_\mu^{-1}\| \cdot \|X - Y + \mu(\hat{A} - \hat{B})\|.$$

As noted before the statement of the theorem, if L_0 is QEP-diagonalizable then the matrix S_μ is indeed diagonalizable: if $\Sigma = \text{diag}(\lambda_i)$ is the diagonal matrix of eigenvalues of \hat{A} , $\Delta = \text{diag}(\delta_i)$ the diagonal matrix of eigenvalues of X , and P their common diagonalization matrix, then $S_\mu = P^{-1}(\mu^2 I - \mu \Sigma - \Delta)P$, and the eigenvalues of S_μ are the complex numbers $\mu^2 - \mu \lambda_k - \delta_k$, so

$$\|S_\mu^{-1}\| \leq \kappa(P) \times \max_{k \in [n]} |\mu^2 - \mu \lambda_k - \delta_k|^{-1}.$$

From this, we infer that there is a $k \in [n]$ such that $\|S_\mu^{-1}\| \leq \kappa(P) \times |\mu^2 - \mu \lambda_k - \delta_k|^{-1}$. Let us denote α_k and β_k the two complex solutions of $0 = z^2 - z \lambda_k - \delta_k$ (they are eigenvalues of L_0), then

$$1 \leq \kappa(P) \times \frac{\|X - Y + \mu(\hat{A} - \hat{B})\|}{|\mu^2 - \mu \lambda_k - \delta_k|} = \kappa(P) \times \frac{\|X - Y + \mu(\hat{A} - \hat{B})\|}{|\mu - \alpha_k| |\mu - \beta_k|}$$

which implies that

$$|\mu - \alpha_k| |\mu - \beta_k| \leq \kappa(P) \|X - Y + \mu(\hat{A} - \hat{B})\| =: x.$$

If $|\mu - \alpha_k|$ and $|\mu - \beta_k|$ were both strictly greater than \sqrt{x} , the preceding inequality would be violated. One of those distances is thus smaller than \sqrt{x} , thus proving (2.3).

The “multiplicities preserved” part is then proven as usual with the complex argument principle, see for instance [14, Appendix A]. \square

When applying the preceding result with $\hat{A} = \hat{B}$, one gets the following corollary.

Corollary 2.1. *Let*

$$L_0 = \begin{bmatrix} \hat{A} & X \\ I & 0 \end{bmatrix} \quad L = \begin{bmatrix} \hat{A} & Y \\ I & 0 \end{bmatrix}, \quad (2.4)$$

where \hat{A}, X, Y are square matrices and are such that L_0 is QEP-diagonalizable with \hat{A} and X diagonalized by the common matrix P . Then, for any eigenvalue μ of L , there is an eigenvalue ν of L_0 such that

$$|\mu - \nu| \leq \varepsilon := \sqrt{\kappa(P)\|X - Y\|}. \quad (2.5)$$

Remark 2.2 (Comparison with Classical Bauer–Fike). Casting the classical Bauer–Fike theorem in this setting would yield an error term of $\varepsilon' = \kappa(Q)\|X - Y\|$, where Q is the diagonalization matrix of L_0 . We thus gain the whole square root, and we do not need to compute the condition number of Q . This improvement is remarkable when the matrix \hat{A} is itself Hermitian, for in this case P is unitary and $\kappa(P) = 1$, thus reducing the error term to $\sqrt{\|X - Y\|}$. If we had invoked the classical Bauer–Fike theorem instead, the error term would be $\kappa(Q)\|X - Y\|$, which can be far bigger than $\|X - Y\|$. In fact, the matrix L_0 is not Hermitian in general, and its diagonalization matrix Q might be either difficult to compute or ill-conditioned: in [31], the bound obtained by the authors is $\kappa(Q) \leq O(\sqrt{1/p})$, where p is the connection probability for an Erdős–Rényi graph $G(n, p)$. Our version of the Bauer–Fike theorem shows that for QEP, the only parameters at stake in perturbations are those of the original matrices \hat{A} and X , not those of the linearization of the QEP.

3. The Stochastic Block Model in the Logarithmic Regime

In this section, we collect results from the literature on stochastic block models or inhomogeneous Erdős–Rényi graphs, based on which we prove several quick results for our models, as given in Propositions 3.1, 3.2 and Corollary 3.1.

3.1. Outliers of the adjacency matrix

The concentration of the spectral norm for the SBMs follows immediately from the spectral norm bounds given in [8, 22] for inhomogeneous random matrices and random graphs. Recall Assumption (A). The following statement can be found for example in [22, Example 4.1]: *assume $\alpha = \omega(\log n)$, then*

$$\mathbb{E}[\|A - \mathbb{E}A\|] \leq (2 + o(1))\sqrt{\alpha}.$$

Also from [8, Eq. (2.4)], there exists a constant $c > 0$ such that

$$\mathbb{P}\left(\left|\frac{\|A - \mathbb{E}A\|}{\sqrt{\alpha}} - \frac{\mathbb{E}[\|A - \mathbb{E}A\|]}{\sqrt{\alpha}}\right| \geq t\right) \leq 2e^{-ca^2t^2}.$$

Taking $t = \sqrt{\log n}/\alpha$ in the inequality above, we have with probability $1 - 2n^{-c}$ that

$$\|A - \mathbb{E}A\| \leq \sqrt{\frac{\log n}{\alpha}} + \mathbb{E}[\|A - \mathbb{E}A\|] \leq (2 + o(1))\sqrt{\alpha}. \quad (3.1)$$

Since all the eigenvalues of $\mathbb{E}[A]$ are $\{-p, \beta, \alpha\}$, and $p = o(1)$ under assumption (A), the Weyl eigenvalue inequalities for Hermitian matrices yields the following proposition.

Proposition 3.1. *Assume $\alpha = \omega(\log n)$, then with high probability the following holds:*

$$|\lambda_1(A) - \alpha| \leq (2 + o(1))\sqrt{\alpha}, \quad |\lambda_2(A) - \beta| \leq (2 + o(1))\sqrt{\alpha},$$

$$\max_{k \geq 3} |\lambda_k(A)| \leq (2 + o(1))\sqrt{\alpha}.$$

3.2. Spectrum of the partially derandomized matrix H_0

We will use the notation

$$\gamma = \frac{n(p+q)}{2} - p - 1 = \alpha - 1,$$

which is the “mean degree minus one”. We introduce the *partial derandomization* of H (defined in (1.3)) as

$$H_0 = \begin{bmatrix} A & -\gamma \mathbf{I} \\ \mathbf{I} & 0 \end{bmatrix}. \quad (3.2)$$

As already mentioned in Lemma 2.1, by elementary operations on H_0 , one finds that the characteristic polynomial of H_0 is indeed equal to

$$\chi_0(z) = \prod_{i=1}^n (z^2 - \lambda_i z + \gamma). \quad (3.3)$$

The eigenvalues of H_0 hence come into conjugate pairs coming from eigenvalues of A . Those eigenvalues λ_k of A for which $|\lambda_k| < 2\sqrt{\gamma}$ give rise to two complex conjugate eigenvalues

$$\frac{\lambda_k - i\sqrt{4\gamma - \lambda_k^2}}{2} \quad \text{and} \quad \frac{\lambda_k + i\sqrt{4\gamma - \lambda_k^2}}{2} \quad (3.4)$$

and the other ones, the outliers $|\lambda_k| > 2\sqrt{\gamma}$ of the spectrum of A , give rise to two “harmonic conjugate” eigenvalues

$$\frac{\lambda_k + \sqrt{\lambda_k^2 - 4\gamma}}{2} \quad \text{and} \quad \frac{\gamma}{\lambda_k + \sqrt{\lambda_k^2 - 4\gamma}}. \quad (3.5)$$

Next we obtain a description of the eigenvalues of H_0 from the discussion on the spectrum of A in Proposition 3.1. The description is illustrated in the second panel

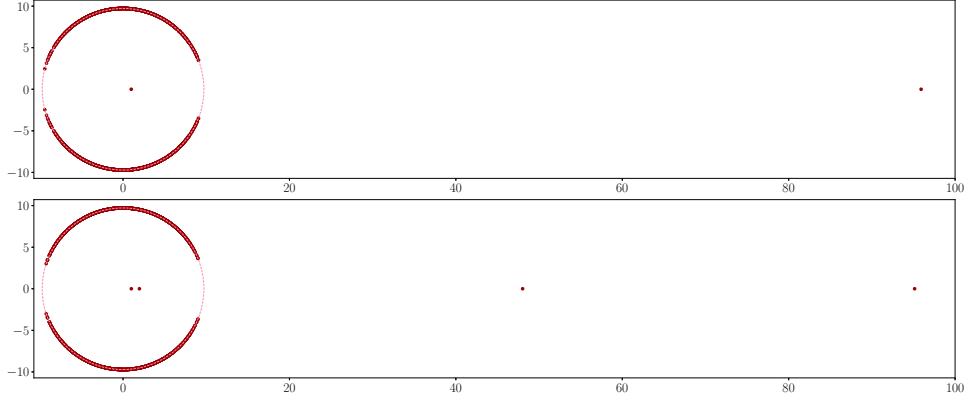


Fig. 2. (Color online) The spectrum of H_0 for the two graphs whose non-backtracking spectrum is depicted in Fig. 1. The circle of radius $\sqrt{\alpha-1}$ is in pink.

of Fig. 2, the first one depicting the same phenomenon but for Erdős–Rényi graphs (with only one outlier in the spectrum).

Proposition 3.2. *Under assumption (A) with high probability the following holds for H_0 .*

(1) *The two eigenvalues with greater modulus, $\lambda_1(H_0)$ and $\lambda_2(H_0)$, are real, and they satisfy*

$$\lambda_1(H_0) = \alpha + O(\alpha^{3/4}) \quad \text{and} \quad \lambda_2(H_0) = \beta + O(\alpha^{3/4}).$$

(2) *The two eigenvalues with smaller modulus, $\lambda_{2n}(H_0)$ and $\lambda_{2n-1}(H_0)$, are real, and they satisfy*

$$\lambda_{2n-1}(H_0) = \frac{\alpha}{\beta} + O(\alpha^{-1/4}), \quad \lambda_{2n}(H_0) = 1 + O(\alpha^{-1/4}).$$

(3) *All the other $2n-4$ eigenvalues have modulus smaller than $\sqrt{\alpha} + o(\sqrt{\alpha})$. Among them, complex eigenvalues lie on a circle of radius $\sqrt{\alpha-1}$ and real ones lie in the intervals $[\sqrt{\alpha} - o(\sqrt{\alpha}), \sqrt{\alpha} + o(\sqrt{\alpha})]$ and $[-\sqrt{\alpha} - o(\sqrt{\alpha}), -\sqrt{\alpha} + o(\sqrt{\alpha})]$.*

Proof. We use Proposition 3.1 and the link described before between the spectrum of A and the spectrum of H_0 given in (3.4) and (3.5).

The greatest eigenvalue of A is $\lambda_1 = \lambda_1(A) = \alpha + O(\sqrt{\alpha})$, from (3.4), it gives rise to two real eigenvalues of H_0 :

$$\mu_1 = \frac{\lambda_1 + \sqrt{\lambda_1^2 - 4\gamma}}{2} = \alpha + O(\alpha^{3/4}), \quad \mu_2 = \frac{\alpha - 1}{\mu_1} = 1 + O(\alpha^{-1/4}).$$

The second greatest eigenvalue (in absolute value) of A is $\lambda_2 = \lambda_2(A) = \beta + O(\sqrt{\alpha})$, which gives rise to two real eigenvalues of H_0 :

$$\mu_3 = \frac{\lambda_2 + \sqrt{\lambda_2^2 - 4\gamma}}{2} = \beta + O(\alpha^{3/4}), \quad \mu_4 = \frac{\alpha - 1}{\mu_3} = \frac{\alpha}{\beta} + O(\alpha^{-1/4}).$$

For the eigenvalues λ_k of A with $2\sqrt{\gamma} < |\lambda_k| \leq (2 + o(1))\sqrt{\alpha}$, from (3.5), the same argument gives

$$|\mu_{2k}| = \left| \frac{\lambda_k + \sqrt{\lambda_k^2 - 4\gamma}}{2} \right| = (1 + o(1))\sqrt{\alpha}, \quad |\mu_{2k+1}| = \left| \frac{\alpha - 1}{\mu_{2k}} \right| = (1 + o(1))\sqrt{\alpha}. \quad (3.6)$$

Finally, from (3.4), all eigenvalues of A with $|\lambda_k| \leq 2\sqrt{\gamma}$ give rise to two complex conjugate eigenvalues of H_0 with magnitude $\sqrt{\gamma} = \sqrt{\alpha - 1}$. This completes the proof. \square

The preceding description in Proposition 3.2 also shows that H_0 is non-singular, and we can quickly describe the eigenvalues of H_0^{-1} . We will later show that the norm of $(H^{-1} - H_0^{-1})$ is very small in Sec. 4. The strategy is then to apply Theorem 2.1 to H_0^{-1} and H^{-1} , which gives a more precise estimate on the location of the outliers in H . See Remark 4.2 for further discussion.

Corollary 3.1 (Inverse Spectrum). *Under the assumption (A), with high probability, in the spectrum of the matrix H_0^{-1} there are exactly two real outliers*

$$\zeta_1 = \frac{1}{1 + O(\alpha^{-1/2})} = 1 + o(1), \quad (3.7)$$

$$\zeta_2 = \frac{1}{\frac{\alpha}{\beta} + O(\alpha^{-1/2})} = \frac{\beta}{\alpha} + o(1), \quad (3.8)$$

and all the other eigenvalues of H_0^{-1} have modulus smaller than $(\sqrt{\alpha}(1 + o(1))^{-1} = o(1))$.

Proof. The location of the two real outliers in the spectrum of H_0^{-1} comes from part (2) in Proposition 3.2. The location of all the other eigenvalues comes from part (1) and (3) in Proposition 3.2. \square

We now turn to the description of the global behavior of the spectrum of A .

3.3. Limiting spectral distribution of A

If we have an SBM with two blocks of equal size, and $p, q = \omega(1/n)$, the empirical spectral distribution of $\frac{A}{\sqrt{\alpha}}$ will converge weakly to the semicircle law: for any bounded continuous test function f , almost surely

$$\frac{1}{n} \sum_{i=1}^n f(\lambda_i(A/\sqrt{\alpha})) \rightarrow \frac{1}{2\pi} \int_{-2}^2 f(t) \sqrt{4 - t^2} dt. \quad (3.9)$$

This can be seen from the graphon representation of SBMs and the result for generalized Wigner matrices (33, Sec. 4]), since each row in $\mathbb{E}A$ has the same row sum or

equivalently, each vertex has the same expected degree. If the degree is not homogeneous, then the limiting spectral distribution will not be the semicircle law. We recall the following result from [33] for generalized Wigner matrices, which includes the regime where the sparsity parameter is $\omega(1/n)$.

Theorem 3.1 ([33, Theorem 4.2]). *Let A_n be a random Hermitian matrix such that entries on and above the diagonal are independent and satisfy the following conditions:*

- (1) $\mathbb{E}[a_{ij}] = 0$, $\mathbb{E}|a_{ij}|^2 = s_{ij}$.
- (2) $\frac{1}{n} \sum_{j=1}^n s_{ij} = 1 + o(1)$ for all $i \in [n]$.
- (3) For any constant $\eta > 0$, $\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{1 \leq i, j \leq n} \mathbb{E}[|a_{ij}|^2 \mathbf{1}(|a_{ij}| \geq \eta \sqrt{n})] = 0$.
- (4) $\sup_{ij} s_{ij} \leq C$ for a constant $C > 0$.

Then the empirical spectral distribution of $\frac{A_n}{\sqrt{n}}$ converges weakly to the semicircle law almost surely, which means that on an event with probability 1, the convergence (3.9) holds for any bounded continuous function f .

We obtain the following theorem for the adjacency matrix A of an SBM, and also for H_0 .

Theorem 3.2. *Assume $\alpha \rightarrow \infty$ and $p, q \rightarrow 0$. The empirical spectral distribution of $\frac{A}{\sqrt{\alpha}}$ converges weakly to the semicircle law supported on $[-2, 2]$ almost surely.*

Moreover, the empirical spectral distribution of $\frac{H_0}{\sqrt{\alpha}}$ converges weakly almost surely to a distribution on the circle of radius 1, and the limiting distribution of the real part of the eigenvalues of H_0 is the semicircle law rescaled on $[-1, 1]$.

Proof. We first consider the centered and scaled matrix

$$M := (m_{ij})_{1 \leq i, j \leq n} = \frac{A - \mathbb{E}A}{\sqrt{(p+q)/2}}.$$

For $i \neq j$ we have

$$\mathbb{E}[|m_{ij}|^2] = \begin{cases} \frac{q(1-q)}{(p+q)/2} & \text{if } i, j \text{ are in the same block,} \\ \frac{p(1-p)}{(p+q)/2} & \text{otherwise.} \end{cases}$$

Then for all $i \in [n]$,

$$\frac{1}{n} \sum_j \mathbb{E}[|m_{ij}|^2] = \frac{1}{n} \left(\left(\frac{n}{2} - 1 \right) \frac{p(1-p)}{(p+q)/2} + \frac{n}{2} \frac{q(1-q)}{(p+q)/2} \right) = 1 + o(1).$$

One can quickly check that all the conditions in Theorem 3.4 hold for M . Therefore the empirical spectral distribution of

$$\frac{M}{\sqrt{n}} = \frac{A - (\mathbb{E}A + p\mathbf{I})}{\sqrt{n(p+q)/2}} + \frac{p\mathbf{I}}{\sqrt{n(p+q)/2}}$$

converges weakly to the semicircle law. Or equivalently the empirical spectral distribution of $\frac{A - (\mathbb{E}A + pI)}{\sqrt{\alpha}}$ converges weakly almost surely to the same distribution. Finally, since the rank of the matrix $(\mathbb{E}A + pI)$ is 2 and by the Cauchy interlacing theorem for the eigenvalue of Hermitian matrices, the empirical spectral distribution of $\frac{A}{\sqrt{\alpha}}$ converges weakly to the semicircle law almost surely.

We now turn to the second part of the theorem. Note that from (B.3), one eigenvalue $\lambda_i(A)$ corresponds to two eigenvalues of H_0 momentarily denoted by $\mu_{2i-1}(H_0), \mu_{2i}(H_0)$, and such that

$$\Re \mu_{2i-1}(H_0) = \Re \mu_{2i}(H_0) = \frac{\lambda_i(A)}{2}.$$

The empirical spectral distribution of the real parts of eigenvalues of $H_0/\sqrt{\alpha}$ satisfies

$$\frac{1}{2n} \sum_{j=1}^{2n} \delta_{\Re(\mu_j/\sqrt{\alpha})} = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i/2\sqrt{\alpha}},$$

which converges weakly almost surely to the semicircle law rescaled on $[-1, 1]$ by the first part of the theorem. \square

3.4. Concentration of the degrees

We finally describe the degrees in the SBM. Let us note d_i the degree of vertex i in the SBM graph. Under the assumption (A), we have $\alpha = \omega(\log n)$, and the degrees are highly concentrated in the following sense.

Lemma 3.1. *With high probability $\max_{i \in [n]} |d_i - \alpha| = o(\alpha)$.*

Remark 3.1. Note that Lemma B.1 is no longer true in other regimes. When $\alpha = O(\log n)$, the event $d_i = (1 + o(1))\alpha$ for all $i \in [n]$ does not happen with high probability (see, for example, [10, Chap. 3]). Then the diagonal degree matrix D is not close to αI , which is a barrier for our perturbation analysis to work.

Lemma B.1 can be found in the literature, but we provide a proof for completeness. We recall Bernstein's inequality: *let $Y_n = \sum_{i=1}^n X_i$ where X_i are independent random variables such that $|X_i| \leq b$. Define $\sigma_n^2 := \text{Var}(Y_n)$. Then for any $x > 0$,*

$$\mathbb{P}(|Y_n - \mathbb{E}[Y_n]| \geq x) \leq 2 \exp \left(\frac{-x^2}{2(\sigma_n^2 + bx/3)} \right).$$

Now we prove Lemma B.1

Proof. Each d_i has the same distribution with mean α , hence we can apply the union bound and get

$$\mathbb{P} \left(\max_{i \in [n]} |d_i - \alpha| \geq x \right) \leq \sum_{i \in [n]} \mathbb{P}(|d_i - \alpha| \geq x) = n \mathbb{P}(|d_1 - \alpha| \geq x). \quad (3.10)$$

Let us write $d_1 = X_2 + \dots + X_{n/2} + X_{n/2+1} + \dots + X_n$, where the X_i are independent, and X_i is a Bernoulli random variable with parameter p if $i \in \{2, \dots, n/2\}$ and q if

$i \in \{n/2 + 1, \dots, n\}$. Those variables are all bounded by 1 so we can take $b = 1$ in Bernstein's inequality. The variance is

$$\sigma_n^2 = \left(\frac{n}{2} - 1\right)p(1-p) + \frac{n}{2}q(1-q) \leq \alpha.$$

From (3.10) and Bernstein's inequality we have

$$\mathbb{P}(|d_1 - \alpha| \geq x) \leq 2 \exp\left(\frac{-x^2}{2(\alpha + x/3)}\right).$$

Let $h(n)$ be any sequence of positive numbers. The choice $x = \frac{\alpha}{h(n)}$ then leads to

$$\mathbb{P}\left(\max_{i \in [n]} |d_i - \alpha| \geq x\right) \leq 2 \exp\left(\log n - \frac{\alpha}{2h^2(n) + 2h(n)/3}\right).$$

Since we know $\alpha = \omega(\log n)$, any choice of $h(n)$ growing to ∞ slowly enough will be sufficient; for instance if $\alpha = \log(n)f(n)$ with $f(n) \rightarrow \infty$, we take $h(n) = f(n)^{1/3}$ and we obtain that $\max_{i \in [n]} |d_i - \alpha| = o(\alpha)$ with probability $1 - o(1)$. \square

4. Proof of Theorem 1.1

4.1. Existence of bulk insiders

In this section we prove the existence of the isolated eigenvalues inside the bulk. To do this, we compare the spectrum of H (defined in (1.3)) and H_0 (defined in (3.2)). We also need to compare the spectrum of H^{-1} and H_0^{-1} to have a more refined estimate compared to [31]. See Remark 4.2 for further discussion.

Fix any non-singular square matrix X . One can easily check that

$$\begin{bmatrix} A & -X \\ I & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & I \\ -X^{-1} & X^{-1}A \end{bmatrix} \quad (4.1)$$

and by conjugation the spectrum of this matrix is the same as the spectrum of the matrix

$$\begin{bmatrix} X^{-1}A & -X^{-1} \\ I & 0 \end{bmatrix}.$$

Let us introduce the matrices

$$K = \begin{bmatrix} (D - I)^{-1}A & -(D - I)^{-1} \\ I & 0 \end{bmatrix} \quad \text{and} \quad K_0 = \begin{bmatrix} (\alpha - 1)^{-1}A & -(\alpha - 1)^{-1}I \\ I & 0 \end{bmatrix}. \quad (4.2)$$

These matrices have the same spectrum as (respectively) H^{-1} and H_0^{-1} and we are going to apply Theorem 2.1 to them. First, one has to note that the spectrum of K is indeed bounded away from zero. More precisely, all the eigenvalues of H are bounded below by 1, as explained in the following statement ([5, Theorem 3.7]), the same result for finite graphs was first given in [20]): *let $d_{\min} \geq 2$ and d_{\max} be the*

minimal and maximal degrees of some finite or infinite graph G . Then the spectrum of B is included in

$$\{\lambda \in \mathbb{C} \setminus \mathbb{R} : \sqrt{d_{\min} - 1} \leq |\lambda| \leq \sqrt{d_{\max} - 1}\} \cup \{\lambda \in \mathbb{R} : 1 \leq |\lambda| \leq d_{\max} - 1\}. \quad (4.3)$$

We see from Lemma 3.1 that with high probability all the degrees in our graph are greater than 2, hence every eigenvalue of H has modulus greater than 1, thus ensuring that every eigenvalue μ of H^{-1} has $|\mu| \leq 1$. We now apply Theorem 2.1 to K and K_0 . It is easily seen from (4.2) that K_0 is QEP-diagonalizable and the change-of-basis matrix P is unitary since A is Hermitian. We take

$$X = (\alpha - 1)^{-1}I, \quad \text{and} \quad Y = (D - I)^{-1}. \quad (4.4)$$

From Theorem 2.1 we have

$$\begin{aligned} \varepsilon := \max_{\mu \in \text{Spec}(H^{-1})} \varepsilon(\mu) &= \max_{\mu \in \text{Spec}(H^{-1})} \sqrt{\kappa(P)} \sqrt{\|X - Y + \mu(Y - X)A\|} \\ &\leq \max_{\mu \in \text{Spec}(H^{-1})} \sqrt{\|X - Y\|(1 + |\mu|\|A\|)} \\ &\leq \max_{\mu \in \text{Spec}(H^{-1})} \sqrt{\|X - Y\|(1 + \|A\|)} \\ &\leq \sqrt{\|X - Y\|(1 + \alpha(1 + o(1))), \end{aligned}$$

where the last line holds with high probability from the description of the spectrum of A in Proposition 3.1. It turns out that $\varepsilon = o(1)$, as a consequence of the following lemma.

Lemma 4.1. *For X and Y defined in (4.4), with high probability $\|X - Y\| = o(\alpha^{-1})$.*

Proof. Since X, Y are diagonal matrices, we have

$$\|X - Y\| = \max_{i \in [n]} \left| \frac{1}{d_i - 1} - \frac{1}{\alpha - 1} \right| = \max_{i \in [n]} \frac{|d_i - \alpha|}{(d_i - 1)(\alpha - 1)}.$$

By Lemma 3.1, with high probability $\max_{i \in [n]} |d_i - \alpha| = o(\alpha)$ and this implies the lemma. \square

We thus have $\varepsilon = o(1)$ and now we can combine the ‘‘multiplicities preserved’’ part of Theorem 2.1 and the description of the spectrum of K_0 in Corollary 3.1.

Remark 4.1. Recall ζ_1, ζ_2 from Corollary 3.1. The crucial fact here is that $\zeta_1 \approx 1$ and $\zeta_2 \approx \beta/\alpha$ are of order 1 and in particular they are bounded away from 0, which is guaranteed by the third inequality in our assumption (A).

Theorem 2.1 implies that there is exactly one eigenvalue of K in $\mathcal{B}(\zeta_1, \varepsilon)$, one in $\mathcal{B}(\zeta_2, \varepsilon)$ and all the other eigenvalues have modulus $o(1)$. In other words, there are

exactly two eigenvalues ξ_1, ξ_2 of H such that

$$\xi_1^{-1} = 1 + o(1), \quad \xi_2^{-1} = \frac{\beta}{\alpha} + o(1),$$

and all the other ones have inverse modulus $o(1)$. By the continuity of $x \mapsto x^{-1}$, we have exactly two eigenvalues of H ,

$$\xi_1 = 1 + o(1), \quad \xi_2 = \frac{\alpha}{\beta} + o(1), \quad (4.5)$$

which are of order 1, and all the other eigenvalues of H have inverse modulus $\omega(1)$.

Since 0 is always an eigenvalue of the Laplacian $D - A$, we have

$$1 \in \{z \in \mathbb{C} : \det(z^2 \mathbf{I} - zA + D - \mathbf{I}) = 0\},$$

which implies 1 is always an eigenvalue of H . So ξ_1 is indeed exactly equal to 1, otherwise we have three eigenvalues of H : 1, ξ_1 and ξ_2 that are of order 1, a contradiction to (4.5).

Moreover, ξ_2 must be a real eigenvalue of H , otherwise from the fact that the spectrum of B is symmetric with respect to the real line, we would see two eigenvalues of K in the ball $\mathcal{B}(\zeta_2, \varepsilon)$, which is a contradiction to Theorem 2.1. This completes the proof of (4.5).

4.2. Existence of the outliers

We now simply apply Theorem 2.1 to the matrices H defined in (1.3) and H_0 defined in (3.2). Here, $A = B$ and in fact we are in the setting of Corollary 2.1 with $X = (\alpha - 1)\mathbf{I}$ and $Y = D - \mathbf{I}$. Hence

$$\varepsilon = \sqrt{\|X - Y\|} = \sqrt{\max_{i \in [n]} |d_i - \alpha|} = o(\sqrt{\alpha})$$

with high probability from Lemma 3.1. From the description of the spectrum of H_0 in Proposition 3.2, we see that there are two outliers located near β and α and all other eigenvalues have order $O(\sqrt{\alpha})$. From this and the ‘‘multiplicities preserved’’ part in Corollary 2.1, we see that H has two outliers located within distance $o(\sqrt{\alpha})$ of

$$\lambda_1(H_0) = \alpha + O(\alpha^{3/4}), \quad \text{and} \quad \lambda_2(H_0) = \beta + O(\alpha^{3/4}).$$

By the symmetry of the spectrum with respect to the real line, those two outliers are real numbers. This completes the proof of (4.4).

Remark 4.2. Note that we could also use this strategy to infer the existence of the bulk insiders: in fact, the result would yield the existence of two eigenvalues located in the balls $\mathcal{B}(1, \varepsilon)$ and $\mathcal{B}(\alpha/\beta, \varepsilon)$. These eigenvalues would be detached from the bulk of eigenvalues of H , which lie within distance $o(\sqrt{\alpha})$ of the circle of radius $\sqrt{\alpha}$; however, no further information can be inferred, since $o(\sqrt{\alpha})$ can go to infinity as well. This is the reason why we had to compare H^{-1} with H_0^{-1} , which

has two effects: first, it isolates the two “insiders” of H_0 and the other eigenvalues close to zero; and secondly, it turns out that the norm of $H^{-1} - H_0^{-1}$ is very small. In addition, our use of the specific Bauer–Fike theorem designed for QEP (Theorem 2.1) yields more precise results than [31].

4.3. Global spectral distribution

We now prove the “bulk” part of Theorem 1.1. The strategy is the same as [31] and we borrow their main theorem.

Theorem 4.1 ([31, Corollary 3.3]). *Let M_m and P_m be $m \times m$ matrices with entries in complex numbers, and let $f(z, m) \geq 1$ be a real function depending on z, m . Let μ_M be the empirical spectral distribution of any square matrix M . Assume that*

$$\frac{1}{m} \|M_m\|_F^2 + \frac{1}{m} \|M_m + P_m\|_F^2 \quad (4.6)$$

is bounded in probability, and

$$f(z, m) \|P_m\| \rightarrow 0 \quad (4.7)$$

in probability, and for almost every complex number $z \in \mathbb{C}$,

$$\|(M_m - zI)^{-1}\| \leq f(z, m) \quad (4.8)$$

with probability tending to 1, then $\mu_{M_m} - \mu_{M_m + P_m}$ converges in probability to zero.

Recall H from (1.3) and H_0 from (3.2). Take

$$M_m = \frac{H_0}{\sqrt{\alpha}}, \quad P_m = \frac{1}{\sqrt{\alpha}}(H - H_0) = \frac{1}{\sqrt{\alpha}} \begin{bmatrix} 0 & \alpha I - D \\ 0 & 0 \end{bmatrix}$$

in Theorem 4.1. If all the conditions in Theorem 4.1 hold, then the “bulk” part of Theorem 1.1 follows from our Theorem 3.2.

The condition (4.6) follows verbatim from the proof of [31, Lemma 3.7]. Condition (4.7) follows from [31, Lemma 3.9] and our Lemma 3.1. Condition (4.8) follows from [31, Lemma 3.9]. This completes the proof of global spectral distribution part of Theorem 1.1.

Acknowledgments

We would like to thank Lorenzo Dall’Amico for sharing the conjecture in [15] and helpful discussions. We also thank Ke Wang for a careful reading of this paper and her useful suggestions. We are grateful to the organizers of the conference Random Matrices and Random Graphs at CIRM, during which this work was initiated. S.C. is supported by ERC NEMO, under the European Union’s Horizon 2020 research and innovation program grant agreement number 788851. Y.Z. is partially supported by NSF DMS-1949617.

References

- [1] E. Abbe, E. Boix-Adserà, P. Ralli and C. Sandon, Graph powering and spectral robustness, *SIAM J. Math. Data Sci.* **2**(1) (2020) 132–157.
- [2] E. Abbe, J. Fan, K. Wang and Y. Zhong, Entrywise eigenvector analysis of random matrices with low expected rank, preprint (2017), arXiv:1709.09565.
- [3] J. Alt, R. Ducez and A. Knowles, Extremal eigenvalues of critical Erdős–Rényi graphs, preprint (2019), arXiv:1905.03243.
- [4] N. Anantharaman, Some relations between the spectra of simple and non-backtracking random walks, preprint (2017), arXiv:1703.03852.
- [5] O. Angel, J. Friedman and S. Hoory, The non-backtracking spectrum of the universal cover of a graph, *Trans. Amer. Math. Soc.* **367**(6) (2015) 4287–4318.
- [6] J. Banks and L. Trevisan, Vector colorings of random, Ramanujan, and large-girth irregular graphs, preprint (2019), arXiv:1907.02539.
- [7] H. Bass, The Ihara–Selberg Zeta function of a tree lattice, *Int. J. Math.* **3**(6) (1992) 717–797.
- [8] F. Benaych-Georges, C. Bordenave and A. Knowles, Spectral radii of sparse random matrices, preprint (2017), arXiv:1704.02945.
- [9] F. Benaych-Georges, C. Bordenave and A. Knowles, Largest eigenvalues of sparse inhomogeneous Erdős–Rényi graphs, *Ann. Probab.* **47**(3) (2019) 1653–1676.
- [10] B. Bollobás, *Random Graphs*, Cambridge Studies in Advanced Mathematics, Vol. 73 (Cambridge University Press, Cambridge, 2001).
- [11] C. Bordenave, A new proof of Friedman’s second eigenvalue theorem and its extension to random lifts, preprint (2015), arXiv:1502.04482.
- [12] C. Bordenave, M. Lelarge and L. Massoulié, Non-backtracking spectrum of random graphs: Community detection and non-regular Ramanujan graphs, *Ann. Probab.* **46**(1) (2018) 1–71.
- [13] G. Brito, I. Dumitriu and K. D. Harris, Spectral gap in random bipartite biregular graphs and its applications, preprint (2018), arXiv:1804.07808.
- [14] S. Coste, The spectral gap of sparse random digraphs, preprint (2017), arXiv:1708.00530.
- [15] L. Dall’Amico, R. Couillet and N. Tremblay, Revisiting the Bethe–Hessian: Improved community detection in sparse heterogeneous graphs, in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox and R. Garnett (eds.) (Curran Associates, Inc., 2019), pp. 4039–4049.
- [16] I. Dumitriu and Y. Zhu, Spectra of random regular hypergraphs, preprint (2019), arXiv:1905.06487.
- [17] P. Grindrod, D. J. Higham and V. Noferini, The deformed graph Laplacian and its applications to network centrality analysis, *SIAM J. Matrix Anal. Appl.* **39**(1) (2018) 310–341.
- [18] L. Gulikers, M. Lelarge and L. Massoulié, Non-backtracking spectrum of degree-corrected stochastic block models, *8th Innovations in Theoretical Computer Science Conf. (ITCS 2017)*, Vol. 67 (Schloss Dagstuhl–Leibniz–Zentrum fuer Informatik, 2017), pp. 44:1–44:27.
- [19] K.-I. Hashimoto, Zeta functions of finite graphs and representations of p-Adic groups, in *Automorphic Forms and Geometry of Arithmetic Varieties* (Elsevier, Amsterdam, 1989), pp. 211–280.
- [20] M. Kotani and T. Sunada, Zeta functions of finite graphs, *J. Math. Sci. Univ. Tokyo* **7**(1) (2000) 7–25.

- [21] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová and P. Zhang, Spectral redemption in clustering sparse networks, *Proc. Natl. Acad. Sci.* **110**(52) (2013) 20935–20940.
- [22] R. Latała, R. van Handel and P. Youssef, The dimension-free structure of non-homogeneous random matrices, *Invent. Math.* **214**(3) (2018) 1031–1080.
- [23] L. Massoulié, Community detection thresholds and the weak Ramanujan property, in *Proc. Forty-Sixth Annual ACM Symposium on Theory of Computing* (ACM, New York, 2014), pp. 694–703.
- [24] E. Mossel, J. Neeman and A. Sly, Reconstruction and estimation in the planted partition model, *Probab. Theory Relat. Fields* **162**(3–4) (2015) 431–461.
- [25] E. Mossel, J. Neeman and A. Sly, A proof of the block model threshold conjecture, *Combinatorica* **38**(3) (2018) 665–708.
- [26] A. Saade, F. Krzakala and L. Zdeborová, Spectral clustering of graphs with the Bethe Hessian, in *Advances in Neural Information Processing Systems* (2014), pp. 406–414.
- [27] H. M. Stark and A. A. Terras, Zeta functions of finite graphs and coverings, *Adv. Math.* **121**(1) (1996) 124–165.
- [28] T. Tao, V. Vu and M. Krishnapur, Random matrices: Universality of ESDs and the circular law, *Ann. Probab.* **38**(5) (2010) 2023–2065.
- [29] A. Terras, *Zeta Functions of Graphs: A Stroll Through the Garden*, Vol. 128 (Cambridge University Press, Cambridge, 2010).
- [30] F. Tisseur and K. Meerbergen, The quadratic eigenvalue problem, *SIAM Rev.* **43**(2) (2001) 235–286.
- [31] K. Wang and P. M. Wood, Limiting empirical spectral distribution for the non-backtracking matrix of an Erdős–Rényi random graph, preprint (2017), arXiv:1710.11015.
- [32] Y. Watanabe and K. Fukumizu, Graph Zeta function in the Bethe free energy and loopy belief propagation, in *Advances in Neural Information Processing Systems*, Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams and A. Culotta (eds.) (Curran Associates, Inc., 2009), pp. 2017–2025.
- [33] Y. Zhu, A graphon approach to limiting spectral distributions of Wigner-type matrices, *Random Struct. Algor.* **56**(1) (2020) 251–279.