Commun. Math. Phys. 388, 507–579 (2021)
Digital Object Identifier (DOI) https://doi.org/10.1007/s00220-021-04167-y

Communications in
Mathematical Physics

Delocalization Transition for Critical Erdős–Rényi Graphs

Johannes Alt, Raphael Ducatez, Antti Knowles

Section of Mathematics, University of Geneva, Rue du Conseil-Général 7-9, 1205 Geneva, Switzerland.
E-mail: antti.knowles@unige.ch; johannes.alt@unige.ch; raphael.ducatez@unige.ch

Received: 26 April 2021 / Accepted: 3 July 2021
Published online: 14 August 2021 – © The Author(s) 2021

Abstract: We analyse the eigenvectors of the adjacency matrix of a critical Erdős–Rényi graph $\mathbb{G}(N, d/N)$, where d is of order $\log N$. We show that its spectrum splits into two phases: a delocalized phase in the middle of the spectrum, where the eigenvectors are completely delocalized, and a semilocalized phase near the edges of the spectrum, where the eigenvectors are essentially localized on a small number of vertices. In the semilocalized phase the mass of an eigenvector is concentrated in a small number of disjoint balls centred around resonant vertices, in each of which it is a radial exponentially decaying function. The transition between the phases is sharp and is manifested in a discontinuity in the localization exponent $\gamma(\mathbf{w})$ of an eigenvector $\mathbf{w}$, defined through $\|\mathbf{w}\|_\infty/\|\mathbf{w}\|_2 = N^{-\gamma(\mathbf{w})}$. Our results remain valid throughout the optimal regime $\sqrt{\log N} \ll d \leqslant O(\log N)$.

1. Introduction

1.1. Overview. Let A be the adjacency matrix of a graph with vertex set $[N] = \{1, \dots, N\}$. We are interested in the geometric structure of the eigenvectors of A, in particular their *spatial localization*. An ℓ^2-normalized eigenvector $\mathbf{w} = (w_x)_{x\in[N]}$ gives rise to a probability measure $\sum_{x\in[N]} w_x^2 \delta_x$ on the set of vertices. Informally, $\mathbf{w}$ is *delocalized* if its mass is approximately uniformly distributed throughout $[N]$, and *localized* if its mass is essentially concentrated in a small number of vertices.

There are several ways of quantifying spatial localization. One is the notion of concentration of mass, sometimes referred to as scarring [49], stating that there is some set $\mathcal{B} \subset [N]$ of small cardinality and a small $\varepsilon > 0$ such that $\sum_{x\in\mathcal{B}} w_x^2 = 1 - \varepsilon$. In this case, it is also of interest to characterize the geometric structure of the vertex set $\mathcal{B}$ and of the eigenvector $\mathbf{w}$ restricted to $\mathcal{B}$. Another convenient quantifier of spatial localization is the *ℓ^p-norm* $\|\mathbf{w}\|_p$ for $2 \leqslant p \leqslant \infty$. It has the following interpretation: if the mass of $\mathbf{w}$ is uniformly distributed over some set $\mathcal{B} \subset [N]$ then $\|\mathbf{w}\|_p^2 = |\mathcal{B}|^{-1+2/p}$. Focusing on

the ℓ^∞-norm for definiteness, we define the *localization exponent* $\gamma(\mathbf{w})$ through

$$\|\mathbf{w}\|_\infty^2 =: N^{-\gamma(\mathbf{w})}. \tag{1.1}$$

Thus, $0 \leqslant \gamma(\mathbf{w}) \leqslant 1$, and $\gamma(\mathbf{w}) = 0$ corresponds to localization at a single vertex while $\gamma(\mathbf{w}) = 1$ to complete delocalization.

In this paper we address the question of spatial localization for the random Erdős–Rényi graph $\mathbb{G}(N, d/N)$. We consider the limit $N \to \infty$ with $d \equiv d_N$. It is well known that $\mathbb{G}(N, d/N)$ undergoes a dramatic change in behaviour at the *critical scale* $d \asymp \log N$, which is the scale at and below which the vertex degrees do not concentrate. Thus, for $d \gg \log N$, with high probability all degrees are approximately equal and the graph is *homogeneous*. On the other hand, for $d \lesssim \log N$, the degrees do not concentrate and the graph becomes highly *inhomogeneous*: it contains for instance hubs of exceptionally large degree, leaves, and isolated vertices. As long as $d > 1$, the graph has with high probability a unique giant component, and we shall always restrict our attention to it.

Here we propose the Erdős–Rényi graph at criticality as a simple and natural model on which to address the question of spatial localization of eigenvectors. It has the following attributes.

(i) Its graph structure provides an intrinsic and nontrivial notion of distance.
(ii) Its spectrum splits into a *delocalized phase* and a *semilocalized phase*. The transition between the phases is sharp, in the sense of a discontinuity in the localization exponent.
(iii) Both phases are amenable to rigorous analysis.

Our results are summarized in the phase diagram of Fig. 1, which is expressed in terms of the parameter b parametrizing $d = b \log N$ on the critical scale and the eigenvalue λ of $A/\sqrt{d}$ associated with the eigenvector $\mathbf{w}$. To the best of our knowledge, the phase coexistence for the critical Erdős–Rényi graph established in this paper had previously not been analysed even in the physics literature.

Throughout the following, we always exclude the largest eigenvalue of A, its Perron–Frobenius eigenvalue, which is an outlier separated from the rest of the spectrum. The *delocalized phase* is characterized by a localization exponent asymptotically equal to 1. It exists for all fixed $b > 0$ and consists asymptotically of energies in $(-2, 0) \cup (0, 2)$. The *semilocalized phase* is characterized by a localization exponent asymptotically less than 1. It exists only when $b < b_*$, where

$$b_* := \frac{1}{2\log 2 - 1} \approx 2.59. \tag{1.2}$$

It consists asymptotically of energies in $(-\lambda_{\max}(b), -2) \cup (2, \lambda_{\max}(b))$, where $\lambda_{\max}(b) > 2$ is an explicit function of b (see (1.14) below). The density of states at energy $\lambda \in \mathbb{R}$ is equal to $N^{\rho_b(\lambda)+o(1)}$, where ρ_b is an explicit exponent defined in (1.14) below and illustrated in Fig. 2. It has a discontinuity at 2 (and similarly at -2), jumping from $\rho_b(2^-) = 1$ to $\rho_b(2^+) = 1 - b/b^*$. The localization exponent $\gamma(\mathbf{w})$ from (1.1) of an eigenvector $\mathbf{w}$ with associated eigenvalue λ satisfies with high probability

$$\gamma(\mathbf{w}) = 1 + o(1) \ \text{ if } \ |\lambda| < 2, \qquad \gamma(\mathbf{w}) \leqslant \rho_b(\lambda) + o(1) \ \text{ if } \ |\lambda| > 2.$$

This establishes a discontinuity, in the limit $N \to \infty$, in the localization exponent $\gamma(\mathbf{w})$ as a function of λ at the energies ± 2. See Fig. 2 for an illustration; we also refer to Appendix A.1 for a simulation depicting the behaviour of $\|\mathbf{w}\|_\infty$ throughout

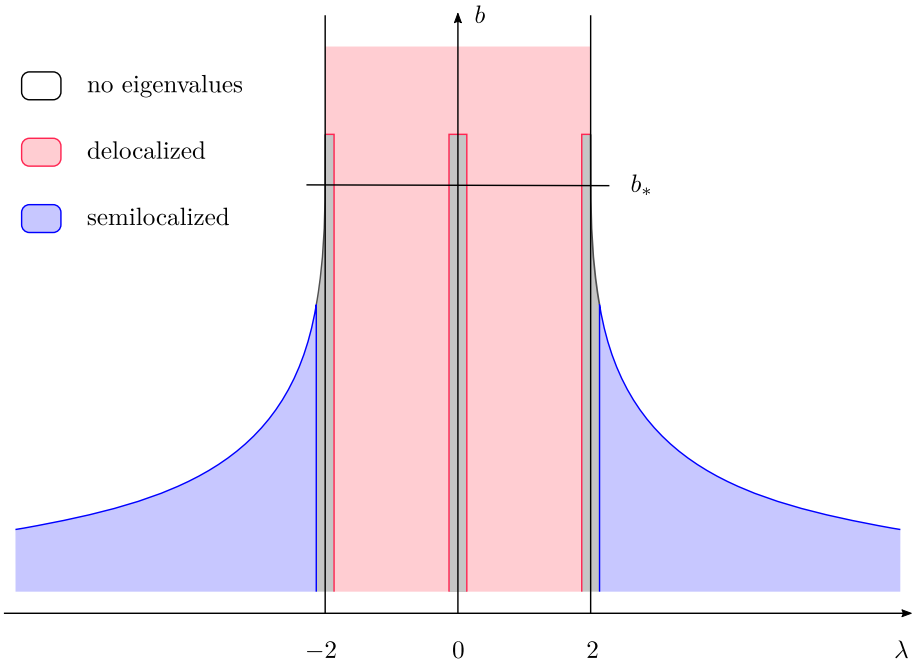

Fig. 1. The phase diagram of the adjacency matrix $A/\sqrt{d}$ of the Erdős–Rényi graph $\mathbb{G}(N, d/N)$ at criticality, where $d = b \log N$ with b fixed. The horizontal axis records the location in the spectrum and the vertical axis the sparseness parameter b. The spectrum is confined to the coloured region. In the red region the eigenvectors are delocalized while in the blue region they are semilocalized. The grey regions have width $o(1)$ and are not analysed in this paper. For $b > b_*$ the spectrum is asymptotically contained in $[-2, 2]$ and the semilocalized phase does not exist. For $b < b_*$ a semilocalized phase emerges in the region $(-\lambda_{\max}(b), -2) \cup (2, \lambda_{\max}(b))$ for some explicit $\lambda_{\max}(b) > 2$

the spectrum. Moreover, in the semilocalized phase scarring occurs in the sense that a fraction $1 - o(1)$ of the mass of the eigenvectors is supported in a set of at most $N^{\rho_b(\lambda)+o(1)}$ vertices.

The eigenvalues in the semilocalized phase were analysed in [10], where it was proved that they arise precisely from vertices x of abnormally large degree, $D_x \geqslant 2d$. More precisely, it was proved in [10] that each vertex x with $D_x \geqslant 2d$ gives rise to two eigenvalues of $A/\sqrt{d}$ near $\pm\Lambda(D_x/d)$, where $\Lambda(\alpha) := \frac{\alpha}{\sqrt{\alpha-1}}$. The same result for the $O(1)$ largest degree vertices was independently proved in [54] by a different method. We refer also to [14,15] for an analysis in the supercritical and subcritical phases.

In the current paper, we prove that the eigenvector $\mathbf{w}$ associated with an eigenvalue λ in the semilocalized phase is highly concentrated around *resonant vertices* at energy λ, which are defined as the vertices x such that $\Lambda(D_x/d)$ is close to λ. For this reason, we also call the resonant vertices *localization centres*. With high probability, and after a small pruning of the graph, all balls $B_r(x)$ of a certain radius $r \gg 1$ around the resonant vertices are disjoint, and within any such ball $B_r(x)$ the eigenvector $\mathbf{w}$ is an approximately radial exponentially decaying function. The number of resonant vertices at energy λ is comparable to the density of states, $N^{\rho_b(\lambda)+o(1)}$, which is much less than N. See Fig. 3 for a schematic illustration of the mass distribution of $\mathbf{w}$.

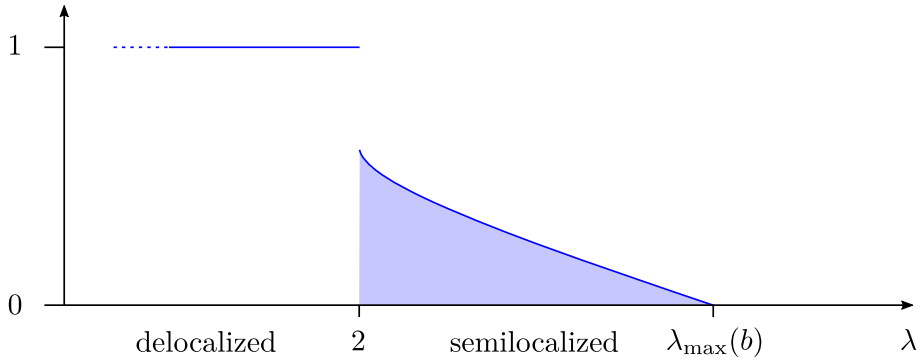

Fig. 2. The behaviour of the exponents ρ_b and γ as a function of the energy λ. The dark blue curve is the exponent $\rho_b(\lambda)$ characterizing the density of states $N^{\rho_b(\lambda)+o(1)}$ of the matrix $A/\sqrt{d}$ at energy λ. The entire blue region (light and dark blue) is the asymptotically allowed region of the localization exponent $\gamma(\mathbf{w})$ of an eigenvector of $A/\sqrt{d}$ as a function of the associated eigenvalue λ. Here $d = b \log N$ with $b = 1$ and $\lambda_{\max}(b) \approx 2.0737$. We only plot a neighbourhood of the threshold energy 2. The discontinuity at 2 of ρ_b is from $\rho_b(2^-) = 1$ to $\rho_b(2^+) = 1 - b/b^* = 2 - 2\log 2$

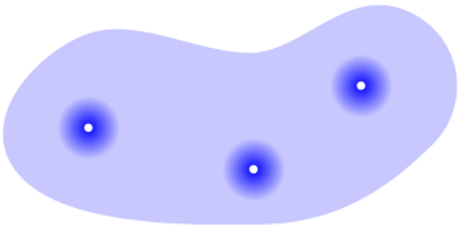

Fig. 3. A schematic representation of the geometric structure of a typical eigenvector in the semilocalized phase. The giant component of the graph is depicted in pale blue. The eigenvector's mass (depicted in dark blue) is concentrated in a small number of disjoint balls centred around resonant vertices (drawn in white), and within each ball the mass decays exponentially in the radius. The mass outside the balls is an asymptotically vanishing proportion of the total mass

The behaviour of the critical Erdős–Rényi graph described above has some similarities but also differences to that of the Anderson model [11]. The Anderson model on $\mathbb{Z}^n$ with $n \geqslant 3$ is conjectured to exhibit a metal-insulator, or delocalization-localization, transition: for weak enough disorder, the spectrum splits into a delocalized phase in the middle of the spectrum and a localized phase near the spectral edges. See e.g. [8, Figure 1.2] for a phase diagram of its conjectured behaviour. So far, only the localized phase of the Anderson model has been understood rigorously, in the landmark works [4,39], as well as contributions of many subsequent developments. The phase diagram for the Anderson model bears some similarity to that of Fig. 1, in which one can interpret $1/b$ as the disorder strength, since smaller values of b lead to stronger inhomogeneities in the graph.

As is apparent from the proofs in [4,39], in the localized phase the local structure of an eigenvector of the Anderson model is similar to that of the critical Erdős–Rényi graph described above: exponentially decaying around well-separated localization centres associated with resonances near the energy λ of the eigenvector. The localization centres arise from exceptionally large local averages of the potential. The phenomenon of localization can be heuristically understood using the following well-known *rule of*

thumb: one expects localization around a single localization centre if *the level spacing is much larger than the tunnelling amplitude between localization centres*. It arises from perturbation theory around the block diagonal model where the complement of balls $B_r(x)$ around localization centres is set to zero. On a very elementary level, this rule is illustrated by the matrix $H(t) = \left(\begin{smallmatrix} 0 & t \\ t & 1 \end{smallmatrix}\right)$, whose eigenvectors are localized for $t = 0$, remain essentially localized for $t \ll 1$, where perturbation theory around $H(0)$ is valid, and become delocalized for $t \gtrsim 1$, where perturbation theory around $H(0)$ fails.

More precisely, it is a general heuristic that the tunnelling amplitude decays exponentially in the distance between the localization centres [25]. Denoting by $\beta(\lambda) > 1$ the rate of exponential decay at energy λ, the rule of thumb hence reads

$$\beta(\lambda)^{-L} \ll \varepsilon(\lambda)\,, \tag{1.3}$$

where L is the distance between the localization centres and $\varepsilon(\lambda)$ the level spacing at energy λ. For the Anderson model restricted to a finite cube of $\mathbb{Z}^n$ with side length $N^{1/n}$, the level spacing $\varepsilon(\lambda)$ is of order N^{-1} (see [57] and [8, Chapter 4]) whereas the diameter of the graph is of order $N^{1/n}$. Hence, the rule of thumb (1.3) becomes

$$\beta(\lambda)^{-N^{1/n}} \ll N^{-1}\,,$$

which is satisfied and one therefore expects localization. For the critical Erdős–Rényi graph, the level spacing $\varepsilon(\lambda)$ is $N^{-\rho(\lambda)+o(1)}$ but the diameter of the giant component is only $\frac{\log N}{\log d}$. Hence, the rule of thumb (1.3) becomes

$$N^{-\frac{\log \beta(\lambda)}{\log d}} \ll N^{-\rho(\lambda)+o(1)}\,,$$

which is never satisfied because $\frac{\log \beta(\lambda)}{\log d} \to 0$ as $N \to \infty$. Thus, the rule of thumb (1.3) is satisfied in the localized phase of the Anderson model but not in the semilocalized phase of the critical Erdős–Rényi graph. The underlying reason behind this difference is that the diameter of the Anderson model is polynomial in N, while the diameter of the critical Erdős–Rényi graph is logarithmic in N. Thus, the critical Erdős–Rényi graph is far more connected than the Anderson model; this property tends to push it more towards the delocalized behaviour of mean-field systems. As noted above, another important difference between the localized phase of the Anderson model and the semilocalized phase of the critical Erdős–Rényi graph is that the density of states is of order N in the former and a fractional power of N in the latter.

Up to now we have focused on the Erdős–Rényi graph on the critical scale $d \asymp \log N$. It is natural to ask whether this assumption can be relaxed without changing its behaviour. The question of the upper bound on d is simple: as explained above, there is no semilocalized phase for $d > b_* \log N$, and the delocalized phase is completely understood up to $d \leqslant N/2$, thanks to Theorem 1.8 below and [35,42]. The lower bound is more subtle. In fact, it turns out that all of our results remain valid throughout the regime

$$\sqrt{\log N} \ll d \leqslant O(\log N)\,. \tag{1.4}$$

The lower bound $\sqrt{\log N}$ is optimal in the sense that below it both phases are disrupted and the phase diagram from Fig. 1 no longer holds. Indeed, for $d \lesssim \sqrt{\log N}$ a new family of localized states, associated with so-called *tuning forks* at the periphery of the graph, appear throughout the delocalized and semilocalized phases. We refer to Sect. 1.5 below for more details.

Previously, strong delocalization with localization exponent $\gamma(\mathbf{w}) = 1 + o(1)$ has been established for many mean-field models, such as Wigner matrices [1,34–37], supercritical Erdős–Rényi graphs [35,42], and random regular graphs [12,13]. All of these models are homogeneous and only have a delocalized phase.

Although a rigorous understanding of the metal-insulator transition for the Anderson model is still elusive, some progress has been made for random band matrices. Random band matrices [23,40,47,58] constitute an attractive model interpolating between the Anderson model and mean-field Wigner matrices. They retain the n-dimensional structure of the Anderson model but have proved somewhat more amenable to rigorous analysis. They are conjectured [40] to have a similar phase diagram as the Anderson model in dimensions $n \geqslant 3$. As for the Anderson model, dimensions $n > 1$ have so far seen little progress, but for $n = 1$ much has been understood both in the localized [48,50] and the delocalized [20–22,28–33,43,51,52,59] phases. A simplification of band matrices is the ultrametric ensemble [41], where the Euclidean metric of $\mathbb{Z}^n$ is replaced with an ultrametric arising from a tree structure. For this model, a phase transition was rigorously established in [56].

Another modification of the n-dimensional Anderson model is the Anderson model on the Bethe lattice, an infinite regular tree corresponding to the case $n = \infty$. For it, the existence of a delocalized phase was shown in [5,38,44]. In [6,7] it was shown that for unbounded random potentials the delocalized phase exists for arbitrarily weak disorder. It extends beyond the spectrum of the unperturbed adjacency matrix into the so-called Lifschitz tails, where the density of states is very small. The authors showed that, through the mechanism of resonant delocalization, the exponentially decaying tunnelling amplitudes between localization centres are counterbalanced by an exponentially large number of possible channels through which tunnelling can occur, so that the rule of thumb (1.3) for localization is violated. As a consequence, the eigenvectors are delocalized across many resonant localization centres. We remark that this analysis was made possible by the absence of cycles on the Bethe lattice. In contrast, the global geometry of the critical Erdős–Rényi graph is fundamentally different from that of the Bethe lattice (through the existence of a very large number of long cycles), which has a defining impact on the nature of the delocalization-semilocalization transition summarized in Fig. 1.

Transitions in the localization behaviour of eigenvectors have also been analysed in several mean-field type models. In [45,46] the authors considered the sum of a Wigner matrix and a diagonal matrix with independent random entries with a large enough variance. They showed that the eigenvectors in the bulk are delocalized while near the edge they are partially localized at a single site. Their partially localized phase can be understood heuristically as a rigorous (and highly nontrivial) verification of the rule of thumb for localization, where the perturbation takes place around the diagonal matrix. Heavy-tailed Wigner matrices, or Lévy matrices, whose entries have α-stable laws for $0 < \alpha < 2$, were proposed in [24] as a simple model that exhibits a transition in the localization of its eigenvectors; we refer to [3] for a summary of the predictions from [24,53]. In [18,19] it was proved that for energies in a compact interval around the origin, eigenvectors are weakly delocalized, and for $0 < \alpha < 2/3$ for energies far enough from the origin, eigenvectors are weakly localized. In [3], full delocalization was proved in a compact interval around the origin, and the authors even established GOE local eigenvalue statistics in the same spectral region. In [2], the law of the eigenvector components of Lévy matrices was computed.

Conventions Throughout the following, every quantity that is not explicitly *constant* depends on the fundamental parameter N. We almost always omit this dependence

from our notation. We use C to denote a generic positive universal constant, and write $X = O(Y)$ to mean $|X| \leqslant CY$. For $X, Y > 0$ we write $X \asymp Y$ if $X = O(Y)$ and $Y = O(X)$. We write $X \ll Y$ or $X = o(Y)$ to mean $\lim_{N\to\infty} X/Y = 0$. A vector is *normalized* if its ℓ^2-norm is one.

1.2. Results—the semilocalized phase. Let $\mathbb{G} = \mathbb{G}(N, d/N)$ be the Erdős–Rényi graph with vertex set $[N] := \{1, \dots, N\}$ and edge probability d/N for $0 \leqslant d \leqslant N$. Let $A = (A_{xy})_{x,y\in[N]} \in \{0,1\}^{N\times N}$ be the adjacency matrix of $\mathbb{G}$. Thus, $A = A^*$, $A_{xx} = 0$ for all $x \in [N]$, and $(A_{xy} : x < y)$ are independent Bernoulli(d/N) random variables.

The entrywise nonnegative matrix $A/\sqrt{d}$ has a *trivial* Perron–Frobenius eigenvalue, which is its largest eigenvalue. In the following we only consider the other eigenvalues, which we call *nontrivial*. In the regime $d \gg \sqrt{\log N/\log\log N}$, which we always assume in this paper, the trivial eigenvalue is located at $\sqrt{d}(1+o(1))$, and it is separated from the nontrivial ones with high probability; see [14]. Moreover, without loss of generality in this subsection we always assume that $d \leqslant 3\log N$, for otherwise the semilocalized phase does not exist (see Sect. 1.1).

For $x \in [N]$ we define the *normalized degree* of x as

$$\alpha_x := \frac{1}{d}\sum_{y\in[N]} A_{xy}\,. \tag{1.5}$$

In Theorem 1.7 below we show that the nontrivial eigenvalues of $A/\sqrt{d}$ outside the interval $[-2, 2]$ are in two-to-one correspondence with vertices with normalized degree greater than 2: each vertex x with $\alpha_x > 2$ gives rise to two eigenvalues of $A/\sqrt{d}$ located with high probability near $\pm\Lambda(\alpha_x)$, where we defined the bijective function $\Lambda : [2,\infty) \to [2,\infty)$ through

$$\Lambda(\alpha) := \frac{\alpha}{\sqrt{\alpha-1}}. \tag{1.6}$$

Our main result in the semilocalized phase is about the eigenvectors associated with these eigenvalues. To state it, we need the following notions.

Definition 1.1. Let $\lambda > 2$ and $0 < \delta \leqslant \lambda - 2$. We define the set of *resonant vertices at energy* λ through

$$\mathcal{W}_{\lambda,\delta} := \big\{x : \alpha_x \geqslant 2, |\Lambda(\alpha_x) - \lambda| \leqslant \delta\big\}\,. \tag{1.7}$$

We denote by $B_r(x)$ the ball around the vertex x of radius r for the graph distance in $\mathbb{G}$. Define

$$r_\star = \big\lfloor c\sqrt{\log N}\big\rfloor\,; \tag{1.8}$$

all of our results will hold provided $c > 0$ is chosen to be a small enough universal constant. The quantity $r_\star$ will play the role of a maximal radius for balls around localization centres.

We introduce the basic control parameters

$$\xi := \frac{\sqrt{\log N}}{d}\log d\,, \qquad \xi_u := \frac{\sqrt{\log N}}{d}\frac{1}{u}\,, \tag{1.9}$$

which under our assumptions will always be small (see Remark 1.5 below). We now state our main result in the semilocalized phase.

Theorem 1.2 (Semilocalized phase). *For any $\nu > 0$ there exists a constant C such that the following holds. Suppose that*

$$C\sqrt{\log N}\log\log N \leqslant d \leqslant 3\log N\,. \tag{1.10}$$

Let $\mathbf{w}$ *be a normalized eigenvector of $A/\sqrt{d}$ with nontrivial eigenvalue $\lambda \geqslant 2 + C\xi^{1/2}$. Let $0 < \delta \leqslant (\lambda - 2)/2$. Then for each $x \in \mathcal{W}_{\lambda,\delta}$ there exists a normalized vector $\mathbf{v}(x)$, supported in $B_{r_\star}(x)$, such that the supports of $\mathbf{v}(x)$ and $\mathbf{v}(y)$ are disjoint for $x \neq y$, and*

$$\sum_{x\in\mathcal{W}_{\lambda,\delta}} \langle \mathbf{v}(x)\,,\mathbf{w}\rangle^2 \geqslant 1 - C\left(\frac{\xi + \xi_{\lambda-2}}{\delta}\right)^2$$

with probability at least $1 - CN^{-\nu}$. Moreover, $\mathbf{v}(x)$ decays exponentially around x in the sense that for any $r \geqslant 0$ we have

$$\sum_{y\notin B_r(x)} (\mathbf{v}(x))_y^2 \leqslant \frac{1}{(\alpha_x - 1)^{r+1}}\,.$$

Remark 1.3. An analogous result holds for negative eigenvalues $-\lambda \leqslant -2 - C\xi^{1/2}$, with a different vector $\mathbf{v}(x)$. See Theorem 3.4 and Remark 3.5 below for a precise statement.

Remark 1.4. The upper bound $d \leqslant 3\log N$ in (1.10) is made for convenience and without loss of generality, because if $d > 3\log N$ then, as explained in Sect. 1.1, with high probability the semilocalized phase does not exist, i.e. eigenvalues satisfying the conditions of Theorem 1.2 do not exist.

Theorem 1.2 implies that $\mathbf{w}$ is almost entirely concentrated in the balls around the resonant vertices, and in each such ball $B_{r_\star}(x)$, $x \in \mathcal{W}_{\lambda,\delta}$, the vector $\mathbf{w}$ is almost collinear to the vector $\mathbf{v}(x)$. Thus, $\mathbf{v}(x)$ has the interpretation of the *localization profile around the localization centre* x. Since it has exponential decay, we deduce immediately from Theorem 1.2 that the radius $r_\star$ can be made smaller at the expense of worse error terms. In fact, in Definition 3.2 and Theorem 3.4 below, we give an explicit definition of $\mathbf{v}(x)$, which shows that it is *radial* in the sense that its value at a vertex y depends only on the distance between x and y, in which it is an exponentially decaying function. To ensure that the supports of the vectors $\mathbf{v}(x)$ for different x do not overlap, $\mathbf{v}(x)$ is in fact defined as the restriction of a radial function around x to a subgraph of $\mathbb{G}$, the *pruned graph*, which differs from $\mathbb{G}$ by only a small number of edges and whose balls of radius $r_\star$ around the vertices of $\mathcal{W}_{\lambda,\delta}$ are disjoint (see Proposition 3.1 below). For positive eigenvalues, the entries of $\mathbf{v}(x)$ are nonnegative, while for negative eigenvalues its entries carry a sign that alternates in the distance to x. The set of resonant vertices $\mathcal{W}_{\lambda,\delta}$ is a small fraction of the whole vertex set $[N]$; its size is analysed in Lemma A.12 below.

Remark 1.5. Note that, by the lower bounds imposed on d and λ in Theorem 1.2, we always have $\xi, \xi_{\lambda-2} \leqslant 1/C$.

Using the exponential decay of the localization profiles, it is easy to deduce from Theorem 1.2 that a positive proportion of the eigenvector mass concentrates at the resonant vertices.

Corollary 1.6. *Under the assumptions of Theorem 1.2 we have*

$$\sum_{y\in\mathcal{W}_{\lambda,\delta}} w_y^2 = \frac{\sqrt{\lambda^2-4}}{\lambda+\sqrt{\lambda^2-4}} + O\left(\frac{\mathcal{C}(\xi+\xi_{\lambda-2})}{\delta} + \frac{\mathcal{C}\delta}{\lambda^{5/2}\sqrt{\lambda-2}}\right)$$

with probability at least $1-\mathcal{C}N^{-\nu}$.

Next, we state a rigidity result on the eigenvalue locations in the semilocalized phase. It generalizes [10, Corollary 2.3] by improving the error bound and extending it to the full regime (1.4) of d, below which it must fail (see Sect. 1.5 below). Its proof is a byproduct of the proof of our main result in the semilocalized phase, Theorem 1.2. We denote the ordered eigenvalues of a Hermitian matrix $M \in \mathbb{C}^{N\times N}$ by $\lambda_1(M) \geqslant \lambda_2(M) \geqslant \cdots \geqslant \lambda_N(M)$. We only consider the nontrivial eigenvalues of $A/\sqrt{d}$, i.e. $\lambda_i(A/\sqrt{d})$ with $2 \leqslant i \leqslant N$. For the following statements we order the normalized degrees by choosing a (random) permutation $\sigma \in S_N$ such that $i \mapsto \alpha_{\sigma(i)}$ is nonincreasing.

Theorem 1.7 (Eigenvalue locations in semilocalized phase). *For any* $\nu > 0$ *there exists a constant* $\mathcal{C}$ *such that the following holds. Suppose that* (1.10) *holds. Let*

$$\mathcal{U} := \{x \in [N] : \Lambda(\alpha_x) \geqslant 2+\xi^{1/2}\}.$$

Then with probability at least $1-\mathcal{C}N^{-\nu}$, *for all* $1 \leqslant i \leqslant |\mathcal{U}|$ *we have*

$$|\lambda_{i+1}(A/\sqrt{d}) - \Lambda(\alpha_{\sigma(i)})| + |\lambda_{N-i+1}(A/\sqrt{d}) + \Lambda(\alpha_{\sigma(i)})| \leqslant \mathcal{C}(\xi+\xi_{\Lambda(\alpha_{\sigma(i)})-2}) \tag{1.11}$$

and for all $|\mathcal{U}|+2 \leqslant i \leqslant N-|\mathcal{U}|$ *we have*

$$|\lambda_i(A/\sqrt{d})| \leqslant 2+\xi^{1/2}. \tag{1.12}$$

We remark that the upper bound on d from (1.10), which is necessary for the existence of a semilocalized phase, can be relaxed in Theorem 1.7 to obtain an estimate on $\max_{2\leqslant i\leqslant N}|\lambda_i(A/\sqrt{d})|$ in the supercritical regime $d \geqslant 3\log N$, which is sharper than the one in [10]. The proof is the same and we do not pursue this direction here.

We conclude this subsection with a discussion on the counting function of the normalized degrees, which we use to give estimates on the number of resonant vertices (1.7). For $b \geqslant 0$ and $\alpha \geqslant 2$ define the exponent

$$\theta_b(\alpha) := [1-b(\alpha\log\alpha-\alpha+1)]_+. \tag{1.13}$$

Define $\alpha_{\max}(b) := \inf\{\alpha \geqslant 2 : \theta_b(\alpha) = 0\}$. Thus, θ_b is a nonincreasing function that is nonzero on $[0, \alpha_{\max}(b))$. Moreover, $\theta_b(2) = [1-b/b_*]_+$, so that $\alpha_{\max}(b) > 2$ if and only if $b < b_*$. From Lemma A.9 below it is easy to deduce that if $d \gg 1$ then $\alpha_{\sigma(1)} = \alpha_{\max}(d/\log N) + O(\zeta/d)$ with probability at least $1-o(1)$ for any $\zeta \gg 1$. Thus, $\alpha_{\max}(d/\log N)$ has the interpretation of the deterministic location of the largest normalized degree. See Fig. 4 for a plot of θ_b.

In Appendix A.4 below, we obtain estimates on the density of the normalized degrees $(\alpha_x)_{x\in[N]}$ and combine it with Theorem 1.2 to deduce a lower bound on the ℓ^p-norm of eigenvectors in the semilocalized phase. The precise statements are given in Lemma A.12 and Corollary A.13, which provide quantitative error bounds throughout the regime

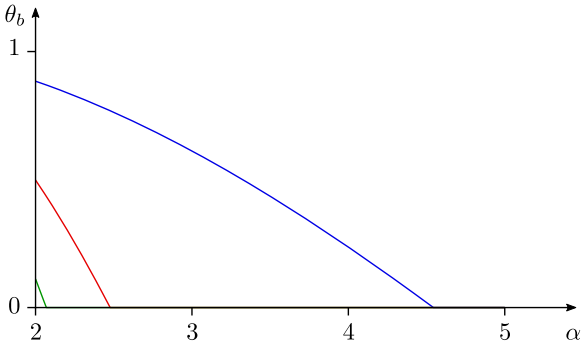

Fig. 4. A plot of the exponent $\theta_b(\alpha)$ as a function of $\alpha \geqslant 2$ for the values $b = 0.3$ (blue), $b = 1.3$ (red), and $b = 2.3$ (green). The graph hits the value 0 at $\alpha_{\max}(b)$

(1.10). Here, we summarize them, for simplicity, in simple qualitative versions in the critical regime $d \asymp \log N$. For $b < b_*$ we abbreviate

$$\lambda_{\max}(b) := \Lambda(\alpha_{\max}(b)), \qquad \rho_b(\lambda) := \begin{cases} \theta_b(\Lambda^{-1}(\lambda)) & \text{if } |\lambda| \geqslant 2 \\ 1 & \text{if } |\lambda| < 2, \end{cases} \tag{1.14}$$

where $\Lambda^{-1}(\lambda) = \frac{\lambda^2}{2}(1 + \sqrt{1 - 4/\lambda^2})$ for $|\lambda| \geqslant 2$. Let $d = b \log N$ with some constant $b < b_*$, and suppose that $2 + \kappa \leqslant \lambda \leqslant \lambda_{\max}(b) - \kappa$ for some constant $\kappa > 0$. Then Lemma A.12 (ii) implies (choosing $1/d \ll \delta \ll 1$)

$$|\mathcal{W}_{\lambda,\delta}| = N^{\rho_b(\lambda)+o(1)} \tag{1.15}$$

with probability $1 - o(1)$. From (1.15) and Theorem 1.2 we obtain, for any $2 \leqslant p \leqslant \infty$,

$$\|\mathbf{w}\|_p^2 \geqslant N^{(2/p-1)\rho_b(\lambda)+o(1)} \tag{1.16}$$

with probability $1 - o(1)$ (see Corollary A.13 below). In other words, the localization exponent $\gamma(\mathbf{w})$ from (1.1) satisfies $\gamma(\mathbf{w}) \leqslant \rho_b(\lambda) + o(1)$. See Fig. 2 for an illustration of the bound (1.16) for $p = \infty$. We remark that the exponent $\rho_b(\lambda)$ also describes the density of states at energy λ: under the above assumptions on b and λ, for any interval I containing λ and satisfying $\xi \ll |I| \ll 1$, the number of eigenvalues in I is equal to $N^{\rho_b(\lambda)+o(1)}|I|$ with probability $1 - o(1)$, as can be seen from Lemma A.12 (i) and Theorem 1.7.

1.3. Results—the delocalized phase. Let A be the adjacency matrix of $\mathbb{G}(N, d/N)$, as in Sect. 1.2. For $0 < \kappa < 1/2$ define the spectral region

$$\mathcal{S}_\kappa := [-2 + \kappa, -\kappa] \cup [\kappa, 2 - \kappa]. \tag{1.17}$$

Theorem 1.8 (Delocalized phase). *For any $\nu > 0$ and $\kappa > 0$ there exists a constant $\mathcal{C} > 0$ such that the following holds. Suppose that*

$$\mathcal{C}\sqrt{\log N} \leqslant d \leqslant (\log N)^{3/2}. \tag{1.18}$$

Let $\mathbf{w}$ *be a normalized eigenvector of* $A/\sqrt{d}$ *with eigenvalue* $\lambda \in \mathcal{S}_\kappa$. *Then*

$$\|\mathbf{w}\|_\infty^2 \leqslant N^{-1+\kappa} \tag{1.19}$$

with probability at least $1 - \mathcal{C}N^{-\nu}$.

In the delocalized phase, i.e. in $\mathcal{S}_\kappa$, we also show that the spectral measure of $A/\sqrt{d}$ at any vertex x is well approximated by the spectral measure at the root of $\mathbb{T}_{d\alpha_x,d}$, the infinite rooted $(d\alpha_x, d)$-regular tree, whose root has $d\alpha_x$ children and all other vertices have d children. This approximation is a local law, valid for intervals containing down to N^κ eigenvalues. See Remark 4.4 as well as Remark 4.3 and Appendix A.2 below for details.

Remark 1.9. In [42] it is shown that (1.19) holds with probability at least $1 - \mathcal{C}N^{-\nu}$ for *all* eigenvectors provided that

$$\mathcal{C} \log N \leqslant d \leqslant N/2\,. \tag{1.20}$$

This shows that the upper bound in (1.18) is in fact not restrictive.

Remark 1.10 (*Optimality of* (1.18) *and* (1.20)). Both lower bounds in (1.18) and (1.20) are optimal (up to the value of $\mathcal{C}$), in the sense that delocalization fails in each case if these lower bounds are relaxed. See Sect. 1.5 below.

We note that the domain $\mathcal{S}_\kappa$ is optimal, up to the choice of $\kappa > 0$. Indeed, as explained in Sect. 1.5 below, delocalization fails in the neighbourhood of the origin, owing to a proliferation highly localized tuning fork states. Similarly, we expect the delocalization to fail in the neighbourhoods of ± 2, where the masses of the eigenvectors become concentrated on vertices x with normalized degrees α_x close to 2. The neighbourhoods of $0, \pm 2$ are also singled out as the regions where the self-consistent equation used to prove Theorem 1.8 (see Lemma 4.16) becomes unstable. This instability is directly related to the appearance of singularities in the spectral measure of the tree $\mathbb{T}_{d\alpha_x,d}$ (see (4.11) and Fig. 8 for an illustration). The singularity near 0 occurs when α_x is close to 0, and the singularities near ± 2 when α_x is close to 2. See Fig. 10 for a simulation that demonstrates numerically the failure of delocalization outside of $\mathcal{S}_\kappa$.

1.4. Extension to general sparse random matrices. Our results, Theorems 1.2, 1.7, and 1.8, hold also for the following family of sparse Wigner matrices. Let $A = (A_{xy})$ be the adjacency matrix of $\mathbb{G}(N, d/N)$ as above and $W = (W_{xy})$ be an independent Wigner matrix with bounded entries. That is, W is Hermitian and its upper triangular entries $(W_{xy} : x \leqslant y)$ are independent complex-valued random variables with mean zero and variance one, $\mathbb{E}|W_{xy}|^2 = 1$, and $|W_{xy}| \leqslant K$ 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 $M_{xy} := A_{xy} W_{xy}$. Since the entries of $M/\sqrt{d}$ are centred, it does not have a trivial eigenvalue like $A/\sqrt{d}$.

Theorem 1.11. *Let* $M = (M_{xy})_{x,y\in[N]}$ *be a sparse Wigner matrix. Define*

$$\alpha_x = \frac{1}{d} \sum_{y\in[N]} |M_{xy}|^2. \tag{1.21}$$

Theorems 1.2 and 1.8 hold with (1.21) if A is replaced with M, and Theorem 1.7 holds with (1.21) if $\lambda_{i+1}(A/\sqrt{d})$, $\lambda_{N-i+1}(A/\sqrt{d})$, and $\lambda_i(A/\sqrt{d})$ are replaced with $\lambda_i(M/\sqrt{d})$, $\lambda_{N-i+1}(M/\sqrt{d})$, and $\lambda_i(M/\sqrt{d})$, respectively. Here, the constants $\mathcal{C}$ depend on K in addition to ν and κ.

The modifications to the proofs of Theorems 1.2 and 1.7 required to establish Theorem 1.11 are minor and follow along the lines of [10, Section 10]. The modification to the proof of Theorem 1.8 is trivial, since the assumptions of the general Theorem 4.2 below include the sparse Wigner matrix M. We also remark that, with some extra work, one can relax the boundedness assumption on the entries of W, which we shall however not do here.

1.5. The limits of sparseness and the scale $d \asymp \sqrt{\log N}$. We conclude this section with a discussion on how sparse $\mathbb{G}$ can be for our results to remain valid. We show that all of our results—Theorems 1.2, 1.7, and 1.8—are wrong below the regime (1.4), i.e. if d is smaller than order $\sqrt{\log N}$. Thus, our sparseness assumptions—the lower bounds on d from (1.10) and (1.18)—are optimal (up to the factor $\log\log N$ in (1.10) and the factor $\mathcal{C}$ in (1.18)). The fundamental reason for this change of behaviour will turn out to be that the ratio $|S_2(x)|/|S_1(x)|$ concentrates if and only if $d \gg \sqrt{\log N}$, where $S_i(x)$ denotes the sphere in $\mathbb{G}$ of radius i around x. This can be easily made precise with a well-known *tuning fork* construction, detailed below.

In the critical and subcritical regime $1 \ll d = O(\log N)$, the graph $\mathbb{G}$ is in general not connected, but with probability $1 - o(1)$ it has a unique giant component $\mathbb{G}_{\text{giant}}$ with at least $N(1-\mathrm{e}^{-d/4})$ vertices (see Corollary A.15 below). Moreover, the spectrum of $A/\sqrt{d}$ restricted to the complement of the giant component is contained in the $O\big(\frac{\sqrt{\log N}}{d}\big)$-neighbourhood of the origin (see Corollary A.16 below). Since we always assume $d \geqslant \mathcal{C}\sqrt{\log N}$ and we only consider eigenvalues in $\mathbb{R} \setminus [-\kappa, \kappa]$, we conclude that all of our results listed above only pertain to the eigenvalues and eigenvectors of the giant component.

For $D = 0, 1, 2, \ldots$ we introduce a *star*[1] *tuning fork of degree D rooted in $\mathbb{G}_{\text{giant}}$*, or *$D$-tuning fork* for short, which is obtained by taking two stars with central degree D and connecting their hubs to a common base vertex in $\mathbb{G}_{\text{giant}}$. We refer to Fig. 5 for an illustration and Definition A.17 below for a precise definition.

It is not hard to see that every D-tuning fork gives rise to two eigenvalues $\pm\sqrt{D/d}$ of $A/\sqrt{d}$ restricted to $\mathbb{G}_{\text{giant}}$, whose associated eigenvectors are supported on the stars (see Lemma A.18 below). We denote by $\Sigma := \{\sqrt{D/d} : \text{a } D\text{-tuning fork exists}\}$ the spectrum of $A/\sqrt{d}$ restricted to $\mathbb{G}_{\text{giant}}$ generated by the tuning forks. Any eigenvector associated with an eigenvalue $\sqrt{D/d} \in \Sigma$ is localized on precisely $2D + 2$ vertices. Thus, D-tuning forks provide a simple way of constructing localized states. Note that this is a very basic form of concentration of mass, supported at the periphery of the graph on special graph structures, and is unrelated to the much more subtle concentration in the semilocalized phase described in Sect. 1.2.

For $d > 0$ and $D \in \mathbb{N}$ we now estimate the number of D-tuning forks in $\mathbb{G}(N, d/N)$, which we denote by $F(d, D)$. The following result is proved in Appendix A.6.

Lemma 1.12 (Number of D-tuning forks). *Suppose that $1 \ll d = b\log N = O(\log N)$ and $0 \leqslant D \ll \log N/\log\log N$. Then $F(d, D) = N^{1-2b-2bD+o(1)}$ with probability $1 - o(1)$.*

[1] For simplicity we only consider stars, but the same argument can be applied to arbitrary trees.

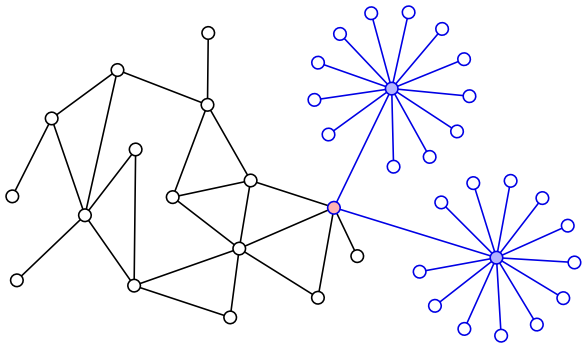

Fig. 5. A star tuning fork of degree 12 rooted in a graph. The tuning fork is highlighted in blue. Its base is filled with red and its two hubs are filled with blue

Defining $D_* := \frac{\log N}{2d} - 1$, we immediately deduce the following result.

Corollary 1.13. *For any constant $\varepsilon > 0$ with probability $1 - o(1)$ the following holds. If $D_* \leqslant -\varepsilon$ then $\Sigma = \emptyset$. If $D_* \geqslant \varepsilon$ then $\Sigma = \{\pm\sqrt{D/d} : D \in \mathbb{N}, D \leqslant D_*(1 + o(1))\}$.*

We deduce that if $d \leqslant (1/2 - \varepsilon) \log N$ then $\Sigma \neq \emptyset$ and hence the delocalization for all eigenvectors from Remark 1.9 fails. Hence, the lower bound (1.20) is optimal up to the value of $\mathcal{C}$.

Similarly, for $d \gg \sqrt{\log N}$ the set Σ is in general nonempty, but we always have $\Sigma \subset [-\kappa, \kappa]$ for any fixed $\kappa > 0$, so that eigenvalues from Σ do not interfere with the statements of Theorems 1.2, 1.7, and 1.8. On the other hand, if $d = \sqrt{\log N}/t$ for constant t, we find that Σ is asymptotically dense in the interval $[-t/\sqrt{2}, t/\sqrt{2}]$. Since the conclusions of Theorems 1.2, 1.7, and 1.8 are obviously wrong for any eigenvalue from Σ, they must all be wrong for large enough t. This shows that the lower bounds d from (1.10) and (1.18) are optimal (up to the factor $\log\log N$ in (1.10) and the factor $\mathcal{C}$ in (1.18)).

In fact, the emergence of the tuning fork eigenvalues of order one and the failure of all of our proofs has the same underlying root cause, which singles out the scale $d \asymp \sqrt{\log N}$ as the scale below which the concentration of the ratio

$$|S_2(x)|/|S_1(x)| = d(1 + o(1)) \tag{1.22}$$

fails for vertices x satisfying $D_x \asymp d$. Clearly, to have a D-tuning fork with $D \asymp d$, (1.22) has to fail at the hubs of the stars. Moreover, (1.22) enters our proofs of both the semilocalized and the delocalized phase in a crucial way. For the former, it is linked to the validity of the local approximation by the (D_x, d)-regular tree from Appendix A.2, which underlies also the construction of the localization profile vectors (see e.g. (3.35) below). For the latter, in the language of Definition 4.6 below, it is linked to the property that most neighbours of any vertex are typical (see Proposition 4.8 (ii) below).

2. Basic Definitions and Overview of Proofs

In this preliminary section we introduce some basic notations and definitions that are used throughout the paper, and give an overview of the proofs of Theorems 1.2 (semilocalized phase) and 1.8 (delocalized phase). These proofs are unrelated and, thus, explained separately. For simplicity, in this overview we only consider qualitative error terms of the form $o(1)$, although all of our estimates are in fact quantitative.

2.1. Basic definitions. We write $\mathbb{N} = \{0, 1, 2, \dots\}$. We set $[n] := \{1, \dots, n\}$ for any $n \in \mathbb{N}^*$ and $[0] := \emptyset$. We write $|X|$ for the cardinality of a finite set X. We use $\mathbb{1}_\Omega$ as symbol for the indicator function of the event Ω.

Vectors in $\mathbb{R}^N$ are denoted by boldface lowercase Latin letters like $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$. We use the notation $\mathbf{v} = (v_x)_{x\in[N]} \in \mathbb{R}^N$ for the entries of a vector. We denote by $\operatorname{supp} \mathbf{v} := \{x \in [N] : v_x \neq 0\}$ the support of a vector $\mathbf{v}$. We denote by $\langle \mathbf{v}, \mathbf{w}\rangle = \sum_{x\in[N]} v_x w_x$ the Euclidean scalar product on $\mathbb{R}^N$ and by $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle}$ the induced Euclidean norm. For a matrix $M \in \mathbb{R}^{N\times N}$, $\|M\|$ is its operator norm induced by the Euclidean norm on $\mathbb{R}^N$. 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 $\mathbf{1}_S := \sum_{x\in S} \mathbf{1}_x$. In particular, $\mathbf{1}_{\{x\}} = \mathbf{1}_x$.

We use blackboard bold letters to denote graphs. Let $\mathbb{H} = (V(\mathbb{H}), E(\mathbb{H}))$ be a (simple, undirected) graph on the vertex set $V(\mathbb{H}) = [N]$. We often identify a graph $\mathbb{H}$ with its set of edges $E(\mathbb{H})$. We denote by $A^{\mathbb{H}} \in \{0, 1\}^{N\times N}$ the adjacency matrix of $\mathbb{H}$. For $r \in \mathbb{N}$ and $x \in [N]$, we denote by $B_r^{\mathbb{H}}(x)$ the closed ball of radius r around x in the graph $\mathbb{H}$, i.e. the set of vertices at distance (with respect to $\mathbb{H}$) at most r from the vertex x. We denote the sphere of radius r around the vertex x by $S_r^{\mathbb{H}}(x) := B_r^{\mathbb{H}}(x) \setminus B_{r-1}^{\mathbb{H}}(x)$. We denote by $D_x^{\mathbb{H}}$ the degree of the vertex x in the graph $\mathbb{H}$. For any subset $V \subset [N]$, we denote by $\mathbb{H}|_V$ the subgraph induced by $\mathbb{H}$ on V. If $\mathbb{H}$ is a subgraph of $\mathbb{G}$ then we denote by $\mathbb{G} \setminus \mathbb{H}$ the graph on $[N]$ with edge set $E(\mathbb{G}) \setminus E(\mathbb{H})$. In the above definitions, if the graph $\mathbb{H}$ is the Erdős–Rényi graph $\mathbb{G}$, we systematically omit the superscript $\mathbb{G}$.

The following notion of very high probability is a convenient shorthand used throughout the paper. It simplifies considerably the probabilistic statements of the kind that appear in Theorems 1.2, 1.7, and 1.8. It also introduces two special symbols, ν and $\mathcal{C}$, which appear throughout the rest of the paper.

Definition 2.1. 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 $\mathcal{C} \equiv \mathcal{C}_\nu$ such that

$$\mathbb{P}(\Xi_{N,\nu}) \geqslant 1 - \mathcal{C}_\nu N^{-\nu}$$

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

Convention 2.2. In statements that hold with very high probability, we use the special symbol $\mathcal{C} \equiv \mathcal{C}_\nu$ to denote a generic positive constant depending on ν such that the statement holds with probability at least $1 - \mathcal{C}_\nu N^{-\nu}$ provided $\mathcal{C}_\nu$ is chosen large enough. Thus, the bound $|X| \leqslant \mathcal{C}Y$ *with very high probability* means that, for each $\nu > 0$, there is a constant $\mathcal{C}_\nu > 0$, depending on ν, such that

$$\mathbb{P}\big(|X| \leqslant \mathcal{C}_\nu Y\big) \geqslant 1 - \mathcal{C}_\nu N^{-\nu}$$

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 $|X| \leqslant \mathcal{C}Y$.

We remark that the notion of very high probability from Definition 2.1 survives a union bound involving $N^{O(1)}$ events. We shall tacitly use this fact throughout the paper. Moreover, throughout the paper, the constant $\mathcal{C} \equiv \mathcal{C}_\nu$ in the assumptions (1.10) and (1.18) is always assumed to be large enough.

2.2. Overview of proof in semilocalized phase. The starting point of the proof of Theorem 1.2 is the following simple observation. Suppose that M is a Hermitian matrix with eigenvalue λ and associated eigenvector $\mathbf{w}$. Let Π be an orthogonal projection and write $\overline{\Pi} := I - \Pi$. If λ is not an eigenvalue of $\overline{\Pi} M \overline{\Pi}$ then from $(M - \lambda)\mathbf{w} = 0$ we deduce

$$\overline{\Pi}\mathbf{w} = -(\overline{\Pi} M \overline{\Pi} - \lambda)^{-1} \overline{\Pi} M \Pi \mathbf{w} . \tag{2.1}$$

If Π is an eigenprojection of M whose range contains the eigenspace of λ (for instance $\Pi = \mathbf{w}\mathbf{w}^*$ if λ is simple) then clearly both sides of (2.1) vanish. The basic idea of our proof is to apply an approximate version of this observation to $M = A/\sqrt{d}$, by choosing Π appropriately, and showing that the left-hand side of (2.1) is small by estimating the right-hand side.

In fact, we choose[2]

$$\Pi := \sum_{x \in \mathcal{W}_{\lambda,\delta}} \mathbf{v}(x)\, \mathbf{v}(x)^* , \tag{2.2}$$

where $\mathcal{W}_{\lambda,\delta}$ is the set (1.7) of resonant vertices at energy λ, and $\mathbf{v}(x)$ is the exponentially decaying localization profile from Theorem 1.2. The proof then consists of two main ingredients:

(a) $\|\overline{\Pi} M \Pi\| = o(1)$;
(b) $\overline{\Pi} M \overline{\Pi}$ has a spectral gap around λ.

Informally, (a) states that Π is close to a spectral projection of M, as $\overline{\Pi} M \Pi = [M, \Pi]\Pi$ quantifies the noncommutativity of M and Π on the range of Π. Similarly, (b) states that Π projects roughly onto an eigenspace of M of energies near λ. Plugging (a) and (b) into (2.1) yields an estimate on $\|\overline{\Pi}\mathbf{w}\|$ from which Theorem 1.2 follows easily. Thus, the main work of the proof is to establish the properties (a) and (b) for the specific choice of Π from (2.2).

The construction of the localization profile $\mathbf{v}(x)$ uses the *pruned graph* $\mathbb{G}_\tau$ from [10], a subgraph of $\mathbb{G}$ depending on a threshold $\tau > 1$, which differs from $\mathbb{G}$ by only a small number of edges and whose balls of radius $r_\star$ around the vertices of $\mathcal{V}_\tau := \{x : \alpha_x \geqslant \tau\}$ are disjoint (see Proposition 3.1 below). Now we define the vector $\mathbf{v}(x) := \mathbf{v}_+^\tau(x)$, where, for $\sigma = \pm$ and $\tau > 1$,

$$\mathbf{v}_\sigma^\tau(x) := \sum_{i=0}^{r_\star} \sigma^i u_i(x) \mathbf{1}_{S_i^{\mathbb{G}_\tau}(x)} / \|\mathbf{1}_{S_i^{\mathbb{G}_\tau}(x)}\| , \qquad u_i(x) := \frac{\sqrt{\alpha_x}}{(\alpha_x - 1)^{i/2}} u_0 \quad (1 \leqslant i \leqslant r_\star) . \tag{2.3}$$

The motivation behind this choice is explained in Appendix A.2: with high probability, the $r_\star$-neighbourhood of x in $\mathbb{G}_\tau$ looks roughly like that of the root of infinite regular tree $\mathbb{T}_{D_x,d}$ whose root has D_x children and all other vertices d children. The adjacency matrix of $\mathbb{T}_{D_x,d}$ has the exact eigenvalues $\pm\sqrt{d}\Lambda(\alpha_x)$ with the corresponding eigenvectors given by (2.3) with $\mathbb{G}_\tau$ replaced with $\mathbb{T}_{D_x,d}$.

The central idea of our proof is the introduction of a block diagonal approximation of the pruned graph. Define the orthogonal projections

$$\Pi^\tau := \sum_{x \in \mathcal{V}_{2+o(1)}} \sum_{\sigma=\pm} \mathbf{v}_\sigma^\tau(x) \mathbf{v}_\sigma^\tau(x)^* , \qquad \overline{\Pi}^\tau := I - \Pi^\tau .$$

[2] This projection Π is denoted by $\Pi^\tau_{\lambda,\delta}$ in the proof of Theorem 3.4 below.

The range of Π from (2.2) is a subspace of the range of Π^τ, i.e. $\Pi\Pi^\tau = \Pi$. The interpretation of Π^τ is the orthogonal projection onto all localization profiles around vertices x with normalized degree at least $2 + o(1)$, which is precisely the set of vertices around which one can define an exponentially decaying localization profile. Now we define the *block diagonal approximation of the pruned graph* as

$$\widehat{H}^\tau := \sum_{x \in \mathcal{V}_{2+o(1)}} \sum_{\sigma=\pm} \sigma \Lambda(\alpha_x) \mathbf{v}_\sigma^\tau(x) \mathbf{v}_\sigma^\tau(x)^* + \overline{\Pi}^\tau H^\tau \overline{\Pi}^\tau ; \tag{2.4}$$

here we defined the centred and scaled adjacency matrix $H^\tau := A^{\mathbb{G}_\tau}/\sqrt{d} - E^\tau$, where E^τ is a suitably chosen matrix that is close to $\mathbb{E}A^{\mathbb{G}}/\sqrt{d}$ and preserves the locality of $A^{\mathbb{G}_\tau}$ in balls around the vertices of $\mathcal{V}_\tau$. In the subspace spanned by the localization profiles $\{\mathbf{v}_\sigma^\tau(x) : \sigma = \pm, x \in \mathcal{V}_{2+o(1)}\}$, $\widehat{H}^\tau$ is diagonal with eigenvalues $\sigma\Lambda(\alpha_x)$. In the orthogonal complement, it is equal to H^τ. The off-diagonal blocks are zero. The main work of our proof consists in an analysis of $\widehat{H}^\tau$.

In terms of $\widehat{H}^\tau$, abbreviating $H := (A^{\mathbb{G}} - \mathbb{E}A^{\mathbb{G}})/\sqrt{d}$, the problem of showing (a) and (b) reduces to showing

(c) $\|H - \widehat{H}^\tau\| = o(1)$,
(d) $\|\overline{\Pi}^\tau H^\tau \overline{\Pi}^\tau\| \leqslant 2 + o(1)$.

Indeed, ignoring minor issues pertaining to the centring $\mathbb{E}A^{\mathbb{G}}$, we replace $M = A^{\mathbb{G}}/\sqrt{d}$ with H in (a) and (b). Then (a) follows immediately from (c), since $\overline{\Pi} H \Pi = \|\overline{\Pi}\widehat{H}^\tau\Pi\| + o(1) = o(1)$, as $\overline{\Pi}\widehat{H}^\tau\Pi = 0$ by the block structure of $\widehat{H}^\tau$ and the relation $\Pi^\tau\Pi = \Pi$. To show (b), we note that the Π^τ-block of $\widehat{H}^\tau$, $\Pi^\tau\widehat{H}^\tau\Pi^\tau = \sum_{x\in\mathcal{V}_{2+o(1)}}\sum_{\sigma=\pm}\sigma\Lambda(\alpha_x)\mathbf{v}_\sigma^\tau(x)\mathbf{v}_\sigma^\tau(x)^*$, trivially has a spectral gap: $\overline{\Pi}\Pi^\tau H^\tau \Pi^\tau\overline{\Pi}$ has no eigenvalues in the δ-neighbourhood of λ, simply because the projection $\overline{\Pi}$ removes the projections $\mathbf{v}_\sigma^\tau(x)\mathbf{v}_\sigma^\tau(x)^*$ with eigenvalues $\sigma\Lambda(\alpha_x)$ in the δ-neighbourhood of λ. Moreover, the $\overline{\Pi}^\tau$-block also has such a spectral gap by (d) and $\lambda > 2 + o(1)$. Hence, by (c), we deduce the desired spectral gap (b).

Thus, what remains is the proof of (c) and (d). To prove (c), we prove $\|H - H^\tau\| = o(1)$ and $\|H^\tau - \widehat{H}^\tau\| = o(1)$. The bound $\|H - H^\tau\| = o(1)$ follows from a detailed analysis of the graph $\mathbb{G} \setminus \mathbb{G}_\tau$ removed from $\mathbb{G}$ to obtain the pruned graph $\mathbb{G}_\tau$, which we decompose as a union of a graph of small maximal degree and a forest, to which standard estimates of adjacency matrices of graphs can be applied (see Lemma 3.8 below). To prove $\|H^\tau - \widehat{H}^\tau\| = o(1)$, we first prove that $\mathbf{v}_\sigma^\tau(x)$ is an approximate eigenvector of H^τ with approximate eigenvalue $\sigma\Lambda(\alpha_x)$ (see Proposition 3.9 below). Then we deduce $\|H^\tau - \widehat{H}^\tau\| = o(1)$ using that the balls $B_{2r_\star}(x)$, $x \in \mathcal{V}_{2+o(1)}$, are disjoint and the locality of the operator H^τ (see Lemma 3.11 below). Thus we obtain (c).

Finally, we sketch the proof of (d). The starting point is an observation going back to [10,15]: from an estimate on the spectral radius of the nonbacktracking matrix associated with H from [15] and an Ihara–Bass-type formula relating the spectra of H and its nonbacktracking matrix from [15], we obtain the quadratic form inequality $|H| \leqslant I + Q + o(1)$ with very high probability, where $Q = \operatorname{diag}(\alpha_x : x \in [N])$, $|H|$ is the absolute value of the Hermitian matrix H, and $o(1)$ is in the sense of operator norm (see Proposition 3.13 below). Using (c), we deduce the inequality

$$|\widehat{H}^\tau| \leqslant I + Q + o(1) . \tag{2.5}$$

To estimate $\|\overline{\Pi}^\tau H^\tau \overline{\Pi}^\tau\|$, we take a normalized eigenvector $\mathbf{w}$ of $\overline{\Pi}^\tau H^\tau \overline{\Pi}^\tau$ with maximal eigenvalue $\lambda > 0$. Thus, $\mathbf{w} \perp \mathbf{v}_\pm^\tau(x)$ for all $x \in \mathcal{V}_{2+o(1)}$. We estimate $\overline{\Pi}^\tau H^\tau \overline{\Pi}^\tau$ from

above (an analogous argument yields an estimate from below) using (2.5) to get

$$\lambda \leqslant 1 + o(1) + \sum_x \alpha_x w_x^2 \leqslant 1 + \tau + o(1) + \max_x \alpha_x \sum_{x \in \mathcal{V}_\tau} w_x^2 \,. \tag{2.6}$$

Choosing $\tau = 1 + o(1)$, we see that (d) follows provided that we can show that

$$\sum_{x \in \mathcal{V}_\tau} w_x^2 = o(1/\log N) \,, \tag{2.7}$$

since $\max_x \alpha_x \leqslant \mathcal{C} \log N$ with very high probability.

The estimate (2.7) is a *delocalization bound*, in the vertex set $\mathcal{V}_\tau$, for any eigenvector $\mathbf{w}$ of $\widehat{H}^\tau$ that is orthogonal to $\mathbf{v}_\pm^\tau(x)$ for all $x \in \mathcal{V}_{2+o(1)}$ and whose associated eigenvalue is larger than $2\tau + o(1)$. It crucially relies on the assumption that $\mathbf{w} \perp \mathbf{v}_\pm^\tau(x)$ for all $x \in \mathcal{V}_{2+o(1)}$, without which it is false (see Proposition 3.14 below). The underlying principle behind its proof is the same as that of the Combes–Thomas estimate [25]: the Green function $((\lambda - Z)^{-1})_{ij}$ of a local operator Z at a spectral parameter λ separated from the spectrum of Z decays exponentially in the distance between i and j, at a rate inversely proportional to the distance from λ to the spectrum of Z. We in fact use a radial form of a Combes–Thomas estimate, where Z is the tridiagonalization of a local restriction of $\widehat{H}^\tau$ around a vertex $x \in \mathcal{V}_\tau$ (see Appendix A.2) and i, j index radii of concentric spheres. The key observation is that, by the orthogonality assumption on $\mathbf{w}$, the Green function $((\lambda - Z)^{-1})_{ir_\star}$, $0 \leqslant i < r_\star$, and the eigenvector components in the radial basis u_i, $0 \leqslant i < r_\star$, satisfy the same linear difference equation. Thus we obtain exponential decay for the components u_i, which yields $u_0^2 \leqslant o(1/\log N) \sum_{i=0}^{r_*} u_i^2$. Going back to the original vertex basis, this implies that $w_x^2 \leqslant o(1/\log N) \|\mathbf{w}|_{B_{2r_\star}^{\mathbb{G}_\tau}(x)}\|^2$ for all $x \in \mathcal{V}_\tau$, from which (2.7) follows since the balls $B_{2r_\star}^{\mathbb{G}_\tau}(x)$, $x \in \mathcal{V}_\tau$, are disjoint.

2.3. Overview of proof in delocalized phase. The delocalization result of Theorem 1.8 is an immediate consequence of a *local law* for the matrix $A/\sqrt{d}$, which controls the entries of the Green function

$$G \equiv G(z) := \left(A/\sqrt{d} - z\right)^{-1}$$

in the form of high-probability estimates, for *spectral scales* $\operatorname{Im} z$ down to the optimal scale $1/N$, which is the typical eigenvalue spacing. Such a local law was first established for $d \gg (\log N)^6$ in [35] and extended down to $d \geqslant \mathcal{C} \log N$ in [42]. In both of these works, the diagonal entries of G are close to the Stieltjes transform of the semicircle law. In contrast, in the regime (1.4) the diagonal entry G_{xx} is close to the Stieltjes transform of the spectral measure at the root of an infinite (D_x, d)-regular tree. Hence, G_{xx} does not concentrate around a deterministic quantity.

The basic approach of the proof is the same as for any local law: derive an approximate self-consistent equation with very high probability, solve it using a stability analysis, and perform a bootstrapping from large to small values of $\operatorname{Im} z$. For a set $T \subset [N]$ denote by $A^{(T)}$ the adjacency matrix of the graph $\mathbb{G}$ where the vertices of T (and all incident edges) have been removed, and denote by $G^{(T)} = \left(A^{(T)}/\sqrt{d} - z\right)^{-1}$ the associated Green function. In order to understand the emergence of the self-consistent equation, it is instructive to consider the toy situation where, for a given vertex x, all neighbours $S_1(x)$ are in different connected components of $A^{(x)}$. This is for instance the case if $\mathbb{G}$ is

a tree. On the *global scale*, where $\operatorname{Im} z$ is large enough, this assumption is in fact valid to a good approximation, since the neighbourhood of x is with high probability a tree. Then a simple application of Schur's complement formula and the resolvent identity yield

$$\frac{1}{G_{xx}} = -z - \frac{1}{d}\sum_{y\in S_1(x)} G^{(x)}_{yy}\,, \qquad G^{(x)}_{yy} - G_{yy} = (G^{(x)}_{yy})^2\frac{1}{d}G_{xx}\,. \tag{2.8}$$

Thus, on the global scale, using that G is bounded, we obtain the self-consistent equation

$$\frac{1}{G_{xx}} = -z - \frac{1}{d}\sum_{y\in S_1(x)} G_{yy} + o(1) \tag{2.9}$$

with very high probability.

It is instructive to solve the self-consistent equation (2.9) in the family $(G_{xx})_{x\in[N]}$ on the global scale. To that end, we introduce the notion of *typical vertices*, which is roughly the set $\mathcal{T} = \{x \in [N] : \alpha_x = 1+o(1)\}$. (In fact, as explained below, the actual definition for local scales has to be different; see (2.12) below.) A simple argument shows that with very high probability most neighbours of any vertex are typical. With this definition, we can try to solve (2.9) on the global scale as follows. From the boundedness of G we obtain a self-consistent equation for the vector $(G_{xx})_{x\in\mathcal{T}}$ that reads

$$\frac{1}{G_{xx}} = -z - \sum_{y\in\mathcal{T}} \frac{1}{d}A_{xy}G_{yy} + \zeta_x\,, \qquad \zeta_x = o(1)\,. \tag{2.10}$$

It is not hard to see that the equation (2.10) has a unique solution, which satisfies $G_{xx} = m + o(1)$ for all $x \in \mathcal{T}$. Here m is the Stieltjes transform of the semicircle law, which satisfies $m = \frac{1}{-z-m}$. Plugging this solution back into (2.9) and using that most neighbours of any vertex are typical shows that for $x \notin \mathcal{T}$ we have $G_{xx} = m_{\alpha_x} + o(1)$, where $m_\alpha := \frac{1}{-z-\alpha m}$. One readily finds (see Appendix A.2 below) that m_{α_x} is Stieltjes transform of the spectral measure of the infinite (D_x, d)-regular tree at the root.

The first main difficulty of the proof is to provide a derivation of identities of the form (2.8) (and hence a self-consistent equation of the form (2.9)) on the local scale $\operatorname{Im} z \ll 1$. We emphasize that the above derivation of (2.8) is completely wrong on the local scale. Unlike on the global scale, on the local scale the behaviour of the Green function is not governed by the local geometry of the graph, and long cycles contribute to G in an essential way. In particular, eigenvector delocalization, which follows from the local law, is a global property of the graph and cannot be addressed using local arguments; it is in fact wrong outside of the region $\mathcal{S}_\kappa$, although the above derivation is insensitive to the real part of z.

We address this difficulty by replacing the identities (2.8) with the following argument, which ultimately provides an a posteriori justification of approximate versions of (2.8) with very high probability, provided we are in the region $\mathcal{S}_\kappa$. We make an a priori assumption that the entries of G are bounded with very high probability; we propagate this assumption from large to small scales using a standard bootstrapping argument and the uniform boundedness of the density of the spectral measure associated with m_α. It is precisely this uniform boundedness requirement that imposes the restriction to $\mathcal{S}_\kappa$ in our local law (as explained in Remark 1.10, this restriction is necessary). The key tool that replaces the simpleminded approximation (2.8) is a series of large deviation estimates for sparse random vectors proved in [42], which, as it turns out, are effective for the full

optimal regime (1.4). Thus, under the bootstrapping assumption that the entries of G are bounded, we obtain (2.8) (and hence also (2.9)), with some additional error terms, with very high probability.

The second main difficulty of the proof is that, on the local scale and for sparse graphs, the self-consistent equation (2.10), which can be derived from (2.9) as explained above, is not stable enough to be solved in $(G_{xx})_{x\in\mathcal{T}}$. This problem stems from the sparseness of the graphs that we are considering, and does not appear in random matrix theory for denser (or even heavy-tailed) matrices. Indeed, the stability estimates of (2.10) carry a logarithmic factor, which is usually of no concern in random matrix theory but is deadly for the sparse regime of this paper. This is a major obstacle and in fact ultimately dooms the self-consistent equation (2.10). To explain the issue, write the sum in (2.10) as $\sum_y S_{xy}G_{yy}$, where S is the $\mathcal{T}\times\mathcal{T}$ matrix $S_{xy} = \frac{1}{d}A_{xy}$. Writing $G_{xx} = m + \varepsilon_x$, plugging it into (2.10), and expanding to first order in ε_x, we obtain, using the definition of m, that $\varepsilon_x = -m^2((I-m^2S)^{-1}\zeta)_x$. Thus, in order to deduce smallness of ε_x from the smallness of ζ_x, we need an estimate on the norm[3] $\|(I-m^2S)^{-1}\|_{\infty\to\infty}$. In Appendix A.10 below we show that for typical S, $\mathrm{Re}\, z\in\mathcal{S}_\kappa$, and small enough $\mathrm{Im}\, z$,we have

$$\frac{\log N}{C(\log\log N)^2} \leqslant \|(I-m^2S)^{-1}\|_{\infty\to\infty} \leqslant C_\kappa \log N \tag{2.11}$$

for some universal constant C and some constant C_κ depending on κ. In our context, where ζ_x is small but much larger than the reciprocal of the lower bound of (2.11), such a logarithmic factor is not affordable.

To address this difficulty, we avoid passing by the form (2.10) altogether, as it is doomed by (2.11). The underlying cause for the instability of (2.10) is the inhomogeneous local structure of the matrix S, which is a multiple of the adjacency matrix of a sparse graph. Thus, the solution is to derive a self-consistent equation of the form (2.10) but with an *unstructured* S, which has constant entries. The basic intuition is to replace the *local average* $\frac{1}{d}\sum_{y\in S_1(x)} G^{(x)}_{yy}$ in the first identity of (2.8) with the *global average* $\frac{1}{N}\sum_{y\neq x} G^{(x)}_{yy}$. Of course, in general these two are not close, but we can include their closeness into the definition of a typical vertex. Thus, we define the set of typical vertices as

$$\mathcal{T} := \left\{ x\in[N] : \alpha_x = 1+o(1)\,,\ \frac{1}{d}\sum_{y\in S_1(x)} G^{(x)}_{yy} = \frac{1}{N}\sum_{y\neq x} G^{(x)}_{yy} + o(1) \right\}. \tag{2.12}$$

The main work of the proof is then to prove the following facts with very high probability.

(a) Most vertices are typical.
(b) Most neighbours of any vertex are typical.

With (a) and (b) at hand, we explain how to conclude the proof. Using (a) and the approximate version of (2.8) established above, we deduce the self-consistent equation for typical vertices,

$$\frac{1}{G_{xx}} = -z - \frac{1}{|\mathcal{T}|}\sum_{y\in\mathcal{T}} G_{yy} + o(1)\,, \qquad x\in\mathcal{T}\,,$$

[3] We write $\|\cdot\|_{p\to p}$ for the operator norm on ℓ^p.

which, unlike (2.10), is stable (see Lemma 4.19 below) and can be easily solved to show that $G_{xx} = m + o(1) = m_{\alpha_x} + o(1)$ for all $x \in \mathcal{T}$. Moreover, if $x \notin \mathcal{T}$ then we obtain from (2.8) and (b) that

$$\frac{1}{G_{xx}} = -z - \frac{1}{d} \sum_{y \in S_1(x) \cap \mathcal{T}} G^{(x)}_{yy} + o(1) = -z - \alpha_x m + o(1) \,,$$

where we used that $G_{yy} = m + o(1)$ for $y \in \mathcal{T}$. This shows that $G_{xx} = m_{\alpha_x} + o(1)$ for all $x \in [N]$ with very high probability, and hence concludes the proof.

What remains, therefore, is the proof of (a) and (b); see Proposition 4.8 below for a precise statement. Using the bootstrapping assumption of boundedness of the entries of G, it is not hard to estimate the probability $\mathbb{P}(x \in \mathcal{T})$, which we prove to be $1 - o(1)$, although $\{x \in \mathcal{T}\}$ does not hold with very high probability (this characterizes the critical and subcritical regimes). Now if the events $\{x \in \mathcal{T}\}$, $x \in [N]$, were all independent, it would then be a simple matter to deduce (a) and (b).

The most troublesome source of dependence among the events $\{x \in \mathcal{T}\}$, $x \in [N]$, is the Green function $G^{(x)}_{yy}$ in the definition of $\mathcal{T}$. Thus, the main difficulty of the proof is a decoupling argument that allows us to obtain good decay for the probability $\mathbb{P}(T \subset \mathcal{T})$ in the size of T. This decay can only work up to a threshold in the size of T, beyond which the correlations among the different events kick in. In fact, we essentially prove that

$$\mathbb{P}(T \subset \mathcal{T}) \leqslant \mathrm{e}^{-o(1)d|T|} + \mathcal{C}N^{-\nu} \qquad \text{for} \quad |T| = o(d) \,; \tag{2.13}$$

see Lemma 4.12. Choosing the largest possible T, $T = o(d)$, we find that the first term on the right-hand side of (2.13) is bounded by $N^{-\nu}$ provided that $o(1)d^2 \geqslant \nu \log N$, which corresponds precisely to the optimal lower bound in (1.18). Using (2.13), we may deduce (a) and (b).

To prove (2.13), we need to decouple the events $\{x \in \mathcal{T}\}$, $x \in T$. We do so by replacing the Green functions $G^{(x)}$ in the definition of $\mathcal{T}$ by $G^{(T)}$, after which the corresponding events are essentially independent. The error that we incur depends on the difference $G^{(T)}_{yy} - G_{yy}$, which we have to show is small with very high probability under the bootstrapping assumption that the entries of G are bounded. For T of fixed size, this follows easily from standard resolvent identities. However, for our purposes it is crucial that T can have size up to $o(d)$, which requires a more careful quantitative analysis. As it turns out, $G^{(T)}_{yy} - G_{yy}$ is small only up to $|T| = o(d)$, which is precisely what we need to reach the optimal scale $d \gg \sqrt{\log N}$ from (1.4).

3. The Semilocalized Phase

In this section we prove the results of Sect. 1.2–Theorems 1.2 and 1.7.

3.1. The pruned graph and proof of Theorem 1.2. The balls $(B_r(x))_{x \in \mathcal{W}_{\lambda,\delta}}$ in Theorem 1.2 are in general not disjoint. For its proof, and in order to give a precise definition of the vector $\mathbf{v}(x)$ in Theorem 1.2, we need to make these balls disjoint by *pruning* the graph $\mathbb{G}$. This is an important ingredient of the proof, and will also allow us to state a more precise version of Theorem 1.2, which is Theorem 3.4 below. This pruning was previously introduced in [10]; it is performed by cutting edges from $\mathbb{G}$ in such a way

that the balls $(B_r(x))_{x\in\mathcal{W}_{\lambda,\delta}}$ are disjoint for appropriate radii, $r = 2r_\star$, by carefully cutting in the right places, thus reducing the number of cut edges. This ensures that the pruned graph is close to the original graph in an appropriate sense. The pruned graph, $\mathbb{G}_\tau$, depends on a parameter $\tau > 1$, and its construction is the subject of the following proposition.

To state it, we introduce the following notations. For a subgraph $\mathbb{G}_\tau$ of $\mathbb{G}$ we abbreviate

$$B_i^\tau(x) := B_i^{\mathbb{G}_\tau}(x)\,, \qquad S_i^\tau(x) := S_i^{\mathbb{G}_\tau}(x)\,.$$

Moreover, we define the set of vertices with large degrees

$$\mathcal{V}_\tau := \{x \in [N] : \alpha_x \geqslant \tau\}\,.$$

Proposition 3.1 (Existence of pruned graph). *Let $1 + \xi^{1/2} \leqslant \tau \leqslant 2$ and $d \leqslant 3\log N$. There exists a subgraph $\mathbb{G}_\tau$ of $\mathbb{G}$ with the following properties.*

(i) *Any path in $\mathbb{G}_\tau$ connecting two different vertices in $\mathcal{V}_\tau$ has length at least $4r_\star + 1$. In particular, the balls $(B^\tau_{2r_\star}(x))_{x\in\mathcal{V}_\tau}$ are disjoint.*
(ii) *The induced subgraph $\mathbb{G}_\tau|_{B^\tau_{2r_\star}(x)}$ 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 $\mathcal{V}_\tau$ incident to it.*
(iv) *For each $x \in \mathcal{V}_\tau$ and each $i \in \mathbb{N}$ satisfying $1 \leqslant i \leqslant 2r_\star$ we have $S_i^\tau(x) \subset S_i(x)$.*
(v) *The degrees induced on $[N]$ by $\mathbb{G} \setminus \mathbb{G}_\tau$ are bounded according to*

$$\max_{x\in[N]} D_x^{\mathbb{G}\setminus\mathbb{G}_\tau} \leqslant \mathcal{C}\frac{\log N}{(\tau-1)^2 d} \tag{3.1}$$

with very high probability.
(vi) *Suppose that $\sqrt{\log N} \leqslant d$. For each $x \in \mathcal{V}_\tau$ and all $2 \leqslant i \leqslant 2r_\star$, the bound*

$$|S_i(x) \setminus S_i^\tau(x)| \leqslant \mathcal{C}\frac{\log N}{(\tau-1)^2} d^{i-2} \tag{3.2}$$

holds with very high probability.

The proof of Proposition 3.1 is postponed to the end of this section, in Sect. 3.5 below. It is essentially [10, Lemma 7.2], the main difference being that (vi) is considerably sharper than its counterpart, [10, Lemma 7.2 (vii)]; this stronger bound is essential to cover the full optimal regime (1.4) (see Sect. 1.5). As a guide for the reader's intuition, we recall the main idea of the pruning. First, for every $x \in \mathcal{V}_\tau$, we make the $2r_\star$-neighbourhood of x a tree by removing appropriate edges incident to x. Second, we take all paths of length less than $4r_\star + 1$ connecting different vertices in $\mathcal{V}_\tau$, and remove all of their edges incident to any vertex in $\mathcal{V}_\tau$. Note that only edges incident to vertices in $\mathcal{V}_\tau$ are removed. This informal description already explains properties (i)–(iv). Properties (v) and (vi) are probabilistic in nature, and express that with very high probability the pruning has a small impact on the graph. See also Lemma 3.8 below for a statement in terms of operator norms of the adjacency matrices. For the detailed algorithm, we refer to the proof of [10, Lemma 7.2].

Using the pruned graph $\mathbb{G}_\tau$, we can give a more precise formulation of Theorem 1.2, where the localization profile vector $\mathbf{v}(x)$ from Theorem 1.2 is explicit. For its statement, we introduce the set of vertices

$$\mathcal{V} := \mathcal{V}_{2+\xi^{1/4}} \tag{3.3}$$

around which a localization profile can be defined.

Definition 3.2 (*Localization profile*). Let $1+\xi^{1/2} \leqslant \tau \leqslant 2$ and $\mathbb{G}_\tau$ be the pruned graph from Proposition 3.1. For $x \in \mathcal{V}$ we introduce positive weights $u_0(x), u_1(x), \ldots, u_{r_\star}(x)$ as follows. Set $u_0(x) > 0$ and define, for $i = 1, \ldots, r_\star - 1$,

$$u_i(x) := \frac{\sqrt{\alpha_x}}{(\alpha_x - 1)^{i/2}}\, u_0(x)\,, \qquad u_{r_\star}(x) := \frac{1}{(\alpha_x - 1)^{(r_\star - 1)/2}}\, u_0(x)\,. \tag{3.4}$$

For $\sigma = \pm$ we define the radial vector

$$\mathbf{v}^\tau_\sigma(x) := \sum_{i=0}^{r_\star} \sigma^i u_i(x) \frac{\mathbf{1}_{S_i^\tau(x)}}{\|\mathbf{1}_{S_i^\tau(x)}\|}\,, \tag{3.5}$$

and choose $u_0(x) > 0$ such that $\mathbf{v}^\tau_\sigma(x)$ is normalized.

Remark 3.3. The family $(\mathbf{v}^\tau_\sigma(x) : x \in \mathcal{V},\ \sigma = \pm)$ is orthonormal. Indeed, if $x, y \in \mathcal{V}$ are distinct, then by Proposition 3.1 (i) the vectors $\mathbf{v}^\tau_\sigma(x)$ and $\mathbf{v}^\tau_{\tilde\sigma}(y)$ are orthogonal for any $\sigma, \tilde\sigma = \pm$ because they are supported on disjoint sets of vertices. Moreover, $\mathbf{v}^\tau_+(x)$ and $\mathbf{v}^\tau_-(x)$ are orthogonal by the choice of $u_{r_\star}(x)$ from (3.4), as can be seen by a simple computation.

The following result restates Theorem 1.2 by identifying $\mathbf{v}(x)$ there as $\mathbf{v}^\tau_+(x)$ given in (3.5). It easily implies Theorem 1.2, and the rest of this section is devoted to its proof.

Theorem 3.4. *The following holds with very high probability. Suppose that d satisfies (1.10). Let $\mathbf{w}$ be a normalized eigenvector of $A/\sqrt{d}$ with nontrivial eigenvalue $\lambda \geqslant 2 + C\xi^{1/2}$. Choose $0 < \delta \leqslant (\lambda - 2)/2$ and set $\tau := 1 + (\lambda - 2)/8 \wedge 1$. Then*

$$\sum_{x \in \mathcal{W}_{\lambda,\delta}} \langle \mathbf{v}^\tau_+(x), \mathbf{w}\rangle^2 \geqslant 1 - C\left(\frac{\xi + \xi_{\tau-1}}{\delta}\right)^2. \tag{3.6}$$

Remark 3.5. An analogous result holds for negative eigenvalues $-\lambda$, where λ is as in Theorem 3.4 and $\mathbf{v}^\tau_+(x)$ in (3.6) is replaced with $\mathbf{v}^\tau_-(x)$.

For the motivation behind Definition 3.2, we refer to the discussion in Sect. 2.2 and Appendix A.2. As explained there, if $\mathbb{G}_\tau$ is sufficiently close to the infinite tree $\mathbb{T}_{D_x,d}$ in a ball of radius $r_\star$ around x, and if $r_\star$ is large enough for $u_{r_\star}(x)$ to be very small, we expect (3.5) to be an approximate eigenvector of A. This will in fact turn out to be true; see Proposition 3.9 below. That $r_\star$ is in fact large enough is easy to see: the definition of $r_\star$ in (1.8) and the bound $\xi \geqslant 1/d$ imply that, for $\alpha_x \geqslant 2 + C(\log d)^2/\sqrt{\log N}$, we have

$$(\alpha_x - 1)^{-(r_\star - 2)/2} \leqslant \xi\,. \tag{3.7}$$

This means that the last element of the sequence $(u_i(x))_{i=0}^{r_\star}$ is bounded by ξ. Note that the lower bound on α_x imposed above always holds for $x \in \mathcal{V}$, since, by (1.10),

$$\frac{C(\log d)^2}{\sqrt{\log N}} \leqslant \xi^{1/4}\,. \tag{3.8}$$

As a guide to the reader, in Fig. 6, we summarize the three main sets of vertices that are used in the proof of Theorem 3.4. We conclude this subsection by proving Theorem 1.2 and Corollary 1.6 using Theorem 3.4.

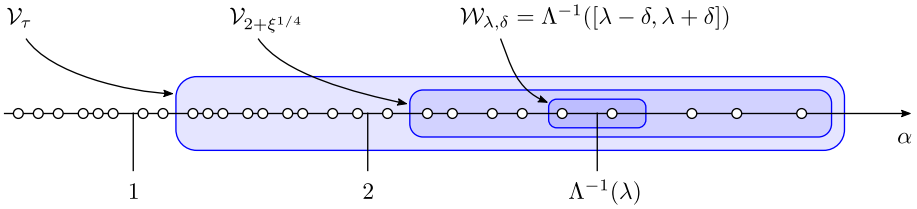

Fig. 6. An illustration of the three sets of vertices of increasing size that enter into the proof of Theorem 3.4. Each vertex x is plotted as a dot at its normalized degree α_x. The largest set is $\mathcal{V}_\tau$ from Proposition 3.1, where $1+\xi^{1/2} \leqslant \tau \leqslant 2$. It is used to define the pruned graph $\mathbb{G}_\tau$. The intermediate set is $\mathcal{V} \equiv \mathcal{V}_{2+\xi^{1/4}}$ from (3.3). It is the set of vertices for which we can define the localization profile vector $\mathbf{v}(x)$ that decays exponentially around x. The smallest set $\mathcal{W}_{\lambda,\delta} = \Lambda^{-1}([\lambda-\delta, \lambda+\delta])$ is the set of resonant vertices at energy λ

Proof of Theorem 1.2. The first claim follows immediately from Theorem 3.4, with $\mathbf{v}(x) = \mathbf{v}_+^\tau(x)$. To verify the claim about the exponential decay of $\mathbf{v}$, we note that the graph distance in $\mathbb{G}$ is bounded by the graph distance in $\mathbb{G}_\tau$, which implies

$$\sum_{y \in B_r(x)^c} (\mathbf{v}_+^\tau(x))_y^2 \leqslant \sum_{y \in B_r^\tau(x)^c} (\mathbf{v}_+^\tau(x))_y^2 = \sum_{i=r+1}^{r_\star} u_i(x)^2 \,,$$

from which the claim easily follows using the definition (3.4). □

Proof of Corollary 1.6. We decompose $\mathbf{w} = \sum_{x \in \mathcal{W}_{\lambda,\delta}} \gamma_x \mathbf{v}_+^\tau(x) + \mathbf{e}$, where $\gamma_x := \langle \mathbf{v}_+^\tau(x), \mathbf{w} \rangle$ and $\mathbf{e}$ is orthogonal to $\mathrm{Span}\{\mathbf{v}_+^\tau(x) : x \in \mathcal{W}_{\lambda,\delta}\}$. By Theorem 3.4 we have $\|\mathbf{e}\| \leqslant \frac{\mathcal{C}(\xi+\xi_{\tau-1})}{\delta}$ and

$$\sum_{x \in \mathcal{W}_{\lambda,\delta}} \gamma_x^2 \geqslant 1 - \frac{\mathcal{C}(\xi + \xi_{\tau-1})}{\delta} \,. \tag{3.9}$$

Moreover, since $\lambda - \delta \geqslant 2 \geqslant \tau$, we have $\mathcal{W}_{\lambda,\delta} \subset \mathcal{V}_\tau$, so that Proposition 3.1 (i) implies $(\mathbf{v}_+^\tau(x))_y = \delta_{xy} u_0(x)$ for $x, y \in \mathcal{W}_{\lambda,\delta}$. Thus we have

$$\begin{aligned} \sum_{y \in \mathcal{W}_{\lambda,\delta}} w_y^2 = \|\mathbf{w}|_{\mathcal{W}_{\lambda,\delta}}\|^2 &= \Big\| \sum_{x \in \mathcal{W}_{\lambda,\delta}} \gamma_x \mathbf{v}_+^\tau(x)|_{\mathcal{W}_{\lambda,\delta}} \Big\|^2 + O(\|\mathbf{e}\|) \\ &= \sum_{y \in \mathcal{W}_{\lambda,\delta}} \gamma_y^2 u_0(y)^2 + \mathcal{O}\left(\frac{\xi + \xi_{\tau-1}}{\delta}\right) . \end{aligned} \tag{3.10}$$

Since $u_0(y)$ was chosen such that $\mathbf{v}_+^\tau(y)$ is normalized, we find

$$u_0(y)^2 = \left(1 + \sum_{i=1}^{r_\star - 1} \frac{\alpha_y}{(\alpha_y - 1)^i} + \frac{1}{(\alpha_y - 1)^{r_\star - 1}}\right)^{-1} = \frac{\alpha_y - 2}{2(\alpha_y - 1)} + O\left(\frac{1}{(\alpha_y - 1)^{r_\star - 1}}\right).$$

Define $\alpha := \Lambda^{-1}(\lambda)$ for $\alpha \geqslant 2$. Since $|\Lambda(\alpha_y) - \lambda| \leqslant \delta$ for $y \in \mathcal{W}_{\lambda,\delta}$, we obtain

$$|\alpha_y - \alpha| \leqslant \delta \max_{t \in [\lambda-\delta, \lambda+\delta]} (\Lambda^{-1})'(t)$$

$$= O\left(\delta\lambda^{3/2}(\lambda-2)^{-1/2}\right),$$

where we used that $\lambda \pm \delta - 2 \asymp \lambda - 2$. Since $\frac{\mathrm{d}}{\mathrm{d}\alpha}\frac{\alpha-2}{2(\alpha-1)} = \frac{1}{2(\alpha-1)^2} \asymp \lambda^{-4}$, we find

$$\begin{aligned} u_0(y)^2 &= \frac{\alpha-2}{2(\alpha-1)} + O\left(\frac{\delta}{\lambda^{5/2}\sqrt{\lambda-2}} + \frac{1}{(\alpha_y-1)^{r_\star-1}}\right) \\ &= \frac{\alpha-2}{2(\alpha-1)} + O\left(\frac{\delta}{\lambda^{5/2}\sqrt{\lambda-2}} + \frac{\xi}{\delta}\right), \end{aligned} \tag{3.11}$$

where we used (3.7) and the upper bound on δ in the last step. By an elementary computation,

$$\frac{\alpha-2}{2(\alpha-1)} = \frac{\sqrt{\lambda^2-4}}{\lambda+\sqrt{\lambda^2-4}},$$

and the claim hence follows by recalling (3.7) and plugging (3.9) and (3.11) into (3.10). □

3.2. Block diagonal approximation of pruned graph and proof of Theorems 3.4 and 1.7. We now introduce the adjacency matrix of $\mathbb{G}_\tau$ and a suitably defined centred version. Then we define a block diagonal approximation of this matrix, called $\widehat{H}^\tau$ in (3.16) below, which is the central construction of our proof.

Definition 3.6. Let A^τ be the adjacency matrix of $\mathbb{G}_\tau$. Let $H := \underline{A}/\sqrt{d}$ and $H^\tau := \underline{A}^\tau/\sqrt{d}$, where

$$\underline{A} := A - \mathbb{E}A, \qquad \underline{A}^\tau := A^\tau - \chi^\tau(\mathbb{E}A)\chi^\tau \tag{3.12}$$

and χ^τ is the orthogonal projection onto $\mathrm{Span}\{\mathbf{1}_y : y \notin \bigcup_{x\in\mathcal{V}_\tau} B^\tau_{2r_\star}(x)\}$.

The definition of $\underline{A}^\tau$ is chosen so that (i) $\underline{A}^\tau$ is close to $\underline{A}$ provided that A^τ is close to A, since the kernel of χ^τ has a relatively low dimension, and (ii) when restricted to vertices at distance at most $2r_\star$ from $\mathcal{V}_\tau$, the matrix $\underline{A}^\tau$ coincides with A^τ. In fact, property (i) is made precise by the simple estimate

$$\|\mathbb{E}A - \chi^\tau(\mathbb{E}A)\chi^\tau\| \leqslant 2 \tag{3.13}$$

with very high probability (see [10, Eq. (8.17)] for details). Property (ii) means that $\underline{A}^\tau$ inherits the locality of the matrix A, meaning that applying $\underline{A}^\tau$ to a vector localized in space to a small enough neighbourhood of $\mathcal{V}_\tau$ yields again a vector localized in space. This property will play a crucial role in the proof, and it can be formalized as follows.

Remark 3.7. Let $i + j \leqslant 2r_\star$. Then for any $x \in \mathcal{V}_\tau$ and vector $\mathbf{v}$ we have

$$\operatorname{supp}\mathbf{v} \subset B^\tau_i(x) \quad \Longrightarrow \quad \operatorname{supp}\left[(H^\tau)^j\mathbf{v}\right] \subset B^\tau_{i+j}(x).$$

The next result states that H^τ is a small perturbation of H.

Lemma 3.8. *Suppose that $d \leqslant 3\log N$. For any $1+\xi^{1/2} \leqslant \tau \leqslant 2$ we have $\|H - H^\tau\| \leqslant C\xi_{\tau-1}$ with very high probability.*

The next result states that $\mathbf{v}^\tau_\sigma(x)$ is an approximate eigenvector of H^τ.

Proposition 3.9. *Let d satisfy (1.10). Let $x \in [N]$ and suppose that $1+\xi^{1/2} \leqslant \tau \leqslant 2$. If $\alpha_x \geqslant 2 + C(\log d)^2/\sqrt{\log N}$ then for $\sigma = \pm$ we have*

$$\|(H^\tau - \sigma\Lambda(\alpha_x))\mathbf{v}_\sigma^\tau(x)\| \leqslant \mathcal{C}\xi \tag{3.14}$$

with very high probability.

The proofs of Lemma 3.8 and Proposition 3.9 are deferred to Sect. 3.3. The following object is the central construction in our proof.

Definition 3.10. (*Block diagonal approximation of pruned graph*) Define the orthogonal projections

$$\Pi^\tau := \sum_{x\in\mathcal{V}}\sum_{\sigma=\pm} \mathbf{v}_\sigma^\tau(x)\mathbf{v}_\sigma^\tau(x)^*, \qquad \overline{\Pi}^\tau := I - \Pi^\tau, \tag{3.15}$$

and the matrix

$$\widehat{H}^\tau := \sum_{x\in\mathcal{V}}\sum_{\sigma=\pm} \sigma\Lambda(\alpha_x)\mathbf{v}_\sigma^\tau(x)\mathbf{v}_\sigma^\tau(x)^* + \overline{\Pi}^\tau H^\tau \overline{\Pi}^\tau. \tag{3.16}$$

That Π^τ and $\overline{\Pi}^\tau$ are indeed orthogonal projections follows from Remark 3.3. Note that $\widehat{H}^\tau$ may be interpreted as a *block diagonal approximation of* H^τ. Indeed, completing the orthonormal family $(\mathbf{v}_\sigma^\tau(x))_{x\in\mathcal{V},\sigma=\pm}$ to an orthonormal basis of $\mathbb{R}^N$, which we write as the columns of the orthogonal matrix R, we have

$$R^*\widehat{H}^\tau R = \begin{bmatrix} \mathrm{diag}(\sigma\Lambda(\alpha_x))_{x\in\mathcal{V},\sigma=\pm} & 0 \\ 0 & [*] \end{bmatrix}.$$

The following estimate states that $\widehat{H}^\tau$ is a small perturbation of H^τ.

Lemma 3.11. *Let d satisfy (1.10). If $1+\xi^{1/2} \leqslant \tau \leqslant 2$ then $\|H^\tau - \widehat{H}^\tau\| \leqslant \mathcal{C}\xi$ with very high probability.*

The proof of Lemma 3.11 is deferred to Sect. 3.3. The following result is the key estimate of our proof; it states that on the range of $\overline{\Pi}^\tau$ the matrix H^τ is bounded by $2\tau + o(1)$.

Proposition 3.12. *Let d satisfy (1.10). If $1+\xi^{1/2} \leqslant \tau \leqslant 2$ then $\|\overline{\Pi}^\tau H^\tau \overline{\Pi}^\tau\| \leqslant 2\tau + \mathcal{C}(\xi + \xi_{\tau-1})$ with very high probability.*

The proof of Proposition 3.12 is deferred to Sect. 3.4. We now use Lemma 3.11 and Proposition 3.12 to conclude Theorems 3.4 and 1.7.

Proof of Theorem 3.4. Define the orthogonal projections

$$\Pi_{\lambda,\delta}^\tau := \sum_{x\in\mathcal{W}_{\lambda,\delta}} \mathbf{v}_+^\tau(x)\,\mathbf{v}_+^\tau(x)^*, \qquad \overline{\Pi}_{\lambda,\delta}^\tau := I - \Pi_{\lambda,\delta}^\tau.$$

By definition, the orthogonal projections Π^τ and $\Pi_{\lambda,\delta}^\tau$ commute. Moreover, under the assumptions of Theorem 3.4 we have the inclusion property

$$\Pi^\tau \Pi_{\lambda,\delta}^\tau = \Pi_{\lambda,\delta}^\tau. \tag{3.17}$$

See also Fig. 6. To show (3.17), we note that the condition on δ and the lower bound on λ in Theorem 3.4 imply $\lambda - \delta \geqslant 2 + \mathcal{C}\xi^{1/2}$. Using $\Lambda(2+x) - 2 \asymp x^2 \wedge x^{1/2}$ for $x \geqslant 0$ we

conclude that for any $\alpha \geqslant 2$ we have the implication $\Lambda(\alpha) \geqslant \lambda - \delta \;\Rightarrow\; \alpha \geqslant 2 + \xi^{1/4}$, which implies (3.17).

Next, we abbreviate $E^\tau := \chi^\tau(\mathbb{E}A/\sqrt{d})\chi^\tau$ and note that $\Pi^\tau E^\tau = 0$ because $\Pi^\tau \chi^\tau = 0$ by construction of $\mathbf{v}^\tau_\sigma(x)$. From (3.17) we obtain $\overline{\Pi}^\tau_{\lambda,\delta} = \overline{\Pi}^\tau_{\lambda,\delta}\Pi^\tau + \overline{\Pi}^\tau$, which yields

$$\overline{\Pi}^\tau_{\lambda,\delta}(\widehat{H}^\tau + E^\tau)\overline{\Pi}^\tau_{\lambda,\delta} = \overline{\Pi}^\tau_{\lambda,\delta}\Pi^\tau \widehat{H}^\tau \Pi^\tau \overline{\Pi}^\tau_{\lambda,\delta} + \left(\overline{\Pi}^\tau \widehat{H}^\tau \overline{\Pi}^\tau + E^\tau\right), \tag{3.18}$$

where we used that the cross terms vanish because of the block diagonal structure of $\widehat{H}^\tau$.

The core of our proof is the *spectral gap*

$$\operatorname{spec}\left(\overline{\Pi}^\tau_{\lambda,\delta}(\widehat{H}^\tau + E^\tau)\overline{\Pi}^\tau_{\lambda,\delta}\right) \subset \mathbb{R} \setminus [\lambda - \delta, \lambda + \delta]\,. \tag{3.19}$$

To establish (3.19), it suffices to establish the same spectral gap for each term on the right-hand side of (3.18) separately, since the right-hand side of (3.18) is a block decomposition of its left-hand side. The first term on the right-hand side of (3.18) is explicit:

$$\overline{\Pi}^\tau_{\lambda,\delta}\Pi^\tau \widehat{H}^\tau \Pi^\tau \overline{\Pi}^\tau_{\lambda,\delta} = \sum_{x\in\mathcal{V}}\sum_{\sigma=\pm} \sigma\,\Lambda(\alpha_x)\,\mathbb{1}_{|\sigma\Lambda(\alpha_x)-\lambda|>\delta}\,\mathbf{v}^\tau_\sigma(x)\mathbf{v}^\tau_\sigma(x)^*\,,$$

which trivially has no eigenvalues in $[\lambda - \delta, \lambda + \delta]$.

In order to establish the spectral gap for the second term of (3.18), we begin by remarking that E^τ has rank one and, by (3.13), its unique nonzero eigenvalue is $\sqrt{d} + O(1/\sqrt{d})$. Hence, by rank-one interlacing and Proposition 3.12, we find

$$\operatorname{spec}\left(\overline{\Pi}^\tau(H^\tau + E^\tau)\overline{\Pi}^\tau\right) \subset \left[-2\tau - \mathcal{C}(\xi + \xi_{\tau-1})\,,\, 2\tau + \mathcal{C}(\xi + \xi_{\tau-1})\right] \cup \{\mu\} \tag{3.20}$$

for some simple eigenvalue $\mu = \sqrt{d} + O(1)$. Thus, to conclude the proof of the spectral gap for the second term of (3.18), it suffices to show that

$$\lambda - \delta > 2\tau + \mathcal{C}(\xi + \xi_{\tau-1}) \tag{3.21}$$

$$\lambda + \delta < \mu\,. \tag{3.22}$$

To prove (3.21), we suppose that $\lambda \geqslant 2 + 8\mathcal{C}\xi^{1/2}$ and, recalling the condition on δ and the choice of τ in Theorem 3.4, obtain

$$\lambda - \delta \geqslant 2 + \frac{\lambda - 2}{2} \geqslant 2\tau + 2\mathcal{C}\xi^{1/2} > 2\tau + \mathcal{C}(\xi + \xi_{\tau-1})\,, \tag{3.23}$$

where in the last step we used that $\xi_{\tau-1} < \xi^{1/2}$ by our choice of τ and the lower bound on λ. This is (3.21).

For the following arguments, we compare $A/\sqrt{d}$ with $\widehat{H}^\tau + E^\tau$ using the estimate

$$\begin{aligned}\|A/\sqrt{d} - (\widehat{H}^\tau + E^\tau)\| &\leqslant \|(H^\tau - \widehat{H}^\tau) + (H - H^\tau)\\ &+(\mathbb{E}A/\sqrt{d} - E^\tau)\| \leqslant \mathcal{C}(\xi + \xi_{\tau-1})\end{aligned} \tag{3.24}$$

with very high probability, which follows from Lemma 3.8, Lemma 3.11, (3.13) and $d^{-1/2} \leqslant \mathcal{C}\xi$.

Next, we use (3.24) to conclude the proof of (3.22). The only nonzero eigenvalue of E^τ is $\sqrt{d}(1 + O(1/d))$, and from Proposition 3.12 and Remark 1.5 we have $\|\widehat{H}^\tau\| \leqslant$

$\Lambda(\max_{x\in\mathcal{V}}\alpha_x)+O(1)$ with very high probability, so that Lemma A.7 and the assumption (1.10) yield $\|\widehat{H}^\tau\|\leqslant\mathcal{C}\sqrt{\frac{\log N}{d}}$ with very high probability. Hence, by first order perturbation theory (e.g. Weyl's inequality), (1.10) and (3.24) imply that $A/\sqrt{d}$ has one eigenvalue bigger than $\sqrt{d}-O(1)$ and all other eigenvalues are at most $\mathcal{C}\sqrt{\frac{\log N}{d}}$. Since λ is nontrivial, we conclude that $\lambda\leqslant\mathcal{C}\sqrt{\frac{\log N}{d}}$. By the upper bound $\delta\leqslant(\lambda-2)/2$ and the lower bound on d in (1.10), this concludes the proof of (3.22) and, thus, the one of the spectral gap (3.19).

Next, from (3.19), and (3.24), we conclude the *spectral gap for the full adjacency matrix*

$$\operatorname{spec}\Big(\overline{\Pi}^\tau_{\lambda,\delta}(A/\sqrt{d})\overline{\Pi}^\tau_{\lambda,\delta}\Big)\subset\mathbb{R}\setminus\big[\lambda-\delta+\mathcal{C}(\xi+\xi_{\tau-1}),\lambda+\delta-\mathcal{C}(\xi+\xi_{\tau-1})\big]. \tag{3.25}$$

Using (3.25) we may conclude the proof. The eigenvalue-eigenvector equation $(A/\sqrt{d}-\lambda)\mathbf{w}=0$ yields

$$\overline{\Pi}^\tau_{\lambda,\delta}\mathbf{w}=-\Big(\overline{\Pi}^\tau_{\lambda,\delta}(A/\sqrt{d})\overline{\Pi}^\tau_{\lambda,\delta}-\lambda\Big)^{-1}\overline{\Pi}^\tau_{\lambda,\delta}(A/\sqrt{d})\Pi^\tau_{\lambda,\delta}\mathbf{w}\,. \tag{3.26}$$

Assuming that $\delta>\mathcal{C}(\xi+\xi_{\tau-1})$, from (3.25) we get

$$\Big\|\Big(\overline{\Pi}^\tau_{\lambda,\delta}(A/\sqrt{d})\overline{\Pi}^\tau_{\lambda,\delta}-\lambda\Big)^{-1}\Big\|\leqslant\frac{1}{\delta-\mathcal{C}(\xi+\xi_{\tau-1})}\,. \tag{3.27}$$

Moreover, since $\overline{\Pi}^\tau_{\lambda,\delta}\widehat{H}^\tau\Pi^\tau_{\lambda,\delta}=0$ and $E^\tau\Pi^\tau_{\lambda,\delta}=0$, we deduce from (3.24) that

$$\|\overline{\Pi}^\tau_{\lambda,\delta}(A/\sqrt{d})\Pi^\tau_{\lambda,\delta}\|\leqslant\mathcal{C}(\xi+\xi_{\tau-1})\,. \tag{3.28}$$

Plugging (3.27) and (3.28) into (3.26) yields

$$\|\overline{\Pi}^\tau_{\lambda,\delta}\mathbf{w}\|\leqslant\frac{\mathcal{C}(\xi+\xi_{\tau-1})}{\delta-\mathcal{C}(\xi+\xi_{\tau-1})}\wedge 1\leqslant\frac{2\mathcal{C}(\xi+\xi_{\tau-1})}{\delta}\,,$$

since $\mathbf{w}$ is normalized. This concludes the proof if $\delta>\mathcal{C}(\xi+\xi_{\tau-1})$ (after a renaming of the constant $\mathcal{C}$), and otherwise the claim is trivial. □

Proposition 3.12 is also the main tool to prove Theorem 1.7.

Proof of Theorem 1.7. The proof uses Proposition 3.12, Lemma 3.8, and Lemma 3.11 for $\tau\in[1+\xi^{1/2}/3,2]$. Note that the lower bound $1+\xi^{1/2}/3$ is smaller than the lower bound $1+\xi^{1/2}$ imposed in these results, but their proofs hold verbatim also in this regime of τ.

We set $E^\tau:=\chi^\tau(\mathbb{E}A/\sqrt{d})\chi^\tau$ with χ^τ from Definition 3.6. We now compare $A/\sqrt{d}$ and $\widehat{H}^\tau+E^\tau$, as in the proof of Theorem 3.4, and use some estimates from its proof. For any $\tau\in[1+\xi^{1/2}/3,2]$, we have

$$\operatorname{spec}(\widehat{H}^\tau+E^\tau)=\{\pm\Lambda(\alpha_x):x\in\mathcal{U}\}\cup\operatorname{spec}\left(\overline{\Pi}^\tau(H^\tau+E^\tau)\overline{\Pi}^\tau\right), \tag{3.29}$$

since $\Pi^\tau\chi^\tau=0$. By first order perturbation theory and the choice $\tau=2$, we get from (3.29), (3.20) and (3.24) that $\lambda_1(A/\sqrt{d})=\mu+O(\xi)=\sqrt{d}+O(1)$ and $\lambda_1(A/\sqrt{d})$ is well separated from the other eigenvalues of $A/\sqrt{d}$ (see the proof of Theorem 3.4).

Combining (3.29), (3.20), and (3.24), choosing $\tau = 1 + \xi^{1/2}/3$ as well as using $\mathcal{C}(\xi + \xi_{\tau-1}) \leqslant \xi^{1/2}/3$ for this choice of τ imply (1.12).

Moreover, we apply first order perturbation theory to (3.29) using (3.20) and (3.24), and obtain

$$|\lambda_{i+1}(A/\sqrt{d}) - \Lambda(\alpha_{\sigma(i)})| + |\lambda_{N-i+1}(A/\sqrt{d}) + \Lambda(\alpha_{\sigma(i)})| \leqslant \mathcal{C}(\xi + \xi_{\tau-1}) \quad (3.30)$$

with very high probability for all $\tau \in [1 + \xi^{1/2}/3, 2]$ and all $i \in [|\mathcal{U}|]$ satisfying

$$2(\tau - 1) + \mathcal{C}(\xi + \xi_{\tau-1}) < \Lambda(\alpha_{\sigma(i)}) - 2. \quad (3.31)$$

What remains is choosing $\tau \equiv \tau_i$, depending on $i \in [|\mathcal{U}|]$, such that the condition (3.31) is satisfied and the error estimate from (3.30) transforms into the form of (1.11). Both are achieved by setting

$$\tau = 1 + \frac{1}{3}\big[(\Lambda(\alpha_{\sigma(i)}) - 2) \wedge 3\big]. \quad (3.32)$$

Note that $\tau \in [1 + \xi^{1/2}/3, 2]$ as $\sigma(i) \in \mathcal{U}$. From $\Lambda(\alpha_{\sigma(i)}) - 2 \geqslant 3(\tau - 1)$ due to (3.32) and $\Lambda(\alpha_{\sigma(i)}) - 2 \geqslant \xi^{1/2}$ by the definition of $\mathcal{U}$, we conclude that

$$\Lambda(\alpha_{\sigma(i)}) - 2 \geqslant \frac{5}{2}(\tau - 1) + \frac{1}{6}\xi^{1/2} \geqslant 2(\tau - 1) + \mathcal{C}(\xi_{\tau-1} + \xi),$$

where we used $\tau - 1 \geqslant 3\xi_{\tau-1} \log d$ as $\tau - 1 \geqslant \xi^{1/2}/3$. This proves (3.31) and, thus, (3.30) for any $\sigma(i) \in \mathcal{U}$ with the choice of τ from (3.32).

In order to show that the right-hand side of (3.30) is controlled by the one in (1.11), we now distinguish the two cases, $\Lambda(\alpha_{\sigma(i)}) - 2 \leqslant 3$ and $\Lambda(\alpha_{\sigma(i)}) - 2 > 3$. In the latter case, $\tau = 2$ by (3.32) and (1.11) follows immediately from (3.30) as $\xi_1 \leqslant \xi$. If $\Lambda(\alpha_{\sigma(i)}) - 2 \leqslant 3$ then $\tau - 1 = (\Lambda(\alpha_{\sigma(i)}) - 2)/3$ and, thus, $\xi_{\tau-1} = 3\xi_{\Lambda(\alpha_{\sigma(i)})-2}$. Hence, (3.30) implies (1.11). This concludes the proof of Theorem 1.7. □

3.3. Proof of Lemma 3.8, Proposition 3.9, and Lemma 3.11.

Proof of Lemma 3.8. To begin with, we reduce the problem to the adjacency matrices by using the estimate (3.13). Hence, with very high probability,

$$\sqrt{d}\|H - H^\tau\| \leqslant \|\mathbb{E}A - \chi^\tau(\mathbb{E}A)\chi^\tau\| + \|A - A^\tau\| \leqslant 2 + \|A^{\mathbb{D}_\tau}\|,$$

where $A^{\mathbb{D}_\tau}$ is the adjacency matrix of the graph $\mathbb{D}_\tau := \mathbb{G} \setminus \mathbb{G}_\tau$. Hence, since $d^{-1/2} \leqslant C\xi_{\tau-1}$ by $d \leqslant 3 \log N$ and the definition (1.9), it suffices to show that $\|A^{\mathbb{D}_\tau}\| \leqslant \mathcal{C}\xi_{\tau-1}\sqrt{d}$.

We know from Proposition 3.1 (iii) and (v) that with very high probability $\mathbb{D}_\tau$ consists of (possibly overlapping) stars[4] around vertices $x \in \mathcal{V}_\tau$ of central degree $D_x^{\mathbb{D}_\tau} \leqslant \mathcal{C}d\xi_{\tau-1}^2$. Moreover, with very high probability,

(i) any ball $B_{2r_\star}(x)$ around $x \in \mathcal{V}_\tau$ has at most $\mathcal{C}$ cycles;
(ii) any ball $B_{2r_\star}(x)$ around $x \in \mathcal{V}_\tau$ contains at most $\mathcal{C}d\xi_{\tau-1}^2$ vertices in $\mathcal{V}_\tau$.

[4] A *star* around a vertex x is a set of edges incident to x.

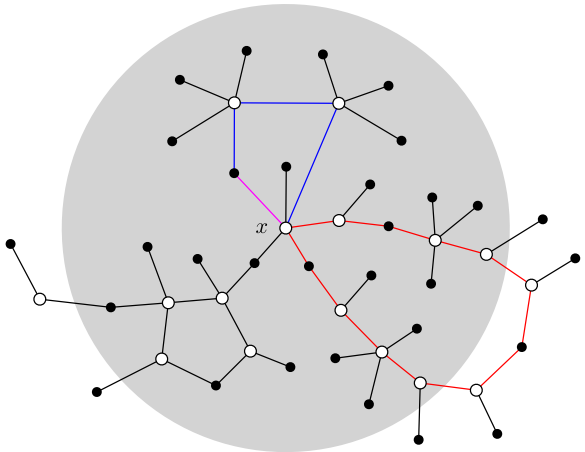

Fig. 7. An illustration of a connected component of $\mathbb{D}_\tau$. Vertices of $\mathcal{V}_\tau$ are drawn in white and the other vertices in black. The ball $B_{2r_\star}(x)$ around a chosen white vertex x is drawn in grey, where $2r_\star = 4$. The illustrated component of $\mathbb{D}_\tau$ has three cycles, two of which are in $B_{2r_\star}(x)$. The blue and red cycles pass through x. The purple edge is removed from the blue cycle, i.e. it is put into the graph $\mathbb{U}_\tau$. With very high probability, the red cycle cannot appear, because it leaves the ball $B_{2r_\star}(x)$ and therefore contains more white vertices in $B_{2r_\star}(x)$ than allowed by property (ii)

Claim (i) follows from [10, Corollary 5.6], the definition (1.8), and Lemma A.7. Claim (ii) follows from [10, Lemma 7.3] and $h((\tau-1)/2) \asymp (\tau-1)^2$ for $1 \leqslant \tau \leqslant 2$.

Let $x \in \mathcal{V}_\tau$. We claim that we can remove at most $\mathcal{C}$ edges of $\mathbb{D}_\tau$ incident to x so that no cycle passes through x. Indeed, if there were more than $\mathcal{C}$ cycles in $\mathbb{D}_\tau$ passing through x, then at least one such cycle would have to leave $B_{2r_\star}(x)$ (by (i)), which would imply that $B_{2r_\star}(x)$ has at least $r_\star$ vertices in $\mathcal{V}_\tau$, which, by (ii), is impossible since $r_\star \geqslant 2\mathcal{C}d\xi_{\tau-1}^2$ by $\tau \geqslant 1 + \xi^{1/2}$. See Fig. 7 for an illustration of $\mathbb{D}_\tau$.

Thus, we can remove a graph $\mathbb{U}_\tau$ from $\mathbb{D}_\tau$ such that $\mathbb{U}_\tau$ has maximal degree $\mathcal{C}$ and $\mathbb{D}_\tau \setminus \mathbb{U}_\tau$ is a forest of maximal degree $\mathcal{C}d\xi_{\tau-1}^2$ (by (ii)). The claim now follows from Lemma A.4. □

Proof of Proposition 3.9. We focus on the case $\sigma = +$; trivial modifications yield (3.14) for $\sigma = -$. The basic strategy is to decompose $(H^\tau - \Lambda(\alpha_x))\mathbf{v}_+^\tau(x)$ into several error terms that are estimated separately. A similar argument was applied in [10, Proposition 5.1] to the original graph $\mathbb{G}$ instead of $\mathbb{G}^\tau$, which however does not yield sharp enough estimates to reach the optimal scale $d \gg \sqrt{\log N}$ (see Sect. 1.5).

We omit x from the notation in this proof and write u_i, $\mathbf{v}_+^\tau$ and S_i^τ instead of $u_i(x)$, $\mathbf{v}_+^\tau(x)$ and $S_i^\tau(x)$. We define

$$\mathbf{s}_i^\tau := \frac{\mathbf{1}_{S_i^\tau}}{\|\mathbf{1}_{S_i^\tau}\|}, \qquad N_i^\tau(y) := |S_1^\tau(y) \cap S_i^\tau|.$$

Note that $(\mathbf{s}_i^\tau)_{i=0}^{2r_\star}$ form an orthonormal system. Defining the vectors

$$\mathbf{w}_2 := \sum_{i=2}^{r_\star} \frac{u_i}{\sqrt{d|S_i^\tau|}} \sum_{y \in S_{i-1}^\tau} \left(N_i^\tau(y) - \frac{|S_i^\tau|}{|S_{i-1}^\tau|} \right) \mathbf{1}_y,$$

$$\mathbf{w}_3 := u_2\left(\frac{\sqrt{|S_2^\tau|}}{\sqrt{d|S_1^\tau|}}-1\right)\mathbf{s}_1^\tau+\sum_{i=2}^{r_\star-1}\left[u_{i+1}\left(\frac{\sqrt{|S_{i+1}^\tau|}}{\sqrt{d|S_i^\tau|}}-1\right)+u_{i-1}\left(\frac{\sqrt{|S_i^\tau|}}{\sqrt{d|S_{i-1}^\tau|}}-1\right)\right]\mathbf{s}_i^\tau,$$

$$\mathbf{w}_4 := u_{r_\star}\left(1-\frac{1}{\sqrt{\alpha_x}}\right)\mathbf{s}_{r_\star-1}^\tau+u_{r_\star-1}\left(\frac{\sqrt{|S_{r_\star}^\tau|}}{\sqrt{d|S_{r_\star-1}^\tau|}}-\frac{1}{\sqrt{\alpha_x-1}}\right)\mathbf{s}_{r_\star}^\tau+u_{r_\star}\frac{\sqrt{|S_{r_\star+1}^\tau|}}{\sqrt{d|S_{r_\star}^\tau|}}\mathbf{s}_{r_\star+1}^\tau, \tag{3.33}$$

a straightforward computation using the definition of $\mathbf{v}_+^\tau$ yields

$$(H^\tau-\Lambda(\alpha_x))\mathbf{v}_+^\tau=\mathbf{w}_2+\mathbf{w}_3+\mathbf{w}_4. \tag{3.34}$$

For a detailed proof of (3.34) in a similar setup, we refer the reader to [10, Lemma 5.2] (note that in the analogous calculation of [10] the left-hand side of (3.34) is multiplied by $\sqrt{d}$). The terms in (3.34) analogous to $\mathbf{w}_0$ and $\mathbf{w}_1$ in [10] vanish, respectively, because the projection χ^τ is included in (3.12) and because $\mathbb{G}_\tau|_{B_{2r_\star}^\tau}$ is a tree by Proposition 3.1 (ii). The vector $\mathbf{w}_4$ from (3.33) differs from the one in [10] due to the special choice of $u_{r_\star}$ in (3.4).

We now complete the proof of (3.14) by showing that each term on the right-hand side of (3.34) is bounded in norm by $\mathcal{C}\xi$ with very high probability. We start with $\mathbf{w}_3$ by first proving the concentration bound

$$\left|\frac{|S_{i+1}^\tau|}{d|S_i^\tau|}-1\right|=\mathcal{O}\left(\frac{\sqrt{\log N}}{d}\right) \tag{3.35}$$

with very high probability, for $i=1,\dots,r_\star$. To prove this, we use Proposition 3.1 (iv) and (vi), as well as [10, Lemma 5.4], to obtain

$$\frac{|S_i^\tau|}{|S_i|}=1-\frac{|S_i\setminus S_i^\tau|}{|S_i|}\geqslant 1-\mathcal{C}\frac{\log N}{(\tau-1)^2d^2} \tag{3.36}$$

with very high probability, where we used that $\alpha_x\geqslant 1$, and the assumption [10, Eq. (5.13)] is satisfied by the definition (1.8). Therefore, invoking [10, Lemma 5.4] in the following expansion yields

$$\frac{|S_{i+1}^\tau|}{d|S_i^\tau|}=\frac{|S_{i+1}|}{d|S_i|}\frac{|S_i|}{|S_i^\tau|}\frac{|S_{i+1}^\tau|}{|S_{i+1}|}=\left(1+\mathcal{O}\left(\frac{\sqrt{\log N}}{d}\right)\right)\left(1+\mathcal{O}\left(\frac{\log N}{d^2(\tau-1)^2}\right)\right) \tag{3.37}$$

with very high probability. Hence, recalling the lower bound $\tau\geqslant 1+\xi^{1/2}$, we obtain (3.35).

We take the norm in the definition of $\mathbf{w}_3$, use the orthonormality of $(\mathbf{s}_i^\tau)_{i=0}^{r_\star}$, and end up with

$$\|\mathbf{w}_3\|^2\leqslant\left[\left(\frac{\sqrt{|S_2^\tau|}}{\sqrt{d|S_1^\tau|}}-1\right)^2u_2^2+2\sum_{i=2}^{r_\star-1}\left(\left(\frac{\sqrt{|S_{i+1}^\tau|}}{\sqrt{d|S_i^\tau|}}-1\right)^2u_{i+1}^2+\left(\frac{\sqrt{|S_i^\tau|}}{\sqrt{d|S_{i-1}^\tau|}}-1\right)^2u_{i-1}^2\right)\right].$$

Consequently, (3.35) and $\sum_{i=0}^{r_\star}u_i^2=1$ yield the desired bound on $\|\mathbf{w}_3\|$.

In order to estimate $\|\mathbf{w}_2\|$, we use the definitions

$$N_i(y):=|S_1(y)\cap S_i|,\qquad Y_i:=\frac{1}{|S_{i-1}^\tau|}\sum_{y\in S_{i-1}^\tau}\left(N_i(y)-\mathbb{E}[N_i(y)|B_{i-1}]\right)^2$$

and the Pythagorean theorem to obtain

$$\begin{aligned}\|\mathbf{w}_2\|^2 &= \sum_{i=2}^{r_\star} \frac{u_i^2}{d|S_i^\tau|} \sum_{y\in S_{i-1}^\tau} \left(N_i^\tau(y) - \frac{|S_i^\tau|}{|S_{i-1}^\tau|}\right)^2 \\ &\leqslant 4\sum_{i=2}^{r_\star} \frac{u_i^2}{d|S_i^\tau|} \sum_{y\in S_{i-1}^\tau} \Big[\big(N_i(y) - \mathbb{E}[N_i(y)|B_{i-1}]\big)^2 + \big(\mathbb{E}[N_i(y)|B_{i-1}] - d\big)^2 \\ &\qquad + \left(d - \frac{|S_i^\tau|}{|S_{i-1}^\tau|}\right)^2 + (N_i^\tau(y) - N_i(y))^2\Big] \\ &\leqslant 4 \max_{2\leqslant i\leqslant r_\star} \frac{|S_{i-1}^\tau|}{d|S_i^\tau|}\Big[Y_i + \mathcal{C}\log N + \big(\max_y D_y^{\mathbb{G}\setminus\mathbb{G}_\tau}\big)^2\Big] \end{aligned} \tag{3.38}$$

with very high probability. Here, in the last step, we used (3.35), $\sum_{i=0}^{r_\star} u_i^2 = 1$ and $|d - \mathbb{E}[N_i(y)|B_{i-1}]| = d|B_{i-1}|/N \leqslant \mathcal{C}$ with very high probability due to [10, Eq. (5.12b)] and Lemma A.7.

Next, we claim that

$$Y_i \leqslant \mathcal{C}\log N \log d \tag{3.39}$$

with very high probability, for $i = 2, \dots, r_\star$. The proof of (3.39) is based on a dyadic decomposition analogous to the one used in the proof of [10, Eq. (5.26)]. We distinguish two regimes and estimate

$$\begin{aligned} Y_i &\leqslant d + \frac{1}{|S_{i-1}^\tau|} \sum_{y\in S_{i-1}^\tau} \mathbb{1}_{|N_i(y)-\mathbb{E}[N_i(y)|B_{i-1}]|>d^{1/2}} \left(N_i(y) - \mathbb{E}[N_i(y)|B_{i-1}]\right)^2 \\ &\leqslant d + \frac{1}{|S_{i-1}^\tau|} \sum_{k=k_{\min}}^{0} d^2\mathrm{e}^{k+1}|\mathcal{N}_{i,k}^\tau| \end{aligned} \tag{3.40}$$

with very high probability, where we introduced

$$\begin{aligned} k_{\min} &:= \lfloor -\log d \rfloor, \\ \mathcal{N}_{i,k}^\tau &:= \left\{ y \in S_{i-1}^\tau : d^2\mathrm{e}^k < \big(N_i(y) - \mathbb{E}[N_i(y)|B_{i-1}]\big)^2 \leqslant d^2\mathrm{e}^{k+1} \right\}. \end{aligned}$$

In (3.40), we used that, with very high probability, $\big(N_i(y) - \mathbb{E}[N_i(y)|B_{i-1}]\big)^2 \leqslant d^2\big((\tau - 1/2)^2 \vee 1\big) \leqslant d^2\mathrm{e}$, because $y \in S_{i-1}^\tau$ implies the conditions $0 \leqslant N_i(y) \leqslant D_y \leqslant \tau d$ due to Proposition 3.1 (i) and $d/2 \leqslant \mathbb{E}[N_i(y)|B_{i-1}] \leqslant d$ with very high probability. By Proposition 3.1 (iv), we have $\mathcal{N}_{i,k}^\tau \subset \mathcal{N}_k^{i-1}$, where $\mathcal{N}_k^{i-1}$ is defined as in the proof of [10, Eq. (5.26)]. (Note that, in the notation of [10], there is a one-to-one mapping between $A_{(B_{i-1})}$ and B_i.) In this proof it is shown that, with very high probability,

$$|\mathcal{N}_k^{i-1}| \leqslant \ell_k, \qquad \ell_k := \frac{\mathcal{C}}{d}(|S_{i-1}| + \log N)\mathrm{e}^{-k}.$$

Using (3.36) and (3.37), and then plugging the resulting bound into (3.40) concludes the proof of (3.39).

Thus, we obtain $\|\mathbf{w}_2\| \leqslant \mathcal{C}\xi$ with very high probability, by starting from (3.38) and using (3.35), (3.39) and Proposition 3.1 (v) as well as the assumption $1+\xi^{1/2} \leqslant \tau \leqslant 2$.

Finally, we estimate $\mathbf{w}_4$. Since $\alpha_x \geqslant 2$ and $u_0 \leqslant 1$ we have that $u_{r_\star} + u_{r_\star - 1} \leqslant 3(\alpha_x - 1)^{-(r_\star - 2)/2}$. The other coefficients of $\mathbf{s}^\tau_{r_\star - 1}$, $\mathbf{s}^\tau_{r_\star}$ and $\mathbf{s}^\tau_{r_\star + 1}$ are bounded by $\mathcal{C}$ with very high probability, due to $\alpha_x \geqslant 2$ and (3.35), respectively. Therefore, (3.7) implies $\|\mathbf{w}_4\| \leqslant \mathcal{C}\xi$. This concludes the proof of Proposition 3.9. □

Proof of Lemma 3.11. We have to estimate the norm of

$$H^\tau - \widehat{H}^\tau = \Pi^\tau H^\tau \Pi^\tau - \sum_{x\in\mathcal{V}} \sum_{\sigma=\pm} \sigma \Lambda(\alpha_x) \mathbf{v}^\tau_\sigma(x) \mathbf{v}^\tau_\sigma(x)^* + \overline{\Pi}^\tau H^\tau \Pi^\tau + (\overline{\Pi}^\tau H^\tau \Pi^\tau)^*. \tag{3.41}$$

Each $x \in \mathcal{V}$ satisfies the condition of Proposition 3.9 since $\xi^{1/4} \geqslant C(\log d)^2/\sqrt{\log N}$ (see (3.8)). Hence, for any $x \in \mathcal{V}$ and $\sigma = \pm$, Proposition 3.9 yields

$$H^\tau \mathbf{v}^\tau_\sigma(x) = \sigma \Lambda(\alpha_x) \mathbf{v}^\tau_\sigma(x) + \mathbf{e}^\tau_\sigma(x), \qquad \operatorname{supp} \mathbf{e}^\tau_\sigma(x) \subset B^\tau_{r_\star+1}(x), \qquad \|\mathbf{e}^\tau_\sigma(x)\| \leqslant \mathcal{C}\xi$$

with very high probability, where the second statement follows from the first together with the definition (3.5) of $\mathbf{v}^\tau_\sigma(x)$ and Remark 3.7. By Proposition 3.1 (i), the balls $B^\tau_{2r_\star}(x)$ and $B^\tau_{2r_\star}(y)$ are disjoint for $x, y \in \mathcal{V}_\tau$ with $x \neq y$. Hence, in this case, $\mathbf{v}^\tau_\sigma(x), \mathbf{e}^\tau_\sigma(x) \perp \mathbf{v}^\tau_{\sigma'}(y), \mathbf{e}^\tau_{\sigma'}(y)$. For any $\mathbf{a} = \sum_{x\in\mathcal{V}} \sum_{\sigma=\pm} a_{x,\sigma} \mathbf{v}^\tau_\sigma(x)$, we obtain

$$\overline{\Pi}^\tau H^\tau \Pi^\tau \mathbf{a} = \sum_{x\in\mathcal{V}} \sum_{\sigma=\pm} a_{x,\sigma} \overline{\Pi}^\tau H^\tau \mathbf{v}^\tau_\sigma(x) = \overline{\Pi}^\tau \sum_{x\in\mathcal{V}} \sum_{\sigma=\pm} a_{x,\sigma} \mathbf{e}^\tau_\sigma(x).$$

Thus, with very high probability, $\|\overline{\Pi}^\tau H^\tau \Pi^\tau \mathbf{a}\|^2 \leqslant \sum_{x\in\mathcal{V}} \|\sum_{\sigma=\pm} a_{x,\sigma} \mathbf{e}^\tau_\sigma(x)\|^2 \leqslant 4\mathcal{C}^2 \sum_{x\in\mathcal{V}} \sum_{\sigma=\pm} a^2_{x,\sigma} \xi^2 = 4\mathcal{C}^2\xi^2 \|\mathbf{a}\|^2$ by orthogonality. Therefore, $\|\overline{\Pi}^\tau H^\tau \Pi^\tau\| \leqslant \mathcal{C}\xi$ with very high probability. Similarly, the representation

$$\left(\Pi^\tau H^\tau \Pi^\tau - \sum_{x\in\mathcal{V}} \sum_{\sigma=\pm} \sigma \Lambda(\alpha_x) \mathbf{v}^\tau_\sigma(x) \mathbf{v}^\tau_\sigma(x)^* \right) \mathbf{a} = \Pi^\tau \sum_{x\in\mathcal{V}} \sum_{\sigma=\pm} a_{x,\sigma} \mathbf{e}^\tau_\sigma(x)$$

yields the desired estimate on the sum of the two first terms on the right-hand side of (3.41). □

3.4. Proof of Proposition 3.12. In this section we prove Proposition 3.12. Its proof relies on two fundamental tools.

The first tool is a quadratic form estimate, which estimates H in terms of the diagonal matrix of the vertex degrees. It is an improvement of [10, Proposition 6.1]. To state it, for two Hermitian matrices X and Y we use the notation $X \leqslant Y$ to mean that $Y - X$ is a nonnegative matrix, and $|X|$ is the absolute value function applied to the matrix X.

Proposition 3.13. *Let $4 \leqslant d \leqslant 3 \log N$. Then, with very high probability, we have*

$$|H| \leqslant I + (1 + 2d^{-1/2})Q + \mathcal{C}\frac{\log N}{d^2} \vee d^{-1/2},$$

where Q is the diagonal matrix with diagonal $(\alpha_x)_{x\in[N]}$.

The second tool is a delocalization estimate for an eigenvector $\mathbf{w}$ of $\widehat{H}^\tau$ associated with an eigenvalue $\lambda > 2$. Essentially, it says that w_x is small at any $x \in \mathcal{V}_\tau$ unless $\mathbf{w}$ happens to be the specific eigenvector $\mathbf{v}^\tau_\pm(x)$ of $\widehat{H}^\tau$, which is by definition localized around x. Thus, in any ball $B^\tau_{2r_\star}(x)$ around $x \in \mathcal{V}_\tau$, all eigenvectors except $\mathbf{v}^\tau_\pm(x)$ are locally delocalized in the sense that their magnitudes at x are small. Using that the balls $(B^\tau_{2r_\star}(x))_{x\in\mathcal{V}_\tau}$ are disjoint, this implies that eigenvectors of $\overline{\Pi}^\tau H^\tau \overline{\Pi}^\tau$ have negligible mass on the set $\mathcal{V}$.

Proposition 3.14. *Let d satisfy (1.10). If $1+\xi^{1/2} \leqslant \tau \leqslant 2$ then the following holds with very high probability. Let λ be an eigenvalue of $\widehat{H}^\tau$ with $\lambda > 2\tau + \mathcal{C}\xi$ and $\mathbf{w} = (w_x)_{x\in[N]}$ its corresponding eigenvector.*

(i) *If $x \in \mathcal{V}$ and $\mathbf{v}^\tau_\pm(x) \perp \mathbf{w}$ or if $x \in \mathcal{V}_\tau \setminus \mathcal{V}$ then*

$$\frac{|w_x|}{\|\mathbf{w}|_{B^\tau_{2r_\star}(x)}\|} \leqslant \frac{\lambda^2}{(\lambda - 2\tau - \mathcal{C}\xi)^2}\left(\frac{2\tau+\mathcal{C}\xi}{\lambda}\right)^{r_\star}.$$

(ii) *Let $\mathbf{w}$ be normalized. If $\mathbf{v}^\tau_\pm(x) \perp \mathbf{w}$ for all $x \in \mathcal{V}$ then*

$$\sum_{x\in\mathcal{V}_\tau} w_x^2 \leqslant \frac{\lambda^4}{(\lambda - 2\tau - \mathcal{C}\xi)^4}\left(\frac{2\tau+\mathcal{C}\xi}{\lambda}\right)^{2r_\star}.$$

Analogous results hold for $\lambda < -2\tau - \mathcal{C}\xi$.

We may now conclude the proof of Proposition 3.12.

Proof of Proposition 3.12. By Proposition 3.13, Lemma 3.11, and Lemma 3.8 we have

$$\begin{aligned}\widehat{H}^\tau &\leqslant I + (1+2d^{-1/2})Q + \mathcal{C}\frac{\log N}{d^2} \vee d^{-1/2} + \|H - H^\tau\| + \|H^\tau - \widehat{H}^\tau\| \\ &\leqslant I + (1+2d^{-1/2})Q + \mathcal{C}(\xi + \xi_{\tau-1})\end{aligned} \tag{3.42}$$

with very high probability, where we used $\frac{\log N}{d^2} \vee d^{-1/2} \leqslant (\xi + \xi_{\tau-1})$.

Arguing by contradiction, we assume that there exists an eigenvalue $\lambda > 2\tau + \mathcal{C}'(\xi + \xi_{\tau-1})$ of $\overline{\Pi}^\tau H^\tau \overline{\Pi}^\tau$ for some $\mathcal{C}' \geqslant 2\mathcal{C}$ to be chosen later. By the lower bound in (1.10), we may assume that $\mathcal{C}'\xi \leqslant 1$. Thus, by the definition of $\widehat{H}^\tau$, there is an eigenvector $\mathbf{w}$ of $\widehat{H}^\tau$ corresponding to λ, which is orthogonal to $\mathbf{v}^\tau_\pm(x)$ for all $x \in \mathcal{V}$. From (3.42), we conclude

$$\begin{aligned}\lambda = \langle \mathbf{w}, \widehat{H}^\tau \mathbf{w}\rangle &\leqslant 1 + (1+2d^{-1/2}) \sum_{x\notin\mathcal{V}_\tau} w_x^2 \tau \\ &+ (1+2d^{-1/2}) \sum_{x\in\mathcal{V}_\tau} w_x^2 \max_{y\in[N]} \alpha_y + \mathcal{C}(\xi + \xi_{\tau-1}).\end{aligned} \tag{3.43}$$

It remains to estimate the two sums on right-hand side of (3.43).

Since $\mathbf{w} \perp \mathbf{v}^\tau_\pm(x)$ for all $x \in \mathcal{V}$, we can apply Proposition 3.14 (ii). We find

$$\begin{aligned}2r_\star \log\left(\frac{2\tau+\mathcal{C}\xi}{\lambda}\right) &\leqslant 2r_\star \log\left(\frac{2\tau+\mathcal{C}\xi}{2\tau+\mathcal{C}'\xi}\right) \leqslant -2r_\star \frac{(\mathcal{C}'-\mathcal{C})\xi}{2\tau+\mathcal{C}'\xi} \\ &\leqslant -\frac{c(\mathcal{C}'-\mathcal{C})}{3}\sqrt{\log N}\,\xi\,,\end{aligned} \tag{3.44}$$

where in the last step we recalled the definition (1.8) and used that $\tau \leqslant 2$ and $\mathcal{C}'\xi \leqslant 1$. Using the estimate

$$\frac{\lambda^4}{(\lambda - 2\tau - \mathcal{C}\xi)^4} \leqslant \frac{C}{(\mathcal{C}' - \mathcal{C})^4 \xi^4},$$

combined with Proposition 3.14 (ii), (3.44) and Lemma A.7, yields

$$\begin{aligned}\frac{1}{\xi} \sum_{x \in \mathcal{V}_\tau} w_x^2 \max_{y \in [N]} \alpha_y &\leqslant \frac{C \log N}{(\mathcal{C}' - \mathcal{C})^4 \xi^5} \exp\left(-\frac{c(\mathcal{C}' - \mathcal{C})}{3} \sqrt{\log N}\, \xi \right) \\ &\leqslant \frac{C d^5 \log N}{(\mathcal{C}' - \mathcal{C})^4} \exp\left(-\frac{c(\mathcal{C}' - \mathcal{C})}{3} \frac{\log N}{d} \log d \right) \\ &\leqslant \frac{C d^5 \log N}{(\mathcal{C}' - \mathcal{C})^4} \frac{1}{d^8} \leqslant 1,\end{aligned}$$

where the third step follows by choosing $\mathcal{C}'$ large enough, depending on $\mathcal{C}$.

Plugging this estimate into (3.43) and using $\sum_x w_x^2 \leqslant 1$ to estimate the first sum in (3.43), we obtain $\lambda \leqslant 2\tau + 2\mathcal{C}(\xi + \xi_{\tau-1})$. This is a contradiction to the assumption $\lambda > 2\tau + \mathcal{C}'(\xi + \xi_{\tau-1})$. The proof of Proposition 3.12 is therefore complete. □

Proof of Proposition 3.13. We only establish an upper bound on H. The proof of the same upper bound on $-H$ is identical and, therefore, omitted.

We introduce the matrices $H(t) = (H_{xy}(t))_{x,y \in [N]}$ and $M(t) = (\delta_{xy} m_x(t))_{x,y \in [N]}$ with entries

$$H_{xy}(t) := \frac{t H_{xy}}{t^2 - H_{xy}^2}, \quad m_x(t) := 1 + \sum_y \frac{H_{xy}^2}{t^2 - H_{xy}^2}$$

By the estimate on the spectral radius of the nonbacktracking matrix associated with H in [15, Theorem 2.5] and the Ihara–Bass-type formula in [15, Lemma 4.1] we have, with very high probability, $\det(M(t) - H(t)) \neq 0$ for all $t \geqslant 1 + \mathcal{C}d^{-1/2}$. Because $(M(t) - H(t)) \to I$ as $t \to \infty$, the matrix $M(t) - H(t)$ is positive definite for large enough t. By continuity of the eigenvalues, we conclude that all eigenvalues of $M(t) - H(t)$ stay positive for $t \geqslant 1 + \mathcal{C}d^{-1/2}$, and hence

$$H(t) \leqslant M(t) \tag{3.45}$$

for all $t \geqslant 1 + \mathcal{C}d^{-1/2}$ with very high probability. We now define the matrix $\Delta = (\Delta_{xy})_{x,y \in [N]}$ with

$$\Delta_{xy} := \begin{cases} H_{xy}(t) - t^{-1} H_{xy} & \text{if } x \neq y \\ \sum_{y'} |H_{xy'}(t) - t^{-1} H_{xy'}| & \text{if } x = y. \end{cases}$$

It is easy to check that Δ is a nonnegative matrix. We also have

$$\sum_{y'} |H_{xy'}(t) - t^{-1} H_{xy'}| \leqslant \sum_{y'} \frac{|H_{xy'}|^3}{t(t^2 - H_{xy'}^2)} \leqslant \frac{2}{t^3 d^{1/2}} \left(\alpha_x + \frac{1}{d} \right),$$

where we used that $|H_{xy}| \leqslant d^{-1/2}$ and $\sum_{y'} H_{xy'}^2 \leqslant \alpha_x + \frac{d}{N}$ by definition of H. We use this to estimate the diagonal entries of Δ and obtain

$$0 \leqslant \Delta \leqslant H(t) - t^{-1}H + \frac{2}{t^3\sqrt{d}}Q + \frac{2}{t^3 d^{3/2}}. \tag{3.46}$$

On the other hand, for the diagonal matrix $M(t)$, we have the trivial upper bound

$$M(t) \leqslant I + t^{-2}Q + \mathcal{C}\frac{\log N}{d^2} \tag{3.47}$$

since $\alpha_x \leqslant \mathcal{C}(\log N)/d$ with very high probability due to Lemma A.7. Finally, combining (3.45), (3.46) and (3.47) yields

$$t^{-1}H \leqslant I + \left(t^{-2} + \frac{2}{t^3\sqrt{d}}\right)Q + \mathcal{C}\frac{\log N}{d^2}$$

and Proposition 3.13 follows by choosing $t = 1 + \mathcal{C}d^{-1/2}$. □

What remains is the proof of Proposition 3.14. The underlying principle behind the proof is the same as that of the Combes–Thomas estimate [25]: the Green function $((\lambda - Z)^{-1})_{ij}$ of a local operator Z at a spectral parameter λ separated from the spectrum of Z decays exponentially in the distance between i and j, at a rate inversely proportional to the distance from λ to the spectrum of Z. Here *local* means that Z_{ij} vanishes if the distance between i and j is larger than 1. Since a graph is equipped with a natural notion of distance and the adjacency matrix is a local operator, a Combes–Thomas estimate would be applicable directly on the level of the graph, at least for the matrix H^τ. For our purposes, however, we need a *radial* version of a Combes–Thomas estimate, obtained by first tridiagonalizing (a modification of) $\widehat{H}^\tau$ around a vertex $x \in \mathcal{V}_\tau$ (see Appendix A.2). In this formulation, the indices i and j have the interpretation of radii around the vertex x, and the notion of distance is simply that of $\mathbb{N}$ on the set of radii. Since Z is tridiagonal, the locality of Z is trivial, although the matrix $\widehat{H}^\tau$ (or its appropriate modification) is not a local operator on the graph $\mathbb{G}_\tau$.

To ensure the separation of $\lambda > 2\tau + o(1)$ and the spectrum of Z, we cannot choose Z to be the tridiagonalization of $\widehat{H}^\tau$, since λ is an eigenvalue of $\widehat{H}^\tau$. In fact, Z is the tridiagonalization of a new matrix $\widehat{H}^{\tau,x}$, obtained by restricting $\widehat{H}^\tau$ to the ball $B^\tau_{2r_\star}(x)$ and possibly subtracting a suitably chosen rank-two matrix, which allows us to show $\|\widehat{H}^{\tau,x}\| \leqslant 2\tau + o(1)$. By the orthogonality assumption on $\mathbf{w}$, we then find that the Green function $((\lambda - Z)^{-1})_{ir_\star}$, $0 \leqslant i < r_\star$, and the eigenvector components in the radial basis u_i, $0 \leqslant i < r_\star$, satisfy the same linear difference equation. The exponential decay of $((\lambda - Z)^{-1})_{ir_\star}$ in $r_\star - i$ then implies that, for each $x \in \mathcal{V}_\tau$, $u_0^2 \leqslant o(1/\log N)\sum_{i=0}^{r_*} u_i^2$. Going back to the original vertex basis, this implies that $w_x^2 \leqslant o(1/\log N)\|\mathbf{w}|_{B^\tau_{2r_\star}(x)}\|^2$ for all $x \in \mathcal{V}_\tau$, from which Proposition 3.14 follows since the balls $B^\tau_{2r_\star}(x)$, $x \in \mathcal{V}_\tau$, are disjoint.

Proof of Proposition 3.14. For a matrix $M \in \mathbb{R}^{N\times N}$ and a set $V \subset [N]$, we use the notation $(M|_V)_{xy} := \mathbb{1}_{x,y\in V}M_{xy}$.

We begin with part (i). We first treat the case $x \in \mathcal{V}$. To that end, we introduce the matrix

$$\widehat{H}^{\tau,x} := \widehat{H}^\tau|_{B^\tau_{2r_\star}(x)} - \Lambda(\alpha_x)\mathbf{v}_+^\tau(x)\mathbf{v}_+^\tau(x)^* + \Lambda(\alpha_x)\mathbf{v}_-^\tau(x)\mathbf{v}_-^\tau(x)^*. \tag{3.48}$$

We claim that, with very high probability,

$$\|\widehat{H}^{\tau,x}\| \leqslant 2\tau + \mathcal{C}\xi\,. \tag{3.49}$$

To show (3.49), we begin by noting that, by Proposition 3.1 (i) and (ii), $\mathbb{G}_\tau$ restricted to $B^\tau_{2r_\star}(x)$ is a tree whose root x has $\alpha_x d$ children and all other vertices have at most τd children. Hence, Lemma A.5 yields $\|H^\tau|_{B^\tau_{2r_\star}(x)}\| \leqslant \sqrt{\tau}\Lambda(\alpha_x/\tau\vee 2)$. Using Lemma 3.11 we find

$$\|\widehat{H}^\tau|_{B^\tau_{2r_\star}(x)} - H^\tau|_{B_{2r_\star}(x)}\| \leqslant \mathcal{C}\xi \tag{3.50}$$

with very high probability, and since $\mathbf{v}^\tau_\pm(x)$ is an eigenvector of $\widehat{H}^\tau|_{B^\tau_{2r_\star}(x)}$ with eigenvalue $\pm\Lambda(\alpha_x)$, we conclude

$$\|\widehat{H}^{\tau,x}\| \leqslant \sqrt{\tau}\Lambda(\alpha_x/\tau \vee 2) + \mathcal{C}\xi \tag{3.51}$$

with very high probability. The estimate (3.51) is rough in the sense that the subtraction of the two last terms of (3.48) is not needed for its validity (since $\Lambda(\alpha_x) \leqslant \sqrt{\tau}\Lambda(\alpha_x/\tau\vee 2)$). Nevertheless, it is sufficient to establish (3.49) in the following cases, which may be considered degenerate.

If $\alpha_x \leqslant 2\tau$ then (3.51) immediately implies (3.49), since $\sqrt{\tau} \leqslant \tau$. Moreover, if $\alpha_x > 2\tau$ and $\Lambda(\alpha_x) \leqslant 2\sqrt{\tau} + \mathcal{C}\xi$, then (3.51) implies

$$\|\widehat{H}^{\tau,x}\| \leqslant \sqrt{\tau}\Lambda(\alpha_x/\tau) + \mathcal{C}\xi \leqslant \sqrt{\tau}\Lambda(\alpha_x) + \mathcal{C}\xi \leqslant 2\tau + 3\mathcal{C}\xi\,,$$

which is (3.49) after renaming the constant $\mathcal{C}$.

Hence, to prove (3.49), it suffices to consider the case $\Lambda(\alpha_x) > 2\sqrt{\tau} + \mathcal{C}\xi$. By Proposition 3.1 (i) and (ii), $\mathbb{G}_\tau$ restricted to $B^\tau_{2r_\star}(x)\setminus\{x\}$ is a forest of maximal degree at most τd. Lemma A.4 therefore yields $\|H^\tau|_{B^\tau_{2r_\star}(x)\setminus\{x\}}\| \leqslant 2\sqrt{\tau}$. Moreover, the adjacency matrix of the star graph consisting of all edges of $\mathbb{G}_\tau$ incident to x has precisely two nonzero eigenvalues, $\pm\sqrt{d\alpha_x}$. By first order perturbation theory, we therefore conclude that $H^\tau|_{B^\tau_{2r_\star}(x)}$ has at most one eigenvalue strictly larger than $2\sqrt{\tau}$ and at most one strictly smaller than $-2\sqrt{\tau}$. Using (3.50) we conclude that $\widehat{H}^\tau|_{B^\tau_{2r_\star}(x)}$ has at most one eigenvalue strictly larger than $2\sqrt{\tau}+\mathcal{C}\xi$ and at most one strictly smaller than $-2\sqrt{\tau}-\mathcal{C}\xi$. Since $\mathbf{v}^\tau_+(x)$ (respectively $\mathbf{v}^\tau_-(x)$) is an eigenvector of $\widehat{H}^\tau|_{B^\tau_{2r_\star}(x)}$ with eigenvalue $\Lambda(\alpha_x)$ (respectively $-\Lambda(\alpha_x)$), and since $\Lambda(\alpha_x) > 2\sqrt{\tau} + \mathcal{C}\xi$, we conclude (3.49).

Next, let $(\mathbf{g}_i)_{i=0}^{r_\star}$ be the Gram–Schmidt orthonormalization of the vectors $((\widehat{H}^{\tau,x})^i \mathbf{1}_x)_{i=0}^{r_\star}$. We claim that

$$\operatorname{supp}\mathbf{g}_i \subset B^\tau_{r_\star+i}(x)\,. \tag{3.52}$$

for $i = 0, \ldots, r_\star$. The proof proceeds by induction. The base case for $i = 0$ holds trivially. For the induction step, it suffices to prove for $0 \leqslant i < r_\star$ that if $\operatorname{supp}\mathbf{g}_i \subset B^\tau_{r_\star+i}(x)$ then

$$\operatorname{supp}(\widehat{H}^{\tau,x}\mathbf{g}_i) \subset B^\tau_{r_\star+i+1}(x) \tag{3.53}$$

To that end, we note that by Proposition 3.1 (i) we have $\widehat{H}^{\tau,x} = \big(\overline{\Pi}^\tau H^\tau \overline{\Pi}^\tau\big)|_{B^\tau_{2r_\star}(x)}$. Hence, by induction assumption, Proposition 3.1 (i), and Remark 3.7,

$$\widehat{H}^{\tau,x}\mathbf{g}_i = \bigg(I - \sum_{\sigma=\pm}\mathbf{v}^\tau_\sigma(x)\mathbf{v}^\tau_\sigma(x)^*\bigg)H^\tau\bigg(I - \sum_{\sigma=\pm}\mathbf{v}^\tau_\sigma(x)\mathbf{v}^\tau_\sigma(x)^*\bigg)\mathbf{g}_i\,,$$

and we conclude (3.53), as $\operatorname{supp} \mathbf{v}_\sigma^\tau(x) \subset B_{r_\star}^\tau(x)$.

Let $Z = (Z_{ij})_{i,j=0}^{r_\star}$, $Z_{ij} := \langle \mathbf{g}_i, \widehat{H}^{\tau,x} \mathbf{g}_j \rangle$, be the tridiagonal representation of $\widehat{H}^{\tau,x}$ up to radius $r_\star$ (see Appendix A.2 below). Owing to (3.49), we have

$$\|Z\| \leqslant 2\tau + \mathcal{C}\xi. \tag{3.54}$$

We set $u_i := \langle \mathbf{g}_i, \mathbf{w} \rangle$ for any $0 \leqslant i \leqslant r_\star$. Because $\mathbf{w}$ is an eigenvector of $\widehat{H}^\tau$ that is orthogonal to $\mathbf{v}_\pm^\tau(x)$, for any $i < r_\star$, (3.52) implies

$$\begin{aligned} \lambda u_i &= \left\langle \mathbf{g}_i, \left(\widehat{H}^\tau - \Lambda(\alpha_x) \mathbf{v}_+^\tau(x) \mathbf{v}_+^\tau(x)^* + \Lambda(\alpha_x) \mathbf{v}_-^\tau(x) \mathbf{v}_-^\tau(x)^* \right) \mathbf{w} \right\rangle \\ &= \left\langle \widehat{H}^{\tau,x} \mathbf{g}_i, \mathbf{w} \right\rangle \\ &= \langle Z_{ii} \mathbf{g}_i + Z_{i\,i+1} \mathbf{g}_{i+1} + Z_{i\,i-1} \mathbf{g}_{i-1}, \mathbf{w} \rangle \\ &= Z_{ii} u_i + Z_{i\,i+1} u_{i+1} + Z_{i\,i-1} u_{i-1} \end{aligned} \tag{3.55}$$

with the conventions $u_{-1} = 0$ and $Z_{0,-1} = 0$. Let $G(\lambda) := (\lambda - Z)^{-1}$ be the resolvent of Z at λ. Note that $\lambda - Z$ is invertible since $\lambda > \|Z\|$ by assumption and (3.54). Since $\big((\lambda - Z)G(\lambda)\big)_{i\,r_\star} = 0$ for $i < r_\star$, we find

$$\lambda G_{i r_\star}(\lambda) = Z_{ii} G_{i r_\star}(\lambda) + Z_{i\,i+1} G_{i+1\,r_\star}(\lambda) + Z_{i\,i-1} G_{i-1\,r_\star}(\lambda).$$

Therefore $(G_{i r_\star}(\lambda))_{i \leqslant r_\star}$ and $(u_i)_{i \leqslant r_\star}$ satisfy the same linear recursive equation (cf. (3.55)); solving them recursively from $i = 0$ to $i = r_\star$ yields

$$\frac{G_{i r_\star}(\lambda)}{G_{r_\star r_\star}(\lambda)} = \frac{u_i}{u_{r_\star}} \tag{3.56}$$

for all $i \leqslant r_\star$. Moreover, as $\lambda > \|Z\|$ by assumption and (3.54), we have the convergent Neumann series $G(\lambda) = \frac{1}{\lambda} \sum_{k \geqslant 0} (Z/\lambda)^k$. Thus, the offdiagonal entries of the resolvent satisfy

$$G_{0 r_\star}(\lambda) = \frac{1}{\lambda} \sum_{k \geqslant 0} \big((Z/\lambda)^k \big)_{0 r_\star}.$$

Since Z is tridiagonal, we deduce that $\big((Z/\lambda)^k\big)_{0 r_\star} = 0$ if $k < r_\star$, so that, by (3.54),

$$|G_{0 r_\star}(\lambda)| \leqslant \left(\frac{2\tau + \mathcal{C}\xi}{\lambda} \right)^{r_\star} \frac{1}{\lambda - 2\tau - \mathcal{C}\xi}. \tag{3.57}$$

On the other hand, for the diagonal entries of the resolvent, we get, by splitting the summation over k into even and odd values,

$$\begin{aligned} G_{r_\star r_\star}(\lambda) &= \frac{1}{\lambda} \sum_{k \geqslant 0} \big((Z/\lambda)^k \big)_{r_\star r_\star} = \frac{1}{\lambda} \sum_{k \geqslant 0} \Big((Z/\lambda)^k (I + Z/\lambda)(Z/\lambda)^k \Big)_{r_\star r_\star} \\ &\geqslant \frac{1}{\lambda} (I + Z/\lambda)_{r_\star r_\star} \geqslant \frac{1}{\lambda} \left(1 - \frac{2\tau + \mathcal{C}\xi}{\lambda} \right), \end{aligned} \tag{3.58}$$

where in the thid step we discarded the terms $k > 0$ to obtain a lower bound using that $I + Z/\lambda \geqslant 0$ by (3.54), and in the last step we used (3.54). Hence, the definition of u_i and (3.52) imply

$$\frac{|w_x|}{\|\mathbf{w}|_{B_{2r_\star}^\tau(x)}\|} \leqslant \frac{|u_0|}{\big(\sum_{i=0}^{r_\star} u_i^2\big)^{1/2}} \leqslant \frac{|u_0|}{|u_{r_\star}|} = \frac{|G_{0 r_\star}(\lambda)|}{G_{r_\star r_\star}(\lambda)} \leqslant \frac{\lambda^2}{(\lambda - 2\tau - \mathcal{C}\xi)^2} \left(\frac{2\tau + \mathcal{C}\xi}{\lambda} \right)^{r_\star}.$$

Here, we used (3.56) in third step and (3.57) as well as (3.58) in the last step. This concludes the proof of (i) for $x \in \mathcal{V}$.

In the case $x \in \mathcal{V}_\tau \setminus \mathcal{V}$, we set $\widehat{H}^{\tau,x} := \widehat{H}^\tau|_{B^\tau_{2r_\star}(x)}$. We claim that (3.49) holds. To see that, we use Proposition 3.1 (i) and (ii) as well as Lemma A.5 with $p = d(2+\xi^{1/4})$ and $q = d\tau$ to obtain

$$\|H^\tau|_{B^\tau_{2r_\star}(x)}\| \leqslant \sqrt{\tau}\Lambda((2+\xi^{1/4})/\tau \vee 2) \leqslant 2\tau.$$

Here, the last step is trivial if $\tau \geqslant 1+\xi^{1/4}/2$ and, if $\tau \in [1+\xi^{1/2}, 1+\xi^{1/4}/2]$, we used that $f(\tau) := \sqrt{\tau}\Lambda((2+\xi^{1/4})/\tau)/(2\tau)$ is monotonically decreasing on this interval and $f(1+\xi^{1/2}) \leqslant 1$, as can be seen by an explicit analysis of the function f. Now we may take over the previous argument verbatim to prove (i) for $x \in \mathcal{V}_\tau \setminus \mathcal{V}$.

Finally, we prove (ii). By (i) we have

$$\sum_{x\in\mathcal{V}_\tau} w_x^2 \leqslant \sum_{x\in\mathcal{V}_\tau} \|\mathbf{w}|_{B^\tau_{2r_\star}(x)}\|^2 \frac{\lambda^4}{(\lambda-2\tau-\mathcal{C}\xi)^4}\left(\frac{2\tau+\mathcal{C}\xi}{\lambda}\right)^{2r_\star}$$
$$\leqslant \frac{\lambda^4}{(\lambda-2\tau-\mathcal{C}\xi)^4}\left(\frac{2\tau+\mathcal{C}\xi}{\lambda}\right)^{2r_\star},$$

where we used that the the balls $\{B^\tau_{2r_\star}(x) : x \in \mathcal{V}_\tau\}$ are disjoint, which implies $1 = \|\mathbf{w}\|^2 \geqslant \sum_{x\in\mathcal{V}_\tau} \|\mathbf{w}|_{B^\tau_{2r_\star}(x)}\|^2$. □

3.5. Proof of Proposition 3.1. We conclude this section with the proof of Proposition 3.1.

Proof of Proposition 3.1. Parts (i)–(v) follow immediately from parts (i)–(iv) and (vi) of [10, Lemma 7.2]. To see this, we remark that the function h from [10] satisfies $h((\tau-1)/2) \asymp (\tau-1)^2$ for $1 < \tau \leqslant 2$. Moreover, by Lemma A.7 and the upper bound on d, we have $\max_x D_x \leqslant \mathcal{C}\log N$ with very high probability. Hence, choosing the universal constant c small enough in (1.8) and recalling the lower bound on $\tau - 1$, in the notation of [10, Equations (5.1) and (7.2)] we obtain for any $x \in \mathcal{V}_\tau$ the inequality $2r_\star \leqslant (\frac{1}{4}r_x) \wedge (\frac{1}{2}r(\tau))$ with very high probability. This yields parts (i)–(v).

It remains to prove (vi), which is the content of the rest of this proof. From now on we systematically omit the argument x from our notation. Part (v) already implies the bound

$$|S_1 \setminus S_1^\tau| = D_x^{\mathbb{G}\setminus\mathbb{G}_\tau} \leqslant \mathcal{C}\frac{\log N}{(\tau-1)^2 d} \tag{3.59}$$

with very high probability, which is (3.2) for $i = 1$.

From [10, Eq. (7.13)] we find

$$|S_i \setminus S_i^\tau| \leqslant \sum_{y\in S_1\setminus S_1^\tau} |S_{i-1}(y)|.$$

(As a guide to the reader, this estimate follows from the construction of $\mathbb{G}_\tau$ given in [10, Proof of Lemma 7.2], which ensures that if a vertex $z \in S_i$ is not in S_i^τ then any path in

$\mathbb{G}$ of length i connecting z to x is cut in $\mathbb{G}_\tau$ at its edge incident to x.) Hence, in order to show (vi) for $i \geqslant 2$, it suffices to prove

$$\sum_{y \in S_1 \setminus S_1^\tau} |S_{i-1}(y)| \leqslant \mathcal{C} \frac{\log N}{(\tau-1)^2} d^{i-2} \tag{3.60}$$

with very high probability, for all $2 \leqslant i \leqslant 2r_\star$.

We start with the case $i = 2$. We shall use the relation

$$\sum_{y \in S_1 \setminus S_1^\tau} |S_1(y)| = \sum_{y \in S_1 \setminus S_1^\tau} N_2(y) + \sum_{y \in S_1 \setminus S_1^\tau} |S_1(y) \cap S_1| + |S_1 \setminus S_1^\tau|\,, \tag{3.61}$$

where, for $y \in S_1$, we introduced $N_2(y) := |S_1(y) \cap S_2|$. Note that $N_2(y)$ is the number of vertices in S_2 connected to x via a path of minimal length passing through y. The identity (3.61) is a direct consequence of $|S_1(y)| = |S_1(y) \cap S_2| + |S_1(y) \cap S_1| + |S_1(y) \cap S_0|$ using the definition of N_2 and $|S_1(y) \cap S_0| = |S_1(y) \cap \{x\}| = 1$.

The second and third terms of (3.61) are smaller than the right-hand side of (3.60) for $i = 2$ due to [10, Eq. (5.23)] and (3.59), respectively. Hence, it remains to estimate the first term on the right-hand side of (3.61) in order to prove (3.60) for $i = 2$.

To that end, we condition on the ball B_1 and abbreviate $\mathbb{P}_{B_1}(\cdot) := \mathbb{P}(\cdot \mid B_1)$. Since

$$N_2(y) = \sum_{z \in [N] \setminus B_1} A_{yz}\,, \tag{3.62}$$

we find that conditioned on B_1 the random variables $(N_2(y))_{y \in S_1}$ are independent $\mathrm{Binom}(N - |B_1|, d/N)$ random variables. We abbreviate $\Gamma := \frac{\log N}{(\tau-1)^2}$. For given $\mathcal{C}, \mathcal{C}'$, we set $\mathcal{C}'' := \mathcal{C}' + 2\mathcal{C}$ and estimate

$$\begin{aligned}
&\mathbb{P}_{B_1}\Big(\sum_{y \in S_1 \setminus S_1^\tau} N_2(y) \geqslant \mathcal{C}''\Gamma \Big) \\
&\leqslant \mathbb{P}_{B_1}\Big(\sum_{y \in S_1 \setminus S_1^\tau} \mathbb{1}_{N_2(y) \geqslant 2d} N_2(y) \geqslant (\mathcal{C}'' - 2\mathcal{C})\Gamma \Big) \\
&\quad + \mathbb{P}_{B_1}\Big(\sum_{y \in S_1 \setminus S_1^\tau} \mathbb{1}_{N_2(y) < 2d} N_2(y) \geqslant 2\mathcal{C}\Gamma \Big) \\
&\leqslant \mathbb{P}_{B_1}\Big(\sum_{y \in S_1} \mathbb{1}_{2d \leqslant N_2(y) \leqslant N^{1/4}} N_2(y) \geqslant \mathcal{C}'\Gamma \Big) \\
&\quad + \sum_{y \in S_1} \mathbb{P}_{B_1}\big(N_2(y) \geqslant N^{1/4}\big) + \mathbb{P}_{B_1}\big(|S_1 \setminus S_1^\tau| \geqslant \mathcal{C}\Gamma d^{-1}\big).
\end{aligned} \tag{3.63}$$

In order to estimate the first term on the right-hand side of (3.63), we shall prove that if $|B_1| \leqslant N^{1/4}$ then

$$\mathbb{E}_{B_1}\Big[\exp\Big(\mathbb{1}_{2d \leqslant N_2(y) \leqslant N^{1/4}} N_2(y) t\Big)\Big] \leqslant 2 \tag{3.64}$$

for all $y \in S_1$ and $t \leqslant 1/8$. To that end, we estimate

$$\mathbb{E}_{B_1}\Big[\exp\Big(\mathbb{1}_{2d \leqslant N_2(y) \leqslant N^{1/4}} N_2(y) t\Big)\Big] \leqslant 1 + \mathbb{E}_{B_1}\Big[\mathbb{1}_{2d \leqslant N_2(y) \leqslant N^{1/4}} \mathrm{e}^{N_2(y) t}\Big].$$

With Poisson approximation, Lemma A.6 below, we obtain (assuming that $2d$ is an integer to simplify notation)

$$\begin{aligned}\mathbb{E}_{B_1}\Big[\mathbb{1}_{2d\leqslant N_2(y)\leqslant N^{1/4}}\mathrm{e}^{N_2(y)t}\Big] &= \sum_{2d\leqslant k\leqslant N^{1/4}}\frac{(d-d|B_1|/N)^k\mathrm{e}^{tk}}{k!}\mathrm{e}^{-d+d|B_1|/N}\big(1+O(N^{-1/2})\big)\\ &\leqslant \sum_{k\geqslant 2d}\frac{d^k\mathrm{e}^{tk}}{k!}\mathrm{e}^{-d}\big(1+O(N^{-1/2})\big)\\ &= \frac{d^{2d}\mathrm{e}^{2td}}{(2d)!}\mathrm{e}^{-d}\sum_{i\geqslant 0}\frac{d^i\mathrm{e}^{ti}}{\prod_{j=2d+1}^{2d+i}j}\big(1+O(N^{-1/2})\big)\\ &\leqslant \frac{d^{2d}\mathrm{e}^{2td}}{(2d)!}\mathrm{e}^{-d}\sum_{i\geqslant 0}\frac{2d^i\mathrm{e}^{ti}}{(2d)^i} = \frac{d^{2d}\mathrm{e}^{2td}}{(2d)!}\mathrm{e}^{-d}\frac{2}{(1-\mathrm{e}^t/2)}.\end{aligned}$$

By Stirling's approximation we get

$$\log\left(\frac{d^{2d}\mathrm{e}^{2td}}{(2d)!}\mathrm{e}^{-d}\right) = d\,(2t-2\log 2+1)-\frac{1}{2}\log(4\pi d)+\mathrm{o}(1).$$

The term in the parentheses on the right-hand side is negative for $t\leqslant 1/8$, and hence

$$\mathbb{E}_{B_1}\Big[\mathbb{1}_{2d\leqslant N_2(y)\leqslant N^{1/4}}\mathrm{e}^{N_2(y)t}\Big]\leqslant 1$$

for large enough d, which gives (3.64). Since the family $(N_2(y))_{y\in S_1}$ is independent conditioned on B_1, we can now use Chebyshev's inequality to obtain, for $0\leqslant t\leqslant 1/8$,

$$\begin{aligned}\mathbb{P}_{B_1}\Big(\sum_{y\in S_1}\mathbb{1}_{2d\leqslant N_2(y)\leqslant N^{1/4}}N_2(y)\geqslant \mathcal{C}'\Gamma\Big) &\leqslant \frac{\max_{y\in S_1}\Big(\mathbb{E}_{B_1}\exp\Big(\mathbb{1}_{2d\leqslant N_2(y)\leqslant N^{1/4}}N_2(y)t\Big)\Big)^{|S_1|}}{\mathrm{e}^{t\mathcal{C}'\Gamma}}\\ &\leqslant \exp\left(|S_1|\log 2-\mathcal{C}'\frac{t}{(\tau-1)^2}\log N\right).\end{aligned}$$

Now we set $t=1/8$, recall the bound $\tau\leqslant 2$, plug this estimate back into (3.63), and take the expectation. We use Lemma A.7 to estimate $|S_1|$, which in particular implies that $|B_1|\leqslant N^{1/4}$ with very high probability; this concludes the estimate of the expectation of the first term of (3.63) by choosing $\mathcal{C}'$ large enough. Next, the expectation of the second term is easily estimated by Lemma A.7 since $N_2(y)$ has law $\mathrm{Binom}(N-|B_1|,d/N)$ when conditioned on B_1. Finally, the expectation of the last term of (3.63) is estimated by (3.59) by choosing $\mathcal{C}$ large enough. This concludes the proof of (3.60) for $i=2$.

We now prove (3.60) for $i+1$ with $i\geqslant 2$ by induction. Using [10, Lemma 5.4 (ii)] combined with Lemma A.7, we deduce that

$$|S_i(y)|\leqslant d|S_{i-1}(y)|+\mathcal{C}\sqrt{d|S_{i-1}(y)|\log N}$$

with very high probability for all $y\in S_1\setminus S_1^\tau$ and all $i\leqslant r_\star$. Therefore, using the induction assumption, i.e. (3.60) for i, we obtain

$$\sum_{y\in S_1\setminus S_1^\tau}|S_i(y)|\leqslant \mathcal{C}\frac{\log N}{(\tau-1)^2}d^{i-1}+\mathcal{C}\sqrt{d\log N}\sum_{y\in S_1\setminus S_1^\tau}\sqrt{|S_{i-1}(y)|}$$

$$\leqslant \mathcal{C}\frac{\log N}{(\tau-1)^2}d^{i-1} + \mathcal{C}\sqrt{d\log N}|S_1 \setminus S_1^\tau|\Bigg(\sum_{y\in S_1\setminus S_1^\tau}\frac{|S_{i-1}(y)|}{|S_1\setminus S_1^\tau|}\Bigg)^{1/2}$$
$$\leqslant \mathcal{C}\frac{\log N}{(\tau-1)^2}d^{i-1} + \mathcal{C}\sqrt{d\log N}\frac{\log N}{d(\tau-1)^2}\sqrt{d^{i-1}}$$

with very high probability, where we used the concavity of $\sqrt{\cdot}$ in the second step, (3.59) and (3.60) for i in the last step. Since $\sqrt{d^i \log N} \leqslant d^{i/2+1} \leqslant d^i$ for $i \geqslant 2$ and the sequence $(d^{1-i/2})_{i\in\mathbb{N}}$ is summable, this proves (3.60) for $i+1$ with a constant $\mathcal{C}$ independent of i. This concludes the proof of Proposition 3.1. □

4. The Delocalized Phase

In this section we prove Theorem 1.8. In fact, we state and prove a more general result, Theorem 4.2 below, which immediately implies Theorem 1.8.

4.1. Local law. Theorem 4.2 is a *local law* for a general class of sparse random matrices of the form

$$M = H + f\mathbf{e}\mathbf{e}^*, \tag{4.1}$$

where $f \geqslant 0$ and $\mathbf{e} := N^{-1/2}(1,1,\dots,1)^*$. Here H is a Hermitian random matrix satisfying the following definition.

Definition 4.1. Let $0 < d < N$. A *sparse matrix* is a complex Hermitian $N \times N$ matrix $H = H^* \in \mathbb{C}^{N\times N}$ whose entries H_{ij} satisfy the following conditions.

(i) The upper-triangular entries $(H_{ij} : 1 \leqslant i \leqslant j \leqslant N)$ are independent.
(ii) We have $\mathbb{E}H_{ij} = 0$ and $\mathbb{E}|H_{ij}|^2 = (1 + O(\delta_{ij}))/N$ for all i, j.
(iii) Almost surely, $|H_{ij}| \leqslant Kd^{-1/2}$ for all i, j and some constant K.

It is easy to check that the set of matrices M defined as in (4.1) and Definition 4.1 contains those from Theorem 1.8 (see the proof of Theorem 1.8 below). From now on we suppose that $K = 1$ to simplify notation.

The local law for the matrix M established in Theorem 4.2 below provides control of the entries of the *Green function*

$$G(z) := \big(M - z\big)^{-1} \tag{4.2}$$

for z in the spectral domain

$$\mathbf{S} \equiv \mathbf{S}_{\kappa,L,N} = \mathcal{S}_\kappa \times [N^{-1+\kappa}, L] \tag{4.3}$$

for some constant $L \geqslant 1$. We also define the Stieltjes transform g of the *empirical spectral measure of* M given by

$$g(z) := \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\lambda_i(M) - z} = \frac{1}{N}\operatorname{Tr} G(z)\,. \tag{4.4}$$

The limiting behaviour of G and g is governed by the following deterministic quantities. Denote by $\mathbb{C}_+ := \{z \in \mathbb{C} : \operatorname{Im} z > 0\}$ the complex upper half-plane. For $z \in \mathbb{C}_+$ we define $m(z)$ as the Stieltjes transform of the semicircle law μ_1,

$$m(z) := \int \frac{\mu_1(\mathrm{d}u)}{u - z}, \qquad \mu_1(\mathrm{d}u) := \frac{1}{2\pi}\sqrt{(4-u^2)_+}\,\mathrm{d}u\,. \tag{4.5}$$

An elementary argument shows that $m(z)$ can be characterized as the unique solution m in $\mathbb{C}_+$ of the equation

$$\frac{1}{m(z)} = -z - m(z)\,. \tag{4.6}$$

For $\alpha \geqslant 0$ and $z \in \mathbb{C}_+$ we define

$$m_\alpha(z) := -\frac{1}{z + \alpha m(z)}\,, \tag{4.7}$$

so that $m_1 = m$ by (4.6). In Lemma A.3 below we show that m_α is bounded in the domain $\mathbf{S}$, with a bound depending only on κ.

For $x \in [N]$ we denote the square Euclidean norm of the xth row of H by

$$\beta_x := \sum_y |H_{xy}|^2\,, \tag{4.8}$$

which should be thought of as the normalized degree of x; see Remark 4.3 below.

Theorem 4.2 (Local law for M). *Fix $0 < \kappa \leqslant 1/2$ and $L \geqslant 1$. Let H be a sparse matrix as in Definition 4.1, define M as in (4.1) for some $0 \leqslant f \leqslant N^{\kappa/6}$, and define G and g as in (4.2) and (4.4) respectively. Then with very high probability, for d satisfying (1.18), for all $z \in \mathbf{S}$ we have*

$$\max_{x,y\in[N]} \left|G_{xy}(z) - \delta_{xy} m_{\beta_x}(z)\right| \leqslant \mathcal{C}\left(\frac{\log N}{d^2}\right)^{1/3}, \tag{4.9}$$

$$\left|g(z) - m(z)\right| \leqslant \mathcal{C}\left(\frac{\log N}{d^2}\right)^{1/3}. \tag{4.10}$$

Proof of Theorem 1.8. Under the assumptions of Theorem 1.8 we find that $M := A/\sqrt{d}$ is of the form (4.1) for some H and f satisfying the assumptions of Theorem 4.2. Now Theorem 1.8 is a well-known consequence of Theorem 4.2 and the boundedness of $m_\alpha(z)$ in (A.4) below. For the reader's convenience, we give the short proof. Denoting the eigenvalues of M by $(\lambda_i(M))_{i\in[N]}$ and the associated eigenvectors by $(\mathbf{w}_i(M))_{i\in[N]}$, setting $z = \lambda + \mathrm{i}\eta$ with $\eta = N^{-1+\kappa}$, by (4.9) and (A.4) we have with very high probability

$$\mathcal{C} \geqslant \operatorname{Im} G_{xx}(z) = \sum_{i\in[N]} \frac{\eta |\langle \mathbf{1}_x, \mathbf{w}_i(M)\rangle|^2}{\eta^2 + (\lambda - \lambda_i(M))^2} \geqslant \frac{1}{\eta}\,|\langle \mathbf{1}_x, \mathbf{w}\rangle|^2\,,$$

where in the last step we omitted all terms except i satisfying $\lambda_i(M) = \lambda$. The claim follows by renaming $\kappa \to \kappa/2$. (Here we used that Theorem 4.2 holds also for random $z \in \mathbf{S}$, as follows form a standard net argument; see e.g. [16, Remark 2.7].) □

Remark 4.3 (**Relation between α_x and β_x**). In the special case $M = d^{-1/2}A$ with A the adjacency matrix of $\mathbb{G}(N, d/N)$, we have

$$\beta_x = \frac{1}{d}\sum_y \left(A_{xy} - \frac{d}{N}\right)^2 = \alpha_x + O\left(\frac{d(1+\alpha_x)}{N}\right) = \alpha_x + \mathcal{O}\left(\frac{d+\log N}{N}\right)$$

with very high probability, by Lemma A.7.

By definition, $m_\alpha(z) \in \mathbb{C}_+$ for $z \in \mathbb{C}_+$, i.e. m_α is a Nevanlinna function, and $\lim_{z\to\infty} z m_\alpha(z) = -1$. By the integral representation theorem for Nevanlinna functions, we conclude that m_α is the Stieltjes transform of a Borel probability measure μ_α on $\mathbb{R}$,

$$m_\alpha(z) = \int \frac{\mu_\alpha(\mathrm{d}u)}{u - z}\,. \tag{4.11}$$

Theorem 4.2 implies that the spectral measure of M at a vertex x is approximately μ_{β_x} with very high probability.

Inverting the Stieltjes transform (4.11) and using the definitions (4.5) and (4.7), we find after a short calculation

$$\mu_\alpha(\mathrm{d}u) = g_\alpha(u)\,\mathrm{d}u + h_\alpha \delta_{s_\alpha}(\mathrm{d}u) + h_\alpha \delta_{-s_\alpha}(\mathrm{d}u)\,, \tag{4.12}$$

where

$$g_\alpha(u) := \frac{\alpha \mathbb{1}_{|u|<2}}{2\pi}\frac{\sqrt{4-u^2}}{(1-\alpha)u^2+\alpha^2}\,, \quad h_\alpha := \mathbb{1}_{\alpha>2}\frac{\alpha-2}{2\alpha-2} + \frac{\mathbb{1}_{\alpha=0}}{2}\,, \quad s_\alpha := \mathbb{1}_{\alpha>2}\Lambda(\alpha)\,.$$

The family $(\mu_\alpha)_{\alpha\geqslant 0}$ contains the semicircle law ($\alpha = 1$), the Kesten-McKay law of parameter d ($\alpha = d/(d-1)$), and the arcsine law ($\alpha = 2$). For rational $\alpha = p/q$, the measure $\mu_{p/q}$ can be interpreted as the spectral measure at the root of the infinite rooted (p, q)-regular tree, whose root has p children and all other vertices have q children. We refer to Appendix A.2 for more details. See Fig. 8 for an illustration of the measure μ_α.

Remark 4.4. Using a standard application the Helffer-Sjöstrand formula (see e.g. [16, Section 8 and Appendix C]), we deduce from Theorem 4.2 the following local law for the spectral measure. Denote by ϱ_x the spectral measure of M at vertex x. Under the assumptions of Theorem 4.2, with very high probability, for any inverval $I \subset \mathcal{S}_\kappa$, we have

$$\varrho_x(I) = \mu_{\beta_x}(I) + \mathcal{O}\left(|I|\left(\frac{\log N}{d^2}\right)^{1/3} + N^{\kappa-1}\right).$$

The error is smaller than the left-hand side provided that $|I| \geqslant \mathcal{C}N^{\kappa-1}$.

The remainder of this section is devoted to the proof of Theorem 4.2. For the rest of this section, we assume that M is as in Theorem 4.2. To simplify notation, we consistently omit the z-dependence from our notation in quantities that depend on $z \in \mathbf{S}$. Unless mentioned otherwise, from now on all statements are uniform in $z \in \mathbf{S}$.

For the proof of Theorem 4.2, it will be convenient to single out the generic constant $\mathcal{C}$ from (1.18) by introducing a new constant $\mathcal{D}$ and replacing (1.18) with

$$\mathcal{D}\sqrt{\log N} \leqslant d \leqslant (\log N)^{3/2}\,. \tag{4.13}$$

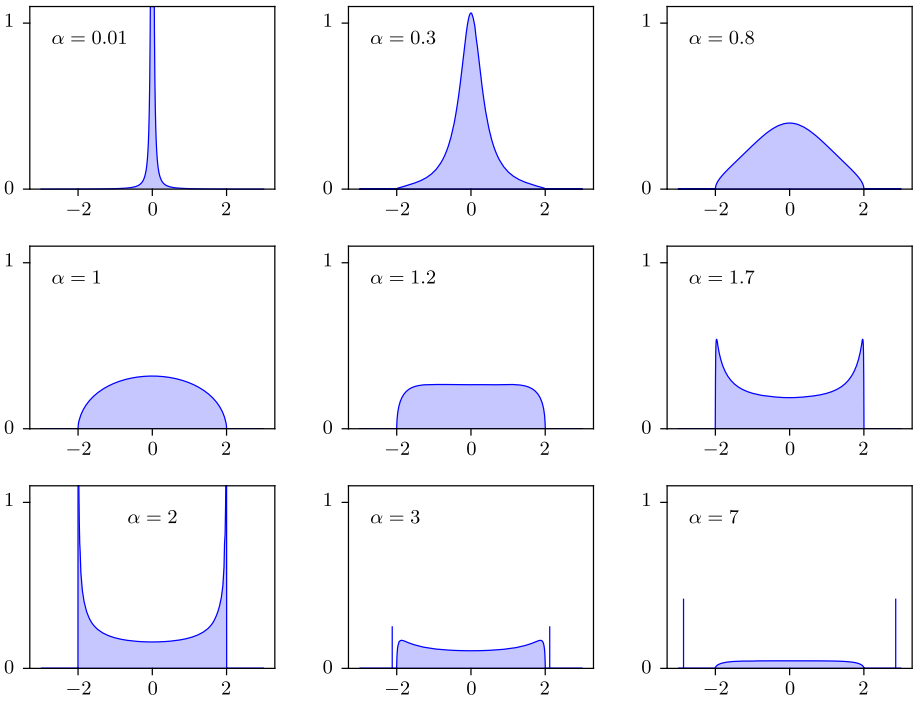

Fig. 8. An illustration of the probability measure μ_α for various values of α. For $\alpha > 2$, μ_α has two atoms which we draw using vertical lines. The measure μ_α is the semicircle law for $\alpha = 1$, the arcsine law for $\alpha = 2$, and the Kesten-McKay law with $d = \frac{\alpha}{\alpha-1}$ for $1 < \alpha < 2$. Note that the density of μ_α is bounded in $\mathcal{S}_\kappa$, uniformly in α. The divergence of the density near 0 is caused by values of α close to 0, and the divergence of the density near ± 2 by values of α close to 2

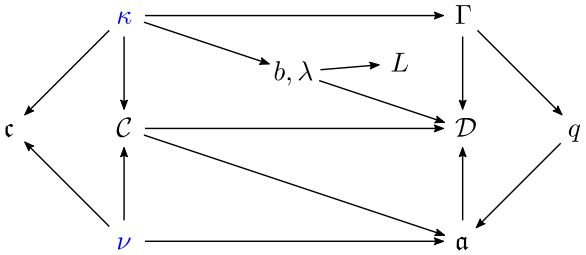

Fig. 9. The dependency graph of the various quantities appearing in the proof of Theorem 4.2. An arrow from x to y means that y is chosen as a function of x. The independent parameters, κ and ν, are highlighted in blue

Our proof will always assume that $\mathcal{C} \equiv \mathcal{C}_\nu$ and $\mathcal{D} \equiv \mathcal{D}_\nu$ are large enough, and the constant $\mathcal{C}$ in (1.18) can be taken to be $\mathcal{C} \vee \mathcal{D}$. For the rest of this section we assume that d satisfies (4.13) for some large enough $\mathcal{D}$, depending on κ and ν. To guide the reader through the proof, in Fig. 9 we include a diagram of the dependencies of the various quantities appearing throughout this section.

4.2. Typical vertices. We start by introducing the key tool in the proof of Theorem 4.2, a decomposition of vertices into *typical vertices* and the complementary *atypical vertices*.

Heuristically, a typical vertex x has close to d neighbours and the spectral measure of M at x is well approximated by the semicircle law. In fact, in order to be applicable to the proof of Proposition 4.18 below, the notion of a typical vertex is somewhat more complicated, and when counting the number of neighbours of a vertex x we also need to weight the neighbours with diagonal entries of a Green function, so that the notion of typical vertex also depends on the spectral parameter z, which in this subsection we allow to be any complex number z with $\operatorname{Im} z \geqslant N^{-1+\kappa}$. This notion is defined precisely using the parameters Φ_x and Ψ_x from (4.18) below. The main result of this subsection is Proposition 4.8 below, which states, in the language of graphs when $M = d^{-1/2}A$ with A the adjacency matrix of $\mathbb{G}(N, d/N)$, that most vertices are typical and most neighbours of any vertex are typical. To state it, we introduce some notation.

Definition 4.5. For any subset $T \subset [N]$, we define the minor $M^{(T)}$ with indices in T as the $(N - |T|) \times (N - |T|)$-matrix

$$M^{(T)} := (M_{xy})_{x,y\in[N]\setminus T}. \tag{4.14}$$

If T consists only of one or two elements, $T = \{x\}$ or $T = \{x, y\}$, then we abbreviate $M^{(x)}$ and $M^{(xy)}$ for $M^{(\{x\})}$ and $M^{(\{x,y\})}$. We also abbreviate $M^{(Tx)}$ for $M^{(T\cup\{x\})}$. The Green function of $M^{(T)}$ is denoted by

$$G^{(T)}(z) := (M^{(T)} - z)^{-1}. \tag{4.15}$$

We use the notation

$$\sum_{x}^{(T)} := \sum_{x\in[N]\setminus T}. \tag{4.16}$$

Definition 4.6 (*Typical vertices*). Let $\mathfrak{a} > 0$ be a constant, and define the set of *typical vertices*

$$\mathcal{T}_{\mathfrak{a}} := \{x \in [N] : |\Phi_x| \vee |\Psi_x| \leqslant \varphi_{\mathfrak{a}}\}, \qquad \varphi_{\mathfrak{a}} := \mathfrak{a}\left(\frac{\log N}{d^2}\right)^{1/3}, \tag{4.17}$$

where

$$\Phi_x := \sum_{y}^{(x)}\left(|H_{xy}|^2 - \frac{1}{N}\right), \qquad \Psi_x := \sum_{y}^{(x)}\left(|H_{xy}|^2 - \frac{1}{N}\right)G^{(x)}_{yy}. \tag{4.18}$$

Note that this notion depends on the spectral parameter z, i.e. $\mathcal{T}_{\mathfrak{a}} \equiv \mathcal{T}_{\mathfrak{a}}(z)$. The constant $\mathfrak{a}$ will depend only on ν and κ. It will be fixed in (4.23) below. The constant $\mathcal{D} \geqslant \mathfrak{a}^{3/2}$ from (4.13) is always chosen large enough so that $\varphi_{\mathfrak{a}} \leqslant 1$.

The following proposition holds on the event $\{\theta = 1\}$, where we introduce the indicator function

$$\theta := \mathbb{1}_{\max_{x,y}|G_{xy}|\leqslant\Gamma} \tag{4.19}$$

depending on some deterministic constant $\Gamma \geqslant 1$. In (4.40) below, we shall choose a constant $\Gamma \equiv \Gamma_\kappa$, depending only on κ, such that the condition $\theta = 1$ can be justified by a bootstrapping argument along the proof of Theorem 4.2 in Sect. 4.3 below.

Throughout the sequel we use the following generalization of Definition 2.1.

Definition 4.7. An event Ξ *holds with very high probability on an event* Ω if for all $\nu > 0$ there exists $C > 0$ such that $\mathbb{P}(\Xi \cap \Omega) \geqslant \mathbb{P}(\Omega) - CN^{-\nu}$ for all $N \in \mathbb{N}$.

We now state the main result of this subsection.

Proposition 4.8. *There are constants* $0 < q \leqslant 1$, *depending only on* Γ, *and* $\mathfrak{a} > 0$, *depending only on* ν *and* q, *such that, on the event* $\{\theta = 1\}$, *the following holds with very high probability.*

(i) *Most vertices are typical:*

$$|\mathcal{T}_{\mathfrak{a}}^c| \leqslant \exp(q\varphi_{\mathfrak{a}}^2 d) + N\exp(-2q\varphi_{\mathfrak{a}}^2 d).$$

(ii) *Most neighbours of any vertex are typical:*

$$\sum_{y \in \mathcal{T}_{\mathfrak{a}}^c}^{(x)} |H_{xy}|^2 \leqslant C\varphi_{\mathfrak{a}} + Cd^4 \exp(-q\varphi_{\mathfrak{a}}^2 d)$$

uniformly for $x \in [N]$.

For the interpretation of Proposition 4.8 (ii), one should think of the motivating example $M = d^{-1/2}A$, for which $d\sum_{y\in\mathcal{T}_{\mathfrak{a}}^c}^{(x)} |H_{xy}|^2$ is the number of atypical neighbours of x, up to an error term $\mathcal{O}\big(\frac{d^2+d\log N}{N}\big)$ by Remark 4.3.

The remainder of Sect. 4.2 is devoted to the proof of Proposition 4.8. We need the following version of $\mathcal{T}_{\mathfrak{a}}$ defined in terms of $H^{(T)}$ instead of H.

Definition 4.9. For any $x \in [N]$ and $T \subset [N]$, we define

$$\Phi_x^{(T)} := \sum_y^{(Tx)} \left(|H_{xy}|^2 - \frac{1}{N} \right), \qquad \Psi_x^{(T)} := \sum_y^{(Tx)} \left(|H_{xy}|^2 - \frac{1}{N} \right) G_{yy}^{(Tx)}$$

and

$$\mathcal{T}_{\mathfrak{a}}^{(T)} := \{x \in [N] \setminus T : |\Phi_x^{(T)}| \vee |\Psi_x^{(T)}| \leqslant \varphi_{\mathfrak{a}}\}.$$

Note that $\Phi_x^{(\emptyset)} = \Phi_x$ and $\Psi_x^{(\emptyset)} = \Psi_x$ with the definitions from (4.18), and hence $\mathcal{T}_{\mathfrak{a}}^{(\emptyset)} = \mathcal{T}_{\mathfrak{a}}$. The proof of Proposition 4.8 relies on the two following lemmas.

Lemma 4.10. *There are constants* $0 < q \leqslant 1$, *depending only on* Γ, *and* $\mathfrak{a} > 0$, *depending only on* ν *and* q, *such that, for any deterministic* $X \subset [N]$, *the following holds with very high probability on the event* $\{\theta = 1\}$.

(i) $|X \cap \mathcal{T}_{\mathfrak{a}/2}^c| \leqslant \exp(q\varphi_{\mathfrak{a}}^2 d) + |X|\exp(-2q\varphi_{\mathfrak{a}}^2 d)$.

(ii) *If* $|X| \leqslant \exp(2q\varphi_{\mathfrak{a}}^2 d)$ *then* $|X \cap \mathcal{T}_{\mathfrak{a}/2}^c| \leqslant \varphi_{\mathfrak{a}} d$.

For any deterministic $x \in [N]$, *the same estimates hold for* $\big(\mathcal{T}_{\mathfrak{a}/2}^{(x)}\big)^c$ *instead of* $\mathcal{T}_{\mathfrak{a}/2}^c$ *and a random set* $X \subset [N] \setminus \{x\}$ *that is independent of* $H^{(x)}$.

Lemma 4.11. *With very high probability, for any constant* $\mathfrak{a} > 0$ *we have*

$$\theta|\Phi_y - \Phi_y^{(x)}| \leqslant \varphi_{\mathfrak{a}/2}, \qquad \theta|\Psi_y - \Psi_y^{(x)}| \leqslant \varphi_{\mathfrak{a}/2}$$

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

Before proving Lemmas 4.10 and 4.11, we use them to establish Proposition 4.8.

Proof of Proposition 4.8. For (i), we choose $X = [N]$ in Lemma 4.10 (i), using that $\mathcal{T}_{\mathfrak{a}/2} \subset \mathcal{T}_{\mathfrak{a}}$.

We now turn to the proof of (ii). By Lemma 4.11, on the event $\{\theta = 1\}$ we have $\mathcal{T}_{\mathfrak{a}}^c \subset \big(\mathcal{T}_{\mathfrak{a}/2}^{(x)}\big)^c$ with very high probability and hence

$$\theta \sum_{y \in \mathcal{T}_{\mathfrak{a}}^c}^{(x)} |H_{xy}|^2 \leqslant \theta \sum_{y \in (\mathcal{T}_{\mathfrak{a}/2}^{(x)})^c}^{(x)} |H_{xy}|^2$$

with very high probability. Since $|H_{xy}|^2 \leqslant 1/d$ almost surely, we obtain the decomposition

$$\begin{aligned}
\sum_{y \in (\mathcal{T}_{\mathfrak{a}/2}^{(x)})^c}^{(x)} |H_{xy}|^2 &\leqslant \sum_{k=0}^{\log N} \sum_{y \in (\mathcal{T}_{\mathfrak{a}/2}^{(x)})^c}^{(x)} |H_{xy}|^2 \mathbb{1}_{d^{-k-2} \leqslant |H_{xy}|^2 \leqslant d^{-k-1}} + \frac{1}{N} \\
&\leqslant \sum_{k=0}^{\log N} \sum_{y \in (\mathcal{T}_{\mathfrak{a}/2}^{(x)})^c}^{(x)} d^{-k-1} \mathbb{1}_{|H_{xy}|^2 \geqslant d^{-k-2}} + \frac{1}{N} \qquad (4.20) \\
&= \sum_{k=0}^{\log N} d^{-k-1} |X_k \cap (\mathcal{T}_{\mathfrak{a}/2}^{(x)})^c| + \frac{1}{N},
\end{aligned}$$

where we defined

$$X_k := \big\{ y \neq x : |H_{xy}|^2 \geqslant d^{-k-2} \big\}.$$

Since $\sum_y^{(x)} |H_{xy}|^2 \leqslant \mathcal{C}d$ with very high probability by Definition 4.1 and Bennett's inequality, we conclude that

$$|X_k| \leqslant \mathcal{C} d^{k+3} \qquad (4.21)$$

with very high probability.

We shall apply Lemma 4.10 to the sets $X = X_k$ and $(\mathcal{T}_{\mathfrak{a}/2}^{(x)})^c$. To that end, note that $X_k \subset [N] \backslash \{x\}$ is a measurable function of the family $(H_{xy})_{y \in [N]}$, and hence independent of $H^{(x)}$. Thus, we may apply Lemma 4.10.

We define $K := \max\big\{k \geqslant 0 : \mathcal{C}d^{k+3} \leqslant \mathrm{e}^{2q\varphi_{\mathfrak{a}}^2 d}\big\}$ and decompose the sum on the right-hand side of (4.20) into

$$\begin{aligned}
\sum_{k=0}^{\log N} d^{-k-1} |X_k \cap \big(\mathcal{T}_{\mathfrak{a}/2}^{(x)}\big)^c| &= \sum_{k=0}^{K} d^{-k-1} |X_k \cap \big(\mathcal{T}_{\mathfrak{a}/2}^{(x)}\big)^c| + \sum_{k=K+1}^{\log N} d^{-k-1} |X_k \cap \big(\mathcal{T}_{\mathfrak{a}/2}^{(x)}\big)^c| \\
&\leqslant \sum_{k=0}^{K} d^{-k-1} \varphi_{\mathfrak{a}} d + \sum_{k=K+1}^{\log N} d^{-k-1} \big(\mathrm{e}^{q\varphi_{\mathfrak{a}}^2 d} + \mathcal{C} d^{k+3} \mathrm{e}^{-2q\varphi_{\mathfrak{a}}^2 d}\big) \\
&\leqslant 2\varphi_{\mathfrak{a}} + \mathcal{C} d^2 \mathrm{e}^{-q\varphi_{\mathfrak{a}}^2 d} \log N
\end{aligned}$$

with very high probability. Here, we used Lemma 4.10 (ii) to estimate the summands if $k \leqslant K$ and Lemma 4.10 (i) and (4.21) for the other summands. Since $\log N \leqslant d^2$, this concludes the proof of (ii). □

The rest of this subsection is devoted to the proofs of Lemmas 4.10 and 4.11. Let θ be defined as in (4.19) for some constant $\Gamma \geqslant 1$. For any subset $T \subset [N]$, we define the indicator function

$$\theta^{(T)} := \mathbb{1}_{\max_{a,b\notin T}|G^{(T)}_{ab}|\leqslant 2\Gamma}.$$

Lemma 4.10 is a direct consequence of the following two lemmas.

The first one, Lemma 4.12, is mainly a decoupling argument for the random variables $(\Psi_x)_{x\in[N]}$. Indeed, the probability that any fixed vertex x is atypical is only small, $o(1)$, and not very small, $N^{-\nu}$; see (4.31) below. If the events of different vertices being atypical were independent, we could deduce that the probability that a sufficiently large set of vertices are atypical is very small. However, these events are not independent. The most serious breach of independence arises from the Green function $G^{(x)}_{yy}$ in the definition of Ψ_x. In order to make this argument work, we have to replace the parameters Φ_x and Ψ_x with their *decoupled* versions $\Phi^{(T)}_x$ and $\Psi^{(T)}_x$ from Definition 4.9. To that end, we have to estimate the error involved, $|\Phi_x - \Phi^{(T)}_x|$ and $|\Psi_x - \Psi^{(T)}_x|$. Unfortunately the error bound on the latter is proportional to β_x (see (4.32)), which is not affordable for vertices of large degree. The solution to this issue involves the observation that if β_x is too large then the vertex is atypical by the condition on Φ_x, which allows us to disregard the size of Ψ_x. The details are given in the proof of Lemma 4.12 below.

The second one, Lemma 4.13, gives a priori bounds on the entries of the Green function $G^{(T)}$, which shows that if the entries of G are bounded then so are those of $G^{(T)}$ for $|T| = o(d)$. For T of fixed size, this fact is a standard application of the resolvent identities from Lemma A.24. For our purposes, it is crucial that T can have size up to $o(d)$, and such a quantitative estimate requires slightly more care.

Lemma 4.12. *There is a constant $0 < q \leqslant 1$, depending only on Γ, such that, for any $\nu > 0$, there is $\mathcal{C} > 0$ such that the following holds for any fixed $\mathfrak{a} > 0$. If $x \notin T \subset [N]$ are deterministic with $|T| \leqslant \varphi_{\mathfrak{a}} d/\mathcal{C}$ then*

$$\mathbb{P}\big(T \subset \mathcal{T}^c_{\mathfrak{a}/2},\ \theta = 1\big) \leqslant \mathrm{e}^{-4q\varphi_{\mathfrak{a}}^2 d|T|} + \mathcal{C}N^{-\nu}, \tag{4.22a}$$

$$\mathbb{P}\big(T \subset \big(\mathcal{T}^{(x)}_{\mathfrak{a}/2}\big)^c,\ \theta^{(x)} = 1\big) \leqslant \mathrm{e}^{-4q\varphi_{\mathfrak{a}}^2 d|T|} + \mathcal{C}N^{-\nu}. \tag{4.22b}$$

Lemma 4.13. *For any subset $T \subset [N]$ satisfying $|T| \leqslant \frac{d}{\mathcal{C}\Gamma^2}$ we have $\theta \leqslant \theta^{(T)}$ with very high probability.*

Before proving Lemma 4.12 and Lemma 4.13, we use them to show Lemma 4.10.

Proof of Lemma 4.10. Throughout the proof we abbreviate $\mathbb{P}_\theta(\Xi) := \mathbb{P}(\Xi \cap \{\theta = 1\})$. Let $\mathcal{C}$ be the constant from Lemma 4.12, and set

$$\mathfrak{a} := \left(\frac{\mathcal{C}\nu}{4q}\right)^{1/3}. \tag{4.23}$$

For the proof of (ii), we choose $k = \varphi_{\mathfrak{a}} d/\mathcal{C}$ and estimate

$$\mathbb{P}_\theta(|X \cap \mathcal{T}^c_{\mathfrak{a}/2}| \geqslant k) \leqslant \sum_{Y\subset X:|Y|=k} \mathbb{P}_\theta(Y \subset \mathcal{T}^c_{\mathfrak{a}/2}) \leqslant \binom{|X|}{k}\left(\mathrm{e}^{-4q\varphi_{\mathfrak{a}}^2 dk} + \mathcal{C}N^{-\nu}\right)$$
$$\leqslant \left(|X|\mathrm{e}^{-4q\varphi_{\mathfrak{a}}^2 d}\right)^k + \mathcal{C}|X|^k N^{-\nu} \leqslant \mathrm{e}^{-2q\varphi_{\mathfrak{a}}^2 dk} + \mathcal{C}\mathrm{e}^{2q\varphi_{\mathfrak{a}}^2 dk} N^{-\nu}$$

$$= N^{-2q\mathfrak{a}^3/\mathcal{C}} + \mathcal{C}N^{2q\mathfrak{a}^3/\mathcal{C}-\nu}.$$

where in the second step we used (4.22a). Thus, by our choice of $\mathfrak{a}$, we have $\mathbb{P}_\theta(|X \cap \mathcal{T}^c_{\mathfrak{a}/2}| \geqslant k) \leqslant (\mathcal{C}+1)N^{-\nu/2}$, from which (ii) follows after renaming ν and $\mathcal{C}$.

To prove (i) we estimate, for $t > 0$ and $l \in \mathbb{N}$,

$$\mathbb{P}_\theta(|X \cap \mathcal{T}^c_{\mathfrak{a}/2}| \geqslant t) \leqslant \frac{1}{t^l}\mathbb{E}\Big(\sum_{x\in X} \mathbb{1}_{x\in\mathcal{T}^c_{\mathfrak{a}/2}}\theta\Big)^l = \frac{1}{t^l}\sum_{x_1,\dots,x_l\in X} \mathbb{P}_\theta(x_1 \in \mathcal{T}^c_{\mathfrak{a}/2}, \dots, x_l \in \mathcal{T}^c_{\mathfrak{a}/2}).$$

Choosing $l = \varphi_\mathfrak{a} d/\mathcal{C}$, regrouping the summation according to the partition of coincidences, and using Lemma 4.12 yield

$$\begin{aligned}\mathbb{P}_\theta(|X \cap \mathcal{T}^c_{\mathfrak{a}/2}| \geqslant t) &\leqslant \frac{1}{t^l}\sum_{\pi\in\mathfrak{P}_l} |X|^{|\pi|}\big(\mathrm{e}^{-4q\varphi_\mathfrak{a}^2 d|\pi|} + \mathcal{C}N^{-\nu}\big)\\ &\leqslant \frac{1}{t^l}\sum_{k=0}^{l}\binom{l}{k} l^{l-k}|X|^k\big(\mathrm{e}^{-4q\varphi_\mathfrak{a}^2 dk} + \mathcal{C}N^{-\nu}\big) = \frac{(l+|X|\mathrm{e}^{-4q\varphi_\mathfrak{a}^2 d})^l + \mathcal{C}N^{-\nu}(l+|X|)^l}{t^l}.\end{aligned}$$

Here, $\mathfrak{P}_l$ denotes the set of partitions of $[l]$, and we denote by $k = |\pi|$ the number of blocks in the partition $\pi \in \mathfrak{P}_l$. We also used that the number of partitions of l elements consisting of k blocks is bounded by $\binom{l}{k} l^{l-k}$. The last step follows from the binomial theorem. Therefore, using $l = \varphi_\mathfrak{a} d/\mathcal{C}$ and choosing $t = \mathrm{e}^{q\varphi_\mathfrak{a}^2 d} + |X|\mathrm{e}^{-2q\varphi_\mathfrak{a}^2 d}$ as well as $\mathcal{C}$ and ν sufficiently large imply the bound in Lemma 4.10 (i) with very high probability, after renaming $\mathcal{C}$ and ν. Here we used (4.13).

To obtain the same statements for $\mathcal{T}^{(x)}_{\mathfrak{a}/2}$ instead of $\mathcal{T}_{\mathfrak{a}/2}$, we estimate

$$\mathbb{P}_\theta\Big(|X \cap (\mathcal{T}^{(x)}_{\mathfrak{a}/2})^c| \geqslant t\Big) \leqslant \mathbb{E}\Big[\mathbb{P}\Big(|X \cap (\mathcal{T}^{(x)}_{\mathfrak{a}/2})^c| \geqslant t, \theta^{(x)} = 1 \Big| X\Big)\Big] + \mathbb{P}\big(\theta^{(x)} = 0, \theta = 1\big).$$

For both parts, (i) and (ii), the conditional probability $\mathbb{P}\big(|X \cap (\mathcal{T}^{(x)}_{\mathfrak{a}/2})^c| \geqslant t, \theta^{(x)} = 1 \big| X\big)$ can be bounded as before using (4.22b) instead of (4.22a) since, by assumption on X, the set $\mathcal{T}^{(x)}_{\mathfrak{a}/2}$ and the indicator function $\theta^{(x)}$ are independent of X. The smallness of $\mathbb{P}(\theta^{(x)} = 0, \theta = 1) \leqslant \mathbb{P}(\theta^{(x)} < \theta)$ is a consequence of Lemma 4.13. This concludes the proof of Lemma 4.10. □

The rest of this subsection is devoted to the proofs of Lemmas 4.11, 4.12, and 4.13.

Lemma 4.14. *There is $\mathfrak{c} \equiv \mathfrak{c}_\nu > 0$, depending on ν and κ, such that for any deterministic $T \subset [N]$ satisfying $|T| \leqslant \mathfrak{c}d/\Gamma^2$ we have with very high probability*

$$\theta \max_{x,y\notin T} \big|G^{(T)}_{xy}\big| \leqslant 2\Gamma. \tag{4.24}$$

Moreover, under the same assumptions on T and for any $u \in [N] \setminus T$, we have

$$\theta \max_{x,y\notin T\cup\{u\}} \big|G^{(Tu)}_{xy} - G^{(T)}_{xy}\big| \leqslant \mathcal{C}d^{-1} \tag{4.25}$$

with very high probability.

Before proving Lemma 4.14, we use it to conclude the proof of Lemma 4.13.

Proof of Lemma 4.13. The bound in (4.24) of Lemma 4.14 implies that $\theta = \theta\theta^{(T)}$ with very high probability. Since $\theta \leqslant 1$, the proof is complete. □

Proof of Lemma 4.14. Throughout the proof we work on the event $\{\theta = 1\}$ exclusively. After a relabelling of the vertices $[N]$, we can suppose that $T = [k]$ with $k \leqslant cd/\Gamma^2$. For $k \in [N]$, we set

$$\Gamma_k := 1 \vee \max_{x,y \notin [k]} |G^{([k])}_{xy}| .$$

Note that $\Gamma_0 \leqslant \Gamma$ by definition of θ.

We now show by induction on k that there is $\mathcal{C} > 0$ such that

$$\Gamma_k \leqslant \Gamma_0 \left(1 + \frac{16\mathcal{C}\Gamma^2}{d}\right)^k \tag{4.26}$$

for all $k \in \mathbb{N}$ satisfying $k \leqslant \frac{d}{32\mathcal{C}\Gamma^2}$. Since $1+x \leqslant \mathrm{e}^x$, (4.26) implies that $\Gamma_k \leqslant \mathrm{e}^{1/2}\Gamma_0 \leqslant 2\Gamma$. This directly implies (4.24) by the definition of θ.

The initial step with $k = 0$ is trivially correct. For the induction step $k \to k+1$, we set $T = [k]$ and $u = k+1$. The algebraic starting point for the induction step is the identities (A.32a) and (A.32b). We shall need the following two estimates. First, from Lemma A.23 and Cauchy–Schwarz, we get

$$\frac{f}{N}\left|G^{(T)}_{uy}\sum_a^{(Tu)} G^{(Tu)}_{xa}\right| \leqslant \frac{f}{N}\Gamma_k\sqrt{\frac{N}{\operatorname{Im} z}}\Gamma_{k+1} \leqslant N^{-\kappa/3}\Gamma_k\Gamma_{k+1} , \tag{4.27}$$

where we used that $\Gamma_{k+1} \geqslant 1$, $f \leqslant N^{\kappa/6}$, and $\operatorname{Im} z \geqslant N^{-1+\kappa}$. Second, the first estimate of (A.28) in Corollary A.21 with $\psi = \Gamma_{k+1}/\sqrt{d}$ and $\gamma = \sqrt{\Gamma_{k+1}/(N \operatorname{Im} z)}$, Lemma A.23, and $\Gamma_{k+1} \geqslant 1$ imply

$$\left|\sum_a^{(Tu)} G^{(Tu)}_{xa} H_{au}\right| \leqslant \frac{\mathcal{C}}{\sqrt{d}}\Gamma_{k+1} \tag{4.28}$$

with very high probability.

Hence, owing to (A.32a) and (A.32b) with $T = [k]$ and $u = k+1$, we get, respectively,

$$\Gamma_{k+1} \leqslant \Gamma_k + \frac{\mathcal{C}}{\sqrt{d}}\Gamma_k\Gamma_{k+1}, \qquad \Gamma_{k+1} \leqslant \Gamma_k + \frac{\mathcal{C}}{d}\Gamma_k\Gamma_{k+1}^2 \tag{4.29}$$

with very high probability.

By the induction assumption (4.26) we have $\mathcal{C}\Gamma_k/\sqrt{d} \leqslant 2\mathcal{C}\Gamma/\sqrt{d} \leqslant 1/2$, so that the first inequality in (4.29) implies the rough a priori bound

$$\Gamma_{k+1} \leqslant 2\Gamma_k \tag{4.30}$$

with very high probability. From the second inequality in (4.29) and (4.30), we deduce that

$$\Gamma_{k+1} \leqslant \Gamma_k\left(1 + \frac{4\mathcal{C}}{d}\Gamma_k^2\right) \leqslant \Gamma_k\left(1 + \frac{16\mathcal{C}\Gamma^2}{d}\right),$$

where in the second step we used $\Gamma_k \leqslant 2\Gamma$, by the induction assumption (4.26). This concludes the proof of (4.26), and, hence, of (4.24).

For the proof of (4.25), we start from (A.32b) and use (4.27), (4.28) as well as (4.24). This concludes the proof of Lemma 4.14. □

The next result provides concentration estimates for the parameters Φ_x and Ψ_x.

Lemma 4.15. *There is a constant $0 < q \leqslant 1$, depending only on Γ, such that the following holds. Let $\mathfrak{c} > 0$ be as in Lemma 4.14, and let $x \in [N]$ and $T \subset [N]$ be deterministic and satisfy $|T| \leqslant \mathfrak{c}d/\Gamma^2$. Then for any $0 < \varepsilon \leqslant 1$ we have*

$$\theta^{(T)}\mathbb{P}\big(|\Phi_x^{(T)}| > \varepsilon \;\big|\; H^{(T)}\big) \leqslant \mathrm{e}^{-32q\varepsilon^2 d}\,, \quad \theta^{(T)}\mathbb{P}\big(|\Psi_x^{(T)}| > \varepsilon \;\big|\; H^{(T)}\big) \leqslant \mathrm{e}^{-32q\varepsilon^2 d}\,, \tag{4.31}$$

and, for any $u \notin T$,

$$\Phi_x^{(Tu)} - \Phi_x^{(T)} = O\left(\frac{1}{d}\right), \qquad \theta^{(T)}\big(\Psi_x^{(Tu)} - \Psi_x^{(T)}\big) = \mathcal{O}\left(\frac{1+\beta_x}{d}\right) \tag{4.32}$$

with very high probability.

Before proving Lemma 4.15, we use it conclude the proof of Lemma 4.11.

Proof of Lemma 4.11. Using (A.27b), we find that $\beta_x \leqslant \mathcal{C}(1 + \frac{\log N}{d})$ with very high probability. The claim now follows from (4.32) with $T = \emptyset$ and the definition of $\varphi_{\mathfrak{a}}$, choosing the constant $\mathcal{D}$ in (4.13) large enough. □

Proof of Lemma 4.15. Set $q := \frac{1}{2^{11}(\mathrm{e}\Gamma)^2}$. We get, using (A.27b) with $r := 32q\varepsilon^2 d \leqslant d$, $\mathbb{E}|H_{xy}|^2 = 1/N$, and Chebyshev's inequality,

$$\theta^{(T)}\mathbb{P}\Big(|\Psi_x^{(T)}| > \varepsilon \;\Big|\; H^{(T)}\Big) = \mathbb{P}\left(\theta^{(T)}\left|\sum_y^{(Tx)}(|H_{xy}|^2 - \mathbb{E}|H_{xy}|^2)G_{yy}^{(T)}\right| > \varepsilon \;\middle|\; H^{(T)}\right)$$
$$\leqslant \left(\frac{8\Gamma}{\varepsilon}\sqrt{\frac{r}{d}}\right)^r = \mathrm{e}^{-32q\varepsilon^2 d}$$

with very high probability for any $0 < \varepsilon \leqslant 1$. This proves the estimate on $\Psi_x^{(T)}$ in (4.31), and the estimate for $\Phi_x^{(T)}$ is proved similarly.

We now turn to the proof of (4.32). If $x = u$ then the statement is trivial. Thus, we assume $x \neq u$. In this case we have

$$\Phi_x^{(Tu)} - \Phi_x^{(T)} = -\left(|H_{xu}|^2 - \frac{1}{N}\right) \tag{4.33}$$

and the claim for Φ follows by Definition 4.1. Next,

$$\Psi_x^{(Tu)} - \Psi_x^{(T)} = \sum_y^{(Tux)}\left(|H_{xy}|^2 - \frac{1}{N}\right)\left(G_{yy}^{(Tux)} - G_{yy}^{(Tx)}\right) - \left(|H_{xu}|^2 - \frac{1}{N}\right)G_{uu}^{(Tx)}\,.$$

The last term multiplied by $\theta^{(T)}$ is estimated by $O(\Gamma/d)$ since $\theta^{(T)}|G_{uu}^{(Tx)}| \leqslant 4\Gamma$ by (4.30). We estimate the first term using (4.25) in Lemma 4.14, which yields

$$\theta^{(T)}\big|\Psi_x^{(Tu)} - \Psi_x^{(T)}\big| \leqslant \sum_y^{(Tux)}|H_{xy}|^2\frac{\mathcal{C}}{d} + \frac{1}{N}\sum_y^{(Tux)}\frac{\mathcal{C}}{d} + O\left(\frac{\Gamma}{d}\right) = \mathcal{O}\left(\frac{1+\beta_x}{d}\right)$$

with very high probability. This concludes the proof of Lemma 4.15. □

Proof of Lemma 4.12. Throughout the proof we abbreviate $\mathbb{P}_\theta(\Xi) := \mathbb{P}(\Xi \cap \{\theta = 1\})$. We have

$$\mathbb{P}\big(T \subset \mathcal{T}_{\mathfrak{a}/2}^c,\, \theta = 1\big) = \mathbb{P}_\theta\bigg(\bigcap_{x \in T} \Omega_x\bigg),$$

where we defined the event

$$\Omega_x := \big\{|\Phi_x| > \varphi_{\mathfrak{a}/2}\big\} \cup \big\{|\Psi_x| > \varphi_{\mathfrak{a}/2}\big\} = \big\{|\Phi_x| > \varphi_{\mathfrak{a}/2}\big\} \cup \big\{|\Phi_x| \leqslant \varphi_{\mathfrak{a}/2}, |\Psi_x| > \varphi_{\mathfrak{a}/2}\big\}.$$

We have the inclusions

$$\begin{aligned}\big\{|\Phi_x| > \varphi_{\mathfrak{a}/2}\big\} &\subset \big\{|\Phi_x^{(T)}| > \varphi_{\mathfrak{a}/4}\big\} \cup \big\{|\Phi_x - \Phi_x^{(T)}| > \varphi_{\mathfrak{a}/4}\big\},\\ \big\{|\Phi_x| \leqslant \varphi_{\mathfrak{a}/2}, |\Psi_x| > \varphi_{\mathfrak{a}/2}\big\} &\subset \big\{|\Psi_x^{(T)}| > \varphi_{\mathfrak{a}/4}\big\} \cup \big\{|\Phi_x| \leqslant \varphi_{\mathfrak{a}/2}, |\Psi_x - \Psi_x^{(T)}| > \varphi_{\mathfrak{a}/4}\big\}.\end{aligned}$$

Defining the event

$$\Omega_x^{(T)} := \big\{|\Phi_x^{(T)}| > \varphi_{\mathfrak{a}/4}\big\} \cup \big\{|\Psi_x^{(T)}| > \varphi_{\mathfrak{a}/4}\big\},$$

we therefore deduce by a union bound that

$$\begin{aligned}\mathbb{P}_\theta\bigg(\bigcap_{x \in T} \Omega_x\bigg) &\leqslant \mathbb{P}_\theta\bigg(\bigcap_{x \in T} \Omega_x^{(T)}\bigg) + \sum_{x \in T} \mathbb{P}_\theta\big(|\Phi_x - \Phi_x^{(T)}| > \varphi_{\mathfrak{a}/4}\big)\\ &\quad + \sum_{x \in T} \mathbb{P}_\theta\big(|\Phi_x| \leqslant \varphi_{\mathfrak{a}/2}, |\Psi_x - \Psi_x^{(T)}| > \varphi_{\mathfrak{a}/4}\big). \qquad (4.34)\end{aligned}$$

We begin by estimating the first term of (4.34). To that end, we observe that, conditioned on $H^{(T)}$, the family $(\Omega_x^{(T)})_{x \in T}$ is independent. Using Lemma 4.13 we therefore get

$$\mathbb{P}_\theta\bigg(\bigcap_{x \in T} \Omega_x^{(T)}\bigg) \leqslant \mathbb{E}\bigg[\theta^{(T)} \mathbb{P}\bigg(\bigcap_{x \in T} \Omega_x^{(T)} \;\bigg|\; H^{(T)}\bigg)\bigg] + \mathcal{C}N^{-\nu} = \mathbb{E}\bigg[\theta^{(T)} \prod_{x \in T} \mathbb{P}(\Omega_x^{(T)} | H^{(T)})\bigg] + \mathcal{C}N^{-\nu},$$

and we estimate each factor using (4.31) from Lemma 4.15 as

$$\begin{aligned}\theta^{(T)} \mathbb{P}(\Omega_x^{(T)} | H^{(T)}) &\leqslant \theta^{(T)} \mathbb{P}\big(|\Phi_x^{(T)}| > \varphi_{\mathfrak{a}/4} \;\big|\; H^{(T)}\big) + \theta^{(T)} \mathbb{P}\big(|\Psi_x^{(T)}| > \varphi_{\mathfrak{a}/4} \;\big|\; H^{(T)}\big)\\ &\leqslant 2\mathrm{e}^{-8q\varphi_{\mathfrak{a}}^2 d} \leqslant \mathrm{e}^{-4q\varphi_{\mathfrak{a}}^2 d},\end{aligned}$$

where in the last step we used that $\mathrm{e}^{-4q\varphi_{\mathfrak{a}}^2 d} \leqslant 1/2$. We conclude that

$$\mathbb{P}_\theta\bigg(\bigcap_{x \in T} \Omega_x^{(T)}\bigg) \leqslant \mathrm{e}^{-4q\varphi_{\mathfrak{a}}^2 d|T|} + \mathcal{C}N^{-\nu}.$$

Next, we estimate the second term of (4.34). After renaming the vertices, we may assume that $T = [k]$ with $k \leqslant \varphi_{\mathfrak{a}} d/\mathcal{C}$, so that we get from (4.32) from Lemma 4.15 (using that $\varphi_{\mathfrak{a}} d/\mathcal{C} \leqslant \mathfrak{c}d/\Gamma^2$ provided that $\mathcal{D}$ in (4.13) is chosen large enough, depending on $\mathfrak{a}$), by telescoping and recalling Lemma 4.13,

$$|\Phi_x - \Phi_x^{(T)}| \leqslant \sum_{i=0}^{k-1} \big|\Phi_x^{([i])} - \Phi_x^{([i+1])}\big| \leqslant O\bigg(\frac{k}{d}\bigg) \leqslant \varphi_{\mathfrak{a}/4} \qquad (4.35)$$

with very high probability on the event $\{\theta = 1\}$, if the constant $\mathcal{C}$ in the upper bound $\varphi_{\mathfrak{a}} d/\mathcal{C}$ on k is large enough.

The last term of (4.34) is estimated analogously, with the additional observation that, by definition of Φ_x and since $\varphi_{\mathfrak{a}/2} \leqslant 1/2$, on the event $\{|\Phi_x| \leqslant \varphi_{\mathfrak{a}/2}\}$ we have $\beta_x \leqslant 2$. Thus, on the event $\{\theta = 1\} \cap \{|\Phi_x| \leqslant \varphi_{\mathfrak{a}/2}\}$ we have, by Lemma 4.13,

$$|\Psi_x - \Psi_x^{(T)}| \leqslant \sum_{i=0}^{k-1} \left|\Psi_x^{([i])} - \Psi_x^{([i+1])}\right| \leqslant \mathcal{O}\left(\frac{k(1+\beta_x)}{d}\right) \leqslant \varphi_{\mathfrak{a}/4} \tag{4.36}$$

with very high probability, for large enough $\mathcal{C}$ in the upper bound on k. We conclude that the two last terms of (4.34) are bounded by $\mathcal{C}N^{-\nu}$, and the proof of (4.22a) is therefore complete.

The proof of (4.22b) is identical, replacing the matrix M with the matrix $M^{(x)}$. □

4.3. Self-consistent equation and proof of Theorem 4.2. In this subsection, we derive an approximate self-consistent equation for the Green function G, and use it to prove Theorem 4.2. The key ingredient is Proposition 4.18 below, which provides a bootstrapping bound stating that if $G_{xx} - m_{\beta_x}$ is smaller than some constant then it is in fact bounded by $\varphi_{\mathfrak{a}}$ with very high probability. It is proved by first deriving and solving a self-consistent equation for the entries G_{xx} indexed by typical vertices $x \in \mathcal{T}_{\mathfrak{a}}$, and using the obtained bounds to analyse G_{xx} for atypical vertices $x \in \mathcal{T}_{\mathfrak{a}}^c$.

We begin with a simple algebraic observation.

Lemma 4.16 (Approximate self-consistent equation). *For any $x \in [N]$ and $z \in \mathbb{C}_+$, we have*

$$\frac{1}{G_{xx}} = -z - \sum_{y}^{(x)} |H_{xy}|^2 G_{yy}^{(x)} + Y_x \,,$$

where we introduced the error term

$$Y_x := H_{xx} + \frac{f}{N} - \sum_{a \neq b}^{(x)} H_{xa} G_{ab}^{(x)} H_{bx} - \sum_{a,b}^{(x)} \left(\frac{f}{N} \left(H_{xa} G_{ab}^{(x)} + G_{ab}^{(x)} H_{bx} \right) + \frac{f^2}{N^2} G_{ab}^{(x)} \right). \tag{4.37}$$

Proof. The lemma follows directly from (A.31) and the definition (4.1). □

Let θ be defined as in (4.19) with some $\Gamma \geqslant 1$. The following lemma provides a priori bounds on the error terms appearing in the self-consistent equation.

Lemma 4.17. *For all $z \in \mathbb{C}$ with $\operatorname{Im} z \geqslant N^{-1+\kappa}$, with very high probability,*

$$\theta \max_x |Y_x| \leqslant \mathcal{C} d^{-1/2}, \tag{4.38a}$$

$$\theta \max_{x \neq y} |G_{xy}| \leqslant \mathcal{C} d^{-1/2}, \tag{4.38b}$$

$$\theta \max_{x \neq a \neq y} |G_{xy} - G_{xy}^{(a)}| \leqslant \mathcal{C} d^{-1}. \tag{4.38c}$$

Proof. We first estimate Y_x. From Definition 4.1, the upper bound on f, and (4.13), we conclude that $|H_{xx}| + f/N = O(d^{-1/2})$ almost surely. Moreover, the Cauchy–Schwarz inequality, Lemma A.23, (4.24) and the upper bound on f imply

$$\theta \frac{f^2}{N^2} \left| \sum_{a,b}^{(x)} G_{ab}^{(x)} \right| \leqslant C_\kappa \frac{f^2}{\sqrt{N \operatorname{Im} z}} \leqslant C_\kappa N^{-\kappa/6} \leqslant \frac{\mathcal{C}}{\sqrt{d}} ,$$

for some constant C_κ depending only on κ. Next, we use the first estimate of (A.28), Lemma A.23, and the upper bound on f to conclude that

$$\frac{f}{N}\theta \left| \sum_{a,b}^{(x)} H_{xa} G_{ab}^{(x)} \right| + \frac{f}{N}\theta \left| \sum_{a,b}^{(x)} G_{ab}^{(x)} H_{bx} \right| \leqslant \frac{\mathcal{C}}{\sqrt{d}} \frac{f}{\sqrt{N \operatorname{Im} z}} \leqslant \frac{\mathcal{C}}{\sqrt{d}} N^{-\kappa/3} \leqslant \frac{\mathcal{C}}{\sqrt{d}}$$

with very high probability (compare the proof of (4.28)). Moreover, from Lemma A.23 and the second estimate of (A.28) we deduce that remaining term in (4.37) is $\mathcal{O}(d^{-1}) = \mathcal{O}(d^{-1/2})$. This concludes the proof of (4.38a).

For the proof of (4.38b), we start from (A.29) and use $M_{xa} = H_{xa} + f/N$ to obtain

$$G_{xy} = -G_{xx} \sum_a^{(x)} H_{xa} G_{ay}^{(x)} - G_{xx} H_{xy} G_{yy}^{(x)} - \frac{f}{N} G_{xx} \sum_a^{(x)} G_{ay}^{(x)} .$$

Similar arguments as in (4.28) and (4.27) show that the first and third term, respectively, are bounded by $\mathcal{C} d^{-1/2}$ with very high probability. The same bound for the second term follows from Definition 4.1 and (4.24) in Lemma 4.14. This proves (4.38b).

Finally, (4.38c) follows directly from (4.25). □

Proposition 4.18 below is the main tool behind the proof of Theorem 4.2. To formulate it, we introduce the z-dependent random control parameters

$$\Lambda_{\mathrm{d}} := \max_x |G_{xx} - m_{\beta_x}| , \qquad \Lambda_{\mathrm{o}} := \max_{x \neq y} |G_{xy}| , \qquad \Lambda := \Lambda_{\mathrm{d}} \vee \Lambda_{\mathrm{o}} ,$$

and, for some constant $\lambda \leqslant 1$, the indicator function

$$\phi := \mathbb{1}_{\Lambda \leqslant \lambda} . \tag{4.39}$$

Proposition 4.18 below provides a strong bound on Λ provided the a priori condition $\phi = 1$ is satisfied. Each step of its proof is valid provided λ is chosen small enough depending on κ. Note that, owing to (A.4), there is a deterministic constant Γ, depending only on κ, such that, for all $z \in \mathbf{S}$, we have

$$\phi \max_{x,y} |G_{xy}| \leqslant \Gamma . \tag{4.40}$$

In particular, if Γ in the definition (4.19) of θ is chosen as in (4.40) then

$$\phi \leqslant \theta . \tag{4.41}$$

Proposition 4.18. *There exists $\lambda > 0$, depending only on κ, such that, for all $z \in \mathbf{S}$, with very high probability,*

$$\phi \Lambda \leqslant \mathcal{C} \varphi_{\mathfrak{a}} .$$

For the proof of Proposition 4.18, we employ the results of the previous subsections to show that the diagonal entries $(G_{xx})_{x\in\mathcal{T}_\mathfrak{a}}$ of the Green function of M at the typical vertices satisfy the approximate self-consistent equation (4.42) below. This is a perturbed version of the relation (4.6) for the Stieltjes transform m of the semicircle law, which holds for all $z \in \mathbb{C}_+$. The stability estimate, (4.43) below, then implies that G_{xx} and m are close for all $x \in \mathcal{T}_\mathfrak{a}$. From this we shall, in a second step, deduce that G_{xx} is close to m_{β_x} for all x; this steps includes also the atypical vertices.

The next lemma is a relatively standard stability estimate of self-consistent equations in random matrix theory (compare e.g. to [27, Lemma 3.5]). It is proved in Appendix A.9.

Lemma 4.19 (Stability of the self-consistent equation for m). *Let $\mathcal{X}$ be a finite set, $\kappa > 0$, and $z \in \mathbb{C}_+$ satisfy $|\mathrm{Re}\, z| \leqslant 2 - \kappa$. We assume that, for two vectors $(g_x)_{x\in\mathcal{X}}$, $(\varepsilon_x)_{x\in\mathcal{X}} \in \mathbb{C}^\mathcal{X}$, the identities*

$$\frac{1}{g_x} = -z - \frac{1}{|\mathcal{X}|}\sum_{y\in\mathcal{X}} g_y + \varepsilon_x \tag{4.42}$$

hold for all $x \in \mathcal{X}$. Then there are constants $b, C \in (0,\infty)$, depending only on κ, such that if $\max_{x\in\mathcal{X}}|g_x - m(z)| \leqslant b$ then

$$\max_{x\in\mathcal{X}}|g_x - m(z)| \leqslant C\max_{x\in\mathcal{X}}|\varepsilon_x|, \tag{4.43}$$

where $m(z)$ satisfies (4.6).

Proof of Proposition 4.18. Throughout the proof, we work on the event $\{\phi = 1\}$, which, by (4.41), is contained in the event $\{\theta = 1\}$. Fix $\mathfrak{a}$ as in Proposition 4.8. Throughout the proof we use that $d^{-1/2} \leqslant \varphi_\mathfrak{a}$ by the upper bound in (4.13). Owing to (4.38b), it suffices to estimate Λ_d. Let b be chosen as in Lemma 4.19, and set $\lambda := b/2$ in the definition (4.39) of ϕ.

For the analysis of G_{xx} we distinguish the two cases $x \in \mathcal{T}_\mathfrak{a}$ and $x \notin \mathcal{T}_\mathfrak{a}$.

If $x \in \mathcal{T}_\mathfrak{a}$ then we write using Lemma 4.16 and the definition (4.18) of Ψ_x that

$$\begin{aligned}\frac{1}{G_{xx}} &= -z - \sum_y^{(x)}|H_{xy}|^2 G_{yy}^{(x)} + Y_x = -z - \frac{1}{N}\sum_y^{(x)} G_{yy}^{(x)} + Y_x - \Psi_x \\ &= -z - \frac{1}{|\mathcal{T}_\mathfrak{a}|}\sum_{y\in\mathcal{T}_\mathfrak{a}} G_{yy} + \varepsilon_x\,,\end{aligned}$$

where the error term ε_x satisfies

$$|\varepsilon_x| = \mathcal{O}\left(d^{-1/2} + \frac{1}{N}\exp(q\varphi_\mathfrak{a}^2 d) + \exp(-2q\varphi_\mathfrak{a}^2 d) + \varphi_\mathfrak{a}\right) = \mathcal{O}(\varphi_\mathfrak{a}) \tag{4.44}$$

with very high probability. Here, in the first step of (4.44) we used (4.38a), (4.38c), Proposition 4.8 (i), and the bound on Ψ_x in the definition (4.17) of $\mathcal{T}_\mathfrak{a}$, and in the second step of (4.44) we used that $\varphi_\mathfrak{a}^2 d = \mathfrak{a}^2(\log N)^{2/3}d^{-1/3}$ and (4.13) imply $(\log N)^{1/6}/\mathcal{C} \leqslant \varphi_\mathfrak{a}^2 d \leqslant \mathcal{C}(\log N)^{1/2}$, which yields

$$\frac{1}{N}\exp(q\varphi_\mathfrak{a}^2 d) + \exp(-2q\varphi_\mathfrak{a}^2 d) \leqslant \mathcal{C}d^{-10} \leqslant \varphi_\mathfrak{a}\,. \tag{4.45}$$

Thus, for $(G_{xx})_{x\in\mathcal{T}_\mathfrak{a}}$ we get the self-consistent equation in (4.42) with $g_x = G_{xx}$ and $\mathcal{X} = \mathcal{T}_\mathfrak{a}$. Moreover, by the bound on Φ_x in the definition (4.17) of $\mathcal{T}_\mathfrak{a}$, we have $\beta_x = 1 + \mathcal{O}(\varphi_\mathfrak{a})$. Hence, by (A.5), the assumption $\phi = 1$ and $d \geqslant \mathcal{C}\sqrt{\log N}$, we find that

$$|G_{xx} - m| \leqslant |G_{xx} - m_{\beta_x}| + |m_{\beta_x} - m| \leqslant b\,,$$

choosing the constant $\mathcal{D}$ in (4.13) large enough that the right-hand side of (A.5), i.e. $C|\beta_x - 1|$, is bounded by $b/2$. Hence Lemma 4.19 is applicable and we obtain $|G_{xx} - m| = O(\max_{y\in\mathcal{T}_\mathfrak{a}}|\varepsilon_y|)$. Therefore, we obtain

$$|G_{xx} - m_{\beta_x}| \leqslant |G_{xx} - m| + |m - m_{\beta_x}| \leqslant \mathcal{C}\varphi_\mathfrak{a} \tag{4.46}$$

with very high probability. This concludes the proof in the case $x \in \mathcal{T}_\mathfrak{a}$.

What remains is the case $x \notin \mathcal{T}_\mathfrak{a}$. In that case, we obtain from Lemma 4.16 that

$$\frac{1}{G_{xx}} = -z - \sum_{y\in\mathcal{T}_\mathfrak{a}}^{(x)} |H_{xy}|^2 G_{yy}^{(x)} - \sum_{y\in\mathcal{T}_\mathfrak{a}^c}^{(x)} |H_{xy}|^2 G_{yy}^{(x)} + Y_x = -z - \beta_x m + \varepsilon_x\,, \tag{4.47}$$

where the error term ε_x satisfies $\varepsilon_x = \mathcal{O}((1+\beta_x)\varphi_\mathfrak{a})$ with very high probability. Here we used (4.38a) as well as (4.38c), (4.45), (4.46) and Proposition 4.8 (ii) twice to conclude that

$$\sum_{y\in\mathcal{T}_\mathfrak{a}}^{(x)} |H_{xy}|^2 G_{yy}^{(x)} = \beta_x m + \mathcal{O}(\beta_x\varphi_\mathfrak{a})\,,$$

$$\sum_{y\in\mathcal{T}_\mathfrak{a}^c}^{(x)} |H_{xy}|^2 G_{yy}^{(x)} = \mathcal{O}\big(\varphi_\mathfrak{a} + d^4\exp(-q\varphi_\mathfrak{a}^2 d)\big) = \mathcal{O}(\varphi_\mathfrak{a})$$

with very high probability. From (4.7) and (4.47) we therefore get

$$G_{xx} - m_{\beta_x} = -m_{\beta_x}\frac{1}{-z - \beta_x m + \varepsilon_x}\,\varepsilon_x\,. \tag{4.48}$$

To estimate the right-hand side of (4.48), we consider the cases $\beta_x \leqslant 1$ and $\beta_x > 1$ separately.

If $\beta_x \leqslant 1$ then, by (A.4), the first factor of (4.48) is bounded by C. Thus, by (4.7), the second factor is bounded by $2C$ provided that $|\varepsilon_x| \leqslant 1/2C$ by choosing $\mathcal{D}$ in (4.13) large enough, and the third factor is bounded by $\mathcal{C}\varphi_\mathfrak{a}$. This yields the claim.

If $\beta_x > 1$, we use that $\operatorname{Im} m \geqslant c$ for some constant $c > 0$ depending only on κ and L. Thus, the right-hand side of (4.48) is bounded in absolute value, again using (A.4), by $C\frac{1}{\beta_x c/2}\mathcal{C}\beta_x\varphi_\mathfrak{a}$, provided that $\mathcal{D}$ in (4.13) is chosen large enough. This yields the claim. □

Proof of Theorem 4.2. After possibly increasing L, we can assume that L in the definition of $\mathbf{S}$ in (4.3) satisfies $L \geqslant 2/\lambda + 1$, where λ is chosen as in Proposition 4.18.

We first show that (4.10) follows from (4.9). Indeed, averaging the estimate on $|G_{xx} - m_{\beta_x}|$ in (4.9) over $x \in [N]$, using that $m_{\beta_x} = m + O(\varphi_\mathfrak{a})$ for $x \in \mathcal{T}_\mathfrak{a}$ by (A.5) and estimating the summands in $\mathcal{T}_\mathfrak{a}^c$ by Proposition 4.8 (i) and (A.4) yield (4.10) due to (4.45).

What remains is the proof of (4.9). Let $z_0 \in \mathbf{S}$, set $J := \min\{j \in \mathbb{N}_0 : \operatorname{Im} z_0 + jN^{-3} \geqslant 2/\lambda\}$, and define $z_j := z_0 + \mathrm{i}jN^{-3}$ for $j \in [J]$. We shall prove the bound in (4.9) at $z = z_j$ by induction on j, starting from $j = J$ and going down to $j = 0$. Since $|G_{xy}(z)| \leqslant (\operatorname{Im} z)^{-1}$ and $|m_{\beta_x}(z)| \leqslant (\operatorname{Im} z)^{-1}$ for all $x, y \in [N]$, we have $\max_x |G_{xx}(z_J) - m_{\beta_x}(z_J)| \leqslant \lambda$ and $\phi(z_J) = 1$.

For the induction step $j \to j-1$, suppose that $\phi(z_j) = 1$ with very high probability. Then, by Proposition 4.18, we deduce that $\Lambda(z_j) \leqslant \mathcal{C}\varphi_{\mathfrak{a}}$ with very high probability. Since G_{xy} and m_{β_x} are Lipschitz-continuous on $\mathbf{S}$ with constant N^2, we conclude that $\Lambda(z_{j-1}) \leqslant \mathcal{C}\varphi_{\mathfrak{a}} + N^{-1}$ with very high probability. If N is sufficiently large and $\varphi_{\mathfrak{a}}$ is sufficiently small, obtained by choosing $\mathcal{D}$ in (4.13) large enough, then we deduce that $\Lambda(z_{j-1}) \leqslant \lambda$ with very high probability and hence $\phi(z_{j-1}) = 1$ with very high probability. Using Proposition 4.18, this concludes the induction step, and hence establishes $\Lambda(z_0) \leqslant \mathcal{C}\varphi_{\mathfrak{a}}$ with very high probability. Here we used that the intersection of J events of very high probability is an event of very high probability, since $J \leqslant CN^3$, where C depends on κ. □

Acknowledgements The authors would like to thank Simone Warzel for helpful discussions. The authors 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 NCCR SwissMAP grant.

Funding Open Access funding provided by Université de Genève.

A. Appendices

In the following appendices we collect various tools and explanations used throughout the paper.

A.1. Simulation of the ℓ^∞-norms of eigenvectors. In Fig. 10 we depict a simulation of the ℓ^∞-norms of the eigenvectors of the adjacency matrix $A/\sqrt{d}$ of the Erdős–Rényi graph $\mathbb{G}(N, d/N)$ restricted to its giant component. We take $d = b \log N$ with $N = 10'000$ and $b = 0.6$. The eigenvalues and eigenvectors are drawn using a scatter plot, where the horizontal coordinate is the eigenvalue and the vertical coordinate the ℓ^∞-norm of the associated eigenvector. The higher a dot is located, the more localized the associated eigenvector is. Complete delocalization corresponds to a vertical coordinate ≈ 0.01, and localization at a single site to a vertical coordinate 1. Note the semilocalization near the origin and outside of $[-2, 2]$. The two semilocalized blips around ± 0.4 are a finite-N effect and tend to 0 as N is increased. The Perron–Frobenius eigenvalue is an outlier near 2.8 with delocalized eigenvector.

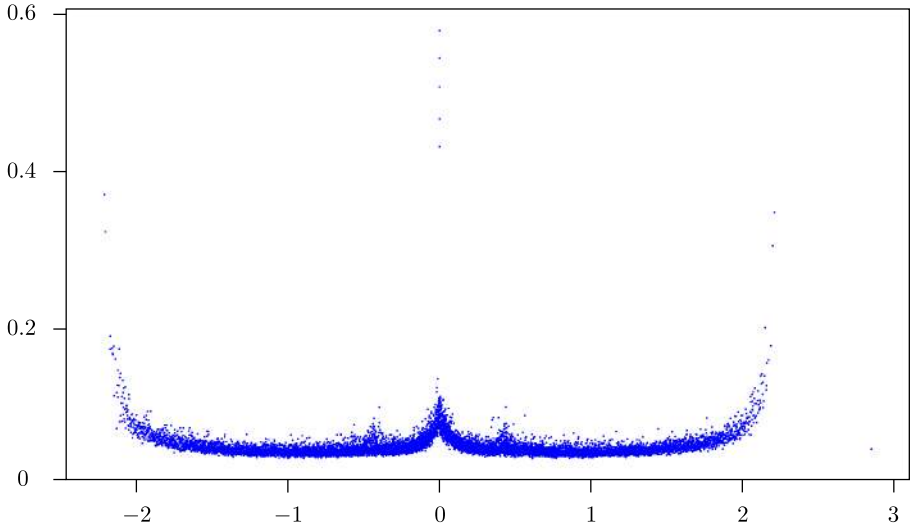

Fig. 10. A scatter plot of $(\lambda, \|\mathbf{w}\|_\infty)$ for all eigenvalue-eigenvector pairs $(\lambda, \mathbf{w})$ of the adjacency matrix $A/\sqrt{d}$ of the critical Erdős–Rényi graph restricted to its giant component, where $N = 10'000$ and $d = 0.6 \log N$

A.2. Spectral analysis of the infinite rooted (p, q)-regular tree. In this appendix we describe the spectrum, eigenvectors, and spectral measure of the following simple graph.

Definition A.1. For $p, q \in \mathbb{N}^*$ we define $\mathbb{T}_{p,q}$ as *the infinite rooted (p, q)-regular tree*, whose root has p children and all other vertices have q children.

A convenient way to analyse the adjacency matrix of $\mathbb{T}_{p,q}$ is by tridiagonalizing it around its root. To that end, we first review the tridiagonalization[5] of a general symmetric matrix $X \in \mathbb{R}^{N\times N}$ around a vertex $x \in [N]$; we refer to [10, Appendices A–C] for details. Let $r \in \mathbb{N}$ and $x \in [N]$. Suppose that the vectors $\mathbf{1}_x, X\mathbf{1}_x, X^2\mathbf{1}_x, \ldots, X^r\mathbf{1}_x$ are linearly independent, and denote by $\mathbf{g}_0, \mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_r$ the associated orthonormalized sequence. Then the tridiagonalization of X around x up to radius r is the $(r + 1) \times (r + 1)$ matrix $Z = (Z_{ij})_{i,j=0}^r$ with $Z_{ij} := \langle \mathbf{g}_i , X\mathbf{g}_j \rangle$. By construction, Z is tridiagonal and conjugate to X restricted to the subspace $\mathrm{Span}\{\mathbf{g}_0, \mathbf{g}_1, \ldots, \mathbf{g}_r\}$.
Let now $X = A \equiv A^{\mathbb{T}_{p,q}}$ be the adjacency matrix of $\mathbb{T}_{p,q}$, whose root we denote by o. Then it is easy to see that $\mathbf{g}_i = \mathbf{1}_{S_i(o)}/\|\mathbf{1}_{S_i(o)}\|$ and the tridiagonalization of A around the root up to radius ∞ is the infinite matrix $\sqrt{q}Z(p/q)$, where

$$Z(\alpha) := \begin{pmatrix} 0 & \sqrt{\alpha} & & & \\ \sqrt{\alpha} & 0 & 1 & & \\ & 1 & 0 & 1 & \\ & & 1 & 0 & \ddots \\ & & & \ddots & \ddots \end{pmatrix}. \tag{A.1}$$

[5] The tridiagonalization algorithm that we use is the Lanczos algorithm. Tridiagonalizing matrices in numerical analysis and random matrix theory [26,55] is usually performed using the numerically more stable Householder algorithm. However, when applied to the adjacency matrix $X = A$ of a graph, the Lanczos algorithm is more convenient because it can exploit the sparseness and local geometry of A.

If $\alpha > 2$, a transfer matrix analysis (see [10, Appendix C]) shows that $Z(\alpha)$ has precisely two eigenvalues in $\mathbb{R} \setminus [-2, 2]$, which are $\pm\Lambda(\alpha)$. The associated eigenvectors are $((\pm)^i u_i)_{i\in\mathbb{N}}$, where $u_0 > 0$ and $u_i := \frac{\sqrt{\alpha}}{(\alpha-1)^{i/2}} u_0$ for $i \geqslant 1$. Note that the eigenvector components are exponentially decaying since $\alpha > 2$, and hence u_0 can be chosen so that the eigenvectors are normalized. Going back to the original vertex basis of $\mathbb{T}_{p,q}$, setting $\alpha = p/q$, we conclude that the adjacency matrix A has eigenvalues $\pm\sqrt{q}\Lambda(\alpha)$ with associated eigenvectors $\sum_{i\in\mathbb{N}}(\pm)^i u_i \mathbf{1}_{S_i(o)}/\|\mathbf{1}_{S_i(o)}\|$.
Next, we show that the measure μ_α from (4.12) is the spectral measure at the root of $A^{\mathbb{T}_{p,q}}/\sqrt{d}$ and the spectral measure at 0 of (A.1).

Lemma A.2. (i) *For any $\alpha \geqslant 0$ the measure μ_α is the spectral measure of $Z(\alpha)$ at 0.*
(ii) *For any $p, q \in \mathbb{N}^*$ the measure $\mu_{p/q}$ is the spectral measure of the normalized adjacency operator $A^{\mathbb{T}_{p,q}}/\sqrt{q}$ at the root.*

Proof. For (i), define the vector $\mathbf{e}_0 = (1, 0, 0, \dots) \in \ell^2(\mathbb{N})$. The spectral measure of $Z(\alpha)$ with respect to $\mathbf{e}_0$ is characterized by its Stieltjes transform

$$\langle \mathbf{e}_0, (Z(\alpha) - z)^{-1}\mathbf{e}_0\rangle = \frac{1}{-z - \alpha\langle \mathbf{e}_0, (Z(1) - z)^{-1}\mathbf{e}_0\rangle}. \tag{A.2}$$

Here, we used Schur's complement formula on the Green function $(Z(\alpha) - z)^{-1}$, observing that the minor of $Z(\alpha)$ obtained by removing the zeroth row and column is $Z(1)$. Setting $\alpha = 1$ in (A.2) and recalling the defining relation (4.6) of the Stieltjes transform m of the semicircle law, we conclude that $\langle \mathbf{e}_0, (Z(1) - z)^{-1}\mathbf{e}_0\rangle = m(z)$ and hence from (4.7) and (A.2) we get $\langle \mathbf{e}_0, (Z(\alpha) - z)^{-1}\mathbf{e}_0\rangle = m_\alpha(z)$, as desired.
The proof of (ii) is analogous. Denote the root of $\mathbb{T}_{p,q}$ by o. Again using Schur's complement formula to remove the oth row and column of $H = A^{\mathbb{T}_{p,q}}/\sqrt{q}$, we deduce that

$$\langle \mathbf{1}_o, \left(A^{\mathbb{T}_{p,q}}/\sqrt{q} - z\right)^{-1}\mathbf{1}_o\rangle = \left(-z - \frac{p}{q}\langle \mathbf{1}_o, \left(A^{\mathbb{T}_{q,q}}/\sqrt{q} - z\right)^{-1}\mathbf{1}_o\rangle\right)^{-1}, \tag{A.3}$$

where we used that $\mathbb{T}_{p,q}$ from which o has been removed consists of p disconnected copies of $\mathbb{T}_{q,q}$. Setting $p = q$ in (A.3) and comparing to (4.6) implies that the left-hand side of (A.3) is equal to $m(z)$ if $p = q$, and hence (ii) for general p follows from (4.7). Finally, we remark that the equality of the spectral measures of $Z(p/q)$ and $A^{\mathbb{T}_{p,q}}/\sqrt{q}$ can also be seen directly, by noting that $Z(p/q)$ is the tridiagonalization of $A^{\mathbb{T}_{p,q}}/\sqrt{q}$ around the root o. □

We conclude with some basic estimates for the Stieltjes transform m_α of μ_α used in Sect. 4.

Lemma A.3. *For each $\kappa > 0$ there is a constant $C > 0$ depending only on κ such that for all $z \in \mathbf{S}$ and all $\alpha \geqslant 0$ we have*

$$|m_\alpha(z)| \leqslant C, \tag{A.4}$$

$$|m_\alpha(z) - m(z)| \leqslant C|\alpha - 1|. \tag{A.5}$$

Proof. The simple facts follow directly from the corresponding properties of the semicircle law and its Stieltjes transform m (see e.g. [16, Lemma 3.3]). We leave the details to the reader. □

A.3. Bounds on adjacency matrices of trees. In this appendix we derive estimates on the operator norm of a tree. We start with a standard estimate on the operator norm of a graph.

Lemma A.4. *Let $\mathbb{T}$ be a graph whose vertices have degree at most $q+1$ for some $q \geqslant 1$. Then $\|A^{\mathbb{T}}\| \leqslant q+1$ and if in addition $\mathbb{T}$ is a tree then $\|A^{\mathbb{T}}\| \leqslant 2\sqrt{q}$.*

Proof. The first claim is obvious by the Schur test for the operator norm. To prove the second claim, choose a root o and denote by C_x the set of children of the vertex x. Then for any vector $\mathbf{w} = (w_x)$ we have

$$\begin{aligned} \left|\langle \mathbf{w}, A^{\mathbb{T}}\mathbf{w}\rangle\right| &= \left|\sum_{x,y} w_x A^{\mathbb{T}}_{xy} w_y\right| = 2\left|\sum_x \sum_{y\in C_x} w_x w_y\right| \leqslant \sum_x \sum_{y\in C_x}\left(\frac{1}{\sqrt{q}} w_x^2 + \sqrt{q} w_y^2\right) \\ &\leqslant \frac{q+1}{\sqrt{q}} w_o^2 + \sum_{x\neq o}\left(\frac{q}{\sqrt{q}} w_x^2 + \sqrt{q} w_x^2\right) \leqslant 2\sqrt{q}\sum_x w_x^2 , \end{aligned}$$

where in third step we used Young's inequality and in the fourth step that each vertex in the sum appears once as a child and at most q times as a parent. This concludes the proof. □

The same proof shows that if $\mathbb{T}$ is a rooted tree whose root has at most p children and all other vertices at most q children, then $\|A^{\mathbb{T}}\| \leqslant \sqrt{q}(p/q \vee 2)$. This bound is sharp for $p \leqslant 2q$ but not for $p > 2q$. The sharp bound in the latter case is established in the following result.

Lemma A.5. *Let $p, q \in \mathbb{N}^*$. Let $\mathbb{T}$ be a tree whose root has p children and all the other vertices have at most q children. Then the adjacency matrix $A^{\mathbb{T}}$ of $\mathbb{T}$ satisfies $\|A^{\mathbb{T}}\| \leqslant \sqrt{q}\Lambda(p/q \vee 2)$.*

Proof. Let $r \in \mathbb{N}$ and denote by $\mathbb{T}_{p,q}(r)$ the rooted (p,q)-regular tree of depth r, whose root x has p children, all vertices at distance $1 \leqslant i \leqslant r$ from x have q children, and all vertices at distance $r+1$ from x are leaves. For large enough r, we can exhibit $\mathbb{T}$ as a subgraph of $\mathbb{T}_{p,q}(r)$. By the Perron–Frobenius theorem,

$$\|A^{\mathbb{T}}\| = \langle \mathbf{w}, A^{\mathbb{T}}\mathbf{w}\rangle \tag{A.6}$$

for the some normalized eigenvector $\mathbf{w}$ whose entries are nonnegative. We extend $\mathbf{w}$ to a vector indexed by the vertex set of $\mathbb{T}_{p,q}(r)$ by setting $w_y = 0$ for y not in the vertex set of $\mathbb{T}$. Clearly,

$$\langle \mathbf{w}, A^{\mathbb{T}}\mathbf{w}\rangle \leqslant \langle \mathbf{w}, A^{\mathbb{T}_{p,q}(r)}\mathbf{w}\rangle . \tag{A.7}$$

Abbreviating $A \equiv A^{\mathbb{T}_{p,q}(r)}$, it therefore remains to estimate the right-hand side of (A.7) for large enough r. To that end, we define Z as the tridiagonalization of A around the root up to radius r (see Appendix A.2). The associated orthonormal set $\mathbf{g}_0, \mathbf{g}_1, \ldots, \mathbf{g}_r$ is given by $\mathbf{g}_i = \mathbf{1}_{S_i(x)}/\|\mathbf{1}_{S_i(x)}\|$, and $Z = \sqrt{q}Z_r(p/q)$, where $Z_r(\alpha)$ is the upper-left $(r+1)\times(r+1)$ block of (A.1). We introduce the orthogonal projections $P_0 := \mathbf{g}_0\mathbf{g}_0^*$ and $P := \sum_{i=0}^r \mathbf{g}_i\mathbf{g}_i^*$. Clearly, $P_0 P = P_0$ and hence $(1-P)(1-P_0) = 1-P$. For large enough r the vectors $\mathbf{g}_r$ and $\mathbf{w}$ have disjoint support, and hence $(1-P)AP\mathbf{w} = (1-P)A\sum_{i=0}^{r-1}\mathbf{g}_i\langle \mathbf{g}_i, \mathbf{w}\rangle = 0$, since $A\mathbf{g}_i \subset \mathrm{Span}\{\mathbf{g}_{i-i}, \mathbf{g}_{i+1}\}$ for $i < r$. Thus we have

$$\langle \mathbf{w}, A\mathbf{w}\rangle = \langle \mathbf{w}, PAP\mathbf{w}\rangle + \langle \mathbf{w}, (1-P)A(1-P)\mathbf{w}\rangle$$

$$= \langle \mathbf{w}, PAP\mathbf{w}\rangle + \langle \mathbf{w}, (1-P)(1-P_0)A(1-P_0)(1-P)\mathbf{w}\rangle . \tag{A.8}$$

From [10, Appendices B and C] we find

$$\lim_{r\to\infty} \|PAP\| = \lim_{r\to\infty} \|Z\| = \sqrt{q}\Lambda(p/q \vee 2) . \tag{A.9}$$

Moreover, the operator $(1-P_0)A(1-P_0)$ is the adjacency matrix of a forest whose vertices have degree at most q. By Lemma A.4, we therefore obtain $\|(1-P_0)A(1-P_0)\| \leqslant 2\sqrt{q}$. From (A.8) we therefore get

$$\limsup_{r\to\infty} \langle \mathbf{w}, A\mathbf{w}\rangle \leqslant \sqrt{q}\Lambda(p/q\vee 2)\|P\mathbf{w}\|^2 + 2\sqrt{q}\|(1-P)\mathbf{w}\|^2 \leqslant \sqrt{q}\Lambda(p/q\vee 2)\|\mathbf{w}\|^2 .$$

By (A.6) and (A.7), the proof is complete. □

A.4. Degree distribution and number of resonant vertices. In this appendix we record some basic facts about the distribution of degrees of the graph $\mathbb{G}(N, d/N)$, and use them to estimate the number of resonant vertices $\mathcal{W}_{\lambda,\delta}$.
The following is a quantitative version of the Poisson approximation of a binomial random variable.

Lemma A.6. (Poisson approximation) *If D is a random variable with law* $\mathrm{Binom}(n, p)$ *then for $k \leqslant \sqrt{n}$ and $p \leqslant 1/\sqrt{n}$ we have*

$$\mathbb{P}(D = k) = \frac{(pn)^k}{k!}\mathrm{e}^{-pn}\left(1 + O\left(\frac{k^2}{n} + p^2 n\right)\right).$$

Proof. Plugging the estimates $(1-p)^{n-k} = \mathrm{e}^{(n-k)\log(1-p)} = \mathrm{e}^{-np+O(pk+p^2n)}$ and

$$\frac{n!}{(n-k)!} = n^k \prod_{i=0}^{k-1}\left(1 - \frac{i}{n}\right) = n^k \mathrm{e}^{\sum_{i=0}^{k-1}\log\left(1-\frac{i}{n}\right)} = n^k \mathrm{e}^{O\left(\frac{k^2}{n}\right)},$$

into $\mathbb{P}(D_x = k) = \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$ yields the claim, since $pk \leqslant k^2/n + p^2 n$. □

Lemma A.7. *For $\mathbb{G}(N, d/N)$ we have $\alpha_x \leqslant \mathcal{C}\left(1 + \frac{\log N}{d}\right)$ with very high probability.*

Proof. This is a simple application of Bennett's inequality; see [10, Lemma 3.3] for details. □

Next, we recall some standard facts about the distribution of the degrees. Define the function $f_d : [1, \infty) \to \left[\frac{1}{2}\log(2\pi d), \infty\right)$ through

$$f_d(\alpha) := d(\alpha\log\alpha - \alpha + 1) + \frac{1}{2}\log(2\pi\alpha d) , \tag{A.10}$$

which is bijective and increasing. For its interpretation, we note that if $Y \overset{\mathrm{d}}{=} \mathrm{Poisson}(d)$ then by Stirling's formula we have $\mathbb{P}(Y = k) = \exp\left(-f_d(k/d) + O\left(\frac{1}{k}\right)\right)$ for any $k \in \mathbb{N}$. There is a universal constant $C > 0$ such that for $1 \leqslant l \leqslant \frac{N}{C\sqrt{d}}$ the equation $f_d(\beta) = \log(N/l)$ has a unique solution $\beta \equiv \beta_l(d)$. The interpretation of $\beta_l(d)$ is the typical location of $\alpha_{\sigma(l)}$. By the implicit function theorem, we find that $d \mapsto \beta_l(d)$ on the interval $\left(0, \frac{N^2}{Cl^2}\right]$ is a decreasing bijective function.

Definition A.8. An event $\Xi \equiv \Xi_N$ holds with *high probability* if $\mathbb{P}(\Xi) = 1 - o(1)$.

The following result is a slight generalization of [10, Proposition D.1], which can be established with the same proof. We note that the qualitative notion of high probability can be made stronger and quantitative with some extra effort, which we however refrain from doing here.

Lemma A.9. *If $d \geqslant 1$ and $l \geqslant 1$ satisfies $\beta_l(d) \geqslant 3/2$ then*

$$|\alpha_{\sigma(l)} - \beta_l(d)| \leqslant \frac{1 \vee (\zeta / \log \beta_l(d))}{d} \tag{A.11}$$

with high probability, where ζ is any sequence tending to infinity with N.

The following result[6] gives bounds on the counting function of the normalized degrees $(\alpha_x)_{x \in [N]}$.

Lemma A.10. *Suppose that ζ satisfies*

$$1 \ll \zeta \leqslant \frac{d}{C \log \log N} \tag{A.12}$$

for some large enough universal constant C. Then for any $\alpha \geqslant 2$ we have with high probability

$$\left\lfloor (N \mathrm{e}^{-f_d(\alpha)} - 1)(\log N)^{-2\zeta} \right\rfloor \leqslant |\{x \in [N] : \alpha_x \geqslant \alpha\}| \leqslant \left\lceil (N \mathrm{e}^{-f_d(\alpha)} + 1)(\log N)^{2\zeta} \right\rceil . \tag{A.13}$$

Proof. If $d > 3 \log N$, then an elementary analysis using Bennett's inequality shows that $|\{x \in [N] : \alpha_x \geqslant \alpha\}| = 0$ with high probability. Since $N \mathrm{e}^{-f_d(\alpha)} \leqslant 1$ for $\alpha \geqslant 2$, the claim follows. Thus, for the following we assume that $d \leqslant 3 \log N$.
Abbreviate $\Upsilon := \frac{3}{2} \frac{\zeta}{d}$, which is an upper bound for the right-hand side of (A.11). For the following we adopt the convention that $\beta_0(d) = \infty$. Choose $l \geqslant 0$ such that

$$\beta_{l+1}(d) < \alpha \leqslant \beta_l(d) , \tag{A.14}$$

and define

$$\underline{k} := \left\lfloor l (\log N)^{-2\zeta} \right\rfloor , \qquad \overline{k} := \left\lceil (l+1)(\log N)^{2\zeta} \right\rceil .$$

We shall show that

$$\beta_{\underline{k}}(d) - \Upsilon \geqslant \beta_l(d) \tag{A.15}$$

for $\underline{k} \geqslant 1$,

$$\beta_{\overline{k}}(d) + \Upsilon \leqslant \beta_{l+1}(d) , \tag{A.16}$$

and

$$\overline{k} \leqslant N \mathrm{e}^{-f_d(3/2)} . \tag{A.17}$$

[6] The assumption $d \gg \log \log N$ in Lemma A.10 is tailored so that it covers the entire range $\alpha \geqslant 2$, which is what we need in this paper. The assumption on d could also be removed at the expense of introducing a nontrivial lower bound on α.

Thus $\beta_{\overline{k}}(d) \geqslant 3/2$ and, assuming $\underline{k} \geqslant 1$, Lemma A.9 is applicable to the indices $\overline{k}$ and $\underline{k}$. We obtain, with high probability,

$$\alpha_{\sigma(\overline{k})} \leqslant \beta_{\overline{k}}(d) + \Upsilon \leqslant \beta_{l+1}(d) \leqslant \alpha \leqslant \beta_l(d) \leqslant \beta_{\underline{k}}(d) - \Upsilon \leqslant \alpha_{\sigma(\underline{k})}\,, \tag{A.18}$$

from which we deduce that

$$\underline{k} \leqslant |\{x \in [N] : \alpha_x \geqslant \alpha\}| \leqslant \overline{k}\,, \tag{A.19}$$

which also holds trivially also for the case $\underline{k} = 0$. By applying the function f_d to (A.14) we obtain $l \leqslant N\mathrm{e}^{-f_d(\alpha)} \leqslant l+1$, so that (A.19) yields (A.13).
Next, we verify (A.17). We consider the cases $l = 0$ and $l \geqslant 1$ separately. If $l = 0$ then, by the definition of $\beta_{\overline{k}}(d)$, for (A.17) we require $(\log N)^{2\zeta} + 1 \leqslant N\mathrm{e}^{-f_d(3/2)}$, which holds by the assumption $d \leqslant 3\log N$ and the upper bound on ζ. Let us therefore suppose that $l \geqslant 1$. By (A.14), $\alpha \geqslant 2$, and the definition of $\beta_l(d)$, we have $l \leqslant N\mathrm{e}^{-f_d(2)}$, and we have to ensure that $(l+2)(\log N)^{2\zeta} \leqslant N\mathrm{e}^{-f_d(3/2)}$. Since $l \geqslant 1$, this is satisfied provided that $3\mathrm{e}^{-f_d(2)}(\log N)^{2\zeta} \leqslant \mathrm{e}^{-f_d(3/2)}$, which holds provided that $f_d(2) - f_d(3/2) \geqslant 3\zeta \log\log N$. This inequality is true because $f_d(2) - f_d(3/2) \geqslant f_d'(3/2)/2 \geqslant d/C$, where we used that $f_d'(\alpha) = d\log\alpha + \frac{1}{2\alpha}$.
What remains, therefore, is the proof of (A.15) and (A.16). We begin with the proof of (A.15). We get from the mean value theorem that

$$\beta_{\underline{k}}(d) - \beta_l(d) = f_d^{-1}\left(\log\left(\frac{N}{\underline{k}}\right)\right) - f_d^{-1}\left(\log\left(\frac{N}{l}\right)\right) \geqslant \frac{3}{4d\log\beta_{\underline{k}}(d)}\log\left(\frac{l}{\underline{k}}\right). \tag{A.20}$$

The right-hand side of (A.20) is bounded from below by Υ provided that

$$\log\left(\frac{l}{\underline{k}}\right) \geqslant 2\zeta\log\beta_{\underline{k}}(d)\,. \tag{A.21}$$

We estimate $\beta_{\underline{k}}(d) \leqslant \beta_1(d)$ using the elementary bound $f_d(\beta) \geqslant \frac{d}{10}\beta$ for $\beta \geqslant 2$, which yields $\log N = f_d(\beta_1(d)) \geqslant \frac{d}{10}\beta_1(d)$. By assumption on d we therefore get

$$\beta_1(d) \leqslant \log N\,. \tag{A.22}$$

Thus, (A.21) holds by $\underline{k} \leqslant l/(\log N)^{2\zeta}$. This concludes the proof of (A.15).
Next, we prove (A.16). As in (A.20), we find

$$\begin{aligned}\beta_{l+1}(d) - \beta_{\overline{k}}(d) &= f_d^{-1}\left(\log\left(\frac{N}{l+1}\right)\right) - f_d^{-1}\left(\log\left(\frac{N}{\overline{k}}\right)\right)\\ &\geqslant \frac{3}{4d\log\beta_{l+1}(d)}\log\left(\frac{\overline{k}}{l+1}\right).\end{aligned} \tag{A.23}$$

Together with $\beta_{l+1}(d) \leqslant \beta_1(d) \leqslant \log N$ from (A.22), we deduce that the right-hand side of (A.23) is bounded from below by Υ provided that $\log\left(\frac{\overline{k}}{l+1}\right) \geqslant 2\zeta\log\log N$, which is true by definition of $\overline{k}$. This concludes the proof of (A.16). □

The following result follows easily from Lemma A.10. Recall the definition (1.13) of the exponent $\theta_b(\alpha)$.

Corollary A.11. *Suppose that ζ satisfies (A.12). Write $d = b \log N$. Then for any $\alpha \geqslant 2$ we have*

$$|\{x \in [N] : \alpha_x \geqslant \alpha\}| \vee 1 = N^{\theta_b(\alpha)+\varepsilon}, \qquad \varepsilon = O\left(\frac{\zeta \log\log N}{\log N}\right)$$

with high probability.

Using the exponent $\theta_b(\alpha)$ from (1.13) and $\alpha_{\max}(b)$ defined below it, we may state the following estimate on the density of the normalized degrees and the number of resonant vertices.

Lemma A.12. *The following holds for a large enough universal constant C. Suppose that ζ satisfies (A.12). Write $d = b \log N$.*

(i) *For $2 \leqslant \alpha < \beta \leqslant \alpha_{\max}(b)$ satisfying $\beta - \alpha \geqslant C\frac{\zeta \log\log N}{d \log \alpha}$, with high probability we have*

$$|\{x \in [N] : \alpha \leqslant \alpha_x \leqslant \beta\}| = N^{\theta_b(\alpha)+\varepsilon}, \qquad \varepsilon = O\left(\frac{\zeta \log\log N}{\log N}\right). \tag{A.24}$$

(ii) *For $\delta \geqslant C\frac{\zeta \log\log N}{d}$ and $2 + \delta \leqslant \lambda \leqslant \Lambda(\alpha_{\max}(b))$, with high probability we have*

$$|\mathcal{W}_{\lambda,\delta}| = N^{\theta_b(\Lambda^{-1}(\lambda-\delta))+\varepsilon}, \qquad \varepsilon = O\left(\frac{\zeta \log\log N}{\log N}\right).$$

Note that, since $\xi \geqslant d^{-1/2}$, if the conclusion of Theorem 1.2 is nontrivial then $\delta \geqslant d^{-1/2}$, and hence the assumption on δ in Lemma A.12 (ii) is automatically satisfied for suitably chosen ζ.

Proof of Lemma A.12. Part (i) follows Corollary A.11 below by noting that the assumption on β implies $\theta_b(\alpha) - \theta_b(\beta) \geqslant C\frac{\zeta \log\log N}{\log N}$ by the mean value theorem.
Part (ii) follows from Part (i), using that $\log(\lambda - \delta) \geqslant \log 2$, that Λ' is bounded on $[2, \infty)$, and the mean value theorem. □

Corollary A.13. *The following holds for large enough universal constants $C, \mathcal{C}$. Suppose that (1.10) holds. Write $d = b \log N$. Let $\mathbf{w} = (w_x)_{x\in[N]}$ be a normalized eigenvector of $A/\sqrt{d}$ with nontrivial eigenvalue $2 + \mathcal{C}\xi^{1/2} \leqslant \lambda \leqslant \Lambda(\alpha_{\max}(b))$. Then with high probability for any $2 \leqslant p \leqslant \infty$ we have*

$$\|\mathbf{w}\|_p^2 \geqslant N^{(2/p-1)\theta_b(\Lambda^{-1}(\lambda))+\varepsilon}, \varepsilon = O\left[\frac{\log\log N}{\sqrt{\log N}} + b(\log\lambda)\left(\lambda + \frac{1}{\sqrt{\lambda-2}}\right)(\xi + \xi_{\lambda-2})\right].$$

Proof. We choose $\delta := C(\xi + \xi_{\lambda-2})$. Then by assumption on λ we have $\delta \leqslant (\lambda-2)/2$, and hence Theorem 1.2 yields, using that $\mathbf{v}(x)$ is supported in $B_{r_\star}(x)$, $\sum_{x\in\mathcal{W}_{\lambda,\delta}} \sum_{y\in B_{r_\star}(x)} w_y^2 \geqslant \frac{1}{2}$ with high probability. Using that for any vector $\mathbf{x} \in \mathbb{R}^n$ we have $\|\mathbf{x}\|_p^2 \geqslant n^{2/p-1}\|\mathbf{x}\|_2^2$ (by Hölder's inequality), with the choice $n = \sum_{x\in\mathcal{W}_{\lambda,\delta}} |B_{r_\star}(x)|$, we get

$$\|\mathbf{w}\|_p^2 \geqslant \frac{1}{2}\left(\sum_{x\in\mathcal{W}_{\lambda,\delta}} |B_{r_\star}(x)|\right)^{2/p-1} \geqslant \frac{1}{2}\left(|\mathcal{W}_{\lambda,\delta}| N^{C\log\log N/\sqrt{\log N}}\right)^{2/p-1} \tag{A.25}$$

with high probability, where we used Lemma A.7 to estimate $\max_{x\in[N]}|B_{r_\star}(x)| \leqslant N^{C\log\log N/\sqrt{\log N}}$ with high probability.
Next, using the mean value theorem and elementary estimates on the derivatives of θ_b and Λ^{-1}, we estimate

$$\theta_b(\Lambda^{-1}(\lambda-\delta)) - \theta_b(\Lambda^{-1}(\lambda)) \leqslant Cb(\log\lambda)\left(\lambda + \frac{1}{\sqrt{\lambda-2}}\right)\delta\,.$$

Invoking Lemma A.12 (ii) with $\zeta := \log\log N$, and recalling (A.25), therefore yields the claim. □

A.5. Connected components of $\mathbb{G}(N, d/N)$. In this appendix we give some basic estimates on the sizes of connected components of $\mathbb{G}(N, d/N)$. These are needed for the analysis of the tuning forks in Appendix A.6 below. The arguments are standard and are tailored to work well in the regime $1 \ll d \leqslant \log N$ that we are interested in. For smaller values of d, see e.g. [17].

Lemma A.14. *Let W_k be the number of connected components that have k vertices and $\widehat{W}_k$ the number of connected components that have k vertices and are not a tree. Then for $k \leqslant N/2$ we have*

$$\mathbb{E}[W_k] \leqslant N\mathrm{e}^{-k(d/2-\log d-1)}\,, \qquad \mathbb{E}[\widehat{W}_k] \leqslant \mathrm{e}^{-k(d/2-\log d-1)}\,.$$

Proof. For a set $X \subset [N]$, denote by $\mathcal{T}(X)$ the set of spanning trees of X. If X is a connected component of $\mathbb{G}$ then there exists $\mathbb{T} \in \mathcal{T}(X)$ a subgraph of $\mathbb{G}$ such that no vertex of X is connected to a vertex of $[N] \setminus X$. Hence,

$$W_k \leqslant \sum_{X\subset[N]} \mathbb{1}_{|X|=k} \sum_{\mathbb{T}\in\mathcal{T}(X)} \mathbb{1}_{\mathbb{T}\subset\mathbb{G}} \prod_{x\in X}\prod_{y\in[N]\setminus X}(1-A_{xy})\,.$$

Taking the expectation now easily yields the claim, using $|\mathcal{T}(X)| = |X|^{|X|-2}$ by Cayley's theorem, that a tree on k vertices has $k-1$ edges, Stirling's approximation, and $1-x \leqslant \mathrm{e}^{-x}$.
The argument to estimate $\widehat{W}_k$ is similar, noting that in addition to a spanning tree $\mathbb{T}$ of X, we also have to have at least one edge not in $\mathbb{T}$ connecting two vertices of X. Thus,

$$\widehat{W}_k \leqslant \sum_{X\subset[N]} \mathbb{1}_{|X|=k} \sum_{\mathbb{T}\in\mathcal{T}(X)} \mathbb{1}_{\mathbb{T}\subset\mathbb{G}} \prod_{x\in X}\prod_{y\in[N]\setminus X}(1-A_{xy}) \sum_{\{u,v\}\in X^2\setminus E(\mathbb{T})} A_{uv}\,,$$

and we may estimate the expectation as before. □

We call a connected component of $\mathbb{G}$ *small* if it is not the giant component. For the following statement we recall the definition of high probability from Definition A.8.

Corollary A.15. *Suppose that $d \gg 1$. All small components of $\mathbb{G}$ have at most $O\big(\frac{\log N}{d}\big)$ vertices with very high probability. All small components of $\mathbb{G}$ are trees with high probability. The giant component of $\mathbb{G}$ has at least $N(1-\mathrm{e}^{-d/4})$ vertices with high probability.*

Proof. Any small component has at most $N/2$ vertices. Using Lemma A.14 we therefore get that the probability that there exists a small component with at least K vertices is bounded by

$$\mathbb{P}(\exists k \in [K, N/2], W_k \geqslant 1) \leqslant \sum_{k=K}^{N/2} \mathbb{E}[W_k] \leqslant 2N\mathrm{e}^{-K(d/2-\log d-1)},$$

by summing the geometric series. Since $d/2-\log d-1 \geqslant cd$ for some universal constant c, we obtain the first claim. To obtain the second claim, we use Lemma A.14 to estimate the probability that there exists a small component that is not a tree by $\sum_{k=1}^{N/2} \mathbb{E}\widehat{W}_k \leqslant \mathrm{e}^{-d/3}$. To obtain the last claim, we estimate the expected number of vertices in small components by $\mathbb{E}\big[\sum_{k=1}^{N/2} kW_k\big] \leqslant N\sum_{k=1}^{\infty} k\mathrm{e}^{-k(d/2-\log d-1)} \leqslant CN\mathrm{e}^{-d/3}$ using Lemma A.14, and the third claim follows from Chebyshev's inequality. □

We may now estimate the adjacency matrix on the small components of $\mathbb{G}(N, d/N)$. The following result follows immediately from Corollary A.15 and Lemma A.4.

Corollary A.16. *Suppose that $d \gg 1$. Then the operator norm of $A/\sqrt{d}$ restricted to the small components of $\mathbb{G}$ is bounded by $O\big(\frac{\sqrt{\log N}}{d}\big)$ with high probability.*

Corollary A.16 makes it explicit that Theorem 1.8 excludes all eigenvectors on small components of $\mathbb{G}$, whose eigenvalues lie outside $\mathcal{S}_\kappa$ precisely under the lower bound from (1.18).

A.6. Tuning forks and proof of Lemma 1.12. In this appendix we give a precise definition of the D-tuning forks from Sect. 1.5 and prove Lemma 1.12.

Definition A.17. A *star of degree $D \in \mathbb{N}$* consists of a vertex, the *hub*, and D leaves adjacent to the hub, the *spokes*. A *star tuning fork of degree D* is obtained by taking two disjoint stars of degree D along with an additional vertex, the *base*, and connecting both hubs to the base. We say that a star tuning fork is *rooted in a graph $\mathbb{H}$* if it is a subgraph of $\mathbb{H}$ in which both hubs have degree $D+1$ and all spokes are leaves.

Lemma A.18. *If a star tuning fork of degree D is rooted in some graph $\mathbb{H}$, then the adjacency matrix of $\mathbb{H}$ has eigenvalues $\pm\sqrt{D}$ with corresponding eigenvectors supported on the stars of the tuning fork, i.e. on $2D+2$ vertices.*

Proof. Suppose first that $D \geqslant 1$. Note first that the adjacency matrix of a star of degree D has rank two and has the two nonzero eigenvalues $\pm\sqrt{D}$, with associated eigenvector equal to $\pm\sqrt{D}$ at the hub and 1 at the spokes. Now take a star tuning fork of degree D rooted in a graph $\mathbb{H}$. Define a vector on the vertex set of $\mathbb{H}$ by setting it to be $\pm\sqrt{D}$ at the hub of the first star, 1 at the spokes of the first star, $\mp\sqrt{D}$ at the hub of the second star, -1 at the spokes of the second star, and 0 everywhere else. Then it is easy to check that this vector is an eigenvector of the adjacency matrix of $\mathbb{H}$ with eigenvalue $\pm\sqrt{D}$. If $D = 0$ the construction is analogous, defining the vector to be +1 at one hub and -1 at the other. □

We recall from Sect. 1.5 that $F(d, D)$ denotes the number of star tuning forks of degree D rooted in $\mathbb{G}_{\mathrm{giant}}$.

Lemma A.19. *Suppose that* $1 \ll d \ll \sqrt{N}$ *and* $0 \leqslant D \ll \sqrt{N}$. *Then*

$$\mathbb{E}[F(d, D)] = \frac{Nd^2 \mathrm{e}^{-2d}}{2D!^2} (d\mathrm{e}^{-d+1})^{2D} (1 + o(1)) \tag{A.26}$$

and $\mathbb{E}[F(d, D)^2] \leqslant \mathbb{E}[F(d, D)]^2 (1 + o(1))$.

Proof of Lemma 1.12. From Lemma A.19 we deduce that if $1 \ll d = b \log N = O(\log N)$ and $D \ll \log N / \log\log N$, then $\mathbb{E}[F(d, D)] = N^{1-2b-2bD+o(1)}$. The claim then follows from the second moment estimate in Lemma A.19 and Chebyshev's inequality. □

Proof of Lemma A.19. Let $x_1, x_2 \in [N]$ be distinct vertices and $R_1, R_2 \subset [N] \setminus \{x_1, x_2\}$ be disjoint subsets of size D. We abbreviate $U = (x_1, x_2, R_1, R_2)$ and sometimes identify U with $\{x_1, x_2\} \cup R_1 \cup R_2$. The family U and a vertex $o \in [N] \setminus U$ define a star tuning fork of degree D with base o, hubs x_1 and x_2, and associated spokes R_1 and R_2. Let $\mathscr{C}_k(\mathbb{H})$ denote the vertex set of the kth largest connected component of the graph $\mathbb{H}$. Then $F(d, D) = \frac{1}{2} \sum_U \sum_{o \in [N] \setminus U} \mathbb{1}_{o \in \mathscr{C}_1(\mathbb{G})} S_{o,U}$, where

$$S_{o,U} := \prod_{i=1}^{2} \Bigg(\prod_{u \in R_i \cup \{o\}} A_{x_i u} \prod_{u \in [N] \setminus (R_i \cup \{o\})} (1 - A_{x_i u}) \prod_{u \in R_i} \prod_{v \in [N] \setminus \{x_i\}} (1 - A_{uv}) \Bigg).$$

The factor $\frac{1}{2}$ corrects the overcounting from the labelling of the two stars.
For disjoint deterministic U, we split the random variables $A = (A', A'')$ into two independent families, where $A' := (A_{uv} : u \in U \text{ or } v \in U)$ and $A'' := (A_{uv} : u, v \in [N] \setminus U)$. Note that $S_{o,U}$ is A'-measurable. We define the event

$$\Xi := \{ |\mathscr{C}_1(\mathbb{G}|_{[N] \setminus U})| > |\mathscr{C}_2(\mathbb{G}|_{[N] \setminus U})| + 2D + 2 \},$$

which is A''-measurable. By Corollary A.15 and the assumption on D, the event Ξ holds with high probability. Moreover, we have $\mathbb{1}_\Xi \mathbb{1}_{o \in \mathscr{C}_1(\mathbb{G})} S_{o,U} = \mathbb{1}_\Xi \mathbb{1}_{o \in \mathscr{C}_1(\mathbb{G}|_{[N] \setminus U})} S_{o,U}$, since the component of o in $\mathbb{G}$ and $\mathbb{G}|_{[N] \setminus U}$ differ by $2D + 2$ vertices. Thus, for fixed $o \in [N] \setminus U$, using the independence of A' and A'', we get

$$\begin{aligned} \mathbb{E}[\mathbb{1}_{o \in \mathscr{C}_1(\mathbb{G})} S_{o,U}] &= \mathbb{E}[\mathbb{1}_\Xi \mathbb{1}_{o \in \mathscr{C}_1(\mathbb{G}|_{[N] \setminus U})} S_{o,U}] + \mathbb{E}[\mathbb{1}_{\Xi^c} \mathbb{1}_{o \in \mathscr{C}_1(\mathbb{G})} S_{o,U}] \\ &= \mathbb{E}[S_{o,U}] \big[\mathbb{P}\big(o \in \mathscr{C}_1(\mathbb{G}|_{[N] \setminus U})\big) + O\big(\mathbb{P}(\Xi^c)\big) \big]. \end{aligned}$$

We have $\mathbb{P}(\Xi^c) = o(1)$ and $\mathbb{P}\big(o \in \mathscr{C}_1(\mathbb{G}|_{[N] \setminus U})\big) = 1 - o(1)$ by Corollary A.15 and the assumption on D. Computing $\mathbb{E}[S_{o,U}]$ and performing the sum over o and U, we therefore conclude that

$$\begin{aligned} \mathbb{E}[F(d, D)] = \frac{N(N-1) \cdots (N - 2D - 3 + 1)}{2D!^2} \\ \left(\frac{d}{N}\right)^{2D+2} \left(1 - \frac{d}{N}\right)^{2(N-D-1)+2D(N-1)} (1 + o(1)), \end{aligned}$$

from which (A.26) follows. The estimate of the second moment is similar; one can even disregard the restriction to the giant component by estimating $\mathbb{E}[F(d, D)^2] \leqslant \frac{1}{4} \sum_{U, \tilde{U}} \sum_{o, \tilde{o} \in [N]} \mathbb{E}[S_{o,U} S_{\tilde{o}, \tilde{U}}]$; we omit the details. □

A.7. Multilinear large deviation bounds for sparse random vectors. In this appendix we collect basic large deviation bounds for multilinear functions of sparse random vectors, which are proved in [42]. The following result is proved in Propositions 3.1, 3.2, and 3.5 of [42]. We denote by $\|X\|_r := (\mathbb{E}|X|^r)^{1/r}$ the L^r-norm of a random variable X.

Proposition A.20. *Let r be even and $1 \leqslant d \leqslant N$. Let $X_1, \dots, X_N$ be independent random variables satisfying*

$$\mathbb{E}X_i = 0, \qquad \mathbb{E}|X_i|^k \leqslant \frac{1}{Nd^{(k-2)/2}}$$

for all $i \in [N]$ and $2 \leqslant k \leqslant r$. Let $a_i \in \mathbb{C}$ and $b_{ij} \in \mathbb{C}$ be deterministic for all $i, j \in [N]$. Suppose that

$$\left(\frac{1}{N}\sum_i |a_i|^2\right)^{1/2} \leqslant \gamma\,, \qquad \frac{\max_i |a_i|}{\sqrt{d}} \leqslant \psi,$$

and

$$\left(\max_i \frac{1}{N}\sum_j |b_{ij}|^2\right)^{1/2} \vee \left(\max_j \frac{1}{N}\sum_i |b_{ij}|^2\right)^{1/2} \leqslant \gamma, \qquad \frac{\max_{i,j}|b_{ij}|}{d} \leqslant \psi$$

for some $\gamma, \psi \geqslant 0$. Then

$$\left\|\sum_i a_i X_i\right\|_r \leqslant \left(\frac{2r}{1 + 2(\log(\psi/\gamma))_+} \vee 2\right)(\gamma \vee \psi), \tag{A.27a}$$

$$\left\|\sum_i a_i \left(|X_i|^2 - \mathbb{E}|X_i|^2\right)\right\|_r \leqslant 2\left(1 + \frac{2d}{N}\right)\max_i |a_i| \left(\frac{r}{d} \vee \sqrt{\frac{r}{d}}\right), \tag{A.27b}$$

$$\left\|\sum_{i \neq j} b_{ij} X_i X_j\right\|_r \leqslant \left(\frac{4r}{1 + (\log(\psi/\gamma))_+} \vee 4\right)^2 (\gamma \vee \psi). \tag{A.27c}$$

The L^r-norm bounds in Proposition A.20 induce bounds that hold with very high probability.

Corollary A.21. *Fix $\kappa \in (0, 1)$. Let the assumptions of Proposition A.20 be satisfied. If $\psi/\gamma \geqslant N^{\kappa/4}$ then with very high probability*

$$\left|\sum_i a_i X_i\right| \leqslant C\psi\,, \qquad \left|\sum_{i \neq j} b_{ij} X_i X_j\right| \leqslant C\psi\,. \tag{A.28}$$

Remark A.22. Our proof of Corollary A.21 shows that C can be chosen as a linear function of ν for the first estimate of (A.28) and as a quadratic function of ν for the second estimate of (A.28).

Proof. Fix $\nu \geqslant 1$. We choose $r = \nu \log N$ in (A.27a) of Proposition A.20 and obtain from Cheybshev's inequality that

$$\mathbb{P}\left(\left|\sum_i a_i X_i\right| > C\psi\right) \leqslant N^{-\nu}, \qquad C := \frac{4\mathrm{e}}{\kappa}\nu$$

as $\kappa \in (0,1)$. Similarly, choosing $r = \frac{1}{2}\nu \log N$ in (A.27c) yields

$$\mathbb{P}\left(\left|\sum_{i\neq j} b_{ij} X_i X_j\right| > 4\mathcal{C}\psi\right) \leqslant N^{-\nu}, \qquad \mathcal{C} := \frac{16\mathrm{e}^2}{\kappa^2}\nu^2 .$$

□

A.8. Resolvent identities. In this appendix we record some well-known identities for the Green function (4.2) and its minors from Definition 4.5. We assume throughout that $z \in \mathbb{C} \setminus \mathbb{R}$.

Lemma A.23 (Ward identity). *For $x \notin T \subset [N]$ we have*

$$\sum_{y}^{(T)} |G_{xy}^{(T)}|^2 = \frac{1}{\operatorname{Im} z} \operatorname{Im} G_{xx}^{(T)} .$$

Proof. This is a standard identity for resolvents, see e.g. [16, Eq. (3.6)]. □

Lemma A.24. *Let $T \subset [N]$. For $x, y \notin T$ and $x \neq y$, we have*

$$G_{xy}^{(T)} = -G_{yy}^{(T)} \sum_{a}^{(Ty)} G_{xa}^{(Ty)} M_{ay} = -G_{xx}^{(T)} \sum_{b}^{(Tx)} M_{xb} G_{by}^{(Tx)}. \tag{A.29}$$

For $x, y, a \notin T$ and $x \neq a \neq y$, we have

$$G_{xy}^{(Ta)} = G_{xy}^{(T)} - \frac{G_{xa}^{(T)} G_{ay}^{(T)}}{G_{aa}^{(T)}}. \tag{A.30}$$

For any $x \in [N]$, we have

$$\frac{1}{G_{xx}} = M_{xx} - z - \sum_{a,b}^{(x)} M_{xa} G_{ab}^{(x)} M_{bx}. \tag{A.31}$$

Proof. All identities are standard and proved e.g. in [16]: (A.29) in [16, Eq. (3.5)], (A.30) in [16, Eq. (3.4)] and (A.31) in [16, Lemma A.1 and (5.1)]. □

We recall (4.1) and derive two expansions used in Sect. 4. For any $T \subset [N]$ and $x, y, u \notin T$, $x \neq u \neq y$, we have

$$G_{xy}^{(Tu)} = G_{xy}^{(T)} + \sum_{a}^{(Tu)} G_{xa}^{(Tu)} H_{au} G_{uy}^{(T)} + \frac{f}{N} G_{uy}^{(T)} \sum_{a}^{(Tu)} G_{xa}^{(Tu)}, \tag{A.32a}$$

which follows from (A.30) and (A.29). Under the same assumptions, applying (A.29) to (A.32a) yields

$$\begin{aligned} G_{xy}^{(Tu)} = {} & G_{xy}^{(T)} - G_{uu}^{(T)} \sum_{a}^{(Tu)} G_{xa}^{(Tu)} H_{au} \sum_{b}^{(Tu)} H_{ub} G_{by}^{(Tu)} \\ & - \frac{f}{N} G_{uu}^{(T)} \sum_{a}^{(Tu)} G_{xa}^{(Tu)} H_{au} \sum_{b}^{(Tu)} G_{by}^{(Tu)} + \frac{f}{N} G_{uy}^{(T)} \sum_{a}^{(Tu)} G_{xa}^{(Tu)}. \end{aligned} \tag{A.32b}$$

A.9. Stability estimate—proof of Lemma 4.19. In this appendix we prove Lemma 4.19. The estimate in [27, Lemma 3.5] corresponding to (4.43) has logarithmic factors, which are not affordable for our purposes: they have to be replaced with constants. The following proof of Lemma 4.19 is analogous to that of the more complicated bulk stability estimate from [9, Lemma 5.11].

Proof of Lemma 4.19. We introduce the vectors $\mathbf{g} := (g_x)_{x\in\mathcal{X}}$ and $\boldsymbol{\varepsilon} := (\varepsilon_x)_{x\in\mathcal{X}}$. Moreover, with the abbreviation $m := m(z)$ we introduce the constant vectors $\mathbf{m} = (m)_{x\in\mathcal{X}}$ and $\mathbf{e} := |\mathcal{X}|^{-1/2}(1)_{x\in\mathcal{X}}$. We regard all vectors as column vectors. A simple computation starting from the difference of (4.6) and (4.42) reveals that

$$B(\mathbf{g}-\mathbf{m}) = m(\mathbf{g}-\mathbf{m})\big(\mathbf{e}\mathbf{e}^*(\mathbf{g}-\mathbf{m})\big) - (\mathbf{g}-\mathbf{m})m\boldsymbol{\varepsilon} - m^2\boldsymbol{\varepsilon}, \tag{A.33}$$

where $B := 1 - m^2\mathbf{e}\mathbf{e}^*$, and column vectors are multiplied entrywise. The inverse of B is

$$B^{-1} = 1 + \frac{m^2}{1-m^2}\mathbf{e}\mathbf{e}^*.$$

For a matrix $R \in \mathbb{C}^{\mathcal{X}\times\mathcal{X}}$, we write $\|R\|_{\infty\to\infty}$ for the operator norm induced by the norm $\|\mathbf{r}\|_\infty = \max_{x\in\mathcal{X}}|r_x|$ on $\mathbb{C}^{\mathcal{X}}$. It is easy to see that there is $c > 0$, depending only on κ, such that $|1-m(w)^2| \geqslant c$ for all $w \in \mathbb{C}_+$ satisfying $|\mathrm{Re}\, w| \leqslant 2-\kappa$. Hence, owing to $\|\mathbf{e}\mathbf{e}^*\|_{\infty\to\infty} = 1$, we obtain $\|B^{-1}\|_{\infty\to\infty} \leqslant 1 + |1-m^2|^{-1} \leqslant 1 + c^{-1}$. Therefore, inverting B in (A.33) and choosing b, depending only on κ, sufficiently small to absorb the term quadratic in $\mathbf{g}-\mathbf{m}$ into the left-hand side of the resulting bound yields (4.43) for some sufficiently large $C > 0$, depending only on κ. This concludes the proof of Lemma 4.19. □

A.10. Instability estimate—proof of (2.11). In this appendix we prove (2.11), which shows that the self-consistent equation (2.10) is unstable with a logarithmic factor, which renders it useless for the analysis of sparse random graphs. More precisely, we show that the norm $\|(I-m^2S)^{-1}\|_{\infty\to\infty}$ is ill-behaved precisely in the situation where we need it. For simplicity, we replace m^2 with a phase $\alpha^{-1} \in S^1$ separated from ± 1, since for $\mathrm{Re}\, z \in \mathcal{S}_\kappa$ we have

$$|m(z)|^2 = 1 - O(\mathrm{Im}\, z)\,, \qquad \mathrm{Im}\, m(z) \asymp 1\,, \tag{A.34}$$

by [33, Lemma 3.5]. Moreover, for definiteness, recalling that with very high probability most of the $d(1+o(1))$ neighbours of any vertex in $\mathcal{T}$ are again in $\mathcal{T}$, we assume that S is the adjacency matrix of a d-regular graph on $\mathcal{T}$ divided by d.
By the spectral theorem and because S is Hermitian, $\|(\alpha - S)^{-1}\|_{2\to 2}$ is bounded, but, as we now show, the same does not apply to $\|(\alpha-S)^{-1}\|_{\infty\to\infty}$. Indeed, the upper bound of (2.11) follows from [34, Proposition A.2], and the lower bound from the following result.

Lemma A.25 (Instability of (2.10)). *Let S be $1/d$ times the adjacency matrix of a graph whose restriction to the ball of radius $r \in \mathbb{N}^*$ around some distinguished vertex is a d-regular tree. Let $\alpha \in S^1$ be an arbitrary phase. Then*

$$\|(\alpha - S)^{-1}\|_{\infty\to\infty} \geqslant c\left(\frac{r}{\log r} \wedge d\right) \tag{A.35}$$

for some universal constant $c > 0$.

In particular, denoting by N the number of vertices in the tree (which may be completed to a d-regular graph by connecting the leaves to each other), for $d \asymp \log N$ and $r \asymp \frac{\log N}{\log d}$ we find

$$\|(\alpha - S)^{-1}\|_{\infty\to\infty} \geqslant \frac{c \log N}{(\log\log N)^2}\,, \tag{A.36}$$

which is the lower bound of (2.11).

Proof of Lemma A.25. After making r smaller if needed, we may assume that $\frac{r}{\log r} \leqslant d$. We shall construct a vector $\mathbf{u}$ satisfying $\|\mathbf{u}\|_\infty = 1$ and $\|(\alpha - S)\mathbf{u}\|_\infty = O\big(\frac{\log r}{r}\big)$, from which (A.35) will follow. To that end, we construct the sequence $a_0, a_1, \ldots, a_r$ by setting

$$a_0 := 1\,, \qquad a_1 := \alpha\,, \qquad a_{k+1} := \frac{d}{d-1}\alpha a_k - \frac{1}{d-1}a_{k-1} \quad \text{for} \quad 1 \leqslant k \leqslant r-1\,.$$

A short transfer matrix analysis shows that $|a_k| \leqslant \mathrm{e}^{C_1 k/d}$ for some constant C_1. Now choose $\mu := C_2\frac{\log r}{r}$ with $C_2 := 2 \vee 2C_1$, and define $b_k := \mathrm{e}^{-\mu k}a_k$. Calling o the distinguished vertex, we define $u_x := b_k$ if $k = \mathrm{dist}(o, x) \leqslant r$ and $u_x = 0$ otherwise. It is now easy to check that $\|(\alpha - S)\mathbf{u}\|_\infty = O\big(\frac{\log r}{r}\big)$, by considering the cases $k = 0$, $1 \leqslant k \leqslant r-1$, and $k \geqslant r$ separately. The basic idea of the construction is that if μ were zero, then $(\alpha - S)\mathbf{u}$ would vanish exactly on $B_{r-1}(o)$, but it would be large on the boundary $S_r(o)$. The factor $\mathrm{e}^{-\mu k}$ introduces exponential decay in the radius which dampens the contribution of the boundary $S_r(o)$ at the expense of introducing errors in the interior $B_{r-1}(o)$. □

References

1. Aggarwal, A.: Bulk universality for generalized Wigner matrices with few moments. Prob. Theor. Rel. Fields **173**(1–2), 375–432 (2019)
2. Aggarwal, A., Lopatto, P., Marcinek, J.: Eigenvector statistics of Lévy matrices. Preprint arXiv:2002.09355 (2020)
3. Aggarwal, A., Lopatto, P., Yau, H.-T.: GOE statistics for Lévy matrices. Preprint arXiv:1806.07363 (2018)
4. Aizenman, M., Molchanov, S.: Localization at large disorder and at extreme energies: an elementary derivation. Commun. Math. Phys. **157**, 245–278 (1993)
5. Aizenman, M., Sims, R., Warzel, S.: Absolutely continuous spectra of quantum tree graphs with weak disorder. Commun. Math. Phys. **264**(2), 371–389 (2006)
6. Aizenman, M., Warzel, S.: Extended states in a Lifshitz tail regime for random Schrödinger operators on trees. Phys. Rev. Lett. **106**(13), 136804 (2011)
7. Aizenman, M., Warzel, S.: Resonant delocalization for random Schrödinger operators on tree graphs. J. Eur. Math. Soc. **15**(4), 1167–1222 (2013)
8. Aizenman, M., Warzel, S.: Random Operators, Graduate Studies in Mathematics. American Mathematical Society, Providence (2015)
9. Ajanki, O.H., Erdős, L., Krüger, T.: Quadratic vector equations on complex upper half-plane. Mem. Am. Math. Soc. **261**(1261), v+133 (2019)
10. Alt, J., Ducatez, R., Knowles, A.: Extremal eigenvalues of critical Erdős–Rényi graphs. Ann. Prob. **49**(3), 1347–1401 (2021)
11. Anderson, P.W.: Absence of diffusion in certain random lattices. Phys. Rev. **109**(5), 1492 (1958)
12. Bauerschmidt, R., Huang, J., Yau, H.-T.: Local Kesten–McKay law for random regular graphs. Commun. Math. Phys. **369**(2), 523–636 (2019)
13. Bauerschmidt, R., Knowles, A., Yau, H.-T.: Local semicircle law for random regular graphs. Commun. Pure Appl. Math. **70**(10), 1898–1960 (2017)
14. Benaych-Georges, F., Bordenave, C., Knowles, A.: Largest eigenvalues of sparse inhomogeneous Erdős–Rényi graphs. Ann. Prob. **47**(3), 1653–1676 (2019)

15. Benaych-Georges, F., Bordenave, C., Knowles, A.: Spectral radii of sparse random matrices. Ann. Inst. Henri Poincaré Probab. Stat. **56**(3), 2141–2161 (2020)
16. Benaych-Georges, F., Knowles, A.: Local Semicircle Law for Wigner Matrices, Advanced Topics in Random Matrices, Panor. Synthèses, vol. 53, pp. 1–90. Soc. Math. France, Paris (2017)
17. Bollobás, B.: Random Graphs. Cambridge University Press, Cambridge (2001)
18. Bordenave, C., Guionnet, A.: Localization and delocalization of eigenvectors for heavy-tailed random matrices. Prob. Theor. Rel. Fields **157**(3–4), 885–953 (2013)
19. Bordenave, C., Guionnet, A.: Delocalization at small energy for heavy-tailed random matrices. Commun. Math. Phys. **354**(1), 115–159 (2017)
20. Bourgade, P., Erdős, L., Yau, H.-T., Yin, J.: Universality for a class of random band matrices. Adv. Theor. Math. Phys. **21**(3), 739–800 (2017)
21. Bourgade, P., Yang, F., Yau, H.-T., Yin, J.: Random band matrices in the delocalized phase, II: generalized resolvent estimates. J. Stat. Phys. **174**(6), 1189–1221 (2019)
22. Bourgade, P., Yau, H.-T., Yin, J.: Random band matrices in the delocalized phase, I: quantum unique ergodicity and universality. Commun. Pure Appl. Math. **73**(7), 1526–1596 (2020)
23. Casati, G., Molinari, L., Izrailev, F.: Scaling properties of band random matrices. Phys. Rev. Lett. **64**(16), 1851 (1990)
24. Cizeau, P., Bouchaud, J.-P.: Theory of Lévy matrices. Phys. Rev. E **50**(3), 1810 (1994)
25. Combes, J.M., Thomas, L.: Asymptotic behaviour of eigenfunctions for multiparticle Schrödinger operators. Commun. Math. Phys. **34**, 251–270 (1973)
26. Dumitriu, I., Edelman, A.: Matrix models for beta ensembles. J. Math. Phys. **43**(11), 5830–5847 (2002)
27. Erdős, L., Yau, H.-T., Yin, J.: Bulk universality for generalized Wigner matrices. Prob. Theor. Rel. Fields **154**(1–2), 341–407 (2012)
28. Erdős, L., Knowles, A.: Quantum diffusion and delocalization for band matrices with general distribution. Ann. H. Poincaré **12**, 1227–1319 (2011)
29. Erdős, L., Knowles, A.: Quantum diffusion and eigenfunction delocalization in a random band matrix model. Commun. Math. Phys. **303**, 509–554 (2011)
30. Erdős, L., Knowles, A.: The Altshuler-Shklovskii formulas for random band matrices II: the general case. Ann. H. Poincaré **16**, 709–799 (2014)
31. Erdős, L., Knowles, A.: The Altshuler-Shklovskii formulas for random band matrices I: the unimodular case. Commun. Math. Phys. **333**, 1365–1416 (2015)
32. Erdős, L., Knowles, A., Yau, H.-T.: Averaging fluctuations in resolvents of random band matrices. Ann. H. Poincaré **14**, 1837–1926 (2013)
33. Erdős, L., Knowles, A., Yau, H.-T., Yin, J.: Delocalization and diffusion profile for random band matrices. Commun. Math. Phys. **323**, 367–416 (2013)
34. Erdős, L., Knowles, A., Yau, H.-T., Yin, J.: The local semicircle law for a general class of random matrices. Electron. J. Probab. **18**, 1–58 (2013)
35. Erdős, L., Knowles, A., Yau, H.-T., Yin, J.: Spectral statistics of Erdős–Rényi graphs I: local semicircle law. Ann. Prob. **41**, 2279–2375 (2013)
36. Erdős, L., Schlein, B., Yau, H.-T.: Local semicircle law and complete delocalization for Wigner random matrices. Commun. Math. Phys. **287**, 641–655 (2009)
37. Erdős, L., Yau, H.-T., Yin, J.: Rigidity of eigenvalues of generalized Wigner matrices. Adv. Math. **229**, 1435–1515 (2012)
38. Froese, R., Hasler, D., Spitzer, W.: Transfer matrices, hyperbolic geometry and absolutely continuous spectrum for some discrete Schrödinger operators on graphs. J. Funct. Anal. **230**(1), 184–221 (2006)
39. Fröhlich, J., Spencer, T.: Absence of diffusion in the Anderson tight binding model for large disorder or low energy. Commun. Math. Phys. **88**, 151–184 (1983)
40. Fyodorov, Y., Mirlin, A.: Scaling properties of localization in random band matrices: a σ-model approach. Phys. Rev. Lett. **67**(18), 2405 (1991)
41. Fyodorov, Y., Ossipov, A., Rodriguez, A.: The Anderson localization transition and eigenfunction multifractality in an ensemble of ultrametric random matrices. J. Stat. Mech. Theor Exper. **2009**(12), L12001 (2009)
42. He, Y., Knowles, A., Marcozzi, M.: Local law and complete eigenvector delocalization for supercritical Erdős–Rényi graphs. Ann. Prob. **47**(5), 3278–3302 (2019)
43. He, Y., Marcozzi, M.: Diffusion profile for random band matrices: a short proof. J. Stat. Phys. **177**(4), 666–716 (2019)
44. Klein, A.: Absolutely continuous spectrum in the Anderson model on the Bethe lattice. Math. Res. Lett. **1**(4), 399–407 (1994)
45. Lee, J.O., Schnelli, K.: Local deformed semicircle law and complete delocalization for Wigner matrices with random potential. J. Math. Phys. **54**(10), 103504 (2013)
46. Lee, J.O., Schnelli, K.: Extremal eigenvalues and eigenvectors of deformed Wigner matrices. Prob. Theor. Rel. Fields **164**(1–2), 165–241 (2016)

47. Mirlin, A.D., Fyodorov, Y.V., Dittes, F.-M., Quezada, J., Seligman, T.H.: Transition from localized to extended eigenstates in the ensemble of power-law random banded matrices. Phys. Rev. E **54**(1), 3221–3230 (1996)
48. Peled, R., Schenker, J., Shamis, M., Sodin, S.: On the Wegner orbital model. Int. Math. Res. Not. **2019**(4), 1030–1058 (2019)
49. Sarnak, P.: Arithmetic Quantum Chaos. The Schur lectures (1992)(Tel Aviv). Israel Math. Conf. Proc., vol. 8, pp. 183–236 (1995)
50. Schenker, J.: Eigenvector localization for random band matrices with power law band width. Commun. Math. Phys. **290**, 1065–1097 (2009)
51. Shcherbina, M., Shcherbina, T.: Universality for 1 D random band matrices. Preprint arXiv:1910.02999 (2019)
52. Sodin, S.: The spectral edge of some random band matrices. Ann. Math. **172**(3), 2223–2251 (2010)
53. Tarquini, E., Biroli, G., Tarzia, M.: Level statistics and localization transitions of Lévy matrices. Phys. Rev. Lett. **116**(1), 010601 (2016)
54. Tikhomirov, K., Youssef, P.: Outliers in spectrum of sparse Wigner matrices. Random Struct. Algorithms **58**(3), 517–605 (2021)
55. Trotter, H.F.: Eigenvalue distributions of large Hermitian matrices; Wigner's semicircle law and a theorem of Kac, Murdock, and Szegő. Adv. Math. **54**(1), 67–82 (1984)
56. von Soosten, P., Warzel, S.: The phase transition in the ultrametric ensemble and local stability of Dyson Brownian motion. Electr. J. Prob. **23**, 1–70 (2018)
57. Wegner, F.: Bounds on the density of states in disordered systems. Z. Phys. B Cond. Mat. **44**(1), 9–15 (1981)
58. Wigner, E.P.: Characteristic vectors of bordered matrices with infinite dimensions. Ann. Math. **62**, 548–564 (1955)
59. Yang, F., Yin, J.: Random band matrices in the delocalized phase, III: averaging fluctuations. Probab. Theory Related Fields **179**(1–2), 451–540 (2021)

Communicated by L. Erdos