LIMITING EMPIRICAL SPECTRAL DISTRIBUTION FOR THE NON-BACKTRACKING MATRIX OF AN ERDŐS-RÉNYI RANDOM GRAPH

KE WANG AND PHILIP MATCHETT WOOD

Abstract. In this note, we give a precise description of the limiting empirical spectral distribution (ESD) for the non-backtracking matrices for an Erdős-Rényi graph $G(n,p)$ assuming $np/\log n$ tends to infinity. We show that derandomizing part of the non-backtracking random matrix simplifies the spectrum considerably, and then we use Tao and Vu's replacement principle and the Bauer-Fike theorem to show that the partly derandomized spectrum is, in fact, very close to the original spectrum.

1. Introduction

For a simple undirected graph $G = (V, E)$, the non-backtracking matrix is defined as follows. For each $(i, j) \in E$, form two directed edges $i \to j$ and $j \to i$. The non-backtracking matrix B is a $2|E| \times 2|E|$ matrix such that

$$B_{i\to j,k\to l} = \begin{cases} 1 & \text{if } j = k \text{ and } i \neq l \\ 0 & \text{otherwise.} \end{cases}$$

The central question of the current paper is the following:

Question 1.1. What can be said about the eigenvalues of the non-backtracking matrix B of random graphs as $|V| \to \infty$?

The non-backtracking matrix was proposed by Hashimoto [Has89a]. The spectrum of the non-backtracking matrix for random graphs was studied by Angel, Friedman, and Hoory [AFH15] in the case where the underlying graph is the tree covering of a finite graph. Motivated by the question of community detection (see [KMM+13, Mas14, MNS13, MNS15]), Bordenave, Lelarge, and Massoulié [BLM15] determined the size of the largest eigenvalue and gave bounds for the sizes of all other eigenvalues for non-backtracking matrices when the underlying graph is drawn from a generalization of Erdős-Rènyi random graphs called the Stochastic Block Model (see [HLL83]), and this work was further extended to the Degree-Corrected Stochastic Block Model (see [KN11]) by Gulikers, Lelarge, and Massoulié [GLM16]. In the recent work of Benaych-Georges, Bordenave and Knowles [BGBK17], they studied the spectral radii of the sparse inhomogeneous Erdős-Rényi graph through a novel application of the non-backtracking matrices.

In the current paper, we give a precise characterization of the limiting distribution of the eigenvalues for the non-backtracking matrix when the underlying graph is the Erdős-Rènyi random graph $G(n,p)$, where each edge ij is present indpendently with probability p, and where we exclude loops (edges of the form ii). We will allow p to be constant or decreasing sublinearly with n, which contrasts to the bounds proved in [BLM15] and [GLM16] corresponding

Date: December 30, 2022.
Ke Wang was supported in part by Hong Kong RGC grants GRF 16301618, GRF 16308219 and ECS 26304920. Philip Matchett Wood was supported in part by NSA grant H98230-16-1-0301.

to the case $p = c/n$ with a constant c. Let A be the adjacency matrix of $G(n, p)$, so $A_{ij} = 1$ exactly when edge ij is part of the graph G and $A_{ij} = 0$ otherwise; and let D the diagonal matrix with $D_{ii} = \sum_{j=1}^{n} A_{ij}$. Much is known about the eigenvalues of A, going back to works of Wigner in the 1950s [Wig55, Wig58] (see also [Gre63] and [Arn67]), who proved that the distribution of eigenvalues follows the semicircular law. More recent results have considered the case where p tends to zero, making the random graph sparse. It is known that assuming $np \to \infty$, the empirical spectral distribution (ESD) of the adjacency matrix A converges to the semicircle distribution (see for example [KP93] or [TVW13]). Actually, much stronger results have been proved about the eigenvalues of A (see the surveys [Vu08] and [BGK16]). For example, Erdős, Knowles, Yau, and Yin [EKYY13] proved that as long as there is a constant C so that $np > (\log n)^{C \log \log n}$ (and thus $np \to \infty$ faster than logarithmic speed), the eigenvalues of the adjacency matrix A satisfy a result called the local semicircle law.

It has been shown in [Has89b, Bas92, AFH15] (for example, Theorem 1.5 from [AFH15]) that the spectrum of B is the set $\{\pm 1\} \cup \{\mu : \det(\mu^2 I - \mu A + D - I) = 0\}$, or equivalently, the set $\{\pm 1\} \cup \{\text{eigenvalues of } H\}$, where

$$H = \begin{pmatrix} A & I - D \\ I & 0 \end{pmatrix}. \tag{1.1}$$

We will call this $2n \times 2n$ matrix H *the non-backtracking spectrum operator for* A, and we will show that the spectrum of H may be precisely described, thus giving a precise description of the eigenvalues of the non-backtracking matrix B. We will study the eigenvalues of H in two regions: the *dense region*, where $p \in (0, 1)$ and p is fixed constant; and the *sparse region*, where $p = o(1)$ and $np \to \infty$. The *diluted region*, where $p = c/n$ for some constant $c > 1$, is the region for which the bounds in [BLM15] and [GLM16] apply, and, as pointed out by [BLM15], it would be interesting to determine the limiting eigenvalue distribution in this region.

Note that $\mathbb{E}(D) = (n - 1)pI$, and so we will denote

$$\alpha = (n - 1)p - 1$$

and consider the partly averaged matrix

$$H_0 = \begin{pmatrix} A & I - \mathbb{E}(D) \\ I & 0 \end{pmatrix} = \begin{pmatrix} A & -\alpha I \\ I & 0 \end{pmatrix}. \tag{1.2}$$

The partly averaged matrix H_0 will be an essential tool in quantifying the eigenvalues of the non-backtracking spectrum operator H. Three main ideas are at the core of this paper: first, that partial derandomization can greatly simplify the spectrum; second, that Tao and Vu's replacement principle [TV10, Theorem 2.1] can be usefully applied to two sequences of random matrices that are highly dependent on each other; and third, that in this case, the partly derandomized matrix may be viewed as a small perturbation of the original matrix, allowing one to apply results from perturbation theory like the Bauer-Fike Theorem. The use of Tao and Vu's replacement principle here is novel, as it is used to compare the spectra of two highly dependent random matrices with the same random inputs; typically, the Tao-Vu replacement principle has been applied in cases where the two sequences of random matrices are independent, see for example [TV10, Woo12, Woo16].

1.1. **Results.** Our first result shows that the spectrum of H_0 can be determined very precisely in terms of the spectrum of the random Hermitian matrix A, which is well-understood.

Proposition 1.2 (Spectrum of the partly averaged matrix)**.** *Let H_0 be defined as in (1.2), and let $0 < p \le p_0 < 1$ for a constant p_0. If $p \ge C/\sqrt{n}$ for some large constant $C > 0$, then, with probability $1 - o(1)$, $\frac{1}{\sqrt{\alpha}} H_0$ has two real eigenvalues μ_1 and μ_2 satisfying $\mu_1 = \sqrt{\alpha}(1 + o(1))$ and $\mu_2 = 1/\sqrt{np}(1 + o(1))$; all other eigenvalues for $\frac{1}{\sqrt{\alpha}} H_0$ are complex with magnitude 1 and occur in complex conjugate pairs. If $np \to \infty$ with n, then the real parts of the eigenvalues in the circular arcs are distributed according to the semicircle law.*

Remark 1.3. When $n^{-1+\epsilon} \le p \le n^{-1/2}$, more real eigenvalues of H_0 will emerge. We provide a short discussion on the real eigenvalues of H_0 in the end of Section 2.1.

The spectrum of the non-backtracking matrix for a degree regular graph was studied in [Bor15], including proving some precise eigenvalue estimates. One can view Proposition 1.2 as extending this general approach by using averaged degree counts, but allowing the graph to no longer be degree regular. Thus, Proposition 1.2 shows that partly averaging H to get H_0 is enough to allow the spectrum to be computed very precisely. Our main results are the theorems below, which show that the bulk distributions of H and H_0 are very close to each other, even for p a decreasing function of n. (The definitions of the measure μ_M for a matrix M and the definition of almost sure convergence of measures are given in Section 1.3).

Theorem 1.4. *Let A be the adjacency matrix for an Erdős-Rènyi random graph $G(n,p)$. Assume $0 < p \le p_0 < 1$ for a constant p_0 and $np/\log n \to \infty$ with n. Let $\frac{1}{\sqrt{\alpha}} H$ be a rescaling of the non-backtracking spectrum operator for A defined in (1.1) with $\alpha = (n-1)p - 1$, and let $\frac{1}{\sqrt{\alpha}} H_0$ be its partial derandomization, defined in (1.2). Then, $\mu_{\frac{1}{\sqrt{\alpha}} H} - \mu_{\frac{1}{\sqrt{\alpha}} H_0}$ converges almost surely (thus, also in probability) to zero as n goes to infinity.*

Remark 1.5. When $p \gg \log n/n$, the graph $G(n,p)$ is almost a random regular graph and thus H_0 appears a good approximation of H. When p becomes smaller, such approximation is no longer accurate. In this sense, Theorem 1.4 is optimal.

In Figure 1, we plot the eigenvalues of $\frac{1}{\sqrt{\alpha}} H$ and $\frac{1}{\sqrt{\alpha}} H_0$ for an Erdős-Rènyi random graph $G(n,p)$, where $n = 500$. The blue dots are the eigenvalues of $H/\sqrt{\alpha}$ and the red crosses are for $H_0/\sqrt{\alpha}$. We can see that the empirical spectral measure of $H/\sqrt{\alpha}$ is very close to those of $H_0/\sqrt{\alpha}$ for p not too small. As p becomes smaller (note that here $\log n/n \approx 0.0054$), the eigenvalues of $H_0/\sqrt{\alpha}$ still lie on the arcs of the unit circle whereas the eigenvalues of $H/\sqrt{\alpha}$ start to escape and be attracted to the inside of the circle.

To prove that the bulk eigenvalue distributions converge in Theorem 1.4, we will use Tao and Vu's replacement principle [TV10, Theorem 2.1] (see also Theorem 3.2), which was a key step in proving the circular law. The replacement principle lets one compare eigenvalue distributions of two sequences of random matrices, and it has often been used in cases where one type of random input—for example, standard Gaussian normal entries—is replaced by a different type of random input—for example, arbitrary mean 0, variance 1 entries. This is how the replacement principle was used to prove the circular law in [TV10], and it was used similarly in, for example, [Woo12, Woo16]. The application of the replacement principle in the current paper is novel in that it compares the eigenvalue distributions of H and H_0, matrices with the same random inputs that are also highly dependent—in fact, H_0 is completely determined by H.

Our third result (Theorem 1.6 below) proves that all eigenvalues of H are close to those of H_0 with high probability when $p \gg \frac{\log^{2/3} n}{n^{1/6}}$, which implies that there are no outlier eigenvalues of

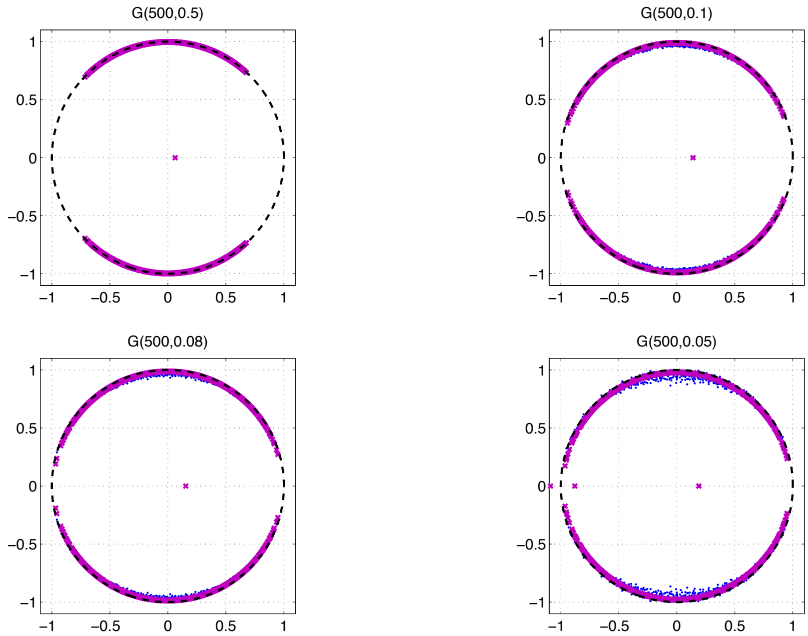

FIGURE 1. The eigenvalues of $H/\sqrt{\alpha}$ defined in (1.1) and $H_0/\sqrt{\alpha}$ defined in (1.2) for a sample of $G(n,p)$ with $n = 500$ and different values of p. The blue crosses are the eigenvalues of $H/\sqrt{\alpha}$ and the red dots are for $H_0/\sqrt{\alpha}$. For comparison, the black dashed line is the unit circle. For the figures from top to bottom and from left to right, the values of p are taken to be $p = 0.5, p = 0.1, p = 0.08$ and $p = 0.05$ respectively.

H, that is, no eigenvalues of H that are far outside the support of the spectrum of H_0 (described in Theorem 1.2).

Theorem 1.6. *Assume* $0 < p \le p_0 < 1$ *for a constant* p_0 *and* $p \ge \frac{\log^{2/3+\varepsilon} n}{n^{1/6}}$ *for* $\varepsilon > 0$. *Let* A *be the adjacency matrix for an Erdős-Rènyi random graph* $G(n,p)$. *Let* $\frac{1}{\sqrt{\alpha}}H$ *be a rescaling of the non-backtracking spectrum operator for* A *defined in* (1.1). *Then, with probability* $1 - o(1)$, *each eigenvalue of* $\frac{1}{\sqrt{\alpha}}H$ *is within distance* $R = 40\sqrt{\frac{\log n}{np^2}}$ *of an eigenvalue of* $\frac{1}{\sqrt{\alpha}}H_0$, *defined in* (1.2).

We would like to mention that the above result could be improved to hold for $p \gg \log n/n$ and that each eigenvalue of of $\frac{1}{\sqrt{\alpha}}H$ is within distance $O((\log n/np)^{1/4})$ of an eigenvalue of $\frac{1}{\sqrt{\alpha}}H_0$, using a variant of Bauer-Fike perturbation theorem that appeared later in [CZ21, Corollary 2.4], as opposed to invoking the classical Bauer-Fike theorem in this paper (see Theorem 4.1).

1.2. **Outline.** We will describe the ESD of the partly averaged matrix H_0 to prove Proposition 1.2 in Section 2. In Section 3, we will show that the bulk ESDs of H and H_0 approach each other as n goes to infinity by using the replacement principle [TV10, Theorem 2.1] and in Section 4 we will use the Bauer-Fike theorem to prove Theorem 1.6, showing that the partly averaged matrix H_0 has eigenvalues close to those of H in the limit as $n \to \infty$.

1.3. **Background definitions.** We give a few definitions to make clear the bulk convergence described in Theorem 1.4 between empirical spectral distribution measures of H and H_0. For an $n \times n$ matrix M_n with eigenvalues $\lambda_1, \ldots, \lambda_n$, the empirical spectral measure μ_{M_n} of M_n is defined to be

$$\mu_{M_n} = \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i},$$

where δ_x is the Dirac delta function with mass 1 at x. Note that μ_{M_n} is a probability measure on the complex numbers $\mathbb{C}$. The empirical spectral distribution (ESD) for M_n is defined to be

$$F^{M_n}(x,y) = \frac{1}{n}\#\left\{\lambda_i : \mathsf{Re}(\lambda_i) \le x \text{ and } \mathsf{Im}(\lambda_i) \le y\right\}.$$

For T a topological space (for example $\mathbb{R}$ or $\mathbb{C}$) and $\mathcal{B}$ its Borel σ-field, we can define convergence of a sequence $(\mu_n)_{n\ge1}$ of random probability measures on $(T,\mathcal{B})$ to a nonrandom probability measure μ also on $(T,\mathcal{B})$ as follows. We say that μ_n *converges weakly to* μ *in probability as* $n\to\infty$ (written $\mu_n \to \mu$ in probability) if for all bounded continuous functions $f : T \to \mathbb{R}$ and all $\epsilon > 0$ we have

$$\Pr\left(\left|\int_T f\,d\mu_n - \int_T f\,d\mu\right| > \epsilon\right) \to 0 \text{ as } n \to \infty.$$

Also, we say that μ_n *converges weakly to* μ *almost surely as* $n\to\infty$ (written $\mu_n \to \mu$ a.s.) if for all bounded continuous functions $f : T \to \mathbb{R}$, we have that $|\int_T f\,d\mu_n - \int_T f\,d\mu| \to 0$ almost surely as $n \to \infty$.

We will use $\|A\|_F := \operatorname{tr}(AA^*)^{1/2}$ to denote the Frobenius norm or Hilbert-Schmidt norm, and $\|A\|$ to denote the operator norm. We denote $\|A\|_{\max} = \max_{ij}|a_{ij}|$. We use the notation $o(1)$ to denote a small quantity that tends to zero as n goes to infinity. We use the asymptotic notations $f(n) \ll g(n)$ if $f(n) = o(g(n))$ and $f(n) = O(g(n))$ if $f(n) \le Cg(n)$ for a constant $C > 0$ when n is sufficiently large.

2. The spectrum of H_0

We are interested in the limiting ESD of H when scaled to have bounded support (except for one outlier eigenvalue), and so we will work with the following rescaled conjugation of H, which has the same eigenvalues as $H/\sqrt{\alpha}$.

$$\widetilde{H} := \frac{1}{\sqrt{\alpha}}\begin{pmatrix} \frac{1}{\sqrt{\alpha}}I & 0 \\ 0 & I\end{pmatrix}\begin{pmatrix} A & I-D \\ I & 0\end{pmatrix}\begin{pmatrix} \sqrt{\alpha}I & 0 \\ 0 & I\end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{\alpha}}A & \frac{1}{\alpha}(I-D) \\ I & 0\end{pmatrix}.$$

Note that the diagonal matrix $\frac{1}{\alpha}(I-D)$ is equal to $-I$ in expectation, and so we will compare the eigenvalues of $\widetilde{H}$ to those of the partly averaged matrix $\widetilde{H}_0$, noting that $\widetilde{H} = \widetilde{H}_0 + E$, where

$$\widetilde{H}_0 := \begin{pmatrix} \frac{1}{\sqrt{\alpha}}A & -I \\ I & 0\end{pmatrix} \qquad \text{and} \qquad E := \begin{pmatrix} 0 & I + \frac{1}{\alpha}(I-D) \\ 0 & 0\end{pmatrix}. \tag{2.1}$$

We will show that $\widetilde{H}_0$ is explicitly diagonalizable in terms of the eigenvectors and eigenvalues of $\frac{1}{\sqrt{\alpha}}A$, and then use this information to find an explicit form for the characteristic polynomial for $\widetilde{H}_0$.

2.1. **Spectrum of $\widetilde{H}_0$: Proof of Proposition 1.2.** Since $\frac{1}{\sqrt{\alpha}}A$ is a real symmetric matrix, it has a set $v_1,\dots,v_n$ of orthonormal eigenvectors with corresponding real eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$. Thus we may write $A = U^T \operatorname{diag}(\lambda_1,\dots,\lambda_n)U$ where U is an orthogonal matrix. Consider the matrix $xI - \widetilde{H}_0$, and note that by adding x times row i of $xI - \widetilde{H}_0$ to row $n+i$

for each $i = 1, 2, \dots n$, we see that $\det(xI - \widetilde{H}_0) = \det(I + (xI - \frac{1}{\sqrt{\alpha}}A)x) = \det(x^2 I - \frac{x}{\sqrt{\alpha}}A + I)$. Conjugating to diagonalize A, we see that

$$\det(xI - \widetilde{H}_0) = \det(x^2 I - x\,\mathrm{diag}(\lambda_1, \dots, \lambda_n) + I) = \prod_{i=1}^{n}(x^2 - \lambda_i x + 1). \tag{2.2}$$

With the characteristic polynomial for $\widetilde{H}_0$ factored into quadratics as in (2.2), we see that for each λ_i of $\frac{1}{\sqrt{\alpha}}A$, there are two eigenvalues μ_{2i-1} and μ_{2i} for $\widetilde{H}_0$ which are are the two solutions to $x^2 - \lambda_i x + 1 = 0$; thus,

$$\mu_{2i-1} = \frac{\lambda_i + \sqrt{\lambda_i^2 - 4}}{2} \qquad \text{and} \qquad \mu_{2i} = \frac{\lambda_i - \sqrt{\lambda_i^2 - 4}}{2}. \tag{2.3}$$

The eigenvalues of A are well-understood. We use the following results that exist in literature.

Theorem 2.1 ([KS03, LS18])**.** *Let A be the adjacency matrix for an Erdős-Rènyi random graph $G(n, p)$. Assume $0 < p \le p_0 < 1$ for a constant p_0 and $p \ge n^{-1+\phi}$ for a small constant $\phi > 0$. Then with probability $1 - o(1)$, the following holds for any $\epsilon > 0$:*

$$\lambda_1(A) = np(1 + o(1));$$

$$\max_{2\le i\le n} |\lambda_i(A) + p| \le L\sqrt{np(1-p)} + n^{\epsilon}\sqrt{np}\left(\frac{1}{(np)^2} + \frac{1}{n^{2/3}}\right),$$

where $L = 2 + \frac{s^{(4)}}{np} + O(\frac{1}{(np)^2})$ and $s^{(4)} = n^2 p \left[\frac{p^3+(1-p)^3}{n^2 p(1-p)} - \frac{3}{n^2}\right]$.

Proof. We collect relevant results regarding the eigenvalues of A from different works in the literature. In [KS03], it is shown that with probability $1-o(1)$, $\lambda_1(A) = (1+o(1))\max\{np, \sqrt{\Delta}\}$ where Δ is the maximum degree. As long as $np/\log n \to \infty$, $\max\{np, \sqrt{\Delta}\} = np$ (for the bounds on Δ see, for instance, the proof of Lemma 3.5 below).

The operator norm of $A - \mathbb{E}A$ and the extreme eigenvalues of A have been studied in various works (see [FK81, Vu07, BGBK17, EKYY13, LS18, HLY20, HK21]). In particular, in [LS18, Theorem 2.9], assuming $p \ge n^{-1+\phi}$, the authors proved that for any $\epsilon > 0$ and $C > 0$, the following estimate holds with probability at least $1 - n^{-C}$:

$$\left|\frac{1}{\sqrt{np(1-p)}}\|A - \mathbb{E}A\| - L\right| \le n^{\epsilon}\left(\frac{1}{(np)^2} + \frac{1}{n^{2/3}}\right)$$

with $L = 2 + \frac{s^{(4)}}{np} + O(\frac{1}{(np)^2})$ and $s^{(4)} = n^2 p\left[\frac{p^3+(1-p)^3}{n^2 p(1-p)} - \frac{3}{n^2}\right] = 1 + O(p)$. The conclusion of the theorem follows immediately from the classical Weyl's inequality that $\max_{2\le i\le n} |\lambda_i(A) + p| = \max_{2\le i\le n} |\lambda_i(A) - \lambda_i(\mathbb{E}A)| \le \|A - \mathbb{E}A\|$. $\square$

Now we are ready to derive Proposition 1.2.

Proof of Proposition 1.2. Note that $\lambda_i = \lambda_i(A)/\sqrt{\alpha}$ and $\alpha = (n-1)p - 1$. We have that

$$\lambda_1 = \sqrt{np}(1 + o(1)) \quad \text{and} \quad \max_{2\le i\le n} |\lambda_i| \le 2\sqrt{1-p}(1 + o(1)) \tag{2.4}$$

with probability $1 - o(1)$. Therefore, for λ_1, we see from (2.3) that μ_1, μ_2 are real eigenvalues and

$$\mu_1 = \sqrt{np}(1 + o(1)) \quad \text{and} \quad \mu_2 = \frac{1}{\sqrt{np}}(1 + o(1))$$

with probability $1-o(1)$. Next, by Theorem 2.1, it holds with probability $1-o(1)$ for any $2 \le i \le n$ that

$$\lambda_i^2 = \frac{\lambda_i^2(A)}{\alpha} \le \frac{1}{\alpha}\left[L\sqrt{np(1-p)} + p + n^{\epsilon}\sqrt{np}\left(\frac{1}{(np)^2}+\frac{1}{n^{2/3}}\right)\right]^2.$$

Since $p \ge C/\sqrt{n}$ for a sufficiently large constant C, we have

$$\begin{aligned}\lambda_i^2 &\le \frac{1}{\alpha}\left[2\sqrt{np(1-p)} + O(\max\{p,(np)^{-1/2},p^{1/2}n^{-1/6+\epsilon}\})\right]^2\\ &= \frac{4np(1-p)+O(\max\{p\sqrt{np},1,pn^{1/3+\epsilon}\})}{np-(p+1)}\\ &= \frac{4(1-p)+O(\max\{\sqrt{p/n},(np)^{-1},n^{-2/3+\epsilon}\})}{1-\frac{p+1}{np}}\\ &= 4(1-p)+O\left(\max\{\sqrt{p/n},(np)^{-1},n^{-2/3+\epsilon}\}\right)+O\left(\frac{1}{np}\right) \le 4-3p,\end{aligned}$$

for all sufficiently large n. Hence, for all $i \ge 2$, we have $\lambda_i^2 < 4$ and thus μ_{2i-1}, μ_{2i} are complex eigenvalues with magnitude 1 (since $|\mu_{2i-1}| = |\mu_{2i}| = 1$). One should also note that $\mu_{2i-1}\mu_{2i} = 1$ for every i, and that whenever μ_{2i-1} is complex (i.e., $i \ge 2$), its complex conjugate is $\overline{\mu}_{2i-1} = \mu_{2i}$.

Furthermore, note that $\mathsf{Re}\mu_{2i-1} = \mathsf{Re}\mu_{2i} = \lambda_i/2 = \lambda_i(A)/2\sqrt{\alpha}$. It is known that the empirical spectral measure of $A/\sqrt{np(1-p)}$ converges to the semicircular law supported on $[-2,2]$ assuming $np \to \infty$ (see for instance [KP93] or [TVW13]). We have the ESD of the scaled real parts of μ_j

$$\frac{1}{2n}\sum_{j=1}^{2n}\delta_{\frac{2\mathsf{Re}\mu_j}{\sqrt{1-p}}} \to \mu_{sc}$$

weakly almost surely where μ_{sc} is the semicircular law supported on $[-2,2]$. The proof of Proposition 1.2 is now complete. □

As mentioned in Remark 1.3, when p gets smaller than $n^{-1/2}$, more real eigenvalues of $\widetilde{H}_0$ will emerge. We could identify some of these eigenvalues, using recent results of [EKYY13, LS18, HLY20, HK21] in the study of the extreme eigenvalues of A. For instance, in [LS18, Corollary 2.13], assume $n^{2\phi-1} \le p \le n^{-2\phi'}$ for $\phi > 1/6$ and $\phi' > 0$. Then

$$\lim_{n\to\infty}\mathbb{P}\left(n^{2/3}\left(\frac{1}{\sqrt{np(1-p)}}\lambda_2(A) - \mathcal{L} - a\right) \le s\right) = F_1^{TW}(s), \tag{2.5}$$

where $\mathcal{L} = 2+\frac{1}{np}+O(\frac{1}{n^2p^2})$, $a = \sqrt{\frac{p}{n(1-p)}}$ and $F_1^{TW}(s)$ is the Tracy-Widom distribution function. Therefore, when $p \ge n^{-2/3+\epsilon}$,

$$\lambda_2(A) = 2\sqrt{np(1-p)} + p + \sqrt{\frac{1-p}{np}} + O(\frac{\sqrt{np}}{n^{2/3}}).$$

Note that if $p < \frac{1-p}{n^{1/3}}$, then $p < \sqrt{\frac{1-p}{np}}$ and thus for $n^{-2/3+\epsilon} \le p \le n^{-1/2} \ll n^{-1/3}$

$$\begin{aligned}\lambda_2^2 - 4 = (\frac{\lambda_2(A)}{\sqrt{\alpha}})^2 - 4 &= \frac{\left(2\sqrt{np(1-p)}+\sqrt{\frac{1-p}{np}}+p+O(\frac{\sqrt{np}}{n^{2/3}})\right)^2}{np-(p+1)} - 4\\ &= \frac{\left(2\sqrt{np(1-p)}+\sqrt{\frac{1-p}{np}}+p\right)^2+O(\frac{np}{n^{2/3}})}{np-(p+1)} - 4\end{aligned}$$

$$= \frac{4np(1-p)+4(1-p)+4p\sqrt{np(1-p)}+O(\frac{np}{n^{2/3}})}{np-(p+1)} - 4$$

$$= \frac{4(1-p)+\frac{4(1-p)}{np}+4p\sqrt{\frac{1-p}{np}}+O(\frac{1}{n^{2/3}})}{1-\frac{p+1}{np}} - 4$$

$$= -4p+4p\sqrt{\frac{1-p}{np}}+\frac{4(1-p)(2+p)}{np}+O(n^{-2/3}) > 0.$$

Hence, from (2.3), both μ_3 and μ_4 are real. The convergence result (2.5) holds for finitely many extreme eigenvalues of A and thus they also generate real eigenvalues for $\widetilde{H}_0$.

The fluctuation of the extreme eigenvalues of A has been obtained in [HLY20, Corollary 1.5] for $n^{-7/9} \ll p \ll n^{-2/3}$ and in [HK21] for the remaining range of p up to $p \geq n^{-1+\epsilon}$. One could use similar discussion as above to extract information about the real eigenvalues of $\widetilde{H}_0$. The details are omitted.

2.2. **$\widetilde{H}_0$ is diagonalizable.** We can now demonstrate an explicit diagonalization for $\widetilde{H}_0$. Since μ_{2i-1} and μ_{2i} are solutions to $\mu^2 - \mu\lambda_i + 1 = 0$, one can check that the following vectors

$$y_{2i-1}^* = \frac{1}{\sqrt{1+|\mu_{2i-1}|^2}}\begin{pmatrix}-\mu_{2i-1}v_i^T & v_i^T\end{pmatrix} \quad \text{and} \quad y_{2i}^* = \frac{1}{\sqrt{1+|\mu_{2i}|^2}}\begin{pmatrix}-\mu_{2i}v_i^T & v_i^T\end{pmatrix} \tag{2.6}$$

satisfy $y_{2i-1}^*\widetilde{H}_0 = \mu_{2i-1}y_{2i-1}^*$ and $y_{2i}^*\widetilde{H}_0 = \mu_{2i}y_{2i}^*$ for all i. Besides, y_{2i-1} and y_{2i} are unit vectors. For $1 \leq i \leq n$, define the vectors

$$x_{2i-1} = \frac{\sqrt{1+|\mu_{2i-1}|^2}}{\mu_{2i}-\mu_{2i-1}}\begin{pmatrix}v_i\\ \mu_{2i}v_i\end{pmatrix} \quad \text{and} \quad x_{2i} = \frac{\sqrt{1+|\mu_{2i}|^2}}{\mu_{2i-1}-\mu_{2i}}\begin{pmatrix}v_i\\ \mu_{2i-1}v_i\end{pmatrix}. \tag{2.7}$$

Denoting

$$Y = \begin{pmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_{2n}^* \end{pmatrix} \quad \text{and} \quad X = \begin{pmatrix} x_1, x_2, \dots, x_{2n} \end{pmatrix}$$

we see that $X = Y^{-1}$ since $v_1, \dots, v_n$ are orthonormal. Also it is easy to check that $Y\widetilde{H}_0X = \operatorname{diag}(\mu_1, \dots, \mu_{2n})$.

3. The bulk distribution: proving Theorem 1.4

We begin be re-stating Theorem 1.4 using the conjugated matrices defined in (2.1).

Theorem 3.1. *Let A be the adjacency matrix for an Erdős-Rènyi random graph $G(n,p)$. Assume $0 < p \leq p_0 < 1$ for a constant p_0 and $np/\log n \to \infty$ with n. Let $\widetilde{H}$ be the rescaled conjugation of the non-backtracking spectrum operator for A defined in (2.1), and let $\widetilde{H}_0$ be its partial derandomization, also defined in (2.1). Then, $\mu_{\widetilde{H}} - \mu_{\widetilde{H}_0}$ converges almost surely (thus, also in probability) to zero as n goes to infinity.*

To prove Theorem 3.1, we will show that the bulk distribution of $\widetilde{H}$ matches that of $\widetilde{H}_0$ using the replacement principle [TV10, Theorem 2.1], which we rephrase slightly as a perturbation result below (see Theorem 3.2). First, we give a few definitions that we will use throughout this section. We say that a random variable $X_n \in \mathbb{C}$ is *bounded in probability* if

$$\lim_{C\to\infty} \liminf_{n\to\infty} \Pr(|X_n| \leq C) = 1$$

and we say that X_n is *almost surely bounded* if

$$\Pr\left(\limsup_{n\to\infty} |X_n| < \infty\right) = 1.$$

Theorem 3.2 (Replacement principle [TV10])**.** *Suppose for each m that M_m and $M_m + P_m$ are random $m \times m$ matrices with entries in the complex numbers. Assume that*

$$\frac{1}{m}\|M_m\|_F^2 + \frac{1}{m}\|M_m + P_m\|_F^2 \text{ is bounded in probability (resp., almost surely)} \tag{3.1}$$

and that, for almost all complex numbers $z \in \mathbb{C}$,

$$\frac{1}{m}\log\left|\det\left(M_m + P_m - zI\right)\right| - \frac{1}{m}\log\left|\det\left(M_m - zI\right)\right| \tag{3.2}$$

converges in probability (resp., almost surely) to zero; in particular, this second condition requires that for almost all $z \in \mathbb{C}$, the matrices $M_m + P_m - zI$ and $M_m - zI$ have non-zero determinant with probability $1 - o(1)$ (resp., almost surely non-zero for all but finitely many m).

Then $\mu_{M_m} - \mu_{M_m+P_m}$ converges in probability (resp., almost surely) to zero.

Note that there is no independence assumption anywhere in Theorem 3.2; thus, entries in P_m may depend on entries in M_m and vice versa.

We will use the following corollary of Theorem 3.2, which essentially says that if the perturbation P_m has largest singular value of order less than the smallest singular value for $M_m - zI$ for almost every $z \in \mathbb{C}$, then adding the perturbation P_m does not appreciably change the bulk distribution of M_m.

Corollary 3.3. *For each m, let M_m and P_m be random $m \times m$ matrices with entries in the complex numbers, and let $f(z, m) \geq 1$ be a real function depending on z and m. Assume that*

$$\frac{1}{m}\|M_m\|_F^2 + \frac{1}{m}\|M_m + P_m\|_F^2 \text{ is bounded in probability (resp., almost surely),} \tag{3.3}$$

and

$$f(z, m)\|P_m\| \text{ converges in probability (resp., almost surely) to zero,} \tag{3.4}$$

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

$$\left\|(M_m - zI)^{-1}\right\| \leq f(z, m), \tag{3.5}$$

with probability tending to 1 (resp., almost surely for all but finitely many m).

Then $\mu_{M_m} - \mu_{M_m+P_m}$ converges in probability (resp., almost surely) to zero.

Proof. We will show that the three conditions (3.3), (3.4), and (3.5) of Corollary 3.3 together imply the two conditions needed to apply Theorem 3.2.

First note that (3.3) directly implies the first condition (3.1) of Theorem 3.2. Next, we will show in the remainder of the proof that (3.2) of Theorem 3.2 holds by noting that sufficiently small perturbations have a small effect on the singular values, and also the absolute value of the determinant is equal to the product of the singular values.

Let z be a complex number for which (3.5) holds, let $M_m - zI$ have singular values $\sigma_1 \geq \cdots \geq \sigma_m$, and let $M_m + P_m - zI$ have singular values $\sigma_1 + s_1 \geq \sigma_2 + s_2 \geq \cdots \geq \sigma_m + s_m$. We will use the following result, which is sometimes called Weyl's perturbation theorem for singular values, to show that the s_i are small.

Lemma 3.4 ([Cha09, Theorem 1.3])**.** *Let A and B be $m \times n$ real or complex matrices with singular values $\sigma_1(A) \geq \cdots \geq \sigma_{\min\{m,n\}}(A) \geq 0$ and $\sigma_1(B) \geq \cdots \geq \sigma_{\min\{m,n\}}(B) \geq 0$, respectively. Then*

$$\max_{1\leq j\leq \min\{m,n\}} |\sigma_j(A) - \sigma_j(B)| \leq \|A - B\|.$$

We then have that

$$\max_{1\leq i\leq m} |s_i| \leq \|P_m\|,$$

and by (3.5),

$$\max_{1\leq i\leq m} \frac{|s_i|}{\sigma_i} \leq f(z,m)\|P_m\|$$

which converges to zero in probability (resp., almost surely) by (3.4). Thus we know that

$$|\log(1 + s_i/\sigma_i)| \leq 2\,|s_i/\sigma_i| \leq 2f(z,m)\|P_m\|,$$

where the inequalities hold with probability tending to 1 (resp., almost surely for all sufficiently large m). Using the fact that the absolute value of the determinant is the product of the singular values, we may write (3.2) as

$$\left|\frac{1}{m}\left(\log\prod_{i=1}^{m}(\sigma_i + s_i) - \log\prod_{i=1}^{m}\sigma_i\right)\right| = \frac{1}{m}\left|\sum_{i=1}^{m}\log\left(1 + \frac{s_i}{\sigma_i}\right)\right| \leq 2f(z,m)\|P_m\|,$$

which converges to zero in probability (resp., almost surely) by (3.4). Thus, we have shown that (3.1) and (3.2) hold, which completes the proof. □

3.1. **Proof of Theorem 3.1.** The proof of Theorem 3.1 will follow from Corollary 3.3 combined with lemmas showing that the conditions (3.3), (3.4), and (3.5) of Corollary 3.3 are satisfied. Indeed, Lemma 3.7 verifies (3.3), Lemma 3.9 verifies (3.5) and (3.4) follows by combining Lemma 3.5 and Lemma 3.9. Note that the assumption $np/\log n \to \infty$ in Theorem 3.1 is only needed to prove conditions (3.3) and (3.4). Condition (3.5) in fact follows for any p and for more general matrices–see the proof of Lemma 3.9.

In Corollary 3.3, we will take M_m to be the partly derandomized matrix $\widetilde{H}_0$ and P_m to be the matrix E (see (2.1)), where we suppress the dependence of $\widetilde{H}_0$ and E on $n = m/2$ to simplify the notation. There are two interesting features: first, the singular values of $\widetilde{H}_0$ may be written out explicitly in terms of the eigenvalues of the Hermitian matrix A (which are well understood; see Lemma 3.9); and second, the matrix E is completely determined by the matrix $\widetilde{H}_0$, making this a novel application of the replacement principle (Theorem 3.2 and Corollary 3.3) where there two matrices are highly dependent.

Lemma 3.5. *Assume $0 < p \leq p_0 < 1$ for a constant p_0. Further assume $np/\log n \to \infty$. For E as defined in (2.1), we have that $\|E\| \leq 20\sqrt{\frac{\log n}{np}}$ almost surely. In particular, $\|E\|$ converges almost surely to zero.*

Proof. First, note that $\mathbb{E}D = (n-1)pI = (\alpha+1)I$ and thus

$$E := \begin{pmatrix} 0 & I + \frac{1}{\alpha}(I - D) \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{\alpha}(\mathbb{E}D - D) \\ 0 & 0 \end{pmatrix}.$$

Since $\mathbb{E}D - D$ is a diagonal matrix, it is easy to check that

$$\|E\| = \|E\|_{\max} = \frac{1}{\alpha}\|D - \mathbb{E}D\|_{\max} = \frac{1}{\alpha}\max_{1\leq i\leq n}|D_{ii} - \mathbb{E}D_{ii}| = \frac{1}{\alpha}\max_{1\leq i\leq n}|D_{ii} - (n-1)p|.$$

Next we denote $d_i := D_{ii}$ for $1 \le i \le n$, the degree of vertex i and consider their order statistics $d_{(1)} \ge d_{(1)} \ge \ldots \ge d_{(n)}$. Namely, $d_{(1)}$ is the largest degree and $d_{(n)}$ is the minimum degree. Since

$$\|E\| = \frac{1}{\alpha} \max_{1\le i\le n} |D_{ii} - (n-1)p| = \max\left\{\frac{1}{\alpha}|d_{(1)} - (n-1)p|, \frac{1}{\alpha}|d_{(n)} - (n-1)p|\right\}, \tag{3.6}$$

it is enough to show bounds for both

$$\frac{1}{\alpha}|d_{(1)} - (n-1)p| \quad \text{and} \quad \frac{1}{\alpha}|d_{(n)} - (n-1)p|.$$

Let us define X_K to be the number of vertices of degree at least K, that is,

$$X_K = \max\{i : d_{(i)} \ge K\} = \sum_{i=1}^{n} \mathbf{1}_{\{d_i \ge K\}}.$$

Note that $\Pr(d_{(1)} < K) = \Pr(X_K = 0)$. Since X_K is always a non-negative integer, by Markov's inequality,

$$\Pr(d_{(1)} \ge K) = 1 - \Pr(X_K = 0) = \Pr(X_K \ge 1) \le \mathbb{E}X_K^2. \tag{3.7}$$

On the other hand, $X_K^2 = X_K + 2\binom{X_K}{2}$, we have $\mathbb{E}X_K^2 = \mathbb{E}X_K + 2\mathbb{E}\binom{X_K}{2}$. Recall that $X_K = \sum_{i=1}^{n} \mathbf{1}_{\{d_i \ge K\}}$. Since d_i's have the same distribution, we obtain that

$$\mathbb{E}X_K = n\Pr(d_1 \ge K) = n\Pr(\sum_{j=2}^{n} a_{1j} \ge K). \tag{3.8}$$

Notice that $\binom{X_K}{2}$ is the number of 2-tuples of vertices with degrees at least K. We have

$$\begin{aligned}
\mathbb{E}\binom{X_K}{2} &= \binom{n}{2}\Pr(d_1 \ge K, d_2 \ge K) = \binom{n}{2}\Pr(\sum_{j\ne 1} a_{1j} \ge K, \sum_{j\ne 2} a_{2j} \ge K) \\
&\le \binom{n}{2}\Pr(\sum_{j=3}^{n} a_{1j} \ge K-1, \sum_{j=3}^{n} a_{2j} \ge K-1) \\
&= \binom{n}{2}\Pr(\sum_{j=3}^{n} a_{1j} \ge K-1)\cdot \Pr(\sum_{j=3}^{n} a_{2j} \ge K-1) \\
&= \binom{n}{2}\Pr(\sum_{j=3}^{n} a_{1j} \ge K-1)^2.
\end{aligned} \tag{3.9}$$

Next we will apply the following general form of Chernoff bound.

Theorem 3.6 (Chernoff bound [Che52]). *Assume $\xi_1, \ldots, \xi_n$ are iid random variables and $\xi_i \in [0,1]$ for all i. Let $p = \mathbb{E}\xi_i$ and $S_n = \sum_{i=1}^{n} \xi_i$, then for any $\varepsilon > 0$,*

$$\Pr(S_n - np \ge n\varepsilon) \le \exp\left(-RE(p+\varepsilon||p)n\right);$$
$$\Pr(S_n - np \le -n\varepsilon) \le \exp\left(-RE(p-\varepsilon||p)n\right)$$

where $RE(p||q) = p\log(\frac{p}{q}) + (1-p)\log(\frac{1-p}{1-q})$ is the relative entropy or Kullback-Leibler divergence.

By our assumption, $np = \omega(n)\log n$ where $\omega(n)$ is a positive function that tends to infinity with n. Now take $K = (n-1)p + npt$ where $t = t(n) = 10\sqrt{\frac{\log n}{np}}$ (say). Our assumption $np/\log n \to \infty$ implies $t \to 0$ with n. Thus

$$\Pr(\sum_{j=2}^{n} a_{1j} \ge K) = \Pr(\sum_{j=1}^{n} a_{1j} - (n-1)p \ge npt) \le \exp(-\mathrm{RE}(p + pt||p)n)$$

where

$$\begin{aligned}\mathrm{RE}(p+pt||p) &= p(1+t)\log(1+t)+(1-p-pt)\log(\frac{1-p-pt}{1-p})\\ &= p(1+t)\log(1+t)-(1-p-pt)\log(1+\frac{pt}{1-p-pt})\\ &> p(1+t)(t-t^2/2)-pt = pt^2(1-t)/2\end{aligned}$$

by the elementary inequalities $x-x^2/2<\log(1+x)<x$ for $x>0$.

Therefore, for n sufficiently large, taking $t=10\sqrt{\frac{\log n}{np}}$, we get

$$\Pr(\sum_{j=2}^{n} a_{1j} \geq (n-1)p+npt) \leq \exp(-\frac{npt^2(1-t)}{2}) \leq \exp(-20\log n) = n^{-20}.$$

By (3.8), we know that $\mathbb{E}X_K \leq n^{-9}$. Using the similar computation, we get from (3.9) that $\mathbb{E}\binom{X_K}{2} \leq n^{-9}$. Therefore $\mathbb{E}X_K^2 = \mathbb{E}X_K + 2\mathbb{E}\binom{X_K}{2} \leq 2n^{-9}$ and by (3.7), we conclude that

$$\Pr(d_{(1)} \geq (n-1)p+npt) \leq 2n^{-9}. \tag{3.10}$$

We can use the same method to compute $\Pr(d_{(n)} \leq (n-1)p-npt)$. We describe it briefly here. For convenience, denote $L=(n-1)p-npt$ where $t=10\sqrt{\frac{\log n}{np}}$. Let Y_L be the number of vertices with degrees at most L. We have the relations

$$Y_L = \max\{i : d_{(i)} \leq L\} = \sum_{i=1}^{n} \mathbf{1}_{\{d_i \leq L\}}$$

and

$$\Pr(d_{(n)} > L) = \Pr(Y_L = 0).$$

Therefore, by Markov's inequality,

$$\Pr(d_{(n)} \leq L) = \Pr(Y_L \geq 1) \leq \mathbb{E}Y_L^2 = \mathbb{E}Y_L + 2\mathbb{E}\binom{Y_L}{2}$$

where

$$\mathbb{E}Y_L = n\Pr(d_1 \leq L) = n\Pr(\sum_{j=2}^{n} a_{1j} \leq L)$$

and

$$\begin{aligned}\mathbb{E}\binom{Y_L}{2} &= \binom{n}{2}\Pr(d_1 \leq L, d_2 \leq L) = \binom{n}{2}\Pr(\sum_{j\neq 1} a_{1j} \leq L, \sum_{j\neq 2} a_{2j} \leq L)\\ &\leq \binom{n}{2}\Pr(\sum_{j=3}^{n} a_{1j} \leq L, \sum_{j=3}^{n} a_{2j} \leq L)\\ &= \binom{n}{2}\Pr(\sum_{j=3}^{n} a_{1j} \leq L)\cdot Pr(\sum_{j=3}^{n} a_{2j} \leq L)\\ &= \binom{n}{2}\Pr(\sum_{j=3}^{n} a_{1j} \leq L)^2.\end{aligned}$$

We apply the Chernoff bound and notice that for n sufficiently large, $t=t(n)<0.01$ (say),

$$\begin{aligned}\mathrm{RE}(p-pt||p) &= p(1-t)\log(1-t)+(1-p+pt)\log(\frac{1-p+pt}{1-p})\\ &= p(1-t)\log(1-t)-(1-p+pt)\log(1-\frac{pt}{1-p+pt})\end{aligned}$$

$$> p(1-t)(-t-\frac{3}{5}t^2) + pt = \frac{1}{5}pt^2(2+3t) \geq \frac{2}{5}pt^2$$

where we use the fact that $\log(1-x) < -x$ for $x \in (0,1)$ and $\log(1-x) > -x - \frac{3}{5}x^2$ for $x \in (0, 0.01)$.

Repeating the similar computation, we get

$$\Pr(d_{(n)} \leq (n-1)p - npt) \leq 2n^{-9}. \tag{3.11}$$

Therefore, by (3.10) and (3.11), we obtain

$$\begin{aligned}\Pr(\frac{1}{npt}|d_{(1)} - (n-1)p| \geq 1) &\leq \Pr(d_{(1)} \geq (n-1)p + npt) + \Pr(d_{(1)} \leq (n-1)p - npt)\\ &\leq \Pr(d_{(1)} \geq (n-1)p + npt) + \Pr(d_{(n)} \leq (n-1)p - npt)\\ &\leq 4n^{-9}.\end{aligned}$$

By the Borel-Cantelli lemma, we have that $\frac{1}{np}|d_{(1)} - (n-1)p| \leq t = 10\sqrt{\frac{\log n}{np}}$ converges to zero almost surely. Since $\alpha/np \to 1$, we get $\frac{1}{\alpha}|d_{(1)} - (n-1)p| \leq 20\sqrt{\frac{\log n}{np}}$ almost surely. Likewise, since

$$\Pr(\frac{1}{npt}|d_{(n)} - (n-1)p| \geq 1) \leq \Pr(d_{(1)} \geq (n-1)p + npt) + \Pr(d_{(n)} \leq (n-1)p - npt) \leq 4n^{-9},$$

with the same $t = 10\sqrt{\frac{\log n}{np}}$, we also get $\frac{1}{\alpha}|d_{(n)} - (n-1)p| \leq 20\sqrt{\frac{\log n}{np}}$ converges to zero almost surely.

□

To show (3.3), we combine Hoeffding's inequality and Lemma 3.5 to prove the following lemma.

Lemma 3.7. *Assume $0 < p \leq p_0 < 1$ for a constant p_0. Further assume $np/\log n \to \infty$. For $\widetilde{H}_0$ and E as defined in (2.1), we have that both $\frac{1}{2n}\|\widetilde{H}_0\|_F^2$ and $\frac{1}{2n}\|\widetilde{H}_0 + E\|_F^2$ are almost surely bounded.*

Proof. We begin by stating Hoeffding's inequality [Hoe63].

Theorem 3.8 (Hoeffding's inequality [Hoe63])**.** *Let $\beta_1, \ldots, \beta_k$ be independent random variables such that for $1 \leq i \leq k$ we have $\Pr(\beta_i \in [a_i, b_i]) = 1$. Let $S := \sum_{i=1}^k \beta_i$. Then for any real t,*

$$\Pr(|S - \mathbb{E}(S)| \geq kt) \leq 2\exp\left(-\frac{2k^2t^2}{\sum_{i=1}^k (b_i - a_i)^2}\right).$$

Recall that $\alpha = (n-1)p - 1$ and $\widetilde{H}_0 = \begin{pmatrix} \frac{1}{\sqrt{\alpha}}A & -I \\ I & 0 \end{pmatrix}$, where $A = (a_{ij})_{1\leq i,j \leq n}$ is the adjacency matrix of an Erdős-Rènyi random graph $G(n,p)$. Thus

$$\|\widetilde{H}_0\|_F^2 = \frac{1}{\alpha}\|A\|_F^2 + 2\|I\|_F^2 = \frac{1}{\alpha}\sum_{i,j} a_{ij}^2 + 2n = \frac{2}{\alpha}\sum_{i<j} a_{ij} + 2n.$$

To apply Hoeffding's inequality, note that a_{ij} $(i < j)$ are iid random variables each taking the value 1 with probability p and 0 otherwise. Let $b_i = 1$ and $a_i = 0$ for all i, and let $k = \binom{n}{2}$, which is the number of random entries in A (recall that the diagonal of A is all zeros by assumption). Letting $S = \sum_{i<j} a_{ij}$, we see that $\mathbb{E}S = kp$ and so

$$\Pr\Big(|S - kp| \geq kt\Big) \leq 2\exp(-2kt^2).$$

Since $\|\widetilde{H}_0\|_F^2 = \frac{2}{\alpha}S + 2n$, we obtain that

$$\Pr\left(\left|\frac{1}{2n}\|\tilde{H}_0\|_F^2 - 1\right| \geq \frac{kt}{n\alpha}\right) \leq 2\exp(-2kt^2).$$

Take $t = p$. For n sufficiently large,

$$\frac{kt}{n\alpha} \leq \frac{\binom{n}{2}t}{n(n-1)p/2} \leq t/p$$

and since $p \geq \omega(n)\log n/n$ for $\omega(n) > 0$ and $\omega(n) \to \infty$ with n, we get

$$\Pr\left(\left|\frac{1}{2n}\|\tilde{H}_0\|_F^2 - 1\right| \geq 1\right) \leq 2\exp(-2kt^2) \leq 2\exp(-\omega(n)^2\log^2 n/2).$$

By the Borel-Cantelli lemma, we conclude that $\frac{1}{2n}\|\tilde{H}_0\|_F^2$ is bounded almost surely. Since $\|E\|_{\max} = \|E\|$, by triangle inequality, we see

$$\frac{1}{2n}\|\widetilde{H}_0 + E\|_F^2 \leq \frac{1}{2n}(\|\tilde{H}_0\|_F + \|E\|_F)^2 \leq \frac{1}{n}\|\tilde{H}_0\|_F^2 + \frac{1}{n}\|E\|_F^2$$
$$\leq \frac{1}{n}\|\tilde{H}_0\|_F^2 + \|E\|.$$

By Lemma 3.5, we get $\frac{1}{2n}\|\widetilde{H}_0 + E\|_F^2$ is bounded almost surely. This completes the proof. □

The last part of proving Theorem 3.1 by way of Corollary 3.3 is proving that (3.5) holds with $M_m = \widetilde{H}_0$ and $f(z,m) = C_z$, a constant depending only on z. The following lemma will be proved by writing a formula for the singular values of $\widetilde{H}_0$ in terms of the eigenvalues of the adjacency matrix A, which are well understood. A number of elementary technical details will be needed to prove that the smallest singular value is bounded away from zero, and these appear in Lemma 3.10.

Lemma 3.9. *Let $\widetilde{H}_0$ be as defined in (2.1) and let z be a complex number such that $\mathsf{Im}(z) \neq 0$ and $|z| \neq 1$ (note that these conditions exclude a set of complex numbers of Lebesgue measure zero). Then there exists a constant C_z depending only on z such that $\left\|(\widetilde{H}_0 - zI)^{-1}\right\| \leq C_z$ with probability 1 for all but finitely many n.*

Proof. We will compute all the singular values of $\widetilde{H}_0 - zI$, showing that they are bounded away from zero by a constant depending on z. The proof does not use randomness and depends only on facts about the determinant and singular values and on the structure of $\widetilde{H}_0$; in fact, the proof is the same if $\widetilde{H}_0$ is replaced with any matrix $\begin{pmatrix} M & -I \\ I & 0 \end{pmatrix}$ with M Hermitian.

To find the singular values of $\widetilde{H}_0$ we will compute the characteristic polynomial $\chi(\tilde{w})$ for $(\widetilde{H}_0 - zI)(\widetilde{H}_0 - zI)^*$, using the definition of $\widetilde{H}_0$ in (2.1), and assuming that $\tilde{w} = w + 1 + |z|^2$; thus,

$$\chi(\tilde{w}) := \det\left((\widetilde{H}_0 - zI)(\widetilde{H}_0 - zI)^* - (w + 1 + |z|^2)I\right)$$
$$= \det\begin{pmatrix} \frac{A^2}{\alpha} - (z+\bar{z})\frac{A}{\sqrt{\alpha}} - wI & \frac{A}{\sqrt{\alpha}} + (\bar{z} - z)I \\ \frac{A}{\sqrt{\alpha}} + (z - \bar{z})I & -wI \end{pmatrix}.$$

We can use the fact that if $\begin{pmatrix} X & Y \\ Z & W \end{pmatrix}$ is a matrix composed of four $n \times n$ square blocks where W and Z commute, then $\det\begin{pmatrix} X & Y \\ Z & W \end{pmatrix} = \det(XW - YZ)$ (see [Sil00, Theorem 3]). Thus, it is equivalent to consider

$$\det\left(w\left(\frac{A^2}{\alpha}-(z+\bar{z})\frac{A}{\sqrt{\alpha}}-wI\right)+\left(\frac{A}{\sqrt{\alpha}}+(\bar{z}-z)I\right)\left(\frac{A}{\sqrt{\alpha}}+(z-\bar{z})I\right)\right).$$

Because $\frac{A}{\sqrt{\alpha}}$ is Hermitian, it can be diagonalized to $L=\operatorname{diag}(\lambda_1,\dots,\lambda_n)$, and thus the above determinant becomes:

$$\begin{aligned}
&\det\left(w\left(\frac{A^2}{\alpha}-(z+\bar{z})\frac{A}{\sqrt{\alpha}}-wI\right)+\left(\frac{A}{\sqrt{\alpha}}+(\bar{z}-z)I\right)\left(\frac{A}{\sqrt{\alpha}}+(z-\bar{z})I\right)\right)\\
&=\det\left(w\left(L^2-(z+\bar{z})L-wI\right)+\left(L+(\bar{z}-z)I\right)\left(L+(z-\bar{z})I\right)\right)\\
&=\prod_{i=1}^{n}\left(w\left(\lambda_i^2-(z+\bar{z})\lambda_i-w\right)+\left(\lambda_i+(z-\bar{z})\right)\left(\lambda_i+(\bar{z}-z)\right)\right)\\
&=\prod_{i=1}^{n}\left(-w^2+w\left(\lambda_i^2-(z+\bar{z})\lambda_i\right)+\lambda_i^2-(z-\bar{z})^2\right).
\end{aligned}$$

The quadratic factors can then be explicitly factored, showing that each λ_i generates two singular values for $\widetilde{H}_0-zI$, each being the positive square root of

$$1+|z|^2+\frac{1}{2}\left(\lambda_i^2-(z+\bar{z})\lambda_i\right)\pm\frac{1}{2}\sqrt{\left(\lambda_i^2-(z+\bar{z})\lambda_i\right)^2+4(\lambda_i^2-(z-\bar{z})^2)}.$$

The proof of Lemma 3.9 is thus completed by Lemma 3.10 (stated and proved below), which shows that the quantity above is bounded from below by a positive constant depending only on z. ☐

Lemma 3.10. *Let z be a complex number satisfying $\mathsf{Im}(z)\neq 0$ and $|z|\neq 1$. Then for any real number λ, we have that*

$$1+|z|^2+\frac{1}{2}\left(\lambda^2-(z+\bar{z})\lambda\right)\pm\frac{1}{2}\sqrt{\left(\lambda^2-(z+\bar{z})\lambda\right)^2+4(\lambda^2-(z-\bar{z})^2)}\geq C_z, \tag{3.12}$$

where C_z is a positive real constant depending only on z.

The proof of Lemma 3.10 is given in Appendix A using elementary calculus, facts about matrices, and case analysis. Lemma 3.10 completes the proof of Lemma 3.5 and thus of Theorem 3.1.

4. Perturbation theory: proving Theorem 1.6

In this section, we study the eigenvalues of $\widetilde{H}$ via perturbation theory. Recall from (2.1) that $\widetilde{H}=\widetilde{H}_0+E$ where $E=\begin{pmatrix}0 & I+\frac{1}{\alpha}(I-D)\\ 0 & 0\end{pmatrix}$. Denote the eigenvalue of matrix M by $\mu(M)$. The spectral variation of $\widetilde{H}_0+E$ with respect to $\widetilde{H}_0$ is defined by

$$S_{\widetilde{H}_0}(\widetilde{H}_0+E)=\max_j\min_i|\mu_j(\widetilde{H}_0+E)-\mu_i(\widetilde{H}_0)|.$$

Theorem 4.1 (Bauer-Fike theorem; see Theorem 6 from [BLM15])**.** *If H_0 is diagonalizable by the matrix Y, then*

$$S_{H_0}(H_0+E)\leq\|E\|\cdot\|Y\|\cdot\|Y^{-1}\|.$$

Denote by $\mathcal{C}_i:=\mathcal{B}(\mu_i(H_0),R)$ the ball centered at $\mu_i(H_0)$ with radius $R=\|E\|\cdot\|Y\|\cdot\|Y^{-1}\|$. Let $\mathcal{I}$ be a set of indices such that

$$(\cup_{i\in\mathcal{I}}\mathcal{C}_i)\cap(\cup_{i\notin\mathcal{I}}\mathcal{C}_i)=\emptyset.$$

Then the number of eigenvalues of H_0+E in $\cup_{i\in\mathcal{I}}\mathcal{C}_i$ is exactly $|\mathcal{I}|$.

We will bound the operator norm of E and the condition number $\|Y\|\,\|Y^{-1}\|$ of Y to prove Theorem 1.6.

By Lemma 3.5, we know that $\|E\| \leq 20\sqrt{\frac{\log n}{np}}$ with probability 1 for all but finitely many n.

To bound the condition number of Y, we note that the square of the condition number of Y is equal to the largest eigenvalue of YY^* divided by the smallest eigenvalue of YY^*. Using the explicit definition of Y from (2.6), we see from the fact that the v_i are orthonormal that

$$YY^* = \operatorname{diag}(Y_1, \ldots, Y_n)$$

where Y_i's are 2×2 block matrices of the following form

$$Y_i = \begin{pmatrix} y_{2i-1}^* y_{2i-1} & y_{2i-1}^* y_{2i} \\ y_{2i}^* y_{2i-1} & y_{2i}^* y_{2i} \end{pmatrix}.$$

Recall that

$$y_{2i-1}^* = \frac{1}{\sqrt{1+|\mu_{2i-1}|^2}} \begin{pmatrix} -\mu_{2i-1} v_i^T & v_i^T \end{pmatrix} \quad \text{and} \quad y_{2i}^* = \frac{1}{\sqrt{1+|\mu_{2i}|^2}} \begin{pmatrix} -\mu_{2i} v_i^T & v_i^T \end{pmatrix}$$

We then have $Y_i = \begin{pmatrix} 1 & \gamma_i \\ \overline{\gamma}_i & 1 \end{pmatrix}$ where

$$\gamma_i := \frac{\mu_{2i-1}\overline{\mu}_{2i} + 1}{\sqrt{(1+|\mu_{2i-1}|^2)(1+|\mu_{2i}|^2)}}.$$

It is easy to check that the eigenvalues of Y_i are $1 \pm |\gamma_i|$. The eigenvalues of YY^* are the union of all the eigenvalues of the blocks, and so we will compute the eigenvalues $1 \pm |\gamma_i|$ based on whether λ_i produced real or complex eigenvalues for $\widetilde{H}_0$.

For $i = 1$, the eigenvalue λ_1 produces two real eigenvalues for $\widetilde{H}_0$. Using the general facts that $\mu_{2i-1}\mu_{2i} = 1$ and $\mu_{2i-1} + \mu_{2i} = \lambda_i$, which together imply that $\mu_{2i-1}^2 + \mu_{2i}^2 = \lambda_i^2 - 2$, we see that in this case $\gamma_1^2 = \frac{4}{\lambda_1^2}$, and so the two eigenvalues corresponding to this block are $1 \pm |\gamma_i| = 1 \pm 2/|\lambda_i|$. By (2.4), we see that $1 \pm |\gamma_i| = 1 \pm \frac{2}{\sqrt{np}}(1 + o(1))$ with probability $1 - o(1)$.

For $i \geq 2$, the eigenvalue λ_i produces two complex eigenvalues for $\widetilde{H}_0$, both with absolute value 1 (see Section 2). In this case, $\gamma_i = \frac{1+\mu_{2i-1}^2}{2}$. Again using the facts that $\mu_{2i-1}\mu_{2i} = 1$ and $\mu_{2i-1}^2 + \mu_{2i}^2 = \lambda_i^2 - 2$, we see that $\overline{\gamma}_i \gamma_i = \lambda_i^2/4$, which shows that the two eigenvalues corresponding to this block are $1 \pm |\lambda_i|/2$.

By [Vu07] (see Theorem 2.1 in Section 2) we know that when $p \geq \frac{\log^{2/3+\varepsilon} n}{n^{1/6}}$, $\max_{2 \leq i \leq n} |\lambda_i| \leq 2\sqrt{1-p} + O(n^{1/4}\log n/\sqrt{np})$ with probability tending to 1, and thus the largest and smallest eigenvalues coming from any of the blocks corresponding to $i \geq 2$ are $1 + \sqrt{1-p} + O(n^{1/4}\log n/\sqrt{np})$ and $1 - \sqrt{1-p} + O(n^{1/4}\log n/\sqrt{np})$ with probability tending to 1. Combining this information with the previous paragraph, we see that the condition number for Y is

$$\sqrt{\frac{1+\sqrt{1-p}+O(n^{1/4}\log n/\sqrt{np})}{1-\sqrt{1-p}+O(n^{1/4}\log n/\sqrt{np})}} = \sqrt{\frac{(1+\sqrt{1-p})^2 + O(n^{-1/4}p^{-1/2}\log n)}{p + O(n^{-1/4}p^{-1/2}\log n)}}$$

$$\leq \sqrt{\frac{5}{p}}\sqrt{\frac{1}{1+O(n^{-1/4}p^{-3/2}\log n)}} \leq \frac{2}{\sqrt{p}}.$$

In the first inequality above, we use that $n^{-1/4}p^{-3/2}\log n \leq \log^{-3/2\varepsilon} n = o(1)$ since $p \geq \frac{\log^{2/3+\varepsilon} n}{n^{1/6}}$. For the second inequality, we use the Taylor expansion.

Finally, we apply Lemma 3.5 and Bauer-Fike Theorem (Theorem 4.1) with $R = \frac{40}{\sqrt{p}}\sqrt{\frac{\log n}{np}} = 40\sqrt{\frac{\log n}{np^2}}$ to complete the proof.

Appendix A. Proof of Lemma 3.10

It is sufficient to show that the left-hand side of (3.12) (replacing $\pm$ with $-$) is bounded below by a positive constant $C_z > 0$ depending only on z. Substituting $z = a + ib$ where $a, b \in \mathbb{R}$, we see that the left-hand side of (3.12) is bounded below by

$$\begin{aligned} 1 + a^2 + b^2 &+ \frac{1}{2}\left(\lambda^2 - 2a\lambda\right) - \frac{1}{2}\sqrt{(\lambda^2 - 2a\lambda)^2 + 4(\lambda^2 + 4b^2)} \\ &= 1 + a^2 + b^2 + \frac{\lambda^2}{2} - a\lambda - \sqrt{\left(\frac{\lambda^2 - 2a\lambda}{2}\right)^2 + \lambda^2 + 4b^2} \end{aligned} \tag{A.1}$$

Note that the quantity in (A.1) is always at least zero because it is a singular value for a matrix.

Define a function g by

$$g(\lambda, a, \gamma) = 1 + a^2 + \gamma + \frac{\lambda^2}{2} - a\lambda - \sqrt{\left(\frac{\lambda^2 - 2a\lambda}{2}\right)^2 + \lambda^2 + 4\gamma} \tag{A.2}$$

(where we replaced b^2 by γ in the quantity (A.1)). It is sufficient to complete the proof if we show that for all real a and γ satisfying $\gamma > 0$ and $a^2 + \gamma \neq 1$ that $g(\lambda, a, \gamma) \geq C_{a,\gamma} > 0$, where $C_{a,\gamma}$ is a constant depending only on a and γ. We will prove this by considering three cases and using calculus.

Case I: $\left(\frac{\lambda^2 - 2a\lambda}{2}\right)^2 + \lambda^2 \geq 4$. In this case we will minimize g over $\gamma > 0$. Note that

$$\frac{\partial}{\partial \gamma} g = 1 - \frac{2}{\sqrt{\left(\frac{\lambda^2 - 2a\lambda}{2}\right)^2 + \lambda^2 + 4\gamma}}.$$

The quantity in the denominator of the fraction above is always greater than 2 since $\left(\frac{\lambda^2 - 2a\lambda}{2}\right)^2 + \lambda^2 \geq 4$ and $\gamma > 0$ by assumption; thus the derivative $\frac{\partial}{\partial \gamma} g$ is positive for all γ, and hence g (viewed as a function of γ) is strictly increasing as γ increases, for any a and any $|\lambda| \geq 2$. Also, note that

$$\frac{\partial^2}{\partial \gamma^2} g = \frac{4}{\left(\left(\frac{\lambda^2 - 2a\lambda}{2}\right)^2 + \lambda^2 + 4\gamma\right)^{3/2}},$$

which is always strictly positive, showing that g is concave upwards as a function of γ, for any a and any $|\lambda| \geq 2$. By the above we know that $\frac{\partial}{\partial \gamma} g(\lambda, a, \gamma/2) \geq 1 - \frac{2}{\sqrt{4+4\gamma}}$, and we also know that $g(\lambda, a, \gamma) \geq 0$ for all λ, a, γ (because it is a singular value); thus, using the fact that the tangent line at $\gamma/2$ is strictly below g (since it is concave upwards), we see that $g(\lambda, a, \gamma) \geq \left(1 - \frac{2}{\sqrt{4+4\gamma}}\right)\frac{\gamma}{2}$, which is a positive constant depending only on $\gamma > 0$, completing the proof in Case I.

Case II: $\left(\frac{\lambda^2 - 2a\lambda}{2}\right)^2 + \lambda^2 < 4$ (which implies $|\lambda| < 2$) and $\left|1 - \frac{\lambda^2}{4} - \gamma\right| \geq \frac{1}{2}\left|a^2 + \gamma - 1\right|$. In this case, we will minimize g over a. Recalling the definition of g in (A.2), we see that

$$\frac{\partial}{\partial a} g = 2a - \lambda - \frac{2\left(\frac{\lambda^2 - 2a\lambda}{2}\right)(-\lambda)}{2\sqrt{\left(\frac{\lambda^2 - 2a\lambda}{2}\right)^2 + \lambda^2 + 4\gamma}}$$

$$= (2a-\lambda)\left(1-\frac{\lambda^2}{2\sqrt{\left(\frac{\lambda^2-2a\lambda}{2}\right)^2+\lambda^2+4\gamma}}\right).$$

Because $|\lambda|<2$, we see that $\frac{\lambda^2}{2\sqrt{\left(\frac{\lambda^2-2a\lambda}{2}\right)^2+\lambda^2+4\gamma}} < \frac{|\lambda|}{\sqrt{\left(\frac{\lambda^2-2a\lambda}{2}\right)^2+\lambda^2+4\gamma}} \le 1$, which shows that $\frac{\partial}{\partial a}g$ has the same sign as $2a-\lambda$ and thus achieves a minimum when $a=\lambda/2$. Thus, using the minimum of g over a in this case, we have

$$g(\lambda,a,\gamma)\ge g(\lambda,\lambda/2,\gamma)=1+\frac{\lambda^2}{4}+\gamma-\sqrt{\lambda^2+4\gamma}.$$

Substituting $x=\sqrt{\frac{\lambda^2}{4}+\gamma}$ into the right-hand side, we see that g is bounded below by the quantity $1+x^2-2x=(x-1)^2$ for positive x satisfying $|x^2-1|\ge\frac{1}{2}|a^2+\gamma-1|$ (which follows from the assumptions on a and γ in this case). We will complete the argument by proving the claim that $|x-1|\ge\min\{1,\frac{1}{6}|a^2+\gamma-1|\}$, which is a positive constant depending only on a and γ. To prove this claim, note that if $|x|\ge 2$, then $|x-1|\ge 1$; and on the other hand, if $|x|<2$, then $|x-1|=\frac{|x^2-1|}{|x+1|}>\frac{1}{3}|x^2-1|\ge\frac{1}{6}|a^2+\gamma-1|$. This completes the proof in Case II.

Case III: $\left(\frac{\lambda^2-2a\lambda}{2}\right)^2+\lambda^2<4$ and $\left|1-\frac{\lambda^2}{4}-\gamma\right|<\frac{1}{2}|a^2+\gamma-1|$. In this case, we will first minimize over γ. Note that

$$\frac{\partial}{\partial\gamma}g=1-\frac{2}{\sqrt{\left(\frac{\lambda^2-2a\lambda}{2}\right)^2+\lambda^2+4\gamma}}.$$

Because $\left(\frac{\lambda^2-2a\lambda}{2}\right)^2+\lambda^2<4$, we see from $\frac{\partial}{\partial\gamma}g$ that g has a minimum when $\gamma=\gamma_0(\lambda,a):=1-\frac{\lambda^2}{4}-\left(\frac{\lambda^2-2a\lambda}{4}\right)^2$. Thus, $g(\lambda,a,\gamma)\ge g(\lambda,a,\gamma_0(a,\lambda))$. Thus, we can find a general lower bound on g in this case by demonstrating a lower bound on

$$\begin{aligned}
g(\lambda,a,\gamma_0(a,\lambda)) &= 1+a^2+1-\frac{\lambda^2}{4}-\left(\frac{\lambda^2-2a\lambda}{4}\right)^2+\frac{\lambda^2}{2}-a\lambda-2\\
&= a^2+\frac{\lambda^2}{4}-\left(\frac{\lambda^2-2a\lambda}{4}\right)^2-a\lambda\\
&= a^2-a\lambda+\frac{\lambda^2}{4}-\frac{1}{16}\left(\lambda^4-4a\lambda^3+4a^2\lambda^2\right)\\
&= a^2\left(1-\frac{\lambda^2}{4}\right)-a\lambda\left(1-\frac{\lambda^2}{4}\right)+\frac{\lambda^2}{4}\left(1-\frac{\lambda^2}{4}\right)\\
&= \left(1-\frac{\lambda^2}{4}\right)\left(a^2-a\lambda+\frac{\lambda^2}{4}\right)\\
&= \left(1-\frac{\lambda^2}{4}\right)\left(a-\frac{\lambda}{2}\right)^2 &\text{(A.3)}\\
&> \frac{\lambda^4}{4}\left(a-\frac{\lambda}{2}\right)^4, &\text{(A.4)}
\end{aligned}$$

where the last inequality used $1-\lambda^2/4>\frac{\lambda^2}{4}(a-\lambda/2)^2$, which is equivalent to one of the Case III assumptions.

We know that $|\lambda|<2$ by assumption. Therefore, $1-\lambda^2/4$ is positive, and thus, the minimum of $g(\lambda,a,\gamma_0(a,\lambda))$ will occur when a is as close as possible to $\lambda/2$. We claim that, given the Case III assumptions, we have $|a-\lambda/2|\ge c_2:=-1+\sqrt{1+\frac{1}{2}|a^2+\gamma-1|}>0$, a constant depending only on a and γ. To prove the claim, assume for a contradiction that $|a-\lambda/2| < c_2$. This implies that $|a| < |\lambda|/2 + c_2$, and so we have

$$\begin{aligned}\left|a^2+\gamma-1\right| &\le \left|\left(\frac{|\lambda|}{2}+c_2\right)^2+\gamma-1\right| = \left|\frac{|\lambda|^2}{4}+c_2|\lambda|+c_2^2+\gamma-1\right| \\ &\le \left|\frac{|\lambda|^2}{4}+\gamma-1\right| + \left|c_2|\lambda|+c_2^2\right| \\ &< \frac{1}{2}\left|a^2+\gamma-1\right| + 2c_2 + c_2^2,\end{aligned}$$

where the last inequality used the assumptions of Case III. The chain of inequalities above implies that $\frac{1}{2}|a^2+\gamma-1|+1 < (1+c_2)^2$, which is a contradiction when combined with the definition of c_2. This proves the claim that $|a-\lambda/2| \ge c_2$.

To complete the proof of Case III, it is sufficient to show that either $(1-\lambda^2/4)c_2^2$ (from (A.3)) or $\lambda^4 c_2^4/4$ (from (A.4)) is bounded from below. This is easily done: if $|\lambda| < 1/2$, then $(1-\lambda^2/4)c_2^2 > \frac{15}{16}c_2^2$; and if $|\lambda| \ge 1/2$, then $\lambda^4 c_2^4/4 > c_2^4/64$. In both cases, the lower bound is a constant depending only on a and γ, which completes the proof of Case III and hence also the proof of Lemma 3.10.

Acknowledgment: The authors would like to thank Zhigang Bao for useful discussions.

References

[AFH15] Omer Angel, Joel Friedman, and Shlomo Hoory. The non-backtracking spectrum of the universal cover of a graph. *Transactions of the American Mathematical Society*, 367(6):4287–4318, 2015.

[Arn67] Ludwig Arnold. On the asymptotic distribution of the eigenvalues of random matrices. *J. Math. Anal. Appl.*, 20:262–268, 1967.

[Bas92] Hyman Bass. The ihara-selberg zeta function of a tree lattice. *International Journal of Mathematics*, 3(06):717–797, 1992.

[BGBK17] Florent Benaych-Georges, Charles Bordenave, and Antti Knowles. Spectral radii of sparse random matrices. *arXiv preprint arXiv:1704.02945*, 2017.

[BGK16] Florent Benaych-Georges and Antti Knowles. Lectures on the local semicircle law for wigner matrices. *arXiv preprint arXiv:1601.04055*, 2016.

[Bor15] Charles Bordenave. A new proof of Friedman's second eigenvalue Theorem and its extension to random lifts. *Annales Scientifiques de l cole Normale Suprieure* 53(6), 2015.

[BLM15] Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Non-backtracking spectrum of random graphs: community detection and non-regular ramanujan graphs. *arXiv preprint arXiv:1501.06087*, 2015.

[Cha09] Djalil Chafaı. Singular values of random matrices. *Lecture Notes*, 2009.

[Che52] Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statistics*, 23:493–507, 1952.

[CZ21] Simon Coste and Yizhe Zhu. Eigenvalues of the non-backtracking operator detached from the bulk. *Random Matrices: Theory and Applications*, 10(03):2150028, 2021.

[EKYY13] László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin. Spectral statistics of Erdős-Rényi graphs I: Local semicircle law. *Ann. Probab.*, 41(3B):2279–2375, 2013.

[FK81] Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981.

[GLM16] Lennart Gulikers, Marc Lelarge, and Laurent Massoulié. Non-backtracking spectrum of degree-corrected stochastic block models. *arXiv preprint arXiv:1609.02487*, 2016.

[Gre63] Ulf Grenander. *Probabilities on algebraic structures.* John Wiley & Sons, Inc., New York-London; Almqvist & Wiksell, Stockholm-Göteborg-Uppsala, 1963.

[Has89a] Ki-ichiro Hashimoto. Zeta functions of finite graphs and representations of p-adic groups. In *Automorphic forms and geometry of arithmetic varieties*, volume 15 of *Adv. Stud. Pure Math.*, pages 211–280. Academic Press, Boston, MA, 1989.

[Has89b] Ki-ichiro Hashimoto. Zeta functions of finite graphs and representations of p-adic groups. *Automorphic forms and geometry of arithmetic varieties.*, pages 211–280, 1989.

[HK21] Yukun He and Antti Knowles. Fluctuations of extreme eigenvalues of sparse Erdős–Rényi graphs. *Probability Theory and Related Fields*, 1–72, 2021.

[HLL83] Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: first steps. *Social Networks*, 5(2):109–137, 1983.

[Hoe63] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13–30, 1963.

[HLY20] Jiaoyang Huang and Benjamin Landon and Horng-Tzer Yau. Transition from Tracy–Widom to Gaussian fluctuations of extremal eigenvalues of sparse Erdős–Rényi graphs. *The Annals of Probability*, 48(2):916–962, 2020.

[Ivc73] Grigorii Ivanovich Ivchenko. On the asymptotic behavior of degrees of vertices in a random graph. *Theory of Probability & Its Applications*, 18(1):188–195, 1973.

[KMM$^+$13] Florent Krzakala, Cristopher Moore, Elchanan Mossel, Joe Neeman, Allan Sly, Lenka Zdeborová, and Pan Zhang. Spectral redemption in clustering sparse networks. *Proc. Natl. Acad. Sci. USA*, 110(52):20935–20940, 2013.

[KN11] Brian Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Phys. Rev. E (3)*, 83(1):016107, 10, 2011.

[KP93] A. M. Khorunzhy and L. A. Pastur. Limits of infinite interaction radius, dimensionality and the number of components for random operators with off-diagonal randomness. *Comm. Math. Phys.*, 153(3):605–646, 1993.

[KS03] Michael Krivelevich and Benny Sudakov. The largest eigenvalue of sparse random graphs. *Combin. Probab. Comput.*, 12(1):61–72, 2003.

[LS18] Ji Oon Lee and Kevin Schnelli. Local law and Tracy–Widom limit for sparse random matrices. *Probability Theory and Related Fields*, 171(1):543–616, 2018.

[LvHY18] Latała, Rafał and van Handel, Ramon and Youssef, Pierre. The dimension-free structure of nonhomogeneous random matrices. *Inventiones mathematicae*, 214(3):1031–1080, 2018.

[Mas14] Laurent Massoulié. Community detection thresholds and the weak Ramanujan property. In *STOC'14—Proceedings of the 2014 ACM Symposium on Theory of Computing*, pages 694–703. ACM, New York, 2014.

[MNS13] Elchanan Mossel, Joe Neeman, and Allan Sly. A proof of the block model threshold conjecture. *arXiv preprint arXiv:1311.4115*, 2013.

[MNS15] Elchanan Mossel, Joe Neeman, and Allan Sly. Reconstruction and estimation in the planted partition model. *Probab. Theory Related Fields*, 162(3-4):431–461, 2015.

[Sil00] John R Silvester. Determinants of block matrices. *The Mathematical Gazette*, 84(501):460–467, 2000.

[TV10] Terence Tao and Van Vu. Random matrices: universality of ESDs and the circular law. *Ann. Probab.*, 38(5):2023–2065, 2010. With an appendix by Manjunath Krishnapur.

[TVW13] Linh V. Tran, Van H. Vu, and Ke Wang. Sparse random graphs: eigenvalues and eigenvectors. *Random Structures Algorithms*, 42(1):110–134, 2013.

[Vu07] Van Vu. *Spectral norm of random matrices. Combinatorica*, 27(6):721–736, 2007.

[Vu08] Van Vu. *Random Discrete Matrices*, pages 257–280. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[Wig55] Eugene P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math. (2)*, 62:548–564, 1955.

[Wig58] Eugene P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)*, 67:325–327, 1958.

[Woo12] Philip Matchett Wood. Universality and the circular law for sparse random matrices. *Ann. Appl. Probab.*, 22(3):1266–1300, 2012.

[Woo16] Philip Matchett Wood. Universality of the ESD for a fixed matrix plus small random noise: a stability approach. *Ann. Inst. Henri Poincaré Probab. Stat.*, 52(4):1877–1896, 2016.

Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
E-mail address: kewang@ust.hk

Department of Mathematics, Harvard University, 1 Oxford Street, Cambridge, MA 02138
E-mail address: pmwood@math.harvard.edu