Local law and complete eigenvector delocalization for supercritical Erdős–Rényi graphs

Yukun He[*]   Antti Knowles[†]   Matteo Marcozzi[‡]

January 14, 2019

We prove a local law for the adjacency matrix of the Erdős-Rényi graph $G(N,p)$ in the supercritical regime $pN \geqslant C \log N$ where $G(N,p)$ has with high probability no isolated vertices. In the same regime, we also prove the complete delocalization of the eigenvectors. Both results are false in the complementary subcritical regime. Our result improves the corresponding results from [11] by extending them all the way down to the critical scale $pN = O(\log N)$.

A key ingredient of our proof is a new family of multilinear large deviation estimates for sparse random vectors, which carefully balance mixed ℓ^2 and ℓ^∞ norms of the coefficients with combinatorial factors, allowing us to prove strong enough concentration down to the critical scale $pN = O(\log N)$. These estimates are of independent interest and we expect them to be more generally useful in the analysis of very sparse random matrices.

1. Introduction

Let $\mathcal{A} \in \{0,1\}^{N \times N}$ be the adjacency matrix of the Erdős-Rényi random graph $G(N,p)$, where $p \equiv p_N \in (0,1)$. That is, $\mathcal{A}$ is real symmetric, and its upper-triangular entries are independent Bernoulli random variables with mean p. The Erdős-Rényi graph exhibits a phase transition in its connectivity around the critical expected degree $pN = \log N$. Indeed, for fixed $\varepsilon > 0$, if $pN \geqslant (1+\varepsilon)\log N$ then $G(N,p)$ is with high probability connected, and if $pN \leqslant (1-\varepsilon)\log N$ then $G(N,p)$ has with high probability isolated vertices. See e.g. [21, Chapter 5] for a clear treatment. The aim of this article is to investigate the spectral and eigenvector properties of $G(N,p)$ in the supercritical regime, where $G(N,p)$ is connected with high probability.

Our main result is a *local law* for the adjacency matrix $\mathcal{A}$ in the supercritical regime $C \log N \leqslant pN \ll N$, where C is some universal constant. In order to describe it, it is convenient to introduce the rescaled adjacency matrix

$$A := \sqrt{\frac{1}{p(1-p)N}}\, \mathcal{A} \tag{1.1}$$

so that the typical eigenvalue spacing of A is of order N^{-1}. A local law provides control of the matrix entries $G_{ij}(z)$ of the Green function

$$G(z) := (A - z)^{-1}, \tag{1.2}$$

---

[*]University of Geneva, Section of Mathematics, `yukun.he@unige.ch`.

[†]University of Geneva, Section of Mathematics, `antti.knowles@unige.ch`

[‡]University of Geneva, Section of Mathematics, `matteo.marcozzi@unige.ch`.

where $z = E + i\eta$ is a spectral parameter with positive imaginary part $\eta \gg N^{-1}$ defining the spectral scale. Our main result states that that the individual entries $G_{ij}(z)$ of the Green function concentrate all the way down to the critical scale $pN = C \log N$: the quantity

$$\max_{i,j}|G_{ij}(z) - \delta_{ij}m(z)|$$

is small with high probability for all $\eta \gg N^{-1}$; here $m(z)$ is the Stieltjes transform of the semicircle law.

Such local laws have become a cornerstone of random matrix theory, ever since the seminal work [12, 13] on Wigner matrices. They serve as fundamental tools in the study of the distribution of eigenvalues and eigenvectors, as well as in establishing universality in random matrix theory.

A local law has two well-known easy consequences, one for the eigenvectors and the other for the eigenvalues of A.

(i) The first consequence is *complete eigenvector delocalization*, which states that the normalized eigenvectors $\{\mathbf{u}_i\}$ of A satisfiy with high probability

$$\max\{\|\mathbf{u}_i\|_\infty : 1 \leqslant i \leqslant N\} \lesssim N^{-1/2} \tag{1.3}$$

where $\lesssim$ denotes a bound up to some factor $N^{o(1)}$.

(ii) The second consequence is a local law for the density of states, which states that the Stieltjes transform of the empirical eigenvalue distribution

$$s(z) := \frac{1}{N}\operatorname{Tr} G(z) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\lambda_i(A) - z}$$

satisfies $|s(z) - m(z)| = o(1)$ with high probability for all $\eta \gg N^{-1}$. Informally, this means that the semicircle law holds down to very small spectral scales.

It is a standard exercise to show that if $pN \to \infty$ then the (global) semicircle law for $s(z)$ holds, stating that $s(z) \to m(z)$ with high probability for any fixed $z \notin \mathbb{R}$. On the other hand, it is not hard to see that both consequences (i) and (ii) are wrong in the subcritical regime $pN \leqslant (1 - \varepsilon)\log N$. Indeed, in the subcritical regime there is with high probability an isolated vertex, with an associated eigenvector localized at that vertex. Hence, the left-hand side of (1.3) is equal to one and (i) fails. To see how (ii) fails in the subcritical regime, let us set $pN = \kappa \log N$ and denote by Y the number of isolated vertices. From [21, Proposition 5.9] we find that

$$\mathbb{E}Y = N^{1-\kappa}\left(1 + O\left(\frac{\kappa^2(\log N)^2}{N}\right)\right),$$

from which we deduce that $\mathbb{P}(Y \geqslant N^{1-\kappa}/4) \geqslant N^{-\kappa}/4$ (since $Y \leqslant N$). Thus we find that with probability at least $N^{-\kappa}/4$, the adjacency matrix A has at least $N^{1-\kappa}/4$ zero eigenvalues. We conclude that with probability at least $N^{-\kappa}/4$ we have

$$\operatorname{Im} s(z) \geqslant \frac{\eta}{4N^\kappa|z|^2}\,.$$

For $\kappa < 1$, setting $z = iN^{-\frac{1+\kappa}{2}}$ yields a contradiction to the estimate (ii), since $|m(z)| \leqslant 1$. (The precise meaning of "high probability" in (ii) is sufficiently strong to rule out events of probability $N^{-\kappa}/4$; see (2.4) below).

Thus, our assumption $pN \geqslant C \log N$ is optimal up to the value of the numerical constant C. A local law for the Erdős-Rényi graph was previously proved in [9, 11] under the assumption $pN \geqslant (\log N)^6$, and the contribution of this paper is therefore to cover the very sparse range $C \log N \leqslant pN \leqslant (\log N)^6$.

Next, we say a few words about the proof. Our strategy is based on the approach introduced in [12–14] and subsequently developed for sparse matrices in [9,11]. Thus, we derive a self-consistent equation for the Green function G using Schur's complement formula and large deviation estimates, which is then bootstrapped in the spectral scale η to reach the smallest scale N^{-1}. The key difficulty in proving local laws for sparse matrices is that the entries are sparse random variables, and hence fluctuate much more strongly than in the Wigner case $p \asymp 1$. To that end, new large deviation estimates for sparse random vectors were developed in [11], which were however ineffective below the scale $(\log N)^6$.

The key novelty of our approach is a new family of multilinear large deviation bounds for sparse random vectors. They are optimal for very sparse vectors, and in particular allow us to reach the critical scale $pN = C \log N$. They provide bounds on multilinear functions of sparse vectors in terms of mixed ℓ^2 and ℓ^∞ norms of their coefficients. We expect them to be more generally useful in a variety of problems on sparse random graphs. To illustrate them and how they are applied, consider a sparse random vector $X \in \mathbb{R}^N$ which is a single row of the matrix $A - \mathbb{E}A$. Let (a_{ij}) be a symmetric deterministic matrix. Then, for example, we have the L^r bound

$$\left\| \sum_{i \neq j} a_{ij} X_i X_j \right\|_r \leqslant \left(\frac{4r}{1 + (\log(\psi/\gamma))_+} \vee 4 \right)^2 (\gamma \vee \psi), \tag{1.4}$$

where

$$\gamma := \left(\max_i \frac{1}{N} \sum_j |a_{ij}|^2 \right)^{1/2}, \qquad \psi := \frac{\max_{i,j} |a_{ij}|}{pN}.$$

We first remark that we have to take r to be at least $\log N$. Indeed, our proof consists of an order $N^{O(1)}$ uses of such large deviation bounds. To compensate the factor $N^{O(1)}$ arising from the union bound, we therefore require bounds smaller than N^{-D} on the error probabilities for any fixed $D > 0$, which we obtain (by Chebyshev's inequality) from the large deviation bounds for $r = \log N$. The crucial feature of the bound (1.4) is the logarithmic factor in the denominator. Without it, there is nothing to compensate the factor $r^2 \geqslant (\log N)^2$ in the numerator, as $\psi \asymp (pN)^{-1} \asymp (\log N)^{-1}$ in the critical regime. Thus, the applicability of (1.4) hinges on the fact that ψ/γ is sufficiently large; this assumption can in fact be verified in all of our applications of (1.4).

Armed with these large deviation estimates, we can follow the basic approach of [11, 14] to conclude the local law by a bootstrapping; the main difference is that we have to be cautious to work as much as possible with L^r norms instead of the more commonly used high-probability bounds.

We remark that, in the recent years, a new approach to proving local laws has emerged [15,16,19], which replaces the row by row operations of the Schur complement formula with operations on individual entries performed by the cumulant expansion (or generalized Stein's formula). This new approach is considerably more versatile and general than the one based on Schur's complement formula. However, it encounters serious difficulties with very sparse matrices, owing to the fact that it requires the computation of high moments of the Green function entries. For the very sparse scales that are of interest to us, these moments may be large, although the entries themselves are small with high probability. The underlying phenomenon is that the exceptional events on which

the Green function is large do not have small enough probability to ensure the good behaviour of high moments.

Our proof is short and self-contained. For conciseness and clarity, we focus on the simple model of the Erdős-Rényi graph, and do not aim for optimal error bounds. However, a straightforward modification of our method, combined with the deterministic analysis of the quadratic vector equation developed in [1, 2], allows to extend our results to more general random graphs with independently chosen edges and general variance profiles, such as the stochastic block model. We do not pursue this direction further in this paper.

We conclude this section with a summary of some related results. The bulk universality for Erdős–Rényi graphs was first proved in [9] for $p \gg N^{-1/3}$, and it was later pushed to $p \gg N^{\varepsilon-1}$ for any fixed $\varepsilon > 0$ in [17]. The edge universality was first proved for $p \gg N^{-1/3}$ in [9], and later extended to $p \gg N^{-2/3}$ in [19]. More recently in [18], it was proved that the extreme eigenvalues has Gaussian fluctuations for $N^{-7/9} \ll p \ll N^{-2/3}$. We emphasize that the local law proved in this paper is, on its own, not precise enough to generalize the above universality results to the entire supercritical regime.

The extreme eigenvalues of A in the supercritical and subcritical regimes were recently investigated in [6, 7], where the authors proved that in the supercritical regime the extreme eigenvalues converge to the spectral edges ± 2, and in the subcritical regime a fraction of eigenvalues leave the bulk $[-2, 2]$ to become a cloud of outlier eigenvalues.

Local laws in the sparse regime have also been established for the random regular graph $G_{N,d}$, defined as the uniform probability distribution over all graphs on N vertices such that every vertex has degree d. Because the degree of any vertex is fixed to be d, the random regular graph $G_{N,d}$ is much more stable in the sparse regime than the Erdős-Rényi graph $G(N, d/N)$ with the same expected degree. In particular, $G_{N,d}$ does not exhibit a connectivity crossover and remains connected down to $d = 3$. A local law for the random regular graph $G_{N,d}$ for $d \geqslant (\log N)^4$ was proved in [4] and for fixed but large d in [3].

We organize the paper as follows. In Section 2 we state our main results. In Section 3 we prove a new family of multilinear large deviation estimates for sparse random vectors. In Section 4 we use the results in Section 3 to complete the proof.

2. Results

For convenience, in the remaining of this paper we introduce the new variable

$$q := \sqrt{pN} \in (0, N^{1/2}).$$

We consider random matrices of the following class; it is an easy exercise to check that A defined in (1.1) in terms of $G(N, p)$ satisfies the following conditions.

Definition 2.1 (Sparse matrix). Let $q \in (0, N^{1/2})$. A sparse matrix is a real symmetric $N \times N$ matrix $H = H^* \in \mathbb{R}^{N \times N}$ whose entries H_{ij} satisfy the following conditions.

(i) The upper-triangular entries $(H_{ij} : 1 \leqslant i \leqslant j \leqslant N)$ are independent.

(ii) We have $\mathbb{E}H_{ij} = 0$ and $\mathbb{E}H_{ij}^2 = (1 + O(\delta_{ij}))/N$ for all i, j.

(iii) For any $k \geqslant 3$, we have $\mathbb{E}|H_{ij}|^k \leqslant 1/(Nq^{k-2})$ for all i, j.

We define the adjacency matrix A by

$$A = H + f\mathbf{e}\mathbf{e}^*,$$

where $\mathbf{e} := N^{-1/2}(1, 1, \ldots, 1)^*$, and $0 \leqslant f \leqslant q$.

We define the spectral domain

$$\mathbf{S} \equiv \mathbf{S} := \{E + i\eta \in \mathbb{C} : N^{-1} < \eta \leqslant 1\}.$$

We always use the notation $z = E + i\eta$ for the real and imaginary parts of a spectral parameter z. For $\operatorname{Im} z \neq 0$, we define the Stieltjes transform m of the semicircle density ϱ by

$$m(z) := \int \frac{\varrho(\mathrm{d}x)}{x - z}, \qquad \varrho(\mathrm{d}x) := \frac{1}{2\pi}\sqrt{(4 - x^2)_+}\,\mathrm{d}x,$$

which is characterized as the unique solution of

$$m(z) + \frac{1}{m(z)} + z = 0$$

satisfying $\operatorname{Im} m(z)\operatorname{Im} z > 0$. We recall the Green function G from (1.2). Finally we define the fundamental error parameter

$$\zeta \equiv \zeta(N, r, q, \eta, f) := \left(\frac{r}{q^2}\right)^{1/4} + \frac{r}{(N\eta)^{1/6}} + \frac{r}{(\log\eta + \log N)q} + \frac{f}{(N\eta)^{1/4}}. \tag{2.1}$$

We now state our main result.

Theorem 2.2 (Local law). *There is a universal constant $C_* \geqslant 1$ such that the following holds. Let A be defined as in Definition 2.1 and G be its Green function (1.2). Let $z \in \mathbf{S}$. Let $r \geqslant 10$, $1 \leqslant q \leqslant N^{1/2}$, and $t > 0$. Suppose that*

$$t\zeta \leqslant 1, \tag{2.2}$$

then

$$\mathbb{P}\left(\max_{i,j}|G_{ij}(z) - m(z)\delta_{ij}| > t\zeta\right) \leqslant N^5\left(\frac{C_*}{t}\right)^r. \tag{2.3}$$

Remark 2.3. The constant C_* in (2.3) can be chosen to be 1000; see (4.49) below. This numerical value can be improved by more careful estimates; we shall not pursue this here.

We draw several consequences from Theorem 2.2. For all of them, we assume the upper bound $q \leqslant (\log N)^{10}$, which can however be relaxed to $q \leqslant N^{1/2}$ without much sweat; it originates from the last term of (2.1), which is a naive estimate of the contribution of the expectation $f\mathbf{e}\mathbf{e}^*$ of A. This upper bound can be removed by a more careful treatment presented in Section 7 of [11]. However, since in this paper we are interested in the regime $q \leqslant (\log N)^3$ not covered by the results of [11], we shall not do so.

We obtain a local law under the condition $q \geqslant C\sqrt{\log N}$, i.e. $pN \geqslant C\log N$. More precisely, by setting $r = \log N$ and $t = C_*\mathrm{e}^{5+D}$ for some $D > 0$ (see Appendix A for details), we deduce from (2.3) that for any $D > 0, \delta \in (0, 1)$ there exist $C \equiv C(\delta, D), N_0 \equiv N_0(\delta, D) > 0$ such that

$$\mathbb{P}\left[\max_{i,j}|G_{ij}(z) - m\delta_{ij}(z)| \leqslant \delta\right] \geqslant 1 - N^{-D} \tag{2.4}$$

5

whenever $\eta \geqslant N^{-1+(\log N)^{-1/2}}$, $C\sqrt{\log N} \leqslant q \leqslant (\log N)^{10}$ and $N \geqslant N_0$. Explicitly, we can choose

$$C(\delta, D) = \left(\frac{4C_* \mathrm{e}^{5+D}}{\delta}\right)^2, \qquad N_0(\delta, D) = \exp\left[\left(\mathrm{e}^{10} \log\left(\frac{4C_* \mathrm{e}^{5+D}}{\delta}\right)\right)^2\right]. \tag{2.5}$$

By (2.4) and a standard complex analysis argument (e.g. see Section 8 of [5]) we have the following result.

Theorem 2.4 (Local law for density of states). *Let* $\mu := \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i(A)}$ *be the empirical eigenvalue density of* A. *For* $D > 0, \delta \in (0,1)$ *there exist constants* $C \equiv C(\delta, D), N_0 \equiv N_0(\delta, D) > 0$, *given by* (2.5), *such that for any interval* $I \subset \mathbb{R}$ *satisfying* $|I| \geqslant N^{-1+(\log N)^{-1/2}}$, *we have*

$$\mathbb{P}\Big[\big|\mu(I) - \varrho(I)\big| \leqslant \delta|I|\Big] \geqslant 1 - N^{-D}$$

whenever $C\sqrt{\log N} \leqslant q \leqslant (\log N)^{10}$ *and* $N \geqslant N_0$.

Remark 2.5. Previously, in [20] a local law for the density of states under the assumption $q \to \infty$ was proved down to scales $|I| \geqslant \left(\frac{\log q}{q}\right)^{1/5}$. This scale is too large to distinguish the presence of isolated vertices, as explained in the introduction. Very recently, in [8] this result was improved to $|I| \geqslant \frac{\log N}{q^2}$ in the supercritical regime, under the assumption that I is separated from the spectral edges.

Remark 2.6. Theorem 2.4 says nothing about the locations of the extreme eigenvalues. This question was addressed in [6, 7], where it was proved that the extreme eigenvalues converge to the spectral edges in the supercritical regime, and they become outliers in the subcritical regime.

A standard consequence of Theorem 2.2 (e.g. in [5, Theorem 2.10]) is the following.

Theorem 2.7 (Complete eigenvector delocalization). *For any* $D > 0$ *there exists* $C \equiv C(D) > 0, N_0 \equiv N_0(D)$ *such that for* $q \geqslant C\sqrt{\log N}$ *and* $N \geqslant N_0(D)$,

$$\mathbb{P}\left(\exists i, \|\mathbf{u}_i\|_\infty \geqslant N^{(\log N)^{-1/2}-1/2}\right) \leqslant N^{-D},$$

where $\mathbf{u}_i \in \mathbb{S}^{N-1}$ *is the* i-th *normalized eigenvector of* A. *(We can take* $C(D) = C(1, D)$ *and* $N_0(D) = N_0(1, D)$ *in* (2.5).*)*

Remark 2.8. Previously, in [20] it was proved that with high probability $\|\mathbf{u}_i\|_\infty = o(1)$ for $q \geqslant C\sqrt{\log N}$. For eigenvectors associated with eigenvalues separated from the spectral edges, the authors obtained the stronger estimate $\|\mathbf{u}_i\|_\infty = O((\log(q/\sqrt{\log N}))^{1.1}\sqrt{\log N}/q)$. Very recently, in [8] this was improved to $\|\mathbf{u}_i\|_\infty = O(\sqrt{\log N}/q)$ for eigenvalues separated from the spectral edges.

Moreover, we obtain the following probabilistic version of local quantum unique ergodicity for the Erdős-Rényi random graph, by combining Theorem 2.7 with [4, Proposition 8.3].

Theorem 2.9 (Probabilistic local quantum unique ergodicity). *Let* A *be the rescaled adjacency matrix of the Erdős-Rényi graph* $G(N,p)$. *Let* $a_1, \dots, a_N \in \mathbb{R}$ *be deterministic numbers satisfying* $\sum_{i=1}^N a_i = 0$. *For any* $D > 0$ *there exists* $C \equiv C(D) > 0, N_0 \equiv N_0(D)$ *such that for* $pN \geqslant C^2 \log N$, $N \geqslant N_0(D)$, $k = 1, \dots, N$, *and* $\theta \geqslant 1$

$$\sum_{i=1}^N a_i u_k(i)^2 = O\left(\frac{\theta^2 N^{(\log N)^{-1/2}}}{N}\left(\sum_{i=1}^N a_i^2\right)^{1/2}\right)$$

6

with probability at least $1 - N^{-D} - \mathrm{e}^{-\theta\sqrt{\log\theta}}$. Here $u_k(i)$ is the i-th component of the k-th eigenvector of A.

Theorem 2.9 states that, on deterministic sets of at least $(\theta^2 N^{(\log N)^{-1/2}})^2$ vertices, all eigenvectors of the random graph A are completely flat with high probability. In other words, with high probability, the random probability measure $i \mapsto v_k(i)^2$ is close (when tested against deterministic test functions) to the uniform probability measure $i \mapsto 1/N$ on $\{1, \ldots, N\}$. For instance, let $I \subset \{1, \ldots, N\}$ be a deterministic subset of vertices. Setting $a_i := \mathbf{1}(i \in I) - |I|/N$ in Theorem 2.9, we obtain

$$\sum_{i \in I} v_k(i)^2 = \sum_{i \in I} \frac{1}{N} + O\left(\frac{\theta^2 N^{(\log N)^{-1/2}} \sqrt{|I|}}{N}\right) \tag{2.6}$$

with probability at least $1 - N^{-D} - \mathrm{e}^{-\theta\sqrt{\log\theta}}$. The main term on the right-hand side of (2.6) is much larger than the error term provided that $|I| \gg (\theta^2 N^{(\log N)^{-1/2}})^2$.

3. Multilinear large deviation bounds for sparse random vectors

The rest of this paper is devoted to the proof of Theorem 2.2. In this section we derive large deviation estimates for multilinear forms of sparse random vectors with independent components. We focus on the linear and bilinear estimates, which are sufficient for our applications. These estimates are designed to be optimal in the regime of very sparse random vectors.

For $n \in \mathbb{N}^*$ denote by $[n] := \{1, 2, \ldots, n\}$, and for a finite set A we denote by $|A|$ the cardinality of A. For two sets U, V, we denote by V^U the set of functions from U to V. For $r \geqslant 1$ we denote by $\|X\|_r := (\mathbb{E}|X|^r)^{1/r}$ the L^r-norm of the random variable X.

Proposition 3.1. *Let r be even and $1 \leqslant q \leqslant N^{1/2}$. Let $X_1, \ldots, X_N \in \mathbb{C}$ be independent random variables satisfying*

$$\mathbb{E}X_i = 0, \qquad \mathbb{E}|X_i|^k \leqslant \frac{1}{Nq^{k-2}} \quad (2 \leqslant k \leqslant r). \tag{3.1}$$

Let $a_1, \ldots, a_N \in \mathbb{C}$ be deterministic, and suppose that

$$\left(\frac{1}{N} \sum_i |a_i|^2\right)^{1/2} \leqslant \gamma, \qquad \frac{\max_i |a_i|}{q} \leqslant \psi$$

for some $\gamma, \psi \geqslant 0$. Then

$$\left\|\sum_i a_i X_i\right\|_r \leqslant \left(\frac{2r}{1 + 2(\log(\psi/\gamma))_+} \vee 2\right)(\gamma \vee \psi).$$

Proof. To avoid cumbersome complex conjugates in our notation, we assume for simplicity that all quantities are real-valued. Denote by $\mathfrak{P}(r)$ the set of partitions of $[r]$, and by $\mathfrak{P}_{\geqslant 2}(r)$ the subset of $\mathfrak{P}(r)$ whose blocks all have size at least two. For $(i_1, \ldots, i_r) = [N]^r$ denote by $P(i_1, \ldots, i_r) \in \mathfrak{P}(r)$ the partition of $[r]$ associated with the equivalence relation $k \sim l$ if and only if $i_k = i_l$. Using the identity

$$1 = \sum_{\Pi \in \mathfrak{P}(r)} \mathbf{1}(P(i_1, \ldots, i_r) = \Pi) = \sum_{\Pi \in \mathfrak{P}(r)} \sum_{s \in [N]^\Pi} \prod_{\pi \neq \tilde{\pi} \in \Pi} \mathbf{1}(s_\pi \neq s_{\tilde{\pi}}) \prod_{\pi \in \Pi} \prod_{k \in \pi} \mathbf{1}(i_k = s_\pi),$$

the independence of the variables X_i, and the fact that $\mathbb{E}X_i = 0$ for all i, we find

$$\left\|\sum_i a_i X_i\right\|_r^r = \sum_{i_1,\dots,i_r=1}^N a_{i_1}\cdots a_{i_r}\,\mathbb{E}X_{i_1}\cdots X_{i_r} \leqslant \sum_{\Pi\in\mathfrak{P}_{\geqslant 2}(r)}\sum_{s\in[N]^\Pi}\prod_{\pi\in\Pi}|a_{s_\pi}|^{|\pi|}\mathbb{E}|X_{s_\pi}|^{|\pi|}\,.$$

Here the restriction $\Pi \in \mathfrak{P}_{\geqslant 2}(r)$ follows from the fact that $\mathbb{E}X_i = 0$, and we dropped the indicator function $\prod_{\pi\neq\tilde\pi\in\Pi}\mathbf{1}(s_\pi \neq s_{\tilde\pi})$ to obtain an upper bound. We deduce from

$$\sum_{s\in[N]^\Pi}\prod_{\pi\in\Pi}|a_{s_\pi}|^{|\pi|}\mathbb{E}|X_{s_\pi}|^{|\pi|} = \prod_{\pi\in\Pi}\sum_{s=1}^N|a_s|^{|\pi|}\mathbb{E}|X_s|^{|\pi|}$$

$$\leqslant \prod_{\pi\in\Pi}\left(\sum_i|a_i|^2\right)(\max_i|a_i|)^{|\pi|-2}\frac{1}{Nq^{|\pi|-2}} \leqslant \gamma^{2|\Pi|}\psi^{r-2|\Pi|}$$

that

$$\left\|\sum_i a_i X_i\right\|_r^r \leqslant R_r(\gamma,\psi)\,, \qquad R_r(\gamma,\psi) := \sum_{k=1}^{r/2}S(r,k)\gamma^{2k}\psi^{r-2k}\,,$$

where $S(r,k) := |\{\Pi \in \mathfrak{P}(r) : |\Pi| = k\}|$ is the Stirling number of the second kind. To conclude the proof, it therefore suffices to prove that

$$R_r(\gamma,\psi)^{1/r} \leqslant \left(\frac{2r}{1+2(\log(\psi/\gamma))_+}\vee 2\right)(\gamma\vee\psi)\,. \tag{3.2}$$

Using the bound $S(r,k) \leqslant \frac{1}{2}\binom{r}{k}k^{r-k}$, we conclude

$$R_r(\gamma,\psi) \leqslant 2^r \max_{1\leqslant k\leqslant r/2}f(k)\,, \qquad f(k) := k^{r-k}\gamma^{2k}\psi^{r-2k}\,.$$

If $\psi \leqslant \gamma$ then we estimate $f(k)^{1/r} \leqslant r\gamma$ and (3.2) follows.

It suffices therefore to assume $\gamma \leqslant \psi$. We consider the function f on the interval $(0,\infty)$. By differentiation, we find that $\log f$ is concave on $(0,\infty)$ and maximized for

$$\frac{r}{k} = 1 + \log k + 2\log(\psi/\gamma)\,. \tag{3.3}$$

Since the left-hand side of (3.3) is decreasing and its right-hand side increasing, it is easy to see that (3.3) has a unique solution k_* in $(0,\infty)$. If $k_* < 1$ then, by concavity of $\log f$, we find that f is decreasing on $[1,\infty)$ and

$$R_r(\gamma,\psi)^{1/r} \leqslant 2f(1)^{1/r} \leqslant 2\psi,$$

which is (3.2). Let us therefore assume that $k_* \geqslant 1$. Then we find

$$R_r(\gamma,\psi)^{1/r} \leqslant 2f(k_*)^{1/r} \leqslant 2k_*\psi\,.$$

Moreover, for $k_* \geqslant 1$ from (3.3) we find that

$$k_* \leqslant \frac{r}{1+2\log(\psi/\gamma)}\,,$$

and (3.2) follows. $\qquad\square$

Proposition 3.2. *Let r be even and $1 \leqslant q \leqslant N^{1/2}$. Let $X_1, \ldots, X_N \in \mathbb{C}$ be independent random variables satisfying* (3.1). *For any deterministic $a_1, \ldots, a_N \in \mathbb{C}$ we have*

$$\left\| \sum_i a_i(|X_i|^2 - \mathbb{E}|X_i|^2) \right\|_r \leqslant 2\left(1 + \frac{2q^2}{N}\right) \max_i |a_i| \left(\frac{r}{q^2} \vee \sqrt{\frac{r}{q^2}}\right).$$

Proof. Since $\mathbb{E}|X_i|^2 \leqslant \frac{1}{N}$, a simple estimate using the binomial theorem yields

$$\mathbb{E}\left(|X_i|^2 - \mathbb{E}|X_i|^2\right)^r \leqslant \frac{1}{Nq^{2r-2}}\left(1 + \frac{2q^2}{N}\right)^r.$$

As in the proof of Proposition 3.1, we conclude that

$$\left\| \sum_i a_i(|X_i|^2 - \mathbb{E}|X_i|^2) \right\|_r^r \leqslant \sum_{k=1}^{r/2} S(r,k)(\max_i |a_i|)^r \frac{1}{q^{2r-2k}}\left(1 + \frac{2q^2}{N}\right)^r$$

$$\leqslant \left[2\max_i |a_i|\left(1 + \frac{2q^2}{N}\right)\right]^r \max_{1 \leqslant k \leqslant r/2}\left(\frac{k}{q^2}\right)^{r-k}$$

and the claim follows. $\qquad\square$

Proposition 3.3. *Let r be even and $1 \leqslant q \leqslant N^{1/2}$. Let $X_1, \ldots, X_N, Y_1, \ldots, Y_N$ be independent random variables, all satisfying* (3.1). *Let $a_{ij} \in \mathbb{C}$, $i, j = 1, \ldots, N$ be deterministic, and suppose that*

$$\left(\max_i \frac{1}{N}\sum_j |a_{ij}|^2\right)^{1/2} \vee \left(\max_j \frac{1}{N}\sum_i |a_{ij}|^2\right)^{1/2} \leqslant \gamma, \qquad \frac{\max_{i,j}|a_{ij}|}{q^2} \leqslant \psi$$

for some $\gamma, \psi \geqslant 0$. Then

$$\left\| \sum_{i,j} a_{ij} X_i Y_j \right\|_r \leqslant \left(\frac{2r}{1 + (\log(\psi/\gamma))_+} \vee 2\right)^2 (\gamma \vee \psi). \tag{3.4}$$

Proof. As in the proof of Proposition 3.1, to simplify notation we assume that all quantities are real-valued. We write

$$\left\| \sum_{i,j} a_{ij} X_i Y_j \right\|_r^r = \sum_{i_1, \ldots, i_r} \sum_{j_1, \ldots, j_r} a_{i_1 j_1} \cdots a_{i_r j_r} \mathbb{E} X_{i_1} Y_{j_1} \cdots X_{i_r} Y_{j_r}.$$

Analogously to the proof of Proposition 3.1, we encode the terms on the right-hand side using two partitions $\Pi_1, \Pi_2 \in \mathfrak{P}_{\geqslant 2}(r)$, by requiring that $k, l \in [r]$ are in the same block of Π_1 if and only if $i_k = i_l$, and $k, l \in [r]$ are in the same block of Π_2 if and only if $j_k = j_l$. Each block of Π_1 and Π_2 has size at least two. We encode each pair $(\Pi_1, \Pi_2) \in \mathfrak{P}_{\geqslant 2}(r)^2$ using a multigraph $G(\Pi_1, \Pi_2)$, whose vertex set is $\Pi_1 \sqcup \Pi_2$ and whose set of edges is obtained by adding, for each $k = 1, \ldots, r$, an edge $\{\pi_1, \pi_2\}$ where $k \in \pi_1 \in \Pi_1$ and $k \in \pi_2 \in \Pi_2$.

It is immediate that $G(\Pi_1, \Pi_2)$ has r edges, that each vertex has degree at least two, and that $G(\Pi_1, \Pi_2)$ is bipartite. Thus we find

$$\left\| \sum_{i,j} a_{ij} X_i Y_j \right\|_r^r \leqslant \sum_{\Pi_1, \Pi_2 \in \mathfrak{P}_{\geqslant 2}(r)} \sum_{s \in [N]^{\Pi_1 \sqcup \Pi_2}} \left(\prod_{\{\pi_1, \pi_2\} \in E(G(\Pi_1, \Pi_2))} |a_{s_{\pi_1} s_{\pi_2}}|\right) \left(\prod_{\pi \in \Pi_1 \sqcup \Pi_2} \frac{1}{Nq^{|\pi|-2}}\right). \tag{3.5}$$

9

To estimate the right-hand side of (3.5), we construct an algorithm that successively sums out the variables s_π. For a multigraph $G = (V(G), E(G))$ define

$$\text{Val}(G) := \frac{1}{N^{|V(G)|}} \sum_{s \in [N]^{V(G)}} \left(\prod_{\{\pi_1, \pi_2\} \in E(G)} |a_{s_{\pi_1} s_{\pi_2}}| \right)$$

so that (3.5) becomes

$$\left\| \sum_{i,j} a_{ij} X_i Y_j \right\|_r^r \leqslant \sum_{\Pi_1, \Pi_2 \in \mathfrak{P}_{\geqslant 2}(r)} \frac{1}{q^{2r - 2|\Pi_1| - 2|\Pi_2|}} \text{Val}(G(\Pi_1, \Pi_2)) \tag{3.6}$$

We shall need the following notions. We denote by $x(G) := \sum_{v \in V(G)} (\deg_G(v) - 2)_+$ the total excess degree of a multigraph G. We fix $\Pi_1, \Pi_2 \in \mathfrak{P}_{\geqslant 2}(r)$ and abbreviate $k := |\Pi_1| + |\Pi_2|$. Abbreviate

$$a := \left(\max_i \frac{1}{N} \sum_j |a_{ij}|^2 \right)^{1/2} \vee \left(\max_j \frac{1}{N} \sum_i |a_{ij}|^2 \right)^{1/2}, \qquad A := \max_{i,j} |a_{ij}|.$$

We construct a sequence $G_0, G_1, \ldots, G_L$ of multigraphs recursively as follows. First, $G_0 := G(\Pi_1, \Pi_2)$. Since $\deg_{G_0}(v) \geqslant 2$ for all $v \in V(G_0)$, we immediately find $x(G_0) = 2r - 2k$. For $\ell \geqslant 0$, the multigraph $G_{\ell+1}$ is constructed from G_ℓ according to the two following cases.

1. G_ℓ has no vertex of degree $\leqslant 2$. Choose an arbitrary $e \in E(G_\ell)$ and let $G_{\ell+1}$ be the multigraph obtained from G_ℓ by removing the edge e. Clearly,

$$x(G_{\ell+1}) = x(G_\ell) - 2, \qquad \text{Val}(G_\ell) \leqslant A \, \text{Val}(G_{\ell+1}).$$

2. G_ℓ has a vertex v of degree $\leqslant 2$. Let $G_{\ell+1}$ be the multigraph obtained from G_ℓ by removing the vertex v and all $\deg_{G_\ell}(v)$ edges incident to it. Clearly,

$$x(G_{\ell+1}) \leqslant x(G_\ell), \qquad \text{Val}(G_\ell) \leqslant a^{\deg_{G_\ell}(v)} \, \text{Val}(G_{\ell+1}),$$

where the second estimate follows by Cauchy-Schwarz and the fact that $\deg_{G_\ell}(v) = 0, 1, 2$.

It is easy to see that the algorithm terminates at $G_L = (\emptyset, \emptyset)$, the empty graph with $\text{Val}(G_L) = 1$.

Next, for $i = 1, 2$, let r_i be the total number of edges removed by the algorithm in all steps of case i. Clearly, $r_1 + r_2 = r$. Moreover, since $x(G_L) \geqslant 0$, we find that $x(G_0) - 2r_1 \geqslant 0$, i.e. $r_1 \leqslant r - k$, which implies $r_2 \geqslant k$. We conclude that

$$\text{Val}(G_0) \leqslant A^{r_1} a^{r_2} \leqslant A^{r-k} a^k,$$

where we used that $a \leqslant A$. Hence, with $k_i := |\Pi_i|$ (so that $k = k_1 + k_2$), we have

$$\frac{1}{q^{2r - 2k_1 - 2k_2}} \text{Val}(G(\Pi_1, \Pi_2)) \leqslant \left(\frac{A}{q^2} \right)^{r - k_1 - k_2} a^{k_1 + k_2}.$$

Summarizing, we have

$$\left\| \sum_{i,j} a_{ij} X_i Y_j \right\|_r^r \leqslant \sum_{k_1, k_2 = 1}^{r/2} S(r, k_1) S(r, k_2) \left(\frac{A}{q^2} \right)^{r - k_1 - k_2} a^{k_1 + k_2} \leqslant \left(\sum_{k=1}^{r/2} S(r, k) \sqrt{\gamma}^{2k} \sqrt{\psi}^{r - 2k} \right)^2.$$

The claim now follows from applying the estimate (3.2) to $R_r(\sqrt{\gamma}, \sqrt{\psi})$. $\qquad\square$

Remark 3.4. It might be tempting to try to prove an analogous result with the parameter γ in (3.4) replaced with the smaller $\frac{1}{N^2}\sum_{i,j}|a_{ij}|^2$. However, in this case the moment-based argument used in the proof of Proposition 3.3 does not lead to a prefactor that is small in the targeted regime $q^2 \gtrsim r$, $\max_{i,j}|a_{ij}| \asymp 1$, $\max_i \frac{1}{N}\sum_j |a_{ij}|^2 \asymp N^{-c}$, which corresponds to the typical behaviour of the parameters q, r, a_{ij} in the applications to proving a local law for supercritical graphs (see Section 4 below). Indeed, consider the case where Π_1 is arbitrary and Π_2 consists of a single block. There are an order $(Cr)^r$ partitions Π_1. If we only allow estimates in terms of $\frac{1}{N^2}\sum_{i,j}|a_{ij}|^2$ and $\max_{i,j}|a_{ij}|$, then each such partition yields a contribution of order

$$\frac{1}{N^2}\sum_{i,j}|a_{ij}|^2\left(\frac{\max_{i,j}|a_{ij}|}{q}\right)^{r-2}$$

Thus, the total contribution of such pairings is of the order $N^{-c}q^2\left(\frac{Cr}{q}\right)^r$, which is much too large.

This observation leads us believe that $\max_i \frac{1}{N}\sum_j |a_{ij}|^2$ is the correct quantity, instead of the smaller $\frac{1}{N^2}\sum_{i,j}|a_{ij}|^2$, when deriving large deviation estimates in the very sparse regime.

Proposition 3.5. *Let r be even and $1 \leqslant q \leqslant N^{1/2}$. Let $X_1, \ldots, X_N$ be independent random variables satisfying (3.1). Let $a_{ij} \in \mathbb{C}$, $i, j = 1, \ldots, N$ be deterministic, and suppose that*

$$\left(\max_i \frac{1}{N}\sum_j |a_{ij}|^2\right)^{1/2} \vee \left(\max_j \frac{1}{N}\sum_i |a_{ij}|^2\right)^{1/2} \leqslant \gamma, \qquad \frac{\max_{i,j}|a_{ij}|}{q^2} \leqslant \psi$$

for some $\gamma, \psi \geqslant 0$. Then

$$\left\|\sum_{i \neq j} a_{ij} X_i X_j\right\|_r \leqslant \left(\frac{4r}{1 + (\log(\psi/\gamma))_+} \vee 4\right)^2 (\gamma \vee \psi). \tag{3.7}$$

Proof. Using the simple decoupling argument from the proof of [10, Lemma B.4], we find

$$\left\|\sum_{i \neq j} a_{ij} X_i X_j\right\|_r \leqslant \frac{1}{2^{N-2}} \sum_{I \sqcup J = [N]} \left\|\sum_{i \in I}\sum_{j \in J} a_{ij} X_i X_j\right\|_r,$$

where on the right-hand side the sum ranges over disjoint nonempty sets I, J whose union is $[N]$. The L^r-norm on the right-hand side can be estimated using Proposition 3.3, and the sum $\sum_{I \sqcup J = [N]}$ yields a factor $2^N - 2$. Hence the claim follows.

Alternatively, we can repeat the proof of Proposition 3.3 almost verbatim, except that instead of having a bipartite multigraph encoded by two partitions, $\Pi_1, \Pi_2 \in \mathfrak{P}_{\geqslant 2}(r)$, we have a multigraph without loops encoded by the single partition $\Pi \in \mathfrak{P}_{\geqslant 2}(2r)$. $\square$

4. Proof of Theorem 2.2

In this section we give the proof of Theorem 2.2. Throughout this section we frequently omit the spectral parameter z from our notation, unless this leads to confusion. We start with a few standard tools.

Definition 4.1. For $k \in [N]$ we define $A^{(k)}$ as the $(N-1) \times (N-1)$ matrix

$$A^{(k)} = (A_{ij})_{i,j \in [N] \setminus \{k\}} \,. \tag{4.1}$$

Moreover, we define the Green function of $A^{(k)}$ through

$$G^{(k)}(z) = (A^{(k)} - z)^{-1} \,. \tag{4.2}$$

We also abbreviate

$$\sum_i^{(k)} := \sum_{i:i \neq k} \,. \tag{4.3}$$

By applying the resolvent identity to $G - G^*$, we get the Ward identity

$$\sum_j |G_{ij}|^2 = \frac{\operatorname{Im} G_{ii}}{\eta} \,, \tag{4.4}$$

which of course also holds for $G^{(k)}$ for any $k \in [N]$.

The following variant of Schur's complement formula is standard and its proof is given e.g. in [5].

Lemma 4.2. *For $i \neq j$ we have*

$$G_{ij} = -G_{jj} \sum_k^{(j)} G_{ik}^{(j)} A_{kj} = -G_{ii} \sum_k^{(i)} A_{ik} G_{kj}^{(i)} \,. \tag{4.5}$$

For $i, j \neq k$ we have

$$G_{ij}^{(k)} = G_{ij} - \frac{G_{ik} G_{kj}}{G_{kk}} \,. \tag{4.6}$$

Finally,

$$\frac{1}{G_{ii}} = A_{ii} - z - \sum_{k,l}^{(i)} A_{ik} G_{kl}^{(i)} A_{li} \tag{4.7}$$

for all i.

Define the Stieltjes transform of the empirical eigenvalue measure as

$$s := \frac{1}{N} \sum_{i=1}^N G_{ii} \,.$$

From (4.7) we easily deduce the following result, which serves as the starting point of our analysis of the diagonal entries of G.

Lemma 4.1. We have

$$\frac{1}{G_{ii}} = -z - s + Y_i \,, \tag{4.8}$$

where

$$Y_i := H_{ii} + \frac{1}{N} \sum_k \frac{G_{ki} G_{ik}}{G_{ii}} - \sum_{k \neq l}^{(i)} H_{ik} G_{kl}^{(i)} H_{li} - \sum_k^{(i)} \left(H_{ik}^2 - \frac{1}{N} \right) G_{kk}^{(i)}$$

$$+ \frac{f}{N} - \left(\frac{f}{N} \right)^2 \sum_{k,l}^{(i)} G_{kl}^{(i)} - \frac{f}{N} \sum_{k,l}^{(i)} G_{kl}^{(i)} (H_{ik} + H_{li}) \,. \tag{4.9}$$

12

Proof. The proof consists of comparing the right-hand side of (4.7) to its conditional expectation $\mathbb{E}(\cdot | H^{(i)})$, and we omit the details. □

Next, we define the random z-dependent error parameter

$$\Gamma := \max_{i,j} |G_{ij}| \vee \max_{i,j \neq k} |G_{ij}^{(k)}|\,,$$

as well as the z-dependent indicator function

$$\phi := \mathbf{1}(\Gamma \leqslant 2)\,. \tag{4.10}$$

The proof consists of a stochastic continuity argument, which establishes control on the entries of G under the bootstrapping assumption $\phi = 1$. The following lemma provides the key probabilistic estimates that allow us to estimate the various fluctuating error terms that appear in the proof.

Lemma 4.3 (Main estimates). *Let* $2 \leqslant r \leqslant q^2 \leqslant N$. *Then*

$$\max_i \|\phi Y_i G_{ii}\|_r \leqslant 48^2 \left[\left(\frac{r}{q^2}\right)^{1/2} + \frac{r^2}{(N\eta)^{1/3}} + \left(\frac{r}{(\log N + \log \eta)q}\right)^2 + \frac{f^2}{\sqrt{N\eta}} \right]\,, \tag{4.11}$$

$$\max_{i \neq j} \|\phi\, G_{ij}\|_r \leqslant 12 \left[\frac{1}{q} + \frac{r}{(N\eta)^{1/6}} + \frac{r}{(\log N + \log \eta)q} + \frac{f}{(N\eta)^{1/4}} \right]\,, \tag{4.12}$$

and

$$\max_{i,j \neq k} \|\phi(G_{ij} - G_{ij}^{(k)})\|_r \leqslant 12 \left[\frac{1}{q} + \frac{r}{(N\eta)^{1/6}} + \frac{r}{(\log N + \log \eta)q} + \frac{f}{(N\eta)^{1/4}} \right] \tag{4.13}$$

for all $z \in \mathbf{S}$.

Proof. We begin with (4.11). Pick $i \in [N]$. We shall first bound $\|\phi Y_i\|_r$ by estimating the L^r-norm of the terms on the right-hand side of (4.9). Definition 2.1 (iii) ensures

$$\|H_{ii}\|_r \leqslant 1/q\,, \tag{4.14}$$

and by the Ward identity (4.4) we get

$$\left| \frac{1}{N} \sum_k \frac{G_{ki} G_{ik}}{G_{ii}} \right| \leqslant \frac{1}{N\eta}\,. \tag{4.15}$$

We shall estimate the L^r-norms of the remaining terms on the right-hand side of (4.9) using the multilinear large deviation estimates from Section 3. We begin with the term $\sum_{k \neq l}^{(i)} H_{ik} G_{kl}^{(i)} H_{li}$. Define

$$\phi^{(i)} := \mathbf{1}\left(\max_{k,l \neq i} |G_{kl}^{(i)}| \leqslant 2\right)\,,$$

and note that $\phi \leqslant \phi^{(i)}$. Denote by $\|\cdot\|_{r|H^{(i)}}$ the conditional L^r norm with respect to the conditional expectation $\mathbb{E}[\cdot | H^{(i)}]$. Then by the nesting property of conditional L^r-norms, we have

$$\left\| \phi \sum_{k \neq l}^{(i)} H_{ik} G_{kl}^{(i)} H_{li} \right\|_r \leqslant \left\| \phi^{(i)} \sum_{k \neq l}^{(i)} H_{ik} G_{kl}^{(i)} H_{li} \right\|_r = \left\| \phi^{(i)} \left\| \sum_{k \neq l}^{(i)} H_{ik} G_{kl}^{(i)} H_{li} \right\|_{r|H^{(i)}} \right\|_r\,.$$

13

The conditional L^r norm is amenable to the large deviation estimate from Proposition 3.5. To that end, we use the Ward identity to estimate

$$\phi^{(i)}\Big(\max_k \frac{1}{N}\sum_l |G_{kl}^{(i)}|^2\Big)^{1/2} \vee \phi\Big(\max_l \frac{1}{N}\sum_k |G_{kl}^{(i)}|^2\Big)^{1/2} \leqslant \sqrt{\frac{2}{N\eta}}, \tag{4.16}$$

so that Proposition 3.5 with the choices $\gamma = \sqrt{\frac{2}{N\eta}}$ and $\psi = \frac{2}{q^2}$ gives

$$\begin{aligned}
\Big\|\phi\sum_{k\neq l}^{(i)} H_{ik}G_{kl}^{(i)}H_{li}\Big\|_r &\leqslant \Big[\Big(\frac{4r}{1+(\log(\sqrt{N\eta}/q^2))_+}\Big)^2 + 16\Big]\Big(\sqrt{\frac{2}{N\eta}} \vee \frac{2}{q^2}\Big) \\
&\leqslant 16r^2 \cdot \frac{2}{(N\eta)^{1/3}} + \Big(\frac{24r}{\log N + \log\eta}\Big)^2 \cdot \frac{2}{q^2} + \frac{32}{\sqrt{N\eta}} + \frac{32}{q^2},
\end{aligned} \tag{4.17}$$

where the second step is obtained by considering the two cases $q^2 \geqslant (N\eta)^{1/3}$ and $q^2 \leqslant (N\eta)^{1/3}$ separately.

Similarly, using Proposition 3.2 on the fourth term on the right-hand side of (4.9) gives

$$\Big\|\phi\sum_k^{(i)}\Big(H_{ik}^2 - \frac{1}{N}\Big)G_{kk}^{(i)}\Big\|_r \leqslant 4\Big(1+\frac{2q^2}{N}\Big)\sqrt{\frac{r}{q^2}} \leqslant 12\sqrt{\frac{r}{q^2}}. \tag{4.18}$$

Next, by the Ward identity we have

$$\Big|\phi\Big(\frac{f}{N}\Big)^2 \sum_{k,l}^{(i)} G_{kl}^{(i)}\Big| \leqslant f^2\sqrt{\frac{2}{N\eta}}. \tag{4.19}$$

Next, we estimate the last term of (4.9). With the abbreviations

$$a_k := \phi^{(i)}\frac{f}{N}\sum_l^{(i)} G_{kl}^{(i)}, \qquad X_k := H_{ik}$$

we have $\max_k |a_k| \leqslant \sqrt{2}f/\sqrt{N\eta}$ by the Ward identity. From Proposition 3.1 with $\gamma = \psi = \sqrt{2}f/\sqrt{N\eta}$ we therefore get

$$\Big\|\phi\frac{f}{N}\sum_{k,l}^{(i)} G_{kl}^{(i)}H_{ik}\Big\|_r = \Big\|\Big\|\sum_k^{(i)} a_k X_k\Big\|_{r|H^{(i)}}\Big\|_r \leqslant 2r \cdot \frac{\sqrt{2}f}{\sqrt{N\eta}} \leqslant (r^2 + f^2)\sqrt{\frac{2}{N\eta}}, \tag{4.20}$$

and the same bound can also be derived for the L^r-norm of $\phi\frac{f}{N}\sum_{k,l}^{(i)} G_{kl}^{(i)}H_{li}$.

Summarizing, by Minkowski's inequality and (4.14)–(4.20), we get

$$\begin{aligned}
\|\phi Y_i G_{ii}\|_r \leqslant 2\|\phi Y_i\|_r \leqslant{} & \frac{2}{q} + \frac{2}{N\eta} + \frac{64r^2}{(N\eta)^{1/3}} + \frac{(48r)^2}{(\log N + \log\eta)^2 q^2} + \frac{64}{\sqrt{N\eta}} + \frac{32}{q^2} \\
& + 24\sqrt{\frac{r}{q^2}} + \frac{2f}{N} + 6f^2\sqrt{\frac{2}{N\eta}} + 4r^2\sqrt{\frac{2}{N\eta}}.
\end{aligned}$$

14

and by $N\eta \geqslant 1$, $r \geqslant 2$ we have

$$\|\phi Y_i G_{ii}\|_r \leqslant \frac{100r^2}{(N\eta)^{1/3}} + \frac{(48r)^2}{(\log N + \log \eta)^2 q^2} + 100\sqrt{\frac{r}{q^2}} + \frac{100f^2}{\sqrt{N\eta}}, \tag{4.21}$$

and (4.11) follows.

Now we turn to (4.12). For $i \neq j$, by (4.5) we have

$$G_{ij} = -G_{ii} \sum_k^{(i)} \left(H_{ik} + \frac{f}{N}\right) G_{kj}^{(i)}. \tag{4.22}$$

Thus, for $i \neq j$ we have

$$\|\phi\, G_{ij}\|_r \leqslant \left\|\phi\, G_{ii} \sum_k^{(i)} \left(H_{ik} + \frac{f}{N}\right) G_{kj}^{(i)}\right\|_r \leqslant 2\left\|\phi^{(i)} \sum_k^{(i)} H_{ik} G_{kj}^{(i)}\right\|_r + \frac{4f}{\sqrt{N\eta}},$$

where in the last step we estimated the term $\phi G_{ii} \sum_k^{(i)} \frac{f}{N} G_{kj}^{(i)}$ by $\frac{4f}{\sqrt{N\eta}}$, using the Ward identity. Invoking Proposition 3.1 with the choices $\gamma = \sqrt{\frac{2}{N\eta}}$ and $\psi = \frac{2}{q}$ yields, by the Ward identity,

$$\|\phi\, G_{ij}\|_r \leqslant \left(\frac{4r}{1 + 2(\log(\sqrt{N\eta}/q))_+} + 4\right)\left(\sqrt{\frac{2}{N\eta}} \vee \frac{2}{q}\right) + \frac{4f}{\sqrt{N\eta}}$$

$$\leqslant 4r \cdot \frac{2}{(N\eta)^{1/6}} + \frac{6r}{\log N + \log \eta} \cdot \frac{2}{q} + \frac{8}{\sqrt{N\eta}} + \frac{8}{q} + \frac{4f}{\sqrt{N\eta}},$$

where the second step is obtained by considering the two cases $q \geqslant (N\eta)^{1/6}$ and $q \leqslant (N\eta)^{1/6}$ separately. Together with $N\eta \geqslant 1$ we obtain (4.12).

Finally, to show (4.13) we use (4.6) and (4.5) to write

$$G_{ij}^{(k)} - G_{ij} = G_{kj} \sum_l^{(k)} G_{il}^{(k)} \left(H_{lk} + \frac{f}{N}\right).$$

The right-hand side can be estimated similarly to the right-hand side of (4.22) to obtain (4.13); we omit the details. $\qquad\square$

The following lemma is a standard stability estimate of the self-consistent equation associated with the semicircle law (see e.g. [5, Lemma 5.5]). Here we give a version with a sharp constant.

Lemma 4.4. *Let $z = E + \mathrm{i}\eta \in \mathbf{S}$. Let $m, \tilde{m}$ be the solutions of the equation*

$$x^2 + zx + 1 = 0.$$

If s satisfies $s^2 + zs + 1 = r$ then

$$|s - m| \wedge |s - \tilde{m}| \leqslant \sqrt{|r|}. \tag{4.23}$$

Proof. We choose the branch cut on the positive real axis, which gives $\operatorname{Im} \sqrt{z} \geqslant 0$ for all $z \in \mathbb{C} \backslash \mathbb{R}_+$. Thus

$$m = \frac{-z + \sqrt{z^2 - 4}}{2} \quad \text{and} \quad \tilde{m} = \frac{-z - \sqrt{z^2 - 4}}{2}.$$

Also, we have

$$s = \frac{-z + \sqrt{z^2 - 4 + 4r}}{2}.$$

Then (4.23) follows from the fact that

$$|\sqrt{a+b} - \sqrt{a}| \wedge |\sqrt{a+b} + \sqrt{a}| \leqslant \frac{|b|}{\sqrt{|a| + |b|}} \leqslant \sqrt{|b|}$$

for any $a, b \in \mathbb{C}$ and any complex square root $\sqrt{\cdot}$. □

Next, we set up the bootstrapping argument. Fix $E \in \mathbb{R}$. Let $K := \max\{k \in \mathbb{N} : 1 - kN^{-2} \geqslant N^{-1}\}$, and for $k = 0, 1, \ldots, K$, we define the spectral parameter $z_k := E + i\eta_k$. We abbreviate

$$\zeta_k := \left(\frac{r}{q^2}\right)^{1/4} + \frac{r}{(N\eta_k)^{1/6}} + \frac{r}{(\log N + \log \eta_k)q} + \frac{f}{(N\eta_k)^{1/4}} \tag{4.24}$$

where $\eta_k := 1 - kN^{-2}$.

The main work is to prove the following result, which states that Theorem 2.2 holds for the spectral parameter z being in the lattice $\{z_0, z_1, \ldots, z_K\}$. To simplify notation, we use the variable $\xi := t/750$ instead of the t from Theorem 2.2.

Proposition 4.5. *Let $\xi > 0$. Suppose that $2 \leqslant r \leqslant q^2 \leqslant N$, and $750\,\xi\zeta_k \leqslant 1$ for all $k = 0, 1, \ldots, K$. We have*

$$\mathbb{P}\Big(\max_{i,j} |G_{ij}(z_k) - m(z_k)\delta_{ij}| > 360\xi\zeta_k \Big) \leqslant (3k+1)N^3\xi^{-r}. \tag{4.25}$$

Proof. Throughout the proof we assume $\xi \geqslant 1$, for otherwise the proof is trivial. We proceed by induction on k. Since the sequence ζ_k is increasing, we may without loss of generality assume that the condition $1200\,\xi\zeta_k \leqslant 1$ holds for all $k = 0, \ldots, K$. Since the spectral parameter z varies in the proof, in this proof we always indicate it explicitly.

We define the events

$$\Omega_k := \big\{|s(z_k) - m(z_k)| \leqslant 50\xi\zeta_k\big\}, \quad \Xi_k := \{\Gamma(z_k) \leqslant 3/2\}$$

for $k = 0, 1, \ldots, K$.

For $k = 0$, we see that $\Gamma(z_0) \leqslant 1$ and hence $\phi(z_0) = 1$. Thus $\mathbb{P}(\Xi_0) = 1$. From (4.8) we find

$$1 + zs + s^2 = \frac{1}{N}\sum_i G_{ii}Y_i,$$

so that, by Minkowski's inequality and (4.11),

$$\|1 + z_0 s(z_0) + s(z_0)^2\|_r = \|\phi(1 + z_0 s(z_0) + s(z_0)^2)\|_r \leqslant \max_i \|\phi(z_0)Y_i(z_0)G_{ii}(z_0)\|_r$$

$$\leqslant 48^2\Big[\left(\frac{r}{q^2}\right)^{1/2} + \frac{r^2}{(N\eta_0)^{1/3}} + \left(\frac{r}{(\log N + \log \eta_0)q}\right)^2 + \frac{f^2}{\sqrt{N\eta_0}}\Big] \leqslant (50\zeta_0)^2.$$

16

Thus Chebyshev's inequality implies

$$\mathbb{P}\Big(|1 + z_0 s(z_0) + s(z_0)^2| > (50\xi\zeta_0)^2\Big) \leqslant \xi^{-2r} \tag{4.26}$$

Hence Lemma 4.4 yields

$$\mathbb{P}\Big(|s(z_0) - m(z_0)| \wedge |s(z_0) - \tilde{m}(z_0)| > 50\xi\zeta_0\Big) \leqslant \xi^{-2r}. \tag{4.27}$$

Note that $\operatorname{Im} s > 0$, and for $\eta_0 = 1$, we easily see that $\operatorname{Im} \tilde{m}(z_0) < -1$. Thus $|s(z_0) - \tilde{m}(z_0)| > 1$, and together with $50\xi\zeta_0 \leqslant 1$ we get $\mathbb{P}(\Omega_0^c) \leqslant \xi^{-2r}$. Similarly, by (4.11) and $50\xi\zeta_0 \leqslant 1$ we have

$$\max_i \mathbb{P}\Big(|Y_i(z_0)G_{ii}(z_0)| > 50\xi\zeta_0\Big) \leqslant \max_i \mathbb{P}\Big(|Y_i(z_0)G_{ii}(z_0)| > (50\xi\zeta_0)^2\Big) \leqslant \xi^{-2r}. \tag{4.28}$$

From (4.8) we get

$$1 + (z + m)G_{ii} = (m - s)G_{ii} + Y_i G_{ii},$$

so that by $m^2 + zm + 1 = 0$ and the elementary estimate $|m| \leqslant 1$ we have

$$|G_{ii} - m| \leqslant |(s - m)G_{ii}| + |Y_i G_{ii}|. \tag{4.29}$$

By $\mathbb{P}(\Omega_0^c) \leqslant \xi^{-2r}$ and (4.28)–(4.29) we have

$$\max_i \mathbb{P}(|G_{ii}(z_0) - m(z_0)| > 150\xi\zeta_0) \leqslant 2\xi^{-2r}. \tag{4.30}$$

Similarly, by (4.12) we have

$$\max_{i \neq j} \mathbb{P}(|G_{ij}(z_0)| > 12\xi\zeta_0) \leqslant \xi^{-r}. \tag{4.31}$$

Thus a union bound shows

$$\mathbb{P}\Big(\max_{i,j} |G_{ij}(z_0) - m(z_0)\delta_{ij}| > 150\xi\zeta_0\Big) \leqslant 2N\xi^{-2r} + N^2\xi^{-r} \leqslant 2N^2\xi^{-r}, \tag{4.32}$$

where in the last step we used without loss of generality $\xi \geqslant 1$ (for otherwise the probability bound is trivial). This proves the case $k = 0$.

For $k \geqslant 1$, we introduce the threshold index $\tilde{K} := \min\{k : k \leqslant K, 105\xi\zeta_k \geqslant |m(z_k) - \tilde{m}(z_k)|\}$. Note that

$$|m - \tilde{m}|^4 = \eta^4 + 2E^2\eta^2 + 8\eta^2 + E^4 + 16 - 8E^2 \tag{4.33}$$

is an increasing function of η. Thus, the sequence $|m(z_k) - \tilde{m}(z_k)|$ is decreasing and the sequence ζ_k is increasing. Hence $105\xi\zeta_k < |m(z_k) - \tilde{m}(z_k)|$ for all $k < \tilde{K}$.

Case 1: $1 \leqslant k < \tilde{K}$. Since $|d\Gamma/d\eta| \leqslant 1/\eta^2$, we have $\phi(z_k) = 1$ on Ξ_{k-1}. As in (4.27), we have

$$\mathbb{P}\Big(\phi|s(z_k) - m(z_k)| \wedge \phi|s(z_k) - \tilde{m}(z_k)| > 50\,\xi\zeta_k\Big) \leqslant \xi^{-2r}. \tag{4.34}$$

By Lipschitz continuity, $N\eta \geqslant 1$, and $\xi \geqslant 1$, we have

$$|s(z_k) - m(z_k)| \leqslant 50\xi\zeta_{k-1} + \frac{2}{N^2\eta^2} \leqslant 52\xi\zeta_k$$

on Ω_{k-1}. Note that for $k < K$ we have

$$|m(z_k) - \tilde{m}(z_k)| \geqslant 105\,\xi\zeta_k,$$

17

thus on Ω_{k-1} we have $|s(z_k) - m(z_k)| \wedge |s(z_k) - \tilde{m}(z_k)| = |s(z_k) - m(z_k)|$. We have therefore proved that

$$\mathbb{P}(\Omega_{k-1} \cap \Xi_{k-1} \cap \Omega_k^c) \leqslant \xi^{-2r}. \tag{4.35}$$

As in (4.30), we can use (4.11) and (4.29) to show that

$$\max_i \mathbb{P}\Big(\Omega_{k-1} \cap \Xi_{k-1} \cap \{|G_{ii}(z_k) - m(z_k)| > 150\,\xi\zeta_k\}\Big) \leqslant 2\xi^{-2r}. \tag{4.36}$$

Also, by (4.12) and (4.13) we see that

$$\max_{i \neq j} \mathbb{P}(\Xi_{k-1} \cap \{|G_{ij}(z_k)| > 12\xi\zeta_k\}) \leqslant \xi^{-r} \tag{4.37}$$

and

$$\max_{i,j \neq l} \mathbb{P}(\Xi_{k-1} \cap \{|G_{ij}(z_k) - G_{ij}^{(l)}| > 12\xi\zeta_k\}) \leqslant \xi^{-r}. \tag{4.38}$$

Thus by $|m| \leqslant 1$ and a union bound we have

$$\mathbb{P}(\Omega_{k-1} \cap \Xi_{k-1} \cap \{\Gamma(z_k) > 1 + 162\,\xi\zeta_k\}) \leqslant 2N\xi^{-2r} + (N^2 + N^3)\xi^{-r} \leqslant 2N^3\xi^{-r}. \tag{4.39}$$

Note that the assumption $750\xi\zeta_k \leqslant 1$ ensures $162\,\xi\zeta_k < 1/2$, we find

$$\mathbb{P}(\Omega_{k-1} \cap \Xi_{k-1} \cap \Xi_k^c) \leqslant 2N^3\xi^{-r}. \tag{4.40}$$

Note that $\mathbb{P}(\Omega_0 \cap \Xi_0) \geqslant 1 - \xi^{-2r} \geqslant 1 - \xi^{-r}$, thus (4.35), (4.40), and an induction argument shows

$$\mathbb{P}(\Omega_k \cap \Xi_k) \geqslant 1 - (2kN^3 + 1)\xi^{-r} \tag{4.41}$$

for all $k < \tilde{K}$. Thus (4.25) for $1 \leqslant k < \tilde{K}$ follows from (4.36), (4.37), and a union bound.

Case 2: $\tilde{K} \leqslant k \leqslant K$. As in Case 1, we have $\phi(z_k) = 1$ on Ξ_{k-1}, and

$$\mathbb{P}\Big(\phi|s(z_k) - m(z_k)| \wedge \phi|s(z_k) - \tilde{m}(z_k)| > 50\,\xi\zeta_k\Big) \leqslant \xi^{-2r}. \tag{4.42}$$

Note that $|m(z_k) - \tilde{m}(z_k)| \leqslant 105\xi\zeta_k$, so that the triangle inequality yields

$$\mathbb{P}(\Xi_{k-1} \cap \{|s(z_k) - m(z_k)| > 155\xi\zeta_k\}) \leqslant \xi^{-2r}. \tag{4.43}$$

As in (4.36), (4.37), and (4.40), we can show that

$$\max_i \mathbb{P}(\Xi_{k-1} \cap \{|G_{ii}(z_k) - m(z_k)| > 360\xi\zeta_k\}) \leqslant 2\xi^{-2r}, \tag{4.44}$$

$$\max_{i \neq j} \mathbb{P}(\Xi_{k-1} \cap \{|G_{ij}(z_k)| > 12\xi\zeta_k\}) \leqslant \xi^{-r}, \tag{4.45}$$

and

$$\mathbb{P}(\Xi_{k-1} \cap \Xi_k^c) \leqslant 2N^3\xi^{-r}, \tag{4.46}$$

where in showing (4.46) we used the fact that $750\xi\zeta_k \leqslant 1$. Note that (4.41) shows $\mathbb{P}(\Xi_{\tilde{K}-1}) \geqslant 1 - (2(\tilde{K} - 1)N^3 + 1)\xi^{-r}$, so that together with (4.46) we have

$$\mathbb{P}(\Xi_k) \geqslant 1 - (2kN^3 + 1)\xi^{-r} \tag{4.47}$$

for all $\tilde{K} \leqslant k \leqslant K$. Thus (4.25) for $\tilde{K} \leqslant k \leqslant K$ follows from (4.44), (4.45), and a union bound. $\quad\square$

18

We may now easily conclude the proof of Theorem 2.2. Without loss of generality, we can assume that $r \leqslant q^2$, for otherwise the condition (2.2) implies that $t \leqslant 1$, in which case Theorem 2.2 is trivially true. By using $K \leqslant N^2 - N$, $\xi = t/750$, and $r \geqslant 10$, we deduce from (4.25) that for $k = 0, 1, \ldots, K$, we have

$$\mathbb{P}\Big(\max_{i,j} |G_{ij}(z_k) - m(z_k)\delta_{ij}| > t\zeta_k/2 \Big) \leqslant N^5 \Big(\frac{1000}{t}\Big)^r. \tag{4.48}$$

Note that $|\mathrm{d}G/\mathrm{d}\eta|, |\mathrm{d}m/\mathrm{d}\eta| \leqslant 1/\eta^2$, and ζ is a decreasing function of η. Thus for any $z \in \mathbf{S}$, we have

$$\mathbb{P}\Big(\max_{i,j} |G_{ij} - m\delta_{ij}| > \frac{t\zeta}{2} + \frac{2}{N^2\eta^2} \Big) \leqslant N^5 \Big(\frac{1000}{t}\Big)^r.$$

Note that the above is trivial for $t \leqslant 1000$, thus by $1/(N\eta)^2 \leqslant \zeta$ we get

$$\mathbb{P}\Big(\max_{i,j} |G_{ij} - m\delta_{ij}| > t\zeta \Big) \leqslant N^5 \Big(\frac{1000}{t}\Big)^r. \tag{4.49}$$

This proves Theorem 2.2 for $C_* = 1000$.

A. Proof of (2.4)

In (2.3), let $r = \log N$ and $t = C_* \mathrm{e}^{5+D}$ for some $D > 0$, which yields

$$\mathbb{P}\Big[\max_{i,j} |G_{ij} - m\delta_{ij}| \leqslant C_* \mathrm{e}^{5+D}\zeta \Big] \geqslant 1 - N^{-D}.$$

Let us work under the assumption $\eta \geqslant N^{-1+\tau}$ for some $\tau > 0$. By the definition of ζ in (2.1), in order to show $C_* \mathrm{e}^{5+D}\zeta < \delta$ for some given $\delta > 0$, it suffices to show

$$C_* \mathrm{e}^{5+D} \max\left\{ \Big(\frac{\log N}{q^2}\Big)^{1/4}, \ \frac{\log N}{(N\eta)^{1/6}}, \ \frac{\log N}{(\log \eta + \log N)q}, \ \frac{f}{(N\eta)^{1/4}} \right\} < \delta/4. \tag{A.1}$$

This leads to the three conditions

$$q \geqslant \max\left\{ \Big(\frac{4C_* \mathrm{e}^{5+D}}{\delta}\Big)^2 \sqrt{\log N}, \ \frac{4C_* \mathrm{e}^{5+D}}{\tau\delta} \right\}, \quad f \leqslant \frac{\delta}{4C_* \mathrm{e}^{5+D}} \cdot N^{\tau/4}, \quad N^\tau \geqslant \Big(\frac{4C_* \mathrm{e}^{5+D} \log N}{\delta}\Big)^6. \tag{A.2}$$

where the first condition comes from the first and third terms in (A.1), the second condition comes from the last term in (A.1), and the last condition comes from the second term in (A.1).

- In order to prove (2.4) for $q \geqslant C\sqrt{\log N}$, we shall assume

$$\tau \geqslant (\log N)^{-1/2} \tag{A.3}$$

  for the rest of this appendix. (One easily checks that for a much smaller choice of $\tau \ll (\log N)^{-1/2}$, the first condition in (A.2) implies $q \gg \sqrt{\log N}$, i.e. q has to be much larger than the critical scale.)

- By the first condition in (A.2), we need

$$q \geqslant C\sqrt{\log N}, \quad \text{where} \quad C \equiv C(\delta, D) := \Big(\frac{4C_* \mathrm{e}^{5+D}}{\delta}\Big)^2. \tag{A.4}$$

- By the second condition in (A.2) and $f \leqslant q$, it suffices to have

$$q \leqslant (\log N)^{10} \tag{A.5}$$

and

$$(\log N)^{10} \leqslant \frac{\delta}{4C_* e^{5+D}} \cdot N^{(\log N)^{-1/2}/4},$$

where the latter can be simplified to

$$N \geqslant \exp\left[\left(e^{10} \log\left(\frac{4C_* e^{5+D}}{\delta}\right)\right)^2\right] =: N_0(\delta, D). \tag{A.6}$$

- One easily checks that $N \geqslant N_0$ satisfies the last condition in (A.2).

  Thus we see that (2.4) is true provided (A.3)–(A.6) holds.

References

[1] O. Ajanki, L. Erdős, and T. Krüger, *Quadratic vector equations on complex upper half-plane*, arXiv preprint arXiv:1506.05095.

[2] ———, *Universality for general Wigner-type matrices*, Prob. Theor. Rel. Fields **169** (2017), 667–727.

[3] R. Bauerschmidt, J. Huang, and H.-T. Yau, *Local Kesten-McKay law for random regular graphs*, Preprint arXiv:1609.09052.

[4] R. Bauerschmidt, A. Knowles, and H.-T. Yau, *Local semicircle law for random regular graphs*, Comm. Pure Appl. Math. **70** (2017), 1898–1960.

[5] F. Benaych-Georges and A. Knowles, *Lectures on the local semicircle law for Wigner matrices*, Preprint arXiv:1601.04055 (2016).

[6] Florent Benaych-Georges, Charles Bordenave, and Antti Knowles, *Largest eigenvalues of sparse inhomogeneous Erdős-Rényi graphs*, Preprint arXiv:1704.02953 (2017).

[7] ———, *Spectral radii of sparse random matrices*, Preprint arXiv:1704.02945 (2017).

[8] I. Dumitriu and Y. Zhu, *Sparse general Wigner-type matrices: Local law and eigenvector delocalization*, Preprint arXiv:1808.07611 (2018).

[9] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, *Spectral statistics of Erdős-Rényi graphs II: Eigenvalue spacing and the extreme eigenvalues*, Comm. Math. Phys. **314** (2012), 587–640.

[10] ———, *Delocalization and diffusion profile for random band matrices*, Comm. Math. Phys. **323** (2013), 367–416.

[11] ———, *Spectral statistics of Erdős-Rényi graphs I: Local semicircle law*, Ann. Prob. **41** (2013), 2279–2375.

[12] L. Erdős, B. Schlein, and H.-T. Yau, *Local semicircle law and complete delocalization for Wigner random matrices*, Comm. Math. Phys. **287** (2009), 641–655.

[13] ———, *Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices*, Ann. Prob. **37** (2009), 815–852.

[14] L. Erdős, H.-T. Yau, and J. Yin, *Bulk universality for generalized Wigner matrices*, Prob. Theor. Rel. Fields **154** (2012), 341–407.

[15] Y. He and A. Knowles, *Mesoscopic eigenvalue statistics of Wigner matrices*, Ann. Appl. Prob. **27** (2017), 1510–1550.

[16] Y. He, A. Knowles, and R. Rosenthal, *Isotropic self-consistent equations for mean-field random matrices*, Preprint arXiv:1611.05364.

[17] J. Huang and B. Landon, *Bulk universality of sparse random matrices*, J. Math. Phys. **56** (2015).

[18] J. Huang, B. Landon, and H.-T. Yau, *Transition from Tracy-Widom to Gaussian fluctuations of extremal eigenvalues of sparse Erdős–Rényi graphs*, Preprint arXiv: 1712.03936 (2017).

[19] J.O. Lee and K. Schnelli, *Local law and Tracy-Widom limit for sparse random matrices*, Preprint arXiv:1605.08767.

[20] L.V. Tran, V.H. Vu, and K. Wang, *Sparse random graphs: Eigenvalues and eigenvectors*, Random Structures & Algorithms (2010).

[21] R. van der Hofstad, *Random graphs and complex networks. Volume 1*, Cambridge Series in Statistical and Probabilistic Mathematics, 2017.