Spectral radii of sparse random matrices

Florent Benaych-Georges   Charles Bordenave   Antti Knowles

September 19, 2019

We establish bounds on the spectral radii for a large class of sparse random matrices, which includes the adjacency matrices of inhomogeneous Erdős-Rényi graphs. Our error bounds are sharp for a large class of sparse random matrices. In particular, for the Erdős-Rényi graph $G(n, d/n)$, our results imply that the smallest and second-largest eigenvalues of the adjacency matrix converge to the edges of the support of the asymptotic eigenvalue distribution provided that $d/\log n \to \infty$. Together with the companion paper [3], where we analyse the extreme eigenvalues in the complementary regime $d/\log n \to 0$, this establishes a crossover in the behaviour of the extreme eigenvalues at $d \asymp \log n$. Our results also apply to non-Hermitian sparse random matrices, corresponding to adjacency matrices of directed graphs. The proof combines (i) a new inequality between the spectral radius of a matrix and the spectral radius of its nonbacktracking version together with (ii) a new application of the method of moments for nonbacktracking matrices.

1. Introduction

The goal of the present paper is to obtain bounds on the spectral radius of a sparse random matrix. Sparse random matrices arise naturally as adjacency matrices of random graphs. In spectral graph theory, obtaining precise bounds on the locations of the extreme eigenvalues, in particular on the spectral gap, is of fundamental importance and has attracted much attention in the past thirty years. See for instance [11, 16, 1] for reviews.

The problem of estimating the spectral radius of a random matrix has a long history, starting with the seminal work of Füredi and Komlós [15], subsequently improved by Vu [25]. For the simple case of the Erdős-Rényi graph $G(n, d/n)$, where each edge of the complete graph on n vertices is kept with probability d/n independently of the others, it is shown in [15, 25] that the smallest and second-largest eigenvalues of the adjacency matrix converge to the edges of the support of the asymptotic eigenvalue distribution provided that $d/(\log n)^4 \to \infty$. In this paper we derive quantitative bounds for the extreme eigenvalues, which are sharp for a large class of sparse random matrices, including $G(n, d/n)$ for small enough expected degree d. An immediate corollary of these bounds is that the extreme eigenvalues converge to the edges of the support of the asymptotic eigenvalue distribution provided that $d/\log n \to \infty$. Together with the companion paper [3], where we analyse the extreme eigenvalues in the complementary regime $d/\log n \to 0$, this establishes a crossover in the behaviour of the extreme eigenvalues of the Erdős-Rényi graph. The location $d \asymp \log n$ of this crossover for the extreme eigenvalues parallels the well-known crossover from connected to disconnected graphs.

Our results hold for a very general class of sparse random matrices with independent mean-zero entries. In particular, we also obtain bounds for the extreme eigenvalues of inhomogeneous Erdős-Rényi graphs and stochastic block models. They hold for Hermitian random matrices, corresponding to adjacency matrices of undirected random graphs, as well as for non-Hermitian random matrices with independent entries, corresponding to adjacency matrices of directed random graphs.

In their seminal work, Füredi and Komlós [15] bound the spectral radius of a random matrix by estimating the expected trace of high powers of this matrix. The technical challenge of this strategy is to derive a sharp upper bound on the number of simple walks with given combinatorial properties. The novelty of our approach lies in a new use of the nonbacktracking matrix. Nonbacktracking matrices and combinatorial estimates on nonbacktracking walks have proved to be powerful tools to study largest eigenvalues of some random matrices, see [14, 24, 7]. Indeed, counting non-backtracking walks turns out to be significantly simpler than counting simple walks. Moreover, on regular graphs, the nonbacktracking walks are counted by polynomials of the adjacency matrix, and the Ihara-Bass formula (see e.g. [19]) implies a simple relation between the spectrum of the adjacency matrix and the nonbacktracking matrix. In this paper, we extend this strategy beyond regular graphs. Let H denote the sparse random matrix we are interested in, and B its associated nonbacktracking matrix (defined in Definition 2.1 below). Our proof consists of two main steps: a deterministic step (i) and a probabilistic step (ii).

(i) The first step is an estimate of the spectral radius of H in terms of the spectral radius of B, summarized in Theorem 2.2. This step also requires bounds on the $\ell^2 \to \ell^\infty$ and $\ell^1 \to \ell^\infty$ norms of H, which are typically very easy to obtain by elementary concentration results. The main algebraic tool behind this step is an Ihara-Bass-type formula given in Lemma 4.1 below. Using this lemma we obtain an estimate relating the eigenvalues of B and H, Proposition 4.2 below, from which Theorem 2.2 easily follows.

(ii) The second step is an estimate of the spectral radius of B, summarized in Theorem 2.5 below. Our starting point is the classical Füredi-Komlós-approach of estimating $\mathbb{E} \operatorname{Tr} B^\ell B^{*\ell}$ for $\ell \gg \log n$, which may be analysed by counting walks on multigraphs. Our main result here is Proposition 5.1. The argument is based on a new soft combinatorial argument which revisits a reduction of walks introduced in Friedman [14].

Our main result for Hermitian sparse matrices, Theorem 2.7, follows from Theorems 2.2 and 2.5 combined with classical concentration estimates. The proof of our main result for non-Hermitian sparse matrices, Theorem 2.11, is proved by an argument analogous to that used to prove Theorem 2.5; the key observation here is that the independence of the entries of H automatically results in a nonbacktracking constraint in the graphical analysis of $\mathbb{E} \operatorname{Tr} H^\ell H^{*\ell}$.

In [2], an approach to estimating spectral radii of random matrices is developed by comparison of the spectral radii of general inhomogeneous random matrices to those of corresponding homogeneous random matrices. After our preprint appeared online, this approach was significantly extended in [22] to cover, among other things, sparse random matrices. For $G(n, d/n)$ in the regime $d/\log n \to \infty$, both our approach and that of [2, 22] yield bounds that are sharp to leading order, although our bound on the subleading error is better than that of [2, 22]. In the complementary very sparse regime $d/\log n \to 0$ our bounds are sharp up to a universal constant while those of [2, 22] are not. See Remark 2.8 below for more details.

We conclude this introduction with a summary of the main contributions of our paper.

(i) We obtain bounds on the spectral radius of a large class of sparse random matrices. For $G(n, d/n)$, our leading order bound is sharp in the regime $d/\log n \to \infty$ and sharp up to a universal constant in the regime $d/\log n \to 0$.

(ii) We establish optimal bounds at leading order for the spectral radius of sparse non-Hermitian matrices that are of mean-field type, which means that any $|H_{ij}|$ is a.s. much smaller than the ℓ^2 row norm $\sqrt{\max_i \sum_j \mathbb{E}|H_{ij}|^2}$.

(iii) We establish optimal bounds at leading order for the spectral radius of the nonbacktracking matrix in the regime where the ratio of the $\ell^2 \to \ell^\infty$ and $\ell^1 \to \ell^\infty$ norms grows. Such bounds are of some independent interest notably in view of recent applications of non-backtracking matrices in community detection; see [20]. Moreover, our estimate on the spectral radius of the nonbacktracking matrix is a key ingredient in the subsequent analysis [4] of the crossover at $d \asymp \log n$.

(iv) We establish a very simple and general relationship between the spectral radii of a Hermitian matrix and its nonbacktracking version, which we believe can be of some use in other contexts as well.

(v) Our proof is simple and fully self-contained, and in particular does not need a priori bounds on the spectral radius of a reference matrix.

2. Main results

For a positive integer n we abbreviate $[n] := \{1, 2, \ldots, n\}$. For a square matrix M we denote by $\sigma(M)$ the spectrum of M and by $\rho(M) := \max_{\lambda \in \sigma(M)} |\lambda|$ the spectral radius of M. Note that for Hermitian M we have $\rho(M) = \|M\| := \|M\|_{2 \to 2}$, the operator norm of M. In some heuristic discussions on the limit $n \to \infty$, we use the usual little o notation, and write $x \ll y$ to mean $x/y \to 0$.

Definition 2.1. Let $H = (H_{ij})_{i,j \in [n]} \in M_n(\mathbb{C})$ be a matrix with complex entries $H_{ij} \in \mathbb{C}$. The *nonbacktracking matrix associated with* H is the matrix $B = (B_{ef})_{e,f \in [n]^2} \in M_{n^2}(\mathbb{C})$ defined for $e = (i, j) \in [n]^2$ and $f = (k, l) \in [n]^2$ by

$$B_{ef} := H_{kl} \mathbf{1}_{j=k} \mathbf{1}_{i \neq l}. \tag{2.1}$$

We shall need the $\ell^2 \to \ell^\infty$ and $\ell^1 \to \ell^\infty$ norms of a matrix $H \in M_n(\mathbb{C})$, defined respectively as

$$\|H\|_{2 \to \infty} := \max_i \sqrt{\sum_j |H_{ij}|^2}, \qquad \|H\|_{1 \to \infty} := \max_{i,j} |H_{ij}|. \tag{2.2}$$

3

The following deterministic result estimates the spectral radius of an arbitrary Hermitian matrix in terms of the spectral radius of its nonbacktracking matrix.

Theorem 2.2. *For $x \geqslant 0$ define $f(x)$ through $f(x) = x + x^{-1}$ if $x \geqslant 1$ and $f(x) = 2$ if $0 \leqslant x \leqslant 1$. Let $H \in M_n(\mathbb{C})$ be a Hermitian matrix with associated nonbacktracking matrix B. Then*

$$\|H\| \leqslant \|H\|_{2\to\infty} f\left(\frac{\rho(B)}{\|H\|_{2\to\infty}}\right) + 7\|H\|_{1\to\infty}.$$

Using $f(x) \leqslant 2 + (x-1)_+^2$ we immediately deduce the following result.

Corollary 2.3. *If H is Hermitian with associated nonbacktracking matrix B then*

$$\|H\| \leqslant 2\|H\|_{2\to\infty} + \frac{(\rho(B) - \|H\|_{2\to\infty})_+^2}{\|H\|_{2\to\infty}} + 7\|H\|_{1\to\infty}.$$

We shall study the spectral radius $\rho(B)$ for the following class of random matrices.

Assumption 2.4. *Let $H \in M_n(\mathbb{C})$ be a Hermitian random matrix whose upper triangular entries $(H_{ij})_{1\leqslant i\leqslant j\leqslant n}$ are independent mean-zero random variables. Moreover, suppose that there exist $q > 0$ and $\kappa \geqslant 1$ such that*

$$\max_i \sum_j \mathbb{E}|H_{ij}|^2 \leqslant 1, \qquad \max_{i,j} \mathbb{E}|H_{ij}|^2 \leqslant \frac{\kappa}{n}, \qquad \max_{i,j}|H_{ij}| \leqslant \frac{1}{q} \ \ a.s. \tag{2.3}$$

For the interpretation of these parameters, consider first the simple case where A is the adjacency matrix of the homogeneous Erdős-Rényi graph $G(n, d/n)$ with $1 \leqslant d \leqslant n$. Define $H := d^{-1/2}(A - \mathbb{E}A)$. Then it is easy to check that Assumption 2.4 holds with $q = \sqrt{d}$ and $\kappa = 1$. Thus, the parameter q controls the sparsity of A (the smaller it is, the sparser A may be). Moreover, if the graph associated with A is inhomogeneous, the variances $\mathbb{E}|H_{ij}|^2$ depend on i and j, and κ has to be chosen larger than 1. Thus, κ controls the structure or inhomogeneity of A (the closer to 1 it is, the less structured, or more homogeneous, A is).

Theorem 2.5. *There are universal constants $C, c > 0$ such that the following holds. Suppose that H satisfies Assumption 2.4 and that B is the nonbacktracking matrix associated with H. Then for $1 \vee q \leqslant n^{1/10}\kappa^{-1/9}$ and $\varepsilon \geqslant 0$ we have*

$$\mathbb{P}\big(\rho(B) \geqslant 1 + \varepsilon\big) \leqslant Cn^{3-cq\log(1+\varepsilon)}.$$

Remark 2.6. Note that if Assumption 2.4 holds for some $q > 0$ then it also holds for any $0 < \hat{q} \leqslant q$. Hence, if $\kappa^{1/9} \leqslant n^{1/10}$, we may apply Theorem 2.5 to $\hat{q} = q \wedge (n^{1/10}\kappa^{-1/9})$.

From Theorems 2.2 and 2.5 it is not hard to conclude an upper bound on the norm of H. We first note that general concentration inequalities imply that there exists a universal constant $c > 0$ such that, if H satisfies Assumption 2.4 then for all $t \geqslant 0$ we have

$$\mathbb{P}\big(\big|\|H\| - \mathbb{E}\|H\|\big| \geqslant t\big) \leqslant 2e^{-cq^2t^2}. \tag{2.4}$$

See for instance [9, Examples 3.14 and 6.8]; the concentration inequality in [9, Example 6.8] is given for the largest eigenvalue $\lambda_{\max}(H) = \sup_{\|x\|_2=1}\langle x, Hx\rangle$, but the same argument applies to $\|H\| = \sup_{\|x\|_2=\|y\|_2=1}\langle y, Hx\rangle$.

The estimate (2.4) shows that, up to subgaussian fluctuations of order $1/q$, it is sufficient to control the expectation of the norm of H. Our main result in this direction is the following.

Theorem 2.7. *There is a universal constant $C > 0$ such that the following holds. Suppose that H satisfies Assumption 2.4. Then for $1 \vee q \leqslant n^{1/10}\kappa^{-1/9}$, we have*

$$\mathbb{E}\|H\| \leqslant 2 + C\frac{\eta}{\sqrt{1 \vee \log \eta}}, \qquad with \quad \eta := \frac{\sqrt{\log n}}{q}. \tag{2.5}$$

In particular,

$$\mathbb{E}\|H\| \leqslant 2 + C\frac{\sqrt{\log n}}{q}. \tag{2.6}$$

Moreover, under the same hypotheses, if in addition $q \geqslant C$ and $\max_i \sum_j \mathbb{E}|H_{ij}|^2 = 1$ then

$$\mathbb{E}\|H\| \leqslant \mathbb{E}\|H\|_{2\to\infty}\left(2 + \frac{C}{q}\right). \tag{2.7}$$

Theorem 2.7 shows that with high probability, $\|H\| \leqslant 2 + o(1)$ as soon as $q^2 \gg \log n$ and $\kappa \ll n^{12/13}$ (See also Remark 2.13 below). As illustrated in [17, Theorem 2.2] and [3, Corollary 1.4], this statement is sharp. We remark that for entries with symmetric distributions, a related bound (without sharp error bounds) was obtained in [17, Theorem 2.1]. (However, the assumption of symmetric distributions rules out random graphs.)

Remark 2.8. Theorem 2.7 also provides an estimate for the very sparse regime $q^2 \leqslant \log n$. Writing $q^2 = \eta^{-2}\log n$ for $\eta \geqslant 1$ we find from Theorem 2.7 and (2.4) that

$$\|H\| \leqslant C\frac{\eta}{\sqrt{\log \eta}}$$

with high probability. On the other hand, in [3] we proved that for A the adjacency matrix of $G(n, d/n)$ with $d = \eta^{-2}\log n$ and $H = (A - \mathbb{E}A)/q$ we have

$$\|H\| \sim \frac{\eta}{\sqrt{2\log \eta}}$$

with high probability for $\eta \gg 1$ (where $a \sim b$ denotes that $a/b \to 1$ as $n \to \infty$). Thus, in the very sparse regime $q^2 \leqslant \log n$ and for general sparse random matrices, Theorem 2.7 yields bounds whose dependence on n and d is optimal up to a universal constant for the example of the Erdős-Rényi graph.

Remark 2.9. For any Hermitian matrix H we always have $\|H\|_{2\to\infty} \leqslant \|H\|$. The estimate (2.7) may be regarded as a probabilistic counterpart to this statement, saying that with high probability the converse estimate is also true up to a constant.

Remark 2.10. Stated explicitly without referring to Assumption 2.4, for instance the estimate (2.6) states that for any Hermitian matrix H with independent centred entries, if

$$\max_{i,j}\mathbb{E}|H_{ij}|^2 \cdot \left(\max_i \sum_j \mathbb{E}|H_{ij}|^2\right)^{7/2} \leqslant n^{-1/10}\left\|\max_{i,j}|H_{ij}|\right\|_{L^\infty}^9$$

then

$$\mathbb{E}\|H\| \leqslant 2\sqrt{\max_i \sum_j \mathbb{E}|H_{ij}|^2} + C\left\|\max_{i,j}|H_{ij}|\right\|_{L^\infty}\sqrt{\log n}.$$

The same remains valid if the quantity $\left\|\max_{i,j}|H_{ij}|\right\|_{L^\infty}$ is replaced by any larger number.

Our techniques also apply to non-Hermitian matrices with independent entries.

Theorem 2.11. *Let $H \in M_n(\mathbb{C})$ be a random matrix whose entries $(H_{ij})_{1 \leqslant i,j \leqslant n}$ are independent mean-zero random variables. Moreover, suppose that there exist $q > 0$ and $\kappa \geqslant 1$ such that (2.3) holds. Then for $1 \vee q \leqslant n^{1/10} \kappa^{-1/9}$ and $\varepsilon \geqslant 0$ we have*

$$\mathbb{P}\big(\rho(H) \geqslant 1 + \varepsilon\big) \leqslant C n^{2 - cq \log(1+\varepsilon)} \,,$$

for some universal positive constants C, c.

Remark 2.12. In Theorems 2.5, 2.7, and 2.11, the almost sure last condition of (2.3) can be easily relaxed by a truncation argument of the entries of H (see for example [27]). We do not pursue this generalization any further.

Remark 2.13. It can easily be seen from the proofs that in Theorems 2.5, 2.7, and 2.11, the condition $q \leqslant n^{1/13} \kappa^{-1/12}$ can be relaxed to $q \leqslant n^{(1-\delta_0)/12} \kappa^{-1/12}$ for any fixed $\delta_0 \in (0,1)$. Then, the constants involved in the statements depend on δ_0. Consequently, our main results hold for $\kappa \leqslant n^{1-c}$ for any fixed $c > 0$. The applications to inhomogeneous Erdős-Rényi graphs in Section 3 can be slightly strengthened in this direction; we shall not pursue this direction further. Moreover, since the parameter q only appears as an upper bound in (2.3), our results clearly apply without an upper bound on q; see Remark 2.6.

2.1. Overview of proofs. We conclude this section by giving an overview of the proofs of our main results – Theorems 2.2, 2.5, 2.7, and 2.11. Theorem 2.2 is proved in Section 4. The first main ingredient is a variant of the Ihara-Bass formula, Lemma 4.1, which relates the spectra of H and B by characterizing the spectrum of B as the singularity set of an explicit matrix function, denoted $M(\lambda) - H(\lambda)$. The second main ingredient is an analysis of the matrix function $M(\lambda) - H(\lambda)$, estimating its singularities in terms of eigenvalues of H, hence concluding the proof of Theorem 2.2. The proof of Theorems 2.5 and 2.7 are given in Section 2.7, where we estimate the spectral radius of B. The main work is to estimate $\mathbb{E} \operatorname{Tr} B^\ell B^{*\ell}$ for large ℓ, which is performed in Proposition 5.1. It is proved using a variant of the classical moment method, exploiting the nonbacktracking property of B to significantly reduce the combinatorics of the resulting graphs. Theorem 2.5 then easily follows using Markov's inequality, and Theorem 2.7 follows by an elementary argument using Theorems 2.2 and 2.5. Finally, in Section 6 we prove Theorem 2.11. Its proof is remarkably similar to that of Proposition 5.1. Indeed, since the entries H_{ij} and H_{ji} for $i \neq j$ are independent, to leading order, the resulting walks on graphs do not backtrack. This results in walks analogous to those of Section 2.7, and we may take over much of the arguments developed there to conclude the proof of Theorem 2.11.

3. Application to inhomogeneous Erdős-Rényi graphs

3.1. Undirected graphs. An important example of a matrix H satisfying Assumption 2.4 is the (centred and rescaled) adjacency matrix of an inhomogeneous (undirected) Erdős-Rényi random graph, where each edge $\{i,j\}$, $1 \leqslant i < j \leqslant n$, is included with probability p_{ij}, independently of the others. Let A be its adjacency matrix and set

$$H := d^{-1/2}(A - \mathbb{E}A) \,,$$

6

where

$$d := \max_i \sum_j p_{ij} \tag{3.1}$$

is the maximal expected degree. Then for each i, j and each $k \geqslant 2$, we have $|H_{ij}| \leqslant 1/\sqrt{d}$ and $\mathbb{E}|H_{ij}|^2 \leqslant p_{ij}/d$. We set

$$q := \sqrt{d} \wedge n^{1/10} \kappa^{-1/9}, \qquad \kappa := \frac{\max_{i,j} p_{ij}}{d/n}. \tag{3.2}$$

Remark 3.1. A scenario of particular interest is when the probabilities p_{ij} are all comparable (i.e. $\kappa = O(1)$) and the typical expected degree d is not too large, in the sense that $d \ll n^{1/5}$. In that case we have $q = \sqrt{d}$.

We deduce the following immediate consequence of Theorem 2.7.

Theorem 3.2. *Let A be the adjacency matrix of an inhomogeneous Erdős-Rényi graph, with d and q defined as in* (3.1) *and* (3.2). *If $q \geqslant 1$ then*

$$\frac{\mathbb{E}\|A - \mathbb{E}A\|}{\sqrt{d}} \leqslant 2 + C \frac{\eta}{\sqrt{1 \vee \log \eta}}, \qquad with \quad \eta := \frac{\sqrt{\log n}}{q}, \tag{3.3}$$

for some universal constant $C > 0$. If the maximal expected degree is not too large in the sense that

$$d \geqslant 1 \quad and \quad d^{7/2} \max_{i,j} p_{ij} \leqslant n^{-1/10}, \tag{3.4}$$

which in particular includes the scenario of Remark 3.1, then

$$\mathbb{E}\|A - \mathbb{E}A\| \leqslant 2\sqrt{d} + C\sqrt{\frac{\log n}{1 \vee \log\left(\frac{\log n}{d}\right)}}. \tag{3.5}$$

In [13, 10], it is proved that $\|H\| \leqslant C$ if d is at least of order $\log n$. Theorem 3.2 retrieves this result and states also that $\|H\| \leqslant 2 + o(1)$ as soon as $d \gg \log n$. In the homogenous case (i.e. when for all i, j, $p_{ij} = d/n$), it is well known [18, 5] that the empirical distribution of the eigenvalues of H converges weakly to the semi-circular law with support $[-2, 2]$ as soon as $d \gg 1$. Theorem 3.2 then also gives the convergence of the extreme eigenvalues to the boundary of $[-2, 2]$ as soon as $d \gg \log n$. In the companion paper [3], we study the largest eigenvalues of H in the regime $d \ll \log n$, showing that a crossover occurs at $d \asymp \log n$. Interestingly, in at least the regime $d \ll \log n$, [3, Corollary 1.4] implies that the upper bound (3.5) is sharp up to the multiplicative constant C; see Remark 2.8.

Theorem 3.2 can be used in statistical inference on graphs, where one wishes to infer information about $\mathbb{E}A$ from a single observation of A. Weyl's inequality for eigenvalues implies that, for any integer $1 \leqslant k \leqslant n$, the k-th largest eigenvalue (counting multiplicities) of $\mathbb{E}A$ and A differ by at most $\|A - \mathbb{E}A\|$ (see for example [6]). Hence, Theorem 3.2 shows that the location of an eigenvalue λ of $\mathbb{E}A$ can be effectively estimated from the spectrum of A as soon as $|\lambda|/\sqrt{d}$ is much larger than the right-hand side of (3.3). In particular, if $|\lambda|$ is of order d, the condition reads

$$d \gg \sqrt{\frac{\log n}{\log \log n}}. \tag{3.6}$$

As mentioned above, [13, 10] require the stronger condition $d \gg \log n$. We also recall that classical tools from perturbation theory assert that also the corresponding eigenspaces of A and $\mathbb{E}A$ are close when $\|A - \mathbb{E}A\|$ is small compared to the spectral gap around the eigenvalue of A under consideration (for instance, the Davis-Kahan Theorem [12] gives a precise quantitative statement of this kind). We may summarize our discussion with the following corollary, which can be used in statistical inference on graphs when (3.6) is satisfied. For simplicity of presentation, we focus on the case where the expected maximal degree is not too large, that is when (3.4) holds.

Corollary 3.3. *Let A be the adjacency matrix of an inhomogeneous Erdős-Rényi graph, with d defined as in (3.1). Suppose that (3.4) holds and $n \geqslant 3$. Then for some universal constant $C > 0$ and for any $0 < \varepsilon < 1$, if*

$$d \geqslant \frac{C}{\varepsilon^2} \sqrt{\frac{\log n}{\log \log n}} \tag{3.7}$$

then

$$\|A - \mathbb{E}A\| \leqslant \varepsilon d \,,$$

with probability at least $1 - 2\exp(-\varepsilon^2 d^3 / C)$.

Proof. This is an elementary argument using (3.5), the condition (3.7), and (2.4). We omit the details. □

3.2. Directed graphs. An important example of a matrix H satisfying the assumptions of Theorem 2.11 is the (centred and rescaled) adjacency matrix of an inhomogeneous directed Erdős-Rényi random graph, where each directed edge (i, j), $1 \leqslant i, j \leqslant n$, is included with probability p_{ij}, independently of the others. Let A be its adjacency matrix and set

$$H := d^{-1/2}(A - \mathbb{E}A) \,,$$

with d given in (3.1). We deduce the following immediate consequence of Theorem 2.7.

Theorem 3.4. *Let A be the adjacency matrix of an inhomogeneous directed Erdős-Rényi graph, with d and q defined as in (3.1) and (3.2). If $q \geqslant 1$ then*

$$\mathbb{P}\big(\rho\big(d^{-1/2}(A - \mathbb{E}A)\big) \geqslant 1 + \varepsilon\big) \leqslant Cn^{2 - cq\log(1+\varepsilon)} \,,$$

for any $\varepsilon \geqslant 0$ and some universal positive constants C, c. If the maximal expected degree is not too large in the sense that (3.4) holds, then we have

$$\mathbb{P}\big(\rho\big(d^{-1/2}(A - \mathbb{E}A)\big) \geqslant 1 + \varepsilon\big) \leqslant Cn^{2 - c\sqrt{d}\log(1+\varepsilon)}$$

for all $\varepsilon \geqslant 0$.

The following corollary can be deduced directly, using the version of the Bauer-Fike theorem given in [8].

Corollary 3.5. *Suppose that the assumptions of Theorem 3.4 hold, and suppose moreover that for all i, j we have $p_{ij} = d/n$ where $d \leqslant n^{1/5}$. Then for any $\varepsilon \geqslant 0$ such that $2(1 + \varepsilon) < \sqrt{d}$, with probability at least $1 - Cn^{2 - c\sqrt{d}\log(1+\varepsilon)}$, A has exactly one eigenvalue at distance at most $(1+\varepsilon)\sqrt{d}$ from d and all other eigenvalues with modulus at most $(1+\varepsilon)\sqrt{d}$.*

4. Comparison of spectra of H and B and proof of Theorem 2.2

The rest of this paper is devoted to the proofs of Theorems 2.2, 2.5, 2.7, and 2.11.

4.1. An Ihara-Bass-type formula. The following lemma is a variant of the Ihara-Bass formula. It is inspired by [23] and generalizes ideas from [21]. It is essentially contained in Theorem 2 of [26]; for the convenience of the reader and to keep this paper self-contained, we give the simple proof.

Lemma 4.1. *Let $H \in M_n(\mathbb{C})$ with associated nonbacktracking matrix B and let $\lambda \in \mathbb{C}$ satisfy $\lambda^2 \neq H_{ij}H_{ji}$ for all $i,j \in [n]$. Define the matrices $H(\lambda)$ and $M(\lambda) = \mathrm{diag}(m_i(\lambda))_{i \in [n]}$ through*

$$H_{ij}(\lambda) \; := \; \frac{\lambda H_{ij}}{\lambda^2 - H_{ij}H_{ji}}, \qquad m_i(\lambda) \; := \; 1 + \sum_k \frac{H_{ik}H_{ki}}{\lambda^2 - H_{ik}H_{ki}} \,. \tag{4.1}$$

Then $\lambda \in \sigma(B)$ if and only if $\det(M(\lambda) - H(\lambda)) = 0$.

Proof. We abbreviate $ij \equiv (i,j) \in [n]^2$. Let $\lambda \in \sigma(B)$ be an eigenvalue of B with eigenvector $x \in \mathbb{C}^{[n]^2}$, i.e. $Bx = \lambda x$, which reads in components

$$\lambda x_{ji} = \sum_{k \neq j} H_{ik} x_{ik} \tag{4.2}$$

for all $i,j \in [n]$. We define $y \in \mathbb{C}^n$ by, for each i,

$$y_i := \sum_k H_{ik} x_{ik} \,.$$

The eigenvalue equation $\lambda x = Bx$ reads

$$\lambda x_{ji} = y_i - H_{ij} x_{ij} \,.$$

Exchanging i and j, we obtain

$$\lambda x_{ij} = y_j - H_{ji} x_{ji} \,,$$

from which we deduce

$$\lambda^2 x_{ji} = \lambda y_i - H_{ij} \lambda x_{ij} = \lambda y_i - H_{ij} y_j + H_{ij} H_{ji} x_{ji} \,.$$

Hence, because $\lambda^2 \neq H_{ij}H_{ji}$,

$$x_{ji} = \frac{\lambda y_i - H_{ij} y_j}{\lambda^2 - H_{ij} H_{ji}} \,. \tag{4.3}$$

We see from this last expression that $y \neq 0$ if $x \neq 0$. We plug this last expression into (4.2) and get

$$\frac{\lambda^2 y_i}{\lambda^2 - H_{ij}H_{ji}} - \frac{\lambda H_{ij} y_j}{\lambda^2 - H_{ij}H_{ji}} = \sum_{k \neq j} \frac{\lambda H_{ik} y_k}{\lambda^2 - H_{ik}H_{ki}} - \sum_{k \neq j} \frac{H_{ik}H_{ki} y_i}{\lambda^2 - H_{ik}H_{ki}} \,,$$

i.e.

$$\frac{\lambda^2 y_i}{\lambda^2 - H_{ij}H_{ji}} = \sum_k \frac{\lambda H_{ik} y_k}{\lambda^2 - H_{ik}H_{ki}} - \sum_{k \neq j} \frac{H_{ik}H_{ki} y_i}{\lambda^2 - H_{ik}H_{ki}} \,.$$

We conclude that

$$y_i = \frac{\lambda^2 y_i}{\lambda^2 - H_{ij}H_{ji}} - \frac{H_{ij}H_{ji}y_i}{\lambda^2 - H_{ij}H_{ji}} = \sum_k \frac{\lambda H_{ik}y_k}{\lambda^2 - H_{ik}H_{ki}} - \sum_k \frac{H_{ik}H_{ki}y_i}{\lambda^2 - H_{ik}H_{ki}}.$$

Hence,

$$y_i\left(1 + \sum_k \frac{H_{ik}H_{ki}}{\lambda^2 - H_{ik}H_{ki}}\right) - \sum_k \frac{\lambda H_{ik}y_k}{\lambda^2 - H_{ik}H_{ki}} = 0,$$

which proves that 0 is an eigenvalue of $M(\lambda) - H(\lambda)$.

Conversely, if 0 is an eigenvalue of $M(\lambda) - H(\lambda)$ with eigenvector y, we define x through (4.3). Then the above computation also implies that x satisfies (4.2), i.e. $Bx = \lambda x$, so that $\lambda \in \sigma(B)$. $\square$

4.2. Comparison of spectra and proof of Theorem 2.2. We use the notation $M \preceq N$ for Hermitian matrices M and N to mean that $N - M$ is a positive matrix. The key estimate behind the proof of Theorem 2.2 is the following result. We thank Sasha Sodin for his help in simplifying its proof.

Proposition 4.2. *Let $H \in M_n(\mathbb{C})$ be a Hermitian matrix with associated nonbacktracking matrix B. Suppose that there exists $\delta \in [0,1]$ such that*

$$\max_{i,j}|H_{ij}| \leqslant \delta \qquad and \qquad \max_i \sum_j |H_{ij}|^2 \leqslant 1 + \delta. \tag{4.4}$$

Let

$$\lambda_0 := \max\{1 + \sqrt{\delta}, \max(\sigma(B) \cap \mathbb{R})\}. \tag{4.5}$$

Then

$$H \preceq \lambda_0 + \frac{1}{\lambda_0} + 6\delta.$$

Proof. Let $H(\lambda)$ and $M(\lambda)$ be the matrices defined in Lemma 4.1. First note that $0 \preceq M(\lambda_0) - H(\lambda_0)$. Indeed, as $\lambda \to +\infty$,

$$M(\lambda) - H(\lambda) = I + O(\lambda^{-1}),$$

so that for λ large enough $M(\lambda) - H(\lambda)$ is positive definite. But by Lemma 4.1, the real zeros of $\lambda \mapsto \det(M(\lambda) - H(\lambda))$ are the real eigenvalues of B, which implies that for $\lambda > \lambda_0$, $\det(M(\lambda) - H(\lambda)) > 0$. By continuity, we conclude that $M(\lambda) - H(\lambda)$ is positive definite for any $\lambda > \lambda_0$.

Next, a direct computation shows that for any $\lambda \geqslant 1 + \sqrt{\delta}$, we have

$$|\lambda H_{ij}(\lambda) - H_{ij}| = |H_{ij}|^3 \frac{1}{\lambda^2 - |H_{ij}|^2} \leqslant \delta|H_{ij}|^2 \frac{1}{\lambda^2 - \delta^2} \leqslant \delta|H_{ij}|^2 \frac{1}{1 + \delta + 2\sqrt{\delta} - \delta^2} \leqslant \delta|H_{ij}|^2.$$

We deduce (by the Schur test or the Gershgorin circle theorem) that $\|\lambda H(\lambda) - H\| \leqslant \delta \max_i \sum_j |H_{ij}|^2 \leqslant 2\delta$. Another computation shows that for any $\lambda \geqslant 1 + \sqrt{\delta}$, we have

$$\lambda m_i(\lambda) - \left(\lambda + \frac{1}{\lambda}\right) = \frac{1}{\lambda}\left(\sum_k \frac{|H_{ik}|^2}{1 - \lambda^{-2}|H_{ik}|^2} - 1\right) \leqslant \frac{1}{\lambda}\left(\frac{1 + \delta}{1 - \lambda^{-2}\delta^2} - 1\right)$$

$$= \delta\frac{\lambda(1 + \lambda^{-2}\delta)}{\lambda^2 - \delta^2} \leqslant \delta\frac{2\lambda}{\lambda^2 - \delta^2} \leqslant 4\delta.$$

10

(In the last step, we considered separately the cases $\lambda \geqslant 2$ and $\lambda < 2$, for which we use $\lambda^2 - \delta^2 \geqslant 1$.) From both previous computations, we deduce that for any $\lambda \geqslant 1 + \sqrt{\delta}$,

$$\lambda(M(\lambda) - H(\lambda)) \preceq \lambda + \frac{1}{\lambda} - H + 6\delta.$$

To sum up, we have

$$0 \preceq \lambda_0(M(\lambda_0) - H(\lambda_0)) \preceq \lambda_0 + \frac{1}{\lambda_0} - H + 6\delta\,,$$

from which the claim follows. □

We may now conclude the proof of Theorem 2.2.

Proof of Theorem 2.2. Set $\delta := \|H\|_{1\to\infty}$. Note that $\delta \leqslant 1$ because $\|H\|_{2\to\infty} \leqslant 1$. By Proposition 4.2, for $\lambda_1 := \max\{1 + \sqrt{\delta}, \rho(B)\}$, we have

$$\|H\| \leqslant f(\lambda_1) + 6\delta\,.$$

By considering the cases $\lambda_1 = \rho(B)$ and $\lambda_1 = 1 + \sqrt{\delta}$ separately, and using that $2 \leqslant f(1+x) \leqslant 2+x^2$, we easily find that if $\|H\|_{2\to\infty} \leqslant 1$ then

$$\|H\| \leqslant f(\rho(B)) + 7\|H\|_{1\to\infty}\,.$$

The claim for arbitrary H now follows by homogeneity. □

5. Estimate of $\rho(B)$ and proof of Theorems 2.5 and 2.7

The main estimate of this section is the following result.

Proposition 5.1. *Suppose that H satisfies Assumption 2.4 and that B is the nonbacktracking matrix associated with H. There exist univeral constants $c_0, C_0 > 0$ such that the following holds. If $\ell \geqslant 1$, $q \geqslant 1$, and $\delta \in (0, 1/3)$ satisfy*

$$\ell \leqslant c_0 \min\left\{\delta q \log n, \frac{n^{1/3-\delta}}{q^2 \kappa^{1/3}}\right\} \tag{5.1}$$

then

$$\mathbb{E}\operatorname{Tr} B^\ell B^{*\ell} \leqslant C_0 n^2 \ell^4 q^2\,.$$

5.1. Proof of Proposition 5.1. Throughout the following we fix ℓ and mostly omit it from our notation. For any $e \in [n]^2$ and $f \in [n]^2$ we have

$$(B^\ell)_{ef} = \sum_{a_1, \ldots, a_{\ell-1} \in [n]^2} B_{ea_1} B_{a_1 a_2} \cdots B_{a_{\ell-1} f}\,.$$

By Definition 2.1, we therefore find

$$(B^\ell)_{ef} = \sum_\xi H_{\xi_0 \xi_1} H_{\xi_1 \xi_2} \cdots H_{\xi_{\ell-1} \xi_\ell}\,,$$

11

where the sum runs over $\xi = (\xi_{-1}, \xi_0, \ldots, \xi_\ell) \in [n]^{\ell+2}$ satisfying $(\xi_{-1}, \xi_0) = e$, $(\xi_{\ell-1}, \xi_\ell) = f$ and $\xi_{i-1} \neq \xi_{i+1}$ for $i = 0, \ldots, \ell - 1$. Hence,

$$\operatorname{Tr} B^\ell B^{\ell*} = \sum_{e,f \in [n]^2} |(B^\ell)_{ef}|^2$$
$$= \sum_{\xi^1, \xi^2} H_{\xi_0^1 \xi_1^1} H_{\xi_1^1 \xi_2^1} \cdots H_{\xi_{\ell-1}^1 \xi_\ell^1} \overline{H}_{\xi_0^2 \xi_1^2} \overline{H}_{\xi_1^2 \xi_2^2} \cdots \overline{H}_{\xi_{\ell-1}^2 \xi_\ell^2},$$

where the sum runs over $\xi^1 = (\xi_{-1}^1, \ldots, \xi_\ell^1)$, $\xi^2 = (\xi_{-1}^2, \ldots, \xi_\ell^2) \in [n]^{\ell+2}$ such that $(\xi_{-1}^1, \xi_0^1) = (\xi_{-1}^2, \xi_0^2)$, $(\xi_{\ell-1}^1, \xi_\ell^1) = (\xi_{\ell-1}^2, \xi_\ell^2)$ and $\xi_{i-1}^1 \neq \xi_{i+1}^1$ and $\xi_{i-1}^2 \neq \xi_{i+1}^2$ for $i = 0, \ldots, \ell - 1$. Note that ξ_{-1}^1 does not appear as an index of H. Fixing all indices except $\xi_{-1}^1 = \xi_{-1}^2$, we find that the sum over $\xi_{-1}^1 = \xi_{-1}^2$ is bounded by n. The remaining sum over $\xi_0^1, \xi_0^2, \ldots, \xi_\ell^1, \xi_\ell^2$ is nonnegative. This yields the estimate

$$\operatorname{Tr} B^\ell B^{\ell*} \leqslant n \sum_{\xi^1, \xi^2} H_{\xi_0^1 \xi_1^1} H_{\xi_1^1 \xi_2^1} \cdots H_{\xi_{\ell-1}^1 \xi_\ell^1} \overline{H}_{\xi_0^2 \xi_1^2} \overline{H}_{\xi_1^2 \xi_2^2} \cdots \overline{H}_{\xi_{\ell-1}^2 \xi_\ell^2}$$
$$= n \sum_{\xi^1, \xi^2} H_{\xi_0^1 \xi_1^1} H_{\xi_1^1 \xi_2^1} \cdots H_{\xi_{\ell-1}^1 \xi_\ell^1} H_{\xi_\ell^2 \xi_{\ell-1}^2} H_{\xi_{\ell-1}^2 \xi_{\ell-2}^2} \cdots H_{\xi_1^2 \xi_0^2},$$

where the sum runs over $\xi^1 = (\xi_0^1, \ldots, \xi_\ell^1), \xi^2 = (\xi_0^2, \ldots, \xi_\ell^2) \in [n]^{\ell+1}$ such that $(\xi_0^1, \xi_{\ell-1}^1, \xi_\ell^1) = (\xi_0^2, \xi_{\ell-1}^2, \xi_\ell^2)$ and $\xi_{i-1}^1 \neq \xi_{i+1}^1$ and $\xi_{i-1}^2 \neq \xi_{i+1}^2$ for $i = 1, \ldots, \ell - 1$. In the second step we used that H is Hermitian. Renaming the summation variables, we have

$$\operatorname{Tr} B^\ell B^{\ell*} \leqslant n \sum_{\xi \in \widetilde{\mathcal{C}}} H_{\xi_0 \xi_1} H_{\xi_1 \xi_2} \cdots H_{\xi_{2\ell-1} \xi_{2\ell}},$$

where

$$\widetilde{\mathcal{C}} := \left\{ \xi = (\xi_0, \ldots, \xi_{2\ell}) \in [n]^{2\ell+1} : \xi_0 = \xi_{2\ell}, \xi_{\ell-1} = \xi_{\ell+1}, \xi_{i-1} \neq \xi_{i+1} \text{ for } i \in [2\ell-1] \backslash \{\ell\} \right\}.$$

Because the entries of H are independent and have mean zero, we find

$$\mathbb{E} \operatorname{Tr} B^\ell B^{\ell*} \leqslant n \mathbb{E} \sum_{\xi \in \mathcal{C}} H_{\xi_0 \xi_1} H_{\xi_1 \xi_2} \cdots H_{\xi_{2\ell-1} \xi_{2\ell}}, \tag{5.2}$$

where we defined $\mathcal{C}$ as the set of $\xi \in \widetilde{\mathcal{C}}$ such that $\left| \left\{ i \in [2\ell] : \{\xi_{i-1}, \xi_i\} = \{a, b\} \right\} \right| \neq 1$ for all $a, b \in [n]$. In words, for $\xi \in \mathcal{C}$ every unordered edge cannot be crossed only once by ξ.

In the following we shall need several basic graph-theoretic notions. Since they involve paths on multigraphs, it is important to introduce them with some care. By definition, a (vertex-labelled undirected) *multigraph* $G = (V(G), E(G), \phi)$ consists of two finite sets, the set of *vertices* $V(G)$ and the set of *edges* $E(G)$, and a map ϕ from $E(G)$ to the unordered sets of one or two elements of $V(G)$. For $e \in E(G)$, the set $\phi(e)$ is the set of vertices *incident* to e. The edge $e \in E(G)$ is a *loop* if $|\phi(e)| = 1$. The *degree* $\deg(v)$ of a vertex $v \in V(G)$ is the number of edges to which it is incident, whereby a loop incident to v counts twice. The *genus* of G is $g(G) := |E(G)| - |V(G)| + 1$. A *path of length* $l \geqslant 1$ *in* G is a word $w = w_0 w_{01} w_1 w_{12} \ldots w_{l-1l} w_l$ such that $w_0, w_1, \ldots, w_l \in V(G)$, $w_{01}, w_{12}, \ldots, w_{l-1l} \in E(G)$, and $\phi(w_{i-1i}) = \{w_{i-1}, w_i\}$ for $i = 1, \ldots, l$. We denote the length l of w by $|w| := l$. The path w is *closed* if $w_0 = w_l$. For $e \in E(G)$ we define the *number of crossings of e by w* to be $m_e(w) := \sum_{i=1}^l \mathbf{1}_{w_{i-1i} = e}$. In particular, we have $\sum_{e \in E(G)} m_e(w) = |w|$.

Moreover, the multigraph G is called simply a *graph* if ϕ is injective, i.e. there are no multiple edges. (Note that in our convention a graph may have loops.) For a graph G we may and shall identify $E(G)$ with a set of unordered pairs of $V(G)$, simply identifying e and $\phi(e)$. Similarly, we identify a path w with the reduced word $w_0 w_1 \ldots w_l$ only containing the vertices, since we must have $w_{i-1 i} = \{w_{i-1}, w_i\}$ for $i = 1, \ldots, l$.

Definition 5.2. For $\xi \in \mathcal{C}$ we define the graph G_ξ as

$$V(G_\xi) := \{\xi_i : i = 0, \ldots, 2\ell\}, \qquad E(G_\xi) := \left\{\{\xi_{i-1}, \xi_i\} : i = 1, \ldots, 2\ell\right\}.$$

Thus, $\xi \equiv \xi_0 \xi_1 \cdots \xi_{2\ell}$ is a closed path in the graph G_ξ.

Next, we introduce an equivalence relation on $\mathcal{C}$ by saying that two paths $\xi, \tilde{\xi} \in \mathcal{C}$ are equivalent, denoted $\xi \sim \tilde{\xi}$, if and only if there exists a permutation τ of $[n]$ such that $\tau(\xi_i) = \tilde{\xi}_i$ for all $i = 0, \ldots, 2\ell$. Clearly, the numbers $|E(G_\xi)|$ and $|V(G_\xi)|$, and hence also $g(G_\xi)$, only depend on the equivalence class of ξ. We denote by $[\xi] \subset \mathcal{C}$ the equivalence class of a path $\xi \in \mathcal{C}$ in the set $\mathcal{C}$.

Lemma 5.3. *For any $\bar{\xi} \in \mathcal{C}$ we have*

$$\mathbb{E} \sum_{\xi \in [\bar{\xi}]} H_{\xi_0 \xi_1} H_{\xi_1 \xi_2} \cdots H_{\xi_{2\ell-1} \xi_{2\ell}} \leqslant n^{1 - g(G_{\bar{\xi}})} \kappa^{g(G_{\bar{\xi}})} q^{2|E(G_{\bar{\xi}})| - 2\ell} .$$

Proof. Abbreviate $g := g(G_{\bar{\xi}})$, $a := |E(G_{\bar{\xi}})|$, and $s := |V(G_{\bar{\xi}})|$, so that $g = a - s + 1$. Since the claim only depends on $\bar{\xi}$ through its equivalence class $[\bar{\xi}]$, we may replace $\bar{\xi}$ with any equivalent path of $[\bar{\xi}]$ obtained by relabelling the vertices. Thus, we may suppose that $V(G_{\bar{\xi}}) = [s]$ and that there is a spanning tree T of $G_{\bar{\xi}}$ such that for all $t \in 2, \ldots, s$ the subgraph $T|_{[t]}$ is a spanning tree of $[t]$. (This amounts to the requirement that the vertices $[s]$ are first explored in increasing order by the path $\bar{\xi}$.) Moreover, we enumerate the edges of $G_{\bar{\xi}}$ as $e_1, \ldots, e_a$, where for $t = 1, \ldots, s - 1$ the edge e_t is the unique edge of $T|_{[t+1]}$ that is not an edge of $T|_{[t]}$, and the edges $e_s, \ldots, e_a$ are the remaining g edges in some arbitrary order.

For $t = 1, \ldots, a$ we abbreviate $m_t := m_{e_t}(\bar{\xi})$ for the number of crossings of e_t by $\bar{\xi}$. Moreover, we denote by $\mathcal{I}_{s,n}$ the set of injective maps from $[s]$ to $[n]$. Then we have

$$\mathbb{E} \sum_{\xi \in [\bar{\xi}]} H_{\xi_0 \xi_1} H_{\xi_1 \xi_2} \cdots H_{\xi_{2\ell-1} \xi_{2\ell}} = \sum_{\tau \in \mathcal{I}_{s,n}} \mathbb{E} H_{\tau(\bar{\xi}_0)\tau(\bar{\xi}_1)} H_{\tau(\bar{\xi}_1)\tau(\bar{\xi}_2)} \cdots H_{\tau(\bar{\xi}_{2\ell-1})\tau(\bar{\xi}_{2\ell})}$$

$$\leqslant \sum_{\tau \in \mathcal{I}_{s,n}} \prod_{t=1}^{a} \mathbb{E} |H_{\tau(e_t)}^{m_t}|,$$

where we used the independence of the entries and the convention $\tau(\{x, y\}) := \{\tau(x), \tau(y)\}$. We use the estimate

$$\max_i \sum_j \mathbb{E} |H_{ij}|^k \leqslant \frac{1}{q^{k-2}} \qquad (k = 2, 3, \ldots) \tag{5.3}$$

for the edges $e_1, \ldots, e_{s-1}$ and the estimate

$$\max_{i,j} \mathbb{E} |H_{ij}|^k \leqslant \frac{\kappa}{n q^{k-2}} \qquad (k = 2, 3, \ldots) \tag{5.4}$$

for the edges $e_s, \ldots, e_a$; both of these estimates follow immediately from (2.3). Thus we find

$$
\mathbb{E} \sum_{\xi \in [\bar{\xi}]} H_{\xi_0 \xi_1} H_{\xi_1 \xi_2} \cdots H_{\xi_{2\ell-1} \xi_{2\ell}} \leqslant \prod_{t=s}^{a} \frac{\kappa}{nq^{m_t-2}} \sum_{\tau \in \mathcal{I}_{s,n}} \prod_{t=1}^{s-1} \mathbb{E}|H_{\tau(e_t)}^{m_t}|
$$

$$
\leqslant \prod_{t=s}^{a} \frac{\kappa}{nq^{m_t-2}} \sum_{\tau \in \mathcal{I}_{s-1,n}} \prod_{t=1}^{s-2} \mathbb{E}|H_{\tau(e_t)}^{m_t}| \times \frac{1}{q^{m_{s-1}-2}}
$$

$$
\leqslant \prod_{t=s}^{a} \frac{\kappa}{nq^{m_t-2}} \sum_{\tau \in \mathcal{I}_{s-2,n}} \prod_{t=1}^{s-3} \mathbb{E}|H_{\tau(e_t)}^{m_t}| \times \frac{1}{q^{m_{s-2}-2}} \frac{1}{q^{m_{s-1}-2}}
$$

$$
\leqslant \cdots
$$

$$
\leqslant \prod_{t=s}^{a} \frac{\kappa}{nq^{m_t-2}} \sum_{\tau \in \mathcal{I}_{1,n}} \frac{1}{q^{m_1-2}} \cdots \frac{1}{q^{m_{s-2}-2}} \frac{1}{q^{m_{s-1}-2}}
$$

$$
= nq^{2a - \sum_{t=1}^{a} m_t} (\kappa/n)^{a-s+1}.
$$

We conclude the proof noting that $g = a - s + 1$ and $\sum_{t=1}^{a} m_t = 2\ell$. $\qquad\square$

Definition 5.4. We say that a pair (G, w) formed by a multigraph G and a path $w = w_0 w_{01} w_1 w_{12} \ldots w_{l-1 l} w_l$ is *normal* if

(i) $V(G) = \{w_0, \ldots, w_l\} = [s]$ where $s := |V(G)|$;

(ii) the vertices of $V(G)$ are visited in increasing order by w, i.e. if $w_i \notin \{w_0, \ldots, w_{i-1}\}$ then $w_i > w_1, \ldots, w_{i-1}$.

Clearly, each equivalence class of $\sim$ in $\mathcal{C}$ has a unique representative ξ such that (G_ξ, ξ) is normal. We denote $\mathcal{C}_0 := \{\xi \in \mathcal{C} : (G_\xi, \xi) \text{ normal}\}$. Thus, from (5.2) and Lemma 5.3 we deduce that

$$
\mathbb{E} \operatorname{Tr} B^\ell B^{\ell*} \leqslant n^2 \sum_{\xi \in \mathcal{C}_0} n^{-g(G_\xi)} \kappa^{g(G_\xi)} q^{2|E(G_\xi)|-2\ell}. \tag{5.5}
$$

Next, we introduce a parametrization of $\mathcal{C}_0$ obtained by deleting vertices of the graph G_ξ that have degree two. In this process the two exceptional vertices $\{\xi_0, \xi_\ell\}$ are not collapsed. This process will result in a multigraph, denoted by $U(\xi)$ below. A similar construction appears in [14]. We refer to Figure 1 for an illustration of the following construction.

Definition 5.5. Let G be a graph and $\mathcal{V} \subset V(G)$. Define the set

$$
\mathcal{I}_\mathcal{V}(G) := \{v \in V(G) \setminus \mathcal{V} : \deg(v) = 2\}.
$$

Define the set $\Sigma_\mathcal{V}(G)$ to be the set of paths $w = w_0 \cdots w_l$ in G such that $w_1, \ldots, w_{l-1}$ are pairwise distinct and belong to $\mathcal{I}_\mathcal{V}(G)$ and $w_0, w_l \notin \mathcal{I}_\mathcal{V}(G)$. We introduce an equivalence relation on $\Sigma_\mathcal{V}(G)$ by saying that $w_0 \cdots w_l$ and $w_l \cdots w_0$ are equivalent, and denote by $\Sigma'_\mathcal{V}(G) := \{[w] : w \in \Sigma_\mathcal{V}(G)\}$ the set of equivalence classes.

The next definition constructs a multigraph $\hat{G}_\xi$ from G_ξ, obtained by replacing every $[w] \in \Sigma'_{\{\xi_0, \xi_\ell\}}(G)$ with an edge of $E(\hat{G}_\xi)$.
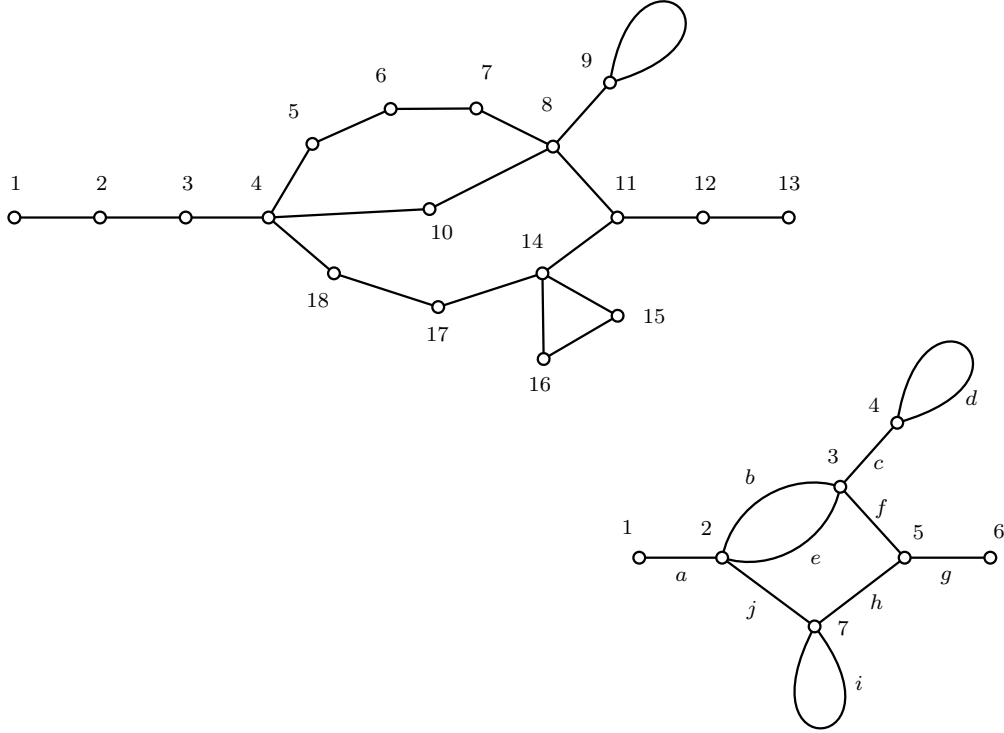
14

Figure 1: In the top diagram we draw the graph G_ξ associated with the path ξ = 1,2,3,4,5,6,7,8,9,9,8,10,4,5,6,7,8,9,9,8,11,12,13,12,11,14,15,16,14,17,18,4,10,8,11,14,16,15,14,17,18,4,3,2,1. Here $\ell = 22$. Note that G_ξ has a loop but no multiple edges. The number of crossings of each edge by ξ is 2, except for $\{8,9\}$ for which it is 4. Note also that (G_ξ, ξ) is normal. In the bottom diagram we draw the multigraph U associated with ξ. We label the vertices of U by $1, \ldots, 7$ and the edges of U by $a, \ldots, j$. The path in U associated with ξ is $\zeta = 1a2b3c4d4c3e2b3c4d4c3f5g6g5h7i7j2e3f5h7i7j2a1$. The number of crossings of each edge by ζ is 2, except for c for which it is 4. Note that ζ is normal in U. Here $\gamma = 6$ and the weights k of the edges of U are given by $k_a = 3$, $k_b = 4$, $k_c = 1$, $k_d = 1$, $k_f = 1$, $k_g = 2$, $k_h = 1$, $k_i = 3$, $k_j = 3$.

Definition 5.6. Let $\xi \in \mathcal{C}_0$. Set $V(\hat{G}_\xi) := V(G_\xi) \setminus \mathcal{I}_{\{\xi_0, \xi_\ell\}}(G_\xi)$ and $E(\hat{G}_\xi) := \Sigma'_{\{\xi_0, \xi_\ell\}}(G)$ with $\phi([w]) := \{w_0, w_l\}$ for each $[w] \in \Sigma'_{\{\xi_0, \xi_\ell\}}(G)$. We also assign to each edge $[w] \in E(\hat{G}_\xi)$ the *weight* $\hat{k}_w$ to be the length of the path w.

Let now $\xi \in \mathcal{C}_0$, which is a closed path $\xi_0 \xi_1 \cdots \xi_{2\ell}$ of length 2ℓ in the graph G_ξ. Because of the nonbacktracking condition in the definition of $\mathcal{C}$ at all vertices of G_ξ except ξ_0 and ξ_ℓ, we find that every pair $\xi_{i-1}\xi_i$ must be contained in a word $w \in \Sigma_{\{\xi_0, \xi_\ell\}}(G_\xi)$. By writing ξ as a concatenation of words from $\Sigma_{\{\xi_0, \xi_\ell\}}(G_\xi)$, we therefore conclude that the closed path $\xi = \xi_0 \xi_1 \cdots \xi_{2\ell}$ in the graph G_ξ gives rise to a closed path $\hat{\xi} = \hat{\xi}_0 \hat{\xi}_{01} \hat{\xi}_1 \hat{\xi}_{12} \cdots \hat{\xi}_{r-1r} \hat{\xi}_r$ on the multigraph $\hat{G}_\xi$. We stress the fundamental role of the nonbacktracking condition in the definition of $\mathcal{C}$ in the construction of $\hat{\xi}$; without it such a construction fails.

Summarizing, for any given $\xi \in \mathcal{C}_0$ we have constructed a triple $(\hat{G}_\xi, \hat{\xi}, \hat{k})$, where $\hat{G}_\xi$ is a multigraph, $\hat{\xi}$ is a closed path in $\hat{G}_\xi$, and $\hat{k} = (\hat{k}_e)_{e \in E(\hat{G}_\xi)}$ is the family of weights of the edges of $\hat{G}_\xi$.

Note that $\hat{\xi}$ and $\hat{G}_\xi$ are in general not normal in the sense of Defintion 5.4. We remedy this by setting τ to be the unique increasing bijection from $V(\hat{G}_\xi)$ to $\{1, \ldots, |V(\hat{G}_\xi)|\}$. Denote by $(U, \zeta, k) \equiv (U(\xi), \zeta(\xi), k(\xi))$ the triple obtained from the triple $(\hat{G}_\xi, \hat{\xi}, \hat{k})$ by relabelling the vertices using τ. By definition, $\xi_0 = \tau(\xi_0) = 1$. Moreover, we set $\gamma \equiv \gamma(\xi) := \tau(\xi_\ell)$.

We refer to Figure 1 for an illustration of the construction of (U, ζ, k). We now collect several basic properties of the mapping $\xi \mapsto (U, \zeta, k)$.

Lemma 5.7. *The mapping $\xi \mapsto (U, \zeta, k)$ satisfies the following properties.*

  (i) *The mapping $\xi \mapsto (U, \zeta, k)$ is injective on $\mathcal{C}_0$.*

 (ii) $g(U) = g(G_\xi)$.

(iii) *ζ is a closed path in the multigraph U. It is normal in U in the sense of Definition 5.4. In particular, $\zeta_0 = \zeta_{|\zeta|} = 1$.*

 (iv) *Every vertex of $V(U) \setminus \{1, \gamma\}$ has degree at least three. The vertices 1 and γ have degree at least one.*

  (v) $|E(G_\xi)| = \sum_{e \in E(U)} k_e$.

 (vi) $m_e(\zeta) \geqslant 2$ for all $e \in E(U)$ and $2\ell = \sum_{e \in E(U)} m_e(\zeta) k_e$.

Proof. All of these properties follow immediately from the construction of (U, ζ, k). For (i), we emphasize that the requirement $\xi \in \mathcal{C}_0$ is crucial, for a relabelling of the vertices of G_ξ will not change the resulting triple (U, ζ, k). $\qquad \square$

Having constructed the triple (U, ζ, k), we may now use it to estimate the right-hand side of (5.5). By property (i) above, it suffices to sum over (U, ζ, k) instead of ξ. We find

$$\mathbb{E} \operatorname{Tr} B^\ell B^{\ell *} \leqslant n^2 \sum_{(U, \zeta, k)} n^{-g(U)} q^{-\sum_{e \in E(U)} m_e(\zeta) k_e + 2 \sum_{e \in E(U)} k_e} \kappa^{g(U)}$$

$$= n^2 \sum_{(U, \zeta, k)} n^{-g(U)} q^{-\sum_{e \in E(U)} k_e (m_e(\zeta) - 2)} \kappa^{g(U)} \,,$$

where the sum ranges over all triples (U, ζ, k) obtained from all $\xi \in \mathcal{C}_0$. Since $k_e \geqslant 1$ and $m_e(\zeta) \geqslant 2$ for all $e \in E(U)$, we find

$$\sum_{e \in E(U)} k_e (m_e(\zeta) - 2) \geqslant \sum_{e \in E(U)} (m_e(\zeta) - 2) = |\zeta| - 2 |E(U)| \,.$$

Since $q \geqslant 1$, we therefore get

$$\mathbb{E} \operatorname{Tr} B^\ell B^{\ell *} \leqslant n^2 \sum_{(U, \zeta, k)} n^{-g(U)} q^{2 |E(U)| - |\zeta|} \kappa^{g(U)} \,.$$

Note that the summand does not depend on k. For fixed (U, ζ), we may therefore estimate the sum over k by estimating from above the number of families $k = (k_e)_{e \in E(U)}$ such that $k_e \geqslant 1$ for all $e \in E(U)$ and $\sum_{e \in E(U)} k_e m_e(\zeta) = 2\ell$. Since $m_e(\zeta) \geqslant 2 \geqslant 1$, this is certainly bounded by the

number of families $k = (k_e)_{e \in E(U)}$ such that $k_e \geqslant 1$ for all $e \in E(U)$ and $\sum_{e \in E(U)} k_e = 2\ell$, which is equal to

$$\binom{2\ell - 1}{|E(U)| - 1} \leqslant \left(\frac{6\ell}{|E(U)|}\right)^{|E(U)|}.$$

We conclude that

$$\mathbb{E}\operatorname{Tr} B^\ell B^{\ell*} \leqslant n^2 \sum_{(U,\zeta)} \left(\frac{6\ell}{|E(U)|}\right)^{|E(U)|} n^{-g(U)} q^{2|E(U)| - |\zeta|} \kappa^{g(U)},$$

where the sum ranges over pairs (U, ζ) obtained from all $\xi \in \mathcal{C}_0$. We estimate this sum using the following bounds on $|E(U)|$.

Lemma 5.8. *For U as above we have*

$$|E(U)| \leqslant 3g(U) + 1,$$

and

$$|V(U)| \leqslant 2g(U) + 2.$$

Proof. In order to prove the upper bound on $|E(U)|$, we write

$$2|E(U)| = \sum_{v \in V(U)} \deg(v) \geqslant 2 + 3(|V(U)| - 2),$$

where we used the fact that each vertex in $V(U) \setminus \{1, \gamma\}$ has degree at least 3, and $\{1, \gamma\}$ have degree at least one. This gives

$$|V(U)| \leqslant \frac{2}{3}|E(U)| + \frac{4}{3},$$

which implies that $g(U) \geqslant \frac{|E(U)| - 1}{3}$ and $g(U) \geqslant \frac{|V(U)|}{2} - 1$. $\qquad\qquad\square$

Using Lemma 5.8 we conclude

$$\mathbb{E}\operatorname{Tr} B^\ell B^{\ell*} \leqslant n^2 \ell q^2 \sum_{(U,\zeta)} \left(\frac{12\ell}{g(U) + 1}\right)^{3g(U)} n^{-g(U)} q^{6g(U) - |\zeta|} \kappa^{g(U)},$$

where we used that $g(U) \leqslant \ell$.

Next, we claim that the number of pairs (U, ζ) such that $|E(U)| \leqslant e$, $|V(U)| \leqslant v$ and ζ has lenght m is at most

$$e^m v^e.$$

Indeed, we may build the path ζ directly as follows: at each of the m steps of the path ζ, we choose an edge numbered from one to e and, for each newly visited edge, we attribute an end vertex among v possible choices. Since (U, ζ) is normal and ζ visits all edges of U, this characterizes uniquely the pair (U, ζ). Using Lemma 5.8, we therefore find that the number of pairs (U, ζ) such that U has genus g and ζ has length m is bounded by

$$(3g + 1)^m (2g + 2)^{3g+1}.$$

Putting everything together, we find

$$\mathbb{E} \operatorname{Tr} B^\ell B^{\ell*} \leqslant n^2 \ell q^2 \sum_{g=0}^{\ell} \sum_{m=1}^{2\ell} (3g+1)^m (2g+2)^{3g+1} \left(\frac{24\ell}{2g+2}\right)^{3g} n^{-g} q^{6g-m} \kappa^g$$

$$\leqslant Cn^2 \ell q^2 \sum_{m=1}^{2\ell} \left(\frac{1}{q}\right)^m + Cn^2 \ell^2 q^2 \sum_{g=1}^{\ell} \left(\frac{C\kappa \ell^3 q^6}{n}\right)^g \sum_{m=1}^{2\ell} \left(\frac{4g}{q}\right)^m, \tag{5.6}$$

where we used that $g \leqslant \ell$. Since $q \geqslant 1$, the first term of (5.6) is bounded by $Cn^2 \ell^2 q^2$. For any $x > 0$ we have $\sum_{m=1}^{2\ell} x^m \leqslant 2\ell(1 + x^\ell)$. Thus, the second term of (5.6) is estimated by

$$Cn^2 \ell^3 q^2 \sum_{g=1}^{\ell} \left(\frac{C\kappa \ell^3 q^6}{n}\right)^g + Cn^2 \ell^3 q^2 \sum_{g=1}^{\ell} \left(\frac{C\kappa \ell^3 q^6}{n}\right)^g \left(\frac{4g}{q}\right)^{2\ell}. \tag{5.7}$$

Now, assume that

$$n \geqslant 2C\kappa \ell^3 q^6, \qquad \ell \leqslant \frac{q}{8} \log \frac{n}{C\kappa \ell^3 q^6} \tag{5.8}$$

(we will then check that (5.1) implies (5.8)). By the assumption (5.8), the first term of (5.7) is estimated by $Cn^2 \ell^3 q^2$. The second term of (5.7) may be written as

$$Cn^2 \ell^3 q^2 \sum_{g=1}^{\ell} \exp\left[-g \log \frac{n}{C\kappa \ell^3 q^6} + 2\ell \log \frac{4g}{q}\right].$$

The argument of the exponential is maximized for

$$g = 2\ell \Big/ \log \frac{n}{C\kappa \ell^3 q^6}.$$

Plugging this back in, we find that if (5.8) holds then this maximum is reached for $g \leqslant 4q$ and we deduce that $\mathbb{E} \operatorname{Tr} B^\ell B^{*\ell} \leqslant Cn^2 \ell^4 q^2$.

What remains, therefore, is to show that (5.1) with large enough C_0 and small enough c_0 implies (5.8). From (5.1) we find

$$n \geqslant \left(\frac{\kappa \ell^3 q^6}{c_0^3}\right)^{\frac{1}{1-3\delta}}, \tag{5.9}$$

which implies the first estimate of (5.8) for small enough c_0. Moreover, (5.9) yields

$$n^{3\delta} \leqslant \frac{c_0^3 n}{\kappa \ell^3 q^6}.$$

We conclude from (5.1) that

$$\ell \leqslant c_0 \delta q \log n = \frac{c_0 q}{3} \log n^{3\delta} \leqslant \frac{c_0 q}{3} \log\left(\frac{c_0^3 n}{\kappa \ell^3 q^6}\right),$$

which implies the second estimate of (5.8) for small enough c_0. This concludes the proof of Proposition 5.1.

18

5.2. Proof of Theorem 2.5. This is a simple application of Proposition 5.1 and Markov's inequality. We have

$$\rho(B) = \rho(B^\ell)^{1/\ell} \leqslant \|B^\ell\|^{1/\ell} = \|B^\ell B^{*\ell}\|^{1/(2\ell)} \leqslant (\mathrm{Tr}(B^\ell B^{*\ell}))^{1/(2\ell)} \,. \tag{5.10}$$

Thus we get

$$\mathbb{P}(\rho(B) > 1 + \varepsilon) \leqslant \mathbb{P}\big(\mathrm{Tr}\, B^\ell B^{*\ell} > (1 + \varepsilon)^{2\ell}\big) \leqslant \frac{\mathbb{E}\,\mathrm{Tr}\, B^\ell B^{*\ell}}{(1 + \varepsilon)^{2\ell}} \,.$$

Choosing $\delta := \frac{1}{30}$ in Proposition 5.1, we find by assumption of Theorem 2.5 that (5.1) holds with $\ell := \lceil \frac{c}{2} q \log n \rceil$, with some universal constant $c > 0$. We therefore find from Proposition 5.1

$$\mathbb{P}(\rho(B) > 1 + \varepsilon) \leqslant Cn^2 q^6 (\log n)^4 (1 + \varepsilon)^{-cq \log n} \leqslant Cn^{3 - cq \log(1+\varepsilon)} \,,$$

as claimed. This concludes the proof of Theorem 2.5.

5.3. Proof of Theorem 2.7. We start by proving the following intermediate result.

Proposition 5.9. *There are universal constants $C, c > 0$ such that the following holds. Suppose that H satisfies Assumption 2.4. Then for $1 \leqslant q \leqslant n^{1/10} \kappa^{-1/9}$ and $\delta \geqslant 0$, we have*

$$\mathbb{P}\left(\|H\| \geqslant 2 + C\delta + \frac{C}{q} \right) \leqslant Cn^{3 - cq \log(1 + \sqrt{\delta})} + ne^{-q^2 h(\delta \vee \delta^2)} \,, \tag{5.11}$$

where $h(\delta) := (1 + \delta) \log(1 + \delta) - \delta$.

Proof. The proof is a combination of Corollary 2.3 and Theorem 2.5. By assumption, we have $\|H\|_{1 \to \infty} \leqslant 1/q$. We estimate $\|H\|_{2 \to \infty}$ using Bennett's inequality [9, Theorem 2.9]. Fix $i \in [n]$ and define the independent random variables $X_1, \ldots, X_n$, where $X_j := |H_{ij}|^2$. By assumption on H we have $X_j \leqslant 1/q^2$ and $\sum_j \mathbb{E} X_j^2 \leqslant 1/q^2$. Since $\sum_j \mathbb{E} X_j \leqslant 1$, we conclude from Bennett's inequality that

$$\mathbb{P}\left(\sum_j |H_{ij}|^2 \geqslant 1 + t \right) \leqslant e^{-q^2 h(t)} \,. \tag{5.12}$$

By a union bound, we therefore deduce that $\mathbb{P}(\|H\|_{2 \to \infty} \geqslant 1 + t) \leqslant \mathbb{P}(\|H\|_{2 \to \infty} \geqslant \sqrt{1 + t \vee t^2}) \leqslant ne^{-q^2 h(t \vee t^2)}$. We apply this last statement to $t = \delta$. The claim now follows from Corollary 2.3 and Theorem 2.5 applied to $\varepsilon^2 = \delta/C$. $\qquad \square$

We are now ready to prove Theorem 2.7. We begin by noting that without loss of generality we can assume $q \geqslant 1$. Indeed, if Theorem 2.7 is already established for $q \geqslant 1$ then in the case $q < 1$ we can apply it to $H' = H/(1/q)$, with corresponding parameters $q' = 1$ and $\kappa' = \kappa q^2 \leqslant \kappa$, and the claim follows easily.

From now on we therefore assume that $q \geqslant 1$. Let us first prove (2.5). We claim that for large enough K and any n, q, κ satisfying

$$1 \leqslant q \leqslant n^{1/10} \kappa^{-1/9} \quad \text{and} \quad \kappa \geqslant 1 \,, \tag{5.13}$$

we have

$$\mathbb{P}\big(\big|\|H\| - \mathbb{E}\|H\|\big| \leqslant \delta\big) + \mathbb{P}\big(\|H\| \leqslant 2 + 2C\delta\big) > 1 \,, \tag{5.14}$$

19

where C is the constant in Proposition 5.9 and

$$\delta := \frac{K\eta}{\sqrt{1 \vee \log \eta}}\,, \qquad \eta := \frac{\sqrt{\log n}}{q}\,.$$

Supposing for now that (5.14) has been proved, we find that the intersection of both events from the left-hand side of (5.14) is nonempty, and hence $\mathbb{E}\|H\| \leqslant 2 + (2C+1)\delta$, which concludes the proof of (2.5).

What remains, therefore, is to prove (5.14). We first remark that, with $c_0 = 2/\mathrm{e}$,

$$q\delta = \frac{K\sqrt{\log n}}{\sqrt{1 \vee \log \eta}} = \frac{K\eta q}{\sqrt{1 \vee \log \eta}} \geqslant c_0 K \tag{5.15}$$

uniformly over all $n \geqslant 2$ and $q \geqslant 1$ (distinguish $\eta \leqslant \mathrm{e}$ and $\eta \geqslant \mathrm{e}$). We shall assume that $c_0 K \geqslant 1$. By (2.4) and (5.11), it suffices to prove that, by choosing K large enough, the three numbers

$$a_1 := q^2\delta^2\,, \qquad a_2 := q\log(1 + \sqrt{\delta})\,, \qquad a_3 := q^2 h(\delta \vee \delta^2) - \log n \tag{5.16}$$

can be made arbitrarily large, uniformly under the conditions (5.13). Indeed, the term C/q from the left-hand side of (5.11) is bounded by $C\delta/(c_0 K) \leqslant C\delta$ by (5.15).

First, from (5.15), $a_1 \geqslant (c_0 K)^2$. Hence, a_1 can be chosen arbitrarily large if K is large enough. Similarly, to prove that a_2 can be chosen arbitrarily large, we note that the function $f(x) = \log(1 + \sqrt{x})/x$ is positive and decreasing on $(0,1)$. Hence, from (5.15), we have

$$a_2 \geqslant (c_0 K)f(c_0 K/q) \geqslant (c_0 K)f(c_0 K) = \log(1 + \sqrt{c_0 K})\,,$$

as desired.

To prove that a_3 can be chosen arbitrarily large, we consider the cases $\eta \leqslant \mathrm{e}$ and $\eta \geqslant \mathrm{e}$ separately. For $\eta \leqslant \mathrm{e}$, we use $h(x) \geqslant c(x^2 \wedge x)$ for all $x \geqslant 0$ and for some universal constant $c > 0$. Hence $h(\delta \vee \delta^2) \geqslant c\delta^2$ and

$$a_3 \geqslant cq^2\delta^2 - \log n = (cK^2 - 1)\log n\,,$$

as desired. For $\eta \geqslant \mathrm{e}$, we use the refined bound $h(x) \geqslant c(x^2 \wedge x)(1 \vee \log x)$ for all $x \geqslant 0$. Hence $h(\delta \vee \delta^2) \geqslant c\delta^2 \log \delta^2$ and, since $\log \delta^2 \geqslant c' \log \eta$ for some constant $c' > 0$,

$$a_3 \geqslant cc'q^2\delta^2 \log \eta - \log n = (cc'K^2 - 1)\log n\,,$$

as desired. This concludes the proof of (2.5).

Finally, we prove (2.7). First, Jensen's inequality implies that $\mathbb{E}\|H\|_{2\to\infty} \geqslant 1$. Note also the triangle inequality gives $\big|\|X\|_{2\to\infty} - \|Y\|_{2\to\infty}\big| \leqslant \|X - Y\|_{2\to\infty} \leqslant \sqrt{\sum_{ij}|X_{ij} - Y_{ij}|^2}$, and the function $X \mapsto \|X\|_{2\to\infty}$ is separately convex in the entries of X (as a maximum (2.2) of separately convex functions). Hence, we may apply Talagrand's concentration inequality in the form of [9, Theorem 6.10]. We find that there exists a universal constant $c > 0$ such that for any $t > 0$,

$$\mathbb{P}\big(\big|\|H\|_{2\to\infty} - \mathbb{E}\|H\|_{2\to\infty}\big| \geqslant t/q\big) \leqslant 2e^{-ct^2}\,. \tag{5.17}$$

This last inequality and (2.4) implies that if we prove that the event

$$\left\{\|H\| \leqslant \|H\|_{2\to\infty}\left(2 + \frac{C}{q}\right)\right\} \tag{5.18}$$

20

has probability at least $1/2$, then

$$\mathbb{E}\|H\| \leqslant \mathbb{E}\|H\|_{2\to\infty}\left(2 + \frac{C}{q}\right) + \frac{2t}{q} \leqslant \mathbb{E}\|H\|_{2\to\infty}\left(2 + \frac{C + 2t}{q}\right),$$

where t is such that $4e^{-ct^2} < 1/2$, the constant c is as in (2.4)–(5.17), and we used that $\mathbb{E}\|H\|_{2\to\infty} \geqslant 1$. Therefore, it suffice to prove that (5.18) holds with probability at least $1/2$. With probability at least $3/4$, for some universal constant $t > 0$, we have

$$\rho(B) \leqslant 1 + \frac{t}{q}.$$

and, from (5.17) with probability at least $3/4$,

$$\|H\|_{2\to\infty} \geqslant \mathbb{E}\|H\|_{2\to\infty} - \frac{t}{q} \geqslant 1 - \frac{t}{q}.$$

We deduce from Corollary 2.3 that with probability at least $1/2$, if $q \geqslant 2t$,

$$\mathbb{E}\|H\| \leqslant 2\|H\|_{2\to\infty} + C\left(\frac{(2t/q)^2}{1 - t/q} + \frac{1}{q}\right) \leqslant 2\|H\|_{2\to\infty} + \frac{C'}{q} \leqslant 2\|H\|_{2\to\infty} + \frac{\|H\|_{2\to\infty}}{1 - C/q}\frac{C'}{q}.$$

Adjusting the constant C, we obtain (5.18). This concludes the proof of (2.7), and hence also of Theorem 2.7.

6. Non-Hermitian matrices: proof of Theorem 2.11

The main estimate of this section is the following result.

Proposition 6.1. *Let $H \in M_n(\mathbb{C})$ be a random matrix whose entries $(H_{ij})_{1\leqslant i,j\leqslant n}$ are independent mean-zero random variables. Moreover, suppose that there exist $q \geqslant 3$ and $\kappa \geqslant 1$ such that (2.3) holds. There exist universal constants $c_0, C_0 > 0$ such that the following holds. If $\ell \geqslant 1$, $q \geqslant 1$, and $\delta \in (0, 1/3)$ satisfy (5.1) then $\mathbb{E}\operatorname{Tr} H^\ell H^{*\ell} \leqslant C_0 n\ell^4 q^2$.*

Once Proposition 6.1 is proved, Theorem 2.11 follows by the argument of Section 5.2.

The rest of this section is devoted to the proof of Proposition 6.1. The proof of Proposition 6.1 is similar to that of Proposition 5.1. Essentially, the nonbacktracking condition in the definition of B is replaced by the independence of the entries of H. We compute

$$\operatorname{Tr} H^\ell H^{*\ell} = \sum_{\xi} H_{\xi_0^1\xi_1^1} H_{\xi_1^1\xi_2^1} \cdots H_{\xi_{\ell-1}^1\xi_\ell^1} \overline{H}_{\xi_0^2\xi_1^2} \overline{H}_{\xi_1^2\xi_2^2} \cdots \overline{H}_{\xi_{\ell-1}^2\xi_\ell^2},$$

where the sum ranges over all $\xi = (\xi_i^\nu : i \in \{0, 1, \ldots, \ell\}, \nu \in \{1, 2\}) \in [n]^{2\ell+2}$ such that $(\xi_0^1, \xi_\ell^1) = (\xi_0^2, \xi_\ell^2)$.

Because the entries of H are independent and have mean zero, we find

$$\mathbb{E}\operatorname{Tr} H^\ell H^{*\ell} = \mathbb{E}\sum_{\xi \in \mathcal{C}} H_{\xi_0^1\xi_1^1} H_{\xi_1^1\xi_2^1} \cdots H_{\xi_{\ell-1}^1\xi_\ell^1} \overline{H}_{\xi_0^2\xi_1^2} \overline{H}_{\xi_1^2\xi_2^2} \cdots \overline{H}_{\xi_{\ell-1}^2\xi_\ell^2}, \tag{6.1}$$

where $\mathcal{C}$ is the set of pairs $\xi = (\xi^1, \xi^2) \in [n]^{2\ell+2}$ satisfying $(\xi_0^1, \xi_\ell^1) = (\xi_0^2, \xi_\ell^2)$ and

$$\sum_{\nu=1}^{2} \sum_{i=1}^{\ell-1} \mathbf{1}_{(\xi_{i-1}^\nu, \xi_i^\nu)=(a,b)} \neq 1$$

for all $a, b \in [n]$.

As in Section 5.1, we estimate the right-hand side of (6.1) using graphs. In contrast to Section 5.1, in this section we always use directed graphs and multigraphs. The following definitions closely mirror those from Section 5.1. By definition, a (vertex-labelled) *directed multigraph* $G = (V(G), E(G), \phi)$ consists of two finite sets, the set of *vertices* $V(G)$ and the set of *edges* $E(G)$, and a map $\phi = (\phi_+, \phi_-)$ from $E(G)$ to the ordered pairs of elements of $V(G)$. The edge $e \in E(G)$ is a *loop* if $\phi_+(e) = \phi_-(e)$. We define the *outdegree* $\deg_+(v) := \sum_{e \in E(G)} \mathbf{1}_{\phi_+(e)=v}$, the *indegree* $\deg_-(v) := \sum_{e \in E(G)} \mathbf{1}_{\phi_-(e)=v}$, and the *degree* $\deg(v) := \deg_+(v) + \deg_-(v)$ of a vertex $v \in V(G)$. As always, the *genus* of G is $g(G) := |E(G)| - |V(G)| + 1$. A *path of length* $l \geqslant 1$ *in* G is a word $w = w_0 w_{01} w_1 w_{12} \ldots w_{l-1l} w_l$ such that $w_0, w_1, \ldots, w_l \in V(G)$, $w_{01}, w_{12}, \ldots, w_{l-1l} \in E(G)$, and $\phi(w_{i-1i}) = (w_{i-1}, w_i)$ for $i = 1, \ldots, l$. We denote the length l of w by $|w| := l$. For $e \in E(G)$ we define the *number of crossings of* e *by* w to be $m_e(w) := \sum_{i=1}^{l} \mathbf{1}_{w_{i-1i}=e}$.

Moreover, the directed multigraph G is called simply a *directed graph* if ϕ is injective, i.e. there are no multiple edges. (Note that in our convention a directed graph may have loops.) For a directed graph G we may and shall identify $E(G)$ with a set of ordered pairs of $V(G)$, simply identifying e and $\phi(e)$. Similarly, we identify a path w with the reduced word $w_0 w_1 \ldots w_l$ only containing the vertices, since we must have $w_{i-1i} = (w_{i-1}, w_i)$ for $i = 1, \ldots, l$.

The following definition is analogous to Definition 5.2.

Definition 6.2. For $\xi \in \mathcal{C}$ we define the directed graph G_ξ as

$$V(G_\xi) := \left\{ \xi_i^\nu : i \in \{0, \ldots, \ell\}, \nu \in \{1, 2\} \right\}, \qquad E(G_\xi) := \left\{ (\xi_{i-1}^\nu, \xi_i^\nu) : i \in \{1, \ldots, \ell\}, \nu \in \{1, 2\} \right\}.$$

Thus, $\xi^1 \equiv \xi_0^1 \xi_1^1 \cdots \xi_\ell^1$ and $\xi^2 \equiv \xi_0^2 \xi_1^2 \cdots \xi_\ell^2$ are paths in G_ξ.

As in Section 5.1, we introduce an equivalence relation on $\mathcal{C}$ by saying that $\xi, \tilde{\xi} \in \mathcal{C}$ are equivalent if and only if there exists a permutation τ of $[n]$ such that $\tau(\xi_i^\nu) = \tilde{\xi}_i^\nu$ for all $i = 0, \ldots, 2\ell$ and $\nu = 1, 2$. The following result is analogous to Lemma 5.3.

Lemma 6.3. *For any* $\bar{\xi} \in \mathcal{C}$ *we have*

$$\mathbb{E} \sum_{\xi \in [\bar{\xi}]} H_{\xi_0^1 \xi_1^1} H_{\xi_1^1 \xi_2^1} \cdots H_{\xi_{\ell-1}^1 \xi_\ell^1} \overline{H}_{\xi_0^2 \xi_1^2} \overline{H}_{\xi_1^2 \xi_2^2} \cdots \overline{H}_{\xi_{\ell-1}^2 \xi_\ell^2} \leqslant n^{1-g(G_{\bar{\xi}})} \kappa^{g(G_{\bar{\xi}})} q^{2|E(G_{\bar{\xi}})|-2\ell}.$$

Proof. The proof is almost identical to that of Lemma 5.3. Using its notations, we suppose without loss of generality that $V(G_{\bar{\xi}}) = [s]$ and pick a (directed) spanning tree T of G_ξ such that, for all $t = 2, \ldots, s$ the subgraph $T|_{[t]}$ is a spanning tree of $[t]$, and enumerate the edges $e_1, \ldots, e_{s-1}$ of T so that for all $t = 1, \ldots, s-2$, the edge e_t is an edge of $T|_{[t+1]}$ that is not an edge of $T|_{[s]}$ and $\phi_-(e_t) = t$. Then the argument from the proof of Lemma 5.3 carries over verbatim, using the estimates (5.3) and (5.4), which also hold under the assumptions of Proposition 6.1. $\qquad\square$

Next, we define normal paths in analogy to Definition 5.4.

Definition 6.4. For $\nu = 1, 2$, let $w^\nu = w^\nu_0 w^\nu_{01} w^\nu_1 w^\nu_{12} \ldots w^\nu_{l^\nu - 1 l^\nu} w^\nu_{l^\nu}$ be a path in a directed multi-graph G. We say that (w^1, w^2) is *normal in* G if

(i) $V(G) = \{w^1_0, \ldots, w^1_{l^1}, w^2_0, \ldots, w^2_{l^2}\} = [s]$ where $s := |V(G)|$;

(ii) the vertices of $V(G)$ are visited in increasing order first by w^1 and then by w^2, i.e. if $w^1_i \notin \{w^1_0, \ldots, w^1_{i-1}\}$ then $w^1_i > w^1_1, \ldots, w^1_{i-1}$, and if $w^2_i \notin \{w^1_0, \ldots, w^1_{l^1}, w^2_0, \ldots, w^2_{i-1}\}$ then $w^1_i > w^1_0, \ldots, w^1_{l^1}, w^2_0, \ldots, w^2_{i-1}$.

Each equivalence class of $\sim$ in $\mathcal{C}$ has a unique representative $\xi = (\xi^1, \xi^2)$ that is normal in G_ξ. We denote $\mathcal{C}_0 := \{\xi \in \mathcal{C} : \xi \text{ normal in } G_\xi\}$. Thus, from (6.1) and Lemma 6.3 we deduce that

$$\mathbb{E} \operatorname{Tr} H^\ell H^{\ell *} \leqslant n \sum_{\xi \in \mathcal{C}_0} n^{-g(G_\xi)} \kappa^{g(G_\xi)} q^{2|E(G_\xi)| - 2\ell} . \tag{6.2}$$
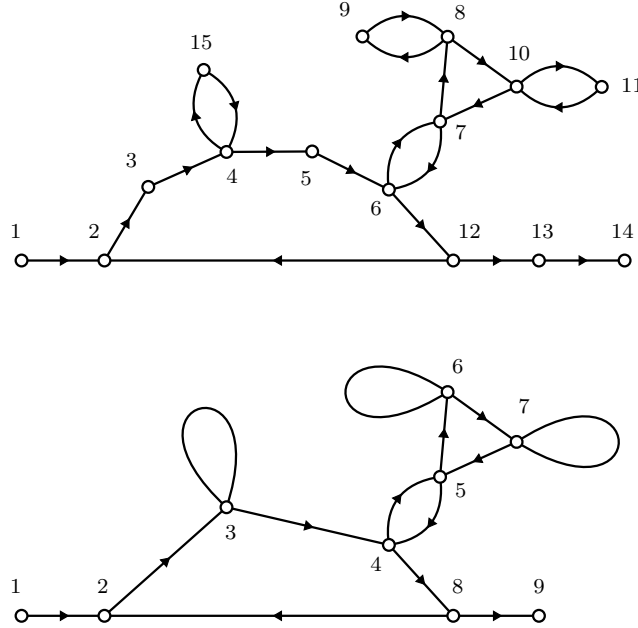

Figure 2: In the top diagram we draw the graph G_ξ associated with the pair $\xi = (\xi^1, \xi^2)$, where $\xi^1 = $ 1,2,3,4,5,6,7,8,9,8,9,8,10,11,10,7,8,10,11,10,7,6,7,6,12,13,14 and $\xi^2 = $ 1,2,3,4,15,4,5,6,12,2,3,4,15,4,15,4,5,6,12,2,3,4,5,6,12,13,14. Here $\ell = 26$. The total number of cross-ings by ξ^1 and ξ^2 of the edges $(2, 3)$, $(3, 4)$, $(4, 5)$, $(5, 6)$, $(6, 12)$, $(12, 2)$, $(4, 15)$, $(15, 4)$ is three, and of all other edges two. Note that ξ is normal in G_ξ. In the bottom diagram we draw the directed multigraph U associated with ξ, which in this example is just a directed graph (i.e. it has no multiple edges). The paths ζ^1, ζ^2 in U associated with ξ are $\zeta^1 = $ 12345666775677545489 and $\zeta^2 = $ 12334823334823489. The pair (ζ^1, ζ^2) is also normal in U. Here $\gamma = 9$.

We now introduce a parametrization (U, ζ, k) of $\mathcal{C}_0$ obtained by deleting vertices of G_ξ that have degree two, except the vertices $1 = \xi^1_0 = \xi^2_0$ and $\xi^1_\ell = \xi^2_\ell$. The construction follows verbatim that of Section 5.1, whereby all graphs and multigraphs are directed. Note that every vertex of the directed graph G_ξ that is not 1 or $\xi^1_\ell = \xi^2_\ell$ and has degree two has indegree one and outdegree one. More formally, we define $\mathcal{I}(\xi) := \{v \in V(G_\xi) \setminus \{1, \xi^1_\ell = \xi^2_\ell\}\}$ as well as $\Sigma(\xi)$ to be the set

of directed paths $w = w_0 \ldots w_l$ in G_ξ such that $w_1, \ldots, w_{l-1} \in \mathcal{I}(\xi)$ and $w_0, w_l \in \mathcal{I}(\xi)$. Then we define $\hat{G}_\xi$ to be the directed multigraph obtained from G_ξ by replacing each directed path $w \in \Sigma(\xi)$ by a directed edge from w_0 to w_l. We denote the resulting paths in $\hat{G}_\xi$ associated with ξ^1, ξ^2 by $\hat{\xi}^1, \hat{\xi}^2$, and the length of the path w associated with $e \in E(\hat{G}_\xi)$ by $\hat{k}_e$. Applying a suitable bijection $\tau : V(\hat{G}_\xi) \to [|V(\hat{G}_\xi)|]$, we obtain the triple (U, ζ, k), where U is a directed multigraph, $\zeta = (\zeta^1, \zeta^2)$ is a pair of paths normal in U of lengths r^1, r^2 satisfying $1 = \zeta_0^1 = \zeta_0^2$ and $\gamma := \zeta_{r^1}^1 = \zeta_{r^2}^2$, and $k = (k_e)_{e \in E(U)}$ is the family of weights of the edges of U. See Figure 2 for an illustration of the construction of (U, ζ, k). As in Section 5.1 every vertex in $V(U) \setminus \{1, \gamma\}$ has degree at least three. The remainder of the proof now follows to the letter the argument from Section 5.1 starting on page 16. This concludes the proof of Proposition 6.1.

References

[1] N. Alon. Spectral techniques in graph algorithms (invited paper). In *LATIN'98: theoretical informatics (Campinas, 1998)*, volume 1380 of *Lecture Notes in Comput. Sci.*, pages 206–215. Springer, Berlin, 1998.

[2] A. S. Bandeira and R. van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Ann. Prob.*, 44(4):2479–2506, 2016.

[3] F. Benaych-Georges, C. Bordenave, and A. Knowles. Largest eigenvalues of sparse Erdős-Rényi random graphs. arXiv:1704.02953, 2017, to appear in Ann. Prob.

[4] J. Alt, R. Ducatez, and A. Knowles. Extremal eigenvalues of critical Erdős-Rényi graphs. arXiv:1905.03243, 2019.

[5] F. Benaych-Georges and A. Knowles. Lectures on the local semicircle law for wigner matrices. *Panoramas et Synthèses*, 53, Société Mathématique de France, 2016.

[6] R. Bhatia. *Matrix analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.

[7] C. Bordenave. A new proof of Friedman's second eigenvalue theorem and its extension to random lifts. arXiv:1502.04482, 2015, to appear in Ann. Sci. Éc. Norm. Sup.

[8] C. Bordenave, M. Lelarge, and L. Massoulié. Non-backtracking spectrum of random graphs: community detection and non-regular ramanujan graphs. *Ann. Prob.*, 46(1):1–71, 2018.

[9] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.

[10] F. Chung and M. Radcliffe. On the spectra of general random graphs. *Electron. J. Combin.*, 18(1):Paper 215, 14, 2011.

[11] F. R. K. Chung. *Spectral graph theory*, volume 92 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 1997.

[12] C. Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.*, 7:1–46, 1970.

[13] U. Feige and E. Ofek. Spectral techniques applied to sparse random graphs. *Random Structures Algorithms*, 27(2):251–275, 2005.

[14] J. Friedman. On the second eigenvalue and random walks in random d-regular graphs. *Combinatorica*, 11(4):331–362, 1991.

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

[16] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)*, 43(4):439–561 (electronic), 2006.

[17] A. Khorunzhy. Sparse random matrices: spectral edge and statistics of rooted trees. *Adv. in Appl. Probab.*, 33(1):124–140, 2001.

[18] A. Khorunzhy, B. Khoruzhenko, L. Pastur, and M. Shcherbina. *Large-n limit in statistical mechanics and the spectral theory of disordered systems*, volume 15 of *Phase Transitions and Critical Phenomena*. Academic Press, London, 1992.

[19] M. Kotani and T. Sunada. Zeta functions of finite graphs. *J. Math. Sci. Univ. Tokyo*, 7(1):7–25, 2000.

[20] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang. Spectral redemption: clustering sparse networks. *Proceedings of the National Academy of Sciences*, (110(52)):20935–20940, 2013.

[21] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang. Spectral redemption in clustering sparse networks. *Proc. Natl. Acad. Sci. USA*, 110(52):20935–20940, 2013.

[22] R. Latała, R. van Handel, and P. Youssef. The dimension-free structure of nonhomogeneous random matrices. *Invent. Math.*, 214(3):1031–1080, 2018.

[23] A. Saade, F. Krzakala, and L. Zdeborova. Spectral clustering of graphs with the bethe hessian. *Advances in Neural Information Processing Systems*, 27:5520, 2014.

[24] S. Sodin. Random matrices, nonbacktracking walks, and orthogonal polynomials. *J. Math. Phys.*, 48(12):123503, 21, 2007.

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

[26] Y. Watanabe and K. Fukumizu. Graph zeta function in the Bethe free energy and loopy belief propagation. In *Advances in Neural Information Processing Systems*, pages 2017–2025, 2009.

[27] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah. On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab. Theory Related Fields*, 78(4):509–521, 1988.

Florent Benaych-Georges, MAP 5 (CNRS UMR 8145) - Université Paris Descartes, 45 rue des Saints-Pères 75270 Paris cedex 6, France. Email: florent.benaych-georges@parisdescartes.fr.

Charles Bordenave, Institut de Mathématiques de Marseille (CNRS UMR 7373) - Aix-Marseille Université, 163 Avenue de Luminy 13009 Marseille. France. Email: charles.bordenave@univ-amu.fr.

Antti Knowles, University of Geneva, Section of Mathematics, 2-4 Rue du Lièvre, 1211 Genève 4, Switzerland. Email: antti.knowles@unige.ch.