

EXTREMAL EIGENVALUES OF CRITICAL ERDŐS–RÉNYI GRAPHS

BY JOHANNES ALT^{*}, RAPHAËL DUCATEZ[†] AND ANTTI KNOWLES[‡]

Section of Mathematics, University of Geneva, ^{*}johannes.alt@unige.ch; [†]raphael.ducatez@unige.ch;
[‡]antti.knowles@unige.ch

We complete the analysis of the extremal eigenvalues of the adjacency matrix A of the Erdős–Rényi graph $\mathbb{G}(N, d/N)$ in the critical regime $d \asymp \log N$ of the transition uncovered in (*Ann. Inst. Henri Poincaré Probab. Stat.* **56** (2020) 2141–2161; *Ann. Probab.* **47** (2019) 1653–1676), where the regimes $d \gg \log N$ and $d \ll \log N$ were studied. We establish a one-to-one correspondence between vertices of degree at least $2d$ and nontrivial (excluding the trivial top eigenvalue) eigenvalues of A/\sqrt{d} outside of the asymptotic bulk $[-2, 2]$. This correspondence implies that the transition characterized by the appearance of the eigenvalues outside of the asymptotic bulk takes place at the critical value $d = d_* = \frac{1}{\log 4 - 1} \log N$. For $d < d_*$, we obtain rigidity bounds on the locations of all eigenvalues outside the interval $[-2, 2]$, and for $d > d_*$, we show that no such eigenvalues exist. All of our estimates are quantitative with polynomial error probabilities.

Our proof is based on a tridiagonal representation of the adjacency matrix and on a detailed analysis of the geometry of the neighbourhood of the large degree vertices. An important ingredient in our estimates is a matrix inequality obtained via the associated nonbacktracking matrix and an Ihara–Bass formula (*Ann. Inst. Henri Poincaré Probab. Stat.* **56** (2020) 2141–2161). Our argument also applies to sparse Wigner matrices, defined as the Hadamard product of A and a Wigner matrix, in which case the role of the degrees is replaced by the squares of the ℓ^2 -norms of the rows.

CONTENTS

1. Introduction	1348
2. Results	1351
3. Notation	1354
Notation related to vectors and matrices	1354
Notation related to graphs	1354
Probabilistic notation and tools	1355
4. Main ideas of the proof	1356
5. Large eigenvalues induced by vertices of large degree	1359
5.1. Decomposition of $(\underline{A} - \sqrt{d}\Lambda(\alpha_x))\mathbf{v}$ —Proof of Lemma 5.2	1362
5.2. Smallness of $\mathbf{w}_0, \dots, \mathbf{w}_4$ —Proof of Lemma 5.3	1363
6. Quadratic form estimates on centred adjacency matrix	1371
6.1. Proof of Proposition 6.1	1372
7. Lower bounds on large eigenvalues	1373
7.1. Proof of Lemma 7.2	1376
8. Upper bounds on large eigenvalues	1381
8.1. Proof of Lemma 8.3	1384
8.2. Proof of Proposition 8.4	1387
9. Proofs of the results in Section 2	1389
10. Modifications for sparse Wigner matrices	1390
10.1. Modifications in Section 5	1390
Definition of the approximate eigenvector, decomposition of the error terms	1390

Received May 2019; revised May 2020.

MSC2020 subject classifications. 60B20, 15B52, 05C80.

Key words and phrases. Erdős–Rényi graphs, sparse random matrix, eigenvalue rigidity.

Concentration of $\|g_{i+1}\|/\|g_i\|$ 1391
 Estimate on w_1 1391
 Estimate on w_2 1392
 Appendix A: Tridiagonalization 1393
 Appendix B: Tridiagonal matrix associated with a regular tree 1394
 Appendix C: Spectral properties of tridiagonal matrices 1395
 Appendix D: Degree distribution of the Erdős–Rényi graph 1399
 Acknowledgments 1401
 References 1401

1. Introduction. This paper is about the extremal eigenvalues of sparse random matrices, such as the adjacency matrix of the Erdős–Rényi graph. In spectral graph theory, obtaining precise bounds on the locations of the extreme eigenvalues, in particular on the spectral gap, is of fundamental importance and has attracted much attention in the past 30 years; see, for instance, [1, 6, 11] for reviews.

The Erdős–Rényi graph $\mathbb{G} = \mathbb{G}(N, d/N)$ is the simplest model of a random graph, where each edge of the complete graph on N vertices is kept independently with probability d/N , with $0 < d < N$. Its adjacency matrix A is the canonical example of a sparse random matrix, and its spectrum has been extensively studied in the random matrix theory literature. In the regime $d \equiv d_N \rightarrow \infty$ as $N \rightarrow \infty$, the empirical eigenvalue measure of A/\sqrt{d} converges to the semicircle law supported on $[-2, 2]$ [18, 21].

The behaviour of the extremal eigenvalues is more subtle and has been investigated in several recent works [3, 4, 7–10, 12–15, 20]. In particular, in [13] it is shown that the largest eigenvalue $\lambda_1(A)$ of A is asymptotically equivalent to the maximum of d and the square root of the largest degree of \mathbb{G} . A more difficult question is that of the other eigenvalues, $\lambda_2(A), \dots, \lambda_N(A)$, which determine in particular the gap $\lambda_1(A) - \lambda_2(A)$ between the largest and second-largest eigenvalues. By a standard eigenvalue interlacing argument, the analysis of the extremal eigenvalues $\lambda_2(A), \dots, \lambda_N(A)$ of A is equivalent to the analysis of the eigenvalues $\lambda_1(\underline{A}), \dots, \lambda_N(\underline{A})$ of the centred adjacency matrix $\underline{A} := A - \mathbb{E}A$.

An important motivation for the present work is a transition in the behaviour of the extremal eigenvalues of \underline{A} observed in [3, 4]. In [4] it is shown that in the regime $d \gg \log N$ the extremal eigenvalues $\lambda_1(\underline{A}/\sqrt{d})$ and $\lambda_N(\underline{A}/\sqrt{d})$ converge with high probability to the edges $+2$ and -2 of the semicircle law’s support. Conversely, in [3] it is shown that, in the regime $d \ll \log N$, the extremal eigenvalues $\lambda_1(\underline{A}/\sqrt{d})$ and $\lambda_N(\underline{A}/\sqrt{d})$ are asymptotically of order $\pm\sqrt{\eta/\log \eta}$ with $\eta := \frac{\log N}{d}$, placing them far outside of the interval $[-2, 2]$.

Based on the two different behaviours observed in [3, 4] we, therefore, expect a transition in the behaviour of the extremal eigenvalues on the *critical* scale $d \asymp \log N$, where the extremal eigenvalues leave the support of the semicircle law.

In this paper we give a detailed analysis of this transition around the critical scale $d \asymp \log N$, which was left open by the works [3, 4], by deriving quantitative high-probability bounds on the locations of all eigenvalues of A/\sqrt{d} and \underline{A}/\sqrt{d} that lie outside the interval $[-2, 2]$. Our analysis also covers the neighbouring sub- and supercritical regimes, $d \ll \log N$ and $d \gg \log N$ and, in particular, provides a complete picture of the transition between these two regimes. Our approach also works for sparse Wigner matrices of the form $M = (M_{xy})_{x,y=1}^N$, where $M_{xy} = W_{xy}A_{xy}$ and $(W_{xy} : x \leq y)$ are uniformly bounded independent random variables with zero expectation and unit variance.

We remark that the critical scale $d \asymp \log N$ is the same as the well-known connectivity threshold for the Erdős–Rényi which happens precisely at the value $d = \log N$. In contrast, although the transition in the locations of the extremal eigenvalues of \underline{A} happens on the same scale $d \asymp \log N$, it happens at a different numerical value, $d = b_* \log N$, where

$b_* := \frac{1}{\log 4 - 1} \approx 2.59$. The underlying cause is the same for both transitions—the lack of concentration of the degree sequence which yields isolated vertices on the one hand and vertices of large degree on the other hand.

Indeed, the mechanism underlying the emergence of eigenvalues outside the support of the semicircle distribution for sufficiently sparse matrices is the appearance of vertices of large degree. This was already observed and exploited in [3] in the subcritical regime $d \ll \log N$. The intuition is that, for sufficiently small d , the concentration of the degrees of the vertices around their mean d fails, and we observe a number of vertices whose degree is much larger than d . This mechanism is also at the heart of our analysis. In fact, our main result is a high-probability correspondence between vertices of large degree and extremal eigenvalues. Roughly, we show that the following holds with probability at least $1 - N^{-\nu}$ for any fixed $\nu > 0$:

- (i) Every vertex x with degree D_x larger than $(2 + o(1))d$ gives rise to exactly one eigenvalue of \underline{A}/\sqrt{d} in $[2 + o(1), \infty)$ and one in $(-\infty, -2 - o(1)]$. These eigenvalues are located near $\pm\Lambda(D_x/d)$, respectively, where $\Lambda(t) := \frac{t}{\sqrt{t-1}}$. The error is bounded by an inverse power of d .
- (ii) There are no other eigenvalues in $(-\infty, -2 - o(1)] \cup [2 + o(1), \infty)$.

Using standard results on the degree distribution of the Erdős–Rényi graph (for the reader’s convenience we review the necessary results in Appendix D), we can then easily conclude rigidity estimates for all eigenvalues of \underline{A}/\sqrt{d} and A/\sqrt{d} in the region $\mathbb{R} \setminus [-2 - o(1), 2 + o(1)]$. Setting $d = b \log N$ for a fixed $b < b_*$, one can check (see Remark 2.6 below) that, with high probability, there are $N^{1-b/b_*+o(1)}$ such eigenvalues.

Our proof is based on the tridiagonal representation [19] of the matrix \underline{A} around some vertex x . Thus, for any vertex $x \in [N]$, we consider the unit vector $\mathbf{1}_x$ supported at x and rewrite \underline{A} in the basis obtained by orthonormalizing the vectors $\mathbf{1}_x, \underline{A}\mathbf{1}_x, \underline{A}^2\mathbf{1}_x, \dots$. The resulting matrix Z is tridiagonal, and its spectrum coincides with that of \underline{A} . Denoting by $S_i(x)$ the sphere of radius i around x , the key intuition behind our proof is that, even though $D_x = |S_1(x)|$ does not concentrate in the critical and subcritical regimes, the *quotients* $|S_{i+1}(x)|/|S_i(x)|$, $i \geq 1$, do. Moreover, we note that balls of sufficiently small radius have only a bounded number of cycles with high probability and can, therefore, be approximated by trees after a removal of a bounded number of edges. Thus, we expect the tridiagonal matrix Z to be close to that of a tree whose root x has D_x children and all other vertices d children (see (4.2) and Figure 2 below). The spectrum of this latter matrix may be analysed using transfer matrix methods. We remark that this approximation requires precise information about the geometry of the neighbourhoods of vertices and is only correct for vertices of large enough degree.

In practice, we proceed as follows. For clarity, let us focus only on the positive eigenvalues. Our proof then consists of two major steps: deriving lower and upper bounds on the extremal eigenvalues of \underline{A} . For the lower bounds we construct approximate eigenvectors $\mathbf{v}(x)$ of \underline{A} around vertices x of high degree, whose definition is motivated by the fact that $\mathbf{v}(x)$ would be an exact eigenvector if the approximation by a regular tree sketched above were exact. In addition to showing that these $\mathbf{v}(x)$ are indeed approximate eigenvectors with a quantitatively controlled error bound, we need to show that all of the associated eigenvalues in $[2 + o(1), \infty)$ are distinct. We do this by a careful pruning of the graph, with the property that all balls (in the pruned graph) of suitable radii around the vertices $\mathcal{V}_2 := \{x \in [N] : D_x \geq 2d\}$ are disjoint and that the degrees of the difference between the original and pruned graphs are not too large. Since $\mathbf{v}(x)$ is supported in a sufficiently small ball around x , this will imply that the family $(\mathbf{v}(x))_{x \in \mathcal{V}_2}$ is orthogonal and, hence, the associated eigenvalues of \underline{A}/\sqrt{d} are distinct.

For the matching upper bounds on the extremal eigenvalues, a fundamental input is an Ihara–Bass type formula and a bound on the spectral radius of the *nonbacktracking matrix*

associated with \underline{A} derived in [4]. This argument allows us to completely bypass typically very complicated combinatorial arguments needed in the moment method for estimating matrix norms. Thanks to the Ihara–Bass formula, the moment method is performed only on the level of the nonbacktracking matrix; this was already performed in [4] using a moment method that was very simple thanks to the nonbacktracking property. In particular, the lack of concentration of the degrees, which has a crucial impact on the extremal eigenvalues of \underline{A} , has no impact on the extremal eigenvalues (in absolute value) of the nonbacktracking matrix of \underline{A} . The outcome of this observation is the matrix inequality $\underline{A}/\sqrt{d} \leq I_N + Q + o(1)$, where Q is the diagonal matrix with entries D_x/d . We apply this inequality to estimate the norm of the matrix \underline{A}/\sqrt{d} restricted to vertices with degrees at most $2d$ and show that it is bounded by $2 + o(1)$. To that end, we need to derive, for the maximal eigenvector of the restricted matrix, a delocalization bound at vertices with degree at least $(1 + o(1))d$. This delocalization bound is derived using a careful analysis of the tridiagonal matrix associated with the restricted adjacency matrix. In fact, all of this analysis has to be done with the pruned adjacency matrix described above in order to obtain simultaneous upper bounds on all eigenvalues down to $2 + o(1)$. We refer to Section 4 below for a more detailed summary of the proof.

The argument sketched above can also be easily applied to the sparse Wigner matrices M described above, essentially by replacing the degree D_x of a vertex by the square ℓ^2 -norm of the x th row of M . The details are explained in Section 10 below.

Our method is rather general and, in particular, it is not tied to the homogeneity of the Erdős–Rényi graph. We, therefore, expect it to be applicable to many other sparse random matrix models, such as inhomogeneous Erdős–Rényi graphs and stochastic block models.

We remark that a related result for the eigenvalues induced by the top degree D_{\max} appeared in the independent work [17] while we were finalizing the current manuscript. In [17] the authors show that, for any fixed $l \in \mathbb{N}$, the largest / smallest l eigenvalues of \underline{A} are, with high probability, equal to $\pm(1 + o(1))\Lambda(D_{\max}/d \vee 2)\sqrt{d}$. This corresponds to a qualitative version of our main result restricted to the top $O(1)$ eigenvalues, which all have the same asymptotic value. In this paper we obtain quantitative rigidity bounds for all eigenvalues in $\mathbb{R} \setminus [-2 - o(1), 2 + o(1)]$. For $d = b \log N$ with some fixed $b < b_*$, there are with high probability $N^{1-b/b_*+o(1)}$ such eigenvalues. The precise location $d = b_* \log N$ for the transition in the behaviour of the top eigenvalues of the Erdős–Rényi graph was also established in [17]. Their argument also works for sparse Wigner matrices M described above. The proof of [17] differs substantially from ours; it relies on suitably chosen trial vectors and an intricate moment method argument controlled using cleverly constructed data structures.

In [2] the eigenvectors of A/\sqrt{d} corresponding to eigenvalues in $\mathbb{R} \setminus [-2 - o(1), 2 + o(1)]$ are shown to be essentially localized on a small number of vertices. In contrast, the eigenvectors of A/\sqrt{d} in the bulk spectrum $[-2 + o(1), 2 - o(1)]$ are completely delocalized [2].

We conclude the introduction with a brief outline of the paper. In Section 2 we state our results. The rest of the paper is devoted to the proofs. In Section 3 we introduce notation used throughout the paper and in Section 4 give a more detailed summary of the proof. In Section 5 we show that a vertex of degree greater than $2d$ induces two approximate eigenvectors for the adjacency matrix. The subsequent Section 6 is devoted to a quadratic form bound on the adjacency matrix in terms of the degree matrix. Lower and upper bounds on large eigenvalues of the adjacency matrix are established in Section 7 and Section 8, respectively. In the short Section 9 we put everything together and conclude our main results for the Erdős–Rényi graph. In Section 10 we explain the minor changes required to handle sparse Wigner matrices. In the appendices we collect some basic results on tridiagonal matrices and the degree distribution of Erdős–Rényi graphs.

CONVENTION. We regard N as the fundamental large parameter. All quantities that are not explicitly *fixed* may depend on N ; we almost always omit the argument N from our notation.

2. Results. Let $A = (A_{xy})_{x,y \in [N]} \in \{0, 1\}^{N \times N}$ be the adjacency matrix of the homogeneous Erdős–Rényi graph with vertex set $[N] := \{1, \dots, N\}$ and edge probability d/N . That is, $A = A^*$, $A_{xx} = 0$ for all $x \in [N]$, and $(A_{xy} : x < y)$ are independent Bernoulli(d/N) random variables. Throughout this paper, N is a large parameter, and $d \equiv d_N$ depends on N . For each $x \in [N]$, we define the *normalized degree* α_x of x through

$$(2.1) \quad \alpha_x := \frac{1}{d} \sum_{y \in [N]} A_{xy}.$$

We also consider the centred adjacency matrix $\underline{A} := A - \mathbb{E}A$. For any Hermitian matrix $M = M^* \in \mathbb{R}^{N \times N}$, we denote by $\lambda_1(M) \geq \lambda_2(M) \geq \dots \geq \lambda_N(M)$ its eigenvalues.

For $t \geq 2$, we define

$$(2.2) \quad \Lambda(t) := \frac{t}{\sqrt{t-1}}.$$

We denote by $\sigma : [N] \rightarrow [N]$ a random permutation such that

$$(2.3) \quad \alpha_{\sigma(1)} \geq \alpha_{\sigma(2)} \geq \dots \geq \alpha_{\sigma(N)}.$$

We can now state our main result.

THEOREM 2.1. *Fix $0 < \kappa < 1/2$. Suppose that $0 < \theta \leq 1/2$ and $(\log N)^{4/(5-2\theta)} \leq d \leq N/2$. Define the random index*

$$L := \max\{l \geq 1 : \alpha_{\sigma(l)} \geq 2 + (\log d)^{-\kappa}\}$$

with the convention that $L = 0$ if $\alpha_{\sigma(1)} < 2 + (\log d)^{-\kappa}$. Then, there is a universal constant $c > 0$ such that, for any $\nu > 0$, there is a constant $\mathcal{C} \equiv \mathcal{C}_{\nu, \kappa}$ such that the following hold with probability at least $1 - \mathcal{C}N^{-\nu}$:

(i) *For $1 \leq l \leq L$, we have*

$$\begin{aligned} & |\lambda_l(\underline{A}) - \sqrt{d}\Lambda(\alpha_{\sigma(l)})| + |\lambda_{N+1-l}(\underline{A}) + \sqrt{d}\Lambda(\alpha_{\sigma(l)})| \\ & \leq \mathcal{C}(d^{-c(\Lambda(\alpha_{\sigma(l)})-2)} + d^{-\theta/3})\sqrt{d}. \end{aligned}$$

(ii) *For $l = L + 1$, we have*

$$\max\{\lambda_l(\underline{A}), -\lambda_{N+1-l}(\underline{A})\} \leq (2 + \mathcal{C}(\log d)^{-2\kappa})\sqrt{d}.$$

REMARK 2.2.

(i) In the supercritical regime $d \gg \log N$, Theorem 2.1 is established in [3], and in the subcritical regime $d \ll \log N$ it is established in [4], in both cases with quantitative error bounds. Hence, in Theorem 2.1 it would be sufficient to assume that $d \asymp \log N$. We allow a larger range $d \geq (\log N)^{4/(5-2\theta)}$ so as to obtain a simple statement that extends to all three regimes, showcasing the full behaviour through the transition at criticality.

(ii) A simple analysis of the degrees shows that $L = 0$ in the supercritical regime, while L is a fractional power of N in the subcritical regime (see Appendix D).

As a consequence of Theorem 2.1, for any $\nu > 0$ there is a constant $C \equiv C_\nu > 0$ such that, with probability at least $1 - CN^{-\nu}$,

$$(2.4) \quad \|\underline{A}\| = \Lambda(\alpha_{\sigma(1)} \vee 2)(1 + o(1))\sqrt{d}.$$

Another easy consequence is the corresponding statement for the noncentred adjacency matrix A , which follows by eigenvalue interlacing.

COROLLARY 2.3. *Under the same conditions and notation as in Theorem 2.1, the following hold with probability at least $1 - CN^{-\nu}$:*

(i) For $1 \leq l \leq L$, we have

$$\begin{aligned} & |\lambda_{l+1}(A) - \sqrt{d}\Lambda(\alpha_{\sigma(l)})| + |\lambda_{N+1-l}(A) + \sqrt{d}\Lambda(\alpha_{\sigma(l)})| \\ & \leq C(d^{-c(\Lambda(\alpha_{\sigma(l)})-2)} + d^{-\theta/3} + (\alpha_{\sigma(l)} - \alpha_{\sigma(l+1)}))\sqrt{d}. \end{aligned}$$

(ii) For $l = L + 1$, we have

$$\max\{\lambda_{l+1}(A), -\lambda_{N+1-l}(A)\} \leq (2 + C(\log d)^{-2\kappa})\sqrt{d}.$$

Note that the additional error term $(\alpha_{\sigma(l)} - \alpha_{\sigma(l+1)})$ is of order $1/d$ with high probability (see Proposition D.1 below). It is well known that the largest eigenvalue $\lambda_1(A)$ is an outlier far outside the bulk spectrum; in fact, a trivial perturbation argument using (2.4) implies that $|\lambda_1(A) - d| \leq \Lambda(\alpha_{\sigma(1)} \vee 2)(1 + o(1))\sqrt{d}$ with probability at least $1 - CN^{-\nu}$, where $\nu > 0$ and $C \equiv C_\nu$.

Theorem 2.1 (and its noncentred counterpart) can be combined with a standard analysis of the distribution of the degree sequence $D_{\sigma(1)}, D_{\sigma(2)}, \dots$ of the Erdős–Rényi graph. For the convenience of the reader, in Appendix D we collect some basic results about the degree distribution. As an illustration, we state such an application for the extremal eigenvalues of A .

For its statement we need the following facts from Appendix D. For any $d > 0$ and $1 \leq l \leq \frac{N}{C\sqrt{d}}$, the equation

$$d(\beta \log \beta - \beta + 1) + \frac{1}{2} \log(2\pi\beta d) = \log(N/l)$$

has a unique solution $\beta_l(d)$ in $[1, \infty)$. (Here, C is a universal constant.) The interpretation of $\beta_l(d)$ is the typical value of the normalized degree $\alpha_{\sigma(l)}$.

Then, Corollary 2.3 and Proposition D.1 imply the following result.

COROLLARY 2.4. *Under the same conditions and notation as in Theorem 2.1, the following holds. Define the deterministic index*

$$\mathcal{L}(d) := \max\{l \geq 1 : \beta_l(d) \geq 2 + (\log d)^{-\kappa}\}$$

with the convention that $\mathcal{L}(d) = 0$ if $\beta_1(d) < 2 + (\log d)^{-\kappa}$:

(i) For $1 \leq l \leq \mathcal{L}(d)$, we have with probability $1 - o(1)$

$$\begin{aligned} & |\lambda_{l+1}(A) - \sqrt{d}\Lambda(\beta_l(d))| + |\lambda_{N+1-l}(A) + \sqrt{d}\Lambda(\beta_l(d))| \\ & \leq C(d^{-c(\Lambda(\beta_l(d))-2)} + d^{-\theta/3})\sqrt{d}. \end{aligned}$$

(ii) For $l = \mathcal{L}(d) + 1$, we have with probability $1 - o(1)$

$$\max\{\lambda_{l+1}(A), -\lambda_{N+1-l}(A)\} \leq (2 + C(\log d)^{-2\kappa})\sqrt{d}.$$

(Here, $C \equiv C_\kappa$ is a constant depending on κ .)

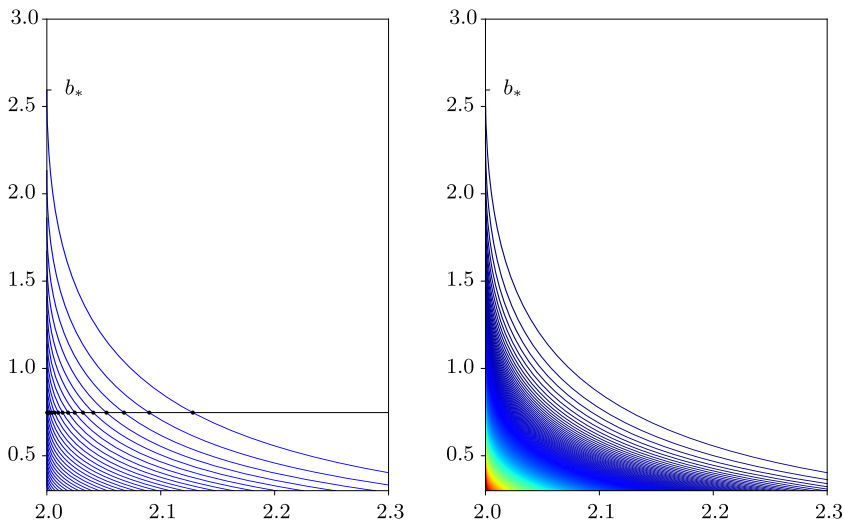


FIG. 1. An illustration of the typical values of the nontrivial eigenvalues $\lambda_2(A/\sqrt{d}), \lambda_3(A/\sqrt{d}), \dots$ in the interval $(2, \infty)$ (horizontal axis) as a function of $b = d/\log N$ (vertical axis). For each $l = 1, 2, \dots$, we plot the function $b \mapsto \Lambda(\beta_l(b \log N))$. Left: $N = 50$; the typical eigenvalue configuration of A/\sqrt{d} in the interval $(2, \infty)$ for $d = b \log N$ is given by a horizontal slice of the graph at b , indicated by black dots. Right: $N = 1000$; we colour the graphs $b \mapsto \Lambda(\beta_l(b \log N))$ depending on l to distinguish them from each other. Note that, for $b > b_* := \frac{1}{\log 4 - 1} \approx 2.59$, there are no typical eigenvalues in $(2, \infty)$, and for $d = b \log N$ with fixed $b < b_*$ there are $N^{1-b/b_*+o(1)}$ typical eigenvalues in $(2, \infty)$.

The errors $o(1)$ in the probabilities can be easily made quantitative by a slight refinement of the argument in Appendix D. See Figure 1 for an illustration of Corollary 2.4. An analogous result holds for the matrix \underline{A} , whose details we omit.

REMARK 2.5. There is a typical normalized degree $\beta_l(d)$ greater than or equal to 2 if and only if $\beta_1(d) \geq 2$. Thus, we introduce the critical value d_* as the unique solution of $\beta_1(d_*) = 2$. It is easy to see that

$$d_* = b_* \log N + O(\log \log N), \quad b_* := \frac{1}{\log 4 - 1}.$$

Since $\mathcal{L}(d) = 0$ for $d > d_*$ and $\mathcal{L}(d) \geq 1$ for $d \leq d_*$, we conclude from Corollary 2.4 that $\lambda_2(A)/\sqrt{d}$ converges to 2 in probability if and only if $\liminf \frac{d-d_*}{\log N} \geq 0$.

REMARK 2.6. Fix $b < b_*$, and set $d = b \log N$. From the definition of $\beta_l(d)$, we deduce that $|\{l : \beta_l(d) \geq 2 + o(1)\}| = N^{1-b/b_*+o(1)}$. Hence, using Corollary 2.4, we conclude that with probability $1 - o(1)$ the matrix A/\sqrt{d} has $N^{1-b/b_*+o(1)}$ eigenvalues in $\mathbb{R} \setminus [-2 - o(1), 2 + o(1)]$.

Our final result is a version of our results for sparse Wigner matrices. Let $A = (A_{xy})$ be as above and $W = (W_{xy})$ be an independent Wigner matrix with bounded entries. That is, W is Hermitian, and its upper triangular entries $(W_{xy} : x \leq y)$ are independent complex-valued random variables with mean zero and variance one, $\mathbb{E}|W_{xy}|^2 = 1$, and $|W_{xy}| \leq K$ almost surely for some constant K . Then, we define the sparse Wigner matrix $M = (M_{xy})$ as the Hadamard product of A and W , with entries $M_{xy} := A_{xy} W_{xy}$.

THEOREM 2.7. *Theorem 2.1 holds also for the eigenvalues $\lambda_l(M)$ of a sparse Wigner matrix M instead of $\lambda_l(\underline{A})$, provided that the normalized degree α_x is replaced by*

$$(2.5) \quad \alpha_x = \frac{1}{d} \sum_{y \in [N]} |M_{xy}|^2.$$

Here, the constant C from Theorem 2.1 depends on K in addition to v and κ .

Similarly, versions of Corollaries 2.3 and 2.4 can be easily obtained if all entries of W have a common positive mean with appropriate upper and lower bounds. Furthermore, for these results and Theorem 2.7 the boundedness assumption on the entries of W can be considerably relaxed with some extra work.

3. Notation. In this section we collect notation and tools used throughout this paper. The reader interested in the strategy of the proof can skip this section at first reading and proceed directly to Section 4, returning to this section as needed for the precise notation.

We denote the positive integers by $\mathbb{N} = \{1, 2, 3, \dots\}$ and define $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. We set $[n] := \{1, \dots, n\}$ for any $n \in \mathbb{N}$, $[0] := \emptyset$ and $[[r]] := \{0, \dots, r\}$ for any $r \in \mathbb{N}_0$. We write $|X|$ for the cardinality of the finite set X . We use $\mathbb{1}_\Omega$ as symbol for the indicator function of the event Ω . Universal constants or estimates involving a universal constant are denoted by C and $O(\cdot)$, respectively.

Notation related to vectors and matrices. Vectors in \mathbb{R}^N are denoted by boldface lowercase Latin letters like \mathbf{u} , \mathbf{v} and \mathbf{w} and their Euclidean norms by $\|\mathbf{u}\|$, $\|\mathbf{v}\|$ and $\|\mathbf{w}\|$, respectively. For a matrix $M \in \mathbb{R}^{N \times N}$, $\|M\|$ is its operator norm induced by the Euclidean norm on \mathbb{R}^N .

Let $M \in \mathbb{R}^{N \times N}$ be a matrix and $V \subset [N]$. We define the matrix $M_V \in \mathbb{R}^{|V| \times |V|}$ and the family $M_{(V)}$ through

$$M_V := (M_{ij})_{i,j \in V}, \quad M_{(V)} := (M_{ij})_{i \in V \text{ or } j \in V}.$$

If $V = \{x\}$ for some $x \in [N]$, then we also write $M_{(x)}$ instead of $M_{\{\{x\}\}}$.

The eigenvalues of a Hermitian matrix $M \in \mathbb{R}^{N \times N}$ are denoted by

$$\lambda_1(M) \geq \lambda_2(M) \geq \dots \geq \lambda_N(M).$$

Moreover, for Hermitian matrices $R, T \in \mathbb{R}^{N \times N}$ we write $R \geq T$ if

$$\langle \mathbf{w}, R\mathbf{w} \rangle \geq \langle \mathbf{w}, T\mathbf{w} \rangle$$

for all $\mathbf{w} \in \mathbb{R}^N$. We remark that this is equivalent to $\lambda_N(R - T) \geq 0$.

For any $x \in [N]$, we define the standard basis vector $\mathbf{1}_x := (\delta_{xy})_{y \in [N]} \in \mathbb{R}^N$. To any subset $S \subset [N]$, we assign the vector $\mathbf{1}_S \in \mathbb{R}^N$ given by $\mathbf{1}_S := \sum_{x \in S} \mathbf{1}_x$. Note that $\mathbf{1}_{\{x\}} = \mathbf{1}_x$. We also introduce the normalized vector $\mathbf{e} := N^{-1/2} \mathbf{1}_{[N]}$. If $V \subset [N]$ and $\mathbf{w} = (w_x)_{x \in [N]} \in \mathbb{R}^N$, then $\mathbf{w}|_V$ denotes the vector in \mathbb{R}^N with components $\langle \mathbf{1}_y, \mathbf{w}|_V \rangle := \langle \mathbf{1}_y, \mathbf{w} \rangle$ for all $y \in V$ and $\langle \mathbf{1}_y, \mathbf{w}|_V \rangle = 0$ for all $y \in [N] \setminus V$.

Notation related to graphs. In the entire paper we consider finite graphs exclusively. Let \mathbb{H} and \mathbb{G} be two graphs. We write $\mathbb{H} \subset \mathbb{G}$ if $V(\mathbb{H}) \subset V(\mathbb{G})$ and $E(\mathbb{H}) \subset E(\mathbb{G})$. If $\mathbb{H} \subset \mathbb{G}$, then we denote by $\mathbb{G} \setminus \mathbb{H}$ the graph on $V(\mathbb{G})$ with edge set $E(\mathbb{G}) \setminus E(\mathbb{H})$. To each graph $\mathbb{G} = (V(\mathbb{G}), E(\mathbb{G}))$, we assign its adjacency matrix $\text{Adj}(\mathbb{G})$. If \mathbb{G} is a graph on $[N]$, then, for any $V \subset [N]$, we denote by $\mathbb{G}|_V$ the subgraph induced by \mathbb{G} on the vertex set V . If A is the adjacency matrix of \mathbb{G} , then $A_V = \text{Adj}(\mathbb{G}|_V)$ is the adjacency matrix of $\mathbb{G}|_V$.

For simplicity, we specialize to the vertex set $[N]$ in the following definitions. Let \mathbb{H} be a graph with vertex set $[N]$ and $M = \text{Adj}(\mathbb{H})$ be its adjacency matrix. Vertices in $[N]$ are usually labelled by x, y, z . The degree of the vertex x is $D_x^{\mathbb{H}} := \sum_{y \in [N]} M_{xy}$. With respect to \mathbb{H} , the graph distance of two vertices $x, y \in [N]$ is denoted by

$$d^{\mathbb{H}}(x, y) := \min\{k \in \mathbb{N}_0 : (M^k)_{xy} \neq 0\}.$$

For $i \in \mathbb{N}_0$, we introduce the i -sphere $S_i^{\mathbb{H}}(x)$ and the i -ball $B_i^{\mathbb{H}}(x)$ around x defined through

$$S_i^{\mathbb{H}}(x) = \{y \in [N] : d^{\mathbb{H}}(x, y) = i\}, \quad B_i^{\mathbb{H}}(x) = \{y \in [N] : d^{\mathbb{H}}(x, y) \leq i\}.$$

For the remainder of this work, \mathbb{G} will be an Erdős–Rényi graph with vertex set $[N]$ and edge probability d/N , where N is a large parameter and $d \equiv d_N$ is a function of N . Moreover, $A = \text{Adj}(\mathbb{G}) = (A_{xy})_{x, y \in [N]} \in \{0, 1\}^{N \times N}$ will always denote the adjacency matrix of \mathbb{G} . In this situation we write $D_x, d(x, y), S_i(x)$ and $B_i(x)$ instead of $D_x^{\mathbb{G}}, d^{\mathbb{G}}(x, y), S_i^{\mathbb{G}}(x)$ and $B_i^{\mathbb{G}}(x)$, respectively. Note the relation $\alpha_x d = D_x$ between the normalized degree α_x , defined in (2.1), and the degree D_x .

Probabilistic notation and tools. We now introduce a notion of very high probability event as well as a notation for bounds which hold with very high probability. Both will be used extensively throughout the present work.

DEFINITION 3.1.

(i) Let $\Xi \equiv \Xi_{N, \nu}$ be a family of events parametrized by $N \in \mathbb{N}$ and $\nu > 0$. We say that Ξ holds with very high probability if for every $\nu > 0$ there exists C_ν such that

$$\mathbb{P}(\Xi_{N, \nu}) \geq 1 - C_\nu N^{-\nu}$$

for all $N \in \mathbb{N}$.

(ii) For a σ -algebra \mathcal{F}_N and an event $E_N \in \mathcal{F}_N$, we extend the definition (i) to Ξ holds with very high probability on E conditioned on \mathcal{F} if for all $\nu > 0$ there exists C_ν such that

$$\mathbb{P}(\Xi_{N, \nu} | \mathcal{F}_N) \geq 1 - C_\nu N^{-\nu}$$

almost surely on E_N , for all $N \in \mathbb{N}$.

We remark that the notion of very high probability survives a union bound involving $N^{O(1)}$ events. We shall tacitly use this fact throughout the paper.

CONVENTION 3.2 (Estimates with very high probability). In statements that hold with very high probability, we use the symbol $\mathcal{C} \equiv C_\nu$ to denote a generic positive constant, depending on ν , such that the statement holds with probability at least $1 - c_\nu N^{-\nu}$, provided C_ν and c_ν are chosen large enough.

We now illustrate the previous convention by explaining in detail the meaning of $|X| \leq CY$ with very high probability. Such estimates often appear throughout the paper. The bound $|X| \leq CY$ with very high probability means that, for each $\nu > 0$, there are constants $C_\nu > 0$ and $c_\nu > 0$, depending on ν , such that

$$\mathbb{P}(|X| \leq C_\nu Y) \geq 1 - c_\nu N^{-\nu}$$

for all $N \in \mathbb{N}$. Here, X and Y are allowed to depend on N .

We also write $X = \mathcal{O}(Y)$ to mean $|X| \leq CY$.

Throughout the following we use the function

$$(3.1) \quad h(\alpha) := (1 + \alpha) \log(1 + \alpha) - \alpha$$

for $\alpha \geq 0$.

To illustrate Definition 3.1 and Convention 3.2, we record the following lemma that we shall need throughout the paper.

LEMMA 3.3 (Upper bound on the degree). *For any $x \in [N]$, we have with very high probability*

$$D_x \leq \Delta \leq C(d + \log N),$$

where $\Delta \equiv \Delta(d, N, C)$ is defined by

$$(3.2) \quad \Delta := \begin{cases} d + C\sqrt{d \log N} & \text{if } d \geq \frac{1}{2} \log N, \\ C \frac{\log N}{\log \log N - \log d} & \text{if } d \leq \frac{1}{2} \log N. \end{cases}$$

PROOF. From Bennett’s inequality we obtain

$$\mathbb{P}(D_x \geq d + \alpha d) \leq e^{-dh(\alpha)}.$$

The claim now follows from an elementary analysis of the right-hand side, by requiring that it be bounded by $N^{-\nu}$. \square

4. Main ideas of the proof. In this section we explain the main ideas of the proof of Theorem 2.1. Let \mathbb{G} be an Erdős-s-Rényi graph with vertex set $[N]$ and edge probability d/N , and let A be its adjacency matrix. In the actual proof, all arguments will be applied to $\underline{A} = A - \mathbb{E}A$. However, in this sketch we explain certain ideas on the level of A for the sake of clarity. In each case a simple adjustment yields the argument for \underline{A} instead of A .

If $d \ll \log N$, then A has many eigenvalues of modulus larger than $2\sqrt{d}$, and they are related to vertices of large degree [3]. On the other hand, if $d \gg \log N$, then there are no eigenvalues whose modulus is larger than $2\sqrt{d}$ [4] (with the exception of the trivial top eigenvalue of A).

In order to understand the relationship between large eigenvalues and vertices of large degree, it is very insightful to analyse the structure of \mathbb{G} in the neighbourhood of a vertex $x \in [N]$ of large normalized degree α_x . (In the following we explain the arguments for large eigenvalues only. Dealing with small eigenvalues requires straightforward modifications.) If α_x is sufficiently large, then there is $r_x \in \mathbb{N}$, depending on α_x , such that \mathbb{G} has with very high probability the following properties:

- (a) For each $1 \leq i \leq r_x$, the ratio $|S_{i+1}(x)|/|S_i(x)|$ concentrates around d (Lemma 5.4 below).
- (b) The subgraph $\mathbb{G}|_{B_{r_x}(x)}$ is a tree up to a bounded number of edges (Lemma 5.5 below).
- (c) The radius r_x tends to infinity with N (cf. (5.1) below).

Owing to the properties (a) and (b) of the local geometry of \mathbb{G} around a vertex x of large degree, it is natural to study the spectral properties of the adjacency matrix of the following idealized graph \mathbb{T} on $[N]$. We suppose that, in the ball $B_{r_x+1}(x)$, the graph \mathbb{T} is a tree where the root vertex x has $d\alpha_x$ children and the vertices in $B_{r_x}^{\mathbb{T}}(x) \setminus \{x\}$ have d children; see Figure 2 for an illustration.

For $i = 0, \dots, r_x$, we have $\mathbf{s}_i = |S_i^{\mathbb{T}}(x)|^{-1/2} \mathbf{1}_{S_i^{\mathbb{T}}(x)}$ (see Lemma B.1 below). Hence, the construction in (4.3) naturally suggests to consider

$$(4.4) \quad \mathbf{v} = \sum_{i=0}^{r_x} u_i |S_i(x)|^{-1/2} \mathbf{1}_{S_i(x)}$$

as approximate eigenvector of A , that is, to replace \mathbf{s}_i in (4.3) by $|S_i(x)|^{-1/2} \mathbf{1}_{S_i(x)}$. In Proposition 5.1 below, we show that \mathbf{v} is an approximate eigenvector of A with approximate eigenvalue $\sqrt{d} \Lambda(\alpha_x)$. The proof heavily relies on the properties (a), (b) and (c) listed above and justified in Section 5.

The proof of Theorem 2.1 requires two additional key steps. Namely,

- (i) two different vertices of large degree induce two different eigenvalues,
- (ii) all eigenvalues of modulus larger than $2\sqrt{d}$ arise from vertices of large degree.

We remark that (i) is equivalent to a lower bound on the l th largest eigenvalue in terms of the l th largest degree of \mathbb{G} while (ii) is equivalent to a corresponding upper bound.

For (i), we construct the *pruned graph* \mathbb{G}_2 . It is a subgraph of \mathbb{G} such that A is well approximated by the adjacency matrix A_2 of \mathbb{G}_2 and $B_{r_x}^{\mathbb{G}_2}(x)$ and $B_{r_y}^{\mathbb{G}_2}(y)$ are disjoint if $x, y \in [N]$, $x \neq y$ and $\alpha_x, \alpha_y \geq 2$ (see Lemma 7.2 below). Hence, the construction in (4.4) yields two orthogonal approximate eigenvectors of A which thus induce two different eigenvalues (or the same eigenvalue with multiplicity at least two). This completes (i) (cf. Proposition 7.1).

Thanks to (i), we now know that $\lambda_1(\underline{A}) \geq \dots \geq \lambda_L(\underline{A}) \geq (2 + o(1))\sqrt{d}$ if $L := N - |V|$ and $V := \{x \in [N] : \alpha_x \leq 2\}$. Hence, (ii) will follow if we can show that $\lambda_{L+1}(\underline{A}) \leq (2 + o(1))\sqrt{d}$. By the min-max principle we have

$$\max_{\mathbf{w} \in \mathbb{S}(U)} \langle \mathbf{w}, \underline{A} \mathbf{w} \rangle \geq \lambda_{L+1}(\underline{A}),$$

where $\mathbb{S}(U)$ is the unit sphere in the linear subspace $U := \text{span}\{\mathbf{1}_x : x \in V\} \subset \mathbb{R}^N$. Thus, it suffices to establish an upper bound on the largest eigenvalue μ of \underline{A}_V . This will be deduced from the matrix inequality

$$(4.5) \quad I_N + Q + o(1) \geq d^{-1/2} \underline{A}$$

which holds with very high probability (see Proposition 6.1 below). Here, $Q = (\alpha_x \delta_{xy})_{x,y \in [N]}$ is the diagonal matrix of normalized degrees. The inequality (4.5) is a consequence of an estimate on the nonbacktracking matrix associated with \underline{A} and an Ihara–Bass type formula from [4].

We now explain how to prove that μ is at most $(2 + o(1))\sqrt{d}$. Let $\tilde{\mathbf{w}} = (\tilde{w}_x)_{x \in V}$ be a normalized eigenvector of \underline{A}_V corresponding to μ . We define a normalized vector $\mathbf{w} = (w_x)_{x \in [N]} \in \mathbb{R}^N$ through $w_x = \tilde{w}_x$ for $x \in V$ and $w_x = 0$ for $x \in [N] \setminus V$. Since $\langle \tilde{\mathbf{w}}, \underline{A}_V \tilde{\mathbf{w}} \rangle = \langle \mathbf{w}, \underline{A} \mathbf{w} \rangle$, we can evaluate the inequality in (4.5) at \mathbf{w} . This yields

$$(4.6) \quad \begin{aligned} \frac{\mu}{\sqrt{d}} - o(1) &\leq \langle \mathbf{w}, (I_N + Q) \mathbf{w} \rangle \\ &= 1 + \sum_{x: \alpha_x > 2} \alpha_x w_x^2 + \sum_{x: 2 \geq \alpha_x > \tau} \alpha_x w_x^2 + \sum_{x: \alpha_x \leq \tau} \alpha_x w_x^2 \end{aligned}$$

for any $\tau \in (1, 2)$, where we used that \mathbf{w} is normalized. The contribution for $\alpha_x > 2$ vanishes as $w_x = 0$ for such x . Since \mathbf{w} is normalized, the contribution for $\alpha_x \leq \tau$ is at most τ . We choose $\tau = 1 + o(1)$.

What remains is estimating the sum in the regime $2 \geq \alpha_x > \tau$. In the following paragraph we shall sketch the proof of the bound

$$(4.7) \quad w_x^2 \leq \varepsilon \|\mathbf{w}\|_{B_{r_x}^{\mathbb{G}_\tau}(x)} \|^2$$

which holds for some $\varepsilon = o(1)$ uniformly for all $x \in [N]$ satisfying $\tau < \alpha_x \leq 2$. Here, \mathbb{G}_τ is the *pruned graph*, a subgraph of \mathbb{G} such that $A_\tau = \text{Adj}(\mathbb{G}_\tau)$, the adjacency matrix of \mathbb{G}_τ , and A are close, and $B_{r_x}^{\mathbb{G}_\tau}(x)$ and $B_{r_y}^{\mathbb{G}_\tau}(y)$ are disjoint for all vertices $x, y \in [N]$ satisfying $x \neq y$ and $\alpha_x, \alpha_y > \tau$ (compare Lemma 7.2 below). Given (4.7), we conclude

$$\sum_{x: 2 \geq \alpha_x > \tau} \alpha_x w_x^2 \leq 2 \sum_{x: 2 \geq \alpha_x > \tau} w_x^2 \leq 2\varepsilon \|\mathbf{w}\|^2,$$

where we employed in the last step that $(\mathbf{1}_{B_{r_x}^{\mathbb{G}_\tau}(x)})_{x: \alpha_x > \tau}$ is a family of orthogonal vectors. Since $\|\mathbf{w}\| = 1$, $\varepsilon = o(1)$ and $\tau = 1 + o(1)$, we obtain from (4.6) that $\mu \leq (1 + o(1) + \tau + 2\varepsilon)\sqrt{d} = (2 + o(1))\sqrt{d}$. Therefore, $\lambda_{L+1}(\underline{A}) \leq \mu \leq (2 + o(1))\sqrt{d}$.

We now sketch the proof of (4.7). For the graph \mathbb{T} described above, the delocalization estimate in (4.7) can be obtained by analysing the tridiagonal matrix $Z^\mathbb{T}$ introduced in (4.1) via a transfer matrix argument. As \mathbb{G}_τ is close to \mathbb{T} locally around a vertex x satisfying $\alpha_x > \tau$, the tridiagonal matrix \widehat{Z} constructed from A_τ around x is well approximated by $Z^\mathbb{T}$. Hence, the transfer matrices associated with \widehat{Z} and $Z^\mathbb{T}$ are also close, and a version of the argument for $Z^\mathbb{T}$ can be used to deduce (4.7). This completes the sketch of the proof of (ii) and thus the sketch of the proof of Theorem 2.1.

5. Large eigenvalues induced by vertices of large degree. Let \mathbb{G} be an Erdős–Rényi graph on the vertex set $[N]$ with edge probability d/N . Let $A = \text{Adj}(\mathbb{G})$ be the adjacency matrix of \mathbb{G} and $\underline{A} := A - \mathbb{E}A$. Proposition 5.1 below, the main result of this section, shows that each vertex of sufficiently large degree induces two approximate eigenvectors of \underline{A} . As explained after the statement of Proposition 5.1, this locates a positive and a negative eigenvalue of \underline{A} of large modulus.

We now introduce the notation necessary to define the approximate eigenvectors. To lighten notation, we fix the vertex x throughout and omit all arguments (x) from our notation. In particular, we just write S_i and B_i instead of $S_i(x)$ and $B_i(x)$. Define

$$(5.1) \quad r_x := \left\lfloor \frac{\log N}{3 \log D_x} \right\rfloor,$$

and let $r \leq r_x$. Let $u_0 > 0$, and define the coefficients

$$(5.2) \quad \begin{aligned} u_1 &:= \frac{\sqrt{D_x}}{\sqrt{D_x - d}} u_0, \\ u_i &:= \frac{d^{(i-1)/2}}{(D_x - d)^{(i-1)/2}} u_1 \quad (i = 2, 3, \dots, r + 1). \end{aligned}$$

Here and in the following, we exclusively consider the event $\{D_x > d\}$ such that u_1, \dots, u_{r+1} are always well defined. Then, on the event $S_i \neq \emptyset$ for $i = 1, \dots, r$, we define the approximate eigenvectors $\mathbf{v} \equiv \mathbf{v}(x, r)$ and $\mathbf{v}_- \equiv \mathbf{v}_-(x, r)$ through

$$(5.3) \quad \mathbf{v} := \sum_{i=0}^r u_i \mathbf{s}_i, \quad \mathbf{v}_- := \sum_{i=0}^r (-1)^i u_i \mathbf{s}_i, \quad \mathbf{s}_i := |S_i|^{-1/2} \mathbf{1}_{S_i}.$$

Finally, we choose u_0 so that the normalization $\|\mathbf{v}\|^2 = \|\mathbf{v}_-\|^2 = \sum_{i=0}^r u_i^2 = 1$ holds.

For the following proposition we recall the definition $\alpha_x = D_x/d$. Also, throughout this section we use $\mathcal{K} \geq 1$ to denote a constant that is chosen large enough depending on ν in the definition of very high probability.

PROPOSITION 5.1 (Eigenvectors induced by vertex of large degree). *Let $x \in [N]$ be a fixed vertex. Suppose that $\mathcal{K}\sqrt{\log N} \leq d \leq N^{1/4}$ and $\log d \leq r \leq r_x$. Then,*

$$\|(\underline{A} - \sqrt{d}\Lambda(\alpha_x))\mathbf{v}\| + \|(\underline{A} + \sqrt{d}\Lambda(\alpha_x))\mathbf{v}_-\| \leq \mathcal{C} \left(\log d + \frac{\log N}{d} \right)^{1/2} \left(1 + \frac{\log N}{D_x} \right)^{1/2}$$

with very high probability on

$$(5.4) \quad \left\{ \left(2 + 2\frac{\log d}{r} \right) d \leq D_x \leq \sqrt{N}(2d)^{-r} \right\}$$

conditioned on D_x .

We remark that if M is a Hermitian matrix and \mathbf{v} a normalized vector such that $\|M\mathbf{v}\| \leq \varepsilon$, then M has an eigenvalue in $[-\varepsilon, \varepsilon]$. Therefore, Proposition 5.1 implies that \underline{A} possesses with very high probability two eigenvalues λ_{\pm} in the vicinity of $\pm\sqrt{d}\Lambda(\alpha_x)$ if α_x is sufficiently large.

We shall show in Lemma 5.4 below that $S_i \neq \emptyset$ for $i = 1, \dots, r$ with very high probability on the event $\{d \leq D_x \leq \sqrt{N}(2d)^{-r}\}$.

To prove Proposition 5.1, we only consider the term $\|(\underline{A} - \sqrt{d}\Lambda(\alpha_x))\mathbf{v}\|$. The other term is treated in the same way. We shall decompose $(\underline{A} - \sqrt{d}\Lambda(\alpha_x))\mathbf{v}$ into a sum $\mathbf{w}_0 + \dots + \mathbf{w}_4$ of vectors which are all proved to have a small norm. (See Lemma 5.2 below and the estimates in Lemma 5.3 below.) Each of the vectors \mathbf{w}_i will turn out to be small for a different reason which is why we treat them individually.

In order to define the vectors \mathbf{w}_i , we introduce the notation

$$(5.5) \quad \mathbf{e} := N^{-1/2}\mathbf{1}_{[N]}, \quad N_i(y) := \langle \mathbf{1}_y, A\mathbf{1}_{S_i} \rangle = |S_i \cap S_1(y)|$$

for all $i = 0, \dots, r$ and $y \in [N]$. Thus, $N_i(y)$ is the number of edges starting in S_i and ending in y . Note that if the graph $\mathbb{G}|_{B_{i+1}}$ is a tree, then it is easy to see that

$$(5.6) \quad N_i(y) = \mathbb{1}_{y \in S_{i-1}}(D_y - \mathbb{1}_{i \geq 2}) + \mathbb{1}_{y \in S_{i+1}}$$

with the convention that $S_{-1} := \emptyset$; see Figure 3 for an illustration of $N_i(y)$.

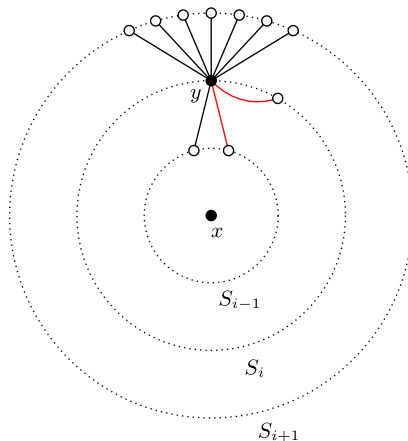


FIG. 3. An illustration of the definition of $N_i(y)$ from (5.5), where $y \in S_i(x)$. The red edges are forbidden in a tree. For a tree, $N_{i+1}(y) = D_y - 1$, $N_i(y) = 0$ and $N_{i-1}(y) = 1$.

Define

$$\begin{aligned}
 \mathbf{w}_0 &:= \frac{d}{N} \mathbf{v} - d(\mathbf{e}, \mathbf{v})\mathbf{e}, \\
 \mathbf{w}_1 &:= \sum_{i=0}^r \frac{u_i}{\sqrt{|S_i|}} \left(\sum_{y \in S_{i+1}} (N_i(y) - 1)\mathbf{1}_y + \sum_{y \in S_i} N_i(y)\mathbf{1}_y \right), \\
 \mathbf{w}_2 &:= \sum_{i=1}^r \frac{u_i}{\sqrt{|S_i|}} \sum_{y \in S_{i-1}} \left(N_i(y) - \frac{|S_i|}{|S_{i-1}|} \right) \mathbf{1}_y, \\
 \mathbf{w}_3 &:= u_2 \left(\frac{\sqrt{|S_2|}}{\sqrt{|S_1|}} - \sqrt{d} \right) \mathbf{s}_1 \\
 &\quad + \sum_{i=2}^{r-1} \left[u_{i+1} \left(\frac{\sqrt{|S_{i+1}|}}{\sqrt{|S_i|}} - \sqrt{d} \right) + u_{i-1} \left(\frac{\sqrt{|S_i|}}{\sqrt{|S_{i-1}|}} - \sqrt{d} \right) \right] \mathbf{s}_i, \\
 \mathbf{w}_4 &:= \left(u_{r-1} \frac{\sqrt{|S_r|}}{\sqrt{|S_{r-1}|}} - u_{r-1}\sqrt{d} - u_{r+1}\sqrt{d} \right) \mathbf{s}_r + u_r \frac{\sqrt{|S_{r+1}|}}{\sqrt{|S_r|}} \mathbf{s}_{r+1}.
 \end{aligned}
 \tag{5.7}$$

LEMMA 5.2 (Decomposition of $(\underline{A} - \sqrt{d}\Lambda(\alpha_x))\mathbf{v}$). *We have the decomposition*

$$(\underline{A} - \sqrt{d}\Lambda(\alpha_x))\mathbf{v} = \mathbf{w}_0 + \mathbf{w}_1 + \mathbf{w}_2 + \mathbf{w}_3 + \mathbf{w}_4.
 \tag{5.8}$$

Lemma 5.2 will be shown in Section 5.1 below. We now explain the origin and interpretation of the different errors $\mathbf{w}_0, \dots, \mathbf{w}_4$:

- The vector \mathbf{w}_0 is equal to $-(\mathbb{E}A)\mathbf{v}$ and hence takes care of the expectation $\mathbb{E}A$ in the definition of $\underline{A} = A - \mathbb{E}A$. It will turn out to be small because the vector \mathbf{v} is localized near the vertex x and hence has a small overlap with \mathbf{e} which is completely delocalized.
- The vector \mathbf{w}_1 quantifies the extent to which $\mathbb{G}|_{B_{r+1}}$ deviates from a tree. Indeed, by (5.6) it vanishes if $\mathbb{G}|_{B_{r+1}}$ is a tree. It will turn out to be small because the number of cycles in $\mathbb{G}|_{B_{r+1}}$ is not too large.
- The vector \mathbf{w}_2 quantifies the extent to which $\mathbb{G}|_{B_r}$ deviates from a tree with the property that, for each $i \geq 2$, all vertices in S_i have the same degree. Indeed, it is immediate that the term $i = 1$ is always zero, and the other terms vanish under the above condition, by (5.6). It will turn out to be small because the number of cycles in $\mathbb{G}|_{B_{r+1}}$ is not too large and because $N_i(y)$ will concentrate around $\frac{|S_i|}{|S_{i-1}|}$ for most vertices $y \in S_{i-1}$, for any $i \geq 2$.
- The vector \mathbf{w}_3 quantifies the extent to which $\mathbb{G}|_{B_r}$ deviates from a graph with the property that $|S_{i+1}| = d|S_i|$ for all $i \geq 1$. The ratios $|S_{i+1}|/|S_i|$ will turn out to concentrate around d with very high probability, thus ensuring the smallness of \mathbf{w}_3 .
- Finally, the vector \mathbf{w}_4 quantifies the error arising from edges connecting the ball B_r , where the tree approximation is valid, and to the rest of the graph $[N] \setminus B_r$, where it is not. It will be small by the exponential decay of the coefficients u_i .

LEMMA 5.3 (Estimates on $\mathbf{w}_0, \dots, \mathbf{w}_4$). *Let $\mathcal{K}\sqrt{\log N} \leq d \leq N^{1/4}$. For any $r \leq r_x$, the estimates*

$$\|\mathbf{w}_0\| = \mathcal{O}(dN^{-1/4}),
 \tag{5.9a}$$

$$\|\mathbf{w}_1\| = \mathcal{O}(d^{-1/2}),
 \tag{5.9b}$$

$$\|\mathbf{w}_2\| = \mathcal{O}\left(\left(\log d + \frac{\log N}{d} \right)^{1/2} \left(1 + \frac{\log N}{D_x} \right)^{1/2} \right),
 \tag{5.9c}$$

$$(5.9d) \quad \|\mathbf{w}_3\| = \mathcal{O}\left(\left(\frac{\log N}{D_x}\right)^{1/2}\right),$$

$$(5.9e) \quad \|\mathbf{w}_4\| = \mathcal{O}\left(\left(\frac{d}{D_x - d}\right)^{(r-2)/2} \left(\frac{d}{\sqrt{D_x - d}} + \left(\frac{\log N}{D_x}\right)^{1/2}\right)\right)$$

hold with very high probability on $\{d < D_x \leq \sqrt{N}(2d)^{-r}\}$, conditioned on $A_{(x)}$.

Lemma 5.3 is proved in Section 5.2 below.

PROOF OF PROPOSITION 5.1. From Lemmas 5.2 and 5.3 we get

$$\|(\underline{A} - \sqrt{d}\Lambda(\alpha_x))\mathbf{v}\| \leq C\left(\log d + \frac{\log N}{d}\right)^{1/2} \left(1 + \frac{\log N}{D_x}\right)^{1/2} + C\sqrt{d}\left(\frac{d}{D_x - d}\right)^{(r-1)/2}$$

with very high probability on $\{2d < D_x \leq \sqrt{N}(2d)^{-r}\}$. Write $D_x = (2 + t)d$ for $t > 0$. To conclude the proof, it suffices to show that

$$(5.10) \quad \sqrt{d}\left(\frac{d}{D_x - d}\right)^{(r-1)/2} \leq 1, \quad \text{that is, } \log(1 + t) \geq \frac{\log d}{r - 1}.$$

Since by assumption $\log d \leq r$, this condition is satisfied, provided that

$$2\frac{\log d}{r} \leq t = \frac{D_x}{d} - 2.$$

This concludes the proof for \mathbf{v} . For \mathbf{v}_- the bound follows in the same way from trivial modifications of Lemma 5.2 and Lemma 5.3 obtained by replacing u_i by $(-1)^i u_i$. We leave the details of these modifications to the reader. \square

5.1. *Decomposition of $(\underline{A} - \sqrt{d}\Lambda(\alpha_x))\mathbf{v}$ —Proof of Lemma 5.2.* In this subsection we prove Lemma 5.2. We recall that $\mathbf{w}_0, \dots, \mathbf{w}_4$ were defined in (5.7) and $N_i(y)$ in (5.5).

PROOF OF LEMMA 5.2. Recalling the definition $\underline{A} = A - \mathbb{E}A$, we have

$$\underline{A}\mathbf{v} = A\mathbf{v} + \frac{d}{N}\mathbf{v} - d\langle \mathbf{e}, \mathbf{v} \rangle \mathbf{e} = \mathbf{w}_0 + \sum_{i=0}^r \frac{u_i}{\sqrt{|S_i|}} A\mathbf{1}_{S_i}.$$

By the definition (5.5) of $N_i(y)$ and the triangle inequality for the graph distance, we have

$$A\mathbf{1}_{S_i} = \mathbb{1}_{i \geq 1} \sum_{y \in S_{i-1}} N_i(y)\mathbf{1}_y + \sum_{y \in S_i} N_i(y)\mathbf{1}_y + \sum_{y \in S_{i+1}} N_i(y)\mathbf{1}_y$$

so that

$$\begin{aligned} \underline{A}\mathbf{v} &= \mathbf{w}_0 + \sum_{i=1}^r \frac{u_i}{\sqrt{|S_i|}} \sum_{y \in S_{i-1}} N_i(y)\mathbf{1}_y + \sum_{i=0}^r \frac{u_i}{\sqrt{|S_i|}} \left[\sum_{y \in S_i} N_i(y)\mathbf{1}_y + \sum_{y \in S_{i+1}} N_i(y)\mathbf{1}_y \right] \\ &= \mathbf{w}_0 + \mathbf{w}_1 + \sum_{i=1}^r \frac{u_i}{\sqrt{|S_i|}} \left[\sum_{y \in S_{i-1}} N_i(y)\mathbf{1}_y + \mathbf{1}_{S_{i+1}} \right] + u_0\mathbf{1}_{S_1} \\ &= \mathbf{w}_0 + \mathbf{w}_1 + \mathbf{w}_2 + \sum_{i=1}^r \frac{u_i}{\sqrt{|S_i|}} \left[\sum_{y \in S_{i-1}} \frac{|S_i|}{|S_{i-1}|} \mathbf{1}_y + \mathbf{1}_{S_{i+1}} \right] + u_0\mathbf{1}_{S_1}. \end{aligned}$$

Thus, we conclude

$$\begin{aligned}
 \underline{A}\mathbf{v} - \sum_{k=0}^2 \mathbf{w}_k &= u_0 \mathbf{1}_{S_1} + \sum_{i=1}^r \frac{u_i}{\sqrt{|S_i|}} \left[\frac{|S_i|}{|S_{i-1}|} \mathbf{1}_{S_{i-1}} + \mathbf{1}_{S_{i+1}} \right] \\
 &= u_0 \sqrt{|S_1|} \mathbf{s}_1 + u_1 \sqrt{|S_1|} \mathbf{s}_0 + u_2 \frac{\sqrt{|S_2|}}{\sqrt{|S_1|}} \mathbf{s}_1 \\
 &\quad + \sum_{i=2}^{r-1} \left(u_{i+1} \frac{\sqrt{|S_{i+1}|}}{\sqrt{|S_i|}} + u_{i-1} \frac{\sqrt{|S_i|}}{\sqrt{|S_{i-1}|}} \right) \mathbf{s}_i \\
 &\quad + u_r \frac{\sqrt{|S_{r+1}|}}{\sqrt{|S_r|}} \mathbf{s}_{r+1} + u_{r-1} \frac{\sqrt{|S_r|}}{\sqrt{|S_{r-1}|}} \mathbf{s}_r.
 \end{aligned}
 \tag{5.11}$$

Since $\sqrt{d}\Lambda(\alpha_x) = \frac{D_x}{\sqrt{D_x-d}}$, from the definition of u_i in (5.2) we get

$$\begin{aligned}
 \sqrt{d}\Lambda(\alpha_x)u_0 &= \sqrt{D_x}u_1, \\
 \sqrt{d}\Lambda(\alpha_x)u_1 &= \sqrt{D_x}u_0 + \sqrt{d}u_2, \\
 \sqrt{d}\Lambda(\alpha_x)u_i &= \sqrt{d}u_{i-1} + \sqrt{d}u_{i+1}
 \end{aligned}$$

for all $i = 2, 3, \dots, r$. This implies

$$\sqrt{d}\Lambda(\alpha_x)\mathbf{v} = u_1 \sqrt{|S_1|} \mathbf{s}_0 + (u_0 \sqrt{|S_1|} + u_2 \sqrt{d}) \mathbf{s}_1 + \sum_{i=2}^r (u_{i-1} \sqrt{d} + u_{i+1} \sqrt{d}) \mathbf{s}_i.$$

Together with (5.11), this yields

$$\begin{aligned}
 (\underline{A} - \sqrt{d}\Lambda(\alpha_x))\mathbf{v} - \sum_{k=0}^2 \mathbf{w}_k &= u_2 \left(\frac{\sqrt{|S_2|}}{\sqrt{|S_1|}} - \sqrt{d} \right) \mathbf{s}_1 \\
 &\quad + \sum_{i=2}^{r-1} \left[u_{i+1} \left(\frac{\sqrt{|S_{i+1}|}}{\sqrt{|S_i|}} - \sqrt{d} \right) + u_{i-1} \left(\frac{\sqrt{|S_i|}}{\sqrt{|S_{i-1}|}} - \sqrt{d} \right) \right] \mathbf{s}_i \\
 &\quad + \left(u_{r-1} \frac{\sqrt{|S_r|}}{\sqrt{|S_{r-1}|}} - u_{r-1} \sqrt{d} - u_{r+1} \sqrt{d} \right) \mathbf{s}_r + u_r \frac{\sqrt{|S_{r+1}|}}{\sqrt{|S_r|}} \mathbf{s}_{r+1} \\
 &= \mathbf{w}_3 + \mathbf{w}_4
 \end{aligned}$$

which concludes the proof. \square

5.2. *Smallness of $\mathbf{w}_0, \dots, \mathbf{w}_4$ —Proof of Lemma 5.3.* This subsection is devoted to the proof of Lemma 5.3.

In order to estimate $\mathbf{w}_0, \dots, \mathbf{w}_4$, we shall make frequent use of the following lemma. We recall from Section 3 that $A_{(x)} = (A_{ij})_{i=x \text{ or } j=x}$.

LEMMA 5.4 (Concentration of $|S_{i+1}|/|S_i|$). *Let $1 \leq d \leq N^{1/4}$:*

(i) *For $r, i \in \mathbb{N}$ satisfying $1 \leq i \leq r$, we have*

$$\left| \frac{|S_{i+1}|}{d|S_i|} - 1 \right| = \mathcal{O} \left(\left(\frac{\log N}{d|S_i|} \right)^{1/2} \right)
 \tag{5.12a}$$

and

$$(5.12b) \quad |S_i| = D_x d^{i-1} \left(1 + \mathcal{O} \left(\left(\frac{\log N}{d D_x} \right)^{1/2} \right) \right)$$

with very high probability on

$$(5.13) \quad \left\{ \mathcal{K} \frac{\log N}{d} \leq D_x \leq \sqrt{N} (2d)^{-r} \right\}$$

conditioned on $A_{(x)}$.

(ii) Moreover, for all $r, i \in \mathbb{N}$ satisfying $1 \leq i \leq r$, the bound

$$(5.14) \quad |S_{i+1}| \leq d |S_i| + \mathcal{C} (d |S_i| \log N)^{1/2}$$

holds with very high probability on $\{D_x \leq \sqrt{N} (2d)^{-r}\}$, conditioned on $A_{(x)}$.

In the applications of Lemma 5.4 below, we shall always work under the assumptions $d \geq \mathcal{K} \sqrt{\log N}$ and $D_x > d$ which imply that the lower bound in (5.13) is always satisfied. Before proving Lemma 5.4, we first conclude (5.9a), (5.9d) and (5.9e) from it.

PROOF OF (5.9a), (5.9d) AND (5.9e). In the whole proof we exclusively work on the event $\{d < D_x \leq \sqrt{N} (2d)^{-r}\}$. For the proof of (5.9a), we start by using the Cauchy–Schwarz inequality to obtain

$$(5.15) \quad \|\mathbf{w}_0\| \leq \frac{d}{N} + \frac{d}{\sqrt{N}} \sum_{y \in B_r} |(\mathbf{1}_y, \mathbf{v})| \leq \frac{d}{N} + \frac{d}{\sqrt{N}} \sqrt{|B_r|}.$$

From Lemma 5.4 we conclude with very high probability

$$|B_r| \leq 1 + \sum_{i=0}^{r-1} D_x (2d)^i \leq 2D_x (2d)^{r-1} \leq (2D_x)^r \leq (2D_x)^{r_x} \leq \sqrt{N}$$

by definition of r_x . Hence, (5.9a) follows.

We now turn to the proof of (5.9d). Estimating the definition of \mathbf{w}_3 yields

$$\begin{aligned} \|\mathbf{w}_3\|^2 &\leq d \left[\left(\frac{\sqrt{|S_2|}}{\sqrt{d|S_1|}} - 1 \right)^2 u_2^2 \right. \\ &\quad \left. + 2 \sum_{i=2}^{r-1} \left(\left(\frac{\sqrt{|S_{i+1}|}}{\sqrt{d|S_i|}} - 1 \right)^2 u_{i+1}^2 + \left(\frac{\sqrt{|S_i|}}{\sqrt{d|S_{i-1}|}} - 1 \right)^2 u_{i-1}^2 \right) \right] \\ &\leq \mathcal{C} \frac{\log N}{D_x} \left[u_2^2 + 2 \sum_{i=2}^{r-1} (u_{i+1}^2 + u_{i-1}^2) \right] \end{aligned}$$

with very high probability conditioned on $A_{(x)}$, where we used (5.12a) and (5.12b) in the last step. This completes the proof of (5.9d).

For the proof of (5.9e), we apply the triangle inequality to \mathbf{w}_4 to obtain

$$(5.16) \quad \|\mathbf{w}_4\| \leq |u_{r-1}| \left| \frac{\sqrt{|S_r|}}{\sqrt{|S_{r-1}|}} - \sqrt{d} \right| + |u_{r+1}| \sqrt{d} + |u_r| \frac{\sqrt{|S_{r+1}|}}{\sqrt{|S_r|}}.$$

In case $D_x \leq 2d$, this implies (5.9e) directly by $|u_i| \leq 1$, Lemma 5.4 and $\sqrt{d} \leq d/\sqrt{D_x - d}$. If $D_x \geq 2d$, then, using Lemma 5.4, we conclude from (5.16) that

$$\|\mathbf{w}_4\| \leq \frac{d^{(r-2)/2}}{(D_x - d)^{(r-2)/2}} \left(\frac{d}{\sqrt{D_x - d}} + \mathcal{C} \left(\frac{\log N}{D_x} \right)^{1/2} \right),$$

where we used the definition of u_{r-1} and u_r as well as $|u_1| \leq 1$ in the last step. This shows the bound (5.9e). \square

PROOF OF LEMMA 5.4. Define

$$\mathcal{E}_i := \frac{d|S_i|}{N} + \frac{1}{\sqrt{N}}.$$

We shall prove below that there are constants $C, c > 0$ such that if there exists $i \geq 1$ satisfying $|B_i| \leq \sqrt{N}$, then, for all $\varepsilon \in [0, 1]$, we have

$$(5.17) \quad \begin{aligned} \mathbb{P}((1 - \varepsilon - C\mathcal{E}_i)d|S_i| \leq |S_{i+1}| \leq (1 + \varepsilon + C\mathcal{E}_i)d|S_i| | A_{(B_{i-1})}) \\ \geq 1 - 2 \exp(-cd|S_i|\varepsilon^2). \end{aligned}$$

From (5.17) we now conclude (5.12a) and

$$(5.18) \quad D_x \left(\frac{d}{2}\right)^{i-1} \leq |S_i| \leq D_x(2d)^{i-1},$$

for $1 \leq i \leq r$, simultaneously, by induction. For $i = 1$, we choose $\varepsilon^2 = C \log N / (dD_x)$ which is bounded by 1 for \mathcal{K} in (5.13) large enough. Since $d \geq 1$, the upper bound on D_x from (5.13) implies $\mathcal{E}_1 \leq \varepsilon$. Thus, we obtain (5.12a) directly from (5.17). The estimate (5.18) is trivial for $i = 1$.

For the induction step, we assume that \mathcal{K} in (5.13) is large enough that the right-hand side of (5.12a) for $i = 1$ is less than $1/2$. Suppose first that with very high probability (5.12a) and (5.18) hold up to i . Since $|S_i| \geq D_x$ by (5.18) $_i$, we conclude from (5.12a) $_i$ that (5.18) $_{i+1}$ holds. Next, suppose that with very high probability (5.12a) holds up to i and (5.18) up to $i + 1$. By (5.18) $_{i+1}$, we deduce that, for $i + 1 \leq r$, we have $|B_{i+1}| \leq D_x(2d)^r \leq \sqrt{N}$, where the last inequality follows by assumption on D_x . Hence, we may apply (5.17) $_{i+1}$ to estimate $|S_{i+2}|$ with the choice $\varepsilon^2 = C \log N / (d|S_{i+1}|)$ with the same C , as in the first induction step. From $d \geq 1$, the upper bound on D_x in (5.13), $i + 1 \leq r$ and (5.18) $_{i+1}$, we obtain $\mathcal{E}_{i+1} \leq \varepsilon$ and, hence, we conclude (5.12a) $_{i+1}$ after taking the conditional expectation with respect to $A_{(x)}$. Note that the necessary union bounds are affordably since the right-hand side of (5.17) is always at least $1 - 2N^{-cC}$.

The expansion in (5.12b) is a direct consequence of

$$(5.19) \quad D_x d^{i-1}(1 - \varepsilon_i) \leq |S_i| \leq D_x d^{i-1}(1 + \varepsilon_i), \quad \varepsilon_i := 2C \left(\frac{\log N}{dD_x}\right)^{1/2} \sum_{j=0}^{i-2} d^{-j/2},$$

for $1 \leq i \leq r$ with very high probability on $\{\mathcal{K} \log N / d \leq D_x \leq \sqrt{N}(2d)^{-r}\}$, conditioned on $A_{(x)}$ as well as the fact that the geometric sum in the definition of ε_i is bounded by 2 uniformly in i .

The estimates in (5.12a) and (5.18) imply (5.19) by induction as follows. The case $i = 1$ is trivial. For the induction step we conclude from (5.12a) that

$$|S_{i+1}| \geq d|S_i| \left(1 - C \left(\frac{\log N}{d|S_i|}\right)^{1/2}\right) \geq D_x d^i \left(1 - \varepsilon_i - C \left(\frac{\log N}{d^i D_x (1 - \varepsilon_i)}\right)^{1/2}\right).$$

Here, we used $|S_i| \neq 0$ by (5.18) in the first step and the induction hypothesis in the second step. As $\varepsilon_i \leq 3/4$ for sufficiently large \mathcal{K} due to $D_x \geq \mathcal{K} \log N / d$, we obtain the lower bound in (5.19). The upper bound is proved completely analogously. This completes the proof of (5.12b).

Next, we also conclude (5.14) from (5.17) by an induction argument. For $i = 1$, we can assume that $D_x = |S_1| \neq 0$, as there is nothing to show otherwise. If $|S_1| \neq 0$, then we choose

$\varepsilon^2 = C \log N / (d|S_1|)$ in (5.17). As $\mathcal{E}_1 \leq \varepsilon$, this directly implies (5.14) for $i = 1$. In the induction step we assume $|S_{i+1}| \neq 0$ as (5.14) $_{i+1}$ is trivial for $|S_{i+1}| = 0$. The induction hypothesis and $D_x \leq \sqrt{N}(2d)^{-r}$ imply $|B_{i+1}| \leq \sqrt{N}$. Hence, $|S_{i+1}| \neq 0$ allows the choice $\varepsilon^2 = C \log N / (d|S_{i+1}|)$ in (5.17) in order to bound $|S_{i+2}|$. As $\mathcal{E}_{i+1} \leq \varepsilon$, we obtain (5.14) $_{i+1}$ by taking the conditional expectation with respect to $A_{(x)}$.

What remains therefore is the proof of (5.17). We condition on $A_{(B_{i-1})}$ and suppose that $|B_i| \leq \sqrt{N}$. (Note that B_i is measurable with respect to $A_{(B_{i-1})}$.) Let us compute the law of $|S_{i+1}|$, conditioned on $A_{(B_{i-1})}$. For $y \in B_i^c$, denote by $F_y = \mathbb{1}_y$ adjacent to S_i . Then, conditioned on $A_{(B_{i-1})}$, $(F_y)_{y \in B_i^c}$ are i.i.d. Bernoulli random variables with expectation $1 - (1 - p)^{|S_i|}$, where $p = d/N$. Thus,

$$|S_{i+1}| = \sum_{y \in B_i^c} F_y \stackrel{d}{=} \text{Binom}(1 - (1 - p)^{|S_i|}, N - |B_i|),$$

conditioned on $A_{(B_{i-1})}$. Thus,

$$(5.20) \quad E_{i+1} := \mathbb{E}[|S_{i+1}| | A_{(B_{i-1})}] = (1 - (1 - p)^{|S_i|})(N - |B_i|) = d|S_i|(1 + O(\mathcal{E}_i)),$$

where, in the last step, we used that $1 - p = e^{-p+O(p^2)}$ as well as the assumptions $|B_i| \leq \sqrt{N}$ and $d \leq N^{1/4}$. We recall the definition of h from (3.1). Applying Bennett’s inequality yields

$$\begin{aligned} \mathbb{P}(|S_{i+1}| - E_{i+1} > \varepsilon E_{i+1} | A_{(B_{i-1})}) &\leq 2 \exp(-E_{i+1} \min\{h(1 + \varepsilon), h(1 - \varepsilon)\}) \\ &\leq 2 \exp\left(-\frac{1}{2}d|S_i| \min\{h(1 + \varepsilon), h(1 - \varepsilon)\}\right), \end{aligned}$$

where we used (5.20) and that $\mathcal{E}_i = O(N^{-1/4})$, as follows from the assumptions $|B_i| \leq \sqrt{N}$ and $d \leq N^{1/4}$. Now, the claim (5.17) follows from (5.20) and the observation that there exists a $c > 0$ such that $\min\{h(1 + \varepsilon), h(1 - \varepsilon)\} \geq c\varepsilon^2$ for all $\varepsilon \in [0, 1]$. \square

LEMMA 5.5 (Few cycles in small balls). *For $k, r \in \mathbb{N}$, we have*

$$(5.21) \quad \mathbb{P}(|E(\mathbb{G}|_{B_r})| - |B_r| + 1 \geq k|S_1|) \leq \frac{1}{N^k} (C(d + |S_1|))^{2kr+k} (2kr)^{2k}.$$

COROLLARY 5.6. *If $d \geq C$ for some universal constant $C > 0$, then the number of cycles in $\mathbb{G}|_{B_r}$ is bounded,*

$$|\{\text{cycles in } \mathbb{G}|_{B_r}\}| = \mathcal{O}(1),$$

with very high probability on $\{d < D_x \leq N^{1/4}\} \cap \{r \leq r_x\}$, conditioned on S_1 .

PROOF. Given $\nu > 0$, it is easy to conclude from (5.21) that there is $k \in \mathbb{N}$ such that the number of cycles in $\mathbb{G}|_{B_r}$ is bounded by k with very high probability on the event $\{d < D_x \leq N^{1/4}\} \cap \{r \leq r_x\}$ by the lower bound on d and the definition of r_x in (5.1). \square

COROLLARY 5.7. *With very high probability on $\{D_x > d\}$, conditioned on S_1 , we have, for all $i \leq r_x + 1$,*

$$(5.22) \quad |S_i| = \sum_{y \in S_i} N_{i-1}(y) + \mathcal{O}(1), \quad |S_i| = \sum_{y \in S_{i-1}} N_i(y) + \mathcal{O}(1)$$

as well as

$$(5.23) \quad \sum_{y \in S_i} N_i(y) = \mathcal{O}(1).$$

PROOF. By choosing a spanning tree of $\mathbb{G}|_{B_{r_x}}$, we conclude that, with very high probability on $\{D_x > d\}$, conditioned on S_1 , we can find $\mathcal{O}(1)$ edges of $\mathbb{G}|_{B_{r_x}}$ such that removing them yields a tree on the vertex set B_{r_x} . Now, (5.22) follows easily by noting that

$$\langle \mathbf{1}_{S_{i-1}}, A\mathbf{1}_{S_i} \rangle = \sum_{y \in S_{i-1}} N_i(y) = \sum_{y \in S_i} N_{i-1}(y),$$

and that the left-hand side equals $|S_i|$ if A is the adjacency matrix of a tree. Finally, (5.23) follows by noting that its left-hand side vanishes if A is the adjacency matrix of a tree. \square

Before the proof of Lemma 5.5, we show how (5.22) and (5.23) are used to bound $\|\mathbf{w}_1\|$ and establish (5.9b).

PROOF OF (5.9b). Since $S_0 = \{x\}$, we have that $N_0(y) = 1$ for all $y \in S_1$. Moreover, $N_0(x) = 0$ since \mathbb{G} has no loops. Hence,

$$(5.24) \quad \mathbf{w}_1 = \sum_{i=1}^r \frac{u_i}{\sqrt{|S_i|}} \left(\sum_{y \in S_{i+1}} (N_i(y) - 1)\mathbf{1}_y + \sum_{y \in S_i} N_i(y)\mathbf{1}_y \right).$$

By the triangle inequality, $|u_i| \leq 1$, and the fact that $N_i(y) \geq 1$ for all $y \in S_{i+1}$, we find

$$\|\mathbf{w}_1\| \leq \sum_{i=1}^r \frac{1}{\sqrt{|S_i|}} \left(\sum_{y \in S_{i+1}} (N_i(y) - 1) + \sum_{y \in S_i} N_i(y) \right) = \sum_{i=1}^r \frac{1}{\sqrt{|S_i|}} \mathcal{O}(1),$$

where in the last step we used (5.22) and (5.23). The claim follows using (5.12b). \square

PROOF OF LEMMA 5.5. Throughout the proof we condition on S_1 . Let $r \leq N$ and $k \in \mathbb{N}$, and without loss of generality $r, k \geq 1$. Define the set \mathcal{H}_k as the set of connected graphs \mathbb{H} satisfying $x \in V(\mathbb{H}) \subset [N]$, $S_1^{\mathbb{H}} \subset S_1$, $|E(\mathbb{H})| = |V(\mathbb{H})| - 1 + k$ and $|V(\mathbb{H})| \leq 2kr + 1$. Let $\mathbb{H} \in \mathcal{H}_k$. Then,

$$\mathbb{P}(E(\mathbb{H}) \subset E(\mathbb{G})|S_1) = \left(\frac{d}{N}\right)^{|E(\mathbb{H})| - |S_1^{\mathbb{H}}|} = \left(\frac{d}{N}\right)^{|V(\mathbb{H})| - 1 + k - |S_1^{\mathbb{H}}|}.$$

Hence, by a union bound,

$$\mathbb{P}(\exists \mathbb{H} \in \mathcal{H}_k, E(\mathbb{H}) \subset E(\mathbb{G})|S_1) \leq \sum_{\mathbb{H} \in \mathcal{H}_k} \mathbb{P}(E(\mathbb{H}) \subset E(\mathbb{G})|S_1).$$

In the sum over $\mathbb{H} \in \mathcal{H}_k$, we shall sum first over the set of vertices $S_1^{\mathbb{H}}$, the set of vertices $V(\mathbb{H}) \setminus (S_1^{\mathbb{H}} \cup \{x\})$ and then over all graphs \mathbb{H} on the vertex set $V(\mathbb{H})$. Writing $q_1 := |S_1^{\mathbb{H}}|$ and $q_2 := |V(\mathbb{H}) \setminus (S_1^{\mathbb{H}} \cup \{x\})|$, so that $|V(\mathbb{H})| = q_1 + q_2 + 1$, we find

$$\begin{aligned} & \mathbb{P}(\exists \mathbb{H} \in \mathcal{H}_k, E(\mathbb{H}) \subset E(\mathbb{G})|S_1) \\ & \leq \sum_{0 \leq q_1 + q_2 \leq 2kr} \binom{|S_1|}{q_1} \binom{N - |S_1| - 1}{q_2} C_{q_1 + q_2 + 1, k} \left(\frac{d}{N}\right)^{q_2 + k}, \end{aligned}$$

where $C_{q,k}$ is the number of connected graphs on q vertices with $q - 1 + k$ edges. To estimate $C_{q,k}$, we note that each such graph can be written as a union of a tree on q vertices and k additional edges. By Cayley’s theorem on the number of trees we, therefore, conclude that

$$C_{q,k} \leq q^{q-2} q^{2k} = q^{q+2k-2}.$$

Putting everything together, we conclude that

$$\begin{aligned} & \mathbb{P}(\exists \mathbb{H} \in \mathcal{H}_k, E(\mathbb{H}) \subset E(\mathbb{G}) | S_1) \\ & \leq \sum_{0 \leq q_1 + q_2 \leq 2kr} \frac{|S_1|^{q_1} N^{q_2}}{q_1! q_2!} (q_1 + q_2 + 1)^{q_1 + q_2 + 2k - 1} \left(\frac{d}{N}\right)^{q_2 + k} \\ & = \frac{d^k}{N^k} \sum_{0 \leq i \leq 2kr} \frac{1}{i!} (|S_1| + d)^i (i + 1)^{i + 2k - 1} \\ & \leq \frac{1}{N^k} (C(d + |S_1|))^{2kr + k} (2kr)^{2k}, \end{aligned}$$

where in the second step we used the binomial theorem and in the last step Stirling’s approximation.

In order to conclude the proof, it suffices to show that

$$(5.25) \quad \{|E(\mathbb{G}|_{B_r})| - |B_r| + 1 \geq k\} \subset \{\exists \mathbb{H} \in \mathcal{H}_k, E(\mathbb{H}) \subset E(\mathbb{G})\}.$$

To show (5.25), we suppose that \mathbb{G} is in the event on the left-hand side of (5.25). Let T be a spanning tree of B_r such that $d_T(x, y) = d(x, y)$ for all $y \in B_r$, where d_T is the graph distance on T . Since $|E(\mathbb{G}|_{B_r})| - |B_r| + 1 \geq k$, we can find k edges of $\mathbb{G}|_{B_r}$ that are not edges of T ; denote these edges by E_1 . Let U_1 denote the vertices incident to the edges of E_1 . Let U_2 denote the vertices in all the (unique) paths of T connecting the vertices of U_1 to x , and E_2 the edges of these paths. Consider now the graph \mathbb{H} with vertex set $U_1 \cup U_2 \cup \{x\}$ and edge set $E_1 \cup E_2$; see Figure 4 for an illustration.

We claim that $\mathbb{H} \in \mathcal{H}_k$ which will conclude the proof of (5.25). The only nonobvious property to verify is that $|V(\mathbb{H})| = |U_1 \cup U_2 \cup \{x\}| \leq 2kr + 1$. This follows easily from the observation that $|U_1| \leq 2k$ and that each of the above paths has at most $r - 1$ vertices in $[N] \setminus (U_1 \cup \{x\})$, and there are at most $2k$ paths. This concludes the proof. \square

PROOF OF (5.9c). We define $\Xi_i := \{d < D_x \leq \sqrt{N}(2d)^{-i}\}$ for $i \in \mathbb{N}$. We start by noting that

$$w_2 = \sum_{i=2}^r \frac{u_i}{\sqrt{|S_i|}} \sum_{y \in S_{i-1}} \left(N_i(y) - \frac{|S_i|}{|S_{i-1}|} \right) \mathbf{1}_y,$$

since $S_0 = \{x\}$ and $N_1(x) = |S_1| = |S_1|/|S_0|$.

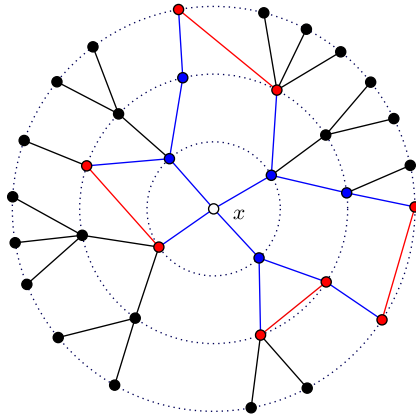


FIG. 4. Graphical representation of the proof of (5.25). We draw the ball B_r for $r = 3$. The spanning tree T is drawn using black and blue edges, and the spheres of radii 1, 2, 3 are drawn using dots. The red edges are E_1 , the blue edges E_2 , the red vertices U_1 and the blue vertices $U_2 \setminus U_1$.

We apply the Pythagorean theorem, use (5.22), $|S_i| \geq D_x d/2$ uniformly for $i \in \{2, \dots, r\}$ with very high probability on Ξ_r conditioned on $A_{(x)}$ by (5.12b) in Lemma 5.4 as well as $\sum_{i=1}^{r-1} u_{i+1}^2 \leq 1$ and obtain

$$\begin{aligned} \|\mathbf{w}_2\|^2 &= \sum_{i=1}^{r-1} \frac{u_{i+1}^2}{|S_{i+1}|} \sum_{y \in S_i} \left(N_{i+1}(y) - \frac{|S_{i+1}|}{|S_i|} \right)^2 \\ &\leq 2 \sum_{i=1}^{r-1} \frac{u_{i+1}^2}{|S_{i+1}|} \sum_{y \in S_i} \left(N_{i+1}(y) - \frac{1}{|S_i|} \sum_{y \in S_i} N_{i+1}(y) \right)^2 + \frac{C}{D_x d} \\ &\leq \frac{4}{d} \max_{1 \leq i \leq r-1} (Y_i + X_i^2) + \frac{C}{D_x d}, \end{aligned}$$

where we defined

$$\begin{aligned} Y_i &:= \frac{1}{|S_i|} \sum_{y \in S_i} (N_{i+1}(y) - \mathbb{E}[N_{i+1}(y)|A_{(B_{i-1})}])^2, \\ X_i &:= \frac{1}{|S_i|} \sum_{y \in S_i} (N_{i+1}(y) - \mathbb{E}[N_{i+1}(y)|A_{(B_{i-1})}]). \end{aligned}$$

Here, we used Young’s inequality and the fact that $\mathbb{E}[N_{i+1}(y)|A_{(B_{i-1})}]$ does not depend on $y \in S_i$. In fact, conditioned on $A_{(B_{i-1})}$, the random variables $(N_{i+1}(y))_{y \in S_i}$ are i.i.d. with law $\text{Binom}(N - |B_i|, d/N)$.

The term X_i can be easily estimated by Bennett’s inequality for the law $\text{Binom}(|S_i|(N - |B_i|), d/N)$, which yields $X_i^2 \leq C \frac{d \log N}{D_x}$ with very high probability on Ξ_r , conditioned on $A_{(x)}$. Here, we used that, by Lemma 5.4, $N - |B_i| \geq N/2$ and $|S_i| \geq D_x$ and, by definition of Ξ_r , $\frac{\log N}{d D_x} \leq 1$.

What remains, therefore, is the estimate of Y_i . We shall prove that, for all $1 \leq i \leq r - 1$,

$$(5.26) \quad Y_i \leq Cd \left(1 + \frac{\log N}{|S_i|} \right) \left(\log d + \frac{\log N}{d} \right)$$

with very high probability on Ξ_i , conditioned on $A_{(x)}$, which will conclude the proof of the lemma.

The estimate (5.26) can be regarded as a concentration result for the degrees of the vertices in S_i ; indeed, by Lemma 5.5, for any $y \in S_i$, we have $D_y = N_{i+1}(y) + \mathcal{O}(1)$ with very high probability. For any vertex $y \in S_i$, we have the variance estimate $\sqrt{\mathbb{E}(D_y - d)^2} \asymp C\sqrt{d}$. On the other hand, in the relevant regime $d \leq C \log N$, the estimate with very high probability (following from Bennett’s inequality) $|D_y - d| \leq C \log N$ is much worse. Essentially, we need an estimate with very high probability of the average $\frac{1}{|S_i|} \sum_{y \in S_i} (D_y - d)^2$, and the trivial bound $C(\log N)^2$ obtained by applying the above estimate is much too large. Instead, we need to use that the typical term of Y_i is much smaller than $(\log N)^2$. We do this using a dyadic classification of the degrees of the vertices in S_i .

For the proof of (5.26), we always condition on $A_{(x)}$ and work on the event Ξ_i . We abbreviate

$$E_y := N_{i+1}(y) - \mathbb{E}[N_{i+1}(y)|A_{(B_{i-1})}]$$

for $y \in S_i$ and introduce the level set sizes

$$(5.27) \quad L_s^i := |\{y \in S_i : E_y^2 > s^2 d^2\}|$$

for any $s > 0$. We have the probabilistic tail bound on L_s^i

$$(5.28) \quad \mathbb{P}(L_s^i \geq \ell | A_{(B_{i-1})}) \leq \binom{|S_i|}{\ell} \exp(-cd\ell(s \wedge s^2))$$

for all $s > 0$ and $\ell \in \mathbb{N}$, with very high probability. Here, $c > 0$ is a universal constant. To prove (5.28), we use a union bound to get

$$(5.29) \quad \begin{aligned} \mathbb{P}(L_s^i \geq \ell | A_{(B_{i-1})}) &\leq \sum_{T \subset S_i, |T|=\ell} \mathbb{P}(E_y^2 > (sd)^2 \text{ for all } y \in T | A_{(B_{i-1})}) \\ &\leq \binom{|S_i|}{\ell} \max_{T \subset S_i, |T|=\ell} \mathbb{P}(E_y^2 > (sd)^2 \text{ for all } y \in T | A_{(B_{i-1})}) \\ &= \binom{|S_i|}{\ell} \mathbb{P}(E_y^2 > (sd)^2 | A_{(B_{i-1})})^\ell \end{aligned}$$

for some $y \in S_i$, since $(N_{i+1}(y))_{y \in S_i}$ are i.i.d., conditioned on $A_{(B_{i-1})}$. By Bennett’s inequality we obtain

$$\mathbb{P}(E_y^2 > (sd)^2 | A_{(B_{i-1})}) \leq \exp(-cd(s \wedge s^2)),$$

since $N - |B_i| \geq N/2$ (by Lemma 5.4) and, hence, (5.28) follows.

Next, we conclude the argument by establishing (5.26). We decompose

$$(5.30) \quad |S_i| Y_i = \sum_{y \in S_i} E_y^2 \leq d|S_i| + \sum_{y \in S_i} E_y^2 \mathbb{1}_{E_y^2 \geq d}.$$

In order to estimate the second summand in (5.30), we now establish the dyadic decomposition

$$(5.31) \quad \sum_{y \in S_i} E_y^2 \mathbb{1}_{E_y^2 \geq d} \leq \sum_{k=k_{\min}}^{k_{\max}} \sum_{y \in \mathcal{N}_k^i} E_y^2 \leq d^2 \sum_{k=k_{\min}}^{k_{\max}} e^{k+1} |\mathcal{N}_k^i|$$

with very high probability, conditioned on $A_{(B_{i-1})}$, where we introduced

$$\begin{aligned} \mathcal{N}_k^i &:= \{y \in S_i : d^2 e^k < E_y^2 \leq e^{k+1} d^2\}, \quad k \in \mathbb{Z}, \\ k_{\min} &:= -\lceil \log(d) \rceil, \\ k_{\max} &:= \left\lceil C + \log\left(\frac{\log N}{d}\right) \vee \left(2 \log\left(\frac{\log N}{d}\right)\right) \right\rceil \end{aligned}$$

with some possibly ν -dependent constant $C > 0$.

We now prove (5.31) by showing that, for all $y \in S_i$, $E_y^2 \leq d^2 e^{k_{\max}+1}$ with very high probability, conditioned on $A_{(B_{i-1})}$. To that end, we note that

$$\mathbb{P}(\exists y \in S_i, E_y^2 > d^2 e^{k_{\max}+1} | A_{(B_{i-1})}) = \mathbb{P}(L_s^i \geq 1 | A_{(B_{i-1})}), \quad s = e^{(k_{\max}+1)/2}.$$

Moreover, from (5.28) with $\ell = 1$ and $|S_i| \leq N$, we obtain

$$\mathbb{P}(L_s^i \geq 1 | A_{(B_{i-1})}) \leq N \exp(-cd(s \wedge s^2)) \leq N^{-\nu},$$

where the last inequality follows for $s = e^{(k_{\max}+1)/2}$ and k_{\max} defined above. This establishes (5.31).

Next, we estimate $|\mathcal{N}_k^i|$. This will allow us to conclude the statement of the lemma from (5.30) and (5.31). In fact, we have $\mathbb{P}(|\mathcal{N}_k^i| \geq \ell | A_{(B_{i-1})}) \leq \mathbb{P}(L_{e^{k/2}}^i \geq \ell | A_{(B_{i-1})})$. We choose $\ell = \ell_k$, where

$$\ell_k := \frac{C}{d} (|S_i| + \log N) \begin{cases} e^{-k/2}, & \text{if } k \geq 0, \\ e^{-k}, & \text{if } k < 0. \end{cases}$$

Using (5.28) and estimating $\binom{|S_i|}{\ell} \leq e^{|S_i|}$, we deduce that $|\mathcal{N}_k^i| \leq \ell_k$ with very high probability on Ξ_i , conditioned on $A_{(x)}$.

With this information we now estimate the right-hand side of (5.31). We conclude from (5.30) and (5.31) that

$$\begin{aligned} |S_i|Y_i &\leq d|S_i| + d^2 \sum_{k=k_{\min}}^{k_{\max}} e^{k+1} \ell_k \\ &\leq d|S_i| + Cd(|S_i| + \log N) \left(\sum_{k=k_{\min}}^0 e^{k+1} e^{-k} + \mathbb{1}_{k_{\max} \geq 0} \sum_{k=0}^{k_{\max}} e^{k+1} e^{-k/2} \right) \\ &\leq d|S_i| + Cd(|S_i| + \log N)(|k_{\min}| + e^{k_{\max}/2}) \\ &\leq Cd(|S_i| + \log N) \left(\log(d) + \left(\frac{\log N}{d} \right)^{1/2} \vee \frac{\log N}{d} \right) \end{aligned}$$

which concludes the proof of (5.26) and hence also of (5.9c). \square

6. Quadratic form estimates on centred adjacency matrix. The main result of the present section is a bound on $\underline{A} = A - \mathbb{E}A$ in Proposition 6.1 below. In the following we write $S \geq T$ for two Hermitian $N \times N$ matrices S, T if $S - T$ is positive semidefinite, that is, $\langle \mathbf{w}, S\mathbf{w} \rangle \geq \langle \mathbf{w}, T\mathbf{w} \rangle$ for all $\mathbf{w} \in \mathbb{R}^N$. We recall the choice of σ from (2.3).

PROPOSITION 6.1 (Upper bound on $d^{-1/2}\underline{A}$). *If $4 \leq d \leq N^{2/13}$, then, with very high probability, we have*

$$I_N + Q + E \geq d^{-1/2}|\underline{A}|,$$

where $|\underline{A}| = \sqrt{\underline{A}^* \underline{A}}$, Q is the diagonal matrix defined through

$$Q := \frac{1}{d} \text{diag}(D_x)_{x \in [N]}$$

and the error matrix E satisfies

$$\|E\| \leq \frac{C(d + D_{\sigma(1)})}{d^{3/2}} \leq Cd^{-1/2} \begin{cases} 1 + d^{-1/2} \sqrt{\log N}, & \text{if } d \geq \frac{1}{2} \log N, \\ \frac{\log N}{d(\log \log N - \log d)}, & \text{if } d \leq \frac{1}{2} \log N, \end{cases}$$

with very high probability.

We postpone the proof of Proposition 6.1 to Section 6.1 below. First, we state and prove the following corollary of Proposition 6.1.

COROLLARY 6.2 (Norm bound on \underline{A}). *If $4 \leq d \leq N^{2/13}$, then we have*

$$\|\underline{A}\| \leq \sqrt{d} + \frac{D_{\sigma(1)}}{\sqrt{d}} + C \left(1 + \frac{D_{\sigma(1)}}{d} \right) \leq \begin{cases} 2\sqrt{d} + C\sqrt{\log N}, & \text{if } d \geq \frac{1}{2} \log N, \\ \frac{C \log N}{d^{1/2}(\log \log N - \log d)}, & \text{if } d \leq \frac{1}{2} \log N, \end{cases}$$

with very high probability.

Corollary 1.3 in [3] and Corollary 3.3 in [4] provide similar statements to Corollary 6.2.

PROOF OF COROLLARY 6.2. Owing to Lemma 3.3, we have $\|Q\| \leq D_{\sigma(1)}/d \leq \Delta/d$ with very high probability. Therefore, Corollary 6.2 follows immediately from Proposition 6.1. \square

6.1. *Proof of Proposition 6.1.* Let B be the *nonbacktracking matrix* associated with $d^{-1/2}\underline{A} = d^{-1/2}(A - \mathbb{E}A)$, that is the $N^2 \times N^2$ matrix with entries $B_{ef} := d^{-1/2}\underline{A}_{uv}\mathbb{1}_{y=u}\mathbb{1}_{x \neq v}$ for $e = (x, y) \in [N]^2$ and $f = (u, v) \in [N]^2$. The next proposition provides a high probability bound on the spectral radius of the nonbacktracking matrix. It is proved in [4], Theorem 2.5.

PROPOSITION 6.3 (Bound on the nonbacktracking matrix of $d^{-1/2}\underline{A}$). *There are universal constants $C > 0$ and $c > 0$ such that, for all $4 \leq d \leq N^{2/13}$ and $\varepsilon \geq 0$, we have*

$$\mathbb{P}(\rho(B) \geq 1 + \varepsilon) \leq CN^{3-c\sqrt{d}\log(1+\varepsilon)}.$$

The Ihara–Bass-type formula in the following lemma relates the spectra of B and \underline{A} . Its formulation is identical to the one of Lemma 4.1 in [4]. Therefore, we shall not present its proof in this paper.

LEMMA 6.4 (Ihara–Bass-type formula). *Let \underline{A} be an $N \times N$ matrix, and let B be the nonbacktracking matrix associated with $d^{-1/2}\underline{A}$. Let $t \in \mathbb{C}$ satisfy $t^2 \neq d^{-1}\underline{A}_{xy}\underline{A}_{yx}$ for all $x, y \in [N]$. We define the matrices $\underline{A}(t) = (\underline{A}_{xy}(t))_{x,y \in [N]}$ and $M(t) = (m_x(t)\delta_{xy})_{x,y \in [N]}$ through*

$$\underline{A}_{xy}(t) := \frac{t\sqrt{d}\underline{A}_{xy}}{t^2d - \underline{A}_{xy}\underline{A}_{yx}}, \quad m_x(t) := 1 + \sum_y \frac{\underline{A}_{xy}\underline{A}_{yx}}{t^2d - \underline{A}_{xy}\underline{A}_{yx}}.$$

Then, $t \in \text{spec}(B)$ if and only if $\det(M(t) - \underline{A}(t)) = 0$.

An argument similar to the following proof of Proposition 6.1 has been used to show Proposition 4.2 in [4].

PROOF OF PROPOSITION 6.1. We only show that $d^{-1/2}\underline{A} \leq I_N + Q + E$. The same proof implies that $-d^{-1/2}\underline{A}$ satisfies the same bound. In this proof we use the matrices $\underline{A}(t)$ and $M(t)$, defined in Lemma 6.4, exclusively for $t \in \mathbb{R}$. Note that $\underline{A}(t)$ and $M(t)$ are Hermitian for all $t \in \mathbb{R}$. If $t \in \mathbb{R}$ converges to $+\infty$, then we have that $M(t) - \underline{A}(t) \rightarrow I_N$. Therefore, $M(t) - \underline{A}(t)$ is strictly positive definite for all sufficiently large $t > 0$. Let t_* be the infimum of all $t > 0$ such that $M(t) - \underline{A}(t)$ is strictly positive definite. Hence, by continuity, the smallest eigenvalue of $M(t_*) - \underline{A}(t_*)$ is zero while all eigenvalues of $M(t) - \underline{A}(t)$ are strictly positive for $t > t_*$. Therefore, Lemma 6.4 implies that $t_* \in \text{spec}(B)$ and $M(t) - \underline{A}(t)$ is strictly positive definite for all $t \in (\rho(B), \infty)$. Hence, Proposition 6.3 yields that

$$(6.1) \quad \mathbb{P}(M(1 + \varepsilon) - \underline{A}(1 + \varepsilon) \geq 0) \geq 1 - CN^{3-c\sqrt{d}\log(1+\varepsilon)}.$$

We shall establish below the existence of a constant $C > 0$ such that

$$(6.2) \quad \begin{aligned} \|\underline{A}(t) - t^{-1}d^{-1/2}\underline{A}\| &\leq \frac{C(1 + D_{\sigma(1)})}{d^{3/2}}, \\ \|M(t) - I_N - t^{-2}Q\| &\leq \frac{C(1 + D_{\sigma(1)})}{d^2} \end{aligned}$$

for each $t \in [1, 2]$. Since $M(t) - \underline{A}(t) \geq 0$ and $Q \geq 0$ imply

$$d^{-1/2}\underline{A} \leq I_N + Q + (t - 1)I_N + t(\|\underline{A}(t) - t^{-1}d^{-1/2}\underline{A}\| + \|M(t) - I_N - t^{-2}Q\|)I_N,$$

choosing $t = 1 + \varepsilon$ with $\varepsilon = Cd^{-1/2}$ and using (6.1), (6.2) as well as $\log(1 + \varepsilon) \geq c\varepsilon$ for some $c > 0$ establish Proposition 6.1 up to showing (6.2).

We now prove (6.2). In order to estimate $\underline{A}(t) - t^{-1}d^{-1/2}\underline{A}$, we use the Schur test to conclude

$$\|\underline{A}(t) - t^{-1}d^{-1/2}\underline{A}\| \leq \max_x \sum_y |\underline{A}_{xy}(t) - t^{-1}d^{-1/2}\underline{A}_{xy}|.$$

A short computation shows that

$$\max_x \sum_y |\underline{A}_{xy}(t) - t^{-1}d^{-1/2}\underline{A}_{xy}| \leq \max_x \sum_y \frac{|\underline{A}_{xy}|^3}{t\sqrt{d}(t^2d - \underline{A}_{xy}^2)} \leq \frac{2}{t^3d^{3/2}}(D_{\sigma(1)} + 1),$$

where we used $|A_{xy}| \leq 1$, $t^2d/2 \geq \underline{A}_{xy}^2$ and $\sum_y \underline{A}_{xy}^2 \leq D_x + 1 \leq D_{\sigma(1)} + 1$ in the last step (recalling that $\underline{A}_{xy} = A_{xy} - d/N$). Thus, the first bound in (6.2) follows due to $t \geq 1$.

As $M(t)$ and Q are diagonal matrices by definition, we obtain

$$\begin{aligned} \|M(t) - I_N - t^{-2}Q\| &= \max_x \left| \sum_y \left(\frac{\underline{A}_{xy}^2}{t^2d - \underline{A}_{xy}^2} - \frac{1}{t^2d}A_{xy}^2 \right) \right| \\ &\leq \max_x \sum_y \left(\frac{\underline{A}_{xy}^4}{t^2d(t^2d - \underline{A}_{xy}^2)} + \frac{2\underline{A}_{xy}^2}{t^2N} + \frac{d}{t^2N^2} \right). \end{aligned}$$

Arguing similarly, as in the proof of the first bound in (6.2), implies the second bound in (6.2). This completes the proof of (6.2) and thus the one of Proposition 6.1. \square

7. Lower bounds on large eigenvalues. The main result of this section is the following proposition. It states that the l th largest eigenvalue of \underline{A} , $\lambda_l(\underline{A})$, is bounded from below by $\sqrt{d}\Lambda(\alpha_{\sigma(l)})$, up to a small error term, as long as $\alpha_{\sigma(l)}$ is sufficiently large. We recall that $\alpha_x := D_x/d$ for any $x \in [N]$ and the permutation σ of $[N]$ is chosen such that $(\alpha_{\sigma(l)})_{l=1}^N$ is nonincreasing (cf. (2.3)). Similarly, up to a small error term, $-\sqrt{d}\Lambda(\alpha_{\sigma(l)})$ bounds the l th smallest eigenvalue, $\lambda_{N+1-l}(\underline{A})$, of \underline{A} from above.

PROPOSITION 7.1. *Let $\mathcal{K}\sqrt{\log N} \leq d \leq \exp(\sqrt{\log N}/4)$. There is a universal constant $C > 0$ such that if the random index L_{\geq} is defined through*

$$(7.1) \quad L_{\geq} := \max\{l \in [N] : \alpha_{\sigma(l)} \geq \tau_*\}, \quad \tau_* := 2 + C \frac{(\log d)^2}{d \wedge \log N},$$

then, for any $l \in [L_{\geq}]$, the bound

$$\begin{aligned} \min\{\lambda_l(\underline{A}), -\lambda_{N+1-l}(\underline{A})\} &\geq \sqrt{d}\Lambda(\alpha_{\sigma(l)}) + \mathcal{O}\left(\left(\log d + \frac{\log N}{d}\right)^{1/2}\right) \\ &\quad \times \left(1 + \frac{\log N}{D_{\sigma(l)}}\right)^{1/2} + \left(1 + \frac{\Delta}{d}\right)\left(\frac{\Delta}{D_{\sigma(l)}} \frac{d + \log N}{d}\right)^{1/2} \end{aligned}$$

holds with very high probability. Here, Δ is defined as in (3.2). In the definition of L_{\geq} , we use the convention that $L_{\geq} = 0$ if $\alpha_{\sigma(1)} < \tau_*$.

The following lemma will be a key ingredient in the proof of the previous proposition. For its formulation we introduce the set \mathcal{V}_{τ} of vertices of large degree given by

$$\mathcal{V}_{\tau} := \{x \in [N] : D_x \geq \tau d\},$$

where $\tau > 1$. We recall the definition of h from (3.1). The following lemma provides a subgraph \mathbb{G}_τ of \mathbb{G} such that, as N goes to infinity, the length of the shortest path in \mathbb{G}_τ of two vertices in \mathcal{V}_τ tends to infinity with a lower bound given in terms of $r(\tau)$, defined through

$$(7.2) \quad r(\tau) := \frac{d}{2 \log d} h\left(\frac{\tau - 1}{2}\right) - 2.$$

The next lemma establishes the existence of the prunder graph \mathbb{G}_τ and lists its properties.

LEMMA 7.2 (Existence of pruned graph). *Let $\tau > 1$ and $r(\tau)$ be defined as in (7.2). For all $x \in \mathcal{V}_\tau$, we set $r_{x,\tau} := (\frac{1}{4}r_x) \wedge (\frac{1}{2}r(\tau))$ with r_x from (5.1). Then, there exists a subgraph \mathbb{G}_τ of \mathbb{G} with the following properties:*

- (i) *If a path p in \mathbb{G}_τ connects two vertices $x, y \in \mathcal{V}_\tau$, $x \neq y$, then p has length at least $r_{x,\tau} + r_{y,\tau} + 1$. In particular, the balls $B_{r_{x,\tau}}^{\mathbb{G}_\tau}(x)$ for $x \in \mathcal{V}_\tau$ are disjoint.*
- (ii) *The induced subgraph $\mathbb{G}_\tau|_{B_{r_{x,\tau}}^{\mathbb{G}_\tau}(x)}$ is a tree for each $x \in \mathcal{V}_\tau$.*
- (iii) *For each edge in $\mathbb{G} \setminus \mathbb{G}_\tau$, there is at least one vertex in \mathcal{V}_τ incident to it.*
- (iv) *For each $x \in \mathcal{V}_\tau$ and each $i \in \mathbb{N}$ satisfying $1 \leq i \leq r_{x,\tau}$, we have $S_i^{\mathbb{G}_\tau}(x) \subset S_i^{\mathbb{G}}(x)$.*
- (v) *For each $x \in \mathcal{V}_\tau$, we have*

$$S_1^{\mathbb{G}_\tau}(y) \cap S_i^{\mathbb{G}_\tau}(x) = S_1^{\mathbb{G}}(y) \cap S_i^{\mathbb{G}}(x)$$

for all $y \in B_{r_{x,\tau}}^{\mathbb{G}_\tau}(x) \setminus \{x\}$ and $1 \leq i \leq r_{x,\tau}$.

- (vi) *Let $C \leq d \leq N^{1/4}$ and $\tau \geq 1 + \mathcal{K}(\frac{\log N}{d^2})^{1/3}$. The degrees induced on $[N]$ by $\mathbb{G} \setminus \mathbb{G}_\tau$ are bounded according to*

$$\max_{x \in [N]} D_x^{\mathbb{G} \setminus \mathbb{G}_\tau} = \mathcal{O}\left(1 + \frac{\log N}{h((\tau - 1)/2)d}\right)$$

with very high probability.

- (vii) *Let $\mathcal{K} \log \log N \leq d \leq \mathcal{K}^{-1}N^{1/4}$. For each $x \in \mathcal{V}_\tau$ and all $2 \leq i \leq \log N / (4 \log d)$, the bound*

$$(7.3) \quad |S_i^{\mathbb{G}}(x) \setminus S_i^{\mathbb{G}_\tau}(x)| \leq D_x^{\mathbb{G} \setminus \mathbb{G}_\tau} d^{i-2} \Delta \left[1 + C \sqrt{\frac{\log N}{d \Delta}}\right]$$

holds with very high probability. Here, Δ is defined as in (3.2).

We postpone the proof to the following subsection. First, we now conclude Proposition 7.1 from Proposition 5.1, Corollary 6.2 and Lemma 7.2.

PROOF OF PROPOSITION 7.1. We always assume that $L_\geq > 0$. Otherwise, there is nothing to prove. We shall only prove the statement about $\lambda_l(\underline{A})$ and leave the necessary modifications for the analogous statement about $\lambda_{N+1-l}(\underline{A})$ to the reader (see the proof of Proposition 5.1). Let \mathbb{G}_2 be a subgraph of \mathbb{G} possessing the properties described in Lemma 7.2 for $\tau = 2$.

We fix $l \in [L_\geq]$ and set $x := \sigma(l)$. Let $\mathbf{v}(x)$ be the associated approximate eigenvector of \underline{A} around x , constructed in (5.3), with $r = r_{x,2} - 2$, where $r_{x,2} := (\frac{1}{4}r_x) \wedge (\frac{1}{2}r(2))$ for $r(2)$ defined in (7.2). We now apply Proposition 5.1. The condition $\log d \leq r$ is satisfied, provided that $\log d \leq \frac{1}{4}r_x$ which, by Lemma 3.3, holds with very high probability under our assumption $\log d \leq \sqrt{\log N}/4$. The upper bound on D_x in (5.4) holds with very high probability due to Lemma 3.3. Finally, the lower bound on D_x in (5.4) follows from $\alpha_x \geq \tau_*$

by definition of L_{\geq} (see (7.1)). Thus, from Proposition 5.1 we conclude for $\lambda_x = \sqrt{d}\Lambda(\alpha_x)$ that

$$(7.4) \quad \|(\underline{A} - \lambda_x)\mathbf{v}(x)\| = \mathcal{O}\left(\left(\log d + \frac{\log N}{d}\right)^{1/2} \left(1 + \frac{\log N}{D_x}\right)^{1/2}\right)$$

with very high probability.

We define

$$\tilde{\mathbf{v}}(x) := \mathbf{v}(x)|_{B_{r_{x,2}-2}^{\mathbb{G}_2}}.$$

We note that the vector $\tilde{\mathbf{v}}(x)$ is not normalized. By the explicit definition of $\mathbf{v}(x)$ in (5.3) we, therefore, conclude that

$$\|\mathbf{v}(x) - \tilde{\mathbf{v}}(x)\|^2 = \sum_{i=1}^{r_{x,2}-2} u_i^2 \frac{|S_i^{\mathbb{G}}(x) \setminus S_i^{\mathbb{G}_2}(x)|}{|S_i^{\mathbb{G}}(x)|} = \mathcal{O}\left(\frac{\Delta}{D_x} \frac{D_x^{\mathbb{G} \setminus \mathbb{G}_2}}{d}\right)$$

with very high probability due to Lemma 7.2(vii), (5.12b) combined with $\mathcal{K} \log N \leq D_x d$ (see the remark below Lemma 5.4) by (7.1) and $\sum_{i=1}^{r_{x,2}} u_i^2 \leq 1$. Here, when applying Lemma 7.2(vii), we also employed that $r_{x,2} \leq r_x/4 \leq \log N / (4 \log d)$ as $D_x \geq 2d$. Hence, we have that

$$(7.5) \quad \|\mathbf{v}(x) - \tilde{\mathbf{v}}(x)\| = \mathcal{O}\left(\left(\frac{\Delta}{D_x} \frac{D_x^{\mathbb{G} \setminus \mathbb{G}_2}}{d}\right)^{1/2}\right)$$

with very high probability. Therefore, from (7.5), Corollary 6.2, and (7.4), we deduce for $\lambda_x = \sqrt{d}\Lambda(\alpha_x)$ that

$$(7.6) \quad \begin{aligned} (\underline{A} - \lambda_x) \frac{\tilde{\mathbf{v}}(x)}{\|\tilde{\mathbf{v}}(x)\|} &= (\underline{A} - \lambda_x) \frac{\mathbf{v}(x)}{\|\tilde{\mathbf{v}}(x)\|} + (\underline{A} - \lambda_x) \frac{\tilde{\mathbf{v}}(x) - \mathbf{v}(x)}{\|\tilde{\mathbf{v}}(x)\|} \\ &= \mathcal{O}\left(\left(\log d + \frac{\log N}{d}\right)^{1/2} \left(1 + \frac{\log N}{D_x}\right)^{1/2}\right. \\ &\quad \left. + \frac{d + \Delta}{\sqrt{d}} \left(\frac{\Delta}{D_x} \frac{D_x^{\mathbb{G} \setminus \mathbb{G}_2}}{d}\right)^{1/2}\right) \\ &= \mathcal{O}\left(\left(\log d + \frac{\log N}{d}\right)^{1/2} \left(1 + \frac{\log N}{D_x}\right)^{1/2}\right. \\ &\quad \left. + \left(1 + \frac{\Delta}{d}\right) \left(\frac{\Delta}{D_x} \frac{d + \log N}{d}\right)^{1/2}\right) \end{aligned}$$

with very high probability. Here, we used Lemma 7.2(vi) in the last step. Hence, $(\tilde{\mathbf{v}}(\sigma(l)))_{l=1}^{L_{\geq}}$ defines a family of orthogonal approximate eigenvectors of \underline{A} , as their supports are disjoint by Lemma 7.2(i).

We set $\tilde{W}_l := \text{span}\{\tilde{\mathbf{v}}(\sigma(k)) : k \in [l]\}$. In the following we write $\mathbb{S}(W)$ for the unit sphere with respect to the Euclidean norm in any linear subspace $W \subset \mathbb{R}^N$. Let $A_2 = \text{Adj}(\mathbb{G}_2)$ be the adjacency matrix of \mathbb{G}_2 . The max-min principle for $\lambda_l(\underline{A})$ yields

$$(7.7) \quad \begin{aligned} \lambda_l(\underline{A}) &= \max_{\dim W=l} \min_{\mathbf{w} \in \mathbb{S}(W)} \langle \mathbf{w}, \underline{A}\mathbf{w} \rangle \\ &\geq \min_{\mathbf{w} \in \mathbb{S}(\tilde{W}_l)} \langle \mathbf{w}, \underline{A}\mathbf{w} \rangle \\ &\geq \min_{\mathbf{w} \in \mathbb{S}(\tilde{W}_l)} \langle \mathbf{w}, A_2\mathbf{w} \rangle - \|A - A_2\| - \|(\mathbb{E}A)|_{\tilde{W}_l}\| \end{aligned}$$

$$\begin{aligned} &\geq \min_{k \in [l]} \lambda_{\sigma(k)} - 2\|A - A_2\| - 2\|(\mathbb{E}A)|_{\tilde{W}_l}\| \\ &\quad - \max_{k \in [l]} \|(A - \lambda_{\sigma(k)})\tilde{\mathbf{v}}(\sigma(k))\| \|\tilde{\mathbf{v}}(\sigma(k))\|^{-1}. \end{aligned}$$

Here, we added and subtracted A_2 in the third step and denote by $(\mathbb{E}A)|_{\tilde{W}_l}$ the restriction of the matrix $\mathbb{E}A$ to the subspace \tilde{W}_l . The last step follows from the definition of \tilde{W}_l , the orthogonality of $\tilde{\mathbf{v}}(\sigma(k))$ and $A_2\tilde{\mathbf{v}}(\sigma(k'))$ for $k \neq k'$ and

$$\begin{aligned} \frac{\langle \tilde{\mathbf{v}}(\sigma(k)), A_2\tilde{\mathbf{v}}(\sigma(k)) \rangle}{\|\tilde{\mathbf{v}}(\sigma(k))\|^2} &\geq \lambda_{\sigma(k)} - \|\tilde{\mathbf{v}}(\sigma(k))\|^{-1} \|(A - \lambda_{\sigma(k)})\tilde{\mathbf{v}}(\sigma(k))\| \\ &\quad - \|A - A_2\| - \|(\mathbb{E}A)|_{\tilde{W}_l}\|. \end{aligned}$$

Now, we estimate the terms on the right-hand side of (7.7) to obtain the lower bound on $\lambda_l(\underline{A})$ in the proposition. Since $t \mapsto \Lambda(t)$ is monotonically increasing for $t \geq 2$, the first term is bounded from below by $\sqrt{d}\Lambda(\alpha_{\sigma(l)})$. For the second term we use $\|A - A_2\| \leq \max_{x \in [N]} D_x^{\mathbb{G} \setminus \mathbb{G}_2} \leq \mathcal{C}(1 + \log N/d)$ with very high probability by Lemma 7.2(vi). If $\mathbf{w} \in \mathbb{S}(\tilde{W}_l)$, then $\text{supp } \mathbf{w} \subset U_l$, where $U_l := \bigcup_{k \in [l]} B_{r_{\sigma(k), 2}^{\mathbb{G}_2}}(\sigma(k))$. Therefore, by the definition of $(\mathbb{E}A)_{U_l}$, we obtain

$$\|(\mathbb{E}A)|_{\tilde{W}_l}\|^2 \leq \|(\mathbb{E}A)_{U_l}\|^2 \leq \frac{d^2}{N^2} |U_l|^2 \leq \frac{4}{d^2}.$$

Here, we estimated the operator norm of $(\mathbb{E}A)_{U_l}$ by its Hilbert–Schmidt norm and used that $\mathbb{E}A_{ij} = \frac{d}{N}$ for $i \neq j$ as well as $|U_l| \leq \frac{2}{d^2} \sum_{k \in [l]} |B_{r_{\sigma(k), 2}^{\mathbb{G}_2}}(\sigma(k))| \leq \frac{2N}{d^2}$ by Lemma 5.4 and Lemma 7.2(i). For the fourth term in (7.7), we use (7.6). This completes the proof of Proposition 7.1. \square

7.1. *Proof of Lemma 7.2.* For the proof of Lemma 7.2, we need the next lemma. For any $x \in \mathcal{V}_\tau$, it provides a bound on the number of other vertices in \mathcal{V}_τ whose distance from x is sufficiently small.

LEMMA 7.3. *There is a universal constant $C > 0$ and a (v -dependent) $\mathcal{K} > 0$ such that if $C \leq d \leq N^{1/4}$ and $\tau \geq 1 + \mathcal{K}(\frac{\log N}{d^2})^{1/3}$, then the following holds.*

For any $r \in \mathbb{N}$ satisfying $r \leq r(\tau)$ with $r(\tau)$ from (7.2) and for any $x \in \mathcal{V}_\tau$, we have

$$(7.8) \quad |\mathcal{V}_\tau \cap B_r(x)| = \mathcal{O}\left(\frac{\log N}{h((\tau - 1)/2)d}\right)$$

with very high probability.

The following lemma controls the growth of $|S_i(z)|$ in terms of d and Δ . In contrast to (5.12b) in Lemma 5.4, no lower bound on D_z is required, and no lower bound on $|S_i(z)|$ is provided.

LEMMA 7.4. *Let $\mathcal{K} \log \log N \leq d \leq \mathcal{K}^{-1}N^{1/4}$, and let $z \in [N]$. For any $i \leq \log N / (4 \log d)$, the bound*

$$|S_i(z)| \leq \Delta d^{i-1} \left[1 + \mathcal{C} \left(\frac{\log N}{d\Delta} \right)^{1/2} \right]$$

holds with very high probability. Here, Δ is defined as in (3.2).

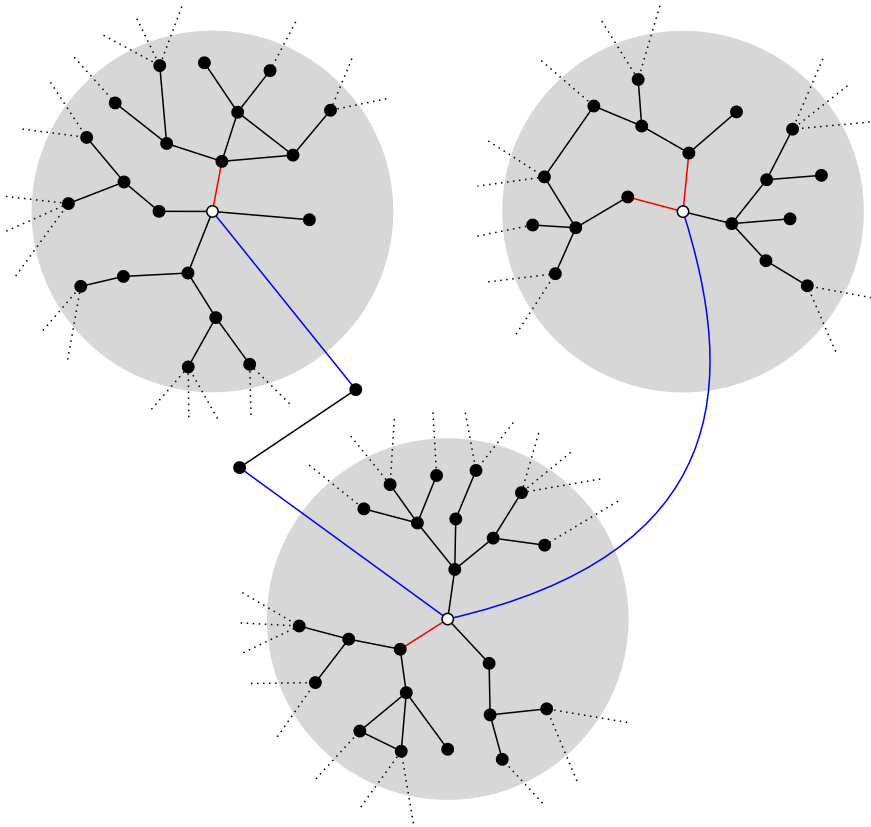


FIG. 5. A schematic illustration of the algorithm in the proof of Lemma 7.2. The vertices of \mathcal{V} are white and the other vertices black. The balls $B_{r_{x,\tau}}^{\mathbb{G}}(x)$, for each $x \in \mathcal{V}$, are indicated using grey balls, and they are disjoint by construction. The edges of the subgraph $\mathbb{H}^{(1)}$ are drawn in red. The edges of the subgraph $\mathbb{H}^{(2)}$ are drawn in blue.

The proofs of the previous two lemmas are postponed until the end of this section.

PROOF OF LEMMA 7.2. In the entire proof we write \mathcal{V} instead of \mathcal{V}_τ . We shall construct a subgraph \mathbb{H}_τ of \mathbb{G} in two steps such that $\mathbb{G}_\tau = \mathbb{G} \setminus \mathbb{H}_\tau$ satisfies the properties stated in the lemma. For a graphical depiction of the following argument, we refer to Figure 5.

In the following construction of \mathbb{H}_τ , we shall identify those edges incident to a vertex $x \in \mathcal{V}_\tau$ that lead to a loop (i.e., prevent the graph from being a tree in the vicinity of \mathcal{V}_τ) or a connection to another vertex in \mathcal{V}_τ . We shall exclusively cut these edges, thus removing the whole corresponding branches in $B_{r_x}(x)$, while leaving the other branches of $B_{r_x}(x)$ unchanged.

First, we construct a subgraph $\mathbb{H}^{(1)} \subset \mathbb{G}$ such that $B_{r_x}^{\mathbb{G} \setminus \mathbb{H}^{(1)}}(x)$ is a tree for each $x \in \mathcal{V}$. Indeed, for any $x \in \mathcal{V}$, we apply the following algorithm. For each $y \in S_1^{\mathbb{G}}(x)$, let T_y be the set of those vertices that are connected to y by a path of length at most r_x not traversing the edge connecting x and y . If $\mathbb{G}|_{T_y}$ is not a tree, that is, $|T_y| < |E(\mathbb{G}|_{T_y})| + 1$ or $x \in T_y$, then we include the edge between x and y into $\mathbb{H}^{(1)}$. We now show that

$$(7.9) \quad \max_{x \in \mathcal{V}} D_x^{\mathbb{H}^{(1)}} \leq C + q_1$$

with very high probability, where q_i denotes the maximal number of vertices in \mathcal{V} that is in the ball of radius i around a vertex in \mathcal{V} , that is,

$$(7.10) \quad q_i := \max_{x \in \mathcal{V}} |\mathcal{V} \cap B_i^{\mathbb{G}}(x) \setminus \{x\}|.$$

Let $x \in \mathcal{V}$. Indeed, owing to Corollary 5.6, with very high probability, there are at most $\mathcal{O}(1)$ edges in $B_{r_x}^{\mathbb{G}}(x)$ that prevent it from being a tree. Moreover, $S_1^{\mathbb{G}}(x)$ contains at most q_1 vertices in \mathcal{V} . This proves (7.9) and, by construction, $(\mathbb{G} \setminus \mathbb{H}^{(1)})|_{B_{r_x/2}^{\mathbb{G} \setminus \mathbb{H}^{(1)}}(x)}$ is a tree for any $x \in \mathcal{V}$.

Second, the subgraph $\mathbb{H}^{(2)} \subset \mathbb{G}$ consists of edges incident to vertices $x \in \mathcal{V}$ that are traversed by paths in $\mathbb{G} \setminus \mathbb{H}^{(1)}$ of length at most $2r_{x,\tau}$ connecting x to another vertex in \mathcal{V} . More precisely, for $x \in \mathcal{V}$, we add the following edges to $\mathbb{H}^{(2)}$. Since $(\mathbb{G} \setminus \mathbb{H}^{(1)})|_{B_{r_x/2}^{\mathbb{G} \setminus \mathbb{H}^{(1)}}(x)}$ is a tree,

for each $y \in (\mathcal{V} \cap B_{2r_{x,\tau}}^{\mathbb{G} \setminus \mathbb{H}^{(1)}}(x)) \setminus \{x\}$, there is a unique vertex $z \in S_1^{\mathbb{G} \setminus \mathbb{H}^{(1)}}(x)$ such that each path in $\mathbb{G} \setminus \mathbb{H}^{(1)}$ of length at most $2r_{x,\tau}$ connecting x and y traverses the edge between x and z . All such edges between x and z are added to $\mathbb{H}^{(2)}$. This algorithm yields that

$$(7.11) \quad \max_{x \in \mathcal{V}} D_x^{\mathbb{H}^{(2)}} \leq q_{r(\tau)},$$

where $q_{r(\tau)}$ is defined in (7.10) with $i = r(\tau)$.

We set $\mathbb{H}_\tau := \mathbb{H}^{(1)} \cup \mathbb{H}^{(2)}$ and $\mathbb{G}_\tau := \mathbb{G} \setminus \mathbb{H}_\tau$. By construction, each path in \mathbb{G}_τ between $x, y \in \mathcal{V}$ with $x \neq y$ has length at least $2(r_{x,\tau} \vee r_{y,\tau}) + 1$. This establishes property (i) of Lemma 7.2. Moreover, since $\mathbb{G}_\tau \subset \mathbb{G} \setminus \mathbb{H}^{(1)}$ is a subgraph and the latter is a tree when restricted to $B_{r_x}^{\mathbb{G} \setminus \mathbb{H}^{(1)}}(x)$, we obtain (ii).

We also note that the construction of \mathbb{H}_τ , explained above, yields

$$(7.12) \quad E(\mathbb{H}_\tau) \subset \bigcup_{x \in \mathcal{V}} \bigcup_{y \in S_1^{\mathbb{G}}(x)} \{x, y\},$$

that is, for each edge in \mathbb{H}_τ , there is at least one vertex $x \in \mathcal{V}$ incident to it. This shows (iii).

For the proof of property (iv), let $x \in \mathcal{V}$ be fixed. The construction of $\mathbb{H}^{(1)}$ implies that $S_i^{\mathbb{G} \setminus \mathbb{H}^{(1)}}(x) \subset S_i^{\mathbb{G}}(x)$ for all $1 \leq i \leq r_x/2$. As $(\mathbb{G} \setminus \mathbb{H}^{(1)})|_{B_{r_x/2}^{\mathbb{G} \setminus \mathbb{H}^{(1)}}(x)}$ is a tree, a vertex lies in $S_i^{\mathbb{G}_\tau}(x)$ only if it was in $S_i^{\mathbb{G} \setminus \mathbb{H}^{(1)}}(x)$ due to the construction of $\mathbb{H}^{(2)}$. Hence, $S_i^{\mathbb{G}_\tau}(x) \subset S_i^{\mathbb{G} \setminus \mathbb{H}^{(1)}}(x) \subset S_i^{\mathbb{G}}(x)$, and we deduce (iv).

Property (v) follows directly from the construction of \mathbb{G}_τ , as it left all branches in $B_{r_x}^{\mathbb{G}_\tau}(x)$ for $x \in \mathcal{V}_\tau$ unchanged.

For each $x \in [N]$, we now verify the bound on $D_x^{\mathbb{G} \setminus \mathbb{H}_\tau} = D_x^{\mathbb{H}_\tau}$ in (vi). For any $x \in \mathcal{V}$, we have

$$D_x^{\mathbb{H}_\tau} = D_x^{\mathbb{H}^{(1)}} + D_x^{\mathbb{H}^{(2)}} \leq \mathcal{C} + q_1 + q_{r(\tau)} \leq \mathcal{C} + 2 \max_{y \in \mathcal{V}} |\mathcal{V} \cap B_{r(\tau)}(y)|,$$

due to (7.9), (7.11) and $q_1 \leq q_{r(\tau)} \leq \max_{y \in \mathcal{V}} |\mathcal{V} \cap B_{r(\tau)}(y)|$. Thus, Lemma 7.3 implies (vi) for all $x \in \mathcal{V}$. Let $x \in [N] \setminus \mathcal{V}$. If $S_1^{\mathbb{G}}(x) \cap \mathcal{V} = \emptyset$, then $D_x^{\mathbb{H}_\tau} = 0$ due to (7.12). If $x \in S_1^{\mathbb{G}}(y)$ for some $y \in \mathcal{V}$, then

$$D_x^{\mathbb{H}_\tau} \leq \mathcal{C} |S_1^{\mathbb{G}}(x) \cap \mathcal{V}| \leq \mathcal{C} |S_2^{\mathbb{G}}(y) \cap \mathcal{V}| \leq \mathcal{C} \frac{\log N}{h(\frac{\tau-1}{2})d}$$

by Lemma 7.3. This completes the proof of (vi).

What remains is the proof of (7.3). We fix $x \in \mathcal{V}$ and conclude from (7.12) that

$$(7.13) \quad |S_i^{\mathbb{G}}(x) \setminus S_i^{\mathbb{G}_\tau}(x)| \leq \sum_{z \in S_1(x) : \{x,z\} \in E(\mathbb{H}_\tau)} |S_{i-1}^{\mathbb{G}}(z)|.$$

Lemma 7.4 provides a uniform bound on the summands in the previous sum. The number of elements in this sum is bounded by $D_x^{\mathbb{G} \setminus \mathbb{G}_\tau}$. Hence, we conclude (vii). This completes the proof of Lemma 7.2. \square

PROOF OF LEMMA 7.3. The lemma will follow from an estimate on the probability of the event $\Xi^{(k)}$ defined through

$$\Xi^{(k)} := \{\exists x \in [N] : x \in \mathcal{V}_\tau, |\mathcal{V}_\tau \cap B_r(x)| \geq k\}$$

for some $k \in \mathbb{N}$ to be chosen later. We decompose $\Xi^{(k)}$, according to

$$(7.14) \quad \begin{aligned} \Xi^{(k)} &= \bigcup_{x, \mathbf{y}, \mathbf{z}} \Xi_{x, \mathbf{y}, \mathbf{z}}^{(k)}, \\ \Xi_{x, \mathbf{y}, \mathbf{z}}^{(k)} &= \{x, y_j \in \mathcal{V}_\tau, \{x, z_1^{(j)}\}, \{z_i^{(j)}, z_{i+1}^{(j)}\}, \{z_{r_j}^{(j)}, y_j\} \in E(\mathbb{G}) \\ &\quad \text{for all } i \in [r_j - 1] \text{ and } j \in [k]\}, \end{aligned}$$

where the union is taken over all $x \in [N]$, k -tuples $\mathbf{y} = (y_1, \dots, y_k)$ of distinct elements of $[N] \setminus \{x\}$ and k -tuples $\mathbf{z} = (z^{(1)}, \dots, z^{(k)})$ of paths $z^{(j)} = (x, z_1^{(j)}, \dots, z_{r_j}^{(j)}, y_j)$ of length $r_j \in \{0, \dots, r\}$ for $j = 1, \dots, k$.

We fix such $x, \mathbf{y} = (y_1, \dots, y_k)$ and $\mathbf{z} = (z^{(1)}, \dots, z^{(k)})$. Then, recalling $p = d/N$, it is easy to see that

$$(7.15) \quad \mathbb{P}(\Xi_{x, \mathbf{y}, \mathbf{z}}^{(k)}) \leq P_{x, \mathbf{y}} \prod_{j=1}^k p^{r_j+1},$$

where $P_{x, \mathbf{y}} := \mathbb{P}(D_x \geq \tau d - k, D_{y_1} \geq \tau d - 1, \dots, D_{y_k} \geq \tau d - 1)$. We now show that

$$(7.16) \quad P_{x, \mathbf{y}} \leq C^{k+1} \exp\left(-d(k+1)h\left(\frac{\tau-1}{2}\right)\right) + (k+1) \left(\frac{(k+1)d}{N}\right)^{d(\tau-1)/2-k}.$$

We start the proof of (7.16) by exploiting the fact that $D_x, D_{y_1}, \dots, D_{y_k}$ are almost independent. Indeed, setting $y_0 := x, a_0 := (\tau - 1)d - k, a_1 := \dots := a_k := (\tau - 1)d - 1$, we obtain

$$(7.17) \quad \begin{aligned} P_{x, \mathbf{y}} &= \mathbb{E}[\mathbb{P}(D_{y_0} - d \geq a_0, \dots, D_{y_k} - d \geq a_k \mid A_Y)] \\ &= \mathbb{E}\left[\prod_{i=0}^k \mathbb{P}(D_{y_i} - d \geq a_i \mid A_Y)\right], \end{aligned}$$

where $Y := \{y_0, \dots, y_k\}$ and in the last step we used that D_{y_0}, \dots, D_{y_k} are independent conditionally on A_Y as

$$(7.18) \quad \begin{aligned} D_{y_i} - d &= X_i - \mathbb{E}[X_i \mid A_Y] + \delta_i, \\ X_i &:= \sum_{z \in [N] \setminus Y} A_{zy_i}, \\ \delta_i &:= \sum_{z \in Y} \left(A_{zy_i} - \frac{d}{N}\right) \end{aligned}$$

and X_0, \dots, X_k are independent while the remainder is measurable with respect to A_Y . Hence, (7.18) and Bennett’s inequality imply

$$\begin{aligned} \mathbb{P}(D_{y_i} - d \geq a_i \mid A_Y) &\leq C \exp(-dh((a_i - \delta_i)/d)) \\ &\leq C \exp\left(-dh\left(\min_{i=0, \dots, k} a_i/d - \max_{i=0, \dots, k} \delta_i/d\right)\right). \end{aligned}$$

Therefore, since $\min_i a_i = (\tau - 1)d - k$, (7.17) yields

$$P_{x,y} \leq C^{k+1} \exp(-d(k+1)h((\tau - 1)/2)) + \mathbb{P}\left(\max_{i=0, \dots, k} \delta_i > (\tau - 1)d/2 - k\right).$$

We choose $n = (\tau - 1)d/2 - k$ and use the bound

$$\mathbb{P}(\delta_i > n) \leq \mathbb{P}\left(\sum_{z \in Y} A_{z,y_i} \geq n\right) \leq \binom{k+1}{n} \left(\frac{d}{N}\right)^n \leq \left(\frac{(k+1)d}{N}\right)^n$$

to conclude (7.16).

We now finish the proof by combining the previous estimates. In fact, from (7.14), (7.15), $p = d/N$ and (7.16), we conclude

$$\begin{aligned} \mathbb{P}(\Xi^{(k)}) &\leq \sum_{x,y,z} \mathbb{P}(\Xi_{x,y,z}^{(k)}) \\ &\leq \binom{N}{k+1} \sum_{r_1=0}^r \cdots \sum_{r_k=0}^r \binom{N-k-1}{r_1} \cdots \binom{N-k-1 - \sum_{l=1}^{k-1} r_l}{r_k} \\ &\quad \times p^{k + \sum_{l=1}^k r_l} \max_{x,y} P_{x,y} \\ &\leq \sum_{r_1, \dots, r_k=0}^r N^{k+1 + \sum_{l=1}^k r_l} p^{k + \sum_{l=1}^k r_l} \max_{x,y} P_{x,y} \\ &\leq N \left(\sum_{l=0}^r d^{l+1}\right)^k \max_{x,y} P_{x,y} \\ &= N \left(\frac{d^{r+2} - 1}{d - 1}\right)^k \left(C^{k+1} \exp\left(-d(k+1)h\left(\frac{\tau - 1}{2}\right)\right)\right) \\ &\quad + (k+1) \left(\frac{(k+1)d}{N}\right)^{d(\tau-1)/2 - k}. \end{aligned}$$

Therefore, in order to obtain (7.8), we now show separately that each of the terms in this upper bound is dominated by $N^{-\nu}$. For the first term the condition

$$(k+1)(\log C - dh((\tau - 1)/2)) + (r+2) \log d + \log N < -\nu \log N$$

has to be satisfied. This condition is met if $k = \mathcal{C} \log N / (dh((\tau - 1)/2))$, $r \leq r(\tau)$ and $d \geq C$. Here, we used that $dh((\tau - 1)/2) \geq 4 \log d$, as we can assume $r(\tau) \geq 0$ without loss of generality (otherwise, there is nothing to be proved).

The upper bound on the second term follows from $kd \leq \mathcal{C} N^{1/3} (\log N)^{2/3}$, $d(\tau - 1) \geq \mathcal{C}k$ and $d(\tau - 1) \geq \mathcal{C}$. These latter estimates are consequences of $d \leq N^{1/4}$, the lower bound on $\tau - 1$ and $h(\tau - 1) \asymp (\tau - 1)^2 \wedge (\tau - 1)$. This completes the proof of Lemma 7.3. \square

PROOF OF LEMMA 7.4. By Lemma 3.3, $D_z \leq \Delta$ with very high probability. Defining $r := \log N / (4 \log d)$, we thus obtain $D_z \leq \sqrt{N} (2d)^{-r}$ with very high probability as $d \leq$

$\mathcal{K}^{-1}N^{1/4}$. Hence, a simple induction argument, starting from $|S_1(z)| = D_z \leq \Delta$ and using (5.14) as well as $i \leq r$ for the induction step, yields

$$(7.19) \quad |S_i(z)| \leq d^{i-1} \Delta \left[1 + 2C \sqrt{\frac{\log N}{d\Delta}} \sum_{k=1}^{i-1} d^{-(k-1)/2} \right]$$

with very high probability for all $1 \leq i \leq \log N/(4 \log d)$. Here, we used that $\sqrt{\log N/(d\Delta)}$ is small if \mathcal{K} is large due to $d \geq \mathcal{K} \log \log N$ and, for $i \geq 1$,

$$(7.20) \quad \sum_{j=1}^{i-1} d^{-(j-1)/2} \leq (1 - d^{-1/2})^{-1} \leq 2.$$

Combining (7.19) and (7.20) completes the proof of Lemma 7.4. \square

8. Upper bounds on large eigenvalues. The following proposition provides the upper bound on the l th largest eigenvalue matching the lower bound of Proposition 7.1. We recall that the permutation σ was chosen in (2.3).

PROPOSITION 8.1. *Let $\kappa \in (0, 1/2)$ be fixed, and suppose that $\theta \in (0, 1/2]$. Suppose that $(\log N)^{4/(5-2\theta)} \leq d \leq \mathcal{K}^{-1}N^{1/4}$. Define the random index L_{\leq} through*

$$(8.1) \quad L_{\leq} := \max\{l \geq 1 : \alpha_{\sigma(l)} \geq 2 + (\log d)^{-\kappa}\}$$

with the convention that $L_{\leq} = 0$ if $\alpha_{\sigma(1)} < 2 + (\log d)^{-\kappa}$. There is a universal constant $c > 0$ such that the following hold with very high probability:

(i) If $L_{\leq} > 0$, then, for all $l \in [L_{\leq}]$,

$$\max\{\lambda_l(\underline{A}), -\lambda_{N+1-l}(\underline{A})\} \leq \sqrt{d} \Lambda(\alpha_{\sigma(l)}) + C\sqrt{d}(d^{-c(\Lambda(\alpha_{\sigma(l)})-2)} + d^{-\theta/3}).$$

(ii) If $L_{\leq} = 0$, then

$$\max\{\lambda_1(\underline{A}), -\lambda_N(\underline{A})\} \leq 2\sqrt{d} + C\sqrt{d}(\log d)^{-2\kappa}.$$

(Here, the constant C depends on κ .)

Let L_{\leq} be defined as in (8.1). If $L_{\leq} > 0$, then, for $l \in [L_{\leq}]$, we set

$$(8.2) \quad V_l := \begin{cases} [N], & \text{if } l = 1, \\ [N] \setminus \{\sigma(1), \dots, \sigma(l-1)\}, & \text{if } l \geq 2. \end{cases}$$

Let \mathbb{G}_{τ} be the subgraph of \mathbb{G} introduced in Lemma 7.2. We denote by $A_{\tau} = \text{Adj}(\mathbb{G}_{\tau})$ the adjacency matrix of \mathbb{G}_{τ} and also define

$$(8.3) \quad \underline{A}_{\tau} := A_{\tau} - \Pi_{\tau}(\mathbb{E}A)\Pi_{\tau}, \quad \mathcal{Z}_{\tau} := \bigcup_{x \in \mathcal{V}_{\tau}} B_{r_{x,\tau}-2}^{\mathbb{G}_{\tau}}(x),$$

where Π_{τ} is the orthogonal projection onto $\text{Span}\{\mathbf{1}_y : y \in [N] \setminus \mathcal{Z}_{\tau}\}$. Moreover, we introduce the $N \times N$ -matrix $\underline{A}_{\tau,l}$ with entries

$$(\underline{A}_{\tau,l})_{xy} := (\underline{A}_{\tau})_{xy} \mathbb{1}_{x \in V_l} \mathbb{1}_{y \in V_l}$$

for $x, y \in [N]$.

For all $\tau, \zeta, \mu > 0$, we define a subset $\mathcal{W}_{\tau,\mu,\zeta}$ of \mathcal{V}_{τ} through

$$(8.4) \quad \mathcal{W}_{\tau,\mu,\zeta} := \{x \in [N] : \alpha_x \geq \tau, \mu \geq \sqrt{d}(\Lambda(\alpha_x \vee 2) + \zeta)\}.$$

The formulation of the following proposition uses the function $\alpha : [2, \infty) \rightarrow [2, \infty)$ defined through

$$(8.5) \quad \alpha(\eta) := \frac{\eta}{2}(\eta + \sqrt{\eta^2 - 4}).$$

Note that α is monotonically increasing and $\Lambda(\alpha(\eta)) = \eta$ for all $\eta \geq 2$.

PROPOSITION 8.2 (Delocalization estimate). *Let $(\log N)^{4/(5-2\theta)} \leq d \leq \mathcal{K}^{-1}N^{1/4}$ for some $\theta \in (0, 1/2]$ and $1 + d^{-\theta/4} \leq \tau \leq 2$. Let L_{\leq} be defined as in (8.1), $l \in [L_{\leq}]$ and V_l be defined as in (8.2). Then, there are a universal constant $c > 0$ and a constant $C > 0$ such that the following holds. If an eigenvalue $\mu > 2\sqrt{d}$ of $\underline{A}_{\tau,l}$ and some $\zeta > 0$ satisfy*

$$(8.6a) \quad \zeta \geq \left(C \left(\frac{\mu}{\mu - 2\sqrt{d}} \right)^{1/2} d^{-\theta/2} \right) \vee d^{-c(\mu/\sqrt{d}-2)},$$

$$(8.6b) \quad \mu \geq \sqrt{d}(2 + \zeta)$$

then, for any normalized eigenvector \mathbf{w} of $\underline{A}_{\tau,l}$ associated with μ , we have

$$\alpha\left(\frac{\mu}{\sqrt{d}}\right) \sum_{x \in \mathcal{W}_{\tau,\mu,\zeta}} \langle \mathbf{w}, \mathbf{1}_x \rangle^2 \leq C d^{-c((\mu/\sqrt{d}-2) \wedge \log(\mu/\sqrt{d}))}$$

with very high probability.

PROOF OF PROPOSITION 8.1. We shall only prove the upper bound on the large eigenvalues. The corresponding lower bound on the small eigenvalues is shown similarly (cf. the proof of Proposition 5.1). We first prove (i) assuming $L_{\leq} > 0$. We fix $l \in [L_{\leq}]$ and define the sphere $\mathbb{S}_l := \{\mathbf{w} \in \mathbb{R}^N : \|\mathbf{w}\| = 1, \mathbf{w}|_{V_l} = \mathbf{w}\}$. The min-max principle implies that

$$(8.7) \quad \max_{\tilde{\mathbf{w}} \in \mathbb{S}_l} \langle \tilde{\mathbf{w}}, \underline{A} \tilde{\mathbf{w}} \rangle \geq \lambda_l(\underline{A}).$$

Let μ be the largest eigenvalue of $\underline{A}_{\tau,l}$ and $\mathbf{w} \in \mathbb{S}_l$ be a corresponding eigenvector. Since $\langle \tilde{\mathbf{w}}, \underline{A}_{\tau,l} \tilde{\mathbf{w}} \rangle = \langle \tilde{\mathbf{w}}, \underline{A}_{\tau} \tilde{\mathbf{w}} \rangle$ for all $\tilde{\mathbf{w}} \in \mathbb{S}_l$, we get

$$(8.8) \quad \mu = \max_{\tilde{\mathbf{w}} \in \mathbb{S}_l} \langle \tilde{\mathbf{w}}, \underline{A}_{\tau,l} \tilde{\mathbf{w}} \rangle = \max_{\tilde{\mathbf{w}} \in \mathbb{S}_l} \langle \tilde{\mathbf{w}}, \underline{A}_{\tau} \tilde{\mathbf{w}} \rangle.$$

Thus, we obtain the lower bound

$$(8.9) \quad \mu \geq \max_{\tilde{\mathbf{w}} \in \mathbb{S}_l} \langle \tilde{\mathbf{w}}, \underline{A} \tilde{\mathbf{w}} \rangle - \|\underline{A}_{\tau} - \underline{A}\|.$$

On the other hand, (8.8) and Proposition 6.1 imply the upper bound

$$(8.10) \quad \sqrt{d} \langle \mathbf{w}, (I_N + Q + E) \mathbf{w} \rangle \geq \langle \mathbf{w}, \underline{A} \mathbf{w} \rangle \geq \mu - \|\underline{A}_{\tau} - \underline{A}\| \geq \lambda_l(\underline{A}) - 2\|\underline{A}_{\tau} - \underline{A}\|.$$

Here, we used (8.9) and (8.7) in the last step.

We choose $\tau = 1 + d^{-\theta/4}$ and apply Proposition 8.2. To that end, let $\zeta > 0$ be defined by the right-hand side of (8.6a). For a proof by contradiction, we now assume that

$$(8.11) \quad \lambda_l(\underline{A}) > \sqrt{d}(\Lambda(\alpha_{\sigma(l)}) + \zeta) + \|\underline{A} - \underline{A}_{\tau}\|.$$

From (8.11), (8.7) and (8.9), we deduce $\mu \geq \sqrt{d}(\Lambda(\alpha_{\sigma(l)}) + \zeta)$. This implies that $V_l \cap \mathcal{V}_{\tau} \subset \mathcal{W}_{\tau,\mu,\zeta}$.

Moreover, as $\mu \geq \sqrt{d}(\Lambda(\alpha_{\sigma(l)}) + \zeta) \geq \sqrt{d} \Lambda(\alpha_{\sigma(l)})$, we conclude that

$$(8.12) \quad \alpha_x \leq \alpha_{\sigma(l)} \leq \alpha\left(\frac{\mu}{\sqrt{d}}\right)$$

for all $x \in V_l$, where we used that α defined in (8.5) is the inverse function of Λ . Since $\langle \mathbf{w}, \mathbf{1}_x \rangle = 0$, for all $x \in [N] \setminus V_l$ and $[N] \setminus V_l \subset \mathcal{V}_\tau$, we have

$$(8.13) \quad \langle \mathbf{w}, Q\mathbf{w} \rangle = \sum_{x \in [N] \setminus \mathcal{V}_\tau} \langle \mathbf{w}, \mathbf{1}_x \rangle^2 \alpha_x + \sum_{x \in \mathcal{V}_\tau \cap V_l} \langle \mathbf{w}, \mathbf{1}_x \rangle^2 \alpha_x \leq \tau + Cd^{-c((\Lambda(\alpha_{\sigma(l)})-2)\wedge 1)}$$

with very high probability due to Proposition 8.2, $V_l \cap \mathcal{V}_\tau \subset \mathcal{W}_{\tau, \mu, \zeta}$ and (8.12). Proposition 8.2 is applicable since ζ and μ satisfy (8.6) due to $\mu \geq \sqrt{d}(\Lambda(\alpha_{\sigma(l)}) + \zeta)$ and the lower bound on $\Lambda(\alpha_{\sigma(l)})$ implied by the definition of L_{\leq} .

We use the assumption (8.11) in (8.10), employ (8.13) and $\tau = 1 + d^{-\theta/4}$ and obtain

$$(8.14) \quad 2 + d^{-\theta/4} + Cd^{-c((\Lambda(\alpha_{\sigma(l)})-2)\wedge 1)} + \|E\| \geq \Lambda(\alpha_{\sigma(l)}) + \zeta - d^{-1/2} \|\underline{A}_\tau - \underline{A}\|$$

with very high probability. Thus, the bound on $\|E\|$ in Proposition 6.1 yields

$$(8.15) \quad d^{-\theta/4} + Cd^{-c((\Lambda(\alpha_{\sigma(l)})-2)\wedge 1)} + Cd^{-1/4-\theta/2} - (\Lambda(\alpha_{\sigma(l)}) - 2) \geq \zeta$$

with very high probability. Here, we used that

$$(8.16) \quad \|\underline{A}_\tau - \underline{A}\| \leq \|A_\tau - A\| + \|\mathbb{E}A - \Pi_\tau(\mathbb{E}A)\Pi_\tau\| \leq \max_{x \in [N]} D_x^{\mathbb{G} \setminus \mathbb{G}_\tau} + 2 = \mathcal{O}(d^{1/4-\theta/2})$$

with very high probability. This bound follows from Lemma 7.2(vi), $h(\varepsilon) \geq c\varepsilon^2$, $d \geq (\log N)^{4/(5-4\theta)}$, $\tau = 1 + d^{-\theta/4}$, $\theta \leq 1/2$ and

$$(8.17) \quad \|\mathbb{E}A - \Pi_\tau(\mathbb{E}A)\Pi_\tau\| \leq 2$$

with very high probability. For the proof of (8.17), we remark that, by construction of \mathbb{G}_τ , all balls $B_{r_{x,\tau}}^{\mathbb{G}_\tau}(x)$ for $x \in \mathcal{V}_\tau$ are disjoint, and we have

$$|\mathcal{Z}_\tau| = \sum_{x \in \mathcal{V}_\tau} |B_{r_{x,\tau}-2}^{\mathbb{G}_\tau}(x)| \leq \frac{2}{d^2} \sum_{x \in \mathcal{V}_\tau} |B_{r_{x,\tau}}^{\mathbb{G}_\tau}(x)| \leq \frac{2N}{d^2}$$

with very high probability, where we used Lemma 5.4 for the middle inequality. Thus, estimating the operator norm by the Hilbert–Schmidt norm yields

$$\|\mathbb{E}A - \Pi_\tau(\mathbb{E}A)\Pi_\tau\|^2 \leq \frac{d^2}{N^2} (|\mathcal{Z}_\tau|^2 + 2|\mathcal{Z}_\tau|(N - |\mathcal{Z}_\tau|)) \leq \frac{d^2 2|\mathcal{Z}_\tau|N}{N^2} \leq \frac{2d^2|\mathcal{Z}_\tau|}{N} \leq 4$$

with very high probability. This proves (8.17) and hence (8.15).

The definition of L_{\leq} in (8.1) and $l \in [L_{\leq}]$ imply $\alpha_{\sigma(l)} \geq 2 + (\log d)^{-\kappa}$ and, hence, $\Lambda(\alpha_{\sigma(l)}) \geq 2 + C(\log d)^{-2\kappa}$ for some $C > 0$. Therefore, we can bound the other error terms from (8.15) by $\frac{1}{2}(\Lambda(\alpha_{\sigma(l)}) - 2)$, multiply the result by 2 and obtain

$$(8.18) \quad Cd^{-c((\Lambda(\alpha_{\sigma(l)})-2)\wedge 1)} - (\Lambda(\alpha_{\sigma(l)}) - 2) \geq 2\zeta$$

with very high probability. Using $\alpha_{\sigma(l)} \geq 2 + (\log d)^{-\kappa}$ and $\Lambda(\alpha_{\sigma(l)}) \geq 2 + C(\log d)^{-2\kappa}$, we see that (8.18), however, contradicts (8.18) and $\zeta > 0$. Therefore, (8.11) is wrong, which implies part (i) of Proposition 8.1, due to (8.16), $\theta \leq 1/2$, and $\Lambda(\alpha_{\sigma(l)}) \geq 2 + C(\log d)^{-2\kappa}$.

We now prove (ii) assuming $L_{\leq} = 0$. We follow the proof of (i) with $l = 1$ and $V_1 = [N]$ and assume for the proof by contradiction that

$$\lambda_1(\underline{A}) > \sqrt{d}(2 + C\zeta) + \|\underline{A}_\tau - \underline{A}\|,$$

for some sufficiently large $C > 0$ and $\zeta := (\log d)^{-2\kappa}$.

Owing to (8.16), we have $\mu \geq \lambda_1(\underline{A}) - \|\underline{A}_\tau - \underline{A}\| \geq 2\sqrt{d} + (C - 1)(\log d)^{-2\kappa}$. Hence, ζ obviously satisfies (8.6a). Moreover, together with $\Lambda(\alpha_{\sigma(1)} \vee 2) \leq 2 + C(\log d)^{-2\kappa}$ for some universal constant $C > 0$, we obtain

$$\begin{aligned} \frac{\mu}{\sqrt{d}} - \Lambda(\alpha_{\sigma(1)} \vee 2) &\geq 2 + (C - 1)(\log d)^{-2\kappa} - (2 + C(\log d)^{-2\kappa}) \\ &\geq (C - 1 - C)(\log d)^{-2\kappa} \geq \zeta, \end{aligned}$$

where we assumed that $C \geq 2 + C$ in the last step. Hence, $\mathcal{V}_\tau \subset \mathcal{W}_{\tau, \mu, \zeta}$.

Similarly to the arguments in part (i), we obtain

$$Cd^{-c\zeta} - C\zeta \geq \zeta$$

as analogue of (8.18). This is a contradiction since $\zeta = (\log d)^{-2\kappa}$ which completes the proof of Proposition 8.1. \square

The condition $\mu \geq \sqrt{d}(\Lambda(\alpha_x \vee 2) + \zeta)$ in the definition of $\mathcal{W}_{\tau,\mu,\zeta}$ in (8.4) is an upper bound on α_x . In fact, we have $\alpha_x \leq \omega$ for all $x \in \mathcal{W}_{\tau,\mu,\zeta}$, where the parameter $\omega \equiv \omega(\mu, \zeta)$ is defined as the unique solution in $(2, \infty)$ of

$$(8.19) \quad \mu = \sqrt{d}(\Lambda(\omega) + \zeta).$$

For the following result we need the definition

$$(8.20) \quad r_\tau^\omega := \frac{\log N}{12 \log(\omega d)} \wedge \left(\frac{d}{4 \log d} h\left(\frac{\tau - 1}{2}\right) - 1 \right) - 2.$$

Note that $r_\tau^\omega \leq r_x/4$ if $x \in [N] \setminus \mathcal{V}_\omega$ and $r_\tau^\omega \leq r(\tau)/2$ with $r(\tau)$, as in (7.2). Hence, owing to Lemma 7.2(i), the balls $(B_r^{\mathbb{G}_\tau}(x))_{x \in \mathcal{V}_\tau \setminus \mathcal{V}_\omega}$ are disjoint for $r = r_\tau^\omega$.

LEMMA 8.3. *Let $(\log N)^{4/(5-2\theta)} \leq d \leq \mathcal{K}^{-1}N^{1/4}$ for some $\theta \in (0, 1/2]$, and let $1 + d^{-\theta/4} \leq \tau \leq 2$. Let $\zeta > 0$. Let $\mu > 2\sqrt{d}$ be an eigenvalue of $\underline{A}_{\tau,l}$. Suppose that ζ and μ satisfy (8.6). Let $\omega \geq 2$ be the unique solution of (8.19). Then, there exist an $r \in [r_\tau^\omega]$, a universal constant $c > 0$ and a constant $C > 0$ such that, for any $x \in \mathcal{W}_{\tau,\mu,\zeta}$ and any eigenvector \mathbf{w} of $\underline{A}_{\tau,\ell}$ associated to μ , we have*

$$\alpha(\mu/\sqrt{d}) \langle \mathbf{w}, \mathbf{1}_x \rangle^2 \leq Cd^{-c((\mu/\sqrt{d}-2) \wedge \log(\mu/\sqrt{d}))} \|\mathbf{w}|_{B_r^{\mathbb{G}_\tau}(x)}\|^2$$

with very high probability.

PROOF OF PROPOSITION 8.2. Let $\eta := \mu/\sqrt{d}$. Lemma 8.3 and $r_\tau^\omega \leq (\frac{1}{4}r_x) \wedge (\frac{1}{2}r(\tau))$ for $\alpha_x \leq \omega$ imply

$$\begin{aligned} \alpha(\eta) \sum_{x \in \mathcal{W}_{\tau,\mu,\zeta}} \langle \mathbf{w}, \mathbf{1}_x \rangle^2 &\leq Cd^{-c((\eta-2) \wedge \log(\eta))} \sum_{x \in \mathcal{W}_{\tau,\mu,\zeta}} \|\mathbf{w}|_{B_r^{\mathbb{G}_\tau}(x)}\|^2 \\ &\leq Cd^{-c((\eta-2) \wedge \log(\eta))} \|\mathbf{w}\|^2. \end{aligned}$$

Here, we used in the second step that $(B_r^{\mathbb{G}_\tau}(x))_{x \in \mathcal{W}_{\tau,\mu,\zeta}}$ are disjoint sets by Lemma 7.2(i). As \mathbf{w} is normalized, Proposition 8.2 follows. \square

The next subsection is devoted to the proof of Lemma 8.3.

8.1. *Proof of Lemma 8.3.* For the rest of this section, we fix $x \in \mathcal{W}_{\tau,\mu,\zeta}$ and omit it from our notation. For the proof of Lemma 8.3, we shall need some basic facts about the tridiagonalization of matrices which are summarized in Appendix A and which we refer to throughout this section. Throughout this section we only work with indices i of tridiagonal matrices satisfying $i \leq m$, where m was defined in Appendix A. This is always a simple application of (5.12b) in Lemma 5.4, and we shall not dwell on this issue further.

In the following result we compare a tridiagonalization of $\underline{A}_{\tau,l}$ with the tridiagonal matrix $Z(\alpha)$ for an appropriately chosen $\alpha > 0$. For $r \in \mathbb{N}$ and $\alpha > 0$, the matrix $Z(\alpha)$ is defined

through

$$(8.21) \quad Z(\alpha) = \begin{pmatrix} 0 & \sqrt{\alpha} & & & & \\ \sqrt{\alpha} & 0 & 1 & & & \\ & 1 & 0 & 1 & & \\ & & 1 & 0 & \ddots & \\ & & & \ddots & \ddots & 1 \\ & & & & 1 & 0 \end{pmatrix} \in \mathbb{R}^{(r+1) \times (r+1)}.$$

PROPOSITION 8.4. *Let $(\log N)^{4/(5-2\theta)} \leq d \leq \mathcal{K}^{-1}N^{1/4}$ for some $\theta \in (0, 5/2)$, and let $1 + d^{-\theta/4} \leq \tau \leq 2$. Let $r \leq r_\tau^\omega$ and $x \in \mathcal{V}_\tau$. For $k \in \mathbb{N}$, we define the error parameter*

$$(8.22) \quad \mathcal{E}_{\tau,k} := \frac{(3\sqrt{\tau} + 2)^k}{\sqrt{d}} \left[\left(\log d + \frac{\log N}{d} \right) \left(1 + \frac{\log N}{d\tau} \right) \right]^{\frac{1}{2}}.$$

Let $\widehat{Z} := Z(\underline{A}_{\tau,l,x})$ be the tridiagonal matrix associated with $\underline{A}_{\tau,l}$ around x , and (\mathbf{g}_k) the associated orthogonal basis (see Appendix A). Then, there exists a constant \mathcal{C} such that if $\mathcal{E}_{\tau,r} \leq \frac{1}{2\mathcal{C}}$, then we have for all $k \leq r$ with very high probability

$$(8.23) \quad \frac{\|\mathbf{g}_k - \mathbf{1}_{S_k^{\mathbb{G}_\tau}}\|}{\|\mathbf{1}_{S_k^{\mathbb{G}_\tau}}\|} \leq \mathcal{C}\mathcal{E}_{\tau,k}$$

and

$$(8.24) \quad \|\widehat{Z}_{\llbracket r \rrbracket} - \sqrt{d}Z(\alpha_x)\| = \mathcal{O}\left(\sqrt{d}\mathcal{E}_{\tau,r} + \frac{\log N}{\tau d^{3/2}} \left(1 + \frac{\log N}{(\tau - 1)^2 d} \right)\right).$$

We postpone the proof of Proposition 8.4 to the following subsection.

PROOF OF LEMMA 8.3. We denote the standard basis vectors of \mathbb{R}^{r+1} by $\mathbf{e}_0, \dots, \mathbf{e}_r$. Let $\widehat{Z} := Z(\underline{A}_{\tau,l,x})$ be the tridiagonal matrix associated with $\underline{A}_{\tau,l}$ around x , and (\mathbf{g}_k) the associated orthogonal basis (see Appendix A). Let \mathbf{w} be an eigenvector of $\underline{A}_{\tau,l}$ with eigenvalue μ . We denote by $\mathbf{b} = (b_k)_{k \in \llbracket N-1 \rrbracket}$ the vector representing \mathbf{w} with respect to the orthonormal basis $(\mathbf{g}_k / \|\mathbf{g}_k\|)_k$. Then, by the tridiagonal property of \widehat{Z} , we have $(\widehat{Z}_{\llbracket r \rrbracket} - \mu I_{r+1})\mathbf{b} \in \text{Span}\{\mathbf{e}_r\}$. Hence, we can apply Proposition C.2 with $\widetilde{Z} := \frac{1}{\sqrt{d}}\widehat{Z}_{\llbracket r \rrbracket}$ to estimate $b_0^2 / (\sum_{i=0}^r b_i^2)$ once we have verified its condition (C.9).

Because $\mathbb{G}_\tau|_{B_r^{\mathbb{G}_\tau}}$ is a tree, $\widetilde{Z}_{00} = \widetilde{Z}_{11} = 0$ by Lemma A.2 and $\widetilde{Z}_{01} = \sqrt{\alpha_x}$. From (8.24) with $r = 2$, $\tau \geq 1 + d^{-\theta/4}$, $\log N \leq d^{5/4-\theta/2}$ and $\theta \leq 1/2$, we conclude

$$(8.25) \quad |1 - \widetilde{Z}_{12}| \leq \varepsilon_2 := \mathcal{C}d^{-\theta/2}.$$

Throughout this proof, we need some Lipschitz-type bounds on $\Lambda(\alpha)$ and $\alpha(\eta)$. We have

$$(8.26) \quad \eta - \Lambda(\alpha \vee 2) \leq C \left(\frac{\alpha(\eta) - 2}{\alpha(\eta)^{3/2}} \right) (\alpha(\eta) - \alpha \vee 2)$$

for all $\alpha > 0$ and $\eta > \Lambda(\alpha \vee 2)$. The bound in (8.26) is a consequence of

$$\Lambda(\alpha(\eta)) - \Lambda(\alpha \vee 2) = \int_{\alpha \vee 2}^{\alpha(\eta)} \Lambda'(t) dt = \int_{\alpha \vee 2}^{\alpha(\eta)} \frac{t - 2}{2(t - 1)^{3/2}} dt$$

and distinguishing cases $\alpha > \alpha(\eta)/2$ and $\alpha \leq \alpha(\eta)/2$.

With the notation $\eta = \frac{\mu}{\sqrt{d}}$, we calculate δ introduced in (C.7) below. From (8.25) we conclude

$$(8.27) \quad |\delta(0, \sqrt{\alpha_x}, 0, \tilde{Z}_{12}, \eta)| = \frac{|\gamma(\eta)\eta^2 - \gamma(\eta)\alpha_x - \eta\tilde{Z}_{12}|}{|\alpha_x + \gamma(\eta)\eta\tilde{Z}_{12} - \eta^2|} \leq \frac{C\eta}{|\alpha_x - \alpha(\eta)|}$$

for some universal constant $C > 0$. Here, we used that the numerator is bounded from above by $C\eta$ due to $\gamma(\eta)\eta \leq 2$, (8.25) and $\sqrt{\alpha_x} \leq \Lambda(\alpha_x \vee 2) \leq \eta$ as $x \in \mathcal{W}_{\tau, \mu, \zeta}$. The denominator is bounded from below by $|\alpha_x - \alpha(\eta)| - \gamma(\eta)\eta|1 - \tilde{Z}_{12}|$ which proves (8.27) since $\alpha(\eta) - \alpha_x \geq \varepsilon_2 \geq 2\gamma(\eta)\eta|1 - \tilde{Z}_{12}|$. This last bounds are consequences of

$$(8.28) \quad \begin{aligned} \alpha(\eta) - \alpha_x \vee 2 &\geq C \frac{\alpha(\eta)^{3/2}}{\alpha(\eta) - 2} (\eta - \Lambda(\alpha_x \vee 2)) \\ &\geq Cd^{-\theta/2} \sqrt{\frac{\eta}{\eta - 2} \frac{\alpha(\eta)^{3/2}}{\alpha(\eta) - 2}} \geq Cd^{-\theta/2} \end{aligned}$$

as well as (8.25) and $\gamma(\eta)\eta \leq 2$. For the proof of (8.28), we use (8.26), $\eta - \Lambda(\alpha_x \vee 2) \geq \zeta$ and the first condition on ζ in (8.6a) and observe that $t \mapsto \frac{t}{t-2}$ and $t \mapsto \frac{t^{3/2}}{t-2}$ have strictly positive lower bounds for $t > 2$.

Let ε be defined as in Proposition C.2 below. Owing to Proposition 8.4, we know that $\varepsilon \leq \varepsilon_r$ with very high probability, where ε_r is defined through

$$(8.29) \quad \varepsilon_r := C \left(\mathcal{E}_{\tau, r} + \frac{\log N}{\tau d^2} \left(1 + \frac{\log N}{(\tau - 1)^2 d} \right) \right).$$

We now verify that the choice

$$(8.30) \quad r = (c \log(\zeta d^{1/5+\theta/2} \sqrt{(\eta - 2)/\eta}) + \tilde{C}) \wedge \frac{\log(d^{1/5+\theta/2}) - \log(4C)}{\log 8}$$

for some sufficiently small universal constant $c > 0$ and some $\tilde{C} > 0$ implies

$$(8.31) \quad \eta^2 \geq 4 + \frac{C(1 + \eta)^2 \varepsilon_r^2}{(1 - \gamma(\eta))^2} (1 + 1 \vee \delta^2), \quad \varepsilon_r \leq 1/2,$$

thus justifying the conditions of Proposition C.2. We remark that $r \geq 1$ due to the lower bound on ζ from (8.6a) and the lower bound on d . Clearly, $r \leq r_\tau^\omega$. From the definitions of ε_r and $\mathcal{E}_{\tau, r}$ in (8.29) and (8.22), respectively, as well as $1 + d^{-\theta/4} \leq \tau \leq 2$ and $\log N \leq d^{5/4-\theta/2}$, we conclude

$$(8.32) \quad \varepsilon_r \leq C \left(8^r d^{-1/5-\theta/2} + \frac{d^{-\theta}}{(\tau - 1)^2} \right) \leq C(8^r d^{-1/5-\theta/2} + d^{-\theta/2}).$$

Hence, the second bound on r in (8.30) yields $\varepsilon_r \leq 1/2$ in (8.31).

The first bound in (8.31) is equivalent to

$$(8.33) \quad \varepsilon_r^2 \leq \frac{(\eta^2 - 4)(1 - \gamma(\eta)^2)^2}{C\eta^2(1 \vee \delta^2)}.$$

Owing to $8^r \leq e^{r/c}$ for a sufficiently small universal $c > 0$, the definition of r in (8.30) and the first condition on ζ in (8.6a), we obtain from (8.32) that

$$\varepsilon_r \leq C \sqrt{\frac{\eta - 2}{\eta}} \zeta \leq C \sqrt{\frac{\eta - 2}{\eta} \frac{(\alpha(\eta) - 2)(\alpha(\eta) - \alpha_x)}{\eta^3}} \leq C \sqrt{\frac{\eta^2 - 4}{\eta^2} \frac{1 - \gamma(\eta)^2}{|\delta|}}.$$

Here, the second step is a consequence of $\zeta \leq \eta - \Lambda(\alpha_x \vee 2)$, (8.26) and $\alpha(\eta) \geq \eta^2/2$ for $\eta > 2$. The third step follows from $(\alpha(\eta) - 2)/\eta^2 \leq 1 - \gamma(\eta)^2$ and (8.27). This proves (8.33)

and thus the remaining estimate in (8.31). Hence, we have justified the conditions of Proposition C.2. It implies

$$\alpha(\eta) \frac{b_0^2}{\sum_{i=0}^r b_i^2} \leq \frac{C\alpha_x \alpha(\eta)}{(\alpha_x - \alpha(\eta))^2} (\gamma_{\geq})^{-2r} \leq \frac{C}{\zeta^2} \left(1 + \frac{1}{4}(\eta - 2)\right)^{-2(r-2)}.$$

Here, we employed $|\alpha(\eta) - \alpha_x| \geq 2\gamma(\eta)\eta|1 - \tilde{Z}_{12}|$ in the first step. In the second step we used $\zeta \leq C\alpha(\eta)^{-1/2}(\alpha(\eta) - \alpha_x)$ due to (8.26) and $x \in \mathcal{W}_{\tau, \mu, \zeta}$. We also used (C.10) as well as $\alpha_x \leq \alpha(\eta) \leq \eta^2$ due to $x \in \mathcal{W}_{\tau, \mu, \zeta}$.

Thus, our choice of r in (8.30), the first condition in (8.6a) and the definition of η yield

$$\begin{aligned} \alpha(\eta) \frac{b_0^2}{\sum_{i=0}^r b_i^2} &\leq \frac{C}{\zeta^2} \exp\left(-2(r-2) \log\left(1 + \frac{1}{4}(\eta - 2)\right)\right) \\ (8.34) \qquad &\leq \frac{C}{\zeta^2} \exp\left(-\frac{c}{2} \log d \log\left(1 + \frac{1}{4}(\eta - 2)\right)\right) \\ &\leq C \exp(-c \log d((\eta - 2) \wedge \log \eta)). \end{aligned}$$

Here, we used the second condition on ζ in (8.6a) for a sufficiently small c in the third step.

Finally, Lemma 8.3 follows from (8.34) since $b_0 = \langle \mathbf{w}, \mathbf{1}_x \rangle$ and

$$\|\mathbf{w}\|_{B_r^{\mathbb{G}_\tau}} \|^2 \geq \sum_{i=0}^r \left\langle \frac{\mathbf{g}_i}{\|\mathbf{g}_i\|}, \mathbf{w} \right\rangle^2 = \sum_{i=0}^r b_i^2.$$

Here, we used that $(\mathbf{g}_i)_{i \in \llbracket r \rrbracket}$ is a family of orthogonal vectors in $\text{Span}\{\mathbf{1}_y : y \in B_r^{\mathbb{G}_\tau}\}$ to obtain the previous inequality. \square

8.2. Proof of Proposition 8.4. PROOF OF PROPOSITION 8.4. We first remark that A_τ and $\underline{A}_{\tau, l}$ agree in the vicinity of $x \in \mathcal{V}_\tau \cap V_l$ in the sense that

$$(\underline{A}_{\tau, l})^i \mathbf{1}_x = (A_\tau)^i \mathbf{1}_x$$

for all $i \in \llbracket r_{x, \tau} - 2 \rrbracket$. This follows from first verifying the same identity with A_τ replaced by \underline{A}_τ and then using that the shift in the definition of \underline{A}_τ in (8.3) vanishes on $B_{r_{x, \tau} - 2}^{\mathbb{G}_\tau}(x)$.

For the proof of (8.23), we now introduce a second family $(\mathbf{f}_k)_k$ of vectors that will turn out to be a good approximation of $(\mathbf{g}_k)_k$. The vectors \mathbf{f}_k are defined through

$$\mathbf{f}_0 = \mathbf{1}_x, \quad \mathbf{f}_1 = \mathbf{1}_{S_1^{\mathbb{G}_\tau}}, \quad \mathbf{f}_2 = \mathbf{1}_{S_2^{\mathbb{G}_\tau}}, \quad \mathbf{f}_{k+2} = Q_0(A_\tau \mathbf{f}_{k+1} - d\mathbf{f}_k)$$

for all $k \geq 1$. Here and in the following, Q_i denotes the orthogonal projection on $\text{Span}\{(A_\tau)^j \mathbf{1}_x : j \in \llbracket i \rrbracket\}^\perp$, as in Appendix A.

The careful analysis of \mathbf{f}_k presented below will imply (8.23) due to the bound

$$(8.35) \qquad \|\mathbf{g}_k - \mathbf{1}_{S_k^{\mathbb{G}_\tau}}\| \leq \|\mathbf{q}_k\|,$$

where we introduced

$$\mathbf{q}_k := \mathbf{f}_k - \mathbf{1}_{S_k^{\mathbb{G}_\tau}}.$$

Before estimating \mathbf{q}_k , we now establish (8.35). It is easy to check that there exists a monic polynomial P_k of degree k such that $\mathbf{f}_k = P_k(A_\tau)\mathbf{1}_x$ and then

$$(8.36) \qquad \mathbf{g}_k = Q_{k-1}(A_\tau^k \mathbf{1}_x) = Q_{k-1}(P_k(A_\tau)\mathbf{1}_x) = Q_{k-1}\mathbf{f}_k.$$

Hence, $\|\mathbf{g}_k\| \leq \|\mathbf{f}_k\|$ and, thus, we have

$$\|\mathbf{g}_k - \mathbf{1}_{S_k^{\mathbb{G}_\tau}}\|^2 = \|\mathbf{g}_k\|^2 - \|\mathbf{1}_{S_k^{\mathbb{G}_\tau}}\|^2 \leq \|\mathbf{f}_k\|^2 - \|\mathbf{1}_{S_k^{\mathbb{G}_\tau}}\|^2 = \|\mathbf{q}_k\|^2.$$

Here, we used in the first step that $\mathbf{g}_k - \mathbf{1}_{S_k^{\mathbb{G}_\tau}}$ is orthogonal to $\mathbf{1}_{S_k^{\mathbb{G}_\tau}}$. This is a consequence of $\text{supp}(\mathbf{g}_k - \mathbf{1}_{S_k^{\mathbb{G}_\tau}}) \subset B_{k-2}^{\mathbb{G}_\tau}$ by Lemma A.4. In the last step we used $\mathbf{q}_k \perp \mathbf{1}_{S_k^{\mathbb{G}_\tau}}$, which follows from $\text{supp } \mathbf{q}_k \subset B_{k-1}^{\mathbb{G}_\tau}$, a consequence of $\langle \mathbf{f}_k, \mathbf{1}_y \rangle = \langle \mathbf{g}_k, \mathbf{1}_y \rangle = 1$ for all $y \in S_k^{\mathbb{G}_\tau}$ by (8.36). This shows (8.35).

Owing to (8.35), the bound in (8.23) follows directly from

$$(8.37) \quad \|\mathbf{q}_k\| \leq \tilde{c}_k \|\mathbf{1}_{S_k^{\mathbb{G}_\tau}}\|,$$

which holds with very high probability for all $k \leq r_{x,\tau} - 2$, as we shall show below. Here, \tilde{c}_k is the unique solution of

$$\tilde{c}_{k+2} = 2\tilde{c}_k + 3\sqrt{\tau}\tilde{c}_{k+1} + \frac{C}{\sqrt{d}} \left(\left(\log d + \frac{\log N}{d} \right)^{1/2} \left(1 + \frac{\log N}{D_x} \right)^{1/2} \right)$$

with the initial choices $\tilde{c}_1 = 0$ and $\tilde{c}_2 = 0$.

We now prove (8.37) by induction on k . The induction basis for $k = 1$ and $k = 2$ is trivial. For the induction step we decompose

$$\begin{aligned} A_\tau \mathbf{f}_{k+1} - d\mathbf{f}_k &= A_\tau (\mathbf{1}_{S_{k+1}^{\mathbb{G}_\tau}} + \mathbf{q}_{k+1}) - d(\mathbf{1}_{S_k^{\mathbb{G}_\tau}} + \mathbf{q}_k) \\ &= \mathbf{1}_{S_{k+2}^{\mathbb{G}_\tau}} + \sum_{y \in S_k^{\mathbb{G}_\tau}} (N_{k+1}(y) - d)\mathbf{1}_y + A_\tau \mathbf{q}_{k+1} - d\mathbf{q}_k \\ &= \mathbf{1}_{S_{k+2}^{\mathbb{G}_\tau}} + \sum_{y \in S_k^{\mathbb{G}_\tau}} \left(N_{k+1}(y) - \frac{|S_{k+1}|}{|S_k|} \right) \mathbf{1}_y + \left(\frac{|S_{k+1}|}{|S_k|} - d \right) \mathbf{1}_{S_k^{\mathbb{G}_\tau}} \\ &\quad + A_\tau \mathbf{q}_{k+1} - d\mathbf{q}_k. \end{aligned}$$

Here, we used in the second step that $\langle \mathbf{1}_y, A_\tau \mathbf{1}_{S_{k+1}^{\mathbb{G}_\tau}} \rangle = |S_1^{\mathbb{G}_\tau}(y) \cap S_{k+1}^{\mathbb{G}_\tau}| = N_{k+1}(y)$ for all $y \in S_k^{\mathbb{G}_\tau}$ by Lemma 7.2(v) as $k > 1$.

Therefore, following the proof of (5.9c) yields

$$(8.38) \quad \begin{aligned} &\frac{1}{\sqrt{|S_{k+1}|}} \left\| \sum_{y \in S_k^{\mathbb{G}_\tau}} \left(N_{k+1}(y) - \frac{|S_{k+1}|}{|S_k|} \right) \mathbf{1}_y \right\| \\ &\leq \frac{1}{\sqrt{|S_{k+1}|}} \left\| \sum_{y \in S_k} \left(N_{k+1}(y) - \frac{|S_{k+1}|}{|S_k|} \right) \mathbf{1}_y \right\| \\ &= \mathcal{O} \left(\left(\log d + \frac{\log N}{d} \right)^{1/2} \left(1 + \frac{\log N}{D_x} \right)^{1/2} \right) \end{aligned}$$

with very high probability. Moreover, using Lemma 5.4 and Lemma 7.2(iv), we get

$$\frac{1}{\sqrt{|S_{k+1}|}} \left\| \left(\frac{|S_{k+1}|}{|S_k|} - d \right) \mathbf{1}_{S_k^{\mathbb{G}_\tau}} \right\| \leq \frac{\sqrt{|S_k|}}{\sqrt{|S_{k+1}|}} \left| \frac{|S_{k+1}|}{|S_k|} - d \right| \leq \mathcal{O} \left(\left[\frac{\log N}{dD_x} \right]^{\frac{1}{2}} \right)$$

for $k \geq 2$ which is smaller than the right-hand side of (8.38). Since A_τ has degree at most τd on $B_r^{\mathbb{G}_\tau} \setminus \{x\}$, then $\|A_\tau \mathbf{v}\| \leq 2\sqrt{\tau d} \|\mathbf{v}\|$ for all \mathbf{v} with $\text{supp}(\mathbf{v}) \subset B_r^{\mathbb{G}_\tau}$ and $\langle \mathbf{v}, \mathbf{1}_x \rangle = 0$ [16, Chapter 11, Ex. 14]. Therefore, we have

$$\|A_\tau \mathbf{q}_{k+1}\| = \|A_\tau Q_0 \mathbf{q}_{k+1}\| \leq 2\sqrt{\tau d} \|\mathbf{q}_{k+1}\|.$$

We put everything together and get

$$\begin{aligned} \|\mathbf{q}_{k+2}\| &= \|Q_0(\mathbf{f}_{k+2} - \mathbf{1}_{S_{k+2}^{\mathbb{G}_\tau})}\| \\ &\leq d\|\mathbf{q}_k\| + 2\sqrt{\tau d}\|\mathbf{q}_{k+1}\| + \sqrt{|S_{k+1}|}\mathcal{O}\left(\left[\left(\log d + \frac{\log N}{d}\right)\left(1 + \frac{\log N}{D_x}\right)\right]^{\frac{1}{2}}\right) \end{aligned}$$

with very high probability. We set $c_k = \frac{\|\mathbf{q}_k\|}{\sqrt{|S_k|}}$. Thus, the previous estimate implies

$$\begin{aligned} c_{k+2} &\leq \frac{d\sqrt{|S_k|}}{\sqrt{|S_{k+2}|}}c_k + \frac{2\sqrt{\tau d|S_{k+1}|}}{\sqrt{|S_{k+2}|}}c_{k+1} \\ &\quad + \frac{\sqrt{|S_{k+1}|}}{\sqrt{|S_{k+2}|}}\mathcal{O}\left(\left[\left(\log d + \frac{\log N}{d}\right)\left(1 + \frac{\log N}{D_x}\right)\right]^{\frac{1}{2}}\right). \end{aligned}$$

We use Lemma 5.4 and obtain

$$\begin{aligned} c_{k+2} &\leq (1 + o(1))c_k + 2\sqrt{\tau}(1 + o(1))c_{k+1} \\ &\quad + \frac{1}{\sqrt{d}}\mathcal{O}\left(\left[\left(\log d + \frac{\log N}{d}\right)\left(1 + \frac{\log N}{D_x}\right)\right]^{\frac{1}{2}}\right). \end{aligned}$$

This completes the induction step. Thus, we have proved (8.37) and hence (8.23) as well.

We now verify (8.24). We have

$$\begin{aligned} \|\mathbf{g}_i\|^2 &= |S_i|\left(1 + \left(\frac{\|\mathbf{g}_i\|^2}{|S_i^{\mathbb{G}_\tau}|} - 1\right)\right)\left(1 + \left(\frac{|S_i^{\mathbb{G}_\tau}|}{|S_i|} - 1\right)\right) \\ &= |S_i|\left(1 + \mathcal{O}\left(\mathcal{E}_{\tau,r}^2 + \frac{\log N}{\tau d^2}\left(1 + \frac{\log N}{(\tau - 1)^2 d}\right)\right)\right). \end{aligned}$$

Here, we used (8.23) and the orthogonality of \mathbf{g}_i and $\mathbf{g}_i - \mathbf{1}_{S_i^{\mathbb{G}_\tau}}$ to see that the second factor equals $1 + \mathcal{O}(\mathcal{E}_{\tau,r}^2)$. To estimate the third factor, we used (7.3), Lemma 7.2(iv), (5.12b) and $D_x \geq \tau d$. Using the assumption on $\mathcal{E}_{\tau,r}$, the lower bound $\tau \geq 1 + d^{-\theta/4}$ and the lower bound on d , we see that error term in this expansion of $\|\mathbf{g}_i\|^2$ is smaller than 1. Therefore, denoting the entries of \hat{Z} by \hat{Z}_{ij} and using Lemma A.2 yield

$$\hat{Z}_{i\ i+1} = \frac{\|\mathbf{g}_{i+1}\|}{\|\mathbf{g}_i\|} = \sqrt{\frac{|S_{i+1}|}{|S_i|}}\left(1 + \mathcal{O}\left(\mathcal{E}_{\tau,r}^2 + \frac{\log N}{\tau d^2}\left(1 + \frac{\log N}{(\tau - 1)^2 d}\right)\right)\right).$$

Therefore, (5.12a) and $\hat{Z}_{ii} = 0$ by Lemma A.3(ii) complete the proof of (8.24) and thus the one of Proposition 8.4. \square

9. Proofs of the results in Section 2. In this short section we state how to conclude the results of Section 2. For $d \leq \exp(\sqrt{\log N}/4)$, Theorem 2.1 follows from Propositions 7.1 and 8.1, noting that $L = L_\leq \leq L_\geq$. For $\exp(\sqrt{\log N}/4) \leq d \leq N/2$, Theorem 2.1 follows immediately from [4], (2.4) and Theorem 2.6. Corollary 2.3 follows from Theorem 2.1 by eigenvalue interlacing $\lambda_l(\underline{A}) \geq \lambda_{l+1}(A) \geq \lambda_{l+1}(\underline{A})$ for $1 \leq l \leq N - 1$ as well as the mean value theorem. Finally, the proof of Theorem 2.7 is very similar to that of Theorem 2.1, and we explain the needed minor modifications in the next section.

10. Modifications for sparse Wigner matrices. In this section, we explain how the arguments in the previous sections can be adapted to yield the proof of Theorem 2.7. We consider a sparse Wigner matrix M with entries $M_{xy} = W_{xy}A_{xy}$. Here, A is the adjacency matrix of an Erdős–Rényi graph \mathbb{G} on $[N]$ with edge probability d/N and $W = (W_{xy})_{x,y \in [N]}$ is an independent Wigner matrix with bounded entries. That is, W is Hermitian and the random variables $(W_{xy} : x \leq y)$ are independent and

$$(10.1) \quad \mathbb{E}W_{xy} = 0, \quad \mathbb{E}|W_{xy}|^2 = 1, \quad |W_{xy}| \leq K$$

for all $x, y \in [N]$ and some constant $K > 0$.

The assumptions imply that M is symmetric, and we consider M as the adjacency matrix of an undirected weighted graph with edge weights W_{xy} . According to this philosophy, we define

$$S_i(x) := \{y \in [N] : \min\{j \geq 0 : (M^j)_{xy} \neq 0\} = i\}, \quad B_i(x) := \bigcup_{j \in [i]} S_j(x)$$

for all $x \in [N]$.

In the remainder of this section, we explain the necessary adjustments in order to conclude Theorem 2.7 along the proof of Theorem 2.1 with the definition of α_x from (2.5) and $D_x = \alpha_x d$. Throughout this section the constant \mathcal{C} as well as the implicit constant in \mathcal{O} are also allowed to depend on K , the uniform bound on W_{xy} in (10.1). With this convention the arguments in Section 6 to Section 8 do not require any changes. They only have to be understood with respect to the new definition of \mathcal{C} and \mathcal{O} . The necessary modifications of Section 5 are explained in the following subsection. Once they are taken into account, Theorem 2.7 follows from Propositions 7.1 and 8.1.

10.1. *Modifications in Section 5.* In this subsection we fix $x \in [N]$ and explain the modifications required in Section 5 to obtain the corresponding results in the setup described above.

Definition of the approximate eigenvector, decomposition of the error terms. We now introduce the analogue of the approximate eigenvector \mathbf{v} from (5.3) in the present setup. We define $\mathbf{g}_0 := \mathbf{1}_x$. For $i \geq 1$, we define

$$\mathbf{g}_i := (M\mathbf{g}_{i-1})|_{S_i(x)}.$$

Note that $\mathbf{g}_0, \dots, \mathbf{g}_i$ are orthogonal. As a heuristic for the following argument, suppose that, for some $r \geq 1$, the graph \mathbb{G} restricted to $B_r(x)$ is a tree and that, for all $i \in [r]$, we have $\sum_{z \in S_{i+1}(x)} \langle \mathbf{1}_z, M\mathbf{1}_y \rangle^2 = d$ for all $y \in S_i(x)$; then the upper-left $(r+1) \times (r+1)$ block of the tridiagonal matrix associated with M around x (see Appendix B) is equal to the right-hand side of (B.1) with $D_x = d\alpha_x$ and α_x given by (2.5). This motivates the construction of \mathbf{v} in the following paragraph.

With the choices of u_i from (5.2), we set

$$\mathbf{v} = \sum_{i=0}^r u_i \hat{\mathbf{g}}_i, \quad \hat{\mathbf{g}}_i := \frac{\mathbf{g}_i}{\|\mathbf{g}_i\|}.$$

Similarly to the proof of Lemma 5.2, we obtain

$$(M - \sqrt{d}\Lambda(\alpha_x))\mathbf{v} = \mathbf{w}_1 + \dots + \mathbf{w}_4,$$

where the error terms $\mathbf{w}_1, \dots, \mathbf{w}_4$ are defined through

$$\begin{aligned} \mathbf{w}_1 &:= \sum_{i=0}^r \frac{u_i}{\|\mathbf{g}_i\|} \left(\sum_{y \in \mathcal{S}_{i+1}} (N_i(y) - \langle \mathbf{1}_y, \mathbf{g}_{i+1} \rangle) \mathbf{1}_y + \sum_{y \in \mathcal{S}_i} N_i(y) \mathbf{1}_y \right), \\ \mathbf{w}_2 &:= \sum_{i=1}^r \frac{u_i}{\|\mathbf{g}_i\|} \left(\sum_{y \in \mathcal{S}_{i-1}} N_i(y) \mathbf{1}_y - \frac{\|\mathbf{g}_i\|^2}{\|\mathbf{g}_{i-1}\|^2} \mathbf{g}_{i-1} \right), \\ \mathbf{w}_3 &:= u_2 \left(\frac{\|\mathbf{g}_2\|}{\|\mathbf{g}_1\|} - \sqrt{d} \right) \hat{\mathbf{g}}_1 + \sum_{i=2}^{r-1} \left[u_{i+1} \left(\frac{\|\mathbf{g}_{i+1}\|}{\|\mathbf{g}_i\|} - \sqrt{d} \right) + u_{i-1} \left(\frac{\|\mathbf{g}_i\|}{\|\mathbf{g}_{i-1}\|} - \sqrt{d} \right) \right] \hat{\mathbf{g}}_i, \\ \mathbf{w}_4 &:= \left(u_{r-1} \frac{\|\mathbf{g}_r\|}{\|\mathbf{g}_{r-1}\|} - u_{r-1} \sqrt{d} - u_{r+1} \sqrt{d} \right) \hat{\mathbf{g}}_r + u_r \frac{\|\mathbf{g}_{r+1}\|}{\|\mathbf{g}_r\|} \hat{\mathbf{g}}_{r+1}. \end{aligned}$$

Here, $N_i(y) := \langle \mathbf{1}_y, M \mathbf{g}_i \rangle$ for all $y \in [N]$. We remark that the analogue of \mathbf{w}_0 vanishes as the entries W_{yz} are centred for all $y, z \in [N]$.

Concentration of $\|\mathbf{g}_{i+1}\|/\|\mathbf{g}_i\|$. In order to establish that the ratio $\frac{\|\mathbf{g}_{i+1}\|}{\|\mathbf{g}_i\|}$ concentrates around \sqrt{d} if $i \geq 1$, we follow the proof of Lemma 5.4. It suffices to verify (5.17) in the new setup. We first prove by induction that if K is the uniform upper bound on the entries of W , then

$$(10.2) \quad \|\mathbf{g}_i\|_\infty \leq (CK)^i$$

for $i \geq 0$ with very high probability. The case $i = 0$ holds trivially. Since $\text{supp } \mathbf{g}_{i-1} \subset \mathcal{S}_{i-1}$ by definition the definition of \mathbf{g}_i implies for $y \in \mathcal{S}_i$ that

$$\begin{aligned} \left| \sum_{z \in [N]} M_{yz} \langle \mathbf{1}_z, \mathbf{g}_{i-1} \rangle \right| &= \left| \sum_{z \in \mathcal{S}_{i-1}} W_{yz} A_{yz} \langle \mathbf{1}_z, \mathbf{g}_{i-1} \rangle \right| \\ &\leq \|\mathbf{g}_{i-1}\|_\infty K N_{i-1}(y) \leq \|\mathbf{g}_{i-1}\|_\infty K C, \end{aligned}$$

where we used Corollary 5.6 in the last step. As $\text{supp } \mathbf{g}_i \subset \mathcal{S}_i$ by definition, this proves (10.2).

Using Bennett’s inequality, it is easy to see that

$$(10.3) \quad \left| \frac{\|\mathbf{g}_{i+1}\|^2}{d \|\mathbf{g}_i\|^2} - 1 \right| \leq C(1 + K^2) \sqrt{\frac{\log N}{d}} \frac{\|\mathbf{g}_i\|_4^2}{\|\mathbf{g}_i\|^2}$$

with very high probability. Using $\|\mathbf{g}_i\|_4^2 \leq (CK)^i \|\mathbf{g}_i\|_2$ by (10.2) in (10.3) yields the desired analogue of (5.17) in the setup of sparse Wigner matrices. This proves the analogue of Lemma 5.4.

Estimate on \mathbf{w}_1 . We remark that the $i = 0$ contribution in the definition of \mathbf{w}_1 vanishes as $N_0(y) = \langle \mathbf{1}_y, M \mathbf{g}_0 \rangle = \langle \mathbf{1}_y, \mathbf{g}_1 \rangle$ for any y as $A_{xx} = 0$. Moreover, $A_{xx} = 0$ also implies that $N_0(y) = 0$ for $y \in \mathcal{S}_0$. Hence,

$$\|\mathbf{w}_1\|^2 = \left\| \sum_{i=1}^r \frac{u_i}{\|\mathbf{g}_i\|} \sum_{y \in \mathcal{S}_i} N_i(y) \mathbf{1}_y \right\|^2 = \sum_{i=1}^r \frac{u_i^2}{\|\mathbf{g}_i\|^2} \sum_{y \in \mathcal{S}_i} N_i(y)^2.$$

Here, we also used that $N_i(y) - \langle \mathbf{1}_y, \mathbf{g}_{i+1} \rangle = \langle \mathbf{1}_y, (M \mathbf{g}_i)|_{[N] \setminus \mathcal{S}_{i+1}(x)} \rangle = 0$ for any $y \in \mathcal{S}_{i+1}$ due to the fact that $(M \mathbf{g}_i)|_{[N] \setminus \mathcal{S}_{i+1}(x)}$ vanishes on $\mathcal{S}_{i+1}(x)$.

Thus, in order to estimate $\|\mathbf{w}_1\|^2$, we use the following version of (5.23) in Corollary 5.7. Namely, for all $i \geq 1$, the bounds

$$\begin{aligned} \sum_{y \in \mathcal{S}_i(x)} N_i(y)^2 &= \sum_{y \in \mathcal{S}_i(x)} \langle \mathbf{1}_y, M \mathbf{g}_i \rangle^2 \\ &= \sum_{y \in \mathcal{S}_i(x)} \left(\sum_{y_1 \in \mathcal{S}_i(x)} \langle M \mathbf{1}_y, \mathbf{1}_{y_1} \rangle \langle \mathbf{1}_{y_1}, \mathbf{g}_i \rangle \right)^2 \leq C \|\mathbf{g}_i\|_\infty^2 \end{aligned}$$

hold with very high probability. In the last step we used Lemma 5.5 to conclude that there are at most $\mathcal{O}(1)$ many nonzero terms. Therefore, (10.2) yields (5.9b) in the current setup due to the growth of $\|\mathbf{g}_i\|$ by the analogue of (5.12b).

Estimate on \mathbf{w}_2 . Here, we follow the proof of (5.9c). By the Pythagorean theorem we have

$$\|\mathbf{w}_2\|^2 = \sum_{i=1}^{r-1} \frac{u_{i+1}^2}{\|\mathbf{g}_{i+1}\|^2} \sum_{y \in \mathcal{S}_i} \left(N_{i+1}(y) - \frac{\|\mathbf{g}_{i+1}\|^2}{\|\mathbf{g}_i\|^2} \langle \mathbf{1}_y, \mathbf{g}_i \rangle \right)^2,$$

where we used that $\mathcal{S}_0 = \{x\}$, $N_1(x) = \langle M \mathbf{1}_x, \mathbf{g}_1 \rangle = \|\mathbf{g}_1\|^2$ as $\mathbf{g}_1 = M \mathbf{1}_x$ due to $A_{xx} = 0$. As $\sum_{i=1}^{r-1} u_{i+1}^2 \leq 1$, we obtain

$$\|\mathbf{w}_2\|^2 \leq \max_{i \in [r-1]} \frac{1}{\|\mathbf{g}_{i+1}\|^2} \sum_{y \in \mathcal{S}_i} \left(N_{i+1}(y) - \frac{\|\mathbf{g}_{i+1}\|^2}{\|\mathbf{g}_i\|^2} \langle \mathbf{1}_y, \mathbf{g}_i \rangle \right)^2 \leq \frac{4}{d} \max_{i \in [r-1]} (Y_i + \tilde{X}_i),$$

where we introduced

$$\begin{aligned} Y_i &:= \frac{1}{\|\mathbf{g}_i\|^2} \sum_{y \in \mathcal{S}_i} (N_{i+1}(y) - \mathbb{E}[N_{i+1}(y) | M_{(B_{i-1})}])^2, \\ \tilde{X}_i &:= \frac{1}{\|\mathbf{g}_i\|^2} \sum_{y \in \mathcal{S}_i} \left(\mathbb{E}[N_{i+1}(y) | M_{(B_{i-1})}] - \frac{\|\mathbf{g}_{i+1}\|^2}{\|\mathbf{g}_i\|^2} \langle \mathbf{1}_y, \mathbf{g}_i \rangle \right)^2. \end{aligned}$$

We first estimate \tilde{X}_i . As $\mathbb{E}[N_{i+1}(y) | M_{(B_{i-1})}] = \langle \mathbf{1}_y, \mathbf{g}_i \rangle d(1 - \frac{|B_i|}{N})$, we conclude

$$\begin{aligned} \tilde{X}_i &= \frac{1}{\|\mathbf{g}_i\|^2} \left(d - \frac{\|\mathbf{g}_{i+1}\|^2}{\|\mathbf{g}_i\|^2} - \frac{d|B_i|}{N} \right)^2 \sum_{y \in \mathcal{S}_i} \langle \mathbf{1}_y, \mathbf{g}_i \rangle^2 \\ &\leq 2 \left(\left(d - \frac{\|\mathbf{g}_{i+1}\|^2}{\|\mathbf{g}_i\|^2} \right)^2 + 1 \right) \leq C \left(\frac{d \log N}{D_x} + 1 \right) \end{aligned}$$

with very high probability for all $i \in [r - 1]$.

In order to estimate Y_i , we follow the proof of (5.26) and explain the necessary changes. We redefine

$$E_y := \frac{1}{\langle \mathbf{1}_y, \mathbf{g}_i \rangle} (N_{i+1}(y) - \mathbb{E}[N_{i+1}(y) | M_{(B_{i-1})}])$$

and use Bennett’s inequality to obtain

$$\mathbb{P}(E_y^2 > (s\kappa d)^2 | M_{(B_{i-1})}) \leq \exp(-c\kappa d(s \wedge s^2)),$$

where $\kappa = \mathbb{E}W_{yz}^4$ for some $y, z \in [N]$. Hence, using this bound in the proof of (5.28) yields

$$\mathbb{P}(L_s^i \geq \ell | M_{(B_{i-1})}) \leq \left(\frac{\|\mathbf{g}_i\|^2}{\ell} \right) \exp(-c\kappa d\ell(s \wedge s^2)).$$

Applying this estimate in the remainder of the proof of (5.26), we deduce

$$|Y_i| \leq C(1 + K)^{2i} d \left(1 + \frac{\log N}{\|\mathbf{g}_i\|^2} \right) \left(\log d + \frac{\log N}{d} \right).$$

Here, we employed that $\|\mathbf{g}_i\|_\infty \leq C(1 + K)^i$. Therefore, using the growth of $\|\mathbf{g}_i\|$, we obtain the same bound on $\|\mathbf{w}_2\|$ as in (5.9c).

When following the proof of (5.9c) in the proof of Proposition 8.4, the same adjustments yield the bound used there.

APPENDIX A: TRIDIAGONALIZATION

Let $M \in \mathbb{R}^{N \times N}$ be a symmetric matrix and $x \in [N]$. Let $m(x) := \dim \text{Span}\{M^n \mathbf{1}_x : n \in \mathbb{N}_0\}$. For $i \in \llbracket m - 2 \rrbracket$, define by induction

$$\mathbf{g}_0 := \mathbf{1}_x, \quad \mathbf{g}_{i+1} := Q_i M \mathbf{g}_i,$$

where Q_i is the orthogonal projection onto the orthogonal complement of $\text{Span}\{M^j \mathbf{1}_x : j \in \llbracket i \rrbracket\}$. We call $(\mathbf{g}_i)_{i \in \llbracket m-1 \rrbracket}$ the orthogonal basis associated with M and x . Note that this basis is, in general, not normalized. For convenience, if $m < N - 1$, that is, $\mathbf{1}_x$ is not a cyclic vector of M , we complete the basis $(\mathbf{g}_i)_{i \in \llbracket m-1 \rrbracket}$ to an orthogonal basis $(\mathbf{g}_i)_{i \in \llbracket N-1 \rrbracket}$ of \mathbb{R}^N in an arbitrary fashion.

We define $Z \equiv Z^{(M,x)}$ as the matrix M in the orthonormal basis $(\mathbf{g}_i / \|\mathbf{g}_i\|)_{i \in \llbracket N-1 \rrbracket}$; that is,

$$Z_{ij} := \left\langle \frac{\mathbf{g}_i}{\|\mathbf{g}_i\|}, M \frac{\mathbf{g}_j}{\|\mathbf{g}_j\|} \right\rangle$$

for $i, j \in \llbracket N - 1 \rrbracket$.

REMARK A.1. It is easy to check that the matrix $Z_{\llbracket m-1 \rrbracket}$ is tridiagonal, that is, $Z_{ij} = 0$ if $|i - j| > 1$ and $i, j \in \llbracket m - 1 \rrbracket$. Hence, we call Z the *tridiagonal matrix* associated with M around x .

LEMMA A.2. If Z is the tridiagonal matrix of M , then

$$Z_{ii+1} = \frac{\|\mathbf{g}_{i+1}\|}{\|\mathbf{g}_i\|}$$

for $i \in \llbracket m - 2 \rrbracket$.

PROOF. We have

$$Z_{ii+1} = \frac{\langle M \mathbf{g}_i, \mathbf{g}_{i+1} \rangle}{\|\mathbf{g}_i\| \|\mathbf{g}_{i+1}\|} = \frac{\langle M \mathbf{g}_i, Q_i M \mathbf{g}_i \rangle}{\|\mathbf{g}_i\| \|\mathbf{g}_{i+1}\|} = \frac{\langle Q_i M \mathbf{g}_i, Q_i M \mathbf{g}_i \rangle}{\|\mathbf{g}_i\| \|\mathbf{g}_{i+1}\|} = \frac{\|\mathbf{g}_{i+1}\|}{\|\mathbf{g}_i\|}. \quad \square$$

LEMMA A.3. Let $M = A$ be the adjacency matrix of a bipartite graph (e.g., a tree) with vertex set $V_0 \cup V_1$ such that $A_{V_0} = 0$ and $A_{V_1} = 0$. Then, the following holds:

- (i) If $x \in V_0$, then $\mathbf{g}_i|_{V_{(i+1) \bmod 2}}$ vanishes for all i .
- (ii) The diagonal of the associated tridiagonal matrix Z vanishes.

PROOF. Part (i) follows directly from the bipartite structure of the graph. Part (ii) is immediate from $Z_{ii} = \langle A \mathbf{g}_i, \mathbf{g}_i \rangle / \|\mathbf{g}_i\|^2$ and the first part. \square

LEMMA A.4. If $M = A$ is the adjacency matrix of a tree, then

$$\mathbf{g}_i = \mathbf{1}_{S_i} + \mathbf{d}_i$$

with some $\mathbf{d}_i \in \mathbb{R}^N$ satisfying $\mathbf{d}_0 = \mathbf{d}_1 = 0$ and $\text{supp } \mathbf{d}_i \subset B_{i-2}$ for all $2 \leq i \leq m$.

as well as

$$(B.3) \quad A\mathbf{1}_{S_l(x)} = \begin{cases} \mathbf{1}_{S_1(x)}, & \text{if } l = 0, \\ \mathbf{1}_{S_2(x)} + D_x\mathbf{1}_x, & \text{if } l = 1, \\ \mathbf{1}_{S_{l+1}(x)} + d\mathbf{1}_{S_{l-1}(x)}, & \text{if } l \in [r-1] \setminus \{1\}. \end{cases}$$

Therefore, (B.2) follows immediately as $\mathbf{1}_{S_i(x)}$ and $\mathbf{1}_{S_j(x)}$ are orthogonal for $i \neq j$ and $\|\mathbf{1}_{S_i(x)}\| = |S_i(x)|^{1/2}$.

We start the proof of (ii) by concluding

$$(B.4) \quad \langle \mathbf{s}_i, A\mathbf{s}_j \rangle = |S_i(x)|^{-1/2} |S_j(x)|^{-1/2} \langle \mathbf{1}_{S_i(x)}, A\mathbf{1}_{S_j(x)} \rangle$$

for $i, j = 0, \dots, r$ from (i). If $|i - j| \neq 1$, then this immediately yields $\langle \mathbf{s}_i, A\mathbf{s}_j \rangle = 0$. Moreover, $|S_0(x)| = 1$ and $|S_i(x)| = D_x d^{i-1}$ for $i \geq 1$ to (B.4). For all $i, j = 0, 1, \dots, r$, we have

$$\langle \mathbf{s}_i, A\mathbf{s}_j \rangle = \begin{cases} \sqrt{D_x}, & \text{if } |i - j| = 1 \text{ and } (i = 0 \text{ or } j = 0), \\ \sqrt{d}, & \text{if } |i - j| = 1 \text{ and } (i > 0 \text{ and } j > 0), \\ 0, & \text{if } |i - j| \neq 1. \end{cases}$$

This yields (ii) and thus completes the proof of Lemma B.1. \square

APPENDIX C: SPECTRAL PROPERTIES OF TRIDIAGONAL MATRICES

In this section we analyse the spectral properties of the tridiagonal matrices $Z(\alpha)$. These $(r + 1) \times (r + 1)$ -matrices were defined in (8.21) for $\alpha > 0$ and $r \in \mathbb{N}$.

In Lemma C.1 below we collect and prove a few spectral properties of $Z(\alpha)$ for large r , in particular, about its extreme eigenvalues and corresponding approximate eigenvectors. Although we shall not need Lemma C.1, it serves as motivation for the approximate eigenvectors introduced in Section 5 for large eigenvalues of the Erdős–Rényi graph. Moreover, the key concepts behind the proof of Lemma C.1 will be needed for the proof of Proposition C.2, and they are most transparent in the simple setting of Lemma C.1.

LEMMA C.1 (Eigenvalues and approximate eigenvectors of $Z(\alpha)$). *If $\alpha > 2$, then the following hold:*

- (i) (Extreme eigenvalues) *The largest and smallest eigenvalues of $Z(\alpha)$, $\lambda_1(Z(\alpha))$ and $\lambda_{r+1}(Z(\alpha))$, converge to $\Lambda(\alpha)$ and $-\Lambda(\alpha)$, respectively, as $r \rightarrow \infty$.*
- (ii) (Bulk eigenvalues) *The eigenvalues $\lambda_2(Z(\alpha)), \dots, \lambda_r(Z(\alpha))$ lie in $[-2, 2]$.*
- (iii) (Approximate eigenvectors) *Let $\mathbf{u} = (u_i)_{i=0}^r$ and $\mathbf{u}_- = ((-1)^i u_i)_{i=0}^r$ have components*

$$(C.1) \quad u_0 \in \mathbb{R} \setminus \{0\}, \quad u_1 := \left(\frac{\alpha}{\alpha - 1}\right)^{1/2} u_0, \quad u_i := \left(\frac{1}{\alpha - 1}\right)^{(i-1)/2} u_1$$

for $i = 2, 3, \dots, r$. Then, \mathbf{u} and \mathbf{u}_- are approximate eigenvectors, as $r \rightarrow \infty$, of $Z(\alpha)$ corresponding to its largest and smallest eigenvalue, respectively.

The eigenvectors of $Z(\alpha)$ can be analysed by a transfer matrix approach. Let η be an eigenvalue of $Z(\alpha)$ and $\mathbf{u} = (u_i)_{i=0}^r$ a corresponding eigenvector. The components of the eigenvalue-eigenvector relation $Z(\alpha)\mathbf{u} = \eta\mathbf{u}$ read

$$(C.2) \quad \sqrt{\alpha}u_1 = \eta u_0, \quad \sqrt{\alpha}u_0 + u_2 = \eta u_1, \quad u_{i-1} + u_{i+1} = \eta u_i, \quad u_{r-1} = \eta u_r$$

for $i = 2, \dots, r - 1$. Hence, for $i = 1, \dots, r - 1$, these relations are equivalent to

$$(C.3) \quad \begin{pmatrix} u_{i+1} \\ u_i \end{pmatrix} = T(\eta)^{i-1} \begin{pmatrix} u_2 \\ u_1 \end{pmatrix},$$

where we introduced the 2×2 transfer matrix $T(\eta)$, defined through

$$(C.4) \quad T(\eta) := \begin{pmatrix} \eta & -1 \\ 1 & 0 \end{pmatrix}.$$

From now on, we suppose that $|\eta| > 2$. In this case we compute the spectrum and the eigenspaces of $T(\eta)$. The eigenvalues of $T(\eta)$ are $\gamma(\eta)$ and $\gamma(\eta)^{-1}$, where we defined

$$(C.5) \quad \gamma(\eta) := \frac{1}{2}(\eta - \text{sign}(\eta)\sqrt{\eta^2 - 4}).$$

Note that $|\gamma(\eta)| < 1$. Moreover, the eigenspaces of $T(\eta)$ associated to $\gamma(\eta)$ and $\gamma(\eta)^{-1}$ are given by

$$(C.6) \quad \begin{aligned} \ker(T(\eta) - \gamma(\eta)I_2) &= \text{span} \left\{ \begin{pmatrix} \gamma(\eta) \\ 1 \end{pmatrix} \right\}, \\ \ker(T(\eta) - \gamma(\eta)^{-1}I_2) &= \text{span} \left\{ \begin{pmatrix} 1 \\ \gamma(\eta) \end{pmatrix} \right\}. \end{aligned}$$

In the following we denote the standard basis vectors of \mathbb{R}^{r+1} by $\mathbf{e}_0, \dots, \mathbf{e}_r$.

PROOF OF LEMMA C.1. We first prove (ii). To that end, we consider $Z(\alpha)$ as a rank-two perturbation of $Z(1)$. It is well known that

$$\text{spec}(Z(1)) = \left\{ 2 \cos\left(\frac{\pi k}{r+2}\right) : k = 1, \dots, r+1 \right\} \subset [-2, 2].$$

This implies (ii) by Weyl’s interlacing inequalities, since the matrix $Z(\alpha) - Z(1)$ has rank two with one positive eigenvalue and one negative eigenvalue.

We now show (i) and (iii) simultaneously. Let \mathbf{u} and \mathbf{u}_- be defined as in (iii). We only focus on the largest eigenvalue of $Z(\alpha)$ and \mathbf{u} . The same arguments work for the smallest eigenvalue and \mathbf{u}_- .

We set $\eta = \Lambda(\alpha)$ and obtain $\gamma(\eta) = (\alpha - 1)^{-1/2}$. Thus, $(u_{i+1}, u_i)^* \in \ker(T(\eta) - \gamma(\eta)I_2)$ for all $i = 1, \dots, r - 1$. Hence, the equivalence between (C.2) for $i = 2, \dots, r - 1$ and (C.3) implies that

$$(Z(\alpha) - \eta I_{r+1})\mathbf{u} = (u_{r-1} - \eta u_r)\mathbf{e}_r.$$

Here, we also used $\eta = \Lambda(\alpha) = \alpha/\sqrt{\alpha - 1}$ and the relation between u_1 and u_0 . Therefore, since $\alpha > 2$, we find $\|(Z(\alpha) - \Lambda(\alpha)I_{r+1})\mathbf{u}\| \rightarrow 0$ as $r \rightarrow \infty$. This completes the proof of Lemma C.1. \square

The following proposition provides an eigenvector delocalization bound for tridiagonal matrices whose structure is similar to the one of $Z(\alpha)$ in the sense that, starting from the second row and column, the diagonal entries are small while the offdiagonal entries are close to one. For its formulation we need some notation, which we define now. For $\eta > 2$, we recall the definition of $\gamma(\eta)$ from (C.5) and introduce

$$(C.7) \quad \delta(z_{00}, z_{01}, z_{11}, z_{12}, \eta) := \frac{\gamma(\eta)(\eta - z_{00})(\eta - z_{11}) - \gamma(\eta)z_{01}^2 - z_{12}(\eta - z_{00})}{z_{01}^2 + \gamma(\eta)(\eta - z_{00})z_{12} - (\eta - z_{00})(\eta - z_{11})},$$

whenever the denominator on the right-hand side is different from zero. For $\eta > 2$ and $\varepsilon > 0$, we also define

$$(C.8) \quad \begin{aligned} \gamma_{\geq}(z_{00}, z_{01}, z_{11}, z_{12}, \eta, \varepsilon) \\ := \gamma(\eta)^{-1} - \frac{8(3 + \eta)\varepsilon}{1 - \gamma(\eta)^2} (1 + 1 \vee (\delta(z_{00}, z_{01}, z_{11}, z_{12}, \eta))^2)^{1/2}. \end{aligned}$$

PROPOSITION C.2 (Delocalization bound for tridiagonal matrices). *Let \tilde{Z} be a symmetric tridiagonal $(r + 1) \times (r + 1)$ matrix and $\mathbf{b} = (b_i)_{i \in [r]} \in \mathbb{R}^{r+1}$. Let $\eta > 2$. We set $\varepsilon := \max_{i \in [r-1]} (|\tilde{Z}_{ii}| \vee |\tilde{Z}_{ii+1} - 1|)$, $\delta := \delta(\tilde{Z}_{00}, \tilde{Z}_{01}, \tilde{Z}_{11}, \tilde{Z}_{12}, \eta)$ and $\gamma_{\geq} = \gamma_{\geq}(\tilde{Z}_{00}, \tilde{Z}_{01}, \tilde{Z}_{11}, \tilde{Z}_{12}, \eta, \varepsilon)$. If $\varepsilon \leq 1/2$ and the condition*

$$(C.9) \quad \eta^2 \geq 4 + \frac{4^5(3 + \eta)^2 \varepsilon^2}{(1 - \gamma(\eta)^2)^2} (1 + 1 \vee \delta^2)$$

is satisfied and $(\tilde{Z} - \eta I_{r+1})\mathbf{b} \in \text{Span}\{\mathbf{e}_r\}$, then

$$\frac{(b_0)^2}{\|\mathbf{b}\|^2} \leq \frac{8(\tilde{Z}_{01}\tilde{Z}_{12})^2[(\gamma_{\geq})^{-2r} \wedge (r - 1)^{-1}]}{((\tilde{Z}_{01})^2 - (\eta - \tilde{Z}_{11})(\eta - \tilde{Z}_{00}) + \gamma(\eta)\tilde{Z}_{12}(\eta - \tilde{Z}_{00}))^2}$$

and

$$(C.10) \quad \gamma_{\geq} \geq 1 + \frac{1}{2}(\gamma(\eta)^{-1} - 1) \geq 1.$$

PROOF. As $(\tilde{Z} - \eta I_{r+1})\mathbf{b} \in \text{Span}\{\mathbf{e}_r\}$, we have

$$(C.11) \quad \tilde{Z}_{00}b_0 + \tilde{Z}_{01}b_1 = \eta b_0, \quad \tilde{Z}_{ii-1}b_{i-1} + \tilde{Z}_{ii}b_i + \tilde{Z}_{ii+1}b_{i+1} = \eta b_i$$

for any $i = 1, \dots, r - 1$. Hence,

$$(C.12) \quad \begin{pmatrix} b_{i+1} \\ b_i \end{pmatrix} = T_i \begin{pmatrix} b_i \\ b_{i-1} \end{pmatrix}, \quad T_i := \frac{1}{\tilde{Z}_{ii+1}} \begin{pmatrix} \eta - \tilde{Z}_{ii} & -\tilde{Z}_{ii-1} \\ \tilde{Z}_{ii+1} & 0 \end{pmatrix}$$

for $i = 1, \dots, r - 1$. For $i \geq 2$, we define $R_i := T_i - T$, where $T = T(\eta)$ is defined as in (C.4). As $\varepsilon = \|\tilde{Z} - Z(\alpha)\| \leq 1/2$, we have $\|R_i\| \leq 2(3 + \eta)\varepsilon$ uniformly for $i \geq 2$.

In the rest of the proof, we write $\gamma \equiv \gamma(\eta)$ for $\gamma(\eta)$ defined in (C.5). For each $i \geq 1$, we denote by p_i and q_i the first and second component of $(b_{i+1}, b_i)^*$ in the eigenbasis of T , respectively. That is,

$$(C.13) \quad \begin{pmatrix} p_i \\ q_i \end{pmatrix} := V^{-1} \begin{pmatrix} b_{i+1} \\ b_i \end{pmatrix}, \quad V = \begin{pmatrix} \gamma & 1 \\ 1 & \gamma \end{pmatrix}.$$

The fact that $V^{-1}TV$ is diagonal can be easily read off from (C.6).

We shall now show that

$$(C.14) \quad \frac{|p_i|}{|q_i|} \leq 1 \vee \left(\frac{|p_1|}{|q_1|} \right), \quad |q_i| \geq (\gamma_{\geq})^{i-1} |q_1|$$

for all $i \geq 1$ by induction on i . The assertion is trivial for $i = 1$. From (C.13) and (C.12), we conclude

$$\begin{aligned} \begin{pmatrix} p_{i+1} \\ q_{i+1} \end{pmatrix} &= V^{-1} \left(V \begin{pmatrix} \gamma & 0 \\ 0 & \gamma^{-1} \end{pmatrix} V^{-1} + R_{i+1} V V^{-1} \right) \begin{pmatrix} b_{i+1} \\ b_i \end{pmatrix} \\ &= \left(\begin{pmatrix} \gamma & 0 \\ 0 & \gamma^{-1} \end{pmatrix} + V^{-1} R_{i+1} V \right) \begin{pmatrix} p_i \\ q_i \end{pmatrix}. \end{aligned}$$

Estimating the first component of this relation implies

$$\begin{aligned} |p_{i+1}| &\leq \gamma |p_i| + \left\| V^{-1} R_{i+1} V \begin{pmatrix} p_i \\ q_i \end{pmatrix} \right\| \\ &\leq \left(\gamma \frac{|p_i|}{|q_i|} + \|V^{-1} R_{i+1} V\| \left(1 + \left(\frac{p_i}{q_i} \right)^2 \right)^{1/2} \right) |q_i| \\ &\leq \left(1 \vee \frac{|p_1|}{|q_1|} \right) (\gamma + \sqrt{2} \|V^{-1} R_{i+1} V\|) |q_i|, \end{aligned}$$

where we used that $1 + p_i^2/q_i^2 \leq 2(1 \vee p_1^2/q_1^2)$ by the induction hypothesis in the last step. Similarly, we bound the second component from below and obtain

$$\begin{aligned}
 |q_{i+1}| &\geq \left(\gamma^{-1} - \|V^{-1}R_{i+1}V\| \left(\left(\frac{p_i}{q_i} \right)^2 + 1 \right)^{1/2} \right) |q_i| \\
 &\geq \left(\gamma^{-1} - \|V^{-1}R_{i+1}V\| \left(1 + 1 \vee \left(\frac{p_1}{q_1} \right)^2 \right)^{1/2} \right) |q_i|,
 \end{aligned}
 \tag{C.15}$$

due to the induction hypothesis.

By dividing the upper bound on $|p_{i+1}|$ by the lower bound on $|q_{i+1}|$, we see that the induction step for the first bound in (C.14) is shown if

$$\gamma + \sqrt{2} \|V^{-1}R_{i+1}V\| \leq \gamma^{-1} - \|V^{-1}R_{i+1}V\| \left(1 + 1 \vee \left(\frac{p_1}{q_1} \right)^2 \right)^{1/2}.
 \tag{C.16}$$

We now deduce this bound from (C.9). To that end, we first compute p_1/q_1 . The definition of p_1 and q_1 in (C.13) yields $p_1 = (-b_1 + \gamma b_2)/(\gamma^2 - 1)$ and $q_1 = (-b_2 + \gamma b_1)/(\gamma^2 - 1)$. We use the first relation in (C.11) and the second relation in (C.11) with $i = 1$ to express b_1 and b_2 in terms of b_0 . Then, an easy computation shows that

$$p_1 = \frac{\gamma b_2 - b_1}{\gamma^2 - 1} = \frac{b_0(\gamma(\eta - \tilde{Z}_{11})(\eta - \tilde{Z}_{00}) - \gamma(\tilde{Z}_{01})^2 - \tilde{Z}_{12}(\eta - \tilde{Z}_{00}))}{(\gamma^2 - 1)\tilde{Z}_{01}\tilde{Z}_{12}},
 \tag{C.17a}$$

$$q_1 = \frac{-b_2 + \gamma b_1}{\gamma^2 - 1} = \frac{b_0((\tilde{Z}_{01})^2 - (\eta - \tilde{Z}_{11})(\eta - \tilde{Z}_{00}) + \gamma\tilde{Z}_{12}(\eta - \tilde{Z}_{00}))}{(\gamma^2 - 1)\tilde{Z}_{01}\tilde{Z}_{12}}.
 \tag{C.17b}$$

Therefore, we obtain

$$\frac{p_1}{q_1} = \frac{-b_1 + \gamma b_2}{-b_2 + \gamma b_1} = \delta := \delta(\tilde{Z}_{00}, \tilde{Z}_{01}, \tilde{Z}_{11}, \tilde{Z}_{12}, \eta),$$

where we used the function δ , defined in (C.7). Thus, as $\|V^{-1}R_{i+1}V\| \leq 4\|R_{i+1}\|/(1 - \gamma^2) \leq 8(3 + \eta)\varepsilon/(1 - \gamma^2)$ and $\gamma^{-1} - \gamma = \sqrt{\eta^2 - 4}$, the definition of $\gamma(\eta)$ in (C.5) shows that (C.16) is a consequence of (C.9). This completes the induction step for the first estimate in (C.14).

From (C.15), $p_1/q_1 = \delta$ and $\|V^{-1}R_{i+1}V\| \leq 8(3 + \eta)\varepsilon/(1 - \gamma^2)$, we deduce $|q_{i+1}| \geq \gamma_{\geq} |q_i|$. Thus, we have completed the proof of (C.14).

We now prove that (C.9) also implies the lower bound on γ_{\geq} in (C.10). The definition of $\gamma(\eta)$ in (C.5) yields $2(\gamma^{-1} - 1) \geq (\gamma^{-1} - \gamma) = \sqrt{\eta^2 - 4}$. Thus, we obtain from (C.9) that

$$\begin{aligned}
 \gamma_{\geq} - 1 - \frac{1}{2}(\gamma^{-1} - 1) &= \frac{1}{2} \left[\gamma^{-1} - 1 - \frac{16(3 + \eta)\varepsilon}{1 - \gamma^2} \left(1 + 1 \vee \left(\frac{p_1}{q_1} \right)^2 \right)^{1/2} \right] \\
 &\geq \frac{1}{4} \left[\sqrt{\eta^2 - 4} - \frac{32(3 + \eta)\varepsilon}{1 - \gamma^2} \left(1 + 1 \vee \left(\frac{p_1}{q_1} \right)^2 \right)^{1/2} \right].
 \end{aligned}$$

Owing to (C.9), the right-hand side is positive. This immediately implies (C.10).

Owing to the second bound in (C.14), we have

$$\begin{aligned}
 2 \sum_{i=0}^r (b_i)^2 &\geq \sum_{i=0}^{r-1} \left\| \begin{pmatrix} b_{i+1} \\ b_i \end{pmatrix} \right\|^2 \\
 &= \sum_{i=0}^{r-1} \left\| V V^{-1} \begin{pmatrix} b_{i+1} \\ b_i \end{pmatrix} \right\|^2 \\
 (C.18) \qquad &\geq \frac{1}{\|V^{-1}\|^2} \sum_{i=1}^{r-1} \left\| \begin{pmatrix} p_i \\ q_i \end{pmatrix} \right\|^2 \\
 &\geq \frac{(\gamma^2 - 1)^2}{4} |q_1|^2 [(\gamma_{\geq})^{2r} \vee (r - 1)].
 \end{aligned}$$

Here, we pulled V out of the norm in the third step and used (C.13). The fourth step is a consequence of $\|V^{-1}\| \leq 2(1 - \gamma^2)^{-1}$, estimating the norm by its second component and using the second bound in (C.14) as well as $\gamma_{\geq} \geq 1$ due to (C.10).

Finally, applying (C.17b) to (C.18) completes the proof of Proposition C.2. \square

APPENDIX D: DEGREE DISTRIBUTION OF THE ERDŐS–RÉNYI GRAPH

The content of this section is standard, and we include it for completeness and the reader’s convenience. It is essentially contained in [5], Chapter 3. We do not aim for sharp estimates of the error probabilities; instead, our goal here is to collect basic qualitative facts about the behaviour of the largest degrees of an Erdős–Rényi graph which, using Theorem 2.1, can be used to understand the key properties of the extremal eigenvalues. We recall the normalized degree (2.1) and the random permutation (2.3).

To formulate qualitative statements conveniently, we use the symbol $o(1)$ to denote any function of N that converges to zero and say that an N -dependent event $\Xi \equiv \Xi_N$ holds with high probability if $\mathbb{P}(\Xi) = 1 - o(1)$.

The distribution of the largest degrees is best analysed using the function

$$(D.1) \qquad f_d(\alpha) := d(\alpha \log \alpha - \alpha + 1) + \frac{1}{2} \log(2\pi \alpha d)$$

for $\alpha \geq 1$. For its interpretation we note that if $Y \stackrel{d}{=} \text{Poisson}(d)$, then by Stirling’s formula we have, for any $k \in \mathbb{N}$,

$$\mathbb{P}(Y = k) = \exp\left(-f_d(k/d) + \mathcal{O}\left(\frac{1}{k}\right)\right).$$

It is easy to see that the function $f_d : [1, \infty) \rightarrow [\frac{1}{2} \log(2\pi d), \infty)$ is bijective and increasing. Therefore, there is a universal constant $C > 0$ such that, for $1 \leq l \leq \frac{N}{C\sqrt{d}}$, the equation

$$f_d(\beta) = \log(N/l)$$

has a unique solution $\beta \equiv \beta_l(d)$. The interpretation of β is the typical location of $\alpha_{\sigma(l)}$. By the implicit function theorem we find that β_l on the interval $(0, \frac{N^2}{C^2}]$ is a decreasing bijective function.

We are interested in normalized degrees greater than or equal to 2. This motivates the definition

$$\mathcal{L}_0(d) := \max\{l \geq 1 : \beta_l(d) \geq 2\},$$

whose interpretation is the typical number of normalized degrees greater than or equal to 2. By definition, $\beta_l(d) \geq 2$ for all $l \leq \mathcal{L}_0(d)$. Note that $\mathcal{L}_0(d)$ is nonzero if and only if $d \leq d_*$, where d_* is defined as the unique solution of $\beta_1(d_*) = 2$. More explicitly, d_* satisfies $f_{d_*}(2) = \log N$.

PROPOSITION D.1. *Let $\xi \equiv \xi_N$ be a positive sequence tending to ∞ . If $1 \leq d \leq d_*$ and $1 \leq l \leq \mathcal{L}_0(d)$, then with high probability we have*

$$(D.2) \quad |\alpha_{\sigma(l)} - \beta_l(d)| \leq \frac{1 \vee (\xi / \log \beta_l(d))}{d}.$$

If $d > d_*$, then with high probability we have

$$(D.3) \quad \alpha_{\sigma(1)} \leq 2 + \frac{\xi}{d}.$$

PROOF. Throughout the proof we suppose that $2 \leq \alpha \leq \frac{\sqrt{N}}{Cd}$ for some large enough universal constant C . From the definition of $\beta_l(d)$, it is easy to check that this condition is satisfied for $\alpha = \beta_l(d)$ for $1 \leq d \leq d_*$ and $1 \leq l \leq \mathcal{L}_0(d)$. The proof of (D.2) consists of an upper and a lower bound. The former is proved using a first moment method and the latter using a second moment method. We make use of the counting function $\mathcal{N}_t := \sum_{x \in [N]} \mathbb{1}_{D_x \geq t}$. Note that by Poisson approximation of the binomial random variable $D_x = d\alpha_x$ (see [3], Lemma 3.3), there is a universal constant C such that

$$(D.4) \quad C^{-1} N e^{-fd(\alpha)} \leq \mathbb{E} \mathcal{N}_{\alpha d} \leq C N e^{-fd(\alpha)}.$$

Let $1 \leq d \leq d_*$ and $1 \leq l \leq \mathcal{L}_0(d)$. We begin by proving an upper bound on $\alpha_{\sigma(1)} = D_{\sigma(1)}/d$. Using (D.4), we get

$$(D.5) \quad \mathbb{P}(\alpha_{\sigma(l)} \geq \alpha) = \mathbb{P}(\mathcal{N}_{\alpha d} \geq l) \leq \frac{\mathbb{E} \mathcal{N}_{\alpha d}}{l} \leq \frac{CN}{l} e^{-fd(\alpha)}.$$

Next, we prove a lower bound on $\alpha_{\sigma(l)}$. Suppose that $\mathbb{E} \mathcal{N}_{\alpha d} \geq 2l$. Then, using a second moment method, we find

$$\mathbb{P}(\alpha_{\sigma(l)} \geq \alpha) = \mathbb{P}(\mathcal{N}_{\alpha d} \geq l) \geq \mathbb{P}(|\mathcal{N}_{\alpha d} - \mathbb{E} \mathcal{N}_{\alpha d}| < \mathbb{E} \mathcal{N}_{\alpha d} / 2) \geq 1 - \frac{4 \text{Var}(\mathcal{N}_{\alpha d})}{(\mathbb{E} \mathcal{N}_{\alpha d})^2}.$$

By [5], Lemma 3.11, we have $\text{Var}(\mathcal{N}_{\alpha d}) \leq C \mathbb{E} \mathcal{N}_{\alpha d}$ for some universal constant C , which yields

$$(D.6) \quad \mathbb{P}(\alpha_{\sigma(l)} \geq \alpha) \geq 1 - \frac{C}{\mathbb{E} \mathcal{N}_{\alpha d}} \geq 1 - \frac{C}{N e^{-fd(\alpha)}},$$

where we used (D.4). From (D.5) we conclude that $\alpha_{\sigma(l)} \leq \alpha$ with high probability if

$$(D.7) \quad f_d(\alpha) - \log(N/l) \rightarrow \infty \quad \text{as } N \rightarrow \infty.$$

From (D.6) we conclude that $\alpha_{\sigma(l)} \geq \alpha$ with high probability if $\mathbb{E} \mathcal{N}_{\alpha d} \geq 2l$ and $f_d(\alpha) - \log N \rightarrow -\infty$ as $N \rightarrow \infty$. By (D.4), both of these conditions are satisfied if

$$(D.8) \quad f_d(\alpha) - \log(N/l) \rightarrow -\infty \quad \text{as } N \rightarrow \infty.$$

Now, (D.2) follows easily by choosing $\alpha = \beta_l(d) \pm \frac{1 \vee (\xi / \log \beta_l(d))}{d}$, using that $f'_d(\alpha) \geq d \log \alpha$.

The proof of (D.3) is analogous to the the proof of the upper bound in (D.2), and we omit the details. \square

Acknowledgments. We gratefully acknowledge funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 715539_RandMat) and from the Swiss National Science Foundation through the SwissMAP grant.

REFERENCES

- [1] ALON, N. (1998). Spectral techniques in graph algorithms (invited paper). In *LATIN’98: Theoretical Informatics (Campinas, 1998)*. *Lecture Notes in Computer Science* **1380** 206–215. Springer, Berlin. MR1635529 <https://doi.org/10.1007/BFb0054322>
- [2] ALT, J., DUCATEZ, R. and KNOWLES, A. (2020). Delocalization transition for critical Erdős–Rényi graphs. Preprint. Available at [arXiv:2005.14180](https://arxiv.org/abs/2005.14180).
- [3] BENAYCH-GEORGES, F., BORDENAVE, C. and KNOWLES, A. (2019). Largest eigenvalues of sparse inhomogeneous Erdős–Rényi graphs. *Ann. Probab.* **47** 1653–1676. MR3945756 <https://doi.org/10.1214/18-AOP1293>
- [4] BENAYCH-GEORGES, F., BORDENAVE, C. and KNOWLES, A. (2020). Spectral radii of sparse random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **56** 2141–2161. MR4116720 <https://doi.org/10.1214/19-AIHP1033>
- [5] BOLLOBÁS, B. (2001). *Random Graphs*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **73**. Cambridge Univ. Press, Cambridge. MR1864966 <https://doi.org/10.1017/CBO9780511814068>
- [6] CHUNG, F. R. K. (1997). *Spectral Graph Theory*. *CBMS Regional Conference Series in Mathematics* **92**. Amer. Math. Soc., Providence, RI. MR1421568
- [7] ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2012). Spectral statistics of Erdős–Rényi Graphs II: Eigenvalue spacing and the extreme eigenvalues. *Comm. Math. Phys.* **314** 587–640. MR2964770 <https://doi.org/10.1007/s00220-012-1527-7>
- [8] ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2013). Spectral statistics of Erdős–Rényi graphs I: Local semicircle law. *Ann. Probab.* **41** 2279–2375. MR3098073 <https://doi.org/10.1214/11-AOP734>
- [9] FEIGE, U. and OFEK, E. (2005). Spectral techniques applied to sparse random graphs. *Random Structures Algorithms* **27** 251–275. MR2155709 <https://doi.org/10.1002/rsa.20089>
- [10] FÜREDI, Z. and KOMLÓS, J. (1981). The eigenvalues of random symmetric matrices. *Combinatorica* **1** 233–241. MR0637828 <https://doi.org/10.1007/BF02579329>
- [11] HOORY, S., LINIAL, N. and WIGDERSON, A. (2006). Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)* **43** 439–561. MR2247919 <https://doi.org/10.1090/S0273-0979-06-01126-8>
- [12] HUANG, J., LANDON, B. and YAU, H.-T. (2020). Transition from Tracy–Widom to Gaussian fluctuations of extremal eigenvalues of sparse Erdős–Rényi graphs. *Ann. Probab.* **48** 916–962. MR4089498 <https://doi.org/10.1214/19-AOP1378>
- [13] KRIVELICH, M. and SUDAKOV, B. (2003). The largest eigenvalue of sparse random graphs. *Combin. Probab. Comput.* **12** 61–72. MR1967486 <https://doi.org/10.1017/S0963548302005424>
- [14] LATAŁA, R., VAN HANDEL, R. and YOUSSEF, P. (2018). The dimension-free structure of nonhomogeneous random matrices. *Invent. Math.* **214** 1031–1080. MR3878726 <https://doi.org/10.1007/s00222-018-0817-x>
- [15] LEE, J. O. and SCHNELLI, K. (2018). Local law and Tracy–Widom limit for sparse random matrices. *Probab. Theory Related Fields* **171** 543–616. MR3800840 <https://doi.org/10.1007/s00440-017-0787-8>
- [16] LOVÁSZ, L. (1993). *Combinatorial Problems and Exercises*, 2nd ed. North-Holland, Amsterdam. MR1265492
- [17] TIKHOMIROV, K. and YOUSSEF, P. (2019). Outliers in spectrum of sparse Wigner matrices. Preprint. Available at [arXiv:1904.07985](https://arxiv.org/abs/1904.07985).
- [18] TRAN, L. V., VU, V. H. and WANG, K. (2013). Sparse random graphs: Eigenvalues and eigenvectors. *Random Structures Algorithms* **42** 110–134. MR2999215 <https://doi.org/10.1002/rsa.20406>
- [19] TROTTER, H. F. (1984). Eigenvalue distributions of large Hermitian matrices; Wigner’s semicircle law and a theorem of Kac, Murdock, and Szegő. *Adv. Math.* **54** 67–82. MR0761763 [https://doi.org/10.1016/0001-8708\(84\)90037-9](https://doi.org/10.1016/0001-8708(84)90037-9)
- [20] VU, V. H. (2007). Spectral norm of random matrices. *Combinatorica* **27** 721–736. MR2384414 <https://doi.org/10.1007/s00493-007-2190-z>
- [21] WIGNER, E. P. (1958). On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)* **67** 325–327. MR0095527 <https://doi.org/10.2307/1970008>