



# Localized Phase for the Erdős–Rényi Graph

Johannes Alt<sup>1</sup>, Raphael Ducatez<sup>2</sup>, Antti Knowles<sup>3</sup> 

<sup>1</sup> Institute for Applied Mathematics, University of Bonn, Bonn, Germany.

E-mail: [johannes.alt@iam.uni-bonn.de](mailto:johannes.alt@iam.uni-bonn.de)

<sup>2</sup> Institut Camille Jordan, Université Lyon 1, Villeurbanne, France.

E-mail: [ducatez@math.univ-lyon1.fr](mailto:ducatez@math.univ-lyon1.fr)

<sup>3</sup> Section of Mathematics, University of Geneva, Geneva, Switzerland.

E-mail: [antti.knowles@unige.ch](mailto:antti.knowles@unige.ch)

Received: 20 June 2023 / Accepted: 16 December 2023

Published online: 19 January 2024 – © The Author(s) 2024

**Abstract:** We analyse the eigenvectors of the adjacency matrix of the Erdős–Rényi graph  $\mathbb{G}(N, d/N)$  for  $\sqrt{\log N} \ll d \lesssim \log N$ . We show the existence of a localized phase, where each eigenvector is exponentially localized around a single vertex of the graph. This complements the completely delocalized phase previously established in Alt et al. (Commun Math Phys 388(1):507–579, 2021). For large enough  $d$ , we establish a mobility edge by showing that the localized phase extends up to the boundary of the delocalized phase. We derive explicit asymptotics for the localization length up to the mobility edge and characterize its divergence near the phase boundary. The proof is based on a rigorous verification of Mott’s criterion for localization, comparing the tunnelling amplitude between localization centres with the eigenvalue spacing. The first main ingredient is a new family of global approximate eigenvectors, for which sharp enough estimates on the tunnelling amplitude can be established. The second main ingredient is a lower bound on the spacing of approximate eigenvalues. It follows from an anticoncentration result for these discrete random variables, obtained by a recursive application of a self-improving anticoncentration estimate due to Kesten.

## 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  $\ell^2$ -normalized eigenvector  $\mathbf{w} = (w_x)_{x \in [N]} \in \mathbb{R}^N$  gives rise to a probability measure  $x \mapsto w_x^2$  on the set of vertices  $[N]$ . Informally,  $\mathbf{w}$  is *delocalized* if its mass is approximately uniformly distributed throughout  $[N]$ , and *localized* if its mass is essentially concentrated on a small number of vertices.

In this paper we study the spatial localization of eigenvectors for the Erdős–Rényi graph  $\mathbb{G} \equiv \mathbb{G}(N, d/N)$ . It is the simplest model of a random graph, where each edge of the complete graph on  $N$  vertices is kept independently with probability  $d/N$ , with  $0 \leq d \leq N$ . Here,  $d \equiv d_N$  is a parameter whose interpretation is the expected degree

of a vertex. It is well known that  $\mathbb{G}$  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. 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 large degree, leaves, and isolated vertices. As long as  $d > 1$ , the graph  $\mathbb{G}$  has with high probability a unique giant component, and we shall always restrict our attention to it.

The Erdős–Rényi graph  $\mathbb{G}$  at and below criticality was proposed in [12] as a simple and natural model on which to address the question of spatial localization of eigenvectors. Its graph structure provides an intrinsic and nontrivial notion of distance, which allows for a study of the geometry of the eigenvectors. It can be interpreted as a model of quantum disorder, where the disorder arises from the random geometry of the graph. Moreover, its phase diagram turns out to be remarkably amenable to rigorous analysis.

In this paper we establish the existence of a fully *localized phase* in a region of the phase diagram of  $\mathbb{G}$  near the spectral edge. This complements the completely *delocalized phase* established in [12, 14, 42, 50]. Our results in both phases are quantitative with essentially optimal bounds.

As a consequence, for a range of critical densities  $d \asymp \log N$ , we establish a *mobility edge* separating the localized and delocalized phases. We derive the explicit behaviour of the localization length on either side of the mobility edge. In particular, we show how the localization length diverges as one approaches the mobility edge from the localized phase (see Fig. 3 below). The Erdős–Rényi graph at criticality is hence one of the very few models where a mobility edge can be rigorously established. Moreover, our proofs yield strong quantitative control of the localization length in the localized phase, as well as complete delocalization in the delocalized phase, all the way up to the mobility edge in both phases. To the best of our knowledge, this is the first time quantitative control is obtained in the vicinity of the mobility edge.

A graphical overview of the main result of this paper, and how it fits into the previous results of [12, 14], is provided by the phase diagram of Fig. 1. It depicts three phases, which are most conveniently characterized by the  $\ell^\infty$ -norm  $\|\mathbf{w}\|_\infty$  of an  $\ell^2$ -normalized eigenvector  $\mathbf{w}$ . Clearly,  $N^{-1} \leq \|\mathbf{w}\|_\infty^2 \leq 1$ , where  $\|\mathbf{w}\|_\infty^2 = 1$  corresponds to localization at a single vertex and  $\|\mathbf{w}\|_\infty^2 = N^{-1}$  to perfect delocalization. We say that an eigenvalue  $\lambda$  of the rescaled adjacency matrix  $H := A/\sqrt{d}$  with eigenvector  $\mathbf{w}$  belongs to

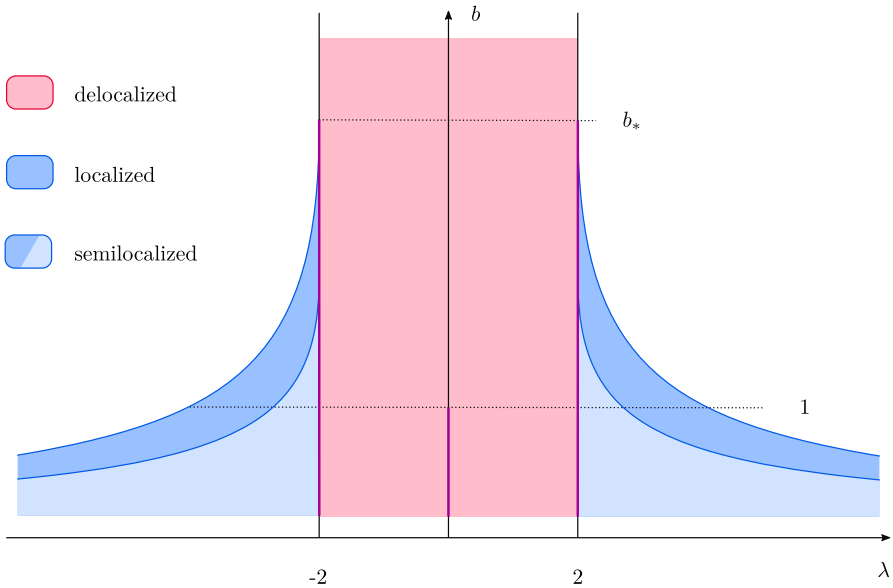
- (i) the *localized phase* if  $\|\mathbf{w}\|_\infty^2 \asymp 1$ ,
- (ii) the *delocalized phase* if  $\|\mathbf{w}\|_\infty^2 = N^{-1+o(1)}$ ,
- (iii) the *semilocalized phase* if  $\|\mathbf{w}\|_\infty^2 \geq N^{-\gamma}$  for some constant  $\gamma < 1$ .

In particular, the localized phase is a subphase of the semilocalized phase. The result of this paper is the existence of phase (i), while phases (ii) and (iii) were previously established in [12, 14].

We now briefly describe the structure of the phase diagram in Fig. 1. It is well known [75] that, as long as  $d \gg 1$ , the global eigenvalue density of  $H$  converges to the semicircle law supported in  $[-2, 2]$ . We write  $d = b \log N$  for some constant<sup>1</sup>  $b > 0$ . The localized

---

<sup>1</sup> Our results hold also for  $d \ll \log N$ , i.e. below the critical scale, but in this overview we suppose for simplicity that  $b$  is a constant.



**Fig. 1.** The phase diagram of the rescaled adjacency matrix  $A/\sqrt{d}$  of the (giant component 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  $\lambda$  in the spectrum and the vertical axis the sparseness parameter  $b$ . The spectrum is confined to the coloured region, which is split into the indicated phases. The thick purple lines correspond to phase boundaries, which are not covered by our results. (The phase boundary at energy 0 for  $b \leq 1$  is discussed in [12, 14].) For large enough  $b$ , there is a mobility edge between the localized and the delocalized phases at energies  $\pm 2$

and semilocalized phases exist only if  $b < b_*$ , where

$$b_* := \frac{1}{2 \log 2 - 1} \approx 2.59. \tag{1.1}$$

For fixed  $b < b_*$ , the spectrum splits into two disjoint parts, the delocalized phase  $(-2, 0) \cup (0, 2)$  and the semilocalized phase  $(-\lambda_{\max}, -2) \cup (2, \lambda_{\max})$ , where  $\lambda_{\max} > 2$  is an explicit function of  $b$  (see (B.3) below). A region  $(-\lambda_{\max}, -\lambda_{\text{loc}}) \cup (\lambda_{\text{loc}}, \lambda_{\max})$  near the spectral edges is in fact fully localized, where  $2 \leq \lambda_{\text{loc}} < \lambda_{\max}$ . In particular, for large enough  $b$ , the semilocalized phase consists entirely of the localized phase, i.e.  $\lambda_{\text{loc}} = 2$ . For smaller values of  $b$ , the diagram in Fig. 1 does not rule out the possibility of an eigenvector  $\mathbf{w}$  in the semilocalized phase satisfying  $\|\mathbf{w}\|_{\infty}^2 = N^{-\gamma+o(1)}$  for some constant  $\gamma \in (0, 1)$ . This latter scenario corresponds to eigenvectors that are neither fully localized nor fully delocalized,<sup>2</sup> where  $\gamma \in (0, 1)$  plays the role of an anomalous fractal dimension. In fact, it is plausible that such fractal eigenvectors occur in the semilocalized phase for small enough  $b$ ; for more details, we refer to [70] and the heuristic discussion on localization phenomena later in this subsection, as well as [15, Appendix G].

The localization-delocalization transition for  $\mathbb{G}$  described above is an example of an *Anderson transition*, where a disordered quantum system exhibits localized or delocalized states depending on the disorder strength and the location in the spectrum, corresponding to an insulator or conductor, respectively. Originally proposed in the 1950s [17] to model conduction in semiconductors with random impurities, this phenomenon

<sup>2</sup> In the literature, this phenomenon is sometimes referred to as nonergodic delocalization.

is now recognized as a general feature of wave transport in disordered media, and is one of the most influential ideas in modern condensed matter physics [3, 43, 60, 63]. It is expected to occur in great generality whenever linear waves, such as quantum particles, propagate through a disordered medium. For weak enough disorder, the stationary states are expected to be delocalized, while a strong enough disorder can give rise to localized states.

The general heuristic behind localization is the following. A disordered quantum system is characterized by its Hamiltonian  $H$ , a large Hermitian random matrix. The disorder inherent in  $H$  gives rise to spatial regions where the environment is in some sense exceptional, such as vertices of unusually large degree for the Erdős–Rényi graph.<sup>3</sup> These regions are possible localization centres, around which localized states may form. Whether they do so is captured by the following well-known rule of thumb, also known as Mott’s criterion: localization occurs whenever *the eigenvalue spacing is much larger than the tunnelling amplitude between localization centres*. The simplest illustration of this rule is for a two-state system whose Hamiltonian is the matrix  $\begin{pmatrix} a & \tau \\ \tau & b \end{pmatrix}$ . Setting the tunnelling amplitude  $\tau$  between the two sites to zero, we have an eigenvalue spacing  $|a - b|$ . Denoting by  $\mathbf{e}_1, \mathbf{e}_2$  the standard basis vectors of  $\mathbb{R}^2$ , we find that if  $|\tau| \ll |a - b|$  then the eigenvectors are approximately  $\mathbf{e}_1, \mathbf{e}_2$ , corresponding to localization, and if  $|\tau| \gg |a - b|$  then the eigenvectors are approximately  $\frac{1}{\sqrt{2}}(\mathbf{e}_1 \pm \mathbf{e}_2)$ , corresponding to delocalization.

More generally, for a disordered Hamiltonian  $H$  defined on some connected graph, a simple yet instructive way to think of the rule of thumb is to suppose that the localization centres are spatially separated, and to construct the Hamiltonian  $\widehat{H}$  from  $H$  by removing edges from the graph so as to disconnect the localization centres from each other. Hence,  $\widehat{H}$  is defined on a union of connected components, each of which is associated with a single localization centre. Denote by  $\mathcal{W}$  the set of localization centres, which are vertices in the underlying graph. A component associated with a localization centre  $x \in \mathcal{W}$  trivially gives rise to a localized state  $\mathbf{w}(x)$  of  $\widehat{H}$  with eigenvalue  $\theta(x)$ , which has the interpretation of the energy of the localization centre  $x$ . Upon putting back the removed edges, one can imagine two scenarios:

- (i) *localization*, where the eigenvectors of  $H$  remain close to the eigenvectors  $\mathbf{w}(x)$  of  $\widehat{H}$ ;
- (ii) *hybridization*, where eigenvectors  $\mathbf{w}(x)$  associated with several resonant localization centres  $x$  with similar energies  $\theta(x)$  are superimposed to form eigenvectors of  $H$ .

We refer to Fig. 2 for an illustration of this dichotomy.

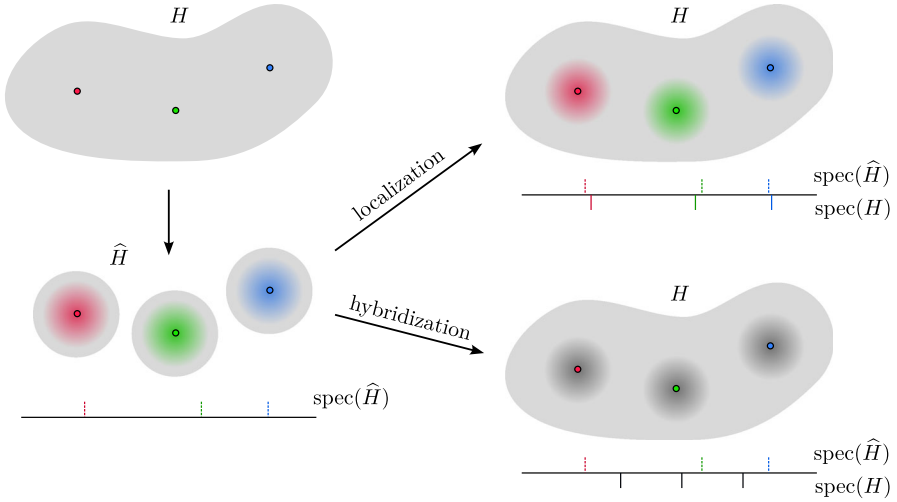
To use Mott’s criterion, we note that one possible way of quantifying the tunnelling amplitude is

$$\tau := \max_{x \in \mathcal{W}} \|(H - \theta(x))\mathbf{w}(x)\|. \tag{1.2}$$

Indeed, this expression clearly generalizes the tunnelling amplitude for the two-level system given above, expressing in general the error arising from putting back the edges of  $H$  missing from  $\widehat{H}$ , with respect to the family  $(\mathbf{w}(x))_{x \in \mathcal{W}}$ . It measures the extent to which the localized eigenvectors  $\mathbf{w}(x)$  of  $\widehat{H}$  are approximate eigenvectors of  $H$  (sometimes also called quasi-modes of  $H$ ). Under some assumptions on the localization

---

<sup>3</sup> For the Anderson model discussed below, these regions arise from an unusually large mean random potential. We note that, just like the large degree vertices for  $\mathbb{G}$ , these regions arise as a result of a collective phenomenon, requiring the conspiracy of a large number of independent random variables to yield an exceptional local configuration.



**Fig. 2.** A schematic illustration of the localization-hybridization dichotomy. A disordered Hamiltonian  $H$  defined on a graph has three localization centres whose energies are in resonance. They are indicated in red, green, and blue. The Hamiltonian  $\hat{H}$  is obtained from  $H$  by splitting the graph into disconnected components around each localization centre. This gives rise to three eigenvalues of  $\hat{H}$  associated with the three components. Each component associated with a centre  $x$  carries a localized state  $\mathbf{w}(x)$  drawn as a decaying density in the colour corresponding to that of the centre. The spectrum of  $\hat{H}$  is drawn above the real line in dotted lines that match the colour of the associated centre. Upon putting back the edges of the graph to return to  $H$ , we can have either localization or hybridization, depending on Mott’s criterion. In each case, we draw the spectrum of  $H$  below the real line. In the case of localization, the eigenvectors of  $H$  remain close to  $\mathbf{w}(x)$  and the eigenvalues are shifted by an amount that is small compared to the eigenvalue spacing. In the case of hybridization, the eigenvectors of  $H$  are delocalized over the three centres, being approximately nontrivial linear combinations of all three vectors  $\mathbf{w}(x)$

components,  $\tau$  also controls the off-diagonal part in a block-diagonal representation of  $H$  in the basis  $(\mathbf{w}(x) : x \in \mathcal{W})$ ; see Sect. 2.3 below. Mott’s criterion states that localization occurs whenever  $\tau \ll \Delta$ , where

$$\Delta := \min_{x \neq y \in \mathcal{W}} |\theta(x) - \theta(y)| \quad (1.3)$$

is the eigenvalue spacing. Otherwise, we expect hybridization. The main difficulty in establishing localization is therefore to control *resonances*, i.e. pairs of vertices  $x, y$  such that  $|\theta(x) - \theta(y)|$  is small, and to rule out hybridization. Although this picture is helpful in gaining intuition about localization, in many instances, such as in the proof for the Erdős–Rényi graph in this paper,<sup>4</sup> it is but a rough caricature and the true picture is considerably more subtle (as we explain later in this subsection and in more detail in Sect. 2.3).

The semilocalized phase of the Erdős–Rényi graph from [12] is a phase where the eigenvectors are concentrated around a small number of localization centres, but where hybridization cannot be ruled out. As shown in [12], the set of all possible localization centres of  $\mathbb{G}$  corresponds to vertices  $x \in [N]$  whose *normalized degree*

$$\alpha_x := \sum_{y \in [N]} H_{xy}^2 = \frac{1}{d} \sum_{y \in [N]} A_{xy}$$

<sup>4</sup> Or for the Anderson model e.g. in [45].

is greater than 2. The energy  $\theta(x)$  of a localization centre  $x$  is approximately equal to  $\Lambda(\alpha_x)$ , where we introduced the function  $\Lambda : [2, \infty) \rightarrow [2, \infty)$  through

$$\Lambda(\alpha) := \frac{\alpha}{\sqrt{\alpha - 1}}. \tag{1.4}$$

As shown in [13] (see also [71]), there is a one-to-one correspondence between eigenvalues  $\lambda > 2$  in the semilocalized phase and vertices  $x$  of normalized degree  $\alpha_x \geq 2$  given by  $\lambda = \Lambda(\alpha_x) + o(1)$ . An eigenvalue  $\lambda$  in the semilocalized phase has an eigenvector that is almost entirely concentrated in small balls around the set of vertices  $\mathcal{W}_\lambda := \{x \in [N] : \Lambda(\alpha_x) = \lambda + o(1)\}$  in resonance with the energy  $\lambda$  [12]. The size of the set of resonant vertices  $\mathcal{W}_\lambda$  is comparable to the density of states, equal to  $N^{\rho_b(\lambda)+o(1)}$  for an explicit exponent  $\rho_b(\lambda) < 1$  given in (B.4) in Appendix B below. Hence, owing to the small size of the set of resonant vertices  $\mathcal{W}_\lambda$ , the semilocalized phase is sharply distinguished from the delocalized phase. However, the key issue of controlling resonances and ruling out hybridization is not addressed in [12] (see Fig. 2).

In this paper we prove localization by ruling out hybridization among the resonant vertices. Our result holds for the largest and smallest  $N^\mu$  eigenvalues for  $\mu < \frac{1}{24}$ . For small enough  $\mu$ , our obtained rate of exponential decay is optimal all the way up to radii of the order of the diameter of the graph. The bound  $\frac{1}{24}$  is not optimal, and we expect that by a refinement of the method developed in this paper it can be improved; for the sake of keeping the argument reasonably simple, we refrain from doing so here. Heuristic arguments suggest that the optimal upper bound for  $\mu$  is  $\frac{1}{4}$ ; see [15, Appendix G] as well as [70].

At this point it is helpful to review the previous works [12, 16], which addressed the tunnelling amplitude (1.2) in the above simple picture of localization based on disjoint neighbourhoods of localization centres. In [12], the estimate  $\tau \lesssim d^{-1/2}$  was established in the entire semilocalized phase, while in [16] it was improved to  $\tau \lesssim d^{-3/2}$  at the spectral edge. In fact, here we argue that the best possible bound on  $\tau$  in terms of the local approximate eigenvectors  $\mathbf{w}(x)$  introduced above is  $e^{-c \frac{\log N}{\log d}}$  for some constant  $c > 0$ . To see this, we recall from [12] that the vector  $\mathbf{w}(x)$  is exponentially decaying around  $x$  at some fixed rate  $C > 0$  depending on  $b$ . Thus, the best possible estimate for  $\tau$  arising from the exponential decay is  $e^{-Cr}$ , where  $r \asymp \text{diam}(\mathbb{G}) = \frac{\log N}{\log d} (1 + o(1))$  (see [32]).

As for the eigenvalue spacing (1.3), it was not addressed at all in [12]. In [16], it was estimated at the spectral edge as  $\Delta \geq d^{-1-\varepsilon}$  with high probability for any constant  $\varepsilon > 0$ . Combined with the bound  $\tau \lesssim d^{-3/2}$  at the spectral edge obtained in [16], one finds that Mott’s criterion is satisfied at the spectral edge. This observation was used in [16] to prove localization for the top  $O(1)$  eigenvectors.

In the interior of the localized phase, the eigenvalue spacing at energy  $\lambda$  is typically of order  $N^{-\rho_b(\lambda)}$ , where  $\rho_b(\lambda) > 0$  is an exponent defined in (B.4) below. This is much smaller than the best possible estimate on the tunnelling amplitude,  $e^{-c \frac{\log N}{\log d}} = N^{-o(1)}$ . Hence, Mott’s criterion is never satisfied inside the localized phase, and thus the simple picture based on local approximate eigenvectors cannot be used to establish localization. In this paper we therefore introduce a new setup for proving localization.

The first key idea of our proof is to abandon the above simple picture of localization, and to replace the *local* approximate eigenvectors  $\mathbf{w}(x)$  by *global* approximate eigenvectors, denoted by  $\mathbf{u}(x)$ , which approximate eigenvectors of  $H$  much more accurately and therefore lead to a much smaller tunnelling amplitude (1.2). To define  $\mathbf{u}(x)$ , we consider the graph obtained from  $\mathbb{G}$  by removing all localization centres except  $x$ ; we

denote by  $\lambda(x)$  the second largest eigenvalue of its adjacency matrix, and by  $\mathbf{u}(x)$  the associated eigenvector. The latter is localized around the vertex  $x$ . Crucially, the quantity (1.2) with  $\mathbf{w}(x)$  replaced by  $\mathbf{u}(x)$  can now be estimated by a *polynomial* error  $N^{-\zeta}$  for some  $\zeta > 0$ .

The price to pay for passing from local approximate eigenvectors  $\mathbf{w}(x)$  to global approximate eigenvectors  $\mathbf{u}(x)$  is a breakdown of orthogonality. Indeed, the vectors  $\mathbf{u}(x)$  have a nonzero overlap, and a significant difficulty in our proof is to control these overlaps and various resulting interactions between localization centres.

To complete the verification of Mott’s criterion, we need to establish a polynomial lower bound  $\min_{x \neq y \in \mathcal{W}} |\lambda(x) - \lambda(y)| \geq N^{-\eta}$  for some  $\eta < \zeta$ . Clearly, the left-hand side cannot be larger than the eigenvalue spacing  $N^{-\rho_b(\lambda)}$  at the energy  $\lambda$  we are considering, which yields the necessary bound  $\eta > \rho_b(\lambda) > 0$ . Hence, we require an anticoncentration result for the eigenvalue difference  $\lambda(x) - \lambda(y)$  on a polynomial scale. Owing to the discrete law of  $\mathbb{G}$  (a product of independent Bernoulli random variables), methods based on smoothness such as Wegner estimates [73] are not available, and obtaining strong enough anticoncentration is the most involved part of our proof. Our basic strategy is to perform a recursive resampling of neighbourhoods of increasingly large balls around  $y$ . At each step, we derive a *concentration* bound for  $\lambda(x)$  and an *anticoncentration* bound for  $\lambda(y)$ . The key tool for the latter is a self-improving version, due to Kesten [57], of a classical anticoncentration result of Doebelin, Lévy, Kolmogorov, and Rogozin. In order to obtain sufficiently strong anticoncentration, it is crucial to perform the recursion up to radii comparable to the diameter of  $\mathbb{G}$ .

We conclude this overview with a survey of related results. The eigenvalues and eigenvectors of the Erdős–Rényi graph have been extensively studied in the denser regime  $d \gg \log N$ . Complete eigenvector delocalization for  $d \gg \log N$  was established in [42, 50]. The local spectral statistics in the bulk were proved to follow the universal GOE statistics in [40, 52] for  $d \geq N^{o(1)}$ . At the spectral edge, the local spectral statistics were proved to be Tracy–Widom for  $d \gg N^{1/3}$  [40, 62], to exhibit a transition from Tracy–Widom to Gaussian at  $d \asymp N^{1/3}$  [53], and to be Gaussian throughout the regime  $N^{o(1)} \leq d \ll N^{1/3}$  [49, 53]. In fact, in the latter regime the Tracy–Widom statistics were recovered in [55, 61] after subtraction of an explicit random shift.

The random  $d$ -regular graph is another canonical model of sparse random graphs. Owing to the regularity constraint, it is much more homogeneous than the Erdős–Rényi graph and only exhibits a delocalized phase. The eigenvectors were proved to be completely delocalized for all  $d \geq 3$  in [21, 22, 54], and the local spectral statistics in the bulk were shown to follow GOE statistics for  $d \geq N^{o(1)}$  in [19] and at the edge Tracy–Widom statistics for  $N^{o(1)} \leq d \ll N^{1/3}$  in [20, 56] or for  $N^{2/3} \ll d \leq N/2$  in [48].

Anderson transitions have been studied in a variety of models. The archetypal example is the tight-binding, or Anderson, model on  $\mathbb{Z}^d$  [1, 2, 11, 18]. In dimensions  $d \leq 2$ , all eigenvectors of the Anderson model are expected to be localized, while for  $d \geq 3$  a coexistence of localized and delocalized phases, separated by a mobility edge, is expected for small enough disorder. So far, only the localized phase of the Anderson model has been understood rigorously, starting from the landmark works [7, 45]; see for instance [11] for a recent survey.

Although a rigorous understanding of the metal-insulator transition for the Anderson tight-binding model is still elusive, some progress has been made for random band matrices. Random band matrices [30, 46, 64, 74] interpolate between the Anderson model and mean-field Wigner matrices. They retain the  $d$ -dimensional structure of the Anderson model but have proved more amenable to rigorous analysis. They are conjectured [46]



to have a similar phase diagram as the Anderson model in dimensions  $d \geq 3$ . For  $d = 1$  much has been understood both in the localized [31, 33, 65, 66] and the delocalized [27–29, 35–39, 41, 51, 67, 68, 79] phases. For large enough  $d$ , recent progress in the delocalized phase has been made in [76–78]. A simplification of band matrices is the ultrametric ensemble [47], where the Euclidean metric of  $\mathbb{Z}^d$  is replaced with an ultrametric arising from a tree structure. For this model, a phase transition was rigorously established in [72].

Another modification of the  $d$ -dimensional Anderson model is the Anderson model on the Bethe lattice, an infinite regular tree corresponding to the case  $d = \infty$ . For it, the existence of a delocalized phase was shown in [8, 44, 58]. In [9, 10] it was shown that for unbounded random potentials the delocalized phase exists for arbitrarily weak disorder. The underlying mechanism is resonant delocalization, in which 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 Mott’s criterion is violated. As a consequence, the eigenvectors hybridize.

Heavy-tailed Wigner matrices, or Lévy matrices, whose entries have  $\alpha$ -stable laws for  $0 < \alpha < 2$ , were proposed in [34] as a simple model that exhibits a transition in the localization of its eigenvectors; we refer to [6] for a summary of the predictions from [34, 69]. In [24, 25] it was proved that eigenvectors are weakly delocalized for energies in a compact interval around the origin, and for  $0 < \alpha < 2/3$  eigenvectors are weakly localized for energies far enough from the origin. In [6] full delocalization, as well as GOE local eigenvalue statistics, were proved in a compact interval around the origin, and in [5] the law of the eigenvector components was computed. Recently, by comparison to a limiting tree model, a mobility edge was established in [4] for  $\alpha$  near 0 or 1.

*Conventions* Every quantity that is not explicitly called *fixed* or a *constant* is a sequence depending on  $N$ . We use the customary notations  $o(\cdot)$  and  $O(\cdot)$  in the limit  $N \rightarrow \infty$ . For nonnegative  $X, Y$ , if  $X = O(Y)$  then we also write  $X \lesssim Y$ , and if  $X = o(Y)$  then we also write  $X \ll Y$ . Moreover, we write  $X \asymp Y$  to mean  $X \lesssim Y$  and  $Y \lesssim X$ . We say that an event  $\Omega$  holds with *high probability* if  $\mathbb{P}(\Omega) = 1 - o(1)$ . Throughout this paper every eigenvector is assumed to be normalized in  $\ell^2([N])$ . Finally, we use  $\kappa \in (0, 1)$  to denote a small positive constant, which is used to state assumptions and definitions; a smaller  $\kappa$  always results in a weaker condition.

*1.2. Results.* Let  $\mathbb{G} \equiv \mathbb{G}(N, d/N)$  be the Erdős-Rényi graph with vertex set  $[N]$  and edge probability  $d/N$  for  $0 \leq d \leq N$ . 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. Define the rescaled adjacency matrix

$$H := A/\sqrt{d}.$$

We always assume that  $d$  satisfies

$$\sqrt{\log N} \log \log N \ll d \leq 3 \log N. \tag{1.5}$$

Owing to the nonzero expectation of  $H$ , it is well known that the largest eigenvalue of  $H$ , denoted by  $\lambda_1(H)$ , is an outlier separated from the rest of the spectrum (see e.g. Proposition 3.4 (iv) below), and we shall always discard it from our discussion. The lower bound of (1.5) is made for convenience, to ensure that  $\lambda_1(H)$  is separated from



the bulk spectrum<sup>5</sup>; the upper bound of (1.5) is made without loss of generality, since for  $d \geq 3 \log N$  the localized phase does not exist and the entire spectrum is known to belong to the delocalized phase [12, 14].

We denote by  $B_r(x)$  (respectively by  $S_r(x)$ ) the closed ball (respectively the sphere) of radius  $r$  with respect to the graph distance of  $\mathbb{G}$  around the vertex  $x$ . We refer to Sect. 2.1 below for a full account of notations used throughout this paper.

The localized phase is characterized by a threshold  $\alpha^*$  defined, for any fixed  $\kappa \in (0, 1)$  and  $\mu \in [0, 1]$ , as

$$\alpha^* \equiv \alpha^*(\mu) := \max \left\{ \inf \{ \alpha > 0 : \mathbb{P}(\alpha_1 \geq \alpha) \leq N^{\mu-1} \}, 2 + \kappa \right\}. \quad (1.6)$$

We refer to Appendix B below for the basic qualitative properties of  $\alpha^*$  as well as a graph. We shall show exponential localization for any eigenvector with eigenvalue  $\lambda$  satisfying the condition

$$\lambda \neq \lambda_1(H), \quad |\lambda| \geq \Lambda(\alpha^*(\mu)) + \kappa, \quad (1.7)$$

for sufficiently small  $\mu > 0$ . In particular, the number of eigenvalues  $\lambda$  satisfying (1.7) is with high probability  $N^{\mu+o(1)}$  as  $\kappa \rightarrow 0$ ; see Remark 1.2 below.

*1.2.1. Exponential localization* Our main result is the following theorem. We recall that by convention all eigenvectors are normalized.

**Theorem 1.1** (Exponential localization). *Suppose that (1.5) holds. Fix  $\kappa \in (0, 1)$  and  $\mu \in (0, 1/24)$ . Then there is a constant  $c \in (0, 1)$  depending on  $\kappa$  such that, with high probability, for any eigenvector  $\mathbf{w} = (w_x)_{x \in [N]}$  of  $H$  with eigenvalue  $\lambda$  satisfying (1.7), there exists a unique vertex  $x \in [N]$  with  $\alpha_x > \alpha^*(\mu)$  such that*

$$w_x^2 = \frac{\alpha_x - 2}{2(\alpha_x - 1)} + o(1), \quad \|\mathbf{w}\|_{B_i(x)^c} \lesssim \sqrt{\alpha_x}(1 - c)^i, \quad (1.8)$$

for all  $i \in \mathbb{N}$  with  $1 \leq i \leq \frac{1}{6} \frac{\log N}{\log d}$ .

*Remark 1.2* (Eigenvalue locations). The eigenvalue  $\lambda$  of the eigenvector  $\mathbf{w}$  and the associated vertex  $x$  from Theorem 1.1 satisfy  $|\lambda| = \Lambda(\alpha_x) + o(1)$  with high probability. This follows from (2.4) and (3.3) below.

The eigenvalue locations in the localized phase were previously studied in [13, Theorem 2.1] (see also [12, Theorem 1.7]). In particular, if  $d > b_* \log N$  then the localized phase does not exist (see [13, Remark 2.5]) and there is no eigenvalue of  $H$  satisfying (1.7). Conversely, if  $d \leq (b_* - \varepsilon) \log N$  for some constant  $\varepsilon > 0$  then for small enough  $\kappa > 0$  there is a polynomial number of eigenvalues satisfying (1.7), by [12, Theorem 1.7]. By the same argument, if  $d/\log N$  is small enough, then with high probability  $N^{\mu+o(1)}$  eigenvalues of  $H$  satisfy (1.7) as  $N \rightarrow \infty$  and  $\kappa \rightarrow 0$ .

*Remark 1.3* (Conditions in Theorem 1.1). The exponential decay in Theorem 1.1 holds up to the scale  $\frac{\log N}{\log d}$  of the diameter of  $\mathbb{G}$ . The upper bound  $1/24$  and the factor  $1/6$  are not optimal (see the discussion in [15, Appendix G]), and they can be improved with some extra effort, which we however refrain from doing here.

<sup>5</sup> In fact, to ensure the separation, the weaker lower bound  $d \gg \sqrt{\log N / \log \log N}$  would be sufficient; see [59] and [16, Remark 1.4] for detailed explanations. We impose the slightly stronger lower bound in (1.5) for convenience, as it allows us to directly import results from [12] that were proved under this condition.

*Remark 1.4* (Optimal exponential decay). If  $\mu$  is sufficiently small then the rate  $1 - c$  of exponential decay from (1.8) can be made explicit. Suppose (1.5). Then for each small enough constant  $\varepsilon > 0$  and small enough constant  $\mu > 0$ , depending on  $\varepsilon$ , with high probability, for any eigenvector  $\mathbf{w}$  of  $H$  with eigenvalue  $\lambda$  satisfying (1.7), there exists a unique vertex  $x \in [N]$  with  $\alpha_x > \alpha^*(\mu)$  such that

$$\|\mathbf{w}|_{B_i(x)^c}\| \lesssim \sqrt{\alpha_x} \left( \frac{1 + O(\varepsilon)}{\sqrt{\alpha_x - 1}} \right)^{i+1} \tag{1.9}$$

for each  $i \in \mathbb{N}$  satisfying  $i \ll \frac{\log N}{\log d} \frac{1}{\log \frac{10 \log N}{d}}$ . This follows from (2.3) and Proposition 4.3 below.

The rate of decay in (1.9) is optimal up to the error term  $O(\varepsilon)$ . Indeed, by Theorem 1.6 and the explicit form (1.10)–(1.11) below, we find that

$$\|\mathbf{w}|_{B_i(x)^c}\| \asymp \sqrt{\alpha_x} \left( \frac{1}{\sqrt{\alpha_x - 1}} \right)^{i+1} + o(1).$$

for any fixed  $i \in \mathbb{N}$ . In particular, (1.9) improves the rate  $\frac{q^i}{(1-q)^2}$  with  $q = (2 + o(1)) \frac{\sqrt{\alpha_x - 1}}{\alpha_x}$  obtained in [16, Theorem 1.7] at the spectral edge, corresponding to  $\alpha_x = (1 + o(1))\alpha^*(0)$ .

*1.2.2. Geometric structure of eigenvectors* Next, we describe the precise geometric structure of the eigenvectors in the localized phase. For any vertex  $x$  with  $\alpha_x > 2$  and radius  $r \in \mathbb{N}^*$ , we shall define two local vectors,  $\mathbf{w}_r(x)$  and  $\mathbf{v}_r(x)$ , which depend only on  $\mathbb{G}$  in the ball  $B_r(x)$ . If  $\mathbf{w}$  is an eigenvector of  $H$  as in Theorem 1.1 with associated vertex  $x$ , then  $\mathbf{w}$  will be well approximated by  $\mathbf{w}_r(x)$  and  $\mathbf{v}_r(x)$  for suitably chosen  $r \gg 1$ .

To define these local vectors, we need the following definitions. Let  $r \in \mathbb{N}^*$ . For  $\alpha > 2$  define the positive sequence  $(u_i(\alpha))_{i=0}^{r-1}$  through

$$u_1(\alpha) := \left( \frac{\alpha}{\alpha - 1} \right)^{1/2} u_0(\alpha), \quad u_i(\alpha) := \left( \frac{1}{\alpha - 1} \right)^{(i-1)/2} u_1(\alpha) \quad (1 \leq i \leq r - 1). \tag{1.10}$$

We normalize the sequence by choosing  $u_0(\alpha) > 0$  such that  $\sum_{i=0}^{r-1} u_i(\alpha)^2 = 1$ .

**Definition 1.5** (*Localization profile vectors  $\mathbf{w}_r(x)$  and  $\mathbf{v}_r(x)$* ). Let  $r \in \mathbb{N}^*$  and  $x \in [N]$ .

- (i) Denote by  $\mathbf{w}_r(x)$  an eigenvector of  $H|_{B_r(x)}$  associated with its largest eigenvalue, chosen so that its value at  $x$  is nonnegative. Here  $H|_{B_r(x)}$  denotes the matrix  $H$  restricted to the vertices in  $B_r(x)$  (See Sect. 2.1 below).
- (ii) For  $\alpha_x > 2$  and  $(u_i(\alpha))_{i=0}^{r-1}$  as in (1.10), define

$$\mathbf{v}_r(x) := \sum_{i=0}^{r-1} u_i(\alpha_x) \frac{\mathbf{1}_{S_i(x)}}{|S_i(x)|^{1/2}}, \tag{1.11}$$

where  $\mathbf{1}_{S_i(x)}$  denotes the indicator function of the sphere  $S_i(x)$ .

Note that  $\mathbf{w}_r(x)$  is unique by the Perron-Frobenius theorem for irreducible matrices with nonnegative entries.

**Theorem 1.6** (Localization profile). *Suppose that (1.5) holds and fix  $\kappa \in (0, 1)$  and  $\mu \in (0, 1/24)$ . With high probability, for any eigenvector  $\mathbf{w}$  of  $H$  with eigenvalue  $\lambda$  satisfying (1.7), there exists a unique vertex  $x \in [N]$  with  $\alpha_x > \alpha^*(\mu)$  such that<sup>6</sup>*

$$\mathbf{w} = \mathbf{w}_r(x) + o(1) = \mathbf{v}_r(x) + o(1) \tag{1.12}$$

for each  $r \in \mathbb{N}^*$  satisfying  $\log d \ll r \leq \frac{1}{6} \frac{\log N}{\log d}$ . Here,  $o(1)$  is meant with respect to the Euclidean norm on  $\mathbb{R}^N$ .

In particular,  $\mathbf{w}$  has locally the radial exponentially decaying structure of  $\mathbf{v}_r(x)$ .

*1.2.3. Mobility edge and localization length* Next, we combine the results of this paper with those obtained for the delocalized phase in [12, 14] to establish a mobility edge at  $\pm 2$  for certain values of  $d$ , and analyse the structure of the eigenvectors quantitatively in the vicinity of the mobility edge.

**Theorem 1.7** (Mobility edge). *Fix  $\kappa > 0$  and suppose that  $\frac{23}{24} + \kappa \leq \frac{d}{b_* \log N} \leq 1 - \kappa$ . Then, with high probability, for any eigenvector  $\mathbf{w}$  of  $H$  with eigenvalue  $\lambda \neq \lambda_1(H)$  we have the following dichotomy.*

- (i) (Localized phase) *If  $|\lambda| \geq 2 + \kappa$  then  $\mathbf{w}$  is exponentially localized as in (1.8) and (1.12).*
- (ii) (Delocalized phase) *If  $|\lambda| \leq 2 - \kappa$  then  $\mathbf{w}$  is completely delocalized in the sense that*

$$\|\mathbf{w}\|_\infty^2 \leq N^{-1+o(1)}. \tag{1.13}$$

Both phases in Theorem 1.7 are nonempty under the assumption on  $d$ ; see Sect. 1.1 and [13, Remark 2.5]. Theorem 1.7 establishes a dichotomy because (1.8) and (1.13) are mutually exclusive, since  $\|\mathbf{w}\|_\infty^2 \geq w_x^2 = \frac{\alpha_x - 2}{2(\alpha_x - 1)} + o(1) \gtrsim 1$  if  $\alpha_x > \alpha^*(\mu) \geq 2 + \kappa$ .

Next, we investigate the spatial extent of the eigenvectors near the mobility edge. To that end, we use the following notion of localization length. With each normalized vector  $\mathbf{w}$  we associate the length

$$\ell(\mathbf{w}) := \min_{x \in [N]} \sum_{y \in [N]} d(x, y) w_y^2, \tag{1.14}$$

where  $d(x, y)$  denotes the distance from  $x$  to  $y$  in the graph  $\mathbb{G}$ . Regarding  $y \mapsto w_y^2$  as a probability measure on  $[N]$ , the quantity  $\ell(\mathbf{w})$  expresses the minimal expected distance from a reference vertex  $x$ . The minimizing vertex  $x$  has the interpretation of a localization centre for  $\mathbf{w}$ .

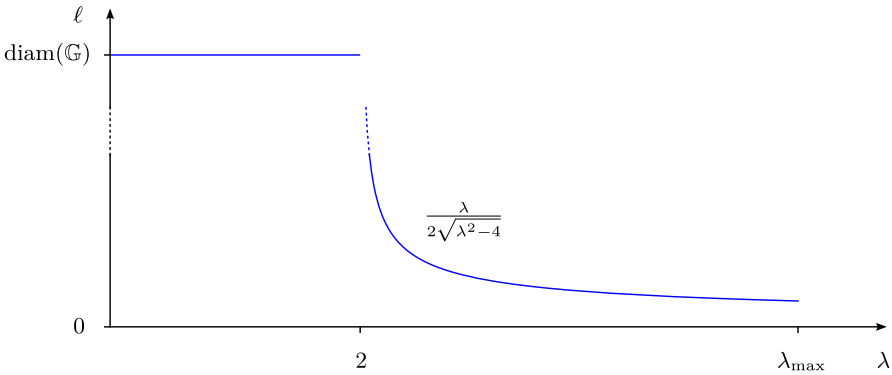
Denote by  $\text{diam}(\mathbb{G})$  the diameter<sup>7</sup> of  $\mathbb{G}$ . It is a classical fact [32] that with high probability  $\text{diam}(\mathbb{G}) = \frac{\log N}{\log d} (1 + o(1))$  as long as  $d \gg 1$ .

**Theorem 1.8** (Localization length). *Fix  $\kappa > 0$  and suppose that  $\frac{23}{24} + \kappa \leq \frac{d}{b_* \log N} \leq 1 - \kappa$ . Then, with high probability, for any eigenvector  $\mathbf{w}$  of  $H$  with eigenvalue  $\lambda \neq \lambda_1(H)$  we have*

$$\ell(\mathbf{w}) = \begin{cases} \frac{|\lambda|}{2\sqrt{\lambda^2 - 4}} + o(1) & \text{if } |\lambda| \geq 2 + \kappa \\ \text{diam}(\mathbb{G})(1 + o(1)) & \text{if } |\lambda| \leq 2 - \kappa. \end{cases} \tag{1.15}$$

<sup>6</sup> We assume that the sign of  $\mathbf{w}$  is chosen suitably.

<sup>7</sup> We recall that the diameter of a connected graph is the length of its longest geodesic. If the graph is disconnected, then its diameter is the maximal diameter of its connected components.



**Fig. 3.** An illustration of the behaviour of the localization length (1.14) around the mobility edge, established in Theorem 1.8. We plot the asymptotic localization length  $\ell$  of an eigenvector with eigenvalue  $\lambda$  as a function of  $\lambda$ . Here  $\frac{d}{b_* \log N}$  is a fixed number in  $[\frac{23}{24} + \kappa, 1 - \kappa]$ . The spectrum is asymptotically given by the interval  $[-\lambda_{\max}, \lambda_{\max}]$ . We only draw a portion of the spectrum near the right edge. Below the mobility edge 2, the localization length is  $\text{diam}(\mathbb{G}) = \frac{\log N}{\log d} (1 + o(1))$ . Above the mobility edge 2, the localization length is finite and diverges as one approaches the mobility edge

*Remark 1.9.* By the same proof, the first estimate of (1.15) holds also for all eigenvectors satisfying the conditions of Theorem 1.1. Moreover, the constant  $\frac{23}{24}$  is not optimal and can be reduced with some extra effort.

Theorem 1.8 shows that the localization length diverges as one approaches the mobility edge from the localized phase, and that it equals the diameter of the graph in the delocalized phase. See Fig. 3 for an illustration.

*1.2.4. Eigenfunction correlator and dynamical localization* Finally, as a consequence of Theorem 1.1, we control quantities commonly used to characterize Anderson localization (see e.g. [11, Section 1.4]). In particular, we establish exponential decay of the eigenfunction correlator and dynamical localization.

**Corollary 1.10.** *Suppose (1.5). Then there is a constant  $\mu > 0$  such that for each fixed  $\kappa \in (0, 1)$ , there exist constants  $c, C > 0$  depending only on  $\kappa$  such that the following holds with high probability. Let  $J \subset [\Lambda(\alpha^*(\mu)) + \kappa, \sqrt{d}/2]$  be an interval with associated spectral projection  $\Pi_J(H)$ . For any  $x \in [N]$ , any measurable function  $F: \mathbb{R} \rightarrow \mathbb{C}$  satisfying  $\|F\|_\infty \leq 1$ , and any  $r \geq 0$ , we have*

$$\|(\Pi_J(H)F(H)\mathbf{1}_x)|_{B_r(x)^c}\| \leq Ce^{-cr}. \tag{1.16}$$

*In particular, denoting by  $\mathbf{w}_\lambda$  the normalized eigenvector of  $H$  associated with  $\lambda \in \text{spec}(H)$ , the eigenfunction correlator satisfies the estimate*

$$\sum_{\lambda \in \text{spec}(H) \cap J} |(\mathbf{1}_x, \mathbf{w}_\lambda)(\mathbf{w}_\lambda, \mathbf{1}_y)| \leq Ce^{-cd(x,y)}, \quad x, y \in [N], \tag{1.17}$$

*and we have dynamical localization,*

$$\sup_{t \in \mathbb{R}} |(\mathbf{1}_y, \Pi_J(H)e^{-itH}\mathbf{1}_x)| \leq Ce^{-cd(x,y)}, \quad x, y \in [N]. \tag{1.18}$$

*Remark 1.11.* By a close inspection of the proof in [15, Appendix F] (using that all error probabilities are polynomially small in  $N$ ), we note that the estimates (1.17) and (1.18) hold also in expectation, provided one multiplies both sides by the factor  $e^{c d(x,y)}$ .

*Structure of the paper* We conclude this section with a short summary of the structure of the paper. In Sect. 2, we collect a few basic notations, then state the three core propositions of the paper: Proposition 2.2, which gives exponential decay of the approximate eigenvectors, Proposition 2.3, which compares the approximate eigenvalues with the true eigenvalues, and Proposition 2.4, which estimates the spacing between neighbouring approximate eigenvalues. After stating them, we use them to deduce Theorem 1.1 in Sect. 2.2. Then, we sketch the proofs of these three propositions in Sect. 2.3. Theorems 1.6, 1.7, and 1.8 are proved in the short Sects. 2.4, 2.5, and 2.6, respectively. Section 3 is devoted to preliminary results on the graph  $\mathbb{G}$ , its spectrum, and its Green function. Sections 4, 5 and 6 are devoted to the proofs of Propositions 2.2, 2.3, and 2.4, respectively. In the appendices, we collect some auxiliary results and basic tools.

## 2. Proof of Main Results

The rest of the paper is devoted to the proofs of Theorems 1.1, 1.6, 1.7, and 1.8, as well as Corollary 1.10. The former four are proved in this section, while Corollary 1.10 is proved in [15, Appendix F].

Throughout,  $\kappa \in (0, 1)$  denotes an arbitrary positive constant.

*2.1. Basic notations.* 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  $X$ . For  $X \subset [N]$  we write  $X^c := [N] \setminus X$ . We use  $\mathbb{1}_\Omega$  to denote the indicator function of an event  $\Omega$ .

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  $\text{supp } \mathbf{v} := \{x \in [N] : v_x \neq 0\}$  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  $X \subset [N]$  we set  $\mathbf{v}|_X := (v_x \mathbb{1}_{x \in X})_{x \in [N]}$ . For any  $x \in [N]$ , we define the standard basis vector  $\mathbf{1}_x := (\delta_{xy})_{y \in [N]} \in \mathbb{R}^N$ , so that  $w_x = \langle \mathbf{1}_x, \mathbf{w} \rangle$ . 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 denote by  $d(x, y)$  the distance between the vertices  $x, y \in [N]$  with respect to the graph  $\mathbb{G}$ , i.e. the number of edges in the shortest path connecting  $x$  and  $y$ . For  $r \in \mathbb{N}$  and  $x \in [N]$ , we denote by  $B_r(x) := \{y \in [N] : d(x, y) \leq r\}$  the closed ball of radius  $r$  around  $x$ , and by  $S_r(x) := \{y \in [N] : d(x, y) = r\}$  the sphere of radius  $r$  around the vertex  $x$ . For  $X \subset [N]$  we denote by  $\mathbb{G}|_X$  the subgraph on  $X$  induced by  $\mathbb{G}$ .

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 an  $N \times N$  Hermitian matrix  $M$ , we denote by  $\lambda_1(M) \geq \lambda_2(M) \geq \dots \geq \lambda_N(M)$  the ordered eigenvalues of  $M$ . For an  $N \times N$  matrix  $M \in \mathbb{R}^{N \times N}$  and a subset  $X \subset [N]$ , we introduce the  $N \times N$  matrices  $M|_X := (M_{xy} \mathbb{1}_{x,y \in X})_{x,y \in [N]}$  as well as  $M^{(X)} := M|_{X^c}$  with entries  $M_{xy}^{(X)} = M_{xy} \mathbb{1}_{x,y \notin X}$ .

*2.2. Exponential localization: proof of Theorem 1.1.* In this section, after introducing some notation and stating the core propositions of the proof, we use them to prove our

main result, Theorem 1.1. Recalling the definition of  $\Lambda$  from (1.4), we introduce the  $\mu$ -dependent sets

$$\mathcal{V} := \{x \in [N] : \alpha_x \geq \alpha^*(\mu)\}, \tag{2.1}$$

$$\mathcal{W} := \{x \in \mathcal{V} : \Lambda(\alpha_x) \geq \Lambda(\alpha^*(\mu)) + \kappa/2\}. \tag{2.2}$$

By definition,  $\mathcal{W} \subset \mathcal{V}$ .

The following definition introduces the fundamental approximate eigenvalues and eigenvectors underlying our proof.

**Definition 2.1** ( $\lambda(x)$  and  $\mathbf{u}(x)$ ). For any  $x \in \mathcal{W}$ , we abbreviate  $\lambda(x) := \lambda_2(H^{(\mathcal{V} \setminus \{x\})})$ . Moreover, we denote by  $\mathbf{u}(x)$  a normalized eigenvector of  $H^{(\mathcal{V} \setminus \{x\})}$  with eigenvalue  $\lambda(x)$  satisfying  $\langle \mathbf{1}_x, \mathbf{u}(x) \rangle \geq 0$ .

As we shall see, with high probability  $\lambda(x)$  is a simple eigenvalue and hence  $\mathbf{u}(x)$  is unique (see Corollary 3.6 below).

The proof of Theorem 1.1 consists of three main steps, which are the content of the three following propositions. The next proposition states that  $\mathbf{u}(x)$  has the exponential decay claimed in Theorem 1.1 for  $\mathbf{w}$ . It is proved in Sect. 4 below.

**Proposition 2.2** (Exponential decay of  $\mathbf{u}(x)$ ). *Suppose that (1.5) holds. Then there is a constant  $c \in (0, 1)$  such that, for each fixed  $\mu \in [0, 1/3)$ , with high probability, for each  $x \in \mathcal{W}$ ,*

$$\langle \mathbf{1}_x, \mathbf{u}(x) \rangle = \sqrt{\frac{\alpha_x - 2}{2(\alpha_x - 1)}} + o(1), \quad \|\mathbf{u}(x)|_{B_i(x)^c}\| \lesssim \sqrt{\alpha_x}(1 - c)^i$$

for all  $i \in \mathbb{N}$  satisfying  $1 \leq i \leq \min\{\frac{1}{5} - \frac{\mu}{4}, \frac{1}{3} - \mu\} \frac{\log N}{\log d} - 2$ .

In the proof of Theorem 1.1, the next two propositions will be used to conclude that any eigenvector of  $H$  whose associated eigenvalue satisfies (1.7) is close to  $\mathbf{u}(x)$  for some  $x \in \mathcal{W}$ . Given Proposition 2.2, this will directly imply Theorem 1.1.

The next proposition, Proposition 2.3, states that  $\lambda(x)$  and  $\mathbf{u}(x)$  are approximate eigenvalues and eigenvectors of  $H$ , respectively, with an error bounded by an inverse power of  $N$ . Moreover, up to such an error, each eigenvalue of  $H$  satisfying (1.7) is approximated by  $\lambda(x)$  for some  $x \in \mathcal{W}$ . In particular, it provides an upper bound for the tunnelling amplitude, in the sense of (1.2), for the global approximate eigenvectors from Definition 2.1. Its proof is given in Sect. 5 below.

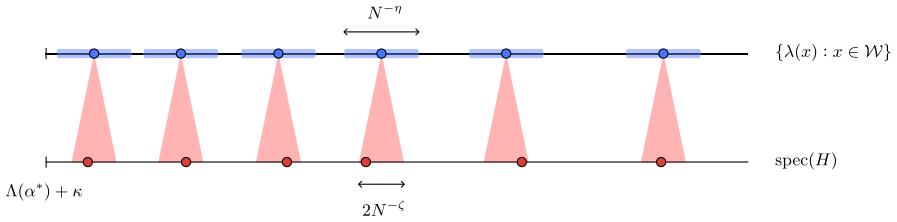
**Proposition 2.3** (Approximate eigenvalues). *Suppose (1.5). Fix  $\mu \in [0, 1/4)$  and  $\zeta \in [0, 1/2 - 3\mu/2)$ . Then, with high probability, for each  $x \in \mathcal{W}$  there exists  $\varepsilon_x \in \mathbb{R}$  such that*

$$\text{spec}(H) \cap \mathcal{I} = \{\lambda(x) + \varepsilon_x : x \in \mathcal{W}\} \cap \mathcal{I}, \quad \mathcal{I} := [\Lambda(\alpha^*) + 3\kappa/4, \sqrt{d}/2]$$

counted with multiplicity and

$$\max_{x \in \mathcal{W}} \max\{|\varepsilon_x|, \|(H - \lambda(x))\mathbf{u}(x)\|\} \leq N^{-\zeta}.$$

The next proposition establishes a spacing of at least  $N^{-\eta}$  between the approximate eigenvalues  $(\lambda(x))_{x \in \mathcal{W}}$ , for large enough  $\eta$ . It is proved in Sect. 6 below.



**Fig. 4.** An illustration of the setup for the perturbation theory in the proof of Theorem 1.1. We draw two instances of the interval  $[\Lambda(\alpha^*) + \kappa, \infty)$ . On the top line, we draw each  $\lambda(x)$  for  $x \in \mathcal{W}$  as blue dot. Each dot is surrounded by a blue buffer of width  $N^{-\eta}$ . By Proposition 2.4, these buffers do not intersect with high probability. On the bottom line, we draw each eigenvalue of  $H$  as red dot. Each dot on the top line gives rise to a red region on the bottom line of width  $2N^{-\zeta}$ . Since  $\zeta > \eta$ , the red regions are disjoint. By Proposition 2.3, each red region contains exactly one eigenvalue of  $H$  and each eigenvalue of  $H$  is contained in a red region. Hence, the eigenvalues of  $H$  are separated by at least  $N^{-\eta}/2$

**Proposition 2.4** (Eigenvalue spacing). *Suppose (1.5). Fix  $\mu \in (0, 1/24)$  and  $\eta > 8\mu$ . Then, with high probability,*

$$|\lambda(x) - \lambda(y)| \geq N^{-\eta}$$

for all  $x, y \in \mathcal{W}$  with  $x \neq y$ .

We now deduce Theorem 1.1 from Propositions 2.2, 2.3, and 2.4.

*Proof of Theorem 1.1.* Let  $\lambda$  be an eigenvalue of  $H$  satisfying (1.7), and  $\mathbf{w}$  an associated, normalized eigenvector. Fix  $\zeta \in (8\mu, 1/3)$  and  $\eta \in (8\mu, \zeta)$ . As  $\mu < 1/24$ , both intervals are nonempty and  $\zeta < 1/2 - 3\mu/2$ . Thus, Propositions 2.3 and 2.4 are applicable with these choices of  $\zeta$  and  $\eta$ . We shall show below that, on the intersection of the high-probability events of Propositions 2.3 and 2.4, there exists a unique  $x \in \mathcal{W}$  such that

$$\mathbf{w} = \mathbf{u}(x) + O(N^{\eta-\zeta}) \tag{2.3}$$

(under suitable choice of the sign of  $\mathbf{w}$ ). Thus, Theorem 1.1 follows from  $\eta < \zeta$  and Proposition 2.2 as  $\exp\left(\frac{1}{6} \frac{\log N}{\log d} \log(1-c)\right) \gg N^{-\varepsilon}$  for any  $\varepsilon > 0$ .

What remains is the proof of (2.3). This is an application of perturbation theory in the form of Lemma D.2, whose conditions we justify now. Note first that, combining (5.24) below and the trivial fact  $\lambda_1(\mathbb{E}H) = \sqrt{d}(1 + o(1))$  with rank-one eigenvalue interlacing (Lemma D.4), we conclude that with high probability  $\lambda_1(H) = \sqrt{d}(1 + o(1))$  and  $\lambda_2(H) \leq \sqrt{d}/2$ . Hence, the eigenvalue  $\lambda$  satisfying (1.7) lies in  $[\Lambda(\alpha^*(\mu)) + \kappa, d/2]$ . From Propositions 2.3 and 2.4 with  $\eta < \zeta$ , we conclude that

$$\text{dist}(\lambda, \text{spec}(H) \setminus \{\lambda\}) \geq N^{-\eta} - 2N^{-\zeta}, \quad |\lambda - \lambda(x)| \leq N^{-\zeta} \tag{2.4}$$

for a unique  $x \in \mathcal{W}$  (see Fig. 4). In particular, as  $\eta < \zeta$ , there is  $\Delta \asymp N^{-\eta}$  such that  $\lambda$  is the unique eigenvalue of  $H$  in  $[\lambda(x) - \Delta, \lambda(x) + \Delta]$ . Moreover,  $\|(H - \lambda(x))\mathbf{u}(x)\| \leq N^{-\zeta}$  by Proposition 2.3. Therefore, all conditions of Lemma D.2 are satisfied, and it implies (2.3). This concludes the proof of Theorem 1.1.  $\square$

By choosing  $\eta$  in (2.4) of the proof of Theorem 1.1 sufficiently small, we conclude the following result.



**Corollary 2.5** (Eigenvalue spacing of  $H$ ). *Suppose (1.5). Fix  $\mu \in (0, 1/24)$  and  $\eta > 8\mu$ . Then, with high probability,*

$$\text{dist}(\lambda, \text{spec}(H) \setminus \{\lambda\}) \geq N^{-\eta},$$

for every  $\lambda \in \text{spec}(H) \cap [\Lambda(\alpha^*(\mu)) + \kappa, \infty)$ .

*Remark 2.6* (Eigenvalue spacing in critical regime). In the critical regime, i.e. when  $d \asymp \log N$ , the lower bound on  $\eta$  in the conditions of Proposition 2.4 and Corollary 2.5 can be weakened to  $\eta > 4\mu$ . Details can be found in Remark 6.2 below.

*Remark 2.7* (Eigenvector mass on vertices in  $\mathcal{V} \setminus \{x\}$ ). With high probability the following holds. Let  $\mathbf{w}$  be an eigenvector of  $H$  associated with the vertex  $x$  as in Theorem 1.1. Then, from  $\mathbf{u}(x)|_{\mathcal{V} \setminus \{x\}} = 0$  and (2.3), we conclude

$$\|\mathbf{w}|_{\mathcal{V} \setminus \{x\}}\| \lesssim N^{-\varepsilon}$$

for any small enough  $\varepsilon > 0$ .

*2.3. Sketch of the proof.* In this subsection we sketch the proof of Theorem 1.1. We use the definitions and notations from Sects. 2.1 and 2.2.

The basic strategy is to find an orthogonal matrix  $U$ , a diagonal  $n \times n$  matrix  $\Theta = \text{diag}(\theta_1, \dots, \theta_n)$ , a symmetric  $(N - n) \times (N - n)$  matrix  $X$ , and a symmetric  $N \times N$  matrix  $E$  such that the following holds. In the basis of the columns  $\mathbf{u}_1, \dots, \mathbf{u}_N$  of  $U$ , the matrix  $H$  has the form

$$U^* H U = \begin{pmatrix} \Theta & 0 \\ 0 & X \end{pmatrix} + E, \tag{2.5}$$

where the matrices  $E$  and  $X$  satisfy

$$\|E\| \ll \min \{|\theta_i - \theta_j| : i \neq j\}, \tag{2.6a}$$

$$\text{dist}(\text{spec}(X), I) \geq \kappa; \tag{2.6b}$$

here  $I$  denotes the interval containing the eigenvalues of  $H$  that we are interested in (cf. (1.7)). We call the first  $n$  columns  $\mathbf{u}_1, \dots, \mathbf{u}_n$  of  $H$  *profile vectors*.

If  $\|E\| = o(1)$  then, for each  $i \in [n]$ , the vector  $\mathbf{u}_i$  is an approximate eigenvector of  $H$  with approximate eigenvalue  $\theta_i$ . Unlike approximate eigenvalues, in general approximate eigenvectors have nothing to do with the actual eigenvectors. For  $\mathbf{u}_i$  to be close to an eigenvector of  $H$ , we require the stronger estimates (2.6), which can be regarded as a version of Mott’s criterion in terms of the profile vectors encoded by  $\mathbf{u}_1, \dots, \mathbf{u}_n$ . Localization then follows provided the  $\mathbf{u}_i$  are shown to be localized.

In [13], it was showed that there is a one-to-one correspondence between eigenvalues of  $H$  in the semilocalized phase  $[2 + o(1), \infty) \setminus \{\lambda_1(H)\}$  and vertices  $x$  of  $\mathbb{G}$  with normalized degree  $\alpha_x \geq 2 + o(1)$ . Subsequently, in [12, 16], the eigenvectors of  $H$  in the semilocalized phase were investigated using the decomposition (2.5). There, the profile vectors  $\mathbf{u}_i$  were supported in balls  $B_r(x)$  around vertices  $x$  of sufficiently large  $\alpha_x$ , where  $r \gg 1$ . We refer to such vectors as *local profile vectors*: they are spatially localized (in the graph distance) and their supports are disjoint. Examples of such local profile vectors are  $\mathbf{w}_r(x)$  and  $\mathbf{v}_r(x)$  for  $x \in \mathcal{W}$ , defined in Definition 1.5 (others were defined in [12, 16]).

The local profile vectors are exponentially decaying with a rate  $c > 0$  depending on  $b$  and the energy. The best possible error estimate for  $\|(H - \theta_i)\mathbf{u}_i\|$  under the condition

that  $\mathbf{u}_i$  is supported in  $B_r(x)$  is obtained by choosing  $\mathbf{u}_i = \mathbf{w}_r(x)$ , the top eigenvector of  $H|_{B_r(x)}$ ; in that case the error is purely a boundary term of order  $e^{-cr}$ . If the supports of the profile vectors are separated by more than 1 (in the graph distance) then it is easy to see that  $\|E\| \leq \max_i \|(H - \theta_i)\mathbf{u}_i\|$  (see also our formulation of Mott’s criterion (1.2)). Hence, the best possible error estimate for  $\|E\|$  is  $e^{-cr}$ . Since the diameter of  $\mathbb{G}$  is  $\frac{\log N}{\log d}(1 + o(1))$  with high probability, the best bound resulting from this approach is  $\|E\| \lesssim N^{-c/\log d} = N^{-o(1)}$  for some constant  $c > 0$ . However, inside the semilocalized phase this bound is always much larger than the typical eigenvalue spacing  $N^{-\eta}$  for some  $\eta > 0$ . Recalling the condition (2.6a), we conclude that any approach to prove localization in the semilocalized phase that uses local profile vectors is doomed to fail.

The reason why any approach based on local profile vectors fails is that local profile vectors (such as  $\mathbf{w}_r(x)$ ) are supported on balls containing a comparatively small set of vertices, and the mass of the true eigenvectors outside of such balls is not small enough to be fully negligible. This leads us to introduce the *global profile vectors*  $\mathbf{u}(x), x \in \mathcal{W}$ , from Definition 2.1, with associated approximate eigenvalues  $\lambda(x)$ . They are defined as the second eigenvector-eigenvalue pair of the matrix  $H^{(\mathcal{V} \setminus \{x\})}$ . Thus,  $\Theta = \text{diag}((\lambda(x))_{x \in \mathcal{W}})$  and the first  $n = |\mathcal{W}|$  columns of  $U$  are given by the orthonormalization of the family  $(\mathbf{u}(x))_{x \in \mathcal{W}}$ . The global profile vector  $\mathbf{u}(x)$  and the best possible local profile vector  $\mathbf{w}_r(x)$  are each defined as eigenvectors of the graph after removal of a set of vertices,  $|\mathcal{V}| - 1 \sim N^\mu \ll N$  vertices for the former and  $|B_r(x)^c| \sim N$  vertices for the latter. This suggests that  $\mathbf{u}(x)$  is a better approximation of a true eigenvector of  $H$ . The price to pay is that its definition is less explicit, and, crucially, the family  $(\mathbf{u}(x))_{x \in \mathcal{W}}$  is not orthogonal owing to the global profile vectors having nonzero overlaps. As explained below, the need to control the overlaps of the global profile vectors presents a serious complication.

The proof of Theorem 1.1 consists of three main steps:

- (i) exponential decay for  $\mathbf{u}(x)$  around  $x$ ,
- (ii)  $\|E\| \leq N^{-\zeta}$  and  $\text{spec}(X)$  is separated from  $I$ ,
- (iii)  $\min_{x \neq y \in \mathcal{W}} |\lambda(x) - \lambda(y)| \geq N^{-\eta}$ ,

for some constants  $\zeta > \eta > 0$  (see also Sect. 2.3.4 below). These items corresponds to Propositions 2.2, 2.3, and 2.4, respectively. We outline their proofs in Sects. 2.3.1, 2.3.2, and 2.3.3, respectively.

**2.3.1. Exponential decay of  $\mathbf{u}(x)$**  First we explain the need to introduce the two different vertex sets  $\mathcal{W} \subset \mathcal{V}$ . In the definition of  $\lambda(x)$  and  $\mathbf{u}(x)$ , all vertices in  $\mathcal{V} \setminus \{x\}$  are removed, while the profile vectors  $\mathbf{u}(x)$  are only considered for  $x$  in the smaller set  $\mathcal{W}$ . The difference between  $\mathcal{W}$  and  $\mathcal{V}$  is used precisely to obtain a spectral gap for  $H^{(\mathcal{V} \setminus \{x\})}$  around  $\lambda(x), x \in \mathcal{W}$ .

To show exponential decay of  $\mathbf{u}(x)$ , we use its definition, a simple computation, and a truncated Neumann series expansion to obtain

$$\begin{aligned} \mathbf{u}(x)|_{\mathcal{V}^c} &= c_x \left( 1 - \frac{H^{(\mathcal{V})}}{\lambda(x)} \right)^{-1} \mathbf{1}_{S_1(x)} \\ &= c_x \sum_{k=0}^{n-1} \left( \frac{H^{(\mathcal{V})}}{\lambda(x)} \right)^k \mathbf{1}_{S_1(x)} + c_x \left( 1 - \frac{H^{(\mathcal{V})}}{\lambda(x)} \right)^{-1} \left( \frac{H^{(\mathcal{V})}}{\lambda(x)} \right)^n \mathbf{1}_{S_1(x)}, \end{aligned} \tag{2.7}$$

where  $c_x := \frac{\langle \mathbf{1}_x, \mathbf{u}(x) \rangle}{\lambda(x)\sqrt{d}}$ . Each term of the sum is supported in  $B_n(x)$  and, thus, vanishes when restricting to  $B_n(x)^c$ . Hence, as  $|\langle \mathbf{1}_x, \mathbf{u}(x) \rangle| \lesssim 1$ ,  $\|\mathbf{1}_{S_1(x)}\| = \sqrt{d\alpha_x}$ , and  $\|(\lambda(x) - H^{(\mathcal{V})})^{-1}\| \lesssim 1$  by the spectral gap of  $H^{(\mathcal{V}\setminus\{x\})}$  around  $\lambda(x) = \lambda_2(H^{(\mathcal{V}\setminus\{x\})})$  mentioned above, we obtain  $\|\mathbf{u}(x)|_{B_n(x)^c}\| \lesssim \sqrt{\alpha_x} q^n$  with  $q = \lambda(x)^{-1} \|H^{(\mathcal{V})}|_{B_{n+1}(x)}\| < 1$ . This is the desired exponential decay. We remark that in (2.7) we establish exponential decay of the Green function of  $H^{(\mathcal{V})}$  evaluated at  $\lambda(x)$ , using that  $\lambda(x)$  is away from the spectrum of  $H^{(\mathcal{V})}$ . This is an instance of a Combes-Thomas estimate, and we translate it to an exponential decay for the eigenvector  $\mathbf{u}(x)$ . Furthermore, we show that  $\lambda(x)$  is isolated in the spectrum of  $H^{(\mathcal{V}\setminus\{x\})}$  and, thus, perturbation theory implies that  $\mathbf{u}(x) = \mathbf{w}_r(x) + o(1)$ .

The rate obtained from the above argument is far from optimal, but an extension of this argument does yield the optimal rate of decay for  $\|\mathbf{u}(x)|_{B_r(x)^c}\|$  for small enough  $\mu$ . To that end, we choose  $n = Ci$  for a large constant  $C > 0$  in (2.7), which makes the remainder term in (2.7) with  $n = Ci$  subleading. It remains to estimate the terms  $(H^{(\mathcal{V})})^k \mathbf{1}_{S_1(x)}$  for  $k = i, \dots, Ci$ , since they vanish for  $k < i$  when restricted to  $B_i(x)^c$ . For  $k \geq i$ , we relate  $(H^{(\mathcal{V})})^k \mathbf{1}_{S_1(x)}$  to the number of a family of walks on the graph  $\mathbb{G}$ . We obtain optimal bounds on this number by a path counting argument, exploiting the tree structure of  $\mathbb{G}|_{B_r(x)}$ , a precise bound on the degrees in  $B_r(x) \setminus \{x\}$  and the concentration of the sphere sizes  $|S_i(x)|$  for  $i \leq r$ .

**2.3.2. Approximate eigenvalues** We now sketch how (ii) is proved. In contrast to the case of local profile vectors discussed above, the proof of  $\|E\| \lesssim \max_{x \in \mathcal{V}} \|(H - \lambda(x))\mathbf{u}(x)\|$  requires also a control of the nonzero overlaps  $\langle \mathbf{u}(x), H\mathbf{u}(y) \rangle$  for  $x \neq y$ . By exponential decay of  $\mathbf{u}(x)$ , it is easy to see that these overlaps are  $N^{-o(1)}$ , but, as explained above, a polynomial bound  $N^{-c}$  is required to prove localization. The construction of  $\mathbf{u}(x)$  and  $\lambda(x)$  and a simple computation reveal that

$$(H - \lambda(x))\mathbf{u}(x) = \sum_{y \in \mathcal{V} \setminus \{x\}} \varepsilon_y(x) \mathbf{1}_y, \quad \varepsilon_y(x) := \frac{1}{\sqrt{d}} \sum_{t \in S_1(y)} \langle \mathbf{1}_t, \mathbf{u}(x) \rangle. \quad (2.8)$$

Then the main idea to estimate  $\varepsilon_y(x)$  is the following elementary bound. Let  $\mathcal{T}$  be a finite set. For any  $(u_t)_{t \in \mathcal{T}} \in \mathbb{C}^{\mathcal{T}}$  and any random  $T \subset \mathcal{T}$  we have

$$\mathbb{E} \left[ \sum_{t \in T} |u_t|^2 \right] = \sum_{t \in \mathcal{T}} \mathbb{E}[\mathbb{1}_{t \in T}] |u_t|^2 \leq \max_{t \in \mathcal{T}} \mathbb{P}(t \in T) \sum_{t \in \mathcal{T}} |u_t|^2. \quad (2.9)$$

Heuristically, by the independence of the edges in the Erdős-Rényi graph, the edges between  $\mathcal{V} \setminus \{x\}$  and  $(\mathcal{V} \setminus \{x\})^c$  are sampled independently of the subgraphs  $\mathbb{G}|_{(\mathcal{V} \setminus \{x\})^c}$  and  $\mathbb{G}|_{\mathcal{V} \setminus \{x\}}$  and are, therefore, independent of  $\mathbf{u}(x)$ . Hence, (2.8), (2.9) and  $|S_1(y)| \lesssim \log N$  yield

$$\begin{aligned} \mathbb{E}[|\varepsilon_y(x)|^2 | A^{(y)}] &\leq \frac{1}{d} \mathbb{E} \left[ |S_1(y)| \sum_{t \in S_1(y)} \langle \mathbf{1}_t, \mathbf{u}(x) \rangle^2 \middle| A^{(y)} \right] \\ &\lesssim \frac{(\log N)^2}{Nd} \|\mathbf{u}(x)\|^2 = N^{-1+o(1)}. \end{aligned}$$

From  $|\mathcal{V}| = N^{\mu+o(1)}$  and Chebyshev's inequality, we therefore conclude  $\|(H - \lambda(x))\mathbf{u}(x)\|^2 \leq N^{2\mu-1+o(1)}$  with high probability.

Since  $(\mathbf{u}(x))_{x \in \mathcal{W}}$  is not an orthogonal family, we choose the first columns of  $U$  in (2.5) as the Gram-Schmidt orthonormalization  $(\mathbf{u}^\perp(x))_{x \in \mathcal{W}}$  of  $(\mathbf{u}(x))_{x \in \mathcal{W}}$  (with respect to a fixed order on  $\mathcal{W}$ ), i.e.

$$\mathbf{u}^\perp(x) := \frac{\mathbf{u}(x) - \Pi_{<x} \mathbf{u}(x)}{\|\mathbf{u}(x) - \Pi_{<x} \mathbf{u}(x)\|}, \tag{2.10}$$

where  $\Pi_{<x}$  denotes the orthogonal projection onto  $\text{span}\{\mathbf{u}(y) : y \in \mathcal{W}, y < x\}$ . It remains to show that, for any  $x \in \mathcal{W}$ ,  $\|(H - \lambda(x))\mathbf{u}^\perp(x)\|$  is also bounded by an inverse power of  $N$ . The denominator in (2.10) is  $\gtrsim 1$  since  $\langle \mathbf{1}_x, \mathbf{u}(x) \rangle \gtrsim 1$  and  $\langle \mathbf{1}_y, \mathbf{u}(x) \rangle = 0$  for all  $y \neq x$ . Moreover,  $H - \lambda(x)$  applied to the numerator of (2.10) is bounded by a negative power of  $N$ , since  $\|(H - \lambda(x))\mathbf{u}(x)\| \leq N^{\mu-1/2+o(1)}$  as shown above, and  $\|\overline{\Pi}_{<x} H \Pi_{<x}\| \leq N^{\mu-1/2+o(1)}$  where  $\overline{\Pi}_{<x} := 1 - \Pi_{<x}$ . The latter bound is proved using the above estimate on  $\|(H - \lambda(x))\mathbf{u}(x)\|$ . Given this construction and the bounds explained above, we extend  $(\mathbf{u}^\perp(x))_{x \in \mathcal{W}}$  to an orthonormal basis and choose these vectors as columns of  $U$ . In particular, the first  $n = |\mathcal{W}|$  columns of  $U$  are given by  $(\mathbf{u}^\perp(x))_{x \in \mathcal{W}}$ .

What remains is to show that  $\text{spec}(X)$  is separated from  $I$ . To that end, we decompose its domain of definition into the span of  $\mathbf{w}_1$ , the eigenvector of  $H$  associated with its largest eigenvalue, and its orthogonal complement. This largest eigenvalue and the overlaps between  $\mathbf{w}_1$  and  $\mathbf{u}^\perp(x)$  for all  $x \in \mathcal{W}$  can be controlled relatively precisely, and we omit  $\mathbf{w}_1$  from the remaining explanations below. It suffices to show that  $\lambda_1(X) \leq \Lambda(\alpha^*) + o(1)$ , by the definitions of  $\mathcal{V}$  and  $\mathcal{W}$ . This upper bound on  $\lambda_1(X)$  is equivalent to

$$\lambda_1(\overline{\Pi}_{\mathcal{W}} H \overline{\Pi}_{\mathcal{W}}) \leq \Lambda(\alpha^*) + o(1), \tag{2.11}$$

where  $\overline{\Pi}_{\mathcal{W}} := 1 - \Pi_{\mathcal{W}}$  and  $\Pi_{\mathcal{W}}$  is the orthogonal projection onto  $\text{span}\{\mathbf{u}^\perp(x) : x \in \mathcal{W}\}$ . An estimate of the form (2.11) was first derived in [12], except that there the projection was defined in terms local profile vectors. Since we are using the global profile vectors  $\mathbf{u}^\perp(x)$ , in our case this estimate is considerably more involved. The rough strategy is to make a link between (2.11) and a corresponding estimate for local profile vectors, which was already established in [12] using bounds on the non-backtracking matrix of  $H$ , an Ihara-Bass type identity, and a local delocalization result for approximate eigenvectors.

To that end, let  $Q$  be the orthogonal projection onto the complement of  $\bigcup_{y \in \mathcal{V} \setminus \mathcal{W}} B_{2r_\star-1}(y)$ , where  $r_\star \asymp \sqrt{\log N}$  is chosen as in [12, eq. (1.8)]. In particular, the local profile vectors  $\mathbf{v}(x)$  from [12] satisfy  $\text{supp } \mathbf{v}(x) \subset B_{r_\star}(x)$  for every  $x \in \mathcal{V}$ . We denote by  $\Pi_Q$  the projection onto  $\text{span}\{Q\mathbf{u}^\perp(x) : x \in \mathcal{W}\}$ . If we obtain a small enough upper bound on  $\|\Pi_{\mathcal{W}} - \Pi_Q\|$  then it suffices to show  $\lambda_1(\overline{\Pi}_Q H \overline{\Pi}_Q) \leq \Lambda(\alpha^*) + o(1)$ , where  $\overline{\Pi}_Q := 1 - \Pi_Q$ . To get a sufficient estimate for  $\|\Pi_{\mathcal{W}} - \Pi_Q\|$ , we need that  $\mathbf{u}^\perp(x)|_{B_r(y)}$  is polynomially small in  $N$  for  $x \neq y$  as a summation over  $x, y \in \mathcal{W}$  is required. This is achieved through an argument motivated by (2.9). Let  $\Pi_v$  be the orthogonal projection onto  $\text{span}\{\mathbf{v}(x) : x \in \mathcal{V} \setminus \mathcal{W}\}$ . By definition of  $Q$  and  $\text{supp } \mathbf{v}(x) \subset B_{r_\star}(x)$ ,  $\Pi_Q$  and  $\Pi_v$  commute. Thus,  $\lambda_1(\overline{\Pi}_Q H \overline{\Pi}_Q) \leq \max\{\lambda_1(\Pi_v H \Pi_v), \lambda_1(\overline{\Pi}_Q \overline{\Pi}_v H \overline{\Pi}_v \overline{\Pi}_Q)\} + o(1)$ , where  $\overline{\Pi}_v := 1 - \Pi_v$ . By [12],  $\lambda_1(\Pi_v H \Pi_v) \leq \Lambda(\alpha^*) + o(1)$ . For eigenvectors of  $H$  which are orthogonal to the local profile vectors  $\mathbf{v}(x)$  and whose associated eigenvalues are large enough, we obtain a weak delocalization estimate by following an argument in [12]. This weak delocalization estimate shows that  $\lambda_1(\overline{\Pi}_Q \overline{\Pi}_v H \overline{\Pi}_v \overline{\Pi}_Q) \leq \Lambda(\alpha^*) + o(1)$  if  $\lambda_1(\overline{\Pi}_{\mathcal{W}} H^{(\mathcal{V} \setminus \mathcal{W})} \overline{\Pi}_{\mathcal{W}}) \leq \Lambda(\alpha^*) + o(1)$ . The last bound is finally obtained by a careful analysis of the spectrum of  $H^{(\mathcal{V} \setminus \mathcal{W})}$ , which is based on viewing it as a perturbation of  $\Pi_{\mathcal{W}} H^{(\mathcal{V} \setminus \mathcal{W})} \Pi_{\mathcal{W}} + \overline{\Pi}_{\mathcal{W}} H^{(\mathcal{V} \setminus \mathcal{W})} \overline{\Pi}_{\mathcal{W}}$  and analysing  $H^{(\mathcal{V} \setminus \mathcal{W})}$  on  $\text{ran } \Pi_{\mathcal{W}}$  in detail.

**2.3.3. Eigenvalue spacing** We now sketch how to prove (iii). To that end, we fix  $a \neq b \in \mathcal{W}$ . To prove that  $\lambda(a)$  and  $\lambda(b)$  are not too close to each other, we choose an appropriate radius  $r$  on the scale  $\frac{\log N}{\log d}$  of the diameter of  $\mathbb{G}$ . Then we fix the two sub-graphs  $\mathbb{G}|_{B_r(b)}$  and  $\mathbb{G}|_{B_r(b)^c}$  and show that resampling the edges between  $S_r(b)$  and  $B_r(b)^c$  results in a substantial change of  $\lambda(b)$  while  $\lambda(a)$  remains almost unchanged: we establish simultaneous *anticoncentration* for  $\lambda(b)$  and *concentration* for  $\lambda(a)$ , which yields anticoncentration for their difference. The edges between  $S_r(b)$  and  $B_r(b)^c$  form an independent family of Bernoulli random variables by definition of the Erdős-Rényi graph.

On a more formal level, we work conditionally on  $\mathcal{F} := \sigma(B_r(b), A|_{B_r(b)}, A|_{B_r(b)^c})$  and prove the following two statements in order to obtain a lower bound on  $|\lambda(a) - \lambda(b)|$ .

- (a)  $\lambda(a)$  fluctuates little under resampling of the edges between  $S_r(b)$  and  $B_r(b)^c$ , i.e. the *concentration* estimate

$$\mathbb{P}(|\lambda(a) - z| \ll N^{-\eta} | \mathcal{F}) \geq 1 - N^{-\eta/2+o(1)}$$

holds if  $z$  is the second largest eigenvalue of  $H^{(\mathcal{V} \cup B_r(b) \setminus \{a\})}$ , which is  $\mathcal{F}$ -measurable.

- (b)  $\lambda(b)$  fluctuates a lot under resampling of the edges between  $S_r(b)$  and  $B_r(b)^c$ , i.e. the *anticoncentration* estimate

$$\mathbb{P}(|\lambda(b) - z| \geq N^{-\eta} | \mathcal{F}) \geq 1 - N^{-\eta/2+o(1)}$$

holds for any  $\mathcal{F}$ -measurable spectral parameter  $z$  in  $I$  (see the definition of  $I$  after (2.6b)).

We justify (a) by replacing  $\mathbf{u}(a)$  with an  $\mathcal{F}$ -measurable version. This allows for a use of (2.9) in a similar fashion as in the first part of Sect. 2.3.2 and reveals that, conditionally on  $\mathcal{F}$ ,  $\lambda(a)$  is concentrated around the second largest eigenvalue of  $H^{(\mathcal{V} \cup B_r(b) \setminus \{a\})}$  since  $|B_r(b)|$  is not too large due to our choice of  $r$ .

The proof of (b) is much more elaborate. We start by noting that  $\lambda(b)$  is characterized by the equation

$$\lambda(b) + \frac{1}{d} \sum_{x,y \in S_1(b)} (H^{(\mathcal{V})} - \lambda(b))_{xy}^{-1} = 0, \tag{2.12}$$

as follows from Schur’s complement formula. The main strategy is to derive a recursive family of equations for the Green function, starting from (2.12) and extending to increasingly large spheres around  $b$ , to which Kesten’s self-improving anticoncentration result can be applied. To quantify anticoncentration, we use *Lévy’s concentration function*

$$Q(X, L) := \sup_{t \in \mathbb{R}} \mathbb{P}(X \in [t - L, t + L]), \tag{2.13}$$

where  $X$  is a random variable and  $L > 0$  is deterministic.

**Proposition 2.8** (Theorem 2 of [57]). *There exists a universal constant  $K$  such that for any independent random variables  $X_1, \dots, X_n$  satisfying  $Q(X_i, L) \leq 1/2$  we have*

$$Q\left(\sum_{i \in [n]} X_i, L\right) \leq \frac{K}{\sqrt{n}} \max_{i \in [n]} Q(X_i, L). \tag{2.14}$$

This result is an improvement due to Kesten [57] of a classical anticoncentration result of Doebelin, Lévy, Kolmogorov, and Rogozin. Kesten’s insight was that such an estimate can be made self-improving, as manifested by the factor  $\max_{i \in [n]} Q(X_i, L)$  on the right-hand side. This factor is crucial for our argument, as it allows us to successively improve the upper bound on  $Q$ .

We now explain more precisely how the expression of  $\lambda(b)$  in terms of a large number of Green function entries is obtained. We shall tacitly use that  $\mathbb{G}|_{B_r(b)}$  is a tree, which can be easily shown to be true with high probability. Applying Schur’s complement formula at  $x \in S_i(b)$ , using standard resolvent identities, and arguing similarly as in (2.9) to control errors yields

$$\frac{1}{G_{xx}(i-1, z)} = -z - \frac{1}{d} \sum_{y \in S_1^+(x)} G_{yy}(i, z) + o(1), \tag{2.15}$$

where  $G_{xx}(i, z) := (H^{(B_i(b) \cup \mathcal{V})} - z)_{xx}^{-1}$  and  $S_1^+(x) = S_1(x) \cap S_{i+1}(b)$  is the set of children of  $x$  in the tree  $\mathbb{G}|_{B_r(x)}$  rooted at  $b$ . The error  $o(1)$  is polynomially small in  $N$ ; it comprises error terms arising from removing vertices from  $H$  and neglecting all off-diagonal Green function entries.

Setting the error term in (2.15) to zero, we obtain a recursive equation for the idealized Green function entries  $(g_x(z))_{x \in B_r(b) \setminus \{b\}}$ , given by

$$g_x(z) := \begin{cases} -z - \frac{1}{d} \sum_{y \in S_1^+(x)} G_{yy}(r, z) & \text{if } x \in S_r(b) \\ -z - \frac{1}{d} \sum_{y \in S_1^+(x)} g_y(z) & \text{if } x \in B_{r-1}(b) \setminus \{b\}, \end{cases} \tag{2.16}$$

which is an approximate version of the recursion (2.15) for the actual Green function entries. The recursion begins at the boundary of the ball  $B_r(b)$  and propagates inwards. We note that, for any  $1 \leq i \leq r$ , conditioned on  $\mathcal{F}$ , the family  $(g_x(z))_{x \in S_i(b)}$  is independent if  $\mathbb{G}|_{B_r(b)}$  is a tree and  $z$  is  $\mathcal{F}$ -measurable. From (2.15), (2.16) and  $r \asymp \frac{\log N}{\log d}$ , it is not hard to conclude by induction that, for large enough  $\eta > 0$ , with high probability,  $g_x(z) = G_{xx}(0, z) + o(N^{-\eta}) = (H^{(\mathcal{V})} - z)_{xx}^{-1} + o(N^{-\eta})$  for all  $x \in S_1(b)$ . Hence, if we can prove that  $Q(g_x(z), N^{-\eta}) \leq N^{-c}$  for all  $x \in S_1(b)$  and some constant  $c > 0$ , then a union bound over  $a \neq b \in \mathcal{W}$ ,  $|\mathcal{W}| \asymp N^\mu$ , the smallness of the off-diagonal entries of  $(H^{(\mathcal{V})} - \lambda(b))^{-1}$  as argued after (2.15), and (2.12) imply anticoncentration for  $\lambda(b) - z$ . This is (b), which together with (a) implies that  $\min_{a \neq b \in \mathcal{W}} |\lambda(a) - \lambda(b)| \geq N^{-\eta}$  with high probability, i.e. (iii).

Therefore, to complete the sketch of (iii), what remains is to prove  $Q(g_x(z), N^{-\eta}) \leq N^{-c}$  for all  $x \in S_1(b)$ , whose proof we sketch now. Throughout the entire argument we condition on  $\mathcal{F}$  and use that  $\mathbb{G}|_{B_r(x)}$  is a tree. To begin the recursion, we first show that  $Q(g_x, d^{-1}) \leq 1/2$  for any  $x \in S_r(b)$ . This follows from the first case of (2.16), using a weak lower bound on the entries  $G_{yy}(r, z)$  and anticoncentration from the fact that the size of  $S_1^+(x)$  is a binomial random variable conditioned on  $\mathcal{F}$ .

Next, let  $x \in S_i(b)$  for  $1 \leq i \leq r - 1$ . By using the second case in (2.16) and Proposition 2.8, we iteratively refine the resolution, i.e. decrease the second argument of  $Q$ , and decrease the upper bound on  $Q$ , which are  $d^{-1}$  and  $1/2$ , respectively, at the starting point. Indeed, conditioning on  $A|_{B_i(b)}$ , the second case in (2.16) and rescaling

the second argument of  $Q$  yield

$$\begin{aligned}
 Q(g_x(z), (T^2d)^{-r+i}) &\leq Q\left(\sum_{y \in S_1^+(x)} g_y(z), (T^2d)^{-r+i+1}\right) \\
 &\leq \frac{K}{\sqrt{|S_1^+(x)|}} \max_{y \in S_1^+(x)} Q(g_y(z), (T^2d)^{-r+i+1}), \tag{2.17}
 \end{aligned}$$

where we applied Proposition 2.8 using the independence of  $(g_y(z))_{y \in S_1^+(x)}$  in the second step. Here, we also used that  $Q(f(X), L) \leq Q(X, T^{-2}L)$  if  $f(t) := \frac{1}{t}$  and  $X \in [T^{-1}, T]$  since the derivative of  $f$  is bounded from below by  $T^{-2}$  on this interval, and that, with high probability,  $g$  lies in  $[T^{-1}, T]$  for  $T \asymp \sqrt{\frac{\log N}{d}}$ .

The estimate (2.17) yields the desired self-improvement provided that  $|S_1^+(x)|$  is large enough. However,  $|S_1^+(x)|$  is not large enough for all vertices  $x$  in  $B_r(b)$  (and in fact consistently applying (2.17) at all vertices yields an anticoncentration bound at the root  $b$  that is far from optimal and too weak to conclude localization). Sometimes, a better bound than (2.17) can be obtained by replacing Kesten’s estimate (2.14) with the trivial estimate

$$Q\left(\sum_{i \in [n]} X_i, L\right) \leq \min_{i \in [n]} Q(X_i, L), \tag{2.18}$$

which follows immediately from the independence of the random variables  $X_i$ . Although this estimate lacks the factor  $K/\sqrt{n}$  from (2.14), it replaces the maximum with a minimum. Thus, an important ingredient in our recursive self-improving anticoncentration argument is an algorithm that determines which of (2.14) or (2.18) is to be used at any given vertex  $x \in B_r(b)$ . It relies on the notion of *robust vertices*.

Recursively, a vertex  $x \in B_r(b)$  is called *robust* if  $x \in S_r(b)$  or  $S_1^+(x)$  contains at least  $d/2$  robust vertices. We denote the set of robust vertices by  $\mathcal{R}$ . An important auxiliary result is that the root  $b$  is robust with high probability, which in particular implies that  $S_i(b) \cap \mathcal{R}$  is large for any  $i \leq r$ . Therefore, we can restrict to  $x \in S_i(b) \cap \mathcal{R}$  and proceed similarly as in (2.17) to obtain

$$\begin{aligned}
 Q(g_x(z), (T^2d)^{-r+i}) &\leq Q\left(\sum_{y \in S_1^+(x) \cap \mathcal{R}} g_y(z), (T^2d)^{-r+i+1}\right) \\
 &\leq \frac{K\sqrt{2}}{\sqrt{d}} \max_{y \in S_1^+(x) \cap \mathcal{R}} Q(g_y(z), (T^2d)^{-r+i+1}),
 \end{aligned}$$

where in the first step we used (2.18). Thus, we obtain  $Q(g_x(z), (T^2d)^{-r}) \leq (K\sqrt{2d^{-1}})^{r-1}$  for all  $x \in S_1(b)$  and, therefore, by choosing  $r$  such that  $(T^2d)^{r+1} = N^\eta$ , we arrive at  $Q(g_x(z), N^{-\eta}) \leq N^{-\eta/2+o(1)}$  for all  $x \in S_1(b)$ . This is the desired anticoncentration bound, (b), which, as explained above, implies (iii).

**2.3.4. The three main exponents of  $N$**  Throughout this paper we use the three exponents  $\mu, \zeta, \eta > 0$  to control three central quantities of the argument. We summarize their roles here for easy reference.



- $N^\mu$  is the typical size of the vertex sets  $\mathcal{V}$  and  $\mathcal{W}$  (cf. Proposition 3.2 (i)), as the parameter  $\mu$  is introduced to control  $\alpha^*$  (see (1.6)). Consequently,  $N^\mu$  is also the typical number of eigenvalues of  $H$  satisfying (1.7). Therefore, the factor  $N^\mu$  emerges from union bounds when a property is required for all  $x \in \mathcal{V}$  or  $x \in \mathcal{W}$ .
- $N^{-\zeta}$  is the upper bound on the eigenvalue approximation that we establish in Proposition 2.3, i.e. on the distance between  $\lambda(x)$  and the eigenvalues of  $H$  or, more precisely, the upper bound on  $\|(H - \lambda(x))\mathbf{u}(x)\|$ . Proposition 2.3 requires the condition  $\zeta < 1/2 - 3\mu/2$ .
- $N^{-\eta}$  is the lower bound on the eigenvalue spacing, or correspondingly  $\min_{x \neq y \in \mathcal{W}} |\lambda(x) - \lambda(y)|$ , that we prove in Proposition 2.4. Since the typical eigenvalue spacing is  $N^{-\mu}$  in the interval we consider, we clearly need  $\mu < \eta$ . In fact, our proof of Proposition 2.4 requires the stronger condition  $8\mu < \eta$  for technical reasons as well as  $\mu < 1/24$ .

To apply perturbation theory in the proof of Theorem 1.1, we need that the error in the eigenvalue approximation  $N^{-\zeta}$  be smaller than the eigenvalue spacing  $N^{-\eta}$ . This means that  $\eta < \zeta$ .

2.4. *Localization profile: proof of Theorem 1.6.* Theorem 1.6 is an immediate consequence of the following result.

**Proposition 2.9** (Local approximation for  $\mathbf{u}(x)$ ). *Let  $\mu \in [0, 1/3)$ . Then, with high probability, the following holds for all  $x \in \mathcal{W}$ . If  $r \in \mathbb{N}$  satisfies*

$$\log d \ll r \leq \min \left\{ \frac{1}{6} \frac{\log N}{\log d}, \min \left\{ \frac{1}{5} - \frac{\mu}{4}, \frac{1}{3} - \mu \right\} \frac{\log N}{\log d} - 2 \right\}, \tag{2.19}$$

then

- (i)  $\mathbf{u}(x) = \mathbf{v}_r(x) + o(1)$ ,
- (ii)  $\mathbf{u}(x) = \mathbf{w}_r(x) + o(1)$ .

Part (i) is proved in Sect. 3.2 below. Part (ii) follows from Corollary 4.2 below, since  $\Lambda(\alpha_x) \geq 2$ .

*Proof of Theorem 1.6.* From (2.3) in the proof of Theorem 1.1, we know that  $\mathbf{w} = \mathbf{u}(x) + O(N^{-\varepsilon})$  for a unique  $x \in \mathcal{W}$  and some small enough  $\varepsilon > 0$ . Therefore, Theorem 1.6 follows from Proposition 2.9. □

2.5. *Mobility edge: proof of Theorem 1.7.* Part (i) follows from Theorem 1.1, since  $\inf\{\alpha > 0 : \mathbb{P}(\alpha_1 \geq \alpha) \leq N^{-23/24}\} \leq 2$  if  $d \geq \frac{23}{24} b_* \log N$ , by Lemma D.1 below. Part (ii) was proved in [14, Theorem 1.1 (ii)].

2.6. *Localization length: proof of Theorem 1.8.* For any eigenvector  $\mathbf{w}$  we introduce the function

$$q(u) := \sum_y d(u, y) \langle \mathbf{1}_y, \mathbf{w} \rangle^2,$$

so that  $\ell(\mathbf{w}) = \min_u q(u)$ .

For the localized phase, suppose that  $\lambda$  is an eigenvalue satisfying  $|\lambda| \geq 2 + \kappa$  with associated eigenvector  $\mathbf{w}$ . Denote by  $x$  the unique vertex associated with  $\mathbf{w}$  from Theorems 1.1 and 1.6. Then the following estimates hold on the intersection of the high-probability events of these two theorems. By Theorem 1.1, there exists a constant  $R \equiv R_\kappa$  such that

$$q(u) \geq d(u, x) \langle \mathbf{1}_x, \mathbf{w} \rangle^2 \geq R \frac{\alpha_x - 2}{2(\alpha_x - 1)} + o(1) \geq \frac{\alpha_x}{\alpha_x - 2} \quad \text{for all } u \in B_R(x)^c, \tag{2.20}$$

where the third inequality holds for large enough constant  $R$ . Moreover, for any  $\varepsilon > 0$  there exists a constant  $R' \in \mathbb{N}$  such that for  $u \in B_{R'}(x)$  we have

$$\begin{aligned} q(u) &= \sum_{y \in B_{R'}(x)} d(u, y) \langle \mathbf{1}_y, \mathbf{w} \rangle^2 + \sum_{y \in B_{R'}(x)^c} d(u, y) \langle \mathbf{1}_y, \mathbf{w} \rangle^2 \\ &= \sum_{y \in B_{R'}(x)} d(u, y) \langle \mathbf{1}_y, \mathbf{w} \rangle^2 + O(\varepsilon) = \sum_{y \in B_{R'}(x)} d(u, y) \langle \mathbf{1}_y, \mathbf{v}_r(x) \rangle^2 + O(\varepsilon) + o(1), \end{aligned} \tag{2.21}$$

where the second step follows from the estimate  $d(u, y) \leq R + d(x, y)$  and the exponential decay from Theorem 1.1, and the third step from Theorem 1.6 with some  $\log d \ll r \ll \frac{\log N}{\log d}$ .

To analyse the sum, we abbreviate  $k := d(x, u)$  and introduce the set  $T_i(u, x) := S_i(x) \cap S_{i+k}(u)$  for  $0 \leq i \leq r$ , which is the set of vertices in  $S_i(x)$  whose geodesic to  $x$  does not pass through  $u$ . By Proposition 3.2 (ii) below, the graph  $\mathbb{G}|_{B_r(x)}$  is a tree, which implies

$$d(u, y) = \begin{cases} k + i & \text{if } y \in T_i(u, x) \\ |k - i| & \text{if } y \in S_i(x) \setminus T_i(u, x). \end{cases} \tag{2.22}$$

Next, we estimate  $|S_i(x) \setminus T_i(u, x)|$ . For  $1 \leq i \leq k$ , the set  $S_i(x) \setminus T_i(u, x)$  consists of the unique vertex on the geodesic from  $x$  to  $u$  at distance  $i$  from  $x$ . For  $i > k$ , we have  $S_i(x) \setminus T_i(u, x) = S_i(x) \cap S_{i-k}(u)$ . Here we used the tree structure of  $\mathbb{G}|_{B_r(x)}$ . Hence, we conclude that for  $i > k$  we have  $|S_i(x) \setminus T_i(u, x)| \leq |S_{i-k}(u)| \lesssim (\log N) d^{i-k-1}$ , by Proposition 3.1.

Next, by [12, Lemma 5.4], with high probability we have  $|S_i(x)| = |S_1(x)| d^{i-1} (1 + o(1))$  for  $1 \leq i \leq r$ . Since  $|S_1(x)| \geq \alpha^* d \geq 2d$ , we conclude that  $|S_i(x)| \geq d^i$  for  $i \leq r$ . Putting all of these estimates together, we conclude for all  $1 \leq i \leq r$  that

$$\frac{|S_i(x) \setminus T_i(u, x)|}{|S_i(x)|} \lesssim \begin{cases} d^{-i} & \text{if } 1 \leq i \leq k \\ (\log N) d^{-k-1} & \text{if } k < i \leq r. \end{cases}$$

Since  $T_i(u, x) = S_i(x)$  for  $k = 0$  or  $i = 0$ , using the condition (1.5) we conclude that

$$\frac{|S_i(x) \setminus T_i(u, x)|}{|S_i(x)|} = o(1) \tag{2.23}$$

for all  $u \in B_R(x)$  and  $0 \leq i \leq r$ .

Next, using (2.22) and recalling the definition (1.11), we write the above sum as

$$\begin{aligned} & \sum_{y \in B_{R'}(x)} d(u, y) \langle \mathbf{1}_y, \mathbf{v}_r(x) \rangle^2 \\ &= \sum_{i=0}^{R'} \left( \sum_{y \in T_i(u, x)} (k+i) \langle \mathbf{1}_y, \mathbf{v}_r(x) \rangle^2 + \sum_{y \in S_i(x) \setminus T_i(u, x)} |k-i| \langle \mathbf{1}_y, \mathbf{v}_r(x) \rangle^2 \right) \\ &= \sum_{i=0}^{R'} \left( (k+i) u_i(\alpha_x)^2 \frac{|T_i(u, x)|}{|S_i(x)|} + |k-i| u_i(\alpha_x)^2 \frac{|S_i(x) \setminus T_i(u, x)|}{|S_i(x)|} \right) \\ &= \sum_{i=0}^{R'} (k+i) u_i(\alpha_x)^2 + o(1), \end{aligned}$$

where in the last step we used (2.23), the fact that  $R'$  is constant, and  $\sum_{i=0}^{r-1} u_i(\alpha_x)^2 = 1$ . The latter sum is clearly minimized for  $k = 0$ . Recalling (2.21), we therefore conclude that for any  $u \in B_R(x)$  we have

$$q(u) \geq q(x) + O(\varepsilon) + o(1), \quad q(x) = \sum_{i=1}^{\infty} i u_i(\alpha_x)^2 + O(\varepsilon) + o(1)$$

for large enough  $R'$  depending on  $\varepsilon$ . Since  $\varepsilon > 0$  was arbitrary, and recalling (2.20), by taking the minimum over  $u \in [N]$ , it therefore suffices to show that

$$\sum_{i=1}^{\infty} i u_i(\alpha_x)^2 = \frac{\alpha_x}{2(\alpha_x - 2)} + o(1) = \frac{|\lambda|}{2\sqrt{\lambda^2 - 4}} + o(1). \tag{2.24}$$

The first equality of (2.24) is an elementary computation using the definition (1.10), recalling the normalization  $\sum_{i=0}^{r-1} u_i(\alpha)^2 = 1$  and that  $r \gg 1$ . The second equality of (2.24) follows from the estimate  $|\lambda| = \Lambda(\alpha_x) + o(1)$  by Remark 1.2, which can be inverted to obtain  $\frac{1}{\alpha_x} = \frac{1}{2} \left( 1 - \frac{\sqrt{\lambda^2 - 4}}{|\lambda|} \right) + o(1)$ .

For the delocalized phase, suppose that  $\lambda$  is an eigenvalue satisfying  $|\lambda| \leq 2 - \kappa$  with associated eigenvector  $\mathbf{w}$ . We use [14, Theorem 1.1 (ii)] to deduce that with probability  $1 - O(N^{-10})$  we have  $\|\mathbf{w}\|_{\infty}^2 \leq N^{-1+o(1)}$ . Hence, with probability  $1 - O(N^{-9})$  we have, for any  $x \in [N]$  and  $r \geq 0$ ,

$$\begin{aligned} q(x) &\geq \sum_{y \in B_r(x)} d(x, y) \langle \mathbf{1}_y, \mathbf{w} \rangle^2 + r \sum_{y \in B_r(x)^c} \langle \mathbf{1}_y, \mathbf{w} \rangle^2 \\ &\geq r \left( 1 - \sum_{y \in B_r(x)} \langle \mathbf{1}_y, \mathbf{w} \rangle^2 \right) \geq r \left( 1 - N^{-1+o(1)} |B_r(x)| \right). \end{aligned}$$

Next, we deduce from [32, Lemma 1] that for any constant  $\varepsilon > 0$  there is a constant  $\delta > 0$  such that, with high probability, if  $r \leq (1 - \varepsilon) \frac{\log N}{\log d}$  then  $|B_r(x)| \leq N^{1-\delta}$  for all  $x \in [N]$ . Choosing  $r := \lfloor (1 - \varepsilon) \frac{\log N}{\log d} \rfloor$ , we conclude that, for any  $\varepsilon > 0$ , with high probability, for all eigenvectors  $\mathbf{w}$  with eigenvalue  $\lambda$  satisfying  $|\lambda| \leq 2 - \kappa$ , we have  $\ell(\mathbf{w}) \geq (1 - \varepsilon + o(1)) \frac{\log N}{\log d}$ . Since  $\varepsilon > 0$  was an arbitrary constant, we conclude

the stronger lower bound  $\ell(\mathbf{w}) \geq (1 - o(1)) \frac{\log N}{\log d} = \text{diam}(\mathbb{G})(1 + o(1))$ , where the last step follows from [32]. The complementary upper bound  $\ell(\mathbf{w}) \leq \text{diam}(\mathbb{G})$  follows by definition of  $\text{diam}(\mathbb{G})$ , since  $d(x, y) \leq \text{diam}(\mathbb{G})$  for all  $x, y \in [N]$  (here we used that under our assumption on  $d$ , the graph  $\mathbb{G}$  is with high probability connected). This concludes the proof.

### 3. Preliminaries

The rest of this paper is devoted to the proofs of Propositions 2.2, 2.3 and 2.4. We begin with a short section that collects some basic properties of the graph  $\mathbb{G}$  and its spectrum.

*3.1. Properties of the graph.* In this subsection, we collect some basic local properties of the Erdős-Rényi graph  $\mathbb{G}$  around vertices in  $\mathcal{V}$ .

**Proposition 3.1.** *Suppose that  $\sqrt{\log N} \ll d \leq 3 \log N$ . With high probability, the following holds.*

- (i)  $\max_{x \in [N]} |S_1(x)| \leq 10 \log N$ .
- (ii)  $|B_i(x)| \lesssim \max\{|S_1(x)|, d\} d^{i-1}$  for all  $x \in [N]$  and all  $i \in \mathbb{N}$  with  $i \leq \frac{1}{3} \frac{\log N}{\log d}$ .

Item (i) is a simple application of Bennett’s inequality and (ii) follows from [12, 13]; a detailed proof is given in [15, Appendix E]. In particular, by the assumption  $d \gg \sqrt{\log N}$ , we get from Proposition 3.1 (i) that

$$\Lambda(\alpha_x) \lesssim \sqrt{\alpha_x} \lesssim \sqrt{\frac{\log N}{d}} \ll \sqrt{d} \tag{3.1}$$

for all  $x \in [N]$  on the high-probability event from Proposition 3.1.

**Proposition 3.2.** *Let  $\mu \in [0, 1/3)$  be a constant. Suppose that  $\sqrt{\log N} \ll d \leq 3 \log N$ . With high probability, for any  $r \in \mathbb{N}$  satisfying*

$$1 \leq r \leq \min \left\{ \frac{1}{5} - \frac{\mu}{4}, \frac{1}{3} - \mu \right\} \frac{\log N}{\log d}, \tag{3.2}$$

*the following holds.*

- (i)  $|\mathcal{V}| \leq N^{\mu+o(1)}$ .
- (ii)  $\mathbb{G}|_{B_r(x)}$  is a tree for all  $x \in \mathcal{V}$ .
- (iii)  $B_r(x) \cap B_r(y) = \emptyset$  for all  $x, y \in \mathcal{V}$  satisfying  $x \neq y$ .
- (iv) Let  $v \in [0, 1]$ . If  $1 - \mu - v - r \frac{\log d}{\log N} \gtrsim 1$  then  $|S_1(y)| \leq \alpha^*(v)d$  for all  $y \in \bigcup_{x \in \mathcal{V}} (B_r(x) \setminus \{x\})$ .

These statements are all consequences of [13, 16]; the details are given in [15, Appendix E].

**3.2. Properties of the spectrum.** In this subsection we collect basic spectral properties of  $H$  and some of its submatrices.

**Definition 3.3.** Let  $\mathbf{w}_1$  be a normalized eigenvector of  $H$  with nonnegative entries associated with the largest eigenvalue  $\lambda_1(H)$  of  $H$ .

Note that, with high probability,  $\mathbf{w}_1$  is unique and coincides with the Perron–Frobenius eigenvector of the giant component of  $\mathbb{G}$ .

**Proposition 3.4.** *Suppose (1.5). With high probability the following holds.*

- (i) For any  $\mu \in [0, 1]$  we have  $\max\{\lambda_2(H^{(\mathcal{V})}), -\lambda_N(H^{(\mathcal{V})})\} \leq \Lambda(\alpha^*) + o(1)$ .
- (ii) Fix  $\mu \in [0, 1/3)$ . If  $X \subset \bigcup_{x \in \mathcal{V}} B_r(x)$  with  $r \in \mathbb{N}$  as in (3.2), then  $\lambda_1(H^{(X)})$  and the corresponding eigenvector  $\mathbf{w}$  of  $H^{(X)}$  satisfy

$$\lambda_1(H^{(X)}) = \sqrt{d}(1 + o(1)), \quad \left\| \mathbf{w} - \frac{\mathbf{1}_{X^c}}{|X^c|^{1/2}} \right\| = o(1).$$

- (iii) Fix  $\mu \in [0, 1)$ . If  $x \in \mathcal{V}$ ,  $r \in \mathbb{N}$  satisfies  $\log d \ll r \leq \frac{1}{6} \frac{\log N}{\log d}$  and  $X \subset [N]$  satisfies  $X \cap B_r(x) = \emptyset$ , then

$$\|(H^{(X)} - \Lambda(\alpha_x))\mathbf{v}_r(x)\| = o(1),$$

where  $\mathbf{v}_r(x)$  was defined in (1.11).

- (iv) Fix  $\mu \in [0, 4/5)$ . For any  $r \in \mathbb{N}$  satisfying  $r \ll \frac{d}{\log \log N}$ , there is a normalized vector  $\mathbf{q}$  with  $\text{supp } \mathbf{q} \subset \left(\bigcup_{x \in \mathcal{V}} B_{r+1}(x)\right)^c$  such that

$$\|(H - \sqrt{d})\mathbf{q}\| \lesssim d^{-1/2}, \quad \|\mathbf{w}_1 - \mathbf{q}\| \lesssim d^{-1}, \quad \|\mathbf{q} - N^{-1/2}\mathbf{1}_{[N]}\| \lesssim d^{-1/2}.$$

These results follow essentially from [12, 13, 16]; the detailed proof is presented in Section C below.

**Definition 3.5.** We denote by  $\Omega$  the intersection of the high-probability events of Propositions 3.1, 3.2 and 3.4.

In particular, on  $\Omega$  the estimate (3.1) holds.

**Corollary 3.6.** Fix  $\mu \in [0, 1/3)$ . On  $\Omega$ , the following holds for all  $x \in \mathcal{W}$  and all  $X \subset [N]$ . If  $X \cap B_r(x) = \emptyset$  and  $\mathcal{V} \setminus \{x\} \subset X \subset \bigcup_{y \in \mathcal{V}} B_r(y)$  for some  $r \in \mathbb{N}$  satisfying  $\log d \ll r \leq \min\left\{\frac{1}{6}, \frac{1}{5} - \frac{\mu}{4}, \frac{1}{3} - \mu\right\} \frac{\log N}{\log d}$  then

$$\lambda_1(H^{(X)}) = \sqrt{d}(1 + o(1)), \quad \lambda_2(H^{(X)}) = \Lambda(\alpha_x) + o(1), \quad \lambda_3(H^{(X)}) \leq \Lambda(\alpha^*) + o(1).$$

Corollary 3.6 with  $X = \mathcal{V} \setminus \{x\}$  directly implies that, on  $\Omega$ ,

$$\lambda(x) = \lambda_2(H^{(\mathcal{V} \setminus \{x\})}) = \Lambda(\alpha_x) + o(1). \quad (3.3)$$

*Proof of Corollary 3.6.* The statement about  $\lambda_1(H^{(X)})$  is identical to Proposition 3.4 (ii). By eigenvalue interlacing (Lemma D.3), we have

$$\lambda_3(H^{(X)}) \leq \lambda_3(H^{(\mathcal{V} \setminus \{x\})}) \leq \lambda_2(H^{(\mathcal{V})}) \leq \Lambda(\alpha^*) + o(1),$$

where for the last inequality we used Proposition 3.4 (i). Finally, Proposition 3.4 (iii) implies that there exists an eigenvalue of  $H^{(X)}$  at distance  $o(1)$  from  $\Lambda(\alpha_x)$ . Because of the estimates on  $\lambda_1(H^{(X)})$  and  $\lambda_3(H^{(X)})$  just proven, and because  $\Lambda(\alpha_x) \geq \Lambda(\alpha^*) + \kappa/2$  for  $x \in \mathcal{W}$  (recall (2.2)) as well as  $\Lambda(\alpha_x) \ll \sqrt{d}$  by (3.1), this eigenvalue has to be  $\lambda_2(H^{(X)})$ .  $\square$

*Proof of Proposition 2.9 (i).* We show that the conclusion of Proposition 2.9 (i) holds on  $\Omega$ . The proof uses a spectral gap of  $H^{(\mathcal{V} \setminus \{x\})}$  around  $\lambda(x) = \lambda_2(H^{(\mathcal{V} \setminus \{x\})})$ , that  $\mathbf{v}_r(x)$  is an approximate eigenvector of  $H^{(\mathcal{V} \setminus \{x\})}$  by Proposition 3.4 (iii), and perturbation theory. Indeed, from Corollary 3.6 with  $X = \mathcal{V} \setminus \{x\}$ , recalling the definition of  $\mathcal{W}$  (see (2.2)), we obtain that  $\lambda_2(H^{(\mathcal{V} \setminus \{x\})})$  is separated from the other eigenvalues of  $H^{(\mathcal{V} \setminus \{x\})}$  by a positive constant. Owing to (3.3), Proposition 3.4 (iii) with  $X = \mathcal{V} \setminus \{x\}$  and Lemma D.2 below imply  $\|\mathbf{u}(x) - \mathbf{v}_r(x)\| = o(1)$ , i.e. Proposition 2.9 (i).  $\square$

We conclude the following corollary from Proposition 2.9 (i) and its proof.

**Corollary 3.7.** Fix  $\mu \in [0, 1/3)$ . On  $\Omega$  we have  $\langle \mathbf{1}_x, \mathbf{u}(x) \rangle = \sqrt{\frac{\alpha_x - 2}{2(\alpha_x - 1)}} + o(1) \gtrsim 1$  for all  $x \in \mathcal{W}$ .

*Proof.* We first note that for any  $r \rightarrow \infty$  as  $N \rightarrow \infty$ , we have  $u_0(\alpha_x) = \sqrt{\frac{\alpha_x - 2}{2(\alpha_x - 1)}} + o(1)$  by (1.10). By Proposition 2.9 (i) and its proof,  $\langle \mathbf{1}_x, \mathbf{u}(x) \rangle = \langle \mathbf{1}_x, \mathbf{v}_r(x) \rangle + o(1) = u_0(\alpha_x) + o(1) \gtrsim 1$  on  $\Omega$ , where in second step we used the definition (1.11), and in the last step we used that  $\alpha_x \geq 2 + \kappa$ , so that the sequence  $(u_i(\alpha_x))_i$  from (1.10) is exponentially decaying in  $i$ , uniformly in  $r$ .  $\square$

**3.3. Properties of the Green function.** In this subsection, we fix  $\mu \in [0, 1/3)$ . Define

$$\mathcal{J} = [\Lambda(\alpha^*) + \kappa/4, \sqrt{d}/2].$$

We shall use that whenever  $z \in \mathcal{J}$ , all Green functions appearing in our proof are bounded, which is the content of the following result.

**Lemma 3.8.** Suppose (1.5) and (3.2). On  $\Omega$ , for  $z, z' \in \mathcal{J}$  and  $X \subset [N]$  satisfying  $\mathcal{V} \subset X \subset \bigcup_{x \in \mathcal{V}} B_r(x)$ , we have

$$\|(H^{(X)} - z)^{-1}\| \leq 8/\kappa, \tag{3.4}$$

$$\|(H^{(X)} - z)^{-1} - (H^{(X)} - z')^{-1}\| \leq (8/\kappa)^2 |z - z'|. \tag{3.5}$$

*Proof of Lemma 3.8.* Eigenvalue interlacing (Lemma D.3) and Proposition 3.4 (i) and (ii) imply

$$\lambda_2(H^{(X)}) \leq \lambda_2(H^{(\mathcal{V})}) \leq \Lambda(\alpha^*) + \kappa/8 \quad \text{and} \quad \lambda_1(H^{(X)}) = \sqrt{d}(1 + o(1)) \geq \sqrt{d}/2 + \kappa/8.$$

Therefore,  $\text{dist}(z, \text{Spec}(H^{(X)})) \geq \kappa/8$  for any  $z \in \mathcal{J}$ , which proves (3.4). The Lipschitz bound (3.5) follows from (3.4) and the resolvent identity.  $\square$

**Lemma 3.9.** *Suppose (1.5) and (3.2). On  $\Omega$ , for any  $\mathcal{V} \subset X \subset \bigcup_{x \in \mathcal{V}} B_r(x)$ ,  $z \in \mathcal{J}$ , and  $y \notin X$ , we have*

$$-(H^{(X)} - z)_{yy}^{-1} \geq (3z)^{-1}.$$

*Proof of Lemma 3.9.* Denoting by  $\lambda_1 \geq \lambda_2 \geq \dots$  and  $\mathbf{w}_1, \mathbf{w}_2, \dots$  the eigenvalues and eigenvectors of  $H^{(X)}$ , respectively, we have

$$-(H^{(X)} - z)_{yy}^{-1} = \frac{\langle \mathbf{1}_y, \mathbf{w}_1 \rangle^2}{z - \lambda_1} + \sum_{2 \leq i \leq N} \frac{\langle \mathbf{1}_y, \mathbf{w}_i \rangle^2}{z - \lambda_i} \geq \frac{1}{z - \lambda_N} (1 - \langle \mathbf{1}_y, \mathbf{w}_1 \rangle^2) - \left| \frac{\langle \mathbf{1}_y, \mathbf{w}_1 \rangle^2}{\lambda_1 - z} \right|, \tag{3.6}$$

where in the second step we used  $z - \lambda_N \geq z - \lambda_i > 0$  for all  $i \geq 2$ , which follows from eigenvalue interlacing (Lemma D.3), Proposition 3.4 (i), and the condition  $z \in \mathcal{J}$ .

To estimate the right-hand side of (3.6), we use Proposition 3.2 (i) and Proposition 3.1 (ii) as well as (3.2) to estimate  $|X| \leq \sum_{x \in \mathcal{V}} |B_r(x)| \leq N^{1/3+o(1)}$  on  $\Omega$ . From Proposition 3.4 (ii) we therefore deduce that

$$\langle \mathbf{1}_y, \mathbf{w}_1 \rangle^2 = |X^c|^{-1/2} + o(1) = o(1). \tag{3.7}$$

We conclude that the first term on the right-hand side of (3.6) is bounded from below by  $((2 + o(1))z)^{-1}$ , as  $z - \lambda_N \leq 2z$  from Proposition 3.4 (i). The second term on the right-hand side of (3.6) is estimated using (3.7) as well as  $|\lambda_1 - z| \gtrsim \sqrt{d} \geq z$  by Proposition 3.4 (ii) and  $z \in \mathcal{J}$ .  $\square$

#### 4. Exponential Decay of $\mathbf{u}(x)$ and Proof of Proposition 2.2

In this section we establish the exponential decay of  $\mathbf{u}(x)$  around the vertex  $x$ . In particular, Proposition 2.2 is a direct consequence of Proposition 4.1 below. Moreover, we prove in Corollary 4.2 below that  $\mathbf{u}(x)$  is well approximated by  $\mathbf{w}_r(x)$ , the eigenvector of  $H|_{B_r(x)}$  corresponding to its largest eigenvalue. This implies Proposition 2.9 (ii).

Throughout this section we use the high-probability event  $\Omega$  from Definition 3.5.

*4.1. Simple exponential decay of  $\mathbf{u}(x)$ .* In this subsection we establish exponential decay at some positive but not optimal rate.

**Proposition 4.1.** *Suppose that (1.5) holds. Then there is a constant  $c \in (0, 1)$  such that, for each fixed  $\mu \in [0, 1/3)$ , on  $\Omega$ , for each  $x \in \mathcal{W}$  there exists  $q_x > 0$  such that*

$$\|\mathbf{u}(x)|_{B_i(x)^c}\| \lesssim \sqrt{\alpha_x} q_x^i, \quad q_x = \frac{\Lambda(\alpha^*(1/2)) + o(1)}{\lambda(x)} \leq 1 - c, \tag{4.1}$$

for all  $i \in \mathbb{N}$  satisfying  $1 \leq i \leq \min \left\{ \frac{1}{5} - \frac{\mu}{4}, \frac{1}{3} - \mu \right\} \frac{\log N}{\log d} - 2$ .

*Proof.* We note that  $\text{supp } \mathbf{u}(x) \subset ([N] \setminus \mathcal{V}) \cup \{x\}$  and decompose  $\mathbf{u}(x) = \langle \mathbf{1}_x, \mathbf{u}(x) \rangle \mathbf{1}_x + Q\mathbf{u}(x)$ , where  $Q := \sum_{y \in [N] \setminus \mathcal{V}} \langle \mathbf{1}_y, \cdot \rangle \mathbf{1}_y$  is the orthogonal projection on the coordinates in  $[N] \setminus \mathcal{V}$ . We apply the projection  $Q$  to the eigenvalue-eigenvector relation

$$\lambda(x)\mathbf{u}(x) = H^{(\mathcal{V} \setminus \{x\})} \mathbf{u}(x) = \langle \mathbf{1}_x, \mathbf{u}(x) \rangle \mathbf{1}_{S_1(x)} / \sqrt{d} + H^{(\mathcal{V} \setminus \{x\})} Q\mathbf{u}(x),$$



solve for  $Q\mathbf{u}(x)$ , and obtain

$$Q\mathbf{u}(x) = \frac{\langle \mathbf{1}_x, \mathbf{u}(x) \rangle}{\sqrt{d}} (\lambda(x) - H^{(\mathcal{V})})^{-1} \mathbf{1}_{S_1(x)}, \tag{4.2}$$

where we used that  $H^{(\mathcal{V} \setminus \{x\})} \mathbf{1}_x = \mathbf{1}_{S_1(x)} / \sqrt{d}$  and  $QH^{(\mathcal{V} \setminus \{x\})}Q = H^{(\mathcal{V})}$ . We also used that  $\lambda(x) - H^{(\mathcal{V})}$  is invertible, which can be seen as follows. From Proposition 3.4 (ii) with  $X = \mathcal{V}$  and  $r = 0$ , we conclude  $\lambda_1(H^{(\mathcal{V})}) = \sqrt{d}(1 + o(1))$  and, hence, (3.3) and Proposition 3.4 (i) yield that  $\lambda(x)$  is not an eigenvalue of  $H^{(\mathcal{V})}$  and

$$\text{dist}(\lambda(x), \text{spec } H^{(\mathcal{V})}) \gtrsim 1. \tag{4.3}$$

Let

$$Q_i := \sum_{y \in [N] \setminus (\mathcal{V} \cup B_i(x))} \langle \mathbf{1}_y, \cdot \rangle \mathbf{1}_y$$

the projection onto the coordinates in  $[N] \setminus (\mathcal{V} \cup B_i(x))$ . As  $\text{supp } \mathbf{u}(x) \subset ([N] \setminus \mathcal{V}) \cup \{x\}$  and  $Q_i Q = Q_i$ , we conclude from (4.2) that

$$\mathbf{u}(x)|_{B_i(x)^c} = Q_i \mathbf{u}(x) = Q_i Q \mathbf{u}(x) = \frac{\langle \mathbf{1}_x, \mathbf{u}(x) \rangle}{\lambda(x)\sqrt{d}} Q_i \left(1 - \frac{H^{(\mathcal{V})}}{\lambda(x)}\right)^{-1} \mathbf{1}_{S_1(x)}. \tag{4.4}$$

For any  $n \in \mathbb{N}$  we have

$$\left(1 - \frac{H^{(\mathcal{V})}}{\lambda(x)}\right)^{-1} = \sum_{k=0}^n \left(\frac{H^{(\mathcal{V})}}{\lambda(x)}\right)^k + \left(1 - \frac{H^{(\mathcal{V})}}{\lambda(x)}\right)^{-1} \left(\frac{H^{(\mathcal{V})}}{\lambda(x)}\right)^{n+1}. \tag{4.5}$$

Since  $H^{(\mathcal{V})}$  is a local operator, we conclude that  $Q_i (H^{(\mathcal{V})})^k \mathbf{1}_{S_1(x)} = 0$  if  $k + 1 \leq i$ . Hence, fixing  $i \geq 1$  and applying (4.5) with  $n = i - 1$  to (4.4), we get

$$\mathbf{u}(x)|_{B_i(x)^c} = \frac{\langle \mathbf{1}_x, \mathbf{u}(x) \rangle}{\sqrt{d}} Q_i (\lambda(x) - H^{(\mathcal{V})})^{-1} \left(\frac{H^{(\mathcal{V})}}{\lambda(x)}\right)^i \mathbf{1}_{S_1(x)} = O\left(\frac{\|(H^{(\mathcal{V})})^i \mathbf{1}_{S_1(x)}\|}{\sqrt{d}\lambda(x)^i}\right). \tag{4.6}$$

Here, in the last step, we employed  $|\langle \mathbf{1}_x, \mathbf{u}(x) \rangle| \leq 1$ ,  $\|Q_i\| \leq 1$ ,  $\|(\lambda(x) - H^{(\mathcal{V})})^{-1}\| \lesssim 1$  by (4.3).

Let  $r \in \mathbb{N}$  be the largest integer satisfying (3.2). We now claim that, on  $\Omega$ ,

$$\|H^{(\mathcal{V})} \mathbf{v}\| \leq (\Lambda(\alpha^*(1/2)) + o(1)) \|\mathbf{v}\| \tag{4.7}$$

for any  $\mathbf{v}$  such that  $\text{supp } \mathbf{v} \subset B_r(x)$ . Before proving (4.7), we use it to conclude the proof of (4.1).

For  $i \leq r - 2$  we have  $\text{supp}(H^{(\mathcal{V})})^i \mathbf{1}_{S_1(x)} \subset B_r(x)$ , so that iterative applications of (4.7) yield

$$\|(H^{(\mathcal{V})})^i \mathbf{1}_{S_1(x)}\| \lesssim (\Lambda(\alpha^*(1/2)) + o(1))^i \sqrt{\alpha_x} \sqrt{d}. \tag{4.8}$$

Hence, the first bound in (4.1) follows from (4.6). We now show the second bound in (4.1). If  $\frac{\log N}{d} \geq T$  for a sufficiently large constant  $T$  then  $q_x \leq 1 - c$  for some constant  $c > 0$  by using Corollary A.3 with  $v = 1/2$ ,  $\lambda(x) \geq \Lambda(\alpha^*(\mu))$ ,  $\mu \leq 1/3$  and possibly increasing its  $T$ . If  $\frac{\log N}{d} < T \lesssim 1$  then  $\alpha_x \lesssim 1$  by Proposition 3.1 (i) and the estimate  $q_x \leq 1 - c$  for some constant  $c > 0$  follows from  $\Lambda(\alpha_x) \leq 2\sqrt{\alpha_x} \lesssim 1$  and

$\lambda(x) = \Lambda(\alpha_x) + o(1) \geq \Lambda(\alpha^*) + \kappa/4 \geq \Lambda(\alpha^*(1/2)) + \kappa/4$ , which is a consequence of (3.3) and the definition of  $\mathcal{W}$ . This completes the proof of the second bound in (4.1).

What remains is the proof of (4.7). For any  $\mathbf{v} \in \mathbb{R}^{[N]}$ , the Cauchy-Schwarz inequality implies  $\|(\mathbb{E}H)^{(\mathcal{V})}\mathbf{v}\| \leq \sqrt{\frac{d|\text{supp } \mathbf{v}|}{N}} \|\mathbf{v}\|$ . By Proposition 3.1 (ii) and (3.2), if  $\text{supp } \mathbf{v} \subset B_r(x)$  then  $|\text{supp } \mathbf{v}| \leq N^{1/5+o(1)}$ . Therefore,

$$\begin{aligned} \|H^{(\mathcal{V})}\mathbf{v}\| &\leq \|(H - \mathbb{E}H)^{(\mathcal{V})}\mathbf{v}\| + o(\|\mathbf{v}\|) \\ &\leq \|(H - \mathbb{E}H)^{(\{y: \alpha_y \geq \alpha^*(1/2)\})}\mathbf{v}\| + o(\|\mathbf{v}\|) \\ &\leq (\Lambda(\alpha^*(1/2)) + o(1))\|\mathbf{v}\|, \end{aligned}$$

where in the second step, we used  $\mathcal{V}^c \cap B_r(x) \subset \{y : \alpha_y \geq \alpha^*(1/2)\}^c$  by Proposition 3.2 (iv) and, in the third step,  $\|(H - \mathbb{E}H)^{(\{y: \alpha_y \geq \alpha^*(1/2)\})}\| \leq \Lambda(\alpha^*(1/2)) + o(1)$  by Lemma C.1.<sup>8</sup> This concludes the proof of (4.7).  $\square$

4.2. *Approximating  $\mathbf{u}(x)$  by  $\mathbf{w}_r(x)$ .* From Proposition 4.1 and its proof, we deduce the following result, which compares  $\mathbf{u}(x)$  and  $\mathbf{w}_r(x)$  from Definition 1.5.

**Corollary 4.2.** *Suppose that (1.5) holds and fix  $\mu \in [0, 1/3)$ . Then, on  $\Omega$ , for all  $x \in \mathcal{W}$  and  $r \in \mathbb{N}$  satisfying (2.19),*

$$\mathbf{u}(x) = \mathbf{w}_r(x) + O(\alpha_x q_x^r \Lambda(\alpha_x)^{-1}) = \mathbf{w}_r(x) + o(\Lambda(\alpha_x)^{-1}),$$

where  $q_x$  is the same as in Proposition 4.1.

*Proof.* We shall apply Lemma D.2 with  $M = H|_{B_r(x)}$ ,  $\widehat{\lambda} = \lambda(x)$  and  $\mathbf{v} = \mathbf{u}(x)$ . First, we check its conditions by studying the spectral gap of  $H|_{B_r(x)}$  around its largest eigenvalue. As  $r \leq \frac{1}{6} \frac{\log N}{\log d}$ , we conclude from Proposition 3.4 (iii) with  $X = B_r(x)^c$  that  $H|_{B_r(x)} = H^{(B_r(x)^c)}$  has an eigenvalue  $\Lambda(\alpha_x) + o(1)$ . The bound (4.7) implies that  $\|H|_{B_r(x) \setminus \{x\}}\| \leq \|H^{(\mathcal{V})}\| \leq \Lambda(\alpha^*(1/2)) + o(1)$ , where in the first step we used Proposition 3.2 (iii). By eigenvalue interlacing (Lemma D.3), we therefore deduce that

$$\lambda_1(H|_{B_r(x)}) = \Lambda(\alpha_x) + o(1), \quad \lambda_2(H|_{B_r(x)}) \leq \lambda_1(H|_{B_r(x) \setminus \{x\}}) \leq \Lambda(\alpha^*(1/2)) + o(1).$$

Thus, the definition of  $\mathcal{W}$  in (2.2) implies

$$\lambda_1(H|_{B_r(x)}) - \lambda_2(H|_{B_r(x)}) \gtrsim \Lambda(\alpha_x) \left(1 - \frac{\Lambda(\alpha^*(1/2))}{\Lambda(\alpha^*(\mu)) + \kappa/2}\right) \gtrsim \Lambda(\alpha_x),$$

where the last inequality follows from Corollary A.3 if  $\frac{\log N}{d} \geq T$  for some large enough constant  $T$ , and from  $\Lambda(\alpha^*(1/2)) + \kappa/2 \leq \Lambda(\alpha^*(\mu)) + \kappa/2 \leq \Lambda(\alpha_x) \lesssim 1$  (see (3.1)) otherwise. Hence, owing to (3.3), there is  $\Delta \gtrsim \Lambda(\alpha_x)$  such that  $H|_{B_r(x)}$  has precisely one eigenvalue in  $[\lambda(x) - \Delta, \lambda(x) + \Delta]$ .

Let  $P_r$  be the orthogonal projection onto the coordinates in  $B_r(x)$ . The eigenvalue-eigenvector relation  $(H^{(\mathcal{V} \setminus \{x\})} - \lambda(x))\mathbf{u}(x) = 0$  and  $P_r H^{(\mathcal{V} \setminus \{x\})} P_r = H|_{B_r(x)}$  imply

$$(H|_{B_r(x)} - \lambda(x))\mathbf{u}(x) = -P_r H^{(\mathcal{V} \setminus \{x\})} (1 - P_r)\mathbf{u}(x) - \lambda(x)(1 - P_r)\mathbf{u}(x).$$

<sup>8</sup> Note that Lemma C.1 holds on  $\Omega$ , since Lemma C.1 is used to prove Proposition 3.4 (i).

Therefore, since  $(1 - P_r)\mathbf{u}(x) = \mathbf{u}(x)|_{B_r(x)^c}$ , we get

$$\|(H|_{B_r(x)} - \lambda(x))\mathbf{u}(x)\| \leq \lambda(x)\|\mathbf{u}(x)|_{B_r(x)^c}\| + \|P_r H^{(\mathcal{V}\setminus\{x\})}(\mathbf{u}(x)|_{B_r(x)^c})\| \lesssim \alpha_x q_x^r. \tag{4.9}$$

In the last step, we used  $\lambda(x) \lesssim \alpha_x^{1/2}$  and Proposition 4.1 to estimate the first term. For the second term, we used that  $P_r H^{(\mathcal{V}\setminus\{x\})}(\mathbf{u}(x)|_{B_r(x)^c}) = P_r H^{(\mathcal{V}\setminus\{x\})}(\mathbf{u}(x)|_{S_{r+1}(x)})$  by the locality of  $H^{(\mathcal{V}\setminus\{x\})}$  as well as the identity  $H^{(\mathcal{V}\setminus\{x\})}(\mathbf{u}(x)|_{S_{r+1}(x)}) = H^{(\mathcal{V})}(\mathbf{u}(x)|_{S_{r+1}(x)})$ , which yield

$$\|H^{(\mathcal{V}\setminus\{x\})}(\mathbf{u}(x)|_{S_{r+1}(x)})\| \lesssim \Lambda(\alpha^*(1/2))\|\mathbf{u}(x)|_{S_{r+1}(x)}\| \lesssim \alpha_x^{1/2}\|\mathbf{u}(x)|_{B_r(x)^c}\| \lesssim \alpha_x q_x^r$$

due to (4.7) as  $r + 1 \leq \min\{\frac{1}{5} - \frac{\mu}{4}, \frac{1}{3} - \mu\} \frac{\log N}{\log d}$ ,  $\Lambda(\alpha^*(1/2))^2 \leq \Lambda(\alpha^*)^2 \lesssim \alpha_x$ , and Proposition 4.1.

Since  $\alpha_x \lesssim \frac{\log N}{d}$  by Proposition 3.1 (i),  $q_x \leq 1 - c$  for some constant  $c > 0$  by Proposition 4.1,  $r \gg \log d \gtrsim \log(\frac{10 \log N}{d})$  by (1.5) and  $\alpha_x \geq 2$ , we obtain

$$\alpha_x q_x^r \ll 1 \lesssim \alpha_x^{1/2}. \tag{4.10}$$

Therefore, owing to (4.9) and  $\alpha_x^{1/2} \asymp \Lambda(\alpha_x) \lesssim \Delta$ , we can now apply Lemma D.2 with  $M = H|_{B_r(x)}$ ,  $\hat{\lambda} = \lambda(x)$  and  $\mathbf{v} = \mathbf{u}(x)$  to conclude  $\mathbf{u}(x) = \mathbf{w}_r(x) + O(\alpha_x q_x^r \Lambda(\alpha_x)^{-1})$  from (4.9) and  $\Delta \gtrsim \Lambda(\alpha_x)$  as well as  $\langle \mathbf{1}_x, \mathbf{u}(x) \rangle \geq 0$  and  $\langle \mathbf{1}_x, \mathbf{w}_r(x) \rangle \geq 0$ . This proves the first equality in Corollary 4.2.

Since  $\alpha_x q_x^r \ll 1$  by (4.10), the last equality in Corollary 4.2 follows immediately.  $\square$

**4.3. Optimal exponential decay of  $\mathbf{u}(x)$ .** In this subsection we establish an explicit rate of exponential decay of  $\mathbf{u}(x)$ . It holds only for a smaller set of eigenvectors near the spectral edge, requiring  $\mu$  to be small enough. As pointed out in Remark 1.4, up to the error  $O(\varepsilon)$ , this rate is optimal.

**Proposition 4.3.** *Suppose (1.5). Then the following holds.*

- (i) *(Subcritical regime) There are constants  $T \geq 1$  and  $c > 0$  such that if  $\frac{\log N}{d} \geq T$  then, on  $\Omega$ , for any small enough constant  $\varepsilon > 0$  and each  $\mu \in [0, \varepsilon]$ ,*

$$\|\mathbf{u}(x)|_{B_i(x)^c}\| \lesssim \sqrt{\alpha_x} \left( \frac{1 + O(\varepsilon)}{\sqrt{\alpha_x - 1}} \right)^{i+1}$$

for all  $x \in \mathcal{W}$  and  $i \leq c \frac{\log N}{(\log d) \log \frac{10 \log N}{d}}$ .

- (ii) *(Critical regime) There is a constant  $c > 0$  such that, for any constants  $T \geq 1$  and  $\varepsilon > 0$  with  $\varepsilon \leq c\kappa$ , if  $\frac{\log N}{d} \leq T$  and  $\mu \in [0, cT^{-1}\varepsilon^2]$  then, with high probability,*

$$\|\mathbf{u}(x)|_{B_i(x)^c}\| \lesssim \sqrt{\alpha_x} \left( \frac{1 + O(\varepsilon)}{\sqrt{\alpha_x - 1}} \right)^{i+1}$$

for all  $x \in \mathcal{W}$  and  $i \leq \frac{c\varepsilon^2 \log N}{T \log d}$ .

*Proof.* Let  $x \in \mathcal{W}$ . For any  $i, n \in \mathbb{N}$ , we conclude from (4.4) and (4.5) (see also (4.6)) that

$$\|\mathbf{u}(x)|_{B_i(x)^c}\| \lesssim \frac{1}{\lambda(x)\sqrt{d}} \left\| Q_i \sum_{k=0}^n \left( \frac{H^{(\mathcal{V})}}{\lambda(x)} \right)^k \mathbf{1}_{S_1(x)} \right\| + \frac{\|(H^{(\mathcal{V})})^{n+1} \mathbf{1}_{S_1(x)}\|}{\sqrt{d}\lambda(x)^{n+1}}$$

since  $|\langle \mathbf{1}_x, \mathbf{u}(x) \rangle| \lesssim 1$ ,  $\|Q_i\| \leq 1$  and  $\|(\lambda(x) - H^{(\mathcal{V})})^{-1}\| \lesssim 1$  by (4.3). We denote the right-hand side of (3.2) by  $r$ . In the following, we always assume that  $n+1 \leq r$  and tacitly use the graph properties listed in Proposition 3.2. As in the proof of Proposition 4.1 (see (4.8) and use  $q_x \leq 1 - \varepsilon$  by (4.1)), we find some constant  $\varepsilon > 0$  such that if  $n + 1 \leq r$  then

$$\frac{\|(H^{(\mathcal{V})})^{n+1} \mathbf{1}_{S_1(x)}\|}{\sqrt{d}\lambda(x)^{n+1}} \lesssim \sqrt{\alpha_x}(1 - \varepsilon)^{n+1}. \tag{4.11}$$

We shall use the following result, whose proof is given at the end of this subsection.

**Claim 4.4** *Suppose that all vertices in  $B_{r-1}(x) \setminus \{x\}$  have degree at most  $\tau d$ , where  $2 \leq 2\sqrt{\tau} < \lambda(x)$ . Then for all  $n \leq r - 1$  and  $i \in \mathbb{N}$  we have on  $\mathcal{W}$*

$$\frac{1}{\lambda(x)\sqrt{d}} \left\| Q_i \sum_{k=0}^n \left( \frac{H^{(\mathcal{V})}}{\lambda(x)} \right)^k \mathbf{1}_{S_1(x)} \right\| \lesssim \sqrt{\alpha_x} \left( \frac{2}{\lambda(x) + \sqrt{\lambda(x)^2 - 4\tau}} \right)^{i+1}. \tag{4.12}$$

We now explain how we choose  $\tau$  and  $n$  in the subcritical and the critical regimes in order to deduce Proposition 4.3 from Claim 4.4. In the subcritical regime, i.e. for the proof of (i), we choose  $\tau := \alpha^*(1 - 3\varepsilon)$  and  $n \leq \varepsilon \frac{\log N}{\log d}$  for some small enough constant  $\varepsilon > 0$ . By arguing similarly as in the proof of Corollary A.3 below, we find a constant  $T \geq 1$  such that

$$\frac{4\tau}{\lambda(x)^2} \leq \frac{4\alpha^*(1 - \varepsilon)(1 + o(1))}{\Lambda(\alpha^*(\mu))^2} = O(\varepsilon) < 1 \tag{4.13}$$

by (3.3) and (2.2) if  $\frac{\log N}{d} \geq T$ . Here, the last inequality holds if  $\varepsilon$  is sufficiently small. Hence, the assumption on  $\tau$  in Claim 4.4 holds. Moreover, with our choice of  $\tau$ , the degrees in  $B_n(x) \setminus \{x\}$  are bounded by  $\tau d$  due to Proposition 3.2 (iv) and our assumption on  $n$ . Hence, the assumptions of Claim 4.4 hold. As  $\lambda(x) \lesssim \sqrt{\frac{\log N}{d}}$  by (3.3) and Proposition 3.1 (i), the right-hand side of (4.11) is bounded by the right-hand side of (4.12) if  $n = iC \log\left(\frac{10 \log N}{d}\right)$  for some sufficiently large constant  $C$ . Since we need that  $n + 1 \leq r$ , this yields the upper bound on  $i$  in (i). From (4.13) and (3.3), we deduce that  $\lambda(x) + \sqrt{\lambda(x)^2 - 4\tau} = \lambda(x)\left(1 + \sqrt{1 - \frac{4\tau}{\lambda(x)^2}}\right) = \Lambda(\alpha_x)(1 + o(1))(2 + O(\varepsilon)) \geq 2\sqrt{\alpha_x - 1}(1 + O(\varepsilon))^{-1}$ , which completes the proof of (i).

For the proof of (ii), we note that  $\mu \leq c\varepsilon^2 T^{-1} < 1/3$  for a sufficiently small constant  $c > 0$  since  $T \geq 1$ ,  $\varepsilon \leq c\kappa$  and  $\kappa \leq 1$ . We choose  $\tau = 1 + \varepsilon$  for some constant  $\varepsilon > 0$ . As  $\lambda(x)^2 - 4\tau = (\lambda(x)^2 - 4)\left(1 - \frac{4\varepsilon}{\lambda(x)^2 - 4}\right) > 0$  due to (3.3) and (2.2) if  $\varepsilon \leq c\kappa$  with a small enough constant  $c > 0$ . This establishes the first condition for (4.12). We set  $R := \frac{c\varepsilon^2 \log N}{T \log d}$  and conclude from Bennett’s inequality (Lemma D.1 below) and Proposition 3.2 that

$$\begin{aligned} & \mathbb{P}\left(\alpha_y \leq \tau \text{ for all } x \in \mathcal{W} \text{ and for all } y \in B_R(x) \setminus \{x\}\right) \\ & \leq \exp\left(\left(\mu + R \frac{\log d}{\log N} - \frac{d}{\log N} h(\varepsilon)\right) \log N\right), \end{aligned}$$

where  $h(\varepsilon) := (1 + \varepsilon) \log(1 + \varepsilon) - \varepsilon$ . Since  $h(\varepsilon) \geq 3c\varepsilon^2$  for some small enough constant  $c > 0$ , the upper bound on  $\mu$  imposed in the statement, the definition of  $R$  and  $\frac{d}{\log N} \geq \frac{1}{T}$  imply that the factor in front of  $\log N$  is negative. Therefore, with high probability, we can apply (4.12) simultaneously for all  $x \in \mathcal{W}$ . We choose  $n = C(i + 1)$  and deduce that if  $C > 0$  is a large enough constant then the error in (4.11) is dominated by the right-hand side of (4.12) as  $\lambda(x) \lesssim 1$  for  $\frac{\log N}{d} \leq T$ . We recall that  $r$  denotes the right-hand side of (3.2) and note that the condition of (4.11),  $n + 1 = C(i + 1) + 1 \leq r$ , can be satisfied by possibly decreasing the constant  $c > 0$ . Finally, we obtain  $\lambda(x) + \sqrt{\lambda(x)^2 - 4\tau} = \lambda(x) + \sqrt{\lambda(x)^2 - 4\tau} + O\left(\frac{\varepsilon}{\sqrt{\lambda(x)^2 - 4\tau}}\right) = 2\sqrt{\alpha_x - 1}(1 + O(\varepsilon))^{-1}$  similarly as argued above and in the proof of (i) using  $\varepsilon \leq c\kappa$  and  $\kappa \leq 1$ . This proves (ii) and, thus, Proposition 4.3.  $\square$

*Proof of Claim 4.4.* For  $y \in S_1(x)$  we denote by  $B_n^+(y)$  the ball of radius  $n$  around  $y$  in the graph  $\mathbb{G}|_{[N] \setminus \{x\}}$ , and we write  $S_n^+(y) := B_n^+(y) \setminus B_{n-1}^+(y)$ . Since  $\mathbb{G}|_{B_{n+1}(x) \setminus \mathcal{V}}$  is a forest on  $\Omega$  and for each  $y \in S_1(x)$ ,  $\mathbb{G}|_{B_n^+(y)}$  is a tree, we obtain

$$(A^{(\mathcal{V})})^k \mathbf{1}_{S_1(x)} = \sum_{y \in S_1(x)} \sum_{j=1}^{k+1} \sum_{z \in S_{j-1}^+(y)} \mathbf{1}_z N_k(y, z),$$

where  $N_k(y, z)$  denotes the number of walks on  $\mathbb{G}|_{B_n^+(y)}$  of length  $k$  between  $y$  and  $z$ . Thus, for  $i \geq 0$ , we conclude

$$\begin{aligned} Q_i \sum_{k=0}^n \frac{1}{\lambda(x)^k d^{k/2}} (A^{(\mathcal{V})})^k \mathbf{1}_{S_1(x)} &= Q_i \sum_{y \in S_1(x)} \sum_{j=1}^{n+1} \sum_{z \in S_{j-1}^+(y)} \sum_{k=j-1}^n \mathbf{1}_z \frac{N_k(y, z)}{\lambda(x)^k d^{k/2}} \\ &= \sum_{y \in S_1(x)} \sum_{j=i+1}^{n+1} \sum_{z \in S_{j-1}^+(y)} \mathbf{1}_z \left( \sum_{k=j-1}^n \frac{N_k(y, z)}{\lambda(x)^k d^{k/2}} \right), \end{aligned}$$

where the second step follows from the definition of  $Q_i$ . Since the sets  $\{B_n^+(y) : y \in S_1(x)\}$ , are disjoint, we conclude that

$$\left\| Q_i \sum_{k=0}^n \left( \frac{H^{(\mathcal{V})}}{\lambda(x)} \right)^k \mathbf{1}_{S_1(x)} \right\|^2 = \sum_{y \in S_1(x)} \sum_{j=i+1}^{n+1} \sum_{z \in S_{j-1}^+(y)} \left( \sum_{k=j-1}^n \frac{N_k(y, z)}{\lambda(x)^k d^{k/2}} \right)^2. \tag{4.14}$$

Next, let  $z \in S_{j-1}^+(y)$  for some  $y \in S_1(x)$ . We note that  $N_k(y, z) = 0$  if  $k - (j - 1)$  is odd due to the bipartite structure of a tree. If this difference is even, then we now show that

$$N_k(y, z) \leq (M^k)_{1j} (\tau d)^{(k-(j-1))/2}, \tag{4.15}$$

where  $M$  is the adjacency matrix of  $\mathbb{N}^*$  (regarded as a graph where consecutive numbers are adjacent), see (D.2) below for a precise definition, under the assumption that the degree of each vertex in  $\mathbb{G}|_{B_n^+(y)}$  is bounded by  $\tau d$ .

For the proof of (4.15), we introduce the set of walks on  $\mathbb{N}^*$

$$W_k(j) := \{\text{walks } \gamma \text{ on } \mathbb{N}^* \text{ of length } k \text{ such that } \gamma(0) = 1 \text{ and } \gamma(k) = j\},$$

and for each  $\gamma \in W_k(j)$  we introduce the set of walks on  $\mathbb{G}|_{B_n^+(y)}$  that project down to  $\gamma$ ,

$$W_k(y, z; \gamma) := \{\text{walks } \Gamma \text{ on } \mathbb{G}|_{B_n^+(y)} \text{ of length } k \\ \text{such that } \Gamma(0) = y, \Gamma(k) = z \text{ and } d(x, \Gamma(\cdot)) = \gamma\}.$$

By definition, any walk  $\Gamma \in W_k(y, z; \gamma)$  projects down to a walk  $\gamma \in W_k(j)$ , which implies

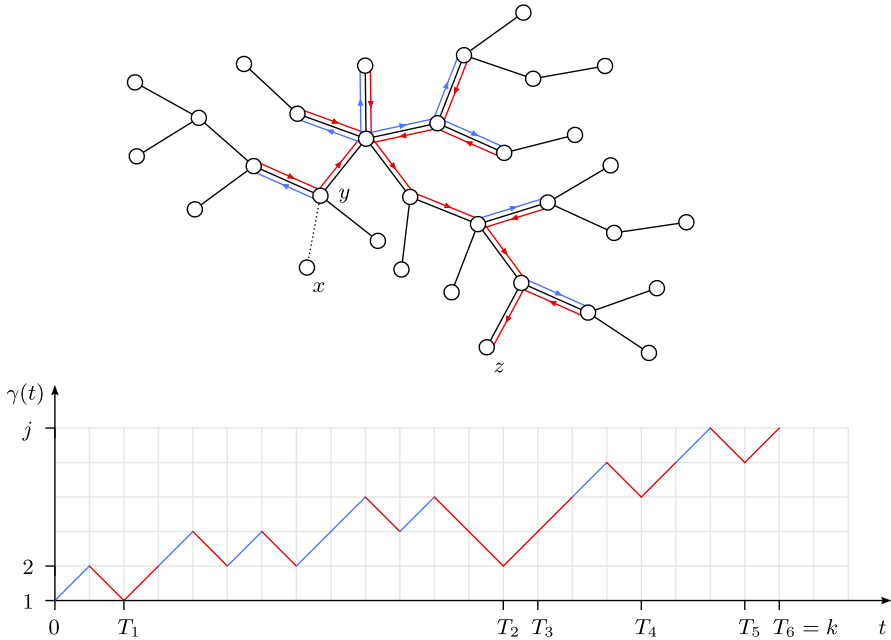
$$N_k(y, z) \leq \sum_{\gamma \in W_k(j)} |W_k(y, z; \gamma)|. \tag{4.16}$$

In order to prove (4.15), we fix  $\gamma \in W_k(j)$  and estimate  $|W_k(y, z; \gamma)|$ . For  $i \in [j]$ , let  $T_i := \max\{t \in \{0, \dots, k\} : \gamma(t) = i\}$ , i.e.  $T_i$  is the last time when  $\gamma$  hits  $i$ . Clearly,  $T_1 < T_2 < \dots < T_{j-1} < T_j = k$ . See Fig. 5 for an illustration of the walks  $\gamma$  and  $\Gamma$ . At each time  $T_i$  with  $i \in [j - 1]$ , the walk  $\gamma$  takes a step to the right, and by definition of the times  $T_i$ , any walk  $\Gamma \in W_k(y, z; \gamma)$  takes a step outwards on the geodesic from  $y$  to  $z$ . This means that  $j - 1$  of the  $k$  steps of  $\Gamma$  are fixed. Of the remaining  $k - (j - 1)$  steps, half correspond to steps to the left of  $\gamma$ , which again correspond to a uniquely determined step of  $\Gamma$  along the unique path back towards  $y$ . Hence, only  $(k - (j - 1))/2$  of the  $k$  steps of  $\Gamma$  are free to choose. Since the degrees in  $\mathbb{G}|_{B_n^+(y)}$  are bounded by  $\tau d$ , we obtain  $|W_k(y, z; \gamma)| \leq (\tau d)^{(k-(j-1))/2}$ . Plugging this estimate into (4.16) implies (4.15), since  $|W_k(j)| = (M^k)_{1j}$  by definition of  $M$ .

Next, applying (4.15) to (4.14), we obtain

$$\begin{aligned} \left\| Q_i \sum_{k=0}^n \left( \frac{H^{(\mathcal{V})}}{\lambda(x)} \right)^k \mathbf{1}_{S_1(x)} \right\|^2 &\leq \sum_{j=i+1}^{n+1} |S_j(x)| \left( \sum_{k=j-1}^n \frac{(M^k)_{1j} (\tau d)^{(k-(j-1))/2}}{\lambda(x)^k d^{k/2}} \right)^2 \\ &\leq \sum_{j=i+1}^{n+1} |S_j(x)| \frac{1}{(\tau d)^{j-1}} \left( \sum_{k=j-1}^n \frac{(M^k)_{1j} \tau^{k/2}}{\lambda(x)^k} \right)^2 \\ &\lesssim \sum_{j=i+1}^{n+1} |S_1(x)| \frac{1}{\tau^{j-1}} \left( \sum_{k=0}^{\infty} \frac{(M^k)_{1j} \tau^{k/2}}{\lambda(x)^k} \right)^2 \\ &= \sum_{j=i+1}^{n+1} |S_1(x)| \frac{1}{\tau^{j-1}} \left( \left( 1 - \frac{\sqrt{\tau}}{\lambda(x)} M \right)_{1j}^{-1} \right)^2 \\ &= \sum_{j=i+1}^{n+1} |S_1(x)| \lambda(x)^2 \left( \frac{2}{\lambda(x) + \sqrt{\lambda(x)^2 - 4\tau}} \right)^{2j} \\ &\lesssim |S_1(x)| \lambda(x)^2 \left( \frac{2}{\lambda(x) + \sqrt{\lambda(x)^2 - 4\tau}} \right)^{2(i+1)}. \end{aligned}$$

Here, in the third step, we used that  $|S_j(x)| \lesssim |S_1(x)| d^{j-1}$  by Proposition 3.1 (ii) and that  $(M^k)_{1j} \geq 0$  for all  $k \in \mathbb{N}$ . The fourth step follows from the condition  $2\sqrt{\tau}/\lambda(x) < 1$ , the invertibility and the Neumann series representation of  $M$  in Lemma D.5 below with  $t = \lambda(x)/\sqrt{\tau}$ . The fifth step is a consequence of the representation of  $(1 - \frac{\sqrt{\tau}}{\lambda(x)} M)_{1j}^{-1}$  in Lemma D.5 below. In the last step, we used that  $\lambda(x) = \Lambda(\alpha_x) + o(1) \geq 2 + \kappa/4$  to sum up the geometric series and conclude that the series is  $\lesssim 1$ . This completes the proof of (4.12) and, thus, the one of Claim 4.4.  $\square$



**Fig. 5.** An illustration of two walks  $\gamma \in W_k(j)$  (bottom) and  $\Gamma \in W_k(y, z; \gamma)$  (top). Here  $k = 21$  and  $j = 6$ . By definition of  $W_k(y, z; \gamma)$ , for each  $t \in \{0, \dots, k\}$  we have  $\gamma(t) = d(x, \Gamma(t))$ . The time  $T_i$  is the last time when  $\gamma$  hits  $i$ . We draw an edge  $\{\Gamma(t), \Gamma(t + 1)\}$  in red if the choice of the vertex  $\Gamma(t + 1)$  is uniquely determined by  $\gamma$  and in blue otherwise. In the latter case, there are at most  $\tau d$  possible choices for  $\Gamma(t + 1)$ . Red edges arise in two ways: (i) a step to the right in  $\gamma$  following a time  $T_i$  ( $j - 1$  in total, in  $\Gamma$  corresponding to a step towards  $z$  along the geodesic from  $y$  to  $z$ ); (ii) a step to the left in  $\gamma$  ( $(k - (j - 1))/2$  in total, in  $\Gamma$  corresponding to a step towards  $y$ )

### 5. Approximate Eigenvalues: Proof of Proposition 2.3

In this section we prove Proposition 2.3, by showing that in the interval  $\mathcal{I}$  there is a one-to-one correspondence between eigenvalues of  $H$  and the points  $\lambda(x)$  for  $x \in \mathcal{W}$ , up to a polynomially small error term.

5.1. Proof of Proposition 2.3. We recall  $\mathbf{w}_1$  from Definition 3.3.

**Definition 5.1.** Let  $\Pi$  be the orthogonal projection onto  $\text{span}(\{\mathbf{w}_1\} \cup \{\mathbf{u}(x) : x \in \mathcal{W}\})$  and  $\bar{\Pi} := 1 - \Pi$ .

Note that the set  $\{\mathbf{w}_1\} \cup \{\mathbf{u}(x) : x \in \mathcal{W}\}$  in the definition of  $\Pi$  is not orthogonal. Throughout the proof, we regard  $H$  as a block matrix associated with the orthogonal sum decomposition  $\text{ran } \Pi \oplus (\text{ran } \Pi)^\perp$ .

**Proposition 5.2.** Suppose (1.5). Fix  $\mu \in [0, 1/3)$  and  $\zeta \in [0, 1/2 - \mu)$ . With high probability, the following holds.

- (i)  $\|(H - \lambda(x))\mathbf{u}(x)\| \leq N^{-\zeta}$  for all  $x \in \mathcal{W}$ .
- (ii) If  $\zeta < 1/2 - 3\mu/2$  then

$$\text{spec}(\Pi H \Pi) \setminus \{0\} = \{\lambda_1(H)\} \cup \{\lambda(x) + \varepsilon_x : x \in \mathcal{W}\}$$



counted with multiplicity, where  $|\varepsilon_x| \leq N^{-\zeta}$  for all  $x \in \mathcal{W}$  and  $\lambda_1(H) = \sqrt{d}(1 + o(1))$ .

(iii) If  $\zeta < 1/2 - 3\mu/2$  then  $\|\overline{\Pi}H\Pi\| \leq N^{-\zeta}$ .

(iv) If  $\mu < 1/4$  then  $\lambda_1(\overline{\Pi}H\Pi) \leq \Lambda(\alpha^*) + \kappa/2 + o(1)$ .

*Proof of Proposition 2.3.* Owing to the block decomposition  $H = \Pi H \Pi + \overline{\Pi} H \overline{\Pi} + \overline{\Pi} H \Pi + \Pi H \overline{\Pi}$ , Proposition 5.2 (iii) yields

$$\text{spec}(H) \setminus \{0\} = \{\lambda + \varepsilon_\lambda : \lambda \in (\text{spec}(\Pi H \Pi) \cup \text{spec}(\overline{\Pi} H \overline{\Pi}))\} \setminus \{0\}$$

counted with multiplicities, where  $|\varepsilon_\lambda| \leq 2\|\overline{\Pi}H\Pi\| \leq N^{-\zeta}$  for all  $\lambda$ . Therefore, Proposition 2.3 follows from the definition of  $\mathcal{I}$  as well as Proposition 5.2 (ii), (iv), and (i).  $\square$

The rest of this section is devoted to the proof of Proposition 5.2. We assume the condition (1.5) throughout. We recall the definition of the high probability event  $\Omega$  from Definition 3.5. For any event  $A$  and random variable  $X$  we write

$$\mathbb{P}_\Omega(A) := \mathbb{P}(\Omega \cap A), \quad \mathbb{E}_\Omega[X] := \mathbb{E}[X \mathbb{1}_\Omega]. \tag{5.1}$$

5.2. *Proof of Proposition 5.2 (i).* The proof of Proposition 5.2 (i) relies on the following result, whose proof is given at the end of this subsection.

**Proposition 5.3.** *Let  $\mu \in [0, 1/3)$ . For any  $x \in \mathcal{W}$ , we have the decomposition*

$$(H - \lambda(x))\mathbf{u}(x) = \sum_{y \in \mathcal{V} \setminus \{x\}} \varepsilon_y(x) \mathbf{1}_y. \tag{5.2}$$

Moreover, for any  $x, y \in [N]$ , we have the estimate

$$\mathbb{E}_\Omega[\mathbb{1}_{x \in \mathcal{W}} \mathbb{1}_{y \in \mathcal{V} \setminus \{x\}} \varepsilon_y(x)^2] \leq d^{-1} (10 \log N)^2 N^{2\mu-3}.$$

From Proposition 5.3, for any  $x \in [N]$ , we obtain

$$\mathbb{E}_\Omega[\mathbb{1}_{x \in \mathcal{W}} \|(H - \lambda(x))\mathbf{u}(x)\|^2] \leq d^{-1} (10 \log N)^2 N^{2\mu-2}. \tag{5.3}$$

*Proof of Proposition 5.2 (i).* From (5.3), a union bound, and Chebyshev’s inequality, we conclude that

$$\mathbb{P}(\exists x \in \mathcal{W}, \|(H - \lambda(x))\mathbf{u}(x)\| > N^{-\zeta}) \leq \mathbb{P}(\Omega^c) + N^{2\zeta+2\mu-1+o(1)}$$

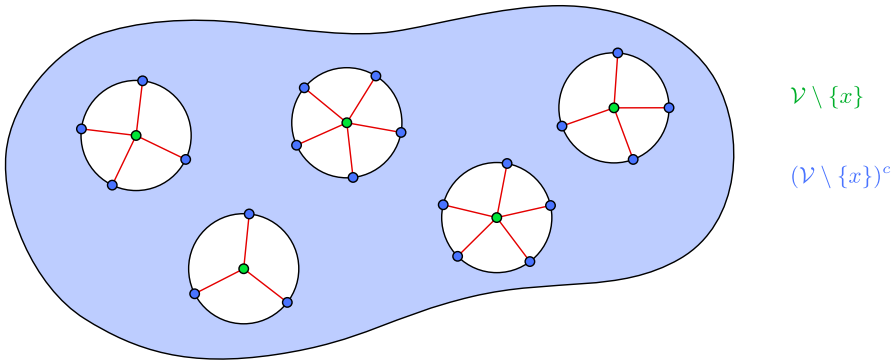
for any  $\zeta > 0$ . This proves Proposition 5.2 (i) since  $\zeta < 1/2 - \mu$  by assumption, and  $\mathbb{P}(\Omega^c) = o(1)$  by Propositions 3.1, 3.2 and 3.4.  $\square$

We shall need modifications of the sets  $\mathcal{V}$  and  $\mathcal{W}$  defined in (2.1) and (2.2), respectively. For  $X \subset [N]$  we define

$$\mathcal{V}^{(X)} := \left\{ y \in [N] \setminus X : \frac{|S_1(x) \setminus X|}{d} \geq \alpha^*(\mu) \right\}, \tag{5.4}$$

$$\mathcal{W}^{(X)} := \left\{ y \in \mathcal{V}^{(X)} : \Lambda\left(\frac{|S_1(x) \setminus X|}{d}\right) \geq \Lambda(\alpha^*(\mu)) + \kappa/2 \right\}. \tag{5.5}$$

The point of these definitions is that  $\mathcal{V}^{(X)}$  and  $\mathcal{W}^{(X)}$  depend only on the edges in  $\mathbb{G}|_{X^c}$ . The following remark states that on the event  $\Omega$  the effect of the upper index in these definitions amounts simply to excluding vertices.



**Fig. 6.** An illustration of the identity (5.6). The set  $\mathcal{V} \setminus \{x\}$  is drawn in green and its complement  $(\mathcal{V} \setminus \{x\})^c$  in blue. The blue neighbours of the green vertices are drawn explicitly, and the remaining blue vertices are represented by the shaded blue region. The edges of incident to  $\mathcal{V} \setminus \{x\}$  are drawn in red. The adjacency matrix of these red edges is  $A - A^{(\mathcal{V} \setminus \{x\})}$ . The vector  $\mathbf{u}(x)$  is supported on the blue vertices, and hence the vector  $(A - A^{(\mathcal{V} \setminus \{x\})})\mathbf{u}(x)$  is supported on the green vertices. Its value at a green vertex  $y$  equals the sum of the entries of  $\mathbf{u}(x)$  at the blue vertices adjacent to  $y$

*Remark 5.4.* On  $\Omega \cap \{y \in \mathcal{V}\}$ , owing to Proposition 3.2 (iii), we have  $\mathcal{V}^{(y)} = \mathcal{V} \setminus \{y\}$  and  $\mathcal{W}^{(y)} = \mathcal{W} \setminus \{y\}$ .

*Proof of Proposition 5.3.* Note that  $H - H^{(\mathcal{V} \setminus \{x\})}$  is  $d^{-1/2}$  times the adjacency matrix of the subgraph of  $\mathbb{G}$  containing the edges incident to  $\mathcal{V} \setminus \{x\}$ . Hence, because  $(H^{(\mathcal{V} \setminus \{x\})} - \lambda(x))\mathbf{u}(x) = 0$  and  $\text{supp } \mathbf{u}(x) \subset (\mathcal{V} \setminus \{x\})^c$ , we find

$$(H - \lambda(x))\mathbf{u}(x) = (H - H^{(\mathcal{V} \setminus \{x\})})\mathbf{u}(x) = \sum_{y \in \mathcal{V} \setminus \{x\}} \varepsilon_y(x)\mathbf{1}_y, \quad \varepsilon_y(x) := \frac{1}{\sqrt{d}} \sum_{t \in \mathcal{S}_1(y)} \langle \mathbf{1}_t, \mathbf{u}(x) \rangle. \tag{5.6}$$

See Fig. 6 for an illustration. The Cauchy-Schwarz inequality implies

$$\varepsilon_y(x)^2 \leq \frac{|\mathcal{S}_1(y)|}{d} \sum_{t \in \mathcal{S}_1(y)} \langle \mathbf{1}_t, \mathbf{u}(x) \rangle^2 \leq \frac{10 \log N}{d} \sum_{t \in \mathcal{S}_1(y)} \langle \mathbf{1}_t, \mathbf{u}(x) \rangle^2 \tag{5.7}$$

on  $\Omega$  due to Proposition 3.1 (i). From now on we fix  $x, y \in [N]$ . Moreover, let  $\tilde{\mathbf{u}}(x)$  be the eigenvector of  $H^{(\mathcal{V}^{(y)} \cup \{y\} \setminus \{x\})}$  associated with its second largest eigenvalue  $\lambda_2(H^{(\mathcal{V}^{(y)} \cup \{y\} \setminus \{x\})})$  and satisfying  $\langle \mathbf{1}_x, \tilde{\mathbf{u}}(x) \rangle \geq 0$ . On  $\Omega \cap \{y \in \mathcal{V}\}$ , we have  $\mathcal{V} = \mathcal{V}^{(y)} \cup \{y\}$  by Remark 5.4 and, thus,  $\mathbf{u}(x) = \tilde{\mathbf{u}}(x)$ . As  $x \neq y$ , on  $\Omega \cap \{y \in \mathcal{V}\}$ , the events  $\{x \in \mathcal{W}\}$  and  $\{x \in \mathcal{W}^{(y)}\}$  coincide by Remark 5.4. Therefore, since  $\mathcal{W}^{(y)}$  and  $\tilde{\mathbf{u}}(x)$  are  $\sigma(A^{(y)})$ -measurable, we obtain

$$\begin{aligned} & \mathbb{E}_\Omega \left[ \mathbb{1}_{x \in \mathcal{W}} \mathbb{1}_{y \in \mathcal{V} \setminus \{x\}} \sum_{t \in \mathcal{S}_1(y)} \langle \mathbf{1}_t, \mathbf{u}(x) \rangle^2 \right] \\ & \leq \mathbb{E} \left[ \mathbb{1}_{x \in \mathcal{W}^{(y)}} \sum_{t \in [N]} \langle \mathbf{1}_t, \tilde{\mathbf{u}}(x) \rangle^2 \mathbb{E} \left[ \mathbb{1}_{\alpha^*(\mu)d \leq |\mathcal{S}_1(y)| \leq 10 \log N} \mathbb{1}_{t \in \mathcal{S}_1(y)} |A^{(y)} \right] \right] \\ & \leq \mathbb{P}(x \in \mathcal{W}^{(y)}) \max_{t \in [N]} \mathbb{P}(t \in \mathcal{S}_1(y), \alpha^*(\mu)d \leq |\mathcal{S}_1(y)| \leq 10 \log N) \\ & \leq \mathbb{P}(x \in \mathcal{W}^{(y)}) \sum_{\alpha^*(\mu)d \leq k \leq 10 \log N} \mathbb{P}(|\mathcal{S}_1(y)| = k) \frac{k}{N} \leq \mathbb{P}(x \in \mathcal{V}) \mathbb{P}(y \in \mathcal{V}) \frac{10 \log N}{N}. \end{aligned} \tag{5.8}$$

Here, in the first step, in addition, we spelled out the condition  $y \in \mathcal{V}$  as  $|S_1(y)| \geq \alpha^*(\mu)d$ , used that  $|S_1(y)| \leq 10 \log N$  on  $\Omega$  by Proposition 3.1 (i) and then dropped the indicator function  $\mathbb{1}_\Omega$ . The second step follows from the independence of the event  $\{\alpha^*(\mu)d \leq |S_1(y)| \leq 10 \log N, t \in S_1(y)\}$  from  $\sigma(A^{(y)})$  and  $\sum_{t \in [N]} \langle \mathbf{1}_t, \tilde{\mathbf{u}}(x) \rangle^2 = \|\tilde{\mathbf{u}}(x)\|^2 = 1$ . (For these two steps, see also (2.9).) For the third step, we conditioned on  $|S_1(y)|$  and used that if  $|S_1(x)| = k$  then  $t$  lies in a uniformly distributed subset of  $[N] \setminus \{y\}$  with  $k$  elements. For the last step, we used that  $\mathcal{W}^{(y)} \subset \mathcal{W} \subset \mathcal{V}$ .

Finally, applying (5.8) to (5.7) and using the estimate  $\mathbb{P}(x \in \mathcal{V}) \leq N^{\mu-1}$  (by the definitions (2.1) and (1.6)) concludes the proof of Proposition 5.3.  $\square$

5.3. *Proof of Proposition 5.2 (ii), (iii).* In this section, we conclude Proposition 5.2 (ii) and (iii) from the following result, which is also proved in this section.

**Definition 5.5.** Order the elements of  $\mathcal{W}$  in some arbitrary fashion, and denote by  $(\mathbf{u}^\perp(x))_{x \in \mathcal{W}}$  the Gram-Schmidt orthonormalization of  $(\mathbf{u}(x))_{x \in \mathcal{W}}$ .

**Proposition 5.6.** Let  $\mu \in [0, 1/3)$ . Then the following holds with high probability. For any  $x \in \mathcal{W}$ , we have

$$\|(H - \lambda(x))\mathbf{u}^\perp(x)\| \lesssim N^{\mu-1/2+o(1)}. \tag{5.9}$$

More generally, denoting  $\mathcal{D} = \sum_{x \in \mathcal{W}} \lambda(x)\mathbf{u}^\perp(x)(\mathbf{u}^\perp(x))^*$  we have

$$\|(H - \mathcal{D})\mathbf{u}\| \lesssim N^{3\mu/2-1/2+o(1)}\|\mathbf{u}\| \tag{5.10}$$

for all  $\mathbf{u} \in \text{span}\{\mathbf{u}(x) : x \in \mathcal{W}\}$ .

*Proof of Proposition 5.2 (ii).* By Definition 5.1,  $\mathbf{w}_1$  is an eigenvector of  $\Pi H \Pi$  with eigenvalue  $\lambda_1(H)$ . Let  $\zeta < 1/2 - 3\mu/2$ . From (5.10), we conclude that, for each  $x \in \mathcal{W}$ , there is  $\varepsilon_x \in [-N^{-\zeta}, N^{-\zeta}]$  such that  $\{\lambda(x) + \varepsilon_x : x \in \mathcal{W}\} \subset \text{spec}(\Pi H \Pi) \setminus \{0\}$  counted with multiplicity. By Proposition 3.4 (ii),  $\lambda_1(H) = \sqrt{d}(1 + o(1))$ . Hence, by (3.1) and (3.3),

$$\lambda_1(H) \gg \Lambda(\alpha_x) + o(1) = \lambda(x) \tag{5.11}$$

for any  $x \in \mathcal{W}$ . Therefore, we have found  $1 + |\mathcal{W}|$  non-zero eigenvalues of  $\Pi H \Pi$  (counted with multiplicity). Since the dimension of  $\text{ran } \Pi$  is at most  $1 + |\mathcal{W}|$ , this completes the proof of Proposition 5.2 (ii).  $\square$

*Proof of Proposition 5.2 (iii).* In order to estimate  $\|\overline{\Pi} H \Pi\|$ , let  $\mathbf{v} \in \text{ran } \Pi$ . We decompose  $\mathbf{v} = \alpha \mathbf{w}_1 + \mathbf{u}$  from some  $\alpha \in \mathbb{R}$  and  $\mathbf{u} \in \text{span}\{\mathbf{u}(x) : x \in \mathcal{W}\}$ . Let  $\zeta < 1/2 - 3\mu/2$ . From (5.10), we obtain that

$$\overline{\Pi} H \mathbf{v} = \overline{\Pi} \alpha \lambda_1(H) \mathbf{w}_1 + \overline{\Pi} \mathcal{D} \mathbf{u} + o(N^{-\zeta} \|\mathbf{u}\|) = o(N^{-\zeta} \|\mathbf{u}\|), \tag{5.12}$$

where the last steps follows from  $\mathbf{w}_1 \in \text{ran } \Pi$  and  $\text{ran } \mathcal{D} \subset \text{ran } \Pi$  due to the definitions of  $\mathcal{D}$  and  $\Pi$ . It remains to show that  $\|\mathbf{u}\| \lesssim \|\mathbf{v}\|$ . From (5.10), we also conclude

$$\lambda_1(H) \langle \mathbf{w}_1, \mathbf{u} \rangle = \langle \mathbf{w}_1, H \mathbf{u} \rangle = \langle \mathbf{w}_1, \mathcal{D} \mathbf{u} \rangle + o(N^{-\zeta} \|\mathbf{u}\|). \tag{5.13}$$

Since  $\|\mathcal{D}\| \leq \max_{x \in \mathcal{W}} \lambda(x) \ll \lambda_1(H)$  by (5.11), we obtain from (5.13) that  $|\langle \mathbf{w}_1, \mathbf{u} \rangle| \ll \|\mathbf{u}\|$  and, thus,  $2|\alpha| |\langle \mathbf{w}_1, \mathbf{u} \rangle| \leq \alpha^2 + |\langle \mathbf{w}_1, \mathbf{u} \rangle|^2 = \alpha^2 + o(\|\mathbf{u}\|^2)$ . Therefore,

$$\|\mathbf{v}\|^2 = \alpha^2 + \|\mathbf{u}\|^2 + 2\alpha \langle \mathbf{w}_1, \mathbf{u} \rangle \geq \alpha^2 + \|\mathbf{u}\|^2 - 2|\alpha| |\langle \mathbf{w}_1, \mathbf{u} \rangle| \geq \alpha^2/2 + (1 - o(1))\|\mathbf{u}\|^2. \tag{5.14}$$

Hence,  $\|\mathbf{u}\| \leq (1 - o(1))\|\mathbf{v}\|$  which completes the proof of Proposition 5.2 (iii) due to (5.12).  $\square$

The main ingredient in the proof of Proposition 5.6 is the following result. It uses the following orthogonal projection.

**Definition 5.7** ( $\Pi_{\mathcal{X}}$ ). For any  $\mathcal{X} \subset \mathcal{W}$ , we denote by  $\Pi_{\mathcal{X}}$  the orthogonal projection onto  $\text{span}\{\mathbf{u}(x) : x \in \mathcal{X}\}$  and we define  $\overline{\Pi}_{\mathcal{X}} := I - \Pi_{\mathcal{X}}$ .

**Proposition 5.8.** *If  $\mu \in [0, 1/3)$  then*

$$\mathbb{E}_{\Omega} \left[ \max_{\mathcal{X} \subset \mathcal{W}} \|\overline{\Pi}_{\mathcal{X}} H \Pi_{\mathcal{X}}\|^2 \right] \lesssim d^{-1} (\log N)^2 N^{2\mu-1}.$$

Before proving Proposition 5.8 we deduce Proposition 5.6 from it.

*Proof of Proposition 5.6.* Recall from Definition 5.5 that the elements of  $\mathcal{W}$  are ordered in an arbitrary fashion. For  $x \in \mathcal{W}$ , let  $\Pi_{<x}$  be the orthogonal projection onto  $\text{span}\{\mathbf{u}(y) : y \in \mathcal{W}, y < x\}$  and  $\overline{\Pi}_{<x} := 1 - \Pi_{<x}$ . Then for any  $x \in \mathcal{W}$  we have

$$\mathbf{u}^{\perp}(x) = \frac{\mathbf{u}(x) - \Pi_{<x}\mathbf{u}(x)}{\|\mathbf{u}(x) - \Pi_{<x}\mathbf{u}(x)\|} = \frac{\overline{\Pi}_{<x}\mathbf{u}(x)}{\|\overline{\Pi}_{<x}\mathbf{u}(x)\|}. \tag{5.15}$$

In order to estimate  $(H - \lambda(x))\overline{\Pi}_{<x}\mathbf{u}(x)$ , we conclude from the definitions of  $\Pi_{<x}$  and  $\overline{\Pi}_{<x}$  that

$$H\overline{\Pi}_{<x} = \overline{\Pi}_{<x}H\overline{\Pi}_{<x} + \Pi_{<x}H\overline{\Pi}_{<x} = \overline{\Pi}_{<x}H - \overline{\Pi}_{<x}H\Pi_{<x} + \Pi_{<x}H\overline{\Pi}_{<x},$$

which then yields

$$(H - \lambda(x))\overline{\Pi}_{<x}\mathbf{u}(x) = \overline{\Pi}_{<x}(H - \lambda(x))\mathbf{u}(x) + \Pi_{<x}H\overline{\Pi}_{<x}\mathbf{u}(x) - \overline{\Pi}_{<x}H\Pi_{<x}\mathbf{u}(x).$$

Hence,

$$\|(H - \lambda(x))\overline{\Pi}_{<x}\mathbf{u}(x)\| \leq \|(H - \lambda(x))\mathbf{u}(x)\| + 2\|\overline{\Pi}_{<x}H\Pi_{<x}\|.$$

From (5.3) combined with a union bound over  $x$ , Proposition 5.8, and Chebyshev’s inequality, we deduce that with high probability

$$\max_{x \in \mathcal{W}} \|(H - \lambda(x))\overline{\Pi}_{<x}\mathbf{u}(x)\| \leq N^{\mu-1/2+o(1)}.$$

Moreover, since  $\langle \mathbf{1}_x, \mathbf{u}(y) \rangle = 0$  for  $x, y \in \mathcal{W}$  satisfying  $y < x$ , we find that with high probability, for all  $x \in \mathcal{W}$ ,

$$\|\overline{\Pi}_{<x}\mathbf{u}(x)\|^2 = \|\mathbf{u}(x) - \Pi_{<x}\mathbf{u}(x)\|^2 \geq \langle \mathbf{1}_x, \mathbf{u}(x) - \Pi_{<x}\mathbf{u}(x) \rangle^2 = \langle \mathbf{1}_x, \mathbf{u}(x) \rangle^2 \gtrsim 1,$$

where the last step follows from Corollary 3.7 and the fact that  $\mathbb{P}(\Omega) = 1 - o(1)$ . Plugging these two estimates into (5.15) yields (5.9).

For the proof of (5.10), we write  $\mathbf{u} = \sum_{x \in \mathcal{W}} a_x \mathbf{u}^{\perp}(x)$  and apply (5.9) and Cauchy-Schwarz,

$$\|(H - \mathcal{D})\mathbf{u}\| = \left\| \sum_{x \in \mathcal{W}} a_x (H - \lambda(x))\mathbf{u}^{\perp}(x) \right\| \leq N^{\mu-1/2+o(1)} |\mathcal{W}|^{1/2} \left( \sum_{x \in \mathcal{W}} a_x^2 \right)^{1/2}.$$

Hence, (5.10) follows from  $\mathcal{W} \subset \mathcal{V}$  and Proposition 3.2 (i). □

*Proof of Proposition 5.8.* Let  $\mathbf{v} \in \text{ran } \Pi_{\mathcal{X}}$ . We write  $\mathbf{v} = \sum_{x \in \mathcal{X}} a_x \mathbf{u}(x)$  with  $a_x \in \mathbb{R}$ , use (5.2) and  $\overline{\Pi}_{\mathcal{X}} \mathbf{u}(x) = 0$  for each  $x \in \mathcal{X}$  to obtain

$$\begin{aligned} \|\overline{\Pi}_{\mathcal{X}} H \mathbf{v}\|^2 &= \left\| \sum_{x \in \mathcal{X}} a_x \overline{\Pi}_{\mathcal{X}} \left( \lambda(x) \mathbf{u}(x) + \sum_{y \in \mathcal{V} \setminus \{x\}} \varepsilon_y(x) \mathbf{1}_y \right) \right\|^2 \\ &= \left\| \overline{\Pi}_{\mathcal{X}} \sum_{x \in \mathcal{X}, y \in \mathcal{V} \setminus \{x\}} \varepsilon_y(x) a_x \mathbf{1}_y \right\|^2 \\ &\leq \sum_{y \in \mathcal{V}} \left( \sum_{x \in \mathcal{X} \setminus \{y\}} \varepsilon_y(x) a_x \right)^2 \leq \left( \sum_{x \in \mathcal{X}} a_x^2 \right) \left( \sum_{x \in \mathcal{W}, y \in \mathcal{V} \setminus \{x\}} \varepsilon_y(x)^2 \right). \end{aligned}$$

Moreover, since  $\langle \mathbf{1}_y, \mathbf{u}(x) \rangle = 0$  for any  $y \in \mathcal{V} \setminus \{x\}$ , on  $\Omega$  we have

$$\|\mathbf{v}\|^2 \geq \sum_{x \in \mathcal{X}} |\langle \mathbf{1}_x, \mathbf{v} \rangle|^2 = \sum_{x \in \mathcal{X}} a_x^2 |\langle \mathbf{1}_x, \mathbf{u}(x) \rangle|^2 \geq c \sum_{x \in \mathcal{X}} a_x^2$$

for some constant  $c > 0$  by Corollary 3.7. Therefore, on  $\Omega$  we have

$$\|\overline{\Pi}_{\mathcal{X}} H \mathbf{v}\|^2 \leq \frac{1}{c} \left( \sum_{x \in \mathcal{W}, y \in \mathcal{V} \setminus \{x\}} \varepsilon_y(x)^2 \right) \|\mathbf{v}\|^2$$

and, in particular,  $\|\overline{\Pi}_{\mathcal{X}} H \Pi_{\mathcal{X}}\|^2 \leq \frac{1}{c} \left( \sum_{x \in \mathcal{W}, y \in \mathcal{V} \setminus \{x\}} \varepsilon_y(x)^2 \right)$  for any  $\mathcal{X} \subset \mathcal{W}$ . By Proposition 5.3, we therefore conclude

$$\mathbb{E}_{\Omega} \left[ \sum_{x \in \mathcal{W}, y \in \mathcal{V} \setminus \{x\}} \varepsilon_y(x)^2 \right] = \sum_{x, y \in [N]} \mathbb{E}_{\Omega} [\mathbb{1}_{x \in \mathcal{W}} \mathbb{1}_{y \in \mathcal{V} \setminus \{x\}} \varepsilon_y(x)^2] \lesssim d^{-1} (\log N)^2 N^{2\mu-1},$$

as claimed. □

*5.4. Proof of Proposition 5.2 (iv).* For the proof of Proposition 5.2 (iv), we shall need several notions from the works [12, 13]. As in [12, eq. (1.8)], we set

$$r_{\star} = \lfloor c \sqrt{\log N} \rfloor \tag{5.16}$$

for the constant  $c > 0$  from [12]. Following [12, eq. (1.9)], we define

$$\xi := \frac{\sqrt{\log N}}{d} \log d, \quad \xi_u := \frac{\sqrt{\log N}}{d} \frac{1}{u} \tag{5.17}$$

for  $u > 0$ . For any  $\tau \in [1 + \xi^{1/2}, 2]$ , we denote by  $\mathbb{G}_{\tau}$  the *pruned graph* introduced in [12, Proposition 3.1]. We denote the balls and spheres in  $\mathbb{G}_{\tau}$  around a vertex  $x \in [N]$  by  $B_i^{\tau}(x)$  and  $S_i^{\tau}(x)$ , respectively. The pruned graph  $\mathbb{G}_{\tau}$  is a subgraph of  $\mathbb{G}$ , which possesses a number of useful properties listed in [12, Proposition 3.1]. In particular, the balls  $B_{2r_{\star}}^{\tau}(x)$  and  $B_{2r_{\star}}^{\tau}(y)$  in  $\mathbb{G}_{\tau}$  are disjoint if  $x, y \in [N]$  satisfy  $x \neq y$  and  $\min\{\alpha_x, \alpha_y\} \geq \tau$ .

Recalling the definition of  $u_i(\alpha)$  from (1.10), for any  $x \in [N]$  with  $\alpha_x \geq 2 + \xi^{1/4}$  and  $\sigma = \pm$ , as in [12, eq. (3.5)], we define

$$\mathbf{v}_\sigma^\tau(x) := \sum_{i=0}^{r_\star} \sigma^i u_i(\alpha_x) \frac{\mathbf{1}_{S_i^\tau(x)}}{\|\mathbf{1}_{S_i^\tau(x)}\|}, \tag{5.18}$$

where, for the last coefficient  $u_{r_\star}(\alpha_x)$  we make the special choice  $u_{r_\star}(\alpha_x) := u_{r_\star-1}(\alpha_x)/\sqrt{\alpha_x}$ , and  $u_0(\alpha_x) > 0$  is chosen such that  $\mathbf{v}_\sigma^\tau(x)$  is normalized, i.e.  $\sum_{i=0}^{r_\star} u_i^2(\alpha_x) = 1$ .

*Remark 5.9.* The family  $(\mathbf{v}_\sigma^\tau(x) : x \in [N], \alpha_x \geq 2 + \xi^{1/4}, \sigma = \pm)$  is orthonormal. See [12, Remark 3.3]

As in [12, Definition 3.6], we denote the adjacency matrix of  $\mathbb{G}_\tau$  by  $A^\tau$  and define the matrix

$$H^\tau := (A^\tau - \chi^\tau(\mathbb{E}A)\chi^\tau)/\sqrt{d}, \tag{5.19}$$

where  $\chi^\tau$  is the orthogonal projection onto  $\text{span}\{\mathbf{1}_y : y \notin \bigcup_{x:\alpha_x \geq \tau} B_{2r_\star}^\tau(x)\}$ . Moreover, we recall [12, Definition 3.10].

**Definition 5.10** ( $\Pi^\tau, \widehat{H}^\tau$ ). Define the orthogonal projections (see Remark 5.9)

$$\Pi^\tau := \sum_{x:\alpha_x \geq 2+\xi^{1/4}} \sum_{\sigma=\pm} \mathbf{v}_\sigma^\tau(x)\mathbf{v}_\sigma^\tau(x)^*, \quad \overline{\Pi}^\tau := 1 - \Pi^\tau$$

and the associated block matrix (recall (5.19))

$$\widehat{H}^\tau := \sum_{x:\alpha_x \geq 2+\xi^{1/4}} \sum_{\sigma=\pm} \sigma \Lambda(\alpha_x) \mathbf{v}_\sigma^\tau(x)\mathbf{v}_\sigma^\tau(x)^* + \overline{\Pi}^\tau H^\tau \overline{\Pi}^\tau. \tag{5.20}$$

We note that, for any  $\tau \in [1 + \xi^{1/2}, 2]$ , by (1.5) we have

$$\xi = o(1), \quad \xi_{\tau-1} = o(1). \tag{5.21}$$

For any  $\tau \in [1 + \xi^{1/2}, 2]$ , the definition of  $\widehat{H}^\tau$  in (5.20) and [12, Proposition 3.12] yield that, with high probability,

$$\|\widehat{H}^\tau\| \leq \max \left\{ \max_{x:\alpha_x \geq 2+\xi^{1/4}} \Lambda(\alpha_x), 2\tau + C\xi \right\} \lesssim \sqrt{\frac{\log N}{d}} \ll \sqrt{d}, \tag{5.22}$$

where we used (3.1), (5.21) and (1.5) in the last two steps.

Owing to (5.21), [12, Lemmas 3.8, 3.11],<sup>9</sup> with high probability, we have

$$\|H - \mathbb{E}H - \widehat{H}^\tau\| = o(1). \tag{5.23}$$

From (5.23) and (5.22), we conclude that, with high probability,

$$\|H - \mathbb{E}H\| \ll \sqrt{d}. \tag{5.24}$$

After these preparations, we can start the proof of Proposition 5.2 (iv). We begin with the following definition.

<sup>9</sup> We stress that the definition of  $H$  in [12] differs from that in the current paper by  $\mathbb{E}H$ ; see [12, Definition 3.6].

**Definition 5.11** ( $Q_r$ ). For  $r \in \mathbb{N}$ , denote by  $Q_r$  the orthogonal projection defined by restriction to the set  $(\bigcup_{y \in \mathcal{V} \setminus \mathcal{W}} B_r(y))^c$ .

**Definition 5.12** ( $Q, \Pi_Q$ ). Let  $r_*$  be as in (5.16). Set  $Q := Q_{2r_*-1}$  and define  $\Pi_Q$  as the orthogonal projection onto  $\text{span}(\{Q\mathbf{u}^\perp(x) : x \in \mathcal{W}\} \cup \{Q\mathbf{w}_1\})$ , and write  $\overline{\Pi}_Q := 1 - \Pi_Q$ .

Then

$$\begin{aligned} \lambda_1(\overline{\Pi}H\overline{\Pi}) &\leq \lambda_1(\overline{\Pi}(H - \mathbb{E}H)\overline{\Pi}) + \|\overline{\Pi}(\mathbb{E}H)\overline{\Pi}\| \\ &\leq \lambda_1(\overline{\Pi}_Q \widehat{H}^\tau \overline{\Pi}_Q) + 2\|H - \mathbb{E}H\| \|\Pi - \Pi_Q\| + \|H - \mathbb{E}H - \widehat{H}^\tau\| \\ &\quad + \|\overline{\Pi}(\mathbb{E}H)\overline{\Pi}\| \\ &\leq \lambda_1(\overline{\Pi}_Q \widehat{H}^\tau \overline{\Pi}_Q) + 2\|H - \mathbb{E}H\| \|\Pi - \Pi_Q\| + o(1), \end{aligned} \quad (5.25)$$

whose last step follows from (5.23) and  $\|\overline{\Pi}(\mathbb{E}H)\overline{\Pi}\| = o(1)$ . The latter bound is a consequence of

$$\begin{aligned} \|\overline{\Pi}(\mathbb{E}H)\overline{\Pi}\| &= \sqrt{d} \|\overline{\Pi}(\mathbf{e}\mathbf{e}^* - 1/N)\overline{\Pi}\| = \sqrt{d} \|\overline{\Pi}(\mathbf{e} - \mathbf{w}_1)(\mathbf{e}^* - \mathbf{w}_1^*)\overline{\Pi}\| \\ &\quad + o(d^{-1/2}) \lesssim d^{-1/2}, \end{aligned} \quad (5.26)$$

where we introduced  $\mathbf{e} := N^{-1/2} \mathbf{1}_{[N]}$ , used  $\Pi \mathbf{w}_1 = \mathbf{w}_1$  in the second step, and in the last step we used Proposition 3.4 (iv) to estimate  $\|\mathbf{e} - \mathbf{w}_1\| \leq \|\mathbf{w}_1 - \mathbf{q}\| + \|\mathbf{q} - \mathbf{e}\| \lesssim d^{-1/2}$ .

Fix  $\mu \in [0, 1/4)$ . We then claim that with high probability

$$\|\Pi - \Pi_Q\| \lesssim d^{-1} \quad (5.27)$$

and

$$\lambda_1(\overline{\Pi}_Q \widehat{H}^\tau \overline{\Pi}_Q) \leq \Lambda(\alpha^*) + \kappa/2. \quad (5.28)$$

Using (5.27) and (5.28), Proposition 5.2 (iv) follows immediately from (5.25) and (5.24). What remains to prove Proposition 5.2 (iv), therefore, is the proof of (5.27) and (5.28).

*5.5. Proof of (5.27).* Clearly,

$$\|\Pi - \Pi_Q\| = \|\Pi \overline{\Pi}_Q - \overline{\Pi} \Pi_Q\| \leq \|\Pi \overline{\Pi}_Q\| + \|\overline{\Pi} \Pi_Q\| = \|\overline{\Pi}_Q \Pi\| + \|\overline{\Pi} \Pi_Q\|, \quad (5.29)$$

where in the last step we used that  $(\Pi \overline{\Pi}_Q)^* = \overline{\Pi}_Q \Pi$ . In order to estimate the terms on the right-hand side, we continue with

$$\|\overline{\Pi}_Q \Pi\|^2 = \sup_{\substack{\mathbf{v} \in \text{ran } \Pi \\ \|\mathbf{v}\|=1}} \|\overline{\Pi}_Q \mathbf{v}\|^2 \leq \sup_{\substack{\mathbf{v} \in \text{ran } \Pi \\ \|\mathbf{v}\|=1}} \inf_{\substack{\mathbf{u} \in \text{ran } \Pi_Q \\ \|\mathbf{u}\|=1}} \|\mathbf{v} - \mathbf{u}\|^2. \quad (5.30)$$

We now apply the next lemma, whose proof is given in Sect. 5.7 below.

**Lemma 5.13.** Fix  $\mu \in [0, 1/3)$ . With high probability, for all  $x \in \mathcal{W}$ ,  $y \in \mathcal{V} \setminus \mathcal{W}$  and  $r \in \mathbb{N}$  satisfying  $r \ll \frac{\log N}{\log d}$ , we have

$$\|\mathbf{u}^\perp(x)|_{B_r(y)}\| \lesssim N^{-1/2+\mu+o(1)}.$$

Owing to Lemma 5.13 as well as parts (iii) and (i) of Proposition 3.2, we obtain

$$\|\mathbf{u}^\perp(x) - Q\mathbf{u}^\perp(x)\| = \left( \sum_{y \in \mathcal{V} \setminus \mathcal{W}} \|\mathbf{u}^\perp(x)|_{B_{2r_\star-1}(y)}\|^2 \right)^{1/2} \leq N^{-1/2+3\mu/2+o(1)}. \quad (5.31)$$

Let  $\mathbf{q}$  be as in Proposition 3.4 (iv) with  $r = 2r_\star - 2 \ll \frac{d}{\log \log N}$  by (1.5). Since  $\text{supp } \mathbf{q} \subset (\bigcup_{x \in \mathcal{V}} B_{r+1}(x))^c = (\bigcup_{x \in \mathcal{V}} B_{2r_\star-1}(x))^c$ , we conclude from Proposition 3.4 (iv) and the definition of  $Q$  that

$$\|\mathbf{w}_1 - Q\mathbf{w}_1\| = \|(1 - Q)(\mathbf{w}_1 - \mathbf{q})\| \leq \|\mathbf{w}_1 - \mathbf{q}\| \lesssim d^{-1}. \quad (5.32)$$

For  $\gamma_1, \gamma_x \in \mathbb{R}$  for  $x \in \mathcal{W}$  write

$$\mathbf{v} := \gamma_1 \mathbf{w}_1 + \sum_{x \in \mathcal{W}} \gamma_x \mathbf{u}^\perp(x) \in \text{ran } \Pi, \quad \mathbf{u} := \gamma_1 Q\mathbf{w}_1 + \sum_{x \in \mathcal{W}} \gamma_x Q\mathbf{u}^\perp(x) \in \text{ran } \Pi_Q. \quad (5.33)$$

Then

$$\begin{aligned} \|\mathbf{v} - \mathbf{u}\| &\lesssim d^{-1}|\gamma_1| + N^{-1/2+3\mu/2+o(1)} \sum_{x \in \mathcal{W}} |\gamma_x| \\ &\lesssim d^{-1}|\gamma_1| + N^{-1/2+2\mu+o(1)} \left( \sum_{x \in \mathcal{W}} |\gamma_x|^2 \right)^{1/2} \lesssim d^{-1}\|\mathbf{v}\|. \end{aligned}$$

Here, we used in (5.32) and (5.31) in the first step, Proposition 3.2 (i) in the second step and, in the fourth step,  $\mu < 1/4$  as well as  $\|\mathbf{v}\|^2 \asymp |\gamma_1|^2 + \sum_{x \in \mathcal{W}} |\gamma_x|^2$  (the inequality  $\gtrsim$  follows from (5.14) and the orthogonality of  $(\mathbf{u}^\perp(x))_{x \in \mathcal{W}}$ ; the inequality  $\lesssim$  is trivial). Hence, if  $\|\mathbf{v}\| = 1$  then  $\|\mathbf{u}\| = 1 + O(d^{-1})$  and, thus,  $\|\mathbf{v} - \frac{\mathbf{u}}{\|\mathbf{u}\|}\| \lesssim d^{-1}$ . Therefore,  $\|\overline{\Pi}_Q \Pi\| \lesssim d^{-1}$  by (5.30).

Finally, similarly to (5.30), we have

$$\|\overline{\Pi} \Pi_Q\|^2 \leq \sup_{\substack{\mathbf{u} \in \text{ran } \Pi_Q \\ \|\mathbf{u}\|=1}} \inf_{\substack{\mathbf{v} \in \text{ran } \Pi \\ \|\mathbf{v}\|=1}} \|\mathbf{v} - \mathbf{u}\|^2,$$

and the same argument as above, with the representation (5.33), implies that the right-hand side is  $O(d^{-1})$ . By (5.29), we therefore conclude (5.27).

5.6. Proof of (5.28). We begin by introducing another orthogonal projection.

**Definition 5.14** ( $\Pi_\mathbf{v}$ ). Let  $\Pi_\mathbf{v}$  be the orthogonal projection onto  $\text{span}\{\mathbf{v}_\pm^\tau(z) : z \in \mathcal{V} \setminus \mathcal{W}\}$ .

For any  $z \in \mathcal{V} \setminus \mathcal{W}$ , we have  $\text{supp } \mathbf{v}_\pm^\tau(z) \subset B_{r_\star}(z)$  and, thus, by definition of  $Q$ ,  $\text{supp } \mathbf{v}_\pm^\tau(z) \cap \text{supp } Q\mathbf{u}^\perp(x) = \emptyset$  for any  $x \in \mathcal{W}$  and  $\text{supp } \mathbf{v}_\pm^\tau(z) \cap \text{supp } Q\mathbf{w}_1 = \emptyset$ . Therefore,  $\mathbf{v}_\pm^\tau(z)$  is orthogonal to  $\text{ran } \Pi_Q$ , i.e.  $\Pi_Q \mathbf{v}_\pm^\tau(x) = 0$ . That implies  $\Pi_\mathbf{v} \Pi_Q = 0 = \Pi_Q \Pi_\mathbf{v}$  and, in particular,  $\Pi_\mathbf{v}$  and  $\Pi_Q$  commute. Since  $\Pi_\mathbf{v}$  and  $\widehat{H}^\tau$  commute and  $\Pi_\mathbf{v}$  and  $\Pi_Q$  commute, we obtain

$$\begin{aligned} \lambda_1(\overline{\Pi}_Q \widehat{H}^\tau \overline{\Pi}_Q) &= \lambda_1(\overline{\Pi}_Q \Pi_\mathbf{v} \widehat{H}^\tau \Pi_\mathbf{v} \overline{\Pi}_Q + \overline{\Pi}_Q \overline{\Pi}_\mathbf{v} \widehat{H}^\tau \overline{\Pi}_\mathbf{v} \overline{\Pi}_Q) \\ &= \max\{\lambda_1(\overline{\Pi}_Q \Pi_\mathbf{v} \widehat{H}^\tau \Pi_\mathbf{v} \overline{\Pi}_Q), \lambda_1(\overline{\Pi}_Q \overline{\Pi}_\mathbf{v} \widehat{H}^\tau \overline{\Pi}_\mathbf{v} \overline{\Pi}_Q)\}. \end{aligned} \quad (5.34)$$



By the definition of  $\widehat{H}^\tau$  in (5.20), we have

$$\lambda_1(\Pi_{\mathcal{V}} \widehat{H}^\tau \Pi_{\mathcal{V}}) = \max_{x \in \mathcal{V} \setminus \mathcal{W}} \Lambda(\alpha_x) \leq \Lambda(\alpha^*) + \kappa/2. \tag{5.35}$$

What remains, therefore, is to estimate  $\lambda_1(\overline{\Pi}_Q \overline{\Pi}_{\mathcal{V}} \widehat{H}^\tau \overline{\Pi}_{\mathcal{V}} \overline{\Pi}_Q)$ .

Suppose that  $\mathbf{w}$  is a normalized eigenvector of  $\overline{\Pi}_Q \overline{\Pi}_{\mathcal{V}} \widehat{H}^\tau \overline{\Pi}_{\mathcal{V}} \overline{\Pi}_Q$  such that the associated eigenvalue  $\lambda$  satisfies  $\lambda \geq \Lambda(\alpha^*) + \kappa/2$ . We now check that the next lemma, whose proof is given in Sect. 5.7 below, is applicable to  $\mathbf{w}$  and  $\lambda$  for any  $x \in \mathcal{V} \setminus \mathcal{W}$ .

**Lemma 5.15.** *Let  $r_\star \in \mathbb{N}$  be as in (5.16). In particular,  $r_\star \asymp \sqrt{\log N}$ . Suppose  $\tau \in [1 + \xi^{1/2}, 2]$ .*

*There is a constant  $C > 0$  such that the following holds with high probability,<sup>10</sup>. Let  $x \in [N]$ ,  $\lambda > 2\tau + C\xi$  and  $\mathbf{w}$  satisfy*

$$(\widehat{H}^\tau \mathbf{w})|_{B_{2r_\star-1}^\tau(x)} = \lambda \mathbf{w}|_{B_{2r_\star-1}^\tau(x)}. \tag{5.36}$$

*If  $\alpha_x \geq 2 + \xi^{1/4}$  and  $\mathbf{v}_-^\tau(x) \perp \mathbf{w} \perp \mathbf{v}_+^\tau(x)$  or  $2 + \xi^{1/4} > \alpha_x \geq \tau$  then*

$$\frac{|\langle \mathbf{1}_x, \mathbf{w} \rangle|}{\|\mathbf{w}|_{B_{2r_\star}^\tau(x)}\|} \lesssim \frac{\lambda^2}{(\lambda - 2\tau - C\xi)^2} \left( \frac{2\tau + C\xi}{\lambda} \right)^{r_\star}. \tag{5.37}$$

*An analogous result holds if  $\lambda < -2\tau - C\xi$ .*

Choosing  $\tau = 1 + \xi^{1/2}$  and taking  $x \in \mathcal{V} \setminus \mathcal{W}$ , we now verify the conditions of Lemma 5.15 for  $\mathbf{w}$  and  $\lambda$ . Owing to (5.21), there is a constant  $c \equiv c_\kappa > 0$  such that  $\frac{2\tau + C\xi}{\lambda} \leq 1 - c$  as  $\lambda \geq \Lambda(\alpha^*) + \kappa/2 \geq 2 + 2c$  due to the definition of  $\alpha^*(\mu)$  in (1.6). In particular,  $\lambda > 2\tau + C\xi$ . From  $\overline{\Pi}_Q \overline{\Pi}_{\mathcal{V}} \widehat{H}^\tau \overline{\Pi}_{\mathcal{V}} \overline{\Pi}_Q \mathbf{w} = \lambda \mathbf{w}$  we conclude that  $\overline{\Pi}_Q \mathbf{w} = \mathbf{w}$  and  $\overline{\Pi}_{\mathcal{V}} \mathbf{w} = \mathbf{w}$ . Thus, as  $\Pi_{\mathcal{V}}$  and  $\widehat{H}^\tau$  commute, we get  $\overline{\Pi}_Q \widehat{H}^\tau \mathbf{w} = \lambda \mathbf{w}$ . Restricting both sides in the last identity to  $B_{2r_\star-1}^\tau(x)$  yields  $(\widehat{H}^\tau \mathbf{w})|_{B_{2r_\star-1}^\tau(x)} = \lambda \mathbf{w}|_{B_{2r_\star-1}^\tau(x)}$  as  $Q$  is the restriction to  $(\bigcup_{y \in \mathcal{V} \setminus \mathcal{W}} B_{2r_\star-1}(y))^c$  and  $B_{2r_\star-1}^\tau(x) \subset B_{2r_\star-1}(x)$  by [12, Proposition 3.1 (iv)]. This proves (5.36). Note that  $\alpha_x \geq 2 + \xi^{1/4}$ . From  $\overline{\Pi}_{\mathcal{V}} \mathbf{w} = \mathbf{w}$ , we conclude  $\mathbf{w} \perp \mathbf{v}_\pm^\tau(x)$ . Since  $(\mathbf{v}_\sigma^\tau(y) : \alpha_y \geq 2 + \xi^{1/4}, \sigma = \pm)$  is an orthonormal family by Remark 5.9, the definition of  $\Pi_{\mathcal{V}}$  implies  $\Pi_{\mathcal{V}} \mathbf{v}_\pm^\tau(y) = 0$  for all  $y \in \mathcal{V}$ . Therefore, as moreover  $\Pi_Q \mathbf{v}_\pm^\tau(x) = 0$ , we have  $\overline{\Pi}_Q \overline{\Pi}_{\mathcal{V}} \widehat{H}^\tau \overline{\Pi}_{\mathcal{V}} \overline{\Pi}_Q \mathbf{v}_\pm^\tau(x) = -\Lambda(\alpha_x) \mathbf{v}_\pm^\tau(x)$ . Hence,  $\mathbf{w} \perp \mathbf{v}_\pm^\tau(x)$ . Therefore, we have verified all assumptions of Lemma 5.15 with  $\tau = 1 + \xi^{1/2}$  for  $\mathbf{w}$ ,  $\lambda$ , and any  $x \in \mathcal{V} \setminus \mathcal{W}$ .

Since  $\frac{2\tau + C\xi}{\lambda} \leq 1 - c$  as shown above, for the right-hand side of (5.37), we get

$$\begin{aligned} \frac{\lambda^2}{(\lambda - 2\tau - C\xi)^2} \left( \frac{2\tau + C\xi}{\lambda} \right)^{r_\star} &= \left( 1 - \frac{2\tau + C\xi}{\lambda} \right)^{-2} \left( \frac{2\tau + C\xi}{\lambda} \right)^{r_\star} \\ &\leq c^{-2} (1 - c)^{r_\star} \ll d^{-1/2} \end{aligned}$$

as  $r_\star \asymp \sqrt{\log N}$ . Therefore, Lemma 5.15, the disjointness of the balls  $(B_{2r_\star}^\tau(x))_{x \in \mathcal{V} \setminus \mathcal{W}}$  (see the paragraph after (5.17) as well as [12, Proposition 3.1 (i)]) and  $\|\mathbf{w}\|^2 = 1$  imply

$$\|\mathbf{w}|_{\mathcal{V} \setminus \mathcal{W}}\|^2 = \sum_{x \in \mathcal{V} \setminus \mathcal{W}} \langle \mathbf{1}_x, \mathbf{w} \rangle^2 \ll d^{-1} \sum_{x \in \mathcal{V} \setminus \mathcal{W}} \|\mathbf{w}|_{B_{2r_\star}^\tau(x)}\|^2 \leq d^{-1}. \tag{5.38}$$

<sup>10</sup> We note that the statement actually holds with *very high probability* meaning that for each  $\nu > 0$ , there is a constant  $C \equiv C_\nu > 0$  such that (5.37) holds with probability at least  $1 - N^{-\nu}$  for all sufficiently large  $N$ .

We recall from Definition 5.11 that  $Q_0$  denotes the orthogonal projection defined by restriction to the set  $(\mathcal{V} \setminus \mathcal{W})^c$ . Since  $\mathbf{w} = \mathbf{w}|_{\mathcal{V} \setminus \mathcal{W}} + Q_0 \mathbf{w}$ , we obtain from (5.38),  $Q_0 \widehat{H}^\tau Q_0 = (\widehat{H}^\tau)^{(\mathcal{V} \setminus \mathcal{W})}$ , and  $\|\widehat{H}^\tau\| \ll \sqrt{d}$  by (5.22) that

$$\begin{aligned} \lambda &= \langle \mathbf{w}, \widehat{H}^\tau \mathbf{w} \rangle = \langle \mathbf{w}, (\widehat{H}^\tau)^{(\mathcal{V} \setminus \mathcal{W})} \mathbf{w} \rangle + o(1) = \langle \mathbf{w}, \overline{\Pi}_Q (\widehat{H}^\tau)^{(\mathcal{V} \setminus \mathcal{W})} \overline{\Pi}_Q \mathbf{w} \rangle + o(1) \\ &\leq \lambda_1(\overline{\Pi}_Q (\widehat{H}^\tau)^{(\mathcal{V} \setminus \mathcal{W})} \overline{\Pi}_Q) + o(1) \leq \lambda_1(\overline{\Pi}(\widehat{H}^\tau + (\mathbb{E}H))^{(\mathcal{V} \setminus \mathcal{W})} \overline{\Pi}) + o(1) \\ &\leq \lambda_1(\overline{\Pi}H^{(\mathcal{V} \setminus \mathcal{W})} \overline{\Pi}) + o(1). \end{aligned} \tag{5.39}$$

Here, the third step is a consequence of  $\overline{\Pi}_Q \mathbf{w} = \mathbf{w}$ . In the fifth step, we used (5.26) and (5.27) and the sixth step follows from (5.23) and  $\|M^{(\mathcal{V} \setminus \mathcal{W})}\| \leq \|M\|$  for any matrix  $M$ .

We now apply the next result, whose proof is given in Sect. 5.7 below.

**Lemma 5.16.** *Fix  $\mu \in [0, 1/3)$ . With high probability,  $\lambda_1(\overline{\Pi}H^{(\mathcal{V} \setminus \mathcal{W})} \overline{\Pi}) \leq \Lambda(\alpha^*) + o(1)$ .*

From Lemma 5.16 and (5.39), we deduce that  $\lambda \leq \Lambda(\alpha^*) + o(1)$ , in contradiction with the assumption  $\lambda \geq \Lambda(\alpha^*) + \kappa/2$ . We conclude that  $\lambda_1(\overline{\Pi}_Q \overline{\Pi}_v \widehat{H}^\tau \overline{\Pi}_v \overline{\Pi}_Q) \leq \Lambda(\alpha^*) + \kappa/2$ . Owing to (5.34) and (5.35), this proves (5.28).

**5.7. Proofs of auxiliary results.** In this final subsection we prove Lemmas 5.13, 5.15, and 5.16.

*Proof of Lemma 5.13.* For fixed  $y \in [N]$ , on the event  $\Omega \cap \{y \in \mathcal{V} \setminus \mathcal{W}\}$ , if  $x \in \mathcal{W}^{(y)}$ , let  $\mathbf{u}^{(y)}(x)$  be the eigenvector of  $H^{((\mathcal{V}^{(y)} \cup \{y\}) \setminus \{x\})}$  with eigenvalue  $\lambda_2(H^{((\mathcal{V}^{(y)} \cup \{y\}) \setminus \{x\})})$  and satisfying  $\langle \mathbf{1}_x, \mathbf{u}^{(y)}(x) \rangle > 0$ . See Corollaries 3.6 and 3.7 for the existence and uniqueness of  $\mathbf{u}^{(y)}(x)$ .

In analogy to Definition 5.5, let  $(\mathbf{u}^{(y)})^\perp(x)_{x \in \mathcal{W}^{(y)}}$  be the Gram-Schmidt orthonormalization of  $(\mathbf{u}^{(y)}(x))_{x \in \mathcal{W}^{(y)}}$ . On  $\Omega \cap \{y \in \mathcal{V} \setminus \mathcal{W}\}$ , we have  $\mathcal{V}^{(y)} \cup \{y\} = \mathcal{V}$  and  $\mathcal{W}^{(y)} = \mathcal{W}$  by Remark 5.4. Therefore,  $\mathbf{u}^{(y)}(x) = \mathbf{u}(x)$  and, thus,  $(\mathbf{u}^{(y)})^\perp(x) = \mathbf{u}^\perp(x)$  for all  $x \in \mathcal{W}^{(y)} = \mathcal{W}$  on  $\Omega \cap \{y \in \mathcal{V} \setminus \mathcal{W}\}$ . Hence, for fixed  $x, y \in [N]$  with  $x \neq y$ , we estimate

$$\begin{aligned} &\mathbb{E}_\Omega \left[ \mathbb{1}_{x \in \mathcal{W}, y \in \mathcal{V} \setminus \mathcal{W}} \|\mathbf{u}^\perp(x)|_{B_r(y)}\|^2 \right] \\ &= \mathbb{E} \left[ \mathbb{1}_{x \in \mathcal{W}^{(y)}} \sum_{a \in [N] \setminus \{y\}} \langle \mathbf{1}_a, (\mathbf{u}^{(y)})^\perp(x) \rangle^2 \mathbb{1}_{a \in B_r(y)} \mathbb{1}_{y \in \mathcal{V} \setminus \mathcal{W}} \right] \\ &\leq \mathbb{E} \left[ \mathbb{1}_{x \in \mathcal{W}^{(y)}} \mathbb{1}_{\Omega_r^{(y)}} \sum_{a \in [N] \setminus \{y\}} \langle \mathbf{1}_a, (\mathbf{u}^{(y)})^\perp(x) \rangle^2 \sum_{b \in B_{r-1}^{(y)}(a)} \right] \\ &\mathbb{E} \left[ \mathbb{1}_{b \in S_1(y)} \mathbb{1}_{\alpha^*(\mu)d \leq |S_1(y)| \leq 10 \log N} \left| A^{(y)} \right| \right] \\ &\leq \mathbb{E} \left[ \mathbb{1}_{x \in \mathcal{W}^{(y)}} \mathbb{1}_{\Omega_r^{(y)}} \sum_{a \in [N] \setminus \{y\}} \langle \mathbf{1}_a, (\mathbf{u}^{(y)})^\perp(x) \rangle^2 |B_{r-1}^{(y)}(a)| \right] \\ &\quad \times \max_{b \in [N] \setminus \{y\}} \mathbb{P}(b \in S_1(y), \alpha^*(\mu)d \leq |S_1(y)| \leq 10 \log N) \end{aligned}$$

$$\begin{aligned} &\lesssim (\log N)d^{r-1}\mathbb{P}(x \in \mathcal{W}^{(y)}) \max_{b \in [N] \setminus \{y\}} \mathbb{P}(b \in S_1(y), \alpha^*(\mu)d \leq |S_1(y)| \leq 10 \log N) \\ &\lesssim \frac{(\log N)^2 d^{r-1}}{N} \mathbb{P}(x \in \mathcal{V}) \mathbb{P}(y \in \mathcal{V}) \leq N^{-3+2\mu+o(1)}. \end{aligned}$$

Here, in the first step, we also used that  $\langle \mathbf{1}_y, (\mathbf{u}^{(y)})^\perp(x) \rangle = 0$  for all  $y \in \mathcal{V} \setminus \mathcal{W}$  and  $x \in \mathcal{W}^{(y)}$ . In the second step, we conditioned on  $A^{(y)}$ , employed the notations

$$\begin{aligned} B_i^{(y)}(a) &:= \text{ball of radius } i \text{ around } a \text{ in the graph } \mathbb{G}_{|[N] \setminus \{y\}} \\ \Omega_r^{(y)} &:= \{ |B_r^{(y)}(z)| \lesssim (\log N)d^{r-1} \text{ for all } z \in [N] \setminus \{y\} \} \end{aligned}$$

and used that  $\mathcal{W}^{(y)}$ ,  $(\mathbf{u}^{(y)})^\perp(x)$ , and  $B_{r-1}^{(y)}(a)$  are  $A^{(y)}$ -measurable, that  $\Omega \subset \Omega_r^{(y)}$  by Proposition 3.1, as  $r \ll \frac{\log N}{\log d}$ , and that  $a \in B_r(y)$  is equivalent to  $b \in S_1(y)$  for some  $b \in B_{r-1}^{(y)}(a)$ . The third step follows from the independence of  $b \in S_1(y)$  and  $|S_1(y)|$  from  $A^{(y)}$ . The normalization of  $(\mathbf{u}^{(y)})^\perp(x)$  and the definition of  $\Omega_r^{(y)}$  imply the fourth step. In the fifth step, we argued as in the last steps of (5.8) and, finally, we used  $r \ll \frac{\log N}{\log d}$  as well as the definitions of  $\mathcal{V}$  in (2.1) and of  $\alpha^*$  in (1.6).

Therefore, a union bound over  $x, y \in [N]$  with  $x \neq y$  and Chebyshev’s inequality complete the proof of Lemma 5.13.  $\square$

*Proof of Lemma 5.15.* In order to prove Lemma 5.15, we follow [12, proof of Proposition 3.14 (i)]<sup>11</sup>, whose assumptions are all satisfied apart from the eigenvalue-eigenvector relation  $\widehat{H}\mathbf{w} = \lambda\mathbf{w}$ .

We now explain the necessary, minor, modifications. We start with the case  $\alpha_x \geq 2 + \xi^{1/4}$  and  $\mathbf{w} \perp \mathbf{v}_\pm^\tau(x)$  which corresponds to the case  $x \in \mathcal{V}$  in [12, Proposition 3.14]. The eigenvalue-eigenvector relation is used in [12, proof of Proposition 3.14 (i)] only in [12, eq. (3.55)]. Using (5.36) instead of the eigenvalue-eigenvector relation and the notation of [12, proof of Proposition 3.14 (i)], we now verify the first two steps in [12, eq. (3.55)]. We write  $P_{2r_\star-1}$  for the orthogonal defined by restriction to the set  $B_{2r_\star-1}^\tau(x)$ . For any  $i < r_\star$ , since  $\text{supp } \mathbf{g}_i \subset B_{2r_\star-1}^\tau(x)$  by [12, eq. (3.52)], we have  $P_{2r_\star-1}\mathbf{g}_i = \mathbf{g}_i$ . Therefore, for any  $i < r_\star$ , we obtain

$$\begin{aligned} \lambda u_i &= \langle \mathbf{g}_i, \lambda P_{2r_\star-1}\mathbf{w} \rangle = \langle (\widehat{H}^\tau - \Lambda(\alpha_x)\mathbf{v}_+^\tau(x)\mathbf{v}_+^\tau(x) + \Lambda(\alpha_x)\mathbf{v}_-^\tau(x)\mathbf{v}_-^\tau(x))\mathbf{g}_i, \mathbf{w} \rangle \\ &= \langle \widehat{H}^{\tau,x}\mathbf{g}_i, \mathbf{w} \rangle. \end{aligned}$$

Here, we used (5.36) and  $\mathbf{w} \perp \mathbf{v}_\pm^\tau(x)$  in the second step and  $\text{supp } \mathbf{g}_i \subset B_{2r_\star-1}^\tau(x)$  as well as the definition of  $\widehat{H}^{\tau,x}$  from [12, eq. (3.48)] in the third step.

The remaining steps in [12, eq. (3.55)] and the remainder of [12, proof of Proposition 3.14 (i)] including the case  $2 + \xi^{1/4} > \alpha_x \geq \tau$ , which corresponds to the case  $x \in \mathcal{V}_\tau \setminus \mathcal{V}$  in [12], are obtained in the same way as in [12].  $\square$

*Proof of Lemma 5.16.* We write  $H^{(\mathcal{V} \setminus \mathcal{W})}$  in the block decomposition

$$H^{(\mathcal{V} \setminus \mathcal{W})} = \Pi H^{(\mathcal{V} \setminus \mathcal{W})} \Pi + \overline{\Pi} H^{(\mathcal{V} \setminus \mathcal{W})} \overline{\Pi} + \overline{\Pi} H^{(\mathcal{V} \setminus \mathcal{W})} \Pi + \Pi H^{(\mathcal{V} \setminus \mathcal{W})} \overline{\Pi}.$$

<sup>11</sup> Note that  $\mathcal{V}_\tau = \{x \in [N] : \alpha_x \geq \tau\}$  and  $\mathcal{V} = \mathcal{V}_{2+\xi^{1/4}}$  in [12], see [12, eq. (3.3)], which differs from the definition of  $\mathcal{V}$  in the present paper, see (2.1).

The nonzero eigenvalues of the block diagonal arise as the eigenvalues of the individual diagonal blocks, i.e.

$$\begin{aligned} & \text{spec}(\Pi H^{(\mathcal{V}\setminus\mathcal{W})} \Pi + \overline{\Pi} H^{(\mathcal{V}\setminus\mathcal{W})} \overline{\Pi}) \setminus \{0\} \\ &= (\text{spec}(\Pi H^{(\mathcal{V}\setminus\mathcal{W})} \Pi) \cup \text{spec}(\overline{\Pi} H^{(\mathcal{V}\setminus\mathcal{W})} \overline{\Pi})) \setminus \{0\}, \end{aligned}$$

counted with multiplicity. Therefore, for any  $1 \leq i, j \leq N$  there are at least  $i + j$  eigenvalues of the block diagonal larger than  $\min\{\lambda_i(\overline{\Pi} H^{(\mathcal{V}\setminus\mathcal{W})} \overline{\Pi}), \lambda_j(\Pi H^{(\mathcal{V}\setminus\mathcal{W})} \Pi)\}$  provided this number is positive. Hence, we conclude

$$\min\{\lambda_1(\overline{\Pi} H^{(\mathcal{V}\setminus\mathcal{W})} \overline{\Pi}), \lambda_{1+|\mathcal{W}|}(\Pi H^{(\mathcal{V}\setminus\mathcal{W})} \Pi)\} \leq \lambda_{2+|\mathcal{W}|}(H^{(\mathcal{V}\setminus\mathcal{W})}) + 2\|\overline{\Pi} H^{(\mathcal{V}\setminus\mathcal{W})} \Pi\|. \tag{5.40}$$

Moreover, using eigenvalue interlacing (Lemma D.3) and Proposition 3.4 (i), we obtain

$$\lambda_{2+|\mathcal{W}|}(H^{(\mathcal{V}\setminus\mathcal{W})}) \leq \lambda_2(H^{(\mathcal{V})}) \leq \Lambda(\alpha^*) + o(1). \tag{5.41}$$

Lemma 5.16 follows from (5.40) and (5.41) provided that we show that there is a constant  $c > 0$  such that

$$\lambda_{1+|\mathcal{W}|}(\Pi H^{(\mathcal{V}\setminus\mathcal{W})} \Pi) \geq \Lambda(\alpha^*) + c, \tag{5.42}$$

and

$$\|\overline{\Pi} H^{(\mathcal{V}\setminus\mathcal{W})} \Pi\| = o(1). \tag{5.43}$$

For the proof of (5.42), we recall the projections  $\Pi_{\mathcal{W}}$  from Definition 5.7 and  $Q_0$  from Definition 5.11. By the definition of  $\mathbf{u}(x)$  for  $x \in \mathcal{W}$ , we have  $\Pi_{\mathcal{W}} = Q_0 \Pi_{\mathcal{W}}$ . Hence, (5.10) implies  $(H - \mathcal{D})Q_0 \Pi_{\mathcal{W}} = o(1)$ . Applying  $Q_0$  to the last relation yields

$$H^{(\mathcal{V}\setminus\mathcal{W})} \Pi_{\mathcal{W}} = \mathcal{D} \Pi_{\mathcal{W}} + o(1) \tag{5.44}$$

as  $Q_0 H Q_0 = H^{(\mathcal{V}\setminus\mathcal{W})}$  and  $Q_0 \mathcal{D} Q_0 = \mathcal{D}$ . Since  $\text{ran } \Pi_{\mathcal{W}} \subset \text{ran } \Pi$  and  $\Pi_{\mathcal{W}} \mathcal{D} \Pi_{\mathcal{W}} = \mathcal{D}$ , the definition of  $\mathcal{D}$  and (5.44) imply that  $\Pi H^{(\mathcal{V}\setminus\mathcal{W})} \Pi$  has at least  $|\mathcal{W}|$  many eigenvalues in  $[\min_{x \in \mathcal{W}} \lambda(x) - o(1), \max_{x \in \mathcal{W}} \lambda(x) + o(1)]$ . Note that  $\min_{x \in \mathcal{W}} \lambda(x) \geq \Lambda(\alpha^*) + c$  for some constant  $c > 0$  by (3.3) and (2.2).

Furthermore, let  $\mathbf{q}$  be as in Proposition 3.4 (iv) with  $r = 1$ . From  $\|H^{(\mathcal{V}\setminus\mathcal{W})}\| \leq \|H\| \lesssim \sqrt{d}$  by (5.24) and  $\|\mathbb{E}H\| \lesssim \sqrt{d}$ , Proposition 3.4 (iv) and  $H^{(\mathcal{V}\setminus\mathcal{W})} \mathbf{q} = H \mathbf{q}$ , we deduce

$$H^{(\mathcal{V}\setminus\mathcal{W})} \mathbf{w}_1 = H^{(\mathcal{V}\setminus\mathcal{W})} \mathbf{q} + o(1) = H \mathbf{q} + o(1) = H \mathbf{w}_1 + o(1) = \lambda_1(H) \mathbf{w}_1 + o(1). \tag{5.45}$$

Hence,  $\Pi H^{(\mathcal{V}\setminus\mathcal{W})} \Pi$  has an eigenvalue at  $\lambda_1(H) + o(1) = \sqrt{d}(1 + o(1)) \gtrsim \sqrt{d} \gg \max_{x \in \mathcal{W}} \lambda(x) + o(1)$  by Proposition 3.4 (ii), (3.1) and (3.3). Therefore,  $\Pi H^{(\mathcal{V}\setminus\mathcal{W})} \Pi$  has  $1 + |\mathcal{W}|$  many eigenvalues larger or equal to  $\Lambda(\alpha^*) + c$  for some constant  $c > 0$ . This proves (5.42).

Finally, (5.43) follows from (5.44),  $\text{ran } \mathcal{D} \subset \text{ran } \Pi_{\mathcal{W}} \subset \text{ran } \Pi$  and (5.45). This completes the proof of Lemma 5.16.  $\square$

### 6. Eigenvalue spacing: Proof of Proposition 2.4

We recall the definition of the high-probability event  $\Omega$  from Definition 3.5. In this section we use the notation from (5.1), as well as the conditional versions

$$\mathbb{P}_\Omega(A \mid \mathcal{F}) := \mathbb{P}(\Omega \cap A \mid \mathcal{F}), \quad \mathbb{E}_\Omega[X \mid \mathcal{F}] := \mathbb{E}[X \mathbb{1}_\Omega \mid \mathcal{F}].$$

Throughout this section, we assume that  $d$  satisfies (1.5),  $\mu \in (0, 1/3)$  and that  $\eta$  satisfies

$$0 < \frac{\eta}{2} < \min \left\{ \frac{1}{6}, \frac{1}{5} - \frac{\mu}{4}, \frac{1}{3} - \mu \right\}. \tag{6.1}$$

Proposition 2.4 follows directly from the following result.

**Proposition 6.1.** *For any  $a \neq b \in [N]$ , we have*

$$\mathbb{P}_\Omega(a, b \in \mathcal{W}, |\lambda(a) - \lambda(b)| \leq N^{-\eta}) \leq N^{-2+2\mu-\eta/4+o(1)}. \tag{6.2}$$

*Remark 6.2.* If we restrict ourselves to the critical regime  $d \asymp \log N$ , then Proposition 6.1 can be improved by replacing the factor  $N^{-\eta}$  inside the probability in (6.2) by  $N^{-\eta/2}$ . See Remark 6.27 below for more details.

*Proof of Proposition 2.4.* The condition (6.1) holds by assumptions on  $\mu$  and  $\eta$ . A union bound and Proposition 6.1 yield

$$\begin{aligned} & \mathbb{P}(\exists x \neq y \in \mathcal{W} : |\lambda(x) - \lambda(y)| \leq N^{-\eta}) \\ & \leq \mathbb{P}(\Omega^c) + \binom{N}{2} \sup_{a \neq b \in [N]} \mathbb{P}_\Omega(a, b \in \mathcal{W}, |\lambda(a) - \lambda(b)| \leq N^{-\eta}) \\ & \leq o(1) + N^{2\mu-\eta/4+o(1)}, \end{aligned}$$

where we used that  $\mathbb{P}(\Omega^c) = o(1)$  by definition of  $\Omega$  (recall Definition 3.5). As  $\eta > 8\mu$ , we conclude that the right-hand side is  $o(1)$ .  $\square$

*6.1. Key tools of the proof of Proposition 6.1.* The rest of this section is devoted to the proof of Proposition 6.1. Throughout, we fix deterministic vertices  $a \neq b \in [N]$  and suppose that  $\eta$  satisfies (6.1). We use the following definitions. Let

$$r := \left\lfloor \frac{\eta}{2} \frac{\log N}{\log d} \right\rfloor - 1. \tag{6.3}$$

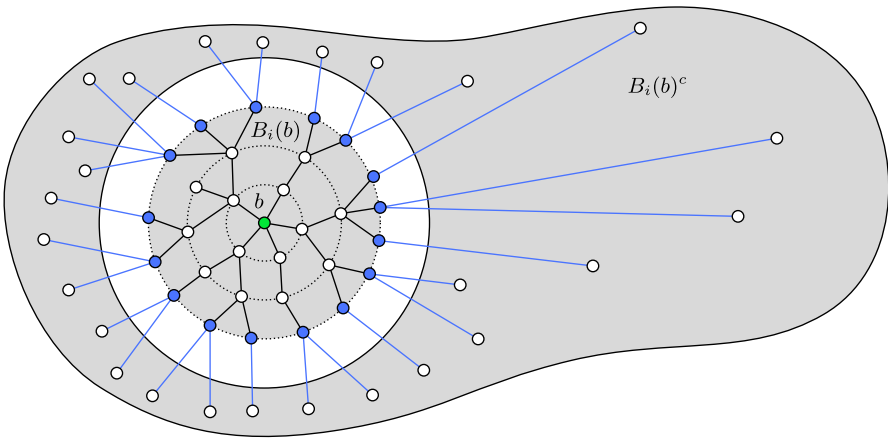
In particular,  $d^{r+1} \leq N^{\eta/2} < d^{r+2}$ . Note that  $r$  from (6.3) satisfies the condition of Proposition 3.1 (ii) and (3.2), i.e. the condition of Proposition 3.2.

**Definition 6.3.** We define the  $\sigma$ -algebra

$$\mathcal{F}_i := \sigma(B_i(b), A|_{B_i(b)}, A|_{B_i(b)^c})$$

for  $0 \leq i \leq r$ , and we abbreviate  $\mathcal{F} \equiv \mathcal{F}_r$ .

More explicitly, we define  $\mathcal{F}_i$  inductively through the filtration  $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \dots \subset \mathcal{G}_i$ , where  $\mathcal{G}_1 = \sigma(A|_{\{b\}})$  and  $\mathcal{G}_{k+1} = \sigma(\mathcal{G}_k, A|_{S_k(b)})$ , since by construction  $S_k(b)$  is  $\mathcal{G}_k$ -measurable. Then we set  $\mathcal{F}_i = \sigma(\mathcal{G}_i, A|_{B_i(b)^c})$ , using that  $B_i(b)$ , and hence  $B_i(b)^c$ , is  $\mathcal{G}_i$ -measurable. See Fig. 7 for an illustration of  $\mathcal{F}_i$ . The following lemma is an immediate consequence of the definition of  $\mathcal{F}_i$  and the independence of the family  $(A_{xy})$ .



**Fig. 7.** An illustration of the  $\sigma$ -algebra  $\mathcal{F}_i$ . Here  $i = 3$ , and the vertex  $b$  is drawn in green. Conditioning on  $\mathcal{F}_i$  means that the graph is fixed in the ball  $B_i(b)$  and its complement, drawn in grey. The only randomness is the choice of the edges from  $S_i(b)$  to  $B_i(b)^c$ , drawn in blue. By Lemma 6.4, these edges are chosen independently with probability  $d/N$

**Lemma 6.4.** *Conditionally on  $\mathcal{F}_i$ , the random variables  $(A_{xy} : x \in S_i(b), y \in B_i(b)^c)$  are independent Bernoulli random variables with mean  $d/N$ .*

**Definition 6.5.** For  $0 \leq i \leq r$  define  $H(i) := H^{(B_i(b) \cup \mathcal{V}^{(B_i(b))})}$  and  $G(i, z) := (H(i) - z)^{-1}$ .

By definition,  $H(i)$  is  $\mathcal{F}_i$ -measurable. We shall need the following regularization of the function  $t \mapsto t^{-1}$ .

**Definition 6.6.** Define  $\iota : \mathbb{R} \rightarrow \mathbb{R}$  through

$$\iota(t) := \begin{cases} t^{-1} & \text{if } t \in [T^{-1}, T] \\ -t + T + T^{-1} & \text{otherwise,} \end{cases} \tag{6.4}$$

with  $T := 10 \max\{\sqrt{d^{-1} \log N}, \kappa^{-1}\}$ .

*Remark 6.7.* The function  $\iota$  is an involution on  $\mathbb{R}$  with Lipschitz constant  $T^2$ .

**Definition 6.8.** For  $x \in S_i(b)$  we define  $S_1^+(x) := S_1(x) \cap S_{i+1}(b)$ .

**Definition 6.9.** For each  $z \in \mathbb{R}$ , we define the family  $(g_x(z))_{x \in B_r(b) \setminus \{b\}}$  recursively through

$$\begin{cases} g_x(z) = -\iota\left(z + \frac{1}{d} \sum_{y \in S_1^+(x) \setminus \mathcal{V}^{(B_r(b))}} G_{yy}(r, z)\right) & \text{if } x \in S_r(b) \\ g_x(z) = -\iota\left(z + \frac{1}{d} \sum_{y \in S_1^+(x)} g_y(z)\right) & \text{if } x \in B_{r-1}(b) \setminus \{b\}. \end{cases} \tag{6.5}$$

The following remark contains a crucial independence property of the family  $(g_x(z))_x$ .

*Remark 6.10.* Conditioned on  $\mathcal{F}$ , if  $\mathbb{G}|_{B_r(b)}$  is a tree and  $z$  is  $\mathcal{F}$ -measurable, then for any  $1 \leq i \leq r$  the family  $(g_x(z))_{x \in S_i(b)}$  is independent. To see this, we first note that  $(g_x(z))_{x \in S_r(b)}$  is independent conditioned on  $\mathcal{F}$  because of Lemma 6.4. For  $i \leq r - 1$ , the statement follows inductively by using the tree structure of  $\mathbb{G}|_{B_r(b)}$ .

*Remark 6.11.* The function  $\iota: \mathbb{R} \rightarrow \mathbb{R}$  is a regularized version of the function  $t \mapsto t^{-1}$  as a function  $\mathbb{R}_+ \rightarrow \mathbb{R}_+$ . The regularization acts on both small values of  $t$  (for  $t < T^{-1}$ ) and large values of  $t$  (for  $t > T$ ). The former regularization is needed to ensure the Lipschitz continuity of the function  $\iota$ , which is used in the proof of Proposition 6.15 to ensure the stability of  $g_x(z)$  under change of the argument  $z$ . The latter regularization is needed to ensure the Lipschitz continuity of the function  $\iota^{-1}$ , which is used in the proof of Proposition 6.26 below to ensure that anticoncentration of a random variable is preserved, up to a factor  $T^2$ , by applying  $\iota$ . In the proof of Proposition 6.15 below, we show and use that with high probability the argument of  $\iota$  is always contained in the interval  $[T^{-1}, T]$  where  $\iota$  coincides with  $t \mapsto t^{-1}$ . Moreover, we choose the lower and upper bounds,  $T^{-1}$  and  $T$ , to be each other’s inverses for convenience, since in that case  $\iota$  is an involution. Actually, an inspection of our proof shows that the lower bound  $T^{-1}$  could be replaced with the larger value  $\kappa/10$ . Finally, we note that in the critical regime  $d \asymp \log N$ , the parameter  $T$  is of order one. This observation can be used to improve Proposition 6.1 somewhat in that regime; see Remarks 6.2 and 6.27.

We now state the three key propositions that underlie the proof of Propositions 6.1–6.12, 6.14, and 6.15. Their proofs are postponed to Sects. 6.5, 6.3, and 6.4, respectively.

Using the independence from Remark 6.10, we obtain the following anticoncentration estimate for the family  $(g_x(z))_{x \in B_r(b) \setminus \{b\}}$ .

**Proposition 6.12** (Anticoncentration of  $g_x$ ). *Let  $z \in \mathcal{J}$  be  $\mathcal{F}$ -measurable and  $(g_x(z))_x$  be defined as in Definition 6.9. Then for any  $a, b \in [N]$  we have*

$$\mathbb{P}_\Omega \left( a, b \in \mathcal{W}, \left| \frac{1}{d} \sum_{x \in S_1(b)} g_x(z) + z \right| \leq N^{-\eta} \right) \leq N^{-2+2\mu-\eta/4+o(1)}. \quad (6.6)$$

For each  $0 \leq i \leq r$  we shall need the following  $\mathcal{F}_i$ -measurable approximation of  $\lambda(a)$ .

**Definition 6.13.** For  $i \in [r]$  we abbreviate  $\lambda(a, i) := \lambda_2(H^{(\mathcal{V}^{(B_i(b)) \cup B_i(b)} \setminus \{a\})})$ .

By definition,  $\lambda(a, i)$  is  $\mathcal{F}_i$ -measurable. The next results states that  $\lambda(a, i)$  is with high probability close to  $\lambda(a)$ .

**Proposition 6.14** (Comparison of  $\lambda(a)$  and  $\lambda(a, i)$ ). *For any small enough  $\varepsilon > 0$ , we have*

$$\mathbb{P}_\Omega(a, b \in \mathcal{W}, \exists i \in [r], |\lambda(a) - \lambda(a, i)| \geq \varepsilon) \leq \varepsilon^{-2} N^{-3+2\mu+\eta/2+o(1)}.$$

The next result states that, when choosing the spectral parameter  $z = \lambda(a, r)$ , the Green function entries of  $H^{(\mathcal{V})}$  on  $S_1(b)$  are well approximated by the family  $(g_x)$  from Definition 6.9.

**Proposition 6.15** (Approximation of Green function by  $g_x$ ). *Let  $(g_x(z))_x$  be defined as in Definition 6.9 with  $z := \lambda(a, r)$ . For any constant  $c > 0$  and any  $\varepsilon \leq N^{-c}$ , we have*

$$\mathbb{P}_\Omega \left( a, b \in \mathcal{W}, \exists x, y \in S_1(b), |(H^{(\mathcal{V})} - z)_{xy}^{-1} - g_x(z)\mathbb{1}_{x=y}| \geq \varepsilon \right) \leq \varepsilon^{-2} N^{-3+2\mu+2\eta+o(1)}.$$

6.2. *Proof of Proposition 6.1.* In this subsection we prove Proposition 6.1. We begin by introducing the following events that we use throughout this section. Recall the definitions of the sets  $\mathcal{V}^{(X)}$  and  $\mathcal{W}^{(X)}$  from (5.4) and (5.5).

**Definition 6.16.** We define

$$\Xi := \{a, b \in \mathcal{W}\},$$

and,  $1 \leq i \leq r$ ,

$$\Xi_i := \{a \in \mathcal{W}^{(B_i(b))}, b \in \mathcal{W}\}.$$

*Remark 6.17.* We record the following straightforward properties of  $\Xi$  and  $\Xi_i$ .

- (i)  $\Xi_i$  is  $\mathcal{F}_i$ -measurable.
- (ii)  $\Xi_i \subset \Xi$  (by definition of  $\mathcal{W}^{(B_i(b))}$ ).
- (iii) On  $\Omega$ , for any  $b \in \mathcal{W}$  and  $1 \leq i \leq r$  we have  $\mathcal{V}^{(B_i(b))} = \mathcal{V} \setminus \{b\}$  and  $\mathcal{W}^{(B_i(b))} = \mathcal{W} \setminus \{b\}$  (see Proposition 3.2 (iii)). In particular,  $\Omega \cap \Xi_i = \Omega \cap \Xi$  for all  $1 \leq i \leq r$ .

Next, we state a basic result that, together with Lemma 3.8, is used throughout this section to establish the boundedness of the Green function for certain spectral parameters.

**Lemma 6.18.** *On  $\Omega \cap \Xi$  we have  $\lambda(a, i) = \Lambda(\alpha_a) + o(1)$  and  $\lambda(a, i) \in \mathcal{J}$  for all  $1 \leq i \leq r$ .*

*Proof.* By Remark 6.17 (iii), on  $\Omega$  the assumptions of Corollary 3.6 are satisfied for  $x = a$  and  $X = (\mathcal{V}^{(B_i(b))} \cup B_i(b)) \setminus \{a\}$ . Therefore,  $\lambda(a, i) = \Lambda(\alpha_a) + o(1)$ , from which we conclude that  $\lambda(a, i) \geq \Lambda(\alpha^*) + \kappa/4$  by the definition (2.2) of  $\mathcal{W}$  as well as  $\lambda(a, i) \ll \sqrt{d}$  by (3.1). □

*Proof of Proposition 6.1.* By spectral decomposition of  $H^{(\mathcal{V} \setminus \{b\})}$ , we have  $\text{Im}(H^{(\mathcal{V} \setminus \{b\})} - \lambda(b) - it)_{bb}^{-1} \geq \langle \mathbf{1}_b, \mathbf{u}(b) \rangle^2 / t$  for any  $t > 0$ . By Corollary 3.7, on  $\Omega$  we have  $\langle \mathbf{1}_b, \mathbf{u}(b) \rangle \neq 0$ , which implies

$$\lim_{t \downarrow 0} \frac{1}{(H^{(\mathcal{V} \setminus \{b\})} - \lambda(b) - it)_{bb}^{-1}} = 0,$$

and hence Schur’s complement formula yields

$$\lambda(b) + \frac{1}{d} \sum_{x, y \in S_1(b)} (H^{(\mathcal{V})} - \lambda(b))_{xy}^{-1} = 0.$$

Therefore, with  $z = \lambda(a, r)$ , we obtain from the definition of the family  $(g_x(\lambda(a, r)))_x$  in Definition 6.9 that

$$\begin{aligned} & \left| \lambda(a, r) + \frac{1}{d} \sum_{x \in S_1(b)} g_x(\lambda(a, r)) \right| \\ & \leq |\lambda(a, r) - \lambda(b)| + \frac{1}{d} \sum_{x, y \in S_1(b)} |(H^{(\mathcal{V})} - \lambda(a, r))_{xy}^{-1} - (H^{(\mathcal{V})} - \lambda(b))_{xy}^{-1}| \\ & \quad + \frac{1}{d} \sum_{x, y \in S_1(b)} |g_x(\lambda(a, r)) \mathbb{1}_{x=y} - (H^{(\mathcal{V})} - \lambda(a, r))_{xy}^{-1}| \\ & \leq C\kappa^{-2}(\log N)^2(|\lambda(a) - \lambda(b)| + |\lambda(a, r) - \lambda(a)|) \\ & \quad + \frac{1}{d} \sum_{x, y \in S_1(b)} |g_x(\lambda(a, r)) \mathbb{1}_{x=y} - (H^{(\mathcal{V})} - \lambda(a, r))_{xy}^{-1}| \end{aligned}$$



on the event  $\Omega$ , where we used (3.5) and Proposition 3.2 (i) for the second inequality. Here  $C$  is some positive constant. Thus, for any  $\gamma \geq 0$ , we obtain

$$\begin{aligned} & \mathbb{P}_{\Omega \cap \Xi}(C\kappa^{-2}(\log N)^2|\lambda(b) - \lambda(a)| \leq \gamma) \\ & \leq \mathbb{P}_{\Omega \cap \Xi}\left(\left|\frac{1}{d} \sum_{x \in S_1(b)} g_x(\lambda(a, r)) + \lambda(a, r)\right| \leq 3\gamma\right) \\ & \quad + \mathbb{P}_{\Omega \cap \Xi}(C\kappa^{-2}(\log N)^2|\lambda(a) - \lambda(a, r)| \geq \gamma) \\ & \quad + \mathbb{P}_{\Omega \cap \Xi}\left(\frac{1}{d} \sum_{x, y \in S_1(b)} \left| (H^{(\mathcal{V})} - \lambda(a, r))_{xy}^{-1} - g_x(\lambda(a, r)) \mathbb{1}_{x=y} \right| \geq \gamma\right). \end{aligned}$$

Now Proposition 6.1 follows with the choice  $\gamma := N^{-\eta}/3$ , applying Proposition 6.12 to the first line, Proposition 6.14 to the second line, and Proposition 6.15 combined with Proposition 3.2 (i) on  $\Omega$  to the third line. Here we used Lemma 6.18 to ensure that  $\lambda(a, i) \in \mathcal{J}$ . □

6.3. Proof of Proposition 6.14. In this subsection we prove Proposition 6.14.

*Proof of Proposition 6.14.* We follow the proof of Proposition 5.3. Let  $\mathbf{u}(a, i)$  be a normalized eigenvector of  $H^{(\mathcal{V}^{(B_i(b))} \cup B_i(b)) \setminus \{a\}}$  associated with the eigenvalue  $\lambda(a, i)$ . Then, as  $\text{supp } \mathbf{u}(a, i) \subset (\mathcal{V}^{(B_i(b))} \cup B_i(b))^c \cup \{a\}$ , on the event  $\Omega \cap \{a, b \in \mathcal{W}\}$ , using Remark 6.17 (iii) we obtain

$$(H^{(\mathcal{V} \setminus \{a\})} - \lambda(a, i))\mathbf{u}(a, i) = (H^{(\mathcal{V} \setminus \{a\})} - H^{(\mathcal{V}^{(B_i(b))} \cup B_i(b)) \setminus \{a\}})\mathbf{u}(a, i) = \sum_{y \in S_i(b)} \varepsilon_{ay}(i)\mathbf{1}_y, \tag{6.7}$$

where  $\varepsilon_{ay}(i) := \frac{1}{\sqrt{d}} \sum_{v \in S_1^+(y)} \langle \mathbf{1}_v, \mathbf{u}(a, i) \rangle$ . From Lemma 6.4 we conclude, using Cauchy-Schwarz,

$$\begin{aligned} & \mathbb{E}_{\Omega} \left[ \sum_{y \in S_i(b)} (\varepsilon_{ay}(i))^2 \middle| \mathcal{F}_i \right] \\ & \lesssim \sum_{y \in S_i(b)} \frac{\log N}{d} \sum_{v \in B_i(b)^c} \mathbb{E}[\mathbb{1}_{v \in S_1^+(y)} \langle \mathbf{1}_v, \mathbf{u}(a, i) \rangle^2 \middle| \mathcal{F}_i] \\ & = \sum_{y \in S_i(b)} \frac{\log N}{d} \sum_{v \in B_i(b)^c} \mathbb{P}(v \in S_1^+(y) \mid \mathcal{F}_i) \langle \mathbf{1}_v, \mathbf{u}(a, i) \rangle^2 \leq \frac{\log N}{N} |S_i(b)|, \end{aligned} \tag{6.8}$$

where we used that on  $\Omega$  we have  $|S_1^+(y)| \lesssim \log N$ , that  $\mathbf{u}(a, i)$  is  $\mathcal{F}_i$ -measurable, and that by Lemma 6.4 we have  $\mathbb{P}(v \in S_1^+(y) \mid \mathcal{F}_i) = \frac{d}{N}$  for any  $y \in S_i(b)$  and  $v \notin B_i(b)$ .

By the definition (2.2) of  $\mathcal{W}$ , we have  $\Lambda(\alpha_a) \geq \Lambda(\alpha^*) + \kappa/2$ , and on  $\Omega$  we have  $\Lambda(\alpha_a) \ll \sqrt{d}$  (recall (3.1)). Hence, by Corollary 3.6, for any small enough  $\varepsilon > 0$ , if there are a scalar  $\hat{\lambda}$  and a normalized vector  $\hat{\mathbf{u}}$  such that, for some  $a \in \mathcal{W}$ ,  $\|(H^{(\mathcal{V} \setminus \{a\})} - \hat{\lambda})\hat{\mathbf{u}}\| \leq \varepsilon$  and  $\hat{\lambda} = \Lambda(\alpha_a) + o(1)$ , then  $|\lambda(a) - \hat{\lambda}| \leq \varepsilon$  (as  $\lambda(a) = \lambda_2(H^{(\mathcal{V} \setminus \{a\})})$ ).

We apply this observation to the choices  $\hat{\lambda} = \lambda(a, i)$  and  $\hat{\mathbf{u}} = \mathbf{u}(a, i)$ , for which Lemma 6.18 yields  $\lambda(a, i) = \Lambda(\alpha_a) + o(1)$ . Hence, for any small enough  $\varepsilon > 0$ , we

have  $|\lambda(a) - \lambda(a, i)| \leq \varepsilon$  provided that  $\|(H^{(\mathcal{V} \setminus \{a\})} - \lambda(a, i))\mathbf{u}(a, i)\| \leq \varepsilon$ . We can then estimate

$$\begin{aligned} & \mathbb{P}_\Omega(a, b \in \mathcal{W}, |\lambda(a) - \lambda(a, i)| > \varepsilon) \\ & \leq \mathbb{P}_\Omega\left(\Xi_i \cap \left\{\|(H^{(\mathcal{V} \setminus \{a\})} - \lambda(a, i))\mathbf{u}(a, i)\|^2 > \varepsilon^2\right\}\right) \\ & \leq \varepsilon^{-2} \mathbb{E}\left[\mathbb{1}_{\Xi_i} \mathbb{1}_{|S_i(b)| \leq N^{\eta/2+o(1)}} \mathbb{E}_\Omega\left[\sum_{y \in S_i(b)} (\varepsilon_{ay}(i))^2 \middle| \mathcal{F}_i\right]\right] \\ & \leq \varepsilon^{-2} \mathbb{P}(\Xi_i) \frac{\log N}{N} N^{\eta/2+o(1)} \leq \varepsilon^{-2} N^{2\mu+\eta/2-3+o(1)}, \end{aligned}$$

where in the first step we used Remark 6.17 (iii), in the second step (6.7), Remark 6.17 (i), and the estimate  $|S_i(b)| \leq N^{\eta/2+o(1)}$  on  $\Omega$  by Proposition 3.2 (ii) and (6.3), in the third step (6.8), and the fourth step Remark 6.17 (ii) and Lemma 6.19 below. Now Proposition 6.14 follows from a union bound over  $i \in [r]$  with  $r \lesssim \log N$ .  $\square$

The following result is used throughout the rest of this section.

**Lemma 6.19.** *For any  $a \neq b \in [N]$  we have  $\mathbb{P}(a, b \in \mathcal{W}) \leq N^{-2+2\mu}$ .*

*Proof.* Since  $0 \leq \Lambda'(\alpha) = \frac{\alpha-2}{2(\alpha-1)^{3/2}} \leq \frac{1}{2}$  for all  $\alpha \geq 2$ , we have

$$\begin{aligned} P(a, b \in \mathcal{W}) & \leq \mathbb{P}\left(\min\{|S_1(a)|, |S_1(b)|\} \geq (\alpha^* + \kappa)d\right) \\ & = \mathbb{E}\left[\mathbb{P}\left(\min\{|S_1(a)|, |S_1(b)|\} \geq (\alpha^* + \kappa)d \middle| A_{ab}\right)\right] \\ & \leq \mathbb{P}(|S_1(a)| \geq (\alpha^* + \kappa)d - 1)^2 \leq \mathbb{P}(|S_1(a)| \geq \alpha^*d)^2 \leq N^{-2+2\mu}, \end{aligned}$$

where we used in the third step that conditionally on  $A_{ab}$ ,  $S_1(a)$  and  $S_1(b)$  are independent and in the last step the definition of  $\alpha^*$ .  $\square$

**6.4. Proof of Proposition 6.15.** This subsection is devoted to the proof of Proposition 6.4. We begin with the following result, which contains two estimates. The first one is an approximate version of Schur’s complement formula, where  $G(i - 1, z)$  is related to  $G(i, z)$  at the cost of an error term; this amounts to removing not just the vertex  $x \in S_i(b)$  at which the Green function is evaluated, but the entire ball  $B_i(b)$ . The second estimate provides an upper bound on the off-diagonal entries of the Green function.

**Lemma 6.20.**

(i) *For  $1 \leq i \leq r$  let  $z_i \in \mathcal{J}$  be  $\mathcal{F}_i$ -measurable and  $x \in S_i(b)$ . Then*

$$-\frac{1}{G_{xx}(i - 1, z_i)} = z_i + \frac{1}{d} \sum_{y \in S_1^+(x)} G_{yy}(i, z_i) + \mathcal{E}_i(x) \tag{6.9}$$

where the error term  $\mathcal{E}_i(x)$  satisfies, for any  $\varepsilon > 0$ ,

$$\mathbb{P}_\Omega(|\mathcal{E}_i(x)| > \varepsilon \mid \mathcal{F}_i) \mathbb{1}_{b \in \mathcal{W}} \lesssim d^2 \kappa^{-4} (\log N)^2 |S_i(b)|^2 N^{-1} \varepsilon^{-2}. \tag{6.10}$$

(ii) *Let  $z \in \mathcal{J}$  be  $\mathcal{F}_1$ -measurable. For any  $x \neq y \in [N]$  and  $\varepsilon > 0$ , we have*

$$\mathbb{P}_\Omega(|(H^{(\mathcal{V})} - z)_{xy}^{-1}| \geq \varepsilon \mid \mathcal{F}_1) \mathbb{1}_{b \in \mathcal{W}} \mathbb{1}_{x, y \in S_1(b)} \lesssim \kappa^{-4} (\log N) N^{-1} \varepsilon^{-2}. \tag{6.11}$$

*Proof of Proposition 6.15.* We choose  $z_i := \lambda(a, i)$ , which is  $\mathcal{F}_i$ -measurable, and set  $z = z_r$ . For  $\varepsilon > 0$  we introduce the event  $\Theta := \Theta_1 \cap \Theta_2 \cap \Theta_3$  given by

$$\begin{aligned} \Theta_1 &:= \{\forall i \leq r, \forall x \in S_i(b), |\mathcal{E}_i(x)| \leq \varepsilon\}, \\ \Theta_2 &:= \{\forall i \leq r, |z - z_i| \leq \varepsilon\}, \\ \Theta_3 &:= \{\forall x \neq y \in S_1(b), |(H^{(\mathcal{V})} - z_1)_{xy}^{-1}| \leq \varepsilon\}, \end{aligned}$$

with  $\mathcal{E}_i(x)$  defined as in Lemma 6.20. We estimate the probability of  $\Theta_1^c$  using  $\Xi_i \in \mathcal{F}_i$ , Lemmas 6.20 (i) and 6.19, Proposition 3.2 (ii), we well as Remark 6.17 as

$$\mathbb{P}_{\Omega \cap \Xi_i}(\exists x \in S_i(b), |\mathcal{E}_i(x)| > \varepsilon) \leq \mathbb{P}_{\Omega}(\Xi) N^{-1+3\eta/2+o(1)} \varepsilon^{-2} \leq N^{-3+2\mu+3\eta/2+o(1)} \varepsilon^{-2}.$$

Similarly, we find using Lemma 6.20 (ii) that

$$\mathbb{P}_{\Omega \cap \Xi_1}(\Theta_3^c) \leq \mathbb{P}_{\Omega}(\Xi) N^{-1+\eta+o(1)} \varepsilon^{-2} \leq N^{-3+2\mu+\eta/2+o(1)} \varepsilon^{-2}.$$

Hence, using Remark 6.17 and Proposition 6.14, we have

$$\begin{aligned} \mathbb{P}_{\Omega}(\Theta^c \cap \Xi) &\leq \sum_{i=1}^r \left( \mathbb{P}_{\Omega \cap \Xi_i}(\exists x \in S_i(b), |\mathcal{E}_i(x)| > \varepsilon) + \mathbb{P}_{\Omega \cap \Xi}(|z - z_i| > \varepsilon) \right) \\ &\quad + \mathbb{P}_{\Omega \cap \Xi_1}(\Theta_3^c) \\ &\leq N^{-3+2\mu+2\eta+o(1)} \varepsilon^{-2}. \end{aligned} \tag{6.12}$$

We shall show below that there is a constant  $C > 0$  such that on the event  $\Omega \cap \Theta \cap \Xi$  we have

$$|g_x(z) - G_{xx}(i-1, z)| \leq C\varepsilon \sum_{j=0}^{r-i+1} (C\kappa^{-4}d^{-1})^j |S_j^+(x)| \leq N^{o(1)}\varepsilon \tag{6.13}$$

for all  $1 \leq i \leq r$  and all  $x \in S_i(b)$ . The second inequality in (6.13) follows from  $r \ll \log N$  and  $d^{-j}|S_j^+(x)| \leq d^{-j}|B_j(x)| \lesssim \log N$  for all  $x \in [N]$  by Proposition 3.2 (ii).

Before proving (6.13), we conclude Proposition 6.15 from (6.13) (after renaming  $\varepsilon \mapsto N^{-o(1)}\varepsilon$ ) with  $i = 1$ , the definition of  $\Theta_3$  and (6.12), where we used that  $G(0, z) = (H^{(\mathcal{V})} - z)^{-1}$  on  $\Omega \cap \{b \in \mathcal{V}\}$  as  $\mathcal{V}^{(B_i(b))} = \mathcal{V} \setminus \{b\}$ .

What remains, therefore, is the proof of (6.13). We prove it by inductively decreasing  $i$  starting from  $i = r+1$ . By convention, for any  $x \in S_{r+1}(b)$  we denote  $g_x(z) := G_{xx}(r, z)$ , so that (6.13) trivially holds for  $i = r+1$ . Therefore, we can assume throughout the following argument that  $g_x(z)$  is defined by the second case in Definition 6.9 for all  $x \in B_r(b) \setminus \{b\}$  (note that on the event  $\Omega \cap \Xi$  we have  $S_1^+(x) \setminus \mathcal{V}^{(B_r(b))} = S_1^+(x)$  for any  $x \in S_r(b)$ , by Proposition 3.2 (iii)).

To verify the induction step, we assume that (6.13) holds on  $S_j(b)$  for all  $i+1 \leq j \leq r+1$  and consider  $x \in S_i(b)$ . We first show that on  $\Omega \cap \Xi \cap \Theta$

$$g_x(z) = - \left( z + \frac{1}{d} \sum_{y \in S_1^+(x)} g_y(z) \right)^{-1}. \tag{6.14}$$

To that end, we conclude from the induction hypothesis, (3.5), the definition of  $\Theta_2$  and Proposition 3.2 (i) that

$$\begin{aligned} & \left| \frac{1}{d} \sum_{y \in S_1^+(x)} G_{yy}(i, z_i) - \frac{1}{d} \sum_{y \in S_1^+(x)} g_y(z) \right| \\ & \leq \frac{1}{d} \sum_{y \in S_1^+(x)} (|G_{yy}(i, z) - g_y(z)| + |G_{yy}(i, z_i) - G_{yy}(i, z)|) \\ & \leq d^{-1} |S_1(x)| \left( \max_{y \in S_1^+(x)} |G_{yy}(i, z) - g_y(z)| + (8/\kappa)^2 |z - z_i| \right) \\ & \leq N^{o(1)} \varepsilon. \end{aligned} \tag{6.15}$$

By Lemma 6.18, on  $\Omega \cap \Xi$  we have  $z_i \in \mathcal{J}$ . We recall from Lemma 3.9 that all Green function entries  $G_{xx}(i - 1, z_i)$  and  $G_{yy}(i, z_i)$  are negative for  $z_i \in \mathcal{J}$ . We use the upper bound (3.4) for  $-G_{xx}(i - 1, z_i)$  as well as (6.9) to obtain, on  $\Omega \cap \Xi \cap \Theta_1 \cap \Theta_2$ ,

$$\begin{aligned} \frac{10}{9T} & \leq \frac{\kappa}{9} \leq -\frac{1}{G_{xx}(i - 1, z_i)} - \mathcal{E}_i(x) = z_i + \frac{1}{d} \sum_{y \in S_1^+(x)} G_{yy}(i, z_i) \\ & \leq \Lambda(\alpha_a) + o(1) \leq \sqrt{2\alpha_a} + o(1) \leq \frac{T}{2}. \end{aligned} \tag{6.16}$$

Here, we also used Corollary 3.6, which is applicable as  $\mathcal{V}^{(B_i(b))} = \mathcal{V} \setminus \{b\}$  on  $\Omega \cap \{b \in \mathcal{W}\}$  by Proposition 3.2 (iii),  $\alpha_a \leq 10d^{-1} \log N$  on  $\Omega$  by Proposition 3.2 (i), and the definitions of  $\Lambda$  and  $T$ . Then, by the assumption  $\varepsilon \leq N^{-c}$ , we find that (6.16) and (6.15) imply  $T^{-1} \leq z + \frac{1}{d} \sum_{y \in S_1^+(x)} g_y(z) \leq T$ , which yields (6.14) by the definitions of  $\iota$  and  $g_x(z)$  in (6.4) and Definition 6.9, respectively. This concludes the proof of (6.14).

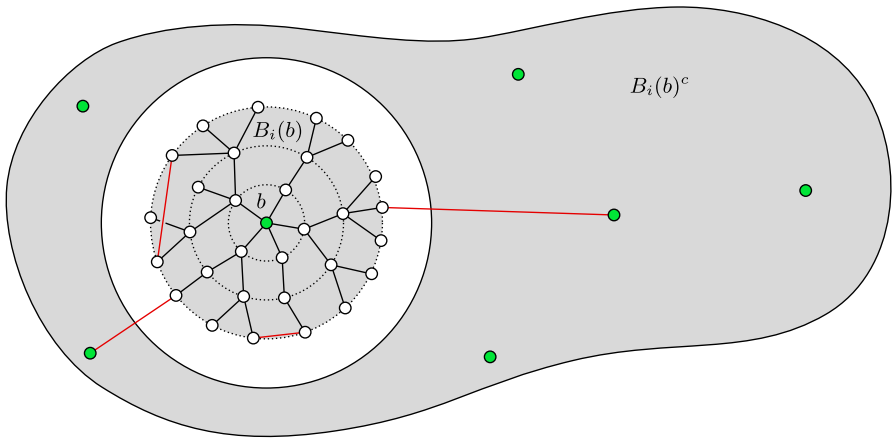
Finally, we find on  $\Omega \cap \Xi \cap \Theta$

$$\begin{aligned} & |g_x(z) - G_{xx}(i - 1, z)| \\ & \leq g_x(z) G_{xx}(i - 1, z_i) |g_x(z)^{-1} - G_{xx}(i - 1, z_i)^{-1}| + (8/\kappa)^2 |z - z_i| \\ & \leq (8/\kappa)^2 \left( \left| \frac{1}{d} \sum_{y \in S_1^+(x)} G_{yy}(i, z_i) - \frac{1}{d} \sum_{y \in S_1^+(x)} g_y(z) \right| + |z - z_i| + |\mathcal{E}_i(x)| \right) \\ & \leq (8/\kappa)^2 \left( \frac{1}{d} \sum_{y \in S_1^+(x)} \sum_{j=0}^{r-i} (C\kappa^{-4} d^{-1})^j |S_j^+(y)| C\varepsilon + ((8/\kappa)^2 d^{-1} |S_1^+(x)| + 1) |z - z_i| + |\mathcal{E}_i(x)| \right) \\ & \leq C^{-1} \left( 64 \sum_{j=1}^{r-i+1} (C\kappa^{-4} d^{-1})^j |S_j^+(x)| + 64^2 \kappa^{-4} (d^{-1} |S_1^+(x)| + 2) \right) C\varepsilon. \end{aligned}$$

Here we used (3.5) in the first inequality, (6.9) and (6.14) in the second, (3.4) and the iteration hypothesis in the third and that  $|S_{j+1}^+(x)| = \sum_{y \in S_1^+(x)} |S_j^+(y)|$  and the definitions of  $\Theta_1$  and  $\Theta_2$  in the last one. We conclude (6.13) for large enough  $C$ .  $\square$

What remains is the proof of Lemma 6.20.

*Proof of Lemma 6.20.* We begin with (i). Throughout the proof, we fix  $i$  and we condition on  $\mathcal{F}_i$ . We always assume that  $b \in \mathcal{W}$ , which is an  $\mathcal{F}_i$ -measurable event since  $i \geq 1$ .



**Fig. 8.** An illustration of the event  $\Gamma_i$  for  $i = 3$ . The vertices of  $\mathcal{V}$  are drawn in green. By definition of  $\Gamma_i$ , the red edges are forbidden

We start by observing that the event

$$\Gamma_i := \{\mathcal{V}^{(B_{i-1}(b))} = \mathcal{V}^{(B_i(b))}\} \cap \{A|_{S_i(b)} = 0\}$$

satisfies  $\Omega \subset \Gamma_i$ , by Proposition 3.2 (ii), (iii). See Fig. 8 for an illustration of  $\Gamma_i$ .

For the proof we abbreviate

$$\mathcal{T}_i := B_{i-1}(b) \cup \mathcal{V}^{(B_i(b))}, \tag{6.17}$$

so that on the event  $\Gamma_i$  we have  $H(i - 1) = H(\mathcal{T}_i)$ . From Schur’s complement formula we get on the event  $\Gamma_i$

$$-\frac{1}{G_{xx}(i - 1, z_i)} = z_i + \frac{1}{d} \sum_{u,v \in S_1^+(x)} (H(\mathcal{T}_i \cup \{x\}) - z_i)_{uv}^{-1}, \tag{6.18}$$

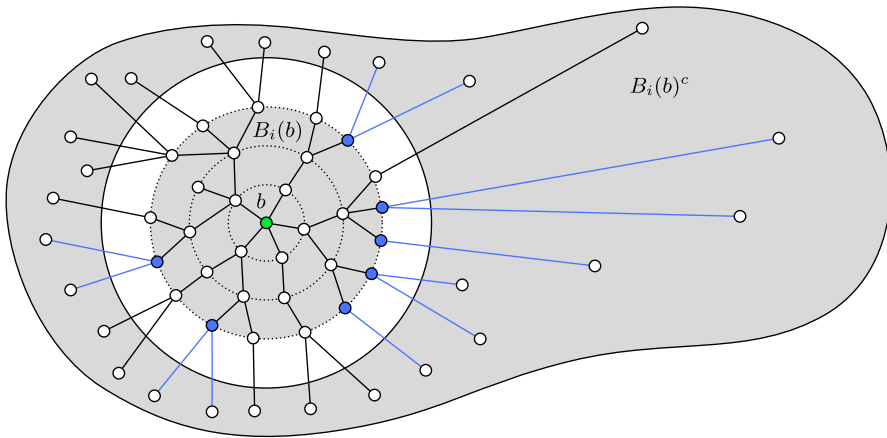
where we used that  $x$  has no neighbours in  $S_i(b)$  by definition of  $\Gamma_i$ .

We now decompose the error term  $\mathcal{E}_i(x)$  from (6.9) into several summands estimated separately. To that end, let  $\{x, y_1, \dots, y_{|S_i(b)|-1}\} = S_i(b)$  be an enumeration of  $S_i(b)$ . We set  $y_0 := x$  and

$$\begin{aligned} \mathcal{E}_i^{(0)} &:= \frac{1}{d} \sum_{u,v \in S_1^+(x), u \neq v} (H(\mathcal{T}_i \cup \{x\}) - z_i)_{uv}^{-1}, \\ \mathcal{E}_i^{(j)} &:= \frac{1}{d} \sum_{u \in S_1^+(x)} \left( (H(\mathcal{T}_i \cup \{y_0, \dots, y_j\}) - z_i)_{uu}^{-1} - (H(\mathcal{T}_i \cup \{y_0, \dots, y_{j-1}\}) - z_i)_{uu}^{-1} \right). \end{aligned}$$

for  $j = 1, \dots, |S_i(b)| - 1$ . Then, from (6.18) we conclude that (6.9) holds with  $\mathcal{E}_i(x) = \mathcal{E}_i^{(0)} - \sum_{j=1}^{|S_i(b)|-1} \mathcal{E}_i^{(j)}$ . Chebyshev’s and the Cauchy-Schwarz inequalities yield

$$\mathbb{P}_\Omega(|\mathcal{E}_i(x)| > \varepsilon | \mathcal{F}_i) \leq \varepsilon^{-2} |S_i(b)| \left( \mathbb{E}_\Omega[|\mathcal{E}_i^{(0)}|^2 | \mathcal{F}_i] + \sum_{j=1}^{|S_i(b)|-1} \mathbb{E}_\Omega[|\mathcal{E}_i^{(j)}|^2 | \mathcal{F}_i] \right). \tag{6.19}$$



**Fig. 9.** An illustration of the  $\sigma$ -algebra  $\mathcal{F}_i(X)$ . Here  $i = 3$ , the vertex  $b$  is drawn in green, and the set  $X \subset S_i(b)$  is drawn in blue. Conditioning on  $\mathcal{F}_i(X)$  means that we fix all edges within  $B_i(b)$ , within  $B_i(b)^c$ , and connecting  $S_i(b) \setminus X$  with  $B_i(b)^c$ . The only randomness is the choice of the edges from  $X$  to  $B_i(b)^c$ , drawn in blue. After removal of the vertices in  $\mathcal{T}_i \cup X$  (see (6.17)), only black edges and edges within  $B_i(b)^c$  remain, which shows that  $H^{(\mathcal{T}_i \cup X)}$  is  $\mathcal{F}_i(X)$ -measurable. Note that for  $X = S_i(b)$  we have  $\mathcal{F}_i(X) = \mathcal{F}_i$ , and we recover the illustration from Fig. 7

Therefore, it remains to estimate  $\mathbb{E}_\Omega[|\mathcal{E}_i^{(j)}|^2 | \mathcal{F}_i]$  for all  $j = 0, \dots, |S_i(b)| - 1$ . To that end, we introduce a  $\sigma$ -algebra refining  $\mathcal{F}_i$  from Definition 6.3. For any subset  $X \subset S_i(b)$  we define the  $\sigma$ -algebra<sup>12</sup>

$$\mathcal{F}_i(X) := \sigma(\mathcal{F}_i, (S_1^+(y))_{y \in S_i(b) \setminus X}),$$

using that  $S_i(b)$  is  $\mathcal{F}_i$ -measurable. See Fig. 9 for an illustration of  $\mathcal{F}_i(X)$ . Moreover, for  $X \subset S_i(b)$  we define the event

$$\Delta_i(X) := \left\{ \| (H^{(\mathcal{T}_i \cup X)} - z_i)^{-1} \| \leq 8\kappa^{-1} \right\}.$$

We note that for any  $X \subset S_i(b)$ , the event  $\Delta_i(X)$  lies in  $\mathcal{F}_i(X)$ , since  $H^{(\mathcal{T}_i \cup X)}$  is  $\mathcal{F}_i(X)$ -measurable (see the definition (6.17) and Fig. 9). Furthermore, by Lemma 3.8 for  $z_i \in \mathcal{J}$  we have

$$\Omega \subset \bigcap_{X \subset S_i(b)} \Delta_i(X). \tag{6.20}$$

For the estimate of  $\mathcal{E}_i^{(0)}$ , we introduce the sets  $\mathcal{Q} := \{(u, v) \in (B_i(b)^c)^2 : u \neq v\}$  and  $\mathcal{Q} := \mathcal{Q} \cap (S_1^+(x))^2$  and the family  $(Z_q)_{q \in \mathcal{Q}}$  defined by  $Z_{(u,v)} := (H^{(\mathcal{T}_i \cup \{x\})} - z_i)_{uv}^{-1}$ . We first note that, for any  $q = (u, v) \in \mathcal{Q}$ , we have

$$\mathbb{P}(q \in \mathcal{Q} | \mathcal{F}_i(\{x\})) = \mathbb{P}(u \in S_1^+(x), v \in S_1^+(x) | \mathcal{F}_i) = \frac{d^2}{N^2}. \tag{6.21}$$

<sup>12</sup> Somewhat more carefully (since the set  $X$  is random), the precise definition of  $\mathcal{F}_i(X)$  is as follows. As the underlying probability space is finite, any  $\sigma$ -algebra, in particular  $\mathcal{F}_i$ , is atomic. Conditioning on  $\mathcal{F}_i$  means that we restrict ourselves to a single atom  $\mathcal{A}$  of  $\mathcal{F}_i$ . On this atom  $S_i(b)$  and  $X$  are deterministic. Then  $\mathcal{F}_i(X)$  is by definition the smallest  $\sigma$ -algebra on  $\mathcal{A}$  such that  $S_1^+(y)$  is measurable for all  $y \in S_i(b) \setminus X$ .

In the last step we used that, by Lemma 6.4, for any  $X \subset S_i(b)$ , conditionally on  $\mathcal{F}_i(X)$ , the random variables  $(A_{xy} : x \in X, y \in B_i(b)^c)$  are independent Bernoulli- $\frac{d}{N}$  random variables. Moreover, we get

$$\sum_{q \in \mathcal{Q}} |Z_q|^2 \mathbb{1}_{\Delta_i(\{x\})} \leq \text{Tr}((H^{(\mathcal{T}_i \cup \{x\})} - z_i)^{-2}) \mathbb{1}_{\Delta_i(\{x\})} \lesssim N\kappa^{-2}. \quad (6.22)$$

Since  $\mathcal{E}_i^{(0)} = \frac{1}{d} \sum_{q \in \mathcal{Q}} Z_q$ , the inclusion  $\Omega \subset \Gamma_i \cap \Delta_i(\{x\})$  and the Cauchy-Schwarz inequality imply

$$\begin{aligned} & \mathbb{E}_\Omega[|\mathcal{E}_i^{(0)}|^2 | \mathcal{F}_i(\{x\})] \\ & \leq \frac{1}{d^2} \mathbb{E}_\Omega \left[ |\mathcal{Q}| \sum_{q \in \mathcal{Q}} |Z_q|^2 \middle| \mathcal{F}_i(\{x\}) \right] \\ & \lesssim \frac{(\log N)^2}{d^2} \mathbb{E} \left[ \sum_{q \in \mathcal{Q}} |Z_q|^2 \mathbb{1}_{\Delta_i(\{x\})} \middle| \mathcal{F}_i(\{x\}) \right] \\ & \lesssim \frac{(\log N)^2}{d^2} \max_{q \in \mathcal{Q}} \mathbb{P}(q \in \mathcal{Q} | \mathcal{F}_i(\{x\})) \sum_{q \in \mathcal{Q}} |Z_q|^2 \mathbb{1}_{\Delta_i(\{x\})} \lesssim \frac{(\log N)^2}{\kappa^2 N}, \end{aligned} \quad (6.23)$$

where we used that  $|S_1(x)| \lesssim \log N$  on  $\Omega$ , by Proposition 3.2 (i), as well as the  $\mathcal{F}_i(\{x\})$ -measurability of  $|Z_q|^2 \mathbb{1}_{\Delta_i(\{x\})}$ . The last step follows from (6.21) and (6.22).

To bound  $\mathcal{E}_i^{(j)}$  for a fixed  $j \in \{1, \dots, |S_i(b)| - 1\}$ , we conclude on the event  $\Gamma_i$  from the resolvent identity that

$$\mathcal{E}_i^{(j)} = \frac{1}{d^{3/2}} \sum_{u \in S_1^+(x)} (H^{(\mathcal{T}_i \cup \{y_0, \dots, y_{j-1}\})} - z_i)_{uy_j}^{-1} \sum_{v \in S_1^+(y_j)} (H^{(\mathcal{T}_i \cup \{y_0, \dots, y_j\})} - z_i)_{vu}^{-1}.$$

Therefore, by applying the Cauchy-Schwarz inequality twice and using (6.20) with  $X = \{y_0, \dots, y_j\}$ , we obtain

$$\begin{aligned} |\mathcal{E}_i^{(j)}|^2 \mathbb{1}_\Omega & \leq \mathbb{1}_\Omega \frac{|S_1^+(x)|}{d^3} \sum_{u \in S_1^+(x)} |(H^{(\mathcal{T}_i \cup \{y_0, \dots, y_{j-1}\})} - z_i)_{uy_j}^{-1}|^2 \\ & \quad \left| \sum_{v \in S_1^+(y_j)} (H^{(\mathcal{T}_i \cup \{y_0, \dots, y_j\})} - z_i)_{vu}^{-1} \right|^2 \\ & \lesssim \mathbb{1}_\Omega \frac{|S_1^+(x)| |S_1^+(y_j)|}{d^3 \kappa^2} \sum_{u \in S_1^+(x)} |(H^{(\mathcal{T}_i \cup \{y_0, \dots, y_{j-1}\})} - z_i)_{uy_j}^{-1}|^2 \\ & \lesssim \mathbb{1}_\Omega \frac{(\log N)^2}{d^3 \kappa^2} \sum_{u \in S_1^+(x)} |(H^{(\mathcal{T}_i \cup \{y_0, \dots, y_{j-1}\})} - z_i)_{uy_j}^{-1}|^2, \end{aligned} \quad (6.24)$$

where in the last step we used  $|S_1(x)| + |S_1(y_j)| \lesssim \log N$  from Proposition 3.2 (i) on  $\Omega$ .

We now set  $Y_u := (H^{\mathcal{T}_i \cup \{y_0, \dots, y_{j-1}\}} - z_i)_{uy_j}^{-1}$  for  $u \in B_i(b)^c$ ,  $U := S_1^+(x)$  and  $\mathcal{U} := B_i(b)^c$ . We apply  $\mathbb{E}[\cdot | \mathcal{F}_i(\{y_0, \dots, y_{j-1}\})]$  to (6.24) and, similarly as in (6.23), obtain

$$\begin{aligned} & \mathbb{E}_\Omega[|\mathcal{E}_i^{(j)}|^2 | \mathcal{F}_i(\{y_0, \dots, y_{j-1}\})] \\ & \lesssim \frac{(\log N)^2}{d^3 \kappa^2} \max_{u \in \mathcal{U}} \mathbb{P}(u \in U | \mathcal{F}_i(\{y_0, \dots, y_{j-1}\})) \sum_{u \in \mathcal{U}} |Y_u|^2 \mathbb{1}_{\Delta_i(\{y_0, \dots, y_{j-1}\})} \\ & \lesssim \frac{(\log N)^2}{d^2 \kappa^4 N}, \end{aligned} \tag{6.25}$$

where we used that  $(Y_u)_{u \in \mathcal{U}}$  is  $\mathcal{F}_i(\{y_0, \dots, y_{j-1}\})$ -measurable, and that  $\sum_{u \in \mathcal{U}} |Y_u|^2 \mathbb{1}_{\Delta_i(\{y_0, \dots, y_{j-1}\})} = \|(Y_u)_{u \in B_i(b)^c}\|^2 \mathbb{1}_{\Delta_i(\{y_0, \dots, y_{j-1}\})} \lesssim \kappa^{-2}$ . We also used the remark following (6.21).

Finally, using the estimates (6.23) and (6.25) in (6.19) together with the tower property of the conditional expectation complete the proof of (6.9) and (6.10). This concludes the proof of (i).

Next, we prove (ii). For the proof of (6.11), we fix  $x, y \in S_1(b)$  and conclude from the resolvent identity [23, eq. (3.5)] that

$$\begin{aligned} (H^{(\mathcal{V})} - z)_{xy}^{-1} &= -(H^{(\mathcal{V})} - z)_{xx}^{-1} \sum_{u \notin \mathcal{V} \cup \{x\}} H_{xu} (H^{(\mathcal{V} \cup \{x\})} - z)_{uy}^{-1} \\ &= -\frac{1}{\sqrt{d}} (H^{(\mathcal{V})} - z)_{xx}^{-1} \sum_{u \notin \mathcal{V} \cup \{x\}} \mathbb{1}_{u \in S_1^+(x)} (H^{(\mathcal{V} \cup \{x\})} - z)_{uy}^{-1}. \end{aligned}$$

Therefore, the Cauchy-Schwarz inequality and Lemma 3.8 imply

$$|(H^{(\mathcal{V})} - z)_{xy}^{-1}|^2 \mathbb{1}_\Omega \lesssim \mathbb{1}_\Omega \frac{|S_1^+(x)|}{\kappa^2 d} \sum_{u \notin \mathcal{T}_1 \cup \{x\}} \mathbb{1}_{u \in S_1^+(x)} |(H^{\mathcal{T}_1 \cup \{x\}} - z)_{uy}^{-1}|^2,$$

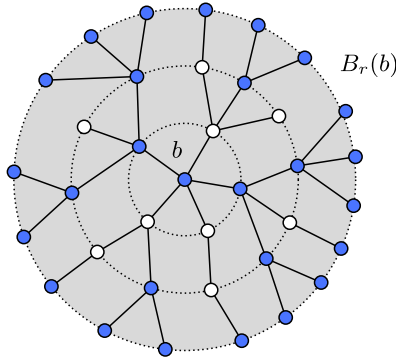
where we also used that  $\mathcal{V} = \{b\} \cup \mathcal{V}^{(B_1(b))} = \mathcal{T}_1$  on  $\Omega$  by Proposition 3.2 (iii). Hence, since  $|S_1(x)| \lesssim \log N$  on  $\Omega$  by Proposition 3.2 (i),  $\Omega \subset \Delta_1(\{x\})$  and  $\Delta_1(\{x\}) \in \mathcal{F}_1(\{x\})$ , we obtain

$$\begin{aligned} & \mathbb{E}[|(H^{(\mathcal{V})} - z)_{xy}^{-1}|^2 \mathbb{1}_\Omega | \mathcal{F}_1(\{x\})] \\ & \lesssim \frac{\log N}{\kappa^2 d} \max_{u \notin \mathcal{T}_1 \cup \{x\}} \mathbb{P}(u \in S_1^+(x) | \mathcal{F}_1(\{x\})) \sum_{u \notin \mathcal{T}_1 \cup \{x\}} |(H^{\mathcal{T}_1 \cup \{x\}} - z)_{uy}^{-1}|^2 \mathbb{1}_{\Delta_1(\{x\})} \\ & \lesssim \frac{\log N}{\kappa^4 N}. \end{aligned}$$

Here, in the last step, we used  $\mathbb{P}(u \in S_1^+(x) | \mathcal{F}_1(\{x\})) \leq d/N$  and  $\sum_{u \notin \mathcal{T}_1 \cup \{x\}} |(H^{\mathcal{T}_1 \cup \{x\}} - z)_{uy}^{-1}|^2 \mathbb{1}_{\Delta_1(\{x\})} \lesssim \kappa^{-2}$ . Thus, Chebyshev's inequality and the tower property of the conditional expectation complete the proof of (6.11) and, therefore, the one of Lemma 6.20.

□





**Fig. 10.** An illustration of the set  $\mathcal{R} \equiv \mathcal{R}_r(\mathbb{G}, b)$  of robust vertices. Here  $r = 3$  and  $d = 4$ . We draw the ball  $B_r(b)$  around the root  $b$ . The robust vertices are drawn in blue and the non-robust vertices in white. In this example, the root  $b$  is robust

6.5. *Proof of Proposition 6.12.* This subsection is devoted to the proof of Proposition 6.12. We start by introducing the notion of a robust vertex.

**Definition 6.21.** Let  $b \in [N]$  and  $r \in \mathbb{N}^*$ . We call a vertex  $y \in B_r(b)$  *robust* if

- (a)  $y \in S_r(b)$  or
- (b)  $y \in B_{r-1}(b)$  and at least  $d/2$  vertices in  $S_1^+(y)$  are robust.

We denote by  $\mathcal{R} \equiv \mathcal{R}_r(\mathbb{G}, b) \subset B_r(b)$  the set of robust vertices.

Note that  $\mathcal{R}$  is an  $\mathcal{F}$ -measurable random set. See Fig. 10 for an illustration of Definition 6.21. The following result states that with high probability the root  $b$  is robust, conditioned on  $S_1(a)$  and  $S_1(b)$ .

**Proposition 6.22.** (The root is robust) *Suppose that  $\sqrt{\log N} \ll d \lesssim \log N$  and that  $r$  satisfies (3.2). Then  $\mathbb{P}_\Omega(b \notin \mathcal{R} \mid S_1(a), S_1(b)) \lesssim N^{-1/2}$  whenever  $a, b \in \mathcal{W}$ .*

The proof of Proposition 6.22 is given at the end of this subsection. From now on, we choose  $r$  as in (6.3).

**Definition 6.23.** Let  $z \in \mathcal{J}$  be an  $\mathcal{F}$ -measurable real random variable. We introduce the event  $\Upsilon$  on which the following conditions hold.

- (A)  $\mathbb{G}|_{B_r(b)}$  is a tree.
- (B)  $b \in \mathcal{R}$ .
- (C)  $-G_{yy}(r, z) \geq (3z)^{-1}$  for all  $y \in (B_r(b) \cup \mathcal{V}^{(B_r(b))})^c$ .
- (D)  $|B_r(b) \cup \mathcal{V}^{(B_r(b))}| \leq N^{1/2}$ .

**Lemma 6.24.** *We have  $\Upsilon \in \mathcal{F}$  and  $\mathbb{P}_\Omega(\Upsilon^c \mid S_1(a), S_1(b)) = O(N^{-1/2})$  whenever  $a, b \in \mathcal{W}$ .*

*Proof.* That  $\Upsilon \in \mathcal{F}$  follows from the Definitions 6.5 and 6.21. The estimate follows from Proposition 6.22 and the facts that on  $\Omega$  the conditions (A), (C), and (D) hold surely. That (C) holds surely on  $\Omega$  follows from Lemma 3.9 and the observation that  $\mathcal{V} = \mathcal{V}^{(B_r(b))}$  on  $\Omega$  (see Proposition 3.2 (iii)). That (D) holds surely on  $\Omega$  follows from the statements (i), (i), (ii) of Proposition 3.2 (recall the choice (6.3)).  $\square$

We recall the definition of Lévy’s concentration function  $Q(X, L)$  from (2.13).

*Remark 6.25.* The concentration function has the following obvious properties.

- (i) For any  $u > 0$  we have  $Q(uX, uL) = Q(X, L)$ . More generally, if  $f$  is a continuous bijection on  $\mathbb{R}$  such that  $f^{-1}$  is  $K$ -Lipschitz, then  $Q(f(X), L) \leq Q(X, KL)$ .
- (ii) If  $X$  and  $Y$  are independent then  $Q(X + Y, L) \leq \min\{Q(X, L), Q(Y, L)\}$ .

Property (ii) is in general not sharp, and in some situations it can be improved considerably; see Proposition 2.8. Nevertheless, in some situations (ii) gives a better bound than Proposition 2.8; this is due to the minimum in (ii) as opposed to the maximum in Proposition 2.8. An important theme in the proof of Proposition 6.12 is a judicious mix of (ii) and Proposition 2.8. How to do this mix is encoded by the set of robust vertices from Definition 6.21.

We denote by  $Q^{\mathcal{F}}$  Lévy’s concentration function with respect to the probability measure  $\mathbb{P}(\cdot | \mathcal{F})$ . The main tool behind the proof of Proposition 6.12 is the following anticoncentration estimate for  $g_x(z)$ .

**Proposition 6.26.** *Let  $z \in \mathcal{J}$  be a  $\mathcal{F}$ -measurable real random variable. There exists a constant  $\chi > 0$  such that, on the event  $\Upsilon$ , for any  $1 \leq i \leq r$  and  $x \in S_i(b) \cap \mathcal{R}$  we have*

$$Q^{\mathcal{F}}\left(g_x(z), \frac{1}{8z(T^2d)^{r-i+1}}\right) \leq \frac{1}{2(\chi d)^{(r-i)/2}}.$$

Before proving Proposition 6.26, we use it to conclude the proof of Proposition 6.12.

*Proof of Proposition 6.12.* We estimate

$$\begin{aligned} & \mathbb{P}_{\Omega}\left(a, b \in \mathcal{W}, \left|\frac{1}{d} \sum_{x \in S_1(b)} g_x(z) + z\right| \leq N^{-\eta}\right) \\ & \leq \mathbb{P}_{\Omega}\left(\Xi_r \cap \Upsilon \cap \left\{\left|\frac{1}{d} \sum_{x \in S_1(b)} g_x(z) + z\right| \leq N^{-\eta}\right\}\right) + \mathbb{P}_{\Omega}(\Xi \cap \Upsilon^c) \\ & = \mathbb{E}\left[\mathbb{1}_{\Xi_r} \mathbb{1}_{\Upsilon} \mathbb{P}\left(\left|\frac{1}{d} \sum_{x \in S_1(b)} g_x(z) + z\right| \leq N^{-\eta} \mid \mathcal{F}\right)\right] + \mathbb{P}_{\Omega}(\Xi \cap \Upsilon^c), \end{aligned} \tag{6.26}$$

where we used Remark 6.17 (i) and (iii) as well as  $\Upsilon \in \mathcal{F}$  by Lemma 6.24. The second term on the right-hand side of (6.26) is estimated as

$$\mathbb{P}_{\Omega}(\Xi \cap \Upsilon^c) = \mathbb{E}[\mathbb{1}_{a \in \mathcal{W}} \mathbb{1}_{b \in \mathcal{W}} \mathbb{P}_{\Omega}(\Upsilon^c | S_1(a), S_1(b))] \lesssim N^{-1/2} \mathbb{P}(a, b \in \mathcal{W}),$$

by Lemma 6.24.

To estimate the first term on the right-hand side of (6.26), we use Proposition 6.26 with  $i = 1$  and the estimate

$$8zd(T^2d)^r \leq d^{2r+3/2} \leq N^{\eta} \tag{6.27}$$

where we used the definitions of  $r$  and  $T$  from (6.3) and Definition 6.6, as well as  $d \gg \sqrt{\log N}$ . On  $\Upsilon$  we have  $b \in \mathcal{R}$  and hence, by Definition 6.21, there exists  $x_* \in S_1(b) \cap \mathcal{R}$ .

Using that  $(g_x(z))_{x \in S_1(b)}$  is an independent family, on the event  $\Upsilon$  conditionally on  $\mathcal{F}$ , this yields, on the event  $\Upsilon$ ,

$$\begin{aligned} & \mathbb{P}\left(\left|\frac{1}{d} \sum_{x \in S_1(b)} g_x(z) + z\right| \leq N^{-\eta} \mid \mathcal{F}\right) \\ & \leq Q^{\mathcal{F}}\left(\frac{1}{d} \sum_{x \in S_1(b)} g_x(z), \frac{1}{8zd(T^2d)^r}\right) \\ & \leq Q^{\mathcal{F}}\left(\frac{1}{d} g_{x_*}(z), \frac{1}{8zd(T^2d)^r}\right) = Q^{\mathcal{F}}\left(g_{x_*}(z), \frac{1}{8z(T^2d)^r}\right), \end{aligned}$$

where in the second step we used Remark 6.25 (ii) and in the last step Remark 6.25 (i). From Proposition 6.26 we therefore conclude that the first term on the right-hand side of (6.26) is bounded by

$$\frac{1}{2(\chi d)^{(r-i)/2}} \mathbb{P}(\Xi_r) \leq N^{-\eta/4+o(1)} \mathbb{P}(a, b \in \mathcal{W}),$$

where we used the definition 6.3 and Remark 6.17 (ii). The claim now follows from Lemma 6.19. □

*Remark 6.27.* If we restrict ourselves to the critical regime  $d \asymp \log N$ , then the factor  $N^{-\eta}$  inside the probability in (6.6) can be improved to  $N^{-\eta/2}$ . To see this, we note that in this regime the parameter  $T$  from Definition 6.6 satisfies  $T = 10\kappa^{-1}$  since, in the critical regime, the estimate (6.16) with small enough  $\kappa$  remains valid for this smaller choice of  $T$ . Thus, the estimate (6.27) in the proof of Proposition 6.12 can be replaced with  $8zd(T^2d)^r \leq (Cd)^r$ , which is bounded by  $N^{\eta/2+o(1)}$ .

The key tool behind the proof of Proposition 6.26 is Proposition 2.8 due to Kesten.

*Proof of Proposition 6.26.* Throughout the proof, the argument  $z$  of  $g_x(z)$  for any  $x \in B_r(b)$  will always be the random variable  $z$  from Definition 6.23. Therefore, we omit this argument from our notation and write  $g_x \equiv g_x(z)$ .

For  $i \in [r]$  we define

$$L_i := \frac{1}{8zT^2d} \left(\frac{1}{T^2d}\right)^{r-i}, \quad P_i := \frac{1}{2} \left(\frac{\sqrt{2}K}{d^{1/2}}\right)^{r-i}.$$

where  $K$  is the universal constant from Proposition 2.8. We prove Proposition 6.26 by showing that, for all  $i \in [r]$  and  $x \in S_i(b) \cap \mathcal{R}$  we have

$$Q^{\mathcal{F}}(g_x, L_i) \leq P_i. \tag{6.28}$$

We show (6.28) by induction on  $i = r, r - 1, \dots, 1$ .

We start the induction at  $i = r$ . Abbreviate  $\mathcal{X} := (B_r(b) \cup \mathcal{V}^{(B_r(b))c})^c$ , which is an  $\mathcal{F}$ -measurable set. For  $x \in S_r(b)$ , conditioned on  $\mathcal{F}$ ,  $(\mathbb{1}_{y \in S_1^+(x)})_{y \in \mathcal{X}}$  are i.i.d. Bernoulli random variables. Hence, conditioned on  $\mathcal{F}$  we have

$$\sum_{y \in S_1^+(x) \setminus \mathcal{V}^{(B_r(b))}} G_{yy}(r, z) \stackrel{d}{=} \sum_{k=0}^{|\mathcal{X}|} \mathbb{1}_{|S_1^+(x) \cap \mathcal{X}|=k} \sum_{i=1}^k G_{\sigma(i)\sigma(i)}(r, z), \tag{6.29}$$

where  $\sigma$  is a uniform random enumeration of  $\mathcal{X}$  (i.e. a bijection  $[|\mathcal{X}|] \rightarrow \mathcal{X}$ ) that is independent of  $|S_1^+(x) \cap \mathcal{X}|$ . Because of the condition (C) in Definition 6.23, for any  $k \neq l$ ,  $|\sum_{i=1}^l G_{\sigma(i)\sigma(i)}(r, z) - \sum_{i=1}^k G_{\sigma(i)\sigma(i)}(r, z)| \geq (3z)^{-1}$ . Therefore, for any  $t \in \mathbb{R}$  we get on  $\Upsilon$

$$\begin{aligned} & \mathbb{P}\left(\left|\sum_{k=0}^{|\mathcal{X}|} \mathbb{1}_{|S_1^+(x) \cap \mathcal{X}|=k} \sum_{i=1}^k G_{\sigma(i)\sigma(i)}(r, z) - t\right| \leq \frac{1}{8z} \mid \sigma, \mathcal{F}\right) \\ & \leq \max_{0 \leq k \leq |\mathcal{X}|} \mathbb{P}(|S_1^+(x) \cap \mathcal{X}| = k \mid \mathcal{F}) \leq \frac{1}{2}, \end{aligned}$$

where in the last step we used that  $|S_1^+(x) \cap \mathcal{X}| \stackrel{d}{=} \text{Binom}(|\mathcal{X}|, d/N)$  conditioned on  $\mathcal{F}$ , that  $|\mathcal{X}| \geq N - N^{1/2}$  by the condition (D) in Definition 6.23, and that  $d \gg 1$ . From (6.29) we therefore conclude that

$$Q^{\mathcal{F}}\left(\sum_{y \in S_1^+(x) \setminus \mathcal{V}^{(B_r(b))}} G_{yy}(r, z), \frac{1}{8z}\right) \leq \frac{1}{2}.$$

Hence, Remarks 6.7 and 6.25 (i) imply  $Q^{\mathcal{F}}\left(g_x, \frac{1}{8zT^2d}\right) \leq \frac{1}{2}$ , which is (6.28) for  $i = r$ .

For the induction step, we let  $i < r$ , choose  $x \in S_i(b) \cap \mathcal{R}$ , and assume that  $Q^{\mathcal{F}}(g_y, L_{i+1}) \leq P_{i+1}$  for all  $y \in S_{i+1}(b) \cap \mathcal{R}$ . Note that  $S_1^+(x)$  and  $\mathcal{R}$  are  $\mathcal{F}$ -measurable, and that the family  $(g_y)_{y \in S_1^+(x)}$  is independent on  $\Upsilon$  conditioned on  $\mathcal{F}$ , by Remark 6.10 and Definition 6.23 (A). Hence, we can apply Proposition 2.8 to the concentration function  $Q^{\mathcal{F}}$  to obtain

$$Q^{\mathcal{F}}\left(\sum_{y \in S_1^+(x) \cap \mathcal{R}} g_y, L_{i+1}\right) \leq \frac{K}{\sqrt{|S_1^+(x) \cap \mathcal{R}|}} P_{i+1} \leq \frac{K\sqrt{2}P_{i+1}}{d^{1/2}}, \tag{6.30}$$

where the last inequality follows from  $|S_1^+(x) \cap \mathcal{R}| \geq d/2$ , by Definition 6.21. Moreover, the conditional independence of the sums  $\sum_{y \in S_1^+(x) \cap \mathcal{R}} g_y$  and  $\sum_{y \in S_1^+(x) \setminus \mathcal{R}} g_y$  combined with Remark 6.25 (i) and (ii) yields

$$\begin{aligned} Q^{\mathcal{F}}\left(\frac{1}{d} \sum_{y \in S_1^+(x)} g_y, \frac{L_{i+1}}{d}\right) &= Q^{\mathcal{F}}\left(\sum_{y \in S_1^+(x)} g_y, L_{i+1}\right) \\ &= Q^{\mathcal{F}}\left(\sum_{y \in S_1^+(x) \cap \mathcal{R}} g_y + \sum_{y \in S_1^+(x) \setminus \mathcal{R}} g_y, L_{i+1}\right) \\ &\leq Q^{\mathcal{F}}\left(\sum_{y \in S_1^+(x) \cap \mathcal{R}} g_y, L_{i+1}\right). \end{aligned}$$

Hence, by Remark 6.7, Remark 6.25 (i), Definition 6.9, and (6.30), we obtain

$$Q^{\mathcal{F}}\left(g_x, \frac{L_{i+1}}{T^2d}\right) \leq \frac{K\sqrt{2}P_{i+1}}{d^{1/2}},$$

which is (6.28). This completes the proof of (6.28) and, hence, the one of Proposition 6.26.  $\square$

*Proof of Proposition 6.22.* The proof proceeds in two steps: first by establishing the claim for the root of a Galton-Watson branching process with Poisson offspring distribution with mean  $d$ , and then concluding by a comparison argument.

Denote by  $\mathcal{P}_s$  a Poisson random variable with expectation  $s$ . Let  $W$  denote the Galton-Watson branching process with Poisson offspring distribution  $\mathcal{P}_d$ , which we regard as a random rooted ordered tree<sup>13</sup> whose root we call  $o$ . We use the graph-theoretic notations (such as  $S_i(x)$ ) from Sect. 2.1 also on rooted ordered trees. Moreover, we extend Definition 6.21 to a rooted ordered tree  $T$  in the obvious fashion, and when needed we use the notation  $\mathcal{R} \equiv \mathcal{R}_r(T, o)$  to indicate the radius  $r$ , the tree  $T$ , and the root  $o$  explicitly.

We define the parameter  $\delta := \mathbb{P}(\mathcal{P}_{3d/4} \leq \frac{d}{2})$ . By Bennett’s inequality (see Lemma D.1 below), we find that  $\delta \leq e^{-cd}$  for some universal constant  $c > 0$ . We shall show by induction on  $i$  that

$$\mathbb{P}(o \notin \mathcal{R}_i(W, o)) \leq \delta \tag{6.31}$$

for all  $i \geq 0$ . For  $i = 0$  we have  $\mathbb{P}(o \notin \mathcal{R}_i(W, o)) = 0$  since  $o \in \mathcal{R}_0(W, o)$  by (the analogue of) Definition 6.21, and (6.31) is trivial.

To advance the induction, we suppose that (6.31) holds for some  $i \geq 0$ . By Definition 6.21,

$$\mathbb{P}(o \notin \mathcal{R}_{i+1}(W, o)) = \mathbb{P}\left(\sum_{x \in S_1(o)} \mathbb{1}_{x \in \mathcal{R}_{i+1}(W, o)} < \frac{d}{2}\right).$$

By definition of the branching process  $W$ , conditioned on  $S_1(o)$ , the random variables  $(\mathbb{1}_{x \in \mathcal{R}_{i+1}(W, o)})_{x \in S_1(o)}$  are independent Bernoulli random variables with expectation

$$\mathbb{P}(x \in \mathcal{R}_{i+1}(W, o) | S_1(o)) = \mathbb{P}(o \in \mathcal{R}_i(W, o)) =: 1 - \zeta_i,$$

where  $x \in S_1(o)$ . We conclude that  $\sum_{x \in S_1(o)} \mathbb{1}_{x \in \mathcal{R}_{i+1}(W, o)} \stackrel{d}{=} \mathcal{P}_{d(1-\zeta_i)}$ . Using the induction assumption  $\zeta_i \leq \delta$  from (6.31) and the bound  $\delta \leq e^{-cd} < 1/4$  for large enough  $d$ , we therefore conclude that

$$\mathbb{P}(o \notin \mathcal{R}_{i+1}(W)) = \mathbb{P}\left(\mathcal{P}_{d(1-\mathbb{P}(o \notin \mathcal{R}_i(W, o)))} < \frac{d}{2}\right) \leq \mathbb{P}\left(\mathcal{P}_{3d/4} \leq \frac{d}{2}\right) = \delta.$$

This concludes the proof of (6.31) for all  $i \geq 0$ .

Hence, denoting by  $\mathcal{B}_{n,p}$  a random variable with law Binom( $n, p$ ), we conclude that if  $|S_1(o)| \geq d$  then

$$\begin{aligned} & \mathbb{P}(o \notin \mathcal{R}_r(W, o) | S_1(o)) \\ &= \mathbb{P}\left(\sum_{x \in S_1(o)} \mathbb{1}_{x \in \mathcal{R}_{r-1}(W, o)} < \frac{d}{2} \mid S_1(o)\right) = \mathbb{P}\left(\mathcal{B}_{|S_1(o)|, 1-\zeta_{r-1}} < \frac{d}{2} \mid S_1(o)\right) \\ &\leq \mathbb{P}\left(\mathcal{B}_{d, 1-\delta} < \frac{d}{2}\right) = \mathbb{P}\left(\mathcal{B}_{d, \delta} \geq \frac{d}{2}\right) \leq e^{-cd\delta \frac{1}{2\delta} \log \frac{1}{2\delta}} \leq e^{-cd^2} \leq N^{-1} \end{aligned} \tag{6.32}$$

for some universal constant  $c > 0$ , where in the third step we used that  $|S_1(o)| \geq d$  and  $\zeta_{r-1} \leq \delta$  by (6.31), in the fifth step Bennett’s inequality (see Lemma D.1 below),

<sup>13</sup> A rooted ordered tree (also called plane tree) is a rooted tree in which an ordering is specified among the children of each vertex.

in the sixth step that  $\delta \leq e^{-cd}$ , and in the last step the assumption  $d \gg \sqrt{\log N}$ . This concludes the estimate for the Galton-Watson process  $W$ .

Next, we analyse  $\mathbb{P}_\Omega(b \notin \mathcal{R}_r(\mathbb{G}, b) \mid S_1(a), S_1(b))$ . We note first that we can assume that  $|B_1(a)| \leq 1 + 10 \log N$  and that  $B_1(a)$  and  $B_1(b)$  are disjoint, for otherwise the above probability vanishes by definition of  $\Omega$  and Proposition 3.2.

We observe that a rooted ordered tree can be regarded as an equivalence class of (labelled) rooted trees up to a relabelling of the vertices that preserves the ordering of the children of each vertex. We denote by  $[\mathbb{T}, x]$  the equivalence class of the labelled rooted tree  $(\mathbb{T}, x)$ , where  $x$  is the root. By convention, if  $\mathbb{T}$  is not a tree (i.e. if it contains a cycle) then its equivalence class is the empty tree. We denote by  $\mathfrak{T}_r$  the set of rooted ordered trees of depth  $r$ . Moreover, we denote by  $\mathfrak{T}_r^* \subset \mathfrak{T}_r$  the subset of rooted ordered trees with at most  $N^{1/5}$  vertices and whose root is a robust vertex with  $|S_1(b)|$  children. Abbreviating  $\Delta := \{B_r(b) \subset [N] \setminus B_1(a)\}$ , we can write

$$\begin{aligned} \mathbb{P}_\Omega(b \notin \mathcal{R}_r(\mathbb{G}, b) \mid S_1(a), S_1(b)) &= \mathbb{P}_\Omega([\mathbb{G}|_{B_r(b)}, b] \notin \mathfrak{T}_r^* \mid S_1(a), S_1(b)) \\ &\leq \mathbb{P}(\{[\mathbb{G}|_{B_r(b)}, b] \notin \mathfrak{T}_r^*\} \cap \Delta \mid S_1(a), S_1(b)), \end{aligned} \tag{6.33}$$

where we used that, since  $a, b \in \mathcal{W} \subset \mathcal{V}$ , Proposition 3.2 implies that on the event  $\Omega$  the graph  $\mathbb{G}|_{B_r(b)}$  is a tree with at most  $N^{1/5}$  vertices, and that  $\Omega \subset \Delta$ .

For the following, let  $T \in \mathfrak{T}_r^*$  and denote by  $o$  its root. For  $1 \leq i \leq r$  we introduce the event

$$\Theta_i := \{[\mathbb{G}|_{B_i(b)}, b] = T|_{B_i(o)}\} \cap \{B_i(b) \subset [N] \setminus B_1(a)\}.$$

In particular,  $\mathbb{P}(\Theta_1 \mid S_1(a), S_1(b)) = 1$  because of the assumed disjointness of  $S_1(a)$  and  $S_1(b)$ . We now estimate

$$\mathbb{P}(\{[\mathbb{G}|_{B_r(b)}, b] = T\} \cap \Delta \mid S_1(a), S_1(b)) = \mathbb{P}(\Theta_r \mid S_1(a), S_1(b)) \tag{6.34}$$

recursively, for  $1 \leq i \leq r - 1$ , using the expression

$$\begin{aligned} &\frac{\mathbb{P}(\Theta_{i+1} \mid S_1(a), S_1(b))}{\mathbb{P}(\Theta_i \mid S_1(a), S_1(b))} \\ &= \frac{(N' - |B_i(o)|)!}{(N' - |B_{i+1}(o)|)! \prod_{x \in S_i(o)} |S_1^+(x)|!} \prod_{x \in S_i(o)} \left(\frac{d}{N}\right)^{|S_1^+(x)|} \left(1 - \frac{d}{N}\right)^{N - |B_i(o)| - |S_1^+(x)|}, \end{aligned}$$

where  $N' := N - |B_1(a)|$ , and the graph-theoretic quantities on the left-hand side are in terms of  $\mathbb{G}$  and on the right-hand side in terms of the deterministic rooted ordered tree  $T$ . Here, the multinomial factor in front arises from a choice of the  $|S_i(o)|$  disjoint subsets representing the children of the vertices in  $S_i(o)$  from  $N' - |B_i(o)|$  available vertices, and the remaining product follows by independence of the edges in  $\mathbb{G}$ . We deduce that

$$\begin{aligned} \frac{\mathbb{P}(\Theta_{i+1} \mid S_1(a), S_1(b))}{\mathbb{P}(\Theta_i \mid S_1(a), S_1(b))} &= (1 + O(N^{-3/4})) \frac{(N' - |B_i(o)|)!}{(N' - |B_{i+1}(o)|)! N^{|S_{i+1}(o)|}} \prod_{x \in S_i(o)} \frac{d^{|S_1^+(x)|}}{|S_1^+(x)|!} e^{-d} \\ &= (1 + O(N^{-3/5})) \prod_{x \in S_i(o)} \frac{d^{|S_1^+(x)|}}{|S_1^+(x)|!} e^{-d}, \end{aligned}$$

where we used that  $N - N' \leq 1 + 10 \log N$  and  $|B_r(o)| \leq N^{1/5}$ . By induction on  $i$  and comparison with the Galton–Watson tree  $W$ , using that

$$\frac{\mathbb{P}(W|_{B_{i+1}(o)} = T|_{B_{i+1}(o)} \mid S_1(o))}{\mathbb{P}(W|_{B_i(o)} = T|_{B_i(o)} \mid S_1(o))} = \prod_{x \in S_r(o)} \frac{d^{|S_1^+(x)|}}{|S_1^+(x)|!} e^{-d},$$

as well as  $\mathbb{P}(\Theta_1 \mid S_1(a), S_1(b)) = 1$ , we therefore conclude from (6.34) that if  $|S_1(b)| = |S_1(o)|$  then

$$\mathbb{P}(\{\mathbb{G}|_{B_r(b)}, b = T\} \cap \Delta \mid S_1(a), S_1(b)) = (1 + O(N^{-1/2})) \mathbb{P}(W|_{B_r(o)} = T \mid S_1(o))$$

for all  $T \in \mathfrak{T}_r^*$ . Thus,

$$\begin{aligned} & \mathbb{P}(\{\mathbb{G}|_{B_r(b)}, b \notin \mathfrak{T}_r^*\} \cap \Delta \mid S_1(a), S_1(b)) \\ & \leq 1 - \sum_{T \in \mathfrak{T}_r^*} \mathbb{P}(\{\mathbb{G}|_{B_r(b)}, b = T\} \cap \Delta \mid S_1(a), S_1(b)) \\ & = 1 - \sum_{T \in \mathfrak{T}_r^*} \mathbb{P}(W|_{B_r(o)} = T \mid S_1(o)) + O(N^{-1/2}) \\ & = \mathbb{P}(o \notin \mathcal{R}_r(W, o) \mid S_1(o)) + O(N^{-1/2}). \end{aligned}$$

The claim now follows from (6.32) and (6.33), noting that if  $b \in \mathcal{W}$  then  $|S_1(b)| = |S_1(o)| \geq d$ . □

**Acknowledgement** JA acknowledges funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 895698 as well as funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy - GZ 2047/1, project-id 390685813. RD acknowledges funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 884584). AK acknowledges funding from the European Research Council (ERC) and the Swiss State Secretariat for Education, Research and Innovation (SERI) through the consolidator grant ProbQuant. JA and AK acknowledge funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 715539), funding from the Swiss National Science Foundation through the NCCR SwissMAP grant, and support from the US National Science Foundation under Grant No. DMS-1928930 during their participation in the program *Universality and Integrability in Random Matrix Theory and Interacting Particle Systems* hosted by the Mathematical Sciences Research Institute in Berkeley, California during the Fall semester of 2021.

**Funding** Open access funding provided by University of Geneva.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### A. Quantitative Behaviour of $\alpha^*$

We recall the definition of  $\alpha^*(\mu)$  from (1.6) and remark that, besides  $\mu$ , it depends on  $N$  and  $d$ . Our analysis of  $\alpha^*(\mu)$  is based on quantitatively approximating the distribution of  $\alpha_1$  by a Poisson distribution. Owing to the Poisson approximation of binomial random variables (see e.g. [12, Lemma A.6]) and the Stirling approximation of factorials, we have

$$\mathbb{P}(d\alpha_1 = k) = \frac{d^k}{k!} e^{-d} \left( 1 + O\left(\frac{k^2}{N^2} + \frac{d^2}{N}\right) \right) = \exp(-f_d(k/d) + O(k^{-1})) \quad (\text{A.1})$$

for  $k \leq \sqrt{N}$ , where  $f_d(\alpha) := d(\alpha \log \alpha - \alpha + 1) + \frac{1}{2} \log(2\pi\alpha d)$  for  $\alpha > 0$ .

**Lemma A.1.** *If  $\mu \in [0, 1 - \varepsilon]$  for some constant  $\varepsilon \in (0, 1)$  then there is a constant  $T \geq 2$  such that if  $d \geq 1$  and  $t := \frac{\log N}{d} \geq T$  then*

$$\alpha^*(\mu) = \frac{(1 - \mu)t}{\log t} \left( 1 + O\left(\frac{\log \log t}{\log t}\right) \right). \quad (\text{A.2})$$

Lemma A.1 is proved in [15, Appendix A]. We use it to derive two simple consequences.

**Lemma A.2.** *Let  $1 \leq d \leq 3 \log N$  and  $\mu \in [0, 1 - \varepsilon]$  for some constant  $\varepsilon \in (0, 1)$ . If  $x \in \mathcal{V}$  then*

$$\alpha_x \gtrsim \frac{\log N}{d \log\left(\frac{10 \log N}{d}\right)}.$$

*Proof.* Let  $T$  be as in Lemma A.1. If  $\frac{\log N}{d} \geq T$  then Lemma A.2 follows directly from Lemma A.1 and the definition of  $\mathcal{V}$  in (2.1). If  $\frac{\log N}{d} < T \lesssim 1$  then  $\alpha_x \geq \alpha^*(\mu) \geq 2 + \kappa$  directly implies the lemma as  $\frac{\log N}{d \log\left(\frac{10 \log N}{d}\right)} \asymp 1$ .  $\square$

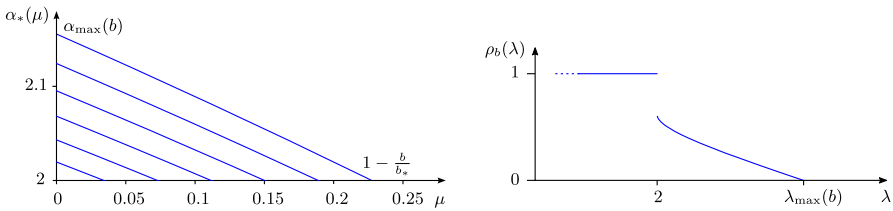
**Corollary A.3.** *Let  $\mu, \nu \in [0, 1 - \varepsilon]$  for some constant  $\varepsilon > 0$ . Then there is a constant  $T \equiv T(\varepsilon) \geq 2$  such that if  $t := \frac{\log N}{d} \geq T$  then*

$$\frac{\Lambda(\alpha^*(\nu))}{\Lambda(\alpha^*(\mu))} = \sqrt{\frac{1 - \nu}{1 - \mu}} \left( 1 + O\left(\frac{\log \log t}{\log t}\right) \right).$$

*Proof.* *Proof.* Choose  $T$  as in Lemma A.1, which then yields that  $\frac{\alpha^*(\nu)}{\alpha^*(\mu)} = \frac{1 - \nu}{1 - \mu} \left( 1 + O\left(\frac{\log \log t}{\log t}\right) \right)$ . Hence, using  $\min\{\alpha^*(\nu), \alpha^*(\mu)\} \gtrsim \frac{t}{\log t}$ , we get

$$\frac{\Lambda(\alpha^*(\nu))}{\Lambda(\alpha^*(\mu))} = \sqrt{\frac{\alpha^*(\nu)}{\alpha^*(\mu)}} \left( 1 + O\left(\frac{1}{\alpha^*(\nu)} + \frac{1}{\alpha^*(\mu)}\right) \right) = \sqrt{\frac{1 - \nu}{1 - \mu}} \left( 1 + O\left(\frac{\log \log t}{\log t}\right) \right). \quad \square$$





**Fig. 11.** Left: an illustration of the function  $\mu \mapsto \alpha_*(\mu)$ , the asymptotic version of  $\alpha^*(\mu)$ . We plot it for the values  $b = 2.0, 2.1, \dots, 2.5$ , corresponding to the graphs from top to bottom. Each graph crosses the horizontal axis at  $\mu = 1 - \frac{b}{b_*}$  and the vertical axis at  $\alpha = \alpha_{\max}(b)$ . Right: the behaviour of the exponent  $\rho_b$  of the density of states from (B.4), as a function of the energy  $\lambda$ . 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 jump at 2 of  $\rho_b$  is from  $\rho_b(2^-) = 1$  to  $\rho_b(2^+) = 1 - b/b_* = 2 - 2 \log 2$

### B. Qualitative Behaviour of Degree Sequence and of $\alpha^*$

In this appendix we describe the qualitative behaviour of the normalized degree sequence  $(\alpha_x)_{x \in [N]}$  and apply it to  $\alpha^*(\mu)$  from (1.6). For definiteness, we focus on the critical regime, where  $d = b \log N$  for some constant  $b$ , and consider the limit  $N \rightarrow \infty$  with  $\kappa = o(1)$ . For detailed proofs, we refer to [12, Appendix A.4].

For  $b \geq 0$  and  $\alpha \geq 2$  define

$$\theta_b(\alpha) := 1 - b(\alpha \log \alpha - \alpha + 1). \tag{B.1}$$

For any  $b \leq b_*$ , it is easy to see that for any  $\mu \in [0, 1 - \frac{b}{b_*}]$  the equation  $\mu = \theta_b(\alpha)$  has a unique solution  $\alpha \geq 2$ , which we denote by  $\alpha_*(\mu)$ . By Poisson approximation from (A.1), we deduce that, with high probability, for any  $\alpha \geq 0$  we have

$$|\{x \in [N] : \alpha_x \geq \alpha\}| = N^{\theta_b(\alpha) + o(1)} + o(1). \tag{B.2}$$

Recalling the definition (1.6), we hence conclude that  $\alpha^*(\mu) = \alpha_*(\mu) + o(1)$ . In other words,  $\alpha_*(\mu)$  is the asymptotic value of  $\alpha^*(\mu)$ . We refer to Fig. 11 for an illustration of the function  $\alpha_*(\mu)$ .

Next, we define

$$\alpha_{\max}(b) := \alpha_*(0), \quad \lambda_{\max}(b) := \Lambda(\alpha_{\max}(b)), \tag{B.3}$$

which, by (B.2) and Remark 1.2, have the interpretation of the asymptotic largest normalized degree and largest nontrivial eigenvalue of  $H$ , respectively. Moreover, for  $b < b_*$  we define

$$\rho_b(\lambda) := \begin{cases} \theta_b(\Lambda^{-1}(\lambda))_+ & \text{if } |\lambda| \geq 2 \\ 1 & \text{if } |\lambda| < 2, \end{cases} \tag{B.4}$$

where  $\Lambda^{-1}(\lambda) = \frac{\lambda^2}{2}(1 + \sqrt{1 - 4/\lambda^2})$  for  $|\lambda| \geq 2$ . Then with high probability the density of states around energy  $\lambda \in \mathbb{R}$  equals  $N^{\rho_b(\lambda) + o(1)}$ . For  $\lambda > 2$ , this follows from Remark 1.2 and (B.2). The function  $\rho_b$  is illustrated in Fig. 11.

### C. Largest Eigenvalues of $H$ and Some Submatrices: Proof of Proposition 3.4

The key tool to prove Proposition 3.4 is the following estimate.

**Lemma C.1.** *Let  $d$  satisfy (1.5). With high probability, for all  $v \in [0, 1]$ , we have*

$$\|(H - \mathbb{E}H)^{(\{x : \alpha_x \geq \alpha^*(v)\})}\| \leq \Lambda(\alpha^*(v)) + o(1). \tag{C.1}$$

The proof of Lemma C.1 is analogous to [12, Proof of Proposition 3.12]. How to deduce Proposition 3.4 from Lemma C.1 is analogous to the corresponding arguments in [13, 16]. In both cases, the details are given in [15, Appendix C].

### D. Tools

In this appendix we summarize several well known results used throughout this paper.

*D.1. Bennett’s inequality.* Define the function  $h : [0, \infty) \rightarrow [0, \infty)$  by

$$h(a) := (1 + a) \log(1 + a) - a. \tag{D.1}$$

We use the notation  $\mathcal{P}_\mu$  to denote a Poisson random variable with parameter  $\mu \geq 0$ , and  $\mathcal{B}_{n,p}$  to denote a binomial random variable with parameters  $n \in \mathbb{N}^*$  and  $p \in [0, 1]$ . The following estimate is proved in [26, Section 2.7].

**Lemma D.1** (Bennett). *For  $0 \leq \mu \leq n$  and  $a > 0$  we have*

$$\mathbb{P}(\mathcal{B}_{n,\mu/n} - \mu \geq a\mu) \leq e^{-\mu h(a)}, \quad \mathbb{P}(\mathcal{B}_{n,\mu/n} - \mu \leq -a\mu) \leq e^{-\mu a^2/2} \leq e^{-\mu h(a)},$$

and  $\frac{a^2}{2(1+a/3)} \leq h(a) \leq \frac{a^2}{2}$ . By taking  $n \rightarrow \infty$ , the same estimates hold with  $\mathcal{B}_{n,\mu/n}$  replaced with  $\mathcal{P}_\mu$ .

*D.2. Perturbation theory.* The following lemma contains simple perturbation estimates for approximate eigenvalues and eigenvectors. Its proof can be found in [16, Lemma 4.10].

**Lemma D.2.** *Let  $M$  be a real symmetric matrix. Let  $\varepsilon, \Delta > 0$  satisfy  $5\varepsilon \leq \Delta$ . Suppose that  $M$  has a unique eigenvalue,  $\lambda$ , in  $[\widehat{\lambda} - \Delta, \widehat{\lambda} + \Delta]$  for some  $\widehat{\lambda} \in \mathbb{R}$ . Let  $\mathbf{w}$  be a corresponding normalized eigenvector of  $M$ . If there exists a normalized vector  $\mathbf{v}$  such that  $\|(M - \widehat{\lambda})\mathbf{v}\| \leq \varepsilon$  then, for some  $\sigma \in \{\pm 1\}$ ,*

$$\lambda - \widehat{\lambda} = \langle \mathbf{v}, (M - \widehat{\lambda})\mathbf{v} \rangle + O\left(\frac{\varepsilon^2}{\Delta}\right), \quad \|\mathbf{w} - \sigma\mathbf{v}\| = O\left(\frac{\varepsilon}{\Delta}\right).$$

*D.3. Eigenvalue interlacing.* For an  $N \times N$  Hermitian matrix  $M$ , we use the convention that  $\lambda_i(M) = -\infty$  for  $i > N$ .

**Lemma D.3** (Interlacing for minors). *Let  $X \subset Y \subset [N]$  and  $M$  be an  $N \times N$  Hermitian matrix. Then, for all  $i \in [N]$ ,*

$$\lambda_{i+|Y \setminus X|}(M^{(X)})_+ \leq \lambda_i(M^{(Y)})_+ \leq \lambda_i(M^{(X)})_+,$$

where  $\lambda_+ := \max\{\lambda, 0\}$ .

**Lemma D.4** (Interlacing for rank-one perturbations). *Let  $M$  be an  $N \times N$  Hermitian matrix and  $V$  a rank-one positive semidefinite  $N \times N$  matrix. Then, for all  $i \in [N]$ ,*

$$\lambda_{i+1}(M + V) \leq \lambda_i(M) \leq \lambda_i(M + V).$$

**D.4. Resolvent of adjacency matrix of  $\mathbb{N}^*$ .** Define the infinite tridiagonal matrix

$$M := \begin{pmatrix} 0 & 1 & 0 & & \\ 1 & 0 & 1 & \ddots & \\ 0 & 1 & 0 & \ddots & \\ & \ddots & \ddots & \ddots & \ddots \end{pmatrix}, \tag{D.2}$$

which we regard as a bounded operator on  $\ell^2(\mathbb{N}^*)$ . Note that  $M$  is the adjacency matrix of  $\mathbb{N}^*$  with the graph structure induced by regarding adjacent numbers as neighbours.

**Lemma D.5.** *Let  $t > 2$ . Then  $1 - t^{-1}M$  is invertible with convergent Neumann series and explicit inverse*

$$(1 - t^{-1}M)_{1j}^{-1} = t \left( \frac{2}{t + \sqrt{t^2 - 4}} \right)^j$$

for any  $j \in \mathbb{N}^*$ .

*Proof.* By the Schur test,  $\|M\| \leq 2$ . Therefore,  $t > 2$  implies the invertibility of  $1 - t^{-1}M$  in  $\ell^2(\mathbb{N}^*)$  and the convergence of the Neumann series representation of its inverse. In particular,  $a = (a_j)_{j \in \mathbb{N}^*} \in \ell^2(\mathbb{N}^*)$  with  $a_j := (1 - t^{-1}M)_{1j}^{-1}$  for any  $j \in \mathbb{N}^*$ . With  $e_1 := (\delta_{1j})_{j \in \mathbb{N}^*}$ , the definition of the resolvent yields  $(1 - t^{-1}M)a = e_1$  and, thus,

$$a_1 - t^{-1}a_2 = 1, \quad \begin{pmatrix} a_{j+1} \\ a_j \end{pmatrix} = T(t) \begin{pmatrix} a_j \\ a_{j-1} \end{pmatrix}, \quad T(t) := \begin{pmatrix} t & -1 \\ 1 & 0 \end{pmatrix} \tag{D.3}$$

for any  $j \in \mathbb{N}$  satisfying  $j \geq 2$ . The transfer matrix  $T(t)$  has the two eigenvalues,  $\gamma$  and  $\gamma^{-1}$ , where  $\gamma := \frac{2}{t + \sqrt{t^2 - 4}}$ . As  $t > 2$ , we have  $\gamma < 1$  and, hence,  $a \in \ell^2(\mathbb{N}^*)$  implies  $a_j = \gamma a_{j-1} = \gamma^{j-1} a_1$  for any  $j \geq 2$ . Together with the first relation in (D.3), we obtain  $a_1 = \frac{2t}{t + \sqrt{t^2 - 4}}$ , which completes the proof.  $\square$

*Proof of Proposition 3.2.* We show that each item holds individually with high probability. For the proof of (i), we estimate  $\mathbb{P}(|\mathcal{V}| \geq tN^\mu) \leq t^{-1}N^{-\mu} \mathbb{E}|\mathcal{V}| = t^{-1}N^{-\mu} N \mathbb{P}(\alpha_1 \geq \alpha^*) \leq t^{-1}$  for  $t > 0$  by the definitions of  $\mathcal{V}$  and  $\alpha^*$  in (2.1) and (1.6), respectively. This proves (i). Item (ii) is a consequence of [13, Lemma 5.5] with  $k = 1$ , Proposition 3.1 (i),  $r \leq (\frac{1}{3} - \frac{\mu}{4}) \frac{\log N}{\log d}$ , and  $d \geq \sqrt{\log N}$ . Item (iii) follows from [16, eq. (9.5)] with the choices  $\tau = \alpha^*$  and  $n = 2$ ,  $\exp(-dh(\alpha^* - 1 - \frac{3}{d})) \leq N^{\mu-1+o(1)}$  and  $r \leq (\frac{1}{3} - \mu) \frac{\log N}{\log d}$ . To prove (iv), we set  $\eta/2 := r \frac{\log d}{\log N}$  and conclude from (i), Proposition 3.1 (i) and (ii) that  $|\bigcup_{x \in \mathcal{V}} B_r(x)| \leq N^{\mu+\eta/2+o(1)}$ . Thus, we obtain (iv) by arguing similarly as in [16, Proof of Proposition 4.4], see especially [16, eq. (9.10)].  $\square$

## References

1. Abou-Chacra, R., Thouless, D.J.: Self-consistent theory of localization. II. Localization near the band edges. J. Phys. C: Solid State Phys. 7(1), 65 (1974)
2. Abou-Chacra, R., Thouless, D.J., Anderson, P.W.: A selfconsistent theory of localization. J. Phys. C: Solid State Phys. 6(10), 1734 (1973)

3. Abrahams, E. (ed.): 50 Years of Anderson Localization, vol. 24. World Scientific, Singapore (2010)
4. Aggarwal, A., Bordenave, C., Lopatto, P.: Mobility edge of Lévy matrices. Preprint [arXiv:2210.09458](https://arxiv.org/abs/2210.09458) (2022)
5. Aggarwal, A., Lopatto, P., Marcinek, J.: Eigenvector statistics of Lévy matrices. *Ann. Probab.* **49**(4), 1778–1846 (2021)
6. Aggarwal, A., Lopatto, P., Yau, H.-T.: GOE statistics for Lévy matrices. *J. Eur. Math. Soc.* **23**(11), 3707–3800 (2018)
7. Aizenman, M., Molchanov, S.: Localization at large disorder and at extreme energies: an elementary derivation. *Commun. Math. Phys.* **157**, 245–278 (1993)
8. Aizenman, M., Sims, R., Warzel, S.: Absolutely continuous spectra of quantum tree graphs with weak disorder. *Commun. Math. Phys.* **264**(2), 371–389 (2006)
9. Aizenman, M., Warzel, S.: Extended states in a Lifshitz tail regime for random Schrödinger operators on trees. *Phys. Rev. Lett.* **106**(13), 136804 (2011)
10. Aizenman, M., Warzel, S.: Resonant delocalization for random Schrödinger operators on tree graphs. *J. Eur. Math. Soc.* **15**(4), 1167–1222 (2013)
11. Aizenman, M., Warzel, S.: *Random Operators*. Graduate Studies in Mathematics. American Mathematical Society, Providence (2015)
12. Alt, J., Ducatez, R., Knowles, A.: Delocalization transition for critical Erdős–Rényi graphs. *Commun. Math. Phys.* **388**(1), 507–579 (2021)
13. Alt, J., Ducatez, R., Knowles, A.: Extremal eigenvalues of critical Erdős–Rényi graphs. *Ann. Probab.* **49**(3), 1347–1401 (2021)
14. Alt, J., Ducatez, R., Knowles, A.: The completely delocalized region of the Erdős–Rényi graph. *Electron. Commun. Probab.* **27**, Paper No. 10, 9 (2022)
15. Alt, J., Ducatez, R., Knowles, A.: Localized phase for the Erdős–Rényi graph, online version of the present article. Preprint [arXiv:2305.16294](https://arxiv.org/abs/2305.16294) (2023)
16. Alt, J., Ducatez, R., Knowles, A.: Poisson statistics and localization at the spectral edge of sparse Erdős–Rényi graphs. *Ann. Probab.* **51**(1), 277–358 (2023)
17. Anderson, P.W.: Absence of diffusion in certain random lattices. *Phys. Rev.* **109**(5), 1492 (1958)
18. Anderson, P.W.: Local moments and localized states. *Science* **201**(4353), 307–316 (1978)
19. Bauerschmidt, R., Huang, J., Knowles, A., Yau, H.-T.: Bulk eigenvalue statistics for random regular graphs. *Ann. Probab.* **45**(6A), 3626–3663 (2017)
20. Bauerschmidt, R., Huang, J., Knowles, A., Yau, H.-T.: Edge rigidity and universality of random regular graphs of intermediate degree. *Geom. Funct. Anal.* **30**(3), 693–769 (2020)
21. Bauerschmidt, R., Huang, J., Yau, H.-T.: Local Kesten–McKay law for random regular graphs. *Commun. Math. Phys.* **369**(2), 523–636 (2019)
22. Bauerschmidt, R., Knowles, A., Yau, H.-T.: Local semicircle law for random regular graphs. *Commun. Pure Appl. Math.* **70**(10), 1898–1960 (2017)
23. Benaych-Georges, F., Knowles, A.: Local semicircle law for Wigner matrices. In: *Advanced Topics in Random Matrices*, Panor. Synthèses, vol. 53, pp. 1–90. Soc. Math. France, Paris (2017)
24. Bordenave, C., Guionnet, A.: Localization and delocalization of eigenvectors for heavy-tailed random matrices. *Probab. Theor. Relat. Fields* **157**(3–4), 885–953 (2013)
25. Bordenave, C., Guionnet, A.: Delocalization at small energy for heavy-tailed random matrices. *Commun. Math. Phys.* **354**(1), 115–159 (2017)
26. Boucheron, S., Lugosi, G., Massart, P.: *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, Oxford (2013)
27. 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)
28. 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)
29. 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)
30. Casati, G., Molinari, L., Izrailev, F.: Scaling properties of band random matrices. *Phys. Rev. Lett.* **64**(16), 1851 (1990)
31. Chen, N., Smart, C.K.: Random band matrix localization by scalar fluctuations. Preprint [arXiv:2206.06439](https://arxiv.org/abs/2206.06439) (2022)
32. Chung, F., Lu, L.: The diameter of sparse random graphs. *Adv. Appl. Math.* **26**(4), 257–279 (2001)
33. Cipolloni, G., Peled, R., Schenker, J., Shapiro, J.: Dynamical localization for random band matrices up to  $w \ll n^{1/4}$ . Preprint [arXiv:2206.05545](https://arxiv.org/abs/2206.05545) (2022)
34. Cizeau, P., Bouchaud, J.-P.: Theory of Lévy matrices. *Phys. Rev. E* **50**(3), 1810 (1994)
35. Erdős, L., Knowles, A.: Quantum diffusion and delocalization for band matrices with general distribution. *Ann. Henri Poincaré* **12**, 1227–1319 (2011)

36. Erdős, L., Knowles, A.: Quantum diffusion and eigenfunction delocalization in a random band matrix model. *Commun. Math. Phys.* **303**, 509–554 (2011)
37. Erdős, L., Knowles, A.: The Altshuler–Shklovskii formulas for random band matrices II: the general case. *Ann. Henri Poincaré* **16**, 709–799 (2014)
38. Erdős, L., Knowles, A.: The Altshuler–Shklovskii formulas for random band matrices I: the unimodular case. *Commun. Math. Phys.* **333**, 1365–1416 (2015)
39. Erdős, L., Knowles, A., Yau, H.-T.: Averaging fluctuations in resolvents of random band matrices. *Ann. Henri Poincaré* **14**, 1837–1926 (2013)
40. Erdős, L., Knowles, A., Yau, H.-T., Yin, J.: Spectral statistics of Erdős–Rényi graphs II: eigenvalue spacing and the extreme eigenvalues. *Commun. Math. Phys.* **314**, 587–640 (2012)
41. Erdős, L., Knowles, A., Yau, H.-T., Yin, J.: Delocalization and diffusion profile for random band matrices. *Commun. Math. Phys.* **323**, 367–416 (2013)
42. Erdős, L., Knowles, A., Yau, H.-T., Yin, J.: Spectral statistics of Erdős–Rényi graphs I: Local semicircle law. *Ann. Probab.* **41**, 2279–2375 (2013)
43. Evers, F., Mirlin, A.D.: Anderson transitions. *Rev. Mod. Phys.* **80**(4), 1355 (2008)
44. 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)
45. 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)
46. Fyodorov, Y., Mirlin, A.: Scaling properties of localization in random band matrices: a  $\sigma$ -model approach. *Phys. Rev. Lett.* **67**(18), 2405 (1991)
47. Fyodorov, Y., Ossipov, A., Rodriguez, A.: The Anderson localization transition and eigenfunction multifractality in an ensemble of ultrametric random matrices. *J. Stat. Mech. Theor. Exp.* **2009**(12), L12001 (2009)
48. He, Y.: Spectral gap and edge universality of dense random regular graphs. Preprint [arXiv:2203.07317](https://arxiv.org/abs/2203.07317) (2022)
49. He, Y., Knowles, A.: Fluctuations of extreme eigenvalues of sparse Erdős–Rényi graphs. *Probab. Theor. Relat. Fields* **180**(3–4), 985–1056 (2021)
50. He, Y., Knowles, A., Marcozzi, M.: Local law and complete eigenvector delocalization for supercritical Erdős–Rényi graphs. *Ann. Probab.* **47**(5), 3278–3302 (2019)
51. He, Y., Marcozzi, M.: Diffusion profile for random band matrices: a short proof. *J. Stat. Phys.* **177**(4), 666–716 (2019)
52. Huang, J., Landon, B., Yau, H.-T.: Bulk universality of sparse random matrices. *J. Math. Phys.* **56**(12), 123301 (2015)
53. Huang, J., Landon, B., Yau, H.-T.: Transition from Tracy–Widom to Gaussian fluctuations of extremal eigenvalues of sparse Erdős–Rényi graphs. *Ann. Probab.* **48**(2), 916–962 (2020)
54. Huang, J., Yau, H.-T.: Spectrum of random  $d$ -regular graphs up to the edge. Preprint [arXiv:2102.00963](https://arxiv.org/abs/2102.00963) (2021)
55. Huang, J., Yau, H.-T.: Edge universality of sparse random matrices. Preprint [arXiv:2206.06580](https://arxiv.org/abs/2206.06580) (2022)
56. Huang, J., Yau, H.-T.: Edge universality of random regular graphs of growing degrees. Preprint [arXiv:2305.01428](https://arxiv.org/abs/2305.01428) (2023)
57. Kesten, H.: A sharper form of the Doeblin–Lévy–Kolmogorov–Rogozin inequality for concentration functions. *Math. Scand.* **25**, 133–144 (1969)
58. Klein, A.: Absolutely continuous spectrum in the Anderson model on the Bethe lattice. *Math. Res. Lett.* **1**(4), 399–407 (1994)
59. Krivelevich, M., Sudakov, B.: The largest eigenvalue of sparse random graphs. *Combin. Probab. Comput.* **12**(1), 61–72 (2003)
60. Legendijk, A., Van Tiggelen, B., Wiersma, D.S.: Fifty years of Anderson localization. *Phys. Today* **62**(8), 24–29 (2009)
61. Lee, J.: Higher order fluctuations of extremal eigenvalues of sparse random matrices. Preprint [arXiv:2108.11634](https://arxiv.org/abs/2108.11634) (2021)
62. Lee, J.O., Schnelli, K.: Local law and Tracy–Widom limit for sparse random matrices. *Probab. Theor. Relat. Fields* **171**(1), 543–616 (2018)
63. Lee, P.A., Ramakrishnan, T.V.: Disordered electronic systems. *Rev. Mod. Phys.* **57**(2), 287 (1985)
64. 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)
65. Peled, R., Schenker, J., Shamis, M., Sodin, S.: On the Wegner orbital model. *Int. Math. Res. Not.* **2019**(4), 1030–1058 (2019)
66. Schenker, J.: Eigenvector localization for random band matrices with power law band width. *Commun. Math. Phys.* **290**, 1065–1097 (2009)

67. Shcherbina, M., Shcherbina, T.: Universality for 1 d random band matrices. *Commun. Math. Phys.* **385**, 667–716 (2021)
68. Sodin, S.: The spectral edge of some random band matrices. *Ann. Math.* **172**(3), 2223–2251 (2010)
69. Tarquini, E., Biroli, G., Tarzia, M.: Level statistics and localization transitions of Lévy matrices. *Phys. Rev. Lett.* **116**(1), 010601 (2016)
70. Tarzia, M.: Fully localized and partially delocalized states in the tails of Erdős–Rényi graphs in the critical regime. *Phys. Rev. B* **105**(17), 174201 (2022)
71. Tikhomirov, K., Youssef, P.: Outliers in spectrum of sparse Wigner matrices. *Rand. Struct. Algorithms* **58**(3), 517–605 (2021)
72. von Soosten, P., Warzel, S.: The phase transition in the ultrametric ensemble and local stability of Dyson Brownian motion. *Electr. J. Probab.* **23**, 1–24 (2018)
73. Wegner, F.: Bounds on the density of states in disordered systems. *Z. Phys. B Condens. Matter* **44**(1), 9–15 (1981)
74. Wigner, E.P.: Characteristic vectors of bordered matrices with infinite dimensions. *Ann. Math.* **62**, 548–564 (1955)
75. Wigner, E.P.: On the distribution of the roots of certain symmetric matrices. *Ann. Math.* **67**, 325–327 (1958)
76. Xu, C., Yang, F., Yau, H.-T., Yin, J.: Bulk universality and quantum unique ergodicity for random band matrices in high dimensions. Preprint [arXiv:2207.14533](https://arxiv.org/abs/2207.14533) (2022)
77. Yang, F., Yau, H.-T., Yin, J.: Delocalization and quantum diffusion of random band matrices in high dimensions I: self-energy renormalization. Preprint [arXiv:2104.12048](https://arxiv.org/abs/2104.12048) (2021)
78. Yang, F., Yau, H.-T., Yin, J.: Delocalization and quantum diffusion of random band matrices in high dimensions II: T-expansion. *Commun. Math. Phys.* **25**, 1–96 (2022)
79. Yang, F., Yin, J.: Random band matrices in the delocalized phase, III: averaging fluctuations. *Probab. Theory Relat. Fields* **179**(1–2), 451–540 (2021)