Spectrum of large random Markov chains: heavy-tailed weights on the oriented complete graph

Charles Bordenave, Pietro Caputo, Djalil Chafaï, Daniele Piras

SPECTRUM OF LARGE RANDOM MARKOV CHAINS: HEAVY-TAILED WEIGHTS ON THE ORIENTED COMPLETE GRAPH

CH. BORDENAVE, P. CAPUTO, D. CHAFAÏ, AND D. PIRAS

Abstract. We consider the random Markov matrix obtained by assigning i.i.d. non-negative weights to each edge of the complete oriented graph. In this study, the weights have unbounded first moment and belong to the domain of attraction of an alpha-stable law. We prove that as the dimension tends to infinity, the empirical measure of the singular values tends to a probability measure which depends only on alpha, characterized as the expected value of the spectral measure at the root of a weighted random tree. The latter is a generalized two-stage version of the Poisson weighted infinite tree (PWIT) introduced by David Aldous. Under an additional smoothness assumption, we show that the empirical measure of the eigenvalues tends to a non-degenerate isotropic probability measure depending only on alpha and supported on the unit disc of the complex plane. We conjecture that the limiting support is actually formed by a strictly smaller disc.

Contents

1. Introduction

A natural way to construct a random $n \times n$ Markov matrix is to assign i.i.d. (independent and identically distributed) non-negative weights $X_{i,j}$ with a given law $\mathcal{L}$ to each ordered pair (i, j), $i, j = 1, \dots, n$, and then consider the stochastic matrix M obtained by normalizing each row with the corresponding row sum:

$$M_{i,j} := \frac{X_{i,j}}{\rho_i} \quad \text{where} \quad \rho_i := \sum_{j=1}^{n} X_{i,j} \,. \tag{1.1}$$

Date: Summer 2016. Compiled October 6, 2016.

2000 *Mathematics Subject Classification.* 47A10; 15A52; 05C80.

Key words and phrases. Spectral theory; Objective method; Operator convergence; Logarithmic potential; Random matrix; Random Graph; Heavy tailed distribution; Stable law.

Support: A*MIDEX project ANR-11-IDEX-0001-02 funded by the "Investissements d'Avenir" French Government program, managed by the French National Research Agency (ANR).

If $\rho_i = 0$ for some i, it is understood that $M_{i,i} = 1$ and $M_{i,j} = 0$ for all $j \neq i$. The random matrix M has independent rows but non-independent columns. The eigenvalues of M, that is the complex roots of $\det(M - z)$, are denoted by $\lambda_1, \ldots, \lambda_n$. Since M is a stochastic matrix, one has $\lambda_1 = 1$ and $|\lambda_j| \leqslant 1$, $j = 2, \ldots, n$. If the symmetry of the weights $X_{i,j} = X_{j,i}$ is imposed, then the resulting Markov chain is reversible with respect to the row sum measure ρ_i, and has real spectrum. Such reversible models have been studied in [4, 5]. Here we consider the case where $X_{i,j}$ and $X_{j,i}$ are independent. In this case the Markov chain is non-reversible and has complex eigenvalues.

If $\mathcal{L}$ has finite second moment, it was shown in [7] that the spectrum is asymptotically uniformly distributed in the disc of radius $\sigma/\sqrt{n}$, where σ^2 denotes the variance of $\mathcal{L}$, i.e.

$$\frac{1}{n}\sum_{k=1}^{n} \delta_{\lambda_k \sqrt{\frac{n}{\sigma^2}}} \overset{\rightsquigarrow}{\scriptstyle n\to\infty} \mathcal{U}\,, \tag{1.2}$$

where $\mathcal{U}$ is the uniform law on the unit disc $\mathbb{D} = \{z \in \mathbb{C} : |z| \leqslant 1\}$ and $\rightsquigarrow$ denotes the weak convergence of probability measures with respect to continuous and bounded test functions. Similar results were recently obtained for discrete matrices with given row sums [15], for the ensemble of uniformly random doubly stochastic matrices [12, 14] and for random matrices with exchangeable entries [1].

In this paper we consider the infinite variance case. We shall actually restrict our attention to the particularly interesting case where the law $\mathcal{L}$ of the entries has infinite first moment. Our main results can be formulated as follows. We assume that for some $\alpha \in (0, 1)$, the random variables $X_{i,j}$ are i.i.d. copies of a random variable $\mathbf{x}$ satisfying the assumptions below:

(H1) $\mathbf{x}$ is a non-negative random variable such that

$$c := \lim_{t\to\infty} t^{\alpha}\, \mathbb{P}(\mathbf{x} \geqslant t) > 0. \tag{1.3}$$

(H2) $\mathbf{x}$ has a bounded probability density function.

It is well known that a random variable satisfying (H1) is in the domain of attraction of an α-stable law. An example of random variable satisfying both (H1) and (H2) is $\mathbf{x} = U^{-1/\alpha}$, where U is any bounded non-negative random variable with continuous probability density φ on $[0, \infty)$ such that $\varphi(0) > 0$.

We recall that for every fixed i, the ordered rearrangement of the random row vector $\{M_{i,j},\, j = 1, \ldots, n\}$ converges weakly to the Poisson-Dirichlet distribution with parameter $\alpha \in (0, 1)$; see [16], [5, Lem. 2.4]. Thus, one expects that the distribution of eigenvalues of M converges to a nontrivial probability measure on the unit disc $\mathbb{D}$ without any further rescaling. This is what we prove in Theorem 1.2 below. The proof, following the "objective method" philosophy [3], will be based on the construction of an infinite random tree that can be identified as a suitable local limit of the random matrix M.

As usual for the non Hermitian setting, the analysis of the eigenvalues starts with the understanding of the asymptotic behavior of the singular values of $M - z$, for $z \in \mathbb{C}$. Here and below, if no confusion arises, we write z for the diagonal matrix zI. Consider the singular values $s_{k,z}$, $k = 1, \ldots, n$, that is the eigenvalues of $\sqrt{(M - z)(M - z)^*}$, and write $\nu_{M,z}$ for the associated empirical distribution:

$$\nu_{M,z} := \frac{1}{n}\sum_{k=1}^{n} \delta_{s_{k,z}}. \tag{1.4}$$

Theorem 1.1 (Singular values)**.** *If (H1) holds, then for each $z \in \mathbb{C}$ there exists a probability measure $\nu_{\alpha,z}$ on $[0, \infty)$ depending only on α and $|z|$ such that almost surely*

$$\nu_{M,z} \overset{\rightsquigarrow}{\scriptstyle n\to\infty} \nu_{\alpha,z}. \tag{1.5}$$

For any $z \in \mathbb{C}$ the measure $\nu_{\alpha,z}$ has unbounded support and satisfies, for all $\lambda > 0$,

$$\int_0^\infty \mathrm{e}^{\lambda s} \nu_{\alpha,z}(\mathrm{d}s) < \infty. \tag{1.6}$$

The proof of Theorem 1.1 is based on the local convergence of the $2n \times 2n$ Hermitian matrix

$$B_z := \begin{pmatrix} 0 & M - z \\ M^\top - \bar{z} & 0 \end{pmatrix} \tag{1.7}$$

to the self-adjoint operator associated to a rooted random weighted infinite tree. In particular, the measure $\nu_{\alpha,z}$ will be interpreted as the expected value of the spectral measure associated to this random rooted tree. While this line of reasoning is entirely parallel to the arguments introduced in [6], an important difference here is that the resulting tree is the outcome of a branching process where two distinct offspring distributions alternate at each generation; see Section 2 below. In contrast with [6], the row sum normalization in M introduces dependencies in the random weights of the limiting tree, making the recursive distributional equation characterizing the spectral measure harder to analyze; see Section 3.4 below.

Note that in contrast with the eigenvalues, the distribution of the singular values of M has unbounded support. On the other hand, unlike the case of singular values of i.i.d. heavy tailed matrices [6], it has finite exponential moments.

Next, we turn to the empirical distribution of the eigenvalues of M:

$$\mu_M := \frac{1}{n} \sum_{k=1}^{n} \delta_{\lambda_k} . \tag{1.8}$$

Theorem 1.2 (Eigenvalues)**.** *If $(H1)$ and $(H2)$ hold then there exists a probability measure μ_α on the unit disc $\mathbb{D}$, depending only on α, such that almost surely*

$$\mu_M \underset{n\to\infty}{\rightsquigarrow} \mu_\alpha. \tag{1.9}$$

Moreover the probability measure μ_α is isotropic and non-degenerate.

The proof of Theorem 1.2 is based on Girko's Hermitization method [13], that is we use logarithmic potentials together with the convergence of the singular values of $M - z$ expressed by Theorem 1.1; see [10] for a survey of the method. A crucial point of this approach is the uniform integrability of the logarithmic function for the measures $\nu_{M,z}$, as $n \to \infty$. Two key facts are needed to establish this property. The first step is the proof that for almost all $z \in \mathbb{C}$ the smallest singular value of $M - z$ is bounded below by an inverse polynomial of n with high probability; see Theorem 4.2 below. This estimate uses the assumption (H2). The second step is an adaptation of the Tao and Vu analysis [19] of the singular values in the bulk. These steps are approached by a combination of methods introduced in [6] and [7]. In both cases, however, the present setting requires a nontrivial extension of the known arguments.

We refer to Figure 1 for simulation plots of the spectrum of M.

Remark 1.3 (Edge behavior: conjectures)**.** *The pictures in Figure 1 prompt the conjecture that the spectral measure μ_α is supported on a disc $\mathbb{D}_\alpha := \{z \in \mathbb{C} : |z| \leqslant r_\alpha\}$ for some $r_\alpha \in (0,1)$, where $r_\alpha \to 1$ as $\alpha \to 0$ and $r_\alpha \to 0$ as $\alpha \to 1$. A closer look at simulations actually suggests an even stronger conjecture, namely that with high probability, except for the eigenvalue $\lambda_1 = 1$, the Markov matrix M has a spectral radius strictly less than 1:*

$$\max\{|\lambda_2|, \ldots, |\lambda_n|\} \leqslant r_\alpha \,, \tag{1.10}$$

for some constant $r_\alpha \in (0,1)$ as above. Heuristic arguments seem to suggest that $r_\alpha \sim \sqrt{1-\alpha}$. A somewhat related question is the long time behavior of the Markov chain with transition matrix M. This question is addressed in [9]*; see also* [8] *for related recent progress concerning sparse random directed graphs.*

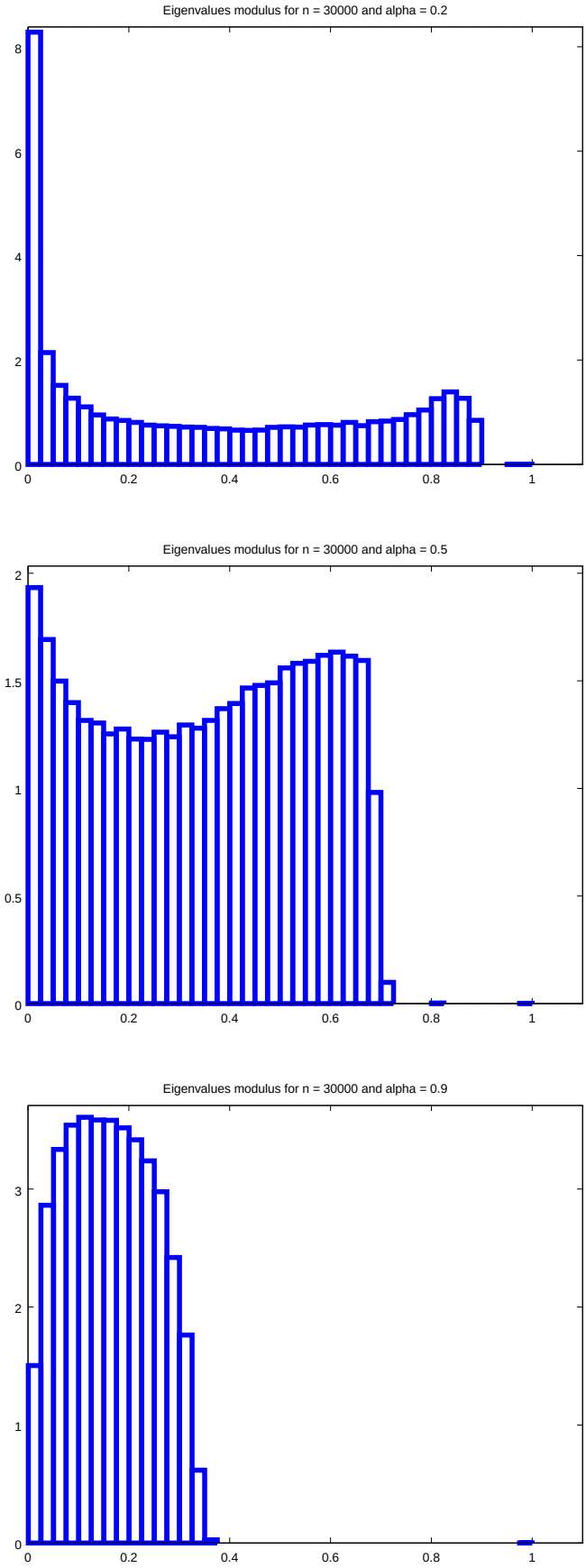

FIGURE 1. Histogram of the modulus of the spectrum of a single simulation of M in dimension $n = 30000$ and with tail index α equal to 0.2, 0.5, and 0.9 respectively, from top to bottom. The simulation uses $\mathbf{x} = U^{-1/\alpha}$, where U is uniform in $[0, 1]$.

A remark concerning the case where the variable $\mathbf{x}$ satisfies (H1) with $\alpha \in (1,2)$: in analogy with results from [5, 6] we expect that in this case the rescaled eigenvalues $\lambda'_k := n^{1-1/\alpha}\lambda_k$ satisfy

$$\frac{1}{n}\sum_{k=1}^{n} \delta_{\lambda'_k} \underset{n\to\infty}{\rightsquigarrow} \mu'_\alpha\,, \tag{1.11}$$

where μ'_α is the isotropic probability measure with unbounded support on $\mathbb{C}$, associated to the i.i.d. matrix X; see [6, Th. 1.2]. The case $\alpha = 1$ should require a logarithmic correction.

The rest of this article is organized as follows. In Section 2 we obtain some preliminary properties of the model and describe the alternate Poisson weighted tree that will be used in the proof of Theorem 1.1. The latter is given in Section 3. Finally, in Section 4 we prove Theorem 1.2.

2. Local convergence to the alternate Poisson weighted tree

The key conceptual step is to associate to our matrix M a limiting random weighted tree. We follow what is by now a well established strategy; see [2, 3, 11, 5]. We start with some facts on convergence to Poisson point processes for sequences of heavy tailed random variables; see e.g. [18] for more background.

2.1. Convergence of rows and columns of the Markov matrix.

If $0 \leqslant x_1 \leqslant x_2 \leqslant \cdots$ is distributed as the Poisson point process with intensity 1 on $[0,\infty)$, then

$$\xi_i = x_i^{-1/\alpha}, \tag{2.1}$$

is distributed as the Poisson point process with intensity $\phi(t) = \alpha t^{-\alpha-1}$ on $[0,\infty)$. In this case we say that the ranked vector $\{\xi_i\}$ has law PPP(α). Moreover, for $\alpha \in (0,1)$, the variable $S = \sum_{i=1}^{\infty} \xi_i$ has the one sided α-stable law, with Laplace transform

$$\mathbb{E}[\exp(-\theta S)] = \exp\left(-\Gamma(1-\alpha)\theta^\alpha\right), \quad \theta \geqslant 0. \tag{2.2}$$

The law of the normalized ranked vector

$$(\zeta_1, \zeta_2, \dots) := S^{-1}(\xi_1, \xi_2, \cdots), \tag{2.3}$$

is called the Poisson-Dirichlet law with index α; see [16] for a detailed account. We shall refer to it as PD(α). The next lemma summarizes the key convergence results for the rows of the matrix M. We recall that here convergence in distribution of a ranked vector coincides with weak convergence, for every fixed $k \in \mathbb{N}$, of the joint law of the first k maxima of the vector. We write $[n]$ for the set $\{1,\dots,n\}$.

Lemma 2.1 (Poisson Point Processes and Poisson Dirichlet distributions)**.** *Let $X_{i,j}$, $i,j \in [n]$ be i.i.d. copies of a random variable $\mathbf{x}$ satisfying (H1). Define $\rho_i = \sum_{j=1}^n X_{i,j}$ and $M_{i,j} = \rho_i^{-1} X_{i,j}$. Set $a_n := (c\,n)^{1/\alpha}$, where c is as in* (1.3)*. Then*

1. *The ranked rearrangement of $\{a_n^{-1} X_{1,j},\, j \in [n]\}$ converges in distribution to $\{\xi_i\}$ which is a Poisson Point Process PPP(α);*
2. *The random variable $a_n^{-1}\rho_1$ converges in distribution to S given in* (2.2)*;*
3. *The ranked rearrangement of $\{M_{1,j},\, j \in [n]\}$ converges in distribution to $\{\zeta_i\}$ which follows the Poisson-Dirichlet law PD(α).*

The above lemma is well known; we refer e.g. to [5, Lem. 2.4] for a proof. We turn to column vectors. Here the convergence result is less immediate. We start with an auxiliary lemma. While Lemma 2.1 does not require the full strength of the assumption (H1), the usual regular variation assumption being sufficient [5], below we do use (H1) in a more stringent way.

Lemma 2.2. *In the setting of Lemma 2.1, for all fixed $t > 0$ one has*

$$\lim_{n\to\infty} n\,\mathbb{P}\left(X_{1,1} > \tfrac{t}{1+t}\,\rho_1\right) = \gamma\, t^{-\alpha}, \tag{2.4}$$

where γ is the constant

$$\gamma = \gamma(\alpha) := \frac{1}{\Gamma(1+\alpha)\Gamma(1-\alpha)}.$$

Proof. For $u > 0$, write $\mathbb{P}(\mathbf{x} > u) = L(u)u^{-\alpha}$, for some function $L(u)$ such that $L(u) \to c$, $u \to \infty$, as in (1.3). Denote by μ_n the law of the variable $a_n^{-1}\hat{\rho}_1$, where $\hat{\rho}_1 := \sum_{i=2}^n X_{1,i}$. Since $X_{1,1} > \frac{t}{1+t}\,\rho_1$ is equivalent to $X_{1,1} > t\hat{\rho}_1$, by the independence of $X_{1,1}$ and $\hat{\rho}_1$ one has

$$n\,\mathbb{P}\left(X_{1,1} > \tfrac{t}{1+t}\,\rho_1\right) = n\,\mathbb{P}\left(X_{1,1} > t\hat{\rho}_1\right) = c^{-1}t^{-\alpha}\int_0^\infty s^{-\alpha}L(a_n ts)\mu_n(\mathrm{d}s). \tag{2.5}$$

From Lemma 2.1 we know that μ_n converges weakly to μ, the law of S in (2.2). We are going to prove that for any $t > 0$:

$$\lim_{n\to\infty} c^{-1}\int_0^\infty s^{-\alpha}L(a_n ts)\mu_n(\mathrm{d}s) = \int_0^\infty s^{-\alpha}\mu(\mathrm{d}s). \tag{2.6}$$

Notice that (2.6) proves (2.4) since by direct computation, using (2.2) and $\alpha\Gamma(\alpha) = \Gamma(1+\alpha)$,

$$\begin{aligned}\int_0^\infty s^{-\alpha}\mu(\mathrm{d}s) &= \frac{1}{\Gamma(\alpha)}\int_0^\infty \mathrm{d}x\, x^{\alpha-1}\int_0^\infty \mathrm{e}^{-xs}\mu(\mathrm{d}s)\\ &= \frac{1}{\Gamma(\alpha)}\int_0^\infty \mathrm{d}x\, x^{\alpha-1}\mathrm{e}^{-x^\alpha\Gamma(1-\alpha)} = \gamma.\end{aligned}$$

To prove (2.6), fix $\varepsilon > 0$ and start by observing that

$$\lim_{n\to\infty}\int_\varepsilon^\infty s^{-\alpha}L(a_n ts)\mu_n(\mathrm{d}s) = c\int_\varepsilon^\infty s^{-\alpha}\mu(\mathrm{d}s), \tag{2.7}$$

Indeed, $|L(a_n ts) - c| \to 0$ as $n \to \infty$, uniformly in $s \geqslant \varepsilon$. Also, $\int_\varepsilon^\infty s^{-\alpha}\mu_n(\mathrm{d}s) \to \int_\varepsilon^\infty s^{-\alpha}\mu(\mathrm{d}s)$ by weak convergence. This implies (2.7).

Next, we show that for any constant $K > 0$,

$$\lim_{n\to\infty}\int_0^{Ka_n^{-1}} s^{-\alpha}L(a_n ts)\mu_n(\mathrm{d}s) = 0. \tag{2.8}$$

The obvious bound $L(u) \leqslant u^\alpha$, $u > 0$, yields $L(a_n ts) \leqslant cs^\alpha t^\alpha n$. Therefore,

$$\int_0^{Ka_n^{-1}} s^{-\alpha}L(a_n ts)\mu_n(\mathrm{d}s) \leqslant c\,t^\alpha n\int_0^{Ka_n^{-1}}\mu_n(\mathrm{d}s). \tag{2.9}$$

Now, $\int_0^{Ka_n^{-1}}\mu_n(\mathrm{d}s) = \mathbb{P}(\hat{\rho}_1 \leqslant K)$, and $\hat{\rho}_1 \geqslant \max_{i=2,\dots,n} X_{1,i}$, so that

$$\int_0^{Ka_n^{-1}}\mu_n(\mathrm{d}s) \leqslant \mathbb{P}(\mathbf{x} \leqslant K)^n, \tag{2.10}$$

which decays exponentially in n for any fixed $K > 0$. This, together with (2.9), implies (2.8).

Finally, consider the integral

$$\int_{Ka_n^{-1}}^\varepsilon s^{-\alpha}L(a_n ts)\mu_n(\mathrm{d}s). \tag{2.11}$$

Since $s \geqslant K/a_n$, (1.3) shows that $L(a_n ts) \leqslant 2c$ if $K = K(t)$ is large enough. Therefore (2.11) is bounded by

$$2c\int_{Ka_n^{-1}}^\varepsilon s^{-\alpha}\mu_n(\mathrm{d}s) \leqslant 2c\int_{Ka_n^{-1}}^\varepsilon s^{-\alpha}\mu_n(\mathrm{d}s).$$

Thus, the proof of (2.6) is complete once we show

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty}\int_{Ka_n^{-1}}^{\varepsilon} s^{-\alpha}\mu_n(\mathrm{d}s)=0\,. \tag{2.12}$$

From (H1) and [6, Lem. 3.5], there exists $\eta>0$, $p\in(0,1)$ such that $\mathbf{x}$ dominates stochastically the product ηDS where D is a Bernoulli(p) variable independent of S defined in (2.2). Thus $a_n^{-1}\hat{\rho}_1$ stochastically dominates $\hat{S}_n := a_n^{-1}(D_1S_1+\cdots D_{n-1}S_{n-1})$ where $\{D_i\}$ and $\{S_i\}$ are independent, D_i are i.i.d. copies of D, S_i are i.i.d. copies of S. If $\hat{\mu}_n$ denotes the law of $\hat{S}_n$, then $\hat{\mu}_n=\sum_{k=0}^{n-1}p(k,n)\nu_k$ where $p(k,n)=\binom{n-1}{k}p^k(1-p)^{n-1-k}$, and ν_k is the law of $\eta a_n^{-1}(S_1+\cdots+S_k)$. By stability $S_1+\cdots+S_k$ has the same law of $k^{1/\alpha}S$, so that ν_k is the law of $\eta(k/cn)^{1/\alpha}S$. Let E be the event that $D_1+\cdots+D_{n-1}\geqslant pn/2$. We estimate

$$\int_{Ka_n^{-1}}^{\varepsilon} s^{-\alpha}\mu_n(\mathrm{d}s)\leqslant (K/a_n)^{-\alpha}\mathbb{P}(E^c)+\sum_{k=pn/2}^{n-1}p(k,n)\int_{Ka_n^{-1}}^{\varepsilon}s^{-\alpha}\nu_k(\mathrm{d}s). \tag{2.13}$$

$\mathbb{P}(E^c)$ decays to zero exponentially in n from the Chernoff bound. Thus the first term above vanishes in the limit $n\to\infty$. We now consider the second term. For any $k\in[pn/2,n-1]$ one has that ν_k is the law of λS, for a constant $\lambda=\eta(k/cn)^{1/\alpha}$ such that $\lambda\in[a,b]$ for some constants $0<a<b<\infty$. Then, uniformly in $k\in[pn/2,n-1]$,

$$\int_{Ka_n^{-1}}^{\varepsilon}s^{-\alpha}\nu_k(\mathrm{d}s)\leqslant a^{-\alpha}\int_0^{\varepsilon/a}x^{-\alpha}\mu(\mathrm{d}x).$$

Since $\int_0^{\varepsilon}x^{-\alpha}\mu(\mathrm{d}x)\to 0$ as $\varepsilon\to 0$, (2.13) implies (2.12). This ends the proof of (2.6). □

We can now state the main results concerning convergence of columns of M.

Lemma 2.3 (Ranked rearrangements)**.** *In the setting of Lemma 2.1, let $\widetilde{X}_{j,1}$, $j\in[n]$ denote the ranked rearrangement of $\{X_{j,1},\,j\in[n]\}$, write $\pi=\pi^n$ for the permutation of $[n]$ such that $\widetilde{X}_{i,1}=X_{\pi(i),1}$ for all $i\in[n]$ (use e.g. lexicographic order to break ties if necessary), and define $\hat{M}_{j,1}:=M_{\pi(j),1}$.*

1. *For any $k\in\mathbb{N}$, $(\hat{M}_{1,1},\dots,\hat{M}_{k,1})$ converges in distribution to $(\hat{\omega}_1,\dots,\hat{\omega}_k)$, defined by*

$$\hat{\omega}_i:=\frac{\xi_i}{\xi_i+S_i}\,, \tag{2.14}$$

   where $\{\xi_i\}$ is the sequence in (2.1) with law PPP(α), and $\{S_i\}$ is an independent sequence of i.i.d. copies of the random variable S in (2.2).
2. *The ranked rearrangement of $\{M_{j,1},\,j\in[n]\}$ converges in distribution to the ranked rearrangement of $\{\hat{\omega}_i\}$. The latter, in turn has the same law of the ranked sequence*

$$\omega_i:=\frac{\xi_i}{\xi_i+q}\,, \tag{2.15}$$

   with $\{\xi_i\}$ as above, and q is the constant $q=\gamma^{-1/\alpha}=(\Gamma(1+\alpha)\Gamma(1-\alpha))^{1/\alpha}$.

Proof. Define $\hat{\rho}_j:=\sum_{i=2}^n X_{j,i}$, so that

$$M_{j,1}=\frac{X_{j,1}}{X_{j,1}+\hat{\rho}_j}. \tag{2.16}$$

It follows that

$$\hat{M}_{j,1}=\frac{X_{\pi(j),1}}{X_{\pi(j),1}+\hat{\rho}_{\pi(j)}}.$$

Since $\{X_{\pi(j),\ell},\ell\neq 1,j\in[n]\}$ are independent of $\{X_{\pi(j),1},j\in[n]\}$, parts 1 and 2 of Lemma 2.1 imply convergence in distribution of $(\hat{M}_{1,1},\dots,\hat{M}_{k,1})$ to $(\hat{\omega}_1,\dots,\hat{\omega}_k)$ for any fixed k. Note that these sequences are not necessarily ranked.

To prove part 2, for any $\ell, k \in [n]$, $\ell < k$, let $E_{\ell,k}(n)$ denote the event that the largest ℓ values of $\{M_{j,1},\, j \in [n]\}$ are not included in the sequence $(\hat{M}_{1,1}, \dots, \hat{M}_{k,1})$, and call $p_{\ell,k}(n)$ its probability. Below, we observe that for any fixed $\ell \in \mathbb{N}$,

$$\lim_{k\to\infty} \limsup_{n\to\infty} p_{\ell,k}(n) = 0. \tag{2.17}$$

Once (2.17) is available, then it is not difficult to see that the largest ℓ values of $\{M_{j,1},\, j \in [n]\}$ and the largest ℓ values of $\{\hat{M}_{j,1},\, j \in [n]\}$ have the same limit in distribution, for any fixed ℓ. It follows from part 1 that the ranked rearrangement of $\{M_{j,1},\, j \in [n]\}$ converges in distribution to the ranked rearrangement of $\{\hat{\omega}_i\}$.

To prove (2.17), we observe that the event $E_{\ell,k}(n)$ implies that there exists $j_* \in [n]$ such that $M_{j_*,1}$ is larger than the ℓ-th maximum of $\{M_{\pi(i),1},\, i = 1, \dots, k\}$ and $X_{j_*,1} < X_{\pi(k),1}$. In particular, there must be $j_* \in [n]$ and $i_* \in \{1, \dots, \ell\}$ such that $M_{j_*,1} > M_{\pi(i_*),1}$ and $X_{j_*,1} < X_{\pi(k),1}$. Therefore, see (2.16), one must have

$$\hat{\rho}_{j_*} < X_{j_*,1} \max_{i=1,\dots,\ell} \frac{\hat{\rho}_{\pi(i)}}{X_{\pi(i),1}}.$$

From part 1 of the lemma it follows that $\max_{i=1,\dots,\ell} \frac{\hat{\rho}_{\pi(i)}}{X_{\pi(i),1}}$ converges in distribution to $\max_{i=1,\dots,\ell} \frac{S_i}{\xi_i}$, and therefore, for all fixed $\ell \in \mathbb{N}$:

$$\lim_{t\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\max_{i=1,\dots,\ell} \frac{\hat{\rho}_{\pi(i)}}{X_{\pi(i),1}} > t \right) = 0.$$

Thus, it suffices to prove that for any $t > 0$:

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\exists j_* \in [n] :\ \hat{\rho}_{j_*} < t X_{j_*,1} \quad \text{and} \quad X_{j_*,1} < X_{\pi(k),1} \right) = 0. \tag{2.18}$$

From Lemma 2.1, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(X_{\pi(k),1} > \varepsilon a_n) = \lim_{k\to\infty} \mathbb{P}(\xi_k > \varepsilon) = 0.$$

Thus, using a union bound, to prove (2.18) we can restrict ourselves to the proof that for any fixed $t > 0$:

$$\lim_{\varepsilon\to 0} \limsup_{n\to\infty} n\, \mathbb{P}\left(\hat{\rho}_1 < t X_{1,1} < \varepsilon a_n\right) = 0. \tag{2.19}$$

However, as in (2.7) we have

$$n\, \mathbb{P}\left(\hat{\rho}_1 < t X_{1,1} < \varepsilon a_n\right) \leqslant c^{-1} t^{\alpha} \int_0^{\varepsilon} s^{-\alpha} L(t^{-1} a_n s) \mu_n(\mathrm{d}s). \tag{2.20}$$

From the estimates (2.8) and (2.12) one obtains (2.19). This concludes the proof of (2.17).

It remains to prove that the ranked values of the limiting sequence have the same law of $\{\omega_i\}$ in (2.15). To this end, define $W_j := X_{j,1}/\hat{\rho}_j$, so that

$$M_{j,1} = \frac{X_{j,1}}{X_{j,1} + \hat{\rho}_j} = \frac{1}{1 + \frac{1}{W_j}}. \tag{2.21}$$

The W_j are i.i.d. random variables. From Lemma 2.2 we know that, as $n \to \infty$,

$$n\, \mathbb{P}(W_1 > t) \to (qt)^{-\alpha}, \quad t > 0. \tag{2.22}$$

Then, by a well known criterion, see [18, Th. 5.3], the ranked rearrangement of $\{q\, W_j,\, j \in [n]\}$ converges in distribution to $\{\xi_i\}$ with law $\mathrm{PPP}(\alpha)$, and from (2.21) it follows that the ranked rearrangement of $(M_{j,1},\, j = 1, \dots, n)$ converges in distribution to $\{\omega_i\}$ as claimed. □

2.2. **The alternate Poisson weighted tree.** Here we define the limiting random rooted tree to be associated to the matrix M. We start with some standard notation. The vertex set V of the tree is deterministic, and it is defined as the countable set of all points $\mathbf{k}$ of the form $\mathbf{k} = (k_1, \dots, k_m)$ for some $m \in \mathbb{N}$ and $k_i \in \mathbb{N}$, for all $i = 1, \dots, m$, together with the root vertex $\varnothing$. If $\mathbf{k} = (k_1, \dots, k_m) \in V$, and $j \in \mathbb{N}$, we write $(\mathbf{k}, j)$ or simply $\mathbf{k}j$ for the vertex $(k_1, \dots, k_m, j) \in V$. The interpretation is that $\mathbf{k}j$ is the j-th child of vertex $\mathbf{k}$. Thus, if $\mathbf{k} = \varnothing$ is the root, then $\varnothing j$ or simply j denotes the j-th vertex in the first generation. Note that for any $\mathbf{k} \in V$, $\mathbf{k} \neq \varnothing$, there is a unique pair $(\mathbf{v_k}, i_\mathbf{k})$, with $\mathbf{v_k} \in V$, and $i_\mathbf{k} \in \mathbb{N}$, such that $\mathbf{k} = \mathbf{v_k}\, i_\mathbf{k}$. The vertex $\mathbf{v_k}$ is called the parent of $\mathbf{k}$. If $\mathbf{k} = (k_1, \dots, k_m)$ for some $m \in \mathbb{N}$, we write $d(\mathbf{k}) := m$, and say that $\mathbf{k}$ belongs to the m-th generation, or equivalently that $\mathbf{k}$ is at distance m from the root.

Let $\xi^{(\mathbf{k})} = \{\xi_i^{(\mathbf{k})}\}$, $\mathbf{k} \in V$, denote independent copies of the Poisson process $\{\xi_i\}$ with law PPP(α), as in (2.1). We can now define a random rooted tree $\mathcal{T}_0$ as the undirected weighted tree with vertex set V, with root at $\varnothing$, with edges connecting vertices $\mathbf{k}, \mathbf{k}' \in V$ if and only if $\mathbf{k}' = \mathbf{k}j$ or $\mathbf{k} = \mathbf{k}'j$ for some $j \in \mathbb{N}$, and with edges $\{\mathbf{k}, \mathbf{k}j\}$ carrying the weight $\xi_j^{(\mathbf{k})}$. We often use the notation

$$\mathcal{T}_0(\mathbf{k}, \mathbf{k}j) = \mathcal{T}_0(\mathbf{k}j, \mathbf{k}) = \xi_j^{(\mathbf{k})}, \tag{2.23}$$

for the weight along the undirected edge $\{\mathbf{k}, \mathbf{k}j\}$.

Next, let us define $S_\mathbf{k} = \sum_{i=1}^\infty \xi_i^{(\mathbf{k})}$. Clearly, $S_\mathbf{k}$, $\mathbf{k} \in V$ are independent copies of the one sided random variable with law (2.2). Let us modify the tree $\mathcal{T}_0$ as follows. We say that $\mathbf{k} \in V$ is even (respectively odd) if $d(\mathbf{k})$ is even (respectively odd). The root counts as an even vertex. Let $\mathcal{T}_+$ denote the tree with the same vertex set and edge set as $\mathcal{T}_0$ but with weights

$$\widehat{\mathcal{T}}_+(\mathbf{k}, \mathbf{k}j) = \widehat{\mathcal{T}}_+(\mathbf{k}j, \mathbf{k}) = \hat\zeta_j^{(\mathbf{k})} \mathbf{1}_{\{\mathbf{k} \text{ even}\}} + \hat\omega_j^{(\mathbf{k})} \mathbf{1}_{\{\mathbf{k} \text{ odd}\}}, \tag{2.24}$$

where

$$\hat\zeta_j^{(\mathbf{k})} := \frac{\xi_j^{(\mathbf{k})}}{\xi_{i_\mathbf{k}}^{(\mathbf{v_k})} + S_\mathbf{k}}, \quad \hat\omega_j^{(\mathbf{k})} := \frac{\xi_j^{(\mathbf{k})}}{\xi_j^{(\mathbf{k})} + S_{\mathbf{k}j}}. \tag{2.25}$$

As before, in (2.25) the pair $(\mathbf{v_k}, i_\mathbf{k})$ is defined by the relation $\mathbf{k} = (\mathbf{v_k}, i_\mathbf{k})$, and we use the convention that if $\mathbf{k} = \varnothing$, then $\xi_{i_\mathbf{k}}^{(\mathbf{v_k})} = 0$. In particular, the weights of the edges from the root to the first generation have the law PD(α) as in (2.3). We remark that with these definitions one has

$$\hat\zeta_j^{(\mathbf{k})} = (1 - \hat\omega_{i_\mathbf{k}}^{(\mathbf{v_k})}) \frac{\xi_j^{(\mathbf{k})}}{S_\mathbf{k}}. \tag{2.26}$$

Also, remark that the weights $\{\hat\omega_j^{(\mathbf{k})},\ j \in \mathbb{N}\}$ are independent of all weights of the form $\widehat{\mathcal{T}}_+(\mathbf{v}, \mathbf{v}i)$ for any $\mathbf{v}$ such that $d(\mathbf{v}) < d(\mathbf{k})$. In particular, the randomness starts afresh at every odd generation.

For any $\mathbf{k} \in V$ one has $\hat\zeta_j^{(\mathbf{k})} \geqslant \hat\zeta_{j+1}^{(\mathbf{k})}$. On the other hand the values $\hat\omega_j^{(\mathbf{k})}$, $j \in \mathbb{N}$ are not necessarily ranked. As in Lemma 2.3 we may consider the ranked rearrangement of the weights $\{\hat\omega_j^{(\mathbf{k})}\}$. Proceeding top to bottom from the root one can then define a new tree $\mathcal{T}_+$ isomorphic to $\hat{\mathcal{T}}_+$ and such that all weights are ranked in the sense that

$$\mathcal{T}_+(\mathbf{k}, \mathbf{k}j) \geqslant \mathcal{T}_+(\mathbf{k}, \mathbf{k}(j+1)), \qquad \mathbf{k} \in V,\ j \in \mathbb{N}.$$

In a similar fashion, we may define the weighted rooted tree

$$\widehat{\mathcal{T}}_-(\mathbf{k}, \mathbf{k}j) = \widehat{\mathcal{T}}_-(\mathbf{k}j, \mathbf{k}) = \hat\zeta_j^{(\mathbf{k})} \mathbf{1}_{\{\mathbf{k} \text{ odd}\}} + \hat\omega_j^{(\mathbf{k})} \mathbf{1}_{\{\mathbf{k} \text{ even}\}}, \tag{2.27}$$

where $\hat{\zeta}_j^{(\mathbf{k})}$ and $\hat{\omega}_j^{(\mathbf{k})}$ are defined by (2.25). Then, by taking ranked rearrangements one defines the tree $\mathcal{T}_-$ such that

$$\mathcal{T}_-(\mathbf{k}, \mathbf{k}j) \geqslant \mathcal{T}_-(\mathbf{k}, \mathbf{k}(j+1)), \qquad \mathbf{k} \in V,\ j \in \mathbb{N}.$$

2.3. **Adjacency operators on trees.** Consider the Hilbert space $\ell^2(V)$ of square integrable sequences $\varphi : V \mapsto \mathbb{C}$, with the scalar product defined by

$$\langle \varphi, \varphi' \rangle = \sum_{\mathbf{k} \in V} \overline{\varphi}(\mathbf{k}) \varphi'(\mathbf{k}).$$

The indicator functions $\mathbf{1}_\mathbf{k}$ defined by $\mathbf{1}_\mathbf{k}(\mathbf{k}') = \delta_{\mathbf{k},\mathbf{k}'}$ form an orthonormal basis. The weights (2.23) produce almost surely a densely defined symmetric operator $\mathcal{T}_0$ acting on $\ell^2(V)$ by the formula

$$\langle \mathbf{1}_\mathbf{k}, \mathcal{T}_0 \mathbf{1}_{\mathbf{k}'} \rangle := \begin{cases} \mathcal{T}_0(\mathbf{k}, \mathbf{k}j) & \text{if } \mathbf{k}' = \mathbf{k}j \\ \mathcal{T}_0(\mathbf{k}', \mathbf{k}'j) & \text{if } \mathbf{k} = \mathbf{k}'j \\ 0 & \text{if } \mathbf{k} \not\sim \mathbf{k}' \end{cases} \tag{2.28}$$

where $\mathbf{k} \not\sim \mathbf{k}'$ indicates that $(\mathbf{k}, \mathbf{k}')$ is not an edge of the tree, that is $\mathbf{k}' \neq \mathbf{k}j$ and $\mathbf{k} \neq \mathbf{k}'j$ for all $j \in \mathbb{N}$. In the same way, using the matrix elements (2.24) and (2.27) we define the operators $\widehat{\mathcal{T}}_+$ and $\widehat{\mathcal{T}}_-$ respectively, together with their ranked versions $\mathcal{T}_+$ and $\mathcal{T}_-$. All the operators defined above have dense domain $\mathcal{D}$, defined as the set of finitely supported vectors, and are symmetric in $\ell^2(V)$. With a slight abuse of notation we identify them with their closure. It is crucial for our purposes that these operators are actually self-adjoint; see e.g. [17] for background on self-adjointness of a symmetric unbounded operator.

Lemma 2.4. *The operators $\mathcal{T}_0$, $\widehat{\mathcal{T}}_\pm$, and $\mathcal{T}_\pm$ are almost surely self-adjoint.*

Proof. Self-adjointness of $\mathcal{T}_0$ is shown in [5, Prop. A.2]. To prove the other statements we may adapt the same argument. Let us consider the case of $\widehat{\mathcal{T}}_+$, the others being very similar. Notice that for any even vertex $\mathbf{k}$ one has

$$\sum_{\mathbf{v} \in V} \widehat{\mathcal{T}}_+(\mathbf{k}, \mathbf{v}) = 1. \tag{2.29}$$

On the other hand, for $\mathbf{k}$ odd and $\kappa > 0$, define the variables

$$\tau_\kappa(\mathbf{k}) = \inf \Big\{ t \geqslant 0 : \sum_{j=t+1}^{\infty} \omega_j^{(\mathbf{k})} \leqslant \kappa \Big\}, \tag{2.30}$$

where $\{\omega_j^{(\mathbf{k})}, j \in \mathbb{N}\}$ denotes the ranked rearrangement of $\{\hat{\omega}_j^{(\mathbf{k})}, j \in \mathbb{N}\}$. For any $\kappa > 0$, the random variables $\{\tau_\kappa(\mathbf{k}),\ \mathbf{k} \text{ odd}\}$ are independent and identically distributed. Thus, the very same proof of [5, Prop. A.2] applies, provided we prove that for fixed $\mathbf{k}$ odd, one has $\mathbb{E}\tau_\kappa(\mathbf{k}) < \infty$ for all $\kappa > 0$ and $\mathbb{E}\tau_\kappa(\mathbf{k}) \to 0$, as $\kappa \to \infty$. Now, by Lemma 2.3, $\tau_\kappa(\mathbf{k})$ has the same law of the first $t \geqslant 0$ such that $\sum_{j=t+1}^{\infty} \frac{\xi_j}{\xi_{j+q}} \leqslant \kappa$; see (2.15). In particular, it is stochastically dominated by the random variable

$$\bar{\tau}_\kappa := \inf \Big\{ t \geqslant 0 : \sum_{j=t+1}^{\infty} \xi_j \leqslant q\kappa \Big\},$$

which satisfies $\mathbb{E}\bar{\tau}_\kappa < \infty$ and $\mathbb{E}\bar{\tau}_\kappa \to 0$, $\kappa \to \infty$, by [5, Lem. A.4]. □

2.4. **The unfolding map.** Fix a vertex $i_0 \in [n]$ and two integers $b, h \in \mathbb{N}$. Let $V_{b,h} \subset V$ denote the set of $\mathbf{k} \in V$ of the form $\mathbf{k} = (k_1, \dots, k_\ell)$, where $k_i \in \{1, \dots, b\}$, and $\ell \leqslant h$, together with the root vertex $\varnothing$. That is, $V_{b,h}$ is the finite subset obtained from V by considering only the first b children of each node, and stopping at the h-th generation. As usual if $\mathbf{k}, \mathbf{k}' \in V_{b,h}$ we write $\mathbf{k} \sim \mathbf{k}'$ iff $\mathbf{k}' = \mathbf{k}j$ for $\mathbf{k} = \mathbf{k}'j$ for some j. We now define a map $\phi_+ : V_{b,h} \to [n]$ (respectively $\phi_- : V_{b,h} \to [n]$) revealing the alternating ranked rearrangements of rows and columns of the matrix X starting with the row labeled i_0 (respectively, starting with the column labeled i_0). The map

$$\phi_+ : V_{b,h} \to [n],$$

is defined as follows. Start with $\phi_+(\varnothing) := i_0$. Let $X_{i_0,\pi(j)}$, $j = 1, \dots, b$, denote the first b terms of the ranked rearrangement of the row $(X_{i_0,j}, j \in [n], j \neq i_0)$ (as usual we break ties using the lexicographic ordering if necessary). Set $\phi_+(j) := \pi(j)$, $j = 1, \dots, b$. Thus far we have defined the map ϕ_+ on the root vertex and the first generation vertices. To uncover the second generation, set $I = \{i_0\}$, and let $X_{\pi_1(j),\pi(1)}$, $j = 1, \dots, b$, denote the first b terms of the ranked rearrangement of the column $(X_{j,\pi(1)}, j \in [n], j \notin I)$, and set $\phi_+(1j) := \pi_1(j)$, $j = 1, \dots, b$. Rename $I = \{i_0, \pi_1(1), \dots, \pi_1(b)\}$, let $X_{\pi_2(j),\pi(2)}$, $j = 1, \dots, b$ denote the first b terms of the ranked rearrangement of the column $(X_{j,\pi(2)}, j \in [n], j \notin I)$, and set $\phi_+(2j) := \pi_2(j)$, $j = 1, \dots, b$. Repeating this procedure for the columns labeled $\pi(3), \dots, \pi(b)$ one completes the definition of ϕ_+ for second generation of $V_{b,h}$. To complete the definition of the map ϕ_+ we proceed recursively (redefining at each step the set I so that all rows revealed so far are excluded from the ranking) in the same way, with the rule that if we are revealing the m-th generation, then we look at rankings of the rows labeled by the vertices revealed in the $(m-1)$-th generation if m is odd, and rankings of the columns labeled by the vertices revealed in the $(m-1)$-th generation if m is even. This ends the definition of ϕ_+. To define the map $\phi_- : V_{b,h} \to [n]$ we proceed in the exact same way, with the role of rows and columns exchanged. Equivalently one may define ϕ_- as the map ϕ_+ obtained by replacing X with $X^\top$.

Example 2.5. Let us give the explicit values of $\phi_\pm$ in a simple example. Suppose $n = 5$ and consider the realization

$$X = \begin{pmatrix} 0.1 & 3.2 & 2.1 & 4 & 0.2 \\ 0 & 1.2 & 3.3 & 3.4 & 1.7 \\ 0.4 & 10.3 & 0.1 & 2 & 3 \\ 0.2 & 3.1 & 1.67 & 5 & 11 \\ 8 & 4.7 & 1.2 & 1.98 & 2 \end{pmatrix}. \tag{2.31}$$

Suppose $i_0 = 3$ so that $\phi_\pm(\varnothing) = 3$. The highest value in row 3 is 10.3 corresponding to column 2, thus $\phi_+(1) = 2$. The second highest value in row 3 is 3 corresponding to column 5, thus $\phi_+(2) = 5$. The two highest values in column 2 are 4.7 and 3.2 corresponding to row 5 and row 1 respectively, thus $\phi_+(11) = 5$ and $\phi_+(12) = 1$. Next, if we eliminate row 1 and row 5, the two highest values in column 5 are 11 and 1.7 corresponding to row 4 and row 2 respectively, thus $\phi_+(21) = 4$ and $\phi_+(22) = 2$. On the other hand, scanning first columns and then rows gives the map ϕ_- depicted below.

$\mathbf{k}$	$\varnothing$	1	2	11	12	21	22
$\phi_+(\mathbf{k})$	3	2	5	5	1	4	2
$\phi_-(\mathbf{k})$	3	2	1	4	5	2	1

(2.32)

We now go back to the general setting. To link our original matrix X with the limiting tree $\mathcal{T}_0$ we need a further step, namely bipartization. Let A denote the $2n \times 2n$ bipartite symmetric matrix

$$A = \begin{pmatrix} 0 & X \\ X^\top & 0 \end{pmatrix}. \tag{2.33}$$

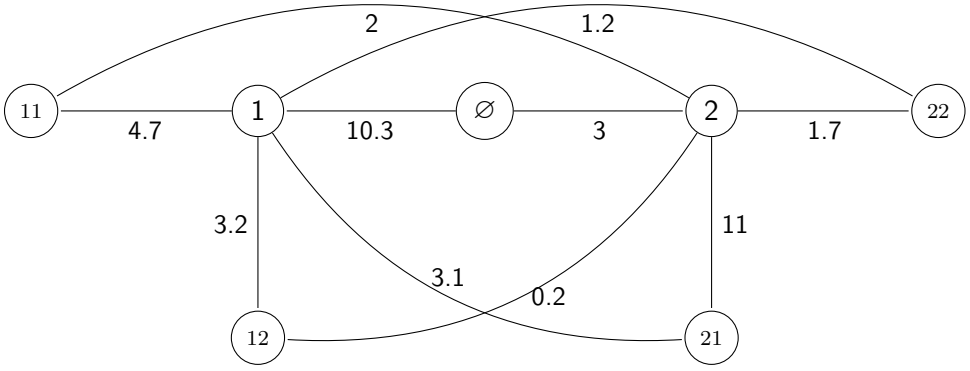

FIGURE 2. The network $A^{2,2}_{+,5}$ corresponding to the matrix X in Example 2.5. If the bended edges are removed, then one obtains the network $\widetilde{A}^{2,2}_{+,5}$ used in the proof of Lemma 2.6.

For any $\mathbf{k} \in V_{b,h}$, set

$$\psi_+(\mathbf{k}) := \begin{cases} \phi_+(\mathbf{k}) & \mathbf{k} \text{ even} \\ \phi_+(\mathbf{k}) + n & \mathbf{k} \text{ odd} \end{cases} \quad \text{and} \quad \psi_-(\mathbf{k}) := \begin{cases} \phi_-(\mathbf{k}) & \mathbf{k} \text{ odd} \\ \phi_-(\mathbf{k}) + n & \mathbf{k} \text{ even} \end{cases} \tag{2.34}$$

The above defines two injective maps $\psi_\pm : V_{b,h} \mapsto [2n]$. A network is a collection of vertices together with undirected weighted edges between them. Consider the finite random network $A^{b,h}_{\pm,n}$ with vertex set $V_{b,h}$ and with weighted edge between vertices $\mathbf{u}, \mathbf{v} \in V_{b,h}$ defined by $a_n^{-1} A_{u,v}$, where $u := \psi_\pm(\mathbf{u})$ and $v := \psi_\pm(\mathbf{v})$. Notice that the network $A^{b,h}_{\pm,n}$ has a fixed number of vertices but the value of edge weights depends on n. See Figure 2 for an explanatory example. At any finite n the network can well have cycles, but these tend to disappear as $n \to \infty$ as the next lemma shows. Weak convergence of finite random networks is defined as weak convergence of the joint law of edge weights in the natural way.

Lemma 2.6 (Convergence of i.i.d. networks to trees)**.** *The random networks $A^{b,h}_{+,n}$ and $A^{b,h}_{-,n}$ both converge weakly, as $n \to \infty$, to the restriction of the random tree $\mathcal{T}_0$ to the vertex set $V_{b,h}$, for any fixed $b, h \in \mathbb{N}$, and for any choice of the initial vertex i_0.*

Proof. By symmetry, $A^{b,h}_{+,n}$ and $A^{b,h}_{-,n}$ have the same law, thus it is sufficient to prove the statement for $A^{b,h}_{+,n}$ only. Let $\widetilde{A}^{b,h}_{+,n}$ denote the sub-network obtained from $A^{b,h}_{+,n}$ by giving weight zero to all edges of the form $\{\mathbf{u}, \mathbf{v}\}$ such that $\mathbf{u} \not\sim \mathbf{v}$. See Figure 2 for an example. Then, by construction, and using repeatedly part 1 of Lemma 2.1 one has that the sequence $\widetilde{A}^{b,h}_{+,n}$ converges weakly to the restriction of $\mathcal{T}_0$ to the vertex set $V_{b,h}$. To prove the desired convergence then it suffices to prove that the discarded weights in $A^{b,h}_{+,n}$ converge to zero in probability. This can be checked by a stochastic domination argument as already done in [5, Prop. 2.6]. □

Next, we turn to our matrix M with normalized rows. Let B denote the $2n \times 2n$ bipartite symmetric matrix

$$B = \begin{pmatrix} 0 & M \\ M^\top & 0 \end{pmatrix}. \tag{2.35}$$

Let $\psi_\pm$ denote the same maps defined in (2.34), and consider the networks $B^{b,h}_{\pm,n}$ defined by the vertex set $V_{b,h}$ as above but with a weighted edge between any pair of vertices $\mathbf{u}, \mathbf{v} \in V_{b,h}$ defined by $B_{u,v}$, where $u := \psi_\pm(\mathbf{k})$ and $v := \psi_\pm(\mathbf{v})$. Note that we do not

require any rescaling now. Also note that in contrast with the case of the matrix A, the law of the finite networks $B^{b,h}_{+,n}$ and $B^{b,h}_{-,n}$ do not coincide.

Lemma 2.7 (Convergence of Markov networks to trees)**.** *The random network $B^{b,h}_{\pm,n}$ converges weakly, as $n \to \infty$, to the restriction of the random tree $\widehat{\mathcal{T}}_\pm$ to the vertex set $V_{b,h}$, for any fixed $b, h \in \mathbb{N}$, and for any choice of the initial vertex i_0.*

Proof. As in the proof of Lemma 2.6 we identify a modified network $\widetilde{B}^{b,h}_{\pm,n}$ which is easily seen to converge to the desired limit by an application of Lemma 2.3 and then prove that the difference between $B^{b,h}_{\pm,n}$ and $\widetilde{B}^{b,h}_{+,n}$ vanishes in probability. Let us start with $B^{b,h}_{+,n}$. Consider the modified network $\widetilde{B}^{b,h}_{+,n}$ obtained from $B^{b,h}_{+,n}$ by giving weight zero to any edge of the form $\{\mathbf{u}, \mathbf{v}\}$ such that $\mathbf{u} \not\sim \mathbf{v}$. Now, a recursive application Lemma 2.3 implies that the network $\widetilde{B}^{b,h}_{+,n}$ converges weakly, as $n \to \infty$, to the restriction of the random tree $\widehat{\mathcal{T}}_+$ to the vertex set $V_{b,h}$. There is a delicate point here due to the fact that while rows of M are independent, the columns are not. The dependence stems from the normalizing sums ρ_u, but with a little care it can be shown to be negligible in the limit. We carry out the details below for the first two generations of the tree. Notice that once the convergence of $\widetilde{B}^{b,h}_{+,n}$ is established, one can conclude by the same stochastic domination argument mentioned in the proof of Lemma 2.6 above.

Let us check the convergence of $\widetilde{B}^{b,h}_{+,n}$ in more detail. For simplicity of exposition we take $b = h = 2$. Here the network consists of 7 vertices, namely the $\varnothing, 1, 2, 11, 12, 21, 22$. Let G_n denote the network $\widetilde{B}^{2,2}_{+,n}$, and let $G_n(\mathbf{u}, \mathbf{v})$ denote the weight of edge $\{\mathbf{u}, \mathbf{v}\}$. Then by definition

$$\begin{aligned}
G_n(\varnothing, 1) &= X_{i_0,\pi(1)}/\rho_{i_0}, \\
G_n(\varnothing, 2) &= X_{i_0,\pi(2)}/\rho_{i_0}, \\
G_n(1, 11) &= X_{\pi_1(1),\pi(1)}/\rho_{\pi_1(1)}, \\
G_n(1, 12) &= X_{\pi_1(2),\pi(1)}/\rho_{\pi_1(2)}, \\
G_n(2, 21) &= X_{\pi_2(1),\pi(2)}/\rho_{\pi_2(1)}, \\
G_n(2, 22) &= X_{\pi_2(2),\pi(2)}/\rho_{\pi_2(2)}.
\end{aligned}$$

All other edges have weight zero. Set $\rho'_i := \rho_i - X_{i,\pi(1)}$, $i \in [n]$, and define a new network G'_n such that G'_n coincides with G_n except that on edges $\{2, 21\}$ and $\{2, 22\}$ one has $G'_n(2, 21) = X_{\pi_2(1),\pi(2)}/\rho'_{\pi_2(1)}$ and $G_n(2, 22) = X_{\pi_2(2),\pi(2)}/\rho'_{\pi_2(2)}$. Then there are no dependencies anymore and an application of Lemma 2.3 yields the desired convergence for the network G'_n. It remains to show that $G_n(2, 21) - G'_n(2, 21)$ and $G_n(2, 22) - G'_n(2, 22)$ converge to zero in probability. However, this follows from the fact that $a_n^{-1} X_{\pi_2(j),\pi(1)} \to 0$ in probability for $j = 1, 2$.

□

2.5. **Operator convergence.** We recall a notion of local convergence that we already used in [5, 6]. Let $\mathcal{D}$ denote the dense subspace of finitely supported vectors in $\ell^2(V)$. Given a sequence of bounded self-adjoint operators T_n, a self-adjoint operator $\mathcal{T}$ such that $\mathcal{D}$ is a core for $\mathcal{T}$, and a sequence of vertices $\mathbf{u}_n \in V$, we write $(T_n, \mathbf{u}_n) \to (\mathcal{T}, \varnothing)$ if there exists a sequence of bijections $\sigma_n : V \mapsto V$ such that $\sigma_n(\varnothing) = \mathbf{u}_n$ for all n and such that

$$\sigma_n^{-1} T_n \sigma_n \varphi \to \mathcal{T}\varphi, \qquad n \to \infty, \tag{2.36}$$

in $\ell^2(V)$, for any $\varphi \in \mathcal{D}$. With slight abuse of notation, above we have used the symbol σ_n to indicate both the bijection on V and the isometry acting on $\ell^2(V)$ induced via $\sigma_n \mathbf{1}_{\mathbf{v}} = \mathbf{1}_{\sigma_n(\mathbf{v})}$. In Section 3.1 below, the convergence of self-adjoint operators defined above will be used to derive convergence of the corresponding spectral measures.

Given the matrix A in (2.33), we define the operator $T_n(X)$ as follows. As usual, let $j \in \mathbb{N}$ denote the elements of V belonging to the first generation. Then, set

$$\langle \mathbf{1}_i, T_n(X)\mathbf{1}_j\rangle := A_{i,j}\mathbf{1}_{1\leqslant i,j\leqslant 2n}\,, \qquad i,j\in\mathbb{N}\,, \tag{2.37}$$

and $\langle \mathbf{1}_\mathbf{u}, T_n(X)\mathbf{1}_\mathbf{v}\rangle = 0$ for all $\mathbf{u},\mathbf{v}\in V$ that are not both in the first generation. By linearity, for every fixed n, (2.37) defines a random bounded self-adjoint operator $T_n(X)$ on $\ell^2(V)$. Similarly, we define the random bounded self-adjoint operators $T_n(M)$ by

$$\langle \mathbf{1}_i, T_n(M)\mathbf{1}_j\rangle := B_{i,j}\mathbf{1}_{1\leqslant i,j\leqslant 2n}\,, \qquad i,j\in\mathbb{N}\,, \tag{2.38}$$

for all $i,j\in\mathbb{N}$, and $\langle \mathbf{1}_\mathbf{u}, T_n(M)\mathbf{1}_\mathbf{v}\rangle = 0$ for all $\mathbf{u},\mathbf{v}\in V$ that are not both in the first generation, where B is the matrix in (2.35).

Theorem 2.8. *It is possible to realize the random operators $T_n(X), T_n^{\pm}(M)$ and $\mathcal{T}_0, \hat{\mathcal{T}}_\pm, \mathcal{T}_\pm$ on the same probability space in such a way that the following holds almost surely, for all $i_0\in\mathbb{N}$, as $n\to\infty$:*

1) $(T_n(X), i_0)\to(\mathcal{T}_0,\varnothing)$;
2) $(T_n(M), i_0)\to(\hat{\mathcal{T}}_+,\varnothing)$ and $(T_n(M), i_0+n)\to(\hat{\mathcal{T}}_-,\varnothing)$.
3) $(T_n(M), i_0)\to(\mathcal{T}_+,\varnothing)$ and $(T_n(M), i_0+n)\to(\mathcal{T}_-,\varnothing)$.

Proof. We proceed as in [6, Th. 2.10]. By Lemma 2.6 and Skorokhod's representation theorem one can realize the operators $T_n(X)$ and $\mathcal{T}_0$ on the same probability space in such a way that for all $b,h\in\mathbb{N}$, for all $\mathbf{u},\mathbf{v}\in V_{b,h}$ as $n\to\infty$, almost surely

$$\langle \mathbf{1}_{\psi_+(\mathbf{u})}, T_n(X)\mathbf{1}_{\psi_+(\mathbf{v})}\rangle \to \langle \mathbf{1}_\mathbf{u}, \mathcal{T}_0\mathbf{1}_\mathbf{v}\rangle, \tag{2.39}$$

where the injective map $\psi_+ : V_{b,h}\mapsto[2n]$ depends on b,h, and n. Thus, by a diagonal extraction, one can find a sequence of bijective maps $\sigma_n : V\mapsto V$ such that $\sigma_n(\varnothing)=i_0$, the fixed initial index, and almost surely

$$\langle \mathbf{1}_{\sigma_n(\mathbf{u})}, T_n(X)\mathbf{1}_{\sigma_n(\mathbf{v})}\rangle \to \langle \mathbf{1}_\mathbf{u}, \mathcal{T}_0\mathbf{1}_\mathbf{v}\rangle. \tag{2.40}$$

Thus, (2.40) expresses the convergence $\sigma_n^{-1}T_n(X)\sigma_n\to\mathcal{T}_0$, in the usual sense of weak convergence of operators on a Hilbert space. To turn this into the required strong convergence in (2.36) it is then sufficient to prove almost sure uniform (in n) square-integrability of the vector $\{\langle \mathbf{1}_\mathbf{u}, \sigma_n^{-1}T_n(X)\sigma_n\mathbf{1}_\mathbf{v}\rangle,\ \mathbf{u}\in V\}$ for any fixed $\mathbf{v}\in V$. The latter can be obtained as in [6, Th. 2.10]. This ends the proof of part 1.

To prove part 2, we proceed in a similar way. By Lemma 2.7 and Skorokhod's representation theorem one has that for all $b,h\in\mathbb{N}$, for all $\mathbf{u},\mathbf{v}\in V_{b,h}$ as $n\to\infty$, almost surely

$$\langle \mathbf{1}_{\psi_+(\mathbf{u})}, T_n(M)\mathbf{1}_{\psi_+(\mathbf{v})}\rangle \to \langle \mathbf{1}_\mathbf{u}, \hat{\mathcal{T}}_+\mathbf{1}_\mathbf{v}\rangle. \tag{2.41}$$

As above one can find a sequence of bijective maps $\sigma_n : V\mapsto V$ such that $\sigma_n(\varnothing)=i_0$, and almost surely

$$\langle \mathbf{1}_{\sigma_n(\mathbf{u})}, T_n(M)\mathbf{1}_{\sigma_n(\mathbf{v})}\rangle \to \langle \mathbf{1}_\mathbf{u}, \hat{\mathcal{T}}_+\mathbf{1}_\mathbf{v}\rangle. \tag{2.42}$$

Now, (2.42) proves that $\sigma_n^{-1}T_n(M)\sigma_n\to\hat{\mathcal{T}}_+$ weakly. The same integrability as above applies here and strong convergence follows. To prove that $(T_n(X), i_0+n)\to(\hat{\mathcal{T}}_-,\varnothing)$, we repeat the steps above, recalling that now $\psi_-(\varnothing)=i_0+n$ and therefore the bijection σ_n satisfies $\sigma_n(\varnothing)=i_0+n$.

To prove part 3, observe that from the argument in the proof of Lemma 2.3 we know that there are bijections $\eta_{b',h} : V_{b',h}\mapsto V_{b',h}$ such that, as $b'\to\infty$, for any fixed $b,h\in\mathbb{N}$, the restriction of $\eta_{b',h}^{-1}\hat{\mathcal{T}}_\pm\eta_{b',h}$ to $V_{b,h}$ converges to the restriction of $\mathcal{T}_\pm$ to $V_{b,h}$. One can then repeat the argument in (2.40) and (2.42) with $\psi_\pm$ replaced by the composed map $\psi_\pm\circ\eta_{b',h}$ with $b'=b'(n)$ a sequence with $b'(n)\to\infty$ as $n\to\infty$. □

3. Convergence of singular values

Here we prove Theorem 1.1. We shall need the following extension of Theorem 2.8. For any $z \in \mathbb{C}$, let $A(z)$ and $B(z)$ denote the matrices (cf. (2.33) and (2.35))

$$A(z) = \begin{pmatrix} 0 & X - z \\ X^\top - \bar{z} & 0 \end{pmatrix}, \qquad B(z) = \begin{pmatrix} 0 & M - z \\ M^\top - \bar{z} & 0 \end{pmatrix}. \tag{3.1}$$

Let also $T_n(X, z)$ and $T_n(M, z)$ denote the associated adjacency operators defined as in (2.37) and (2.38) respectively with A replaced by $A(z)$ and B by $B(z)$. To describe the local convergence of these new networks we proceed as follows. Let us start with $A(z)$. We define the limiting tree $\mathcal{T}_0(z)$ recursively. Consider the previously defined tree $\mathcal{T}_0$ rooted at $\varnothing$. Call $\mathcal{T}^1(z)$ the new tree obtained by appending an outgoing directed edge with weight $-z$ (respectively $-\bar{z}$) to every even vertex (respectively, odd vertex) of the tree, including the root. Thus $\mathcal{T}_0^1(z)$ is a tree with pending nodes (leaves) at the endpoints of the added directed edges. Next take i.i.d. copies of $\mathcal{T}_0$ and append one of them to each pending node in $\mathcal{T}_0^1(z)$. Now, call $\mathcal{T}_0^2(z)$ the tree obtained by appending an outgoing directed edge with weight $-z$ (respectively $-\bar{z}$) to every new even vertex (respectively, odd vertex) of the tree, where new means that it was not a vertex of $\mathcal{T}_0^1(z)$ (the leaves of $\mathcal{T}_0^1(z)$ do not count as new). Next take i.i.d. copies of $\mathcal{T}_0$ and append one of them to each pending node in $\mathcal{T}_0^2(z)$. Recursively, we then construct trees $\mathcal{T}_0^k(z)$, $k = 1, 2, \ldots$. Repeating ad libitum this procedure yields the random tree that we call $\mathcal{T}_0(z)$.

The trees $\hat{\mathcal{T}}_\pm(z)$ are defined in a similar way. For $\hat{\mathcal{T}}_+(z)$, we start with $\hat{\mathcal{T}}_+$ and produce the tree $\mathcal{T}_+^1(z)$ by appending an outgoing directed edge with weight $-z$ (respectively $-\bar{z}$) to every even vertex (respectively, odd vertex) of the tree, including the root. Then we append i.i.d. copies of $\hat{\mathcal{T}}_+$ (respectively $\hat{\mathcal{T}}_-$) to every even (respectively odd) pending node, and we proceed recursively as above. Here a leaf is even (respectively odd) if it is the endpoint of a directed edge emanating from an odd (respectively even) vertex. To define $\hat{\mathcal{T}}_-(z)$ we proceed exactly as above with the role of even and odd interchanged. If instead we use $\mathcal{T}_\pm$, the ranked versions of $\hat{\mathcal{T}}_\pm$, we obtain trees that we will call $\mathcal{T}_\pm(z)$.

All the above defined trees naturally define, via adjacency, linear operators on the Hilbert space $\ell^2(V)$. To ensure the symmetry we use the convention that when there is a directed edge $(\mathbf{k}, \mathbf{k}')$ with weight z there is also the opposite edge $(\mathbf{k}', \mathbf{k})$ with weight $\bar{z}$. The following theorem is a direct generalization of Lemma 2.4 and Theorem 2.8. We omit the details of the proof.

Theorem 3.1. *It is possible to realize the random operators $T_n(X, z), T_n^\pm(M, z)$ and $\mathcal{T}_0(z)$, $\hat{\mathcal{T}}_\pm(z)$, $\mathcal{T}_\pm(z)$ on the same probability space in such a way that the following holds almost surely, for all fixed $z \in \mathbb{C}$, $i_0 \in \mathbb{N}$, as $n \to \infty$:*

1) The operators $\mathcal{T}_0(z), \hat{\mathcal{T}}_\pm(z), \mathcal{T}_\pm(z)$ are self-adjoint;
2) $(T_n(X, z), i_0) \to (\mathcal{T}_0(z), \varnothing)$;
3) $(T_n(M, z), i_0) \to (\hat{\mathcal{T}}_+(z), \varnothing)$ and $(T_n(M, z), i_0 + n) \to (\hat{\mathcal{T}}_-(z), \varnothing)$.
4) $(T_n(M, z), i_0) \to (\mathcal{T}_+(z), \varnothing)$ and $(T_n(M, z), i_0 + n) \to (\mathcal{T}_-(z), \varnothing)$.

3.1. **The spectral measure.** For every $z \in \mathbb{C}$, let $\mu_\pm^{(z)}$ denote the spectral measure at the root of the self-adjoint operator $\hat{\mathcal{T}}_\pm(z)$. This is the unique probability measure on $\mathbb{R}$ such that for every $\eta \in \mathbb{C}_+ := \{w \in \mathbb{C} : \mathfrak{Im}(w) > 0\}$:

$$\langle \delta_\varnothing, (\hat{\mathcal{T}}_\pm(z) - \eta)^{-1} \delta_\varnothing \rangle = m_{\mu_\pm^{(z)}}(\eta) = \int \frac{\mu_\pm^{(z)}(\mathrm{d}x)}{x - \eta}, \tag{3.2}$$

where $(\hat{\mathcal{T}}_\pm(z) - \eta)^{-1}$ denotes the resolvent of $\hat{\mathcal{T}}_\pm(z)$ at $\eta \in \mathbb{C}_+$, and for any probability measure μ on $\mathbb{R}$ we let $m_\mu : \mathbb{C}_+ \mapsto \mathbb{C}_+$ denote the Cauchy-Stieltjes transform defined by $\int \frac{\mu(\mathrm{d}x)}{x - \eta}$. With these notation, $m_{\mu_\pm^{(z)}}(\eta)$ is a bounded random variable for each $\eta \in \mathbb{C}_+$.

From elementary spectral theory we know that for each fixed $n \in \mathbb{N}$ one has

$$\frac{1}{2n}\sum_{i=1}^{2n}\langle \delta_i, (T_n(M,z)-\eta)^{-1}\delta_i\rangle = \mathrm{tr}[(T_n(M,z)-\eta)^{-1}] = m_{\mu_n^{(M,z)}}(\eta), \tag{3.3}$$

where $\mu_n^{(M,z)}$ denotes the empirical spectral distribution of $T_n(M,z)$, i.e.

$$\mu_n^{(M,z)} = \frac{1}{2n}\sum_{i=1}^{2n}\delta_{\lambda_{i,z}}, \tag{3.4}$$

if $\lambda_{i,z} \geqslant \lambda_{i+1,z}$ are the ranked eigenvalues of $B(z)$ from (3.1). Moreover, it is well known that $\mu_n^{(M,z)}$ is symmetric on $\mathbb{R}$ and the singular values $s_{i,z}$ of $M - z$ satisfy $s_{i,z} = \lambda_{i,z}$, $i = 1, \ldots, n$.

From [17, Th. VIII.25(a)], see also [5, Th. 2.2], it is known that the local convergence $(T_n(M,z), i_0) \to (\hat{\mathcal{T}}_+(z), \varnothing)$ and $(T_n(M,z), i_0+n) \to (\hat{\mathcal{T}}_-(z), \varnothing)$ expressed in Theorem 3.1 implies convergence of resolvents, i.e. almost surely

$$\langle \delta_\varnothing, (\hat{\mathcal{T}}_+(z)-\eta)^{-1}\delta_\varnothing\rangle = \lim_{n\to\infty}\langle \delta_1, (T_n(M,z)-\eta)^{-1}\delta_1\rangle \tag{3.5}$$

$$\langle \delta_\varnothing, (\hat{\mathcal{T}}_-(z)-\eta)^{-1}\delta_\varnothing\rangle = \lim_{n\to\infty}\langle \delta_{1+n}, (T_n(M,z)-\eta)^{-1}\delta_{1+n}\rangle \tag{3.6}$$

In particular, the above convergence holds in distributions. Since the random variables are bounded, one has convergence of the expected values. Thus, using linearity and exchangeability in (3.3), it follows that for any $\eta \in \mathbb{C}_+$

$$\begin{aligned} & m_{\mathbb{E}\,\mu_n^{(M,z)}}(\eta) = \mathbb{E}\, m_{\mu_n^{(M,z)}}(\eta) \\ & = \tfrac{1}{2}\mathbb{E}\langle \delta_1, (T_n(M,z)-\eta)^{-1}\delta_1\rangle + \tfrac{1}{2}\mathbb{E}\langle \delta_{1+n}, (T_n(M,z)-\eta)^{-1}\delta_{1+n}\rangle \\ & \to \tfrac{1}{2}\mathbb{E}\, m_{\mu_+^{(z)}}(\eta) + \tfrac{1}{2}\mathbb{E}\, m_{\mu_-^{(z)}}(\eta) = m_{\bar{\mu}^{(z)}}(\eta)\,, \qquad n\to\infty\,, \end{aligned} \tag{3.7}$$

where $\bar{\mu}^{(z)}$ denotes the probability measure

$$\bar{\mu}^{(z)} = \tfrac{1}{2}\mathbb{E}\,\mu_+^{(z)} + \tfrac{1}{2}\mathbb{E}\,\mu_-^{(z)}. \tag{3.8}$$

3.2. **Proof of Theorem 1.1.** From (3.7) we know that

$$m_{\mathbb{E}\,\mu_n^{(M,z)}}(\eta) \to m_{\bar{\mu}^{(z)}}(\eta),$$

for all $\eta \in \mathbb{C}_+$. It is well known that convergence of the Cauchy-Stieltjes transform implies weak convergence. Therefore $\mathbb{E}\mu_n^{(M,z)} \rightsquigarrow \bar{\mu}^{(z)}$ as $n \to \infty$. Since M has independent rows, thanks to the general concentration result in [6, Lem. C.2], it follows that $\mu_n^{(M,z)} \rightsquigarrow \bar{\mu}^{(z)}$ as $n \to \infty$, almost surely. Finally, as observed after (3.4), $\nu_{M,z}$ is the image of $\mu_n^{(M,z)}$ under reflection. Letting $\nu_{\alpha,z}$ denote the image of $\bar{\mu}^{(z)}$ under reflection, we have proved that $\nu_{M,z} \rightsquigarrow \nu_{\alpha,z}$ as $n \to \infty$, almost surely, for each $z \in \mathbb{C}$. That $\nu_{\alpha,z}$ depends on $z \in \mathbb{C}$ only through the absolute value $|z|$, and has bounded exponential moments is shown in Lemma 3.2 below. Unbounded support is proven in Lemma 3.3. This will complete the proof of Theorem 1.1. In Section 3.4 we shall discuss the recursive distributional equations satisfied by the Cauchy-Stieltjes transforms $\mu_\pm^{(z)}(\eta)$.

3.3. **Properties of the singular values distribution.** Here we use estimates on moments of $\nu_{\alpha,z}$ to prove that all exponential moments of $\nu_{\alpha,z}$ are finite, that $\nu_{\alpha,z}$ depends on $z \in \mathbb{C}$ only through the absolute value $|z|$, and that the support of $\nu_{\alpha,z}$ is unbounded.

Lemma 3.2 (Singular values moments)**.** *For any $\lambda > 0$ and any $z \in \mathbb{C}$, $\alpha \in (0,1)$,*

$$\int_0^\infty \mathrm{e}^{\lambda t}\nu_{\alpha,z}(dt) < \infty. \tag{3.9}$$

Moreover $\nu_{\alpha,z} = \nu_{\alpha,w}$ for all $z, w \in \mathbb{C}$ such that $|z| = |w|$.

Proof. Define the moments

$$m_n = \int_0^\infty x^n \nu_{\alpha,z}(\mathrm{d}x).$$

For n even, we have

$$m_n = \int_{-\infty}^\infty x^n \bar{\mu}^{(z)}(\mathrm{d}x) = \frac{1}{2}\,\mathbb{E}\langle \delta_\varnothing, \hat{\mathcal{T}}_+(z)^n \delta_\varnothing\rangle + \frac{1}{2}\,\mathbb{E}\langle \delta_\varnothing, \hat{\mathcal{T}}_+(z)^n \delta_\varnothing\rangle. \tag{3.10}$$

Let Γ_k denote the set of all paths of length $n = 2k$ on the tree which start and end at the root. Let also $w_\pm(\gamma)$ denote the weight of one path $\gamma \in \Gamma_k$, so that

$$\mathbb{E}\langle \delta_\varnothing, \hat{\mathcal{T}}_\pm(z)^{2k} \delta_\varnothing\rangle = \sum_{\gamma\in\Gamma_k} \mathbb{E}[w_\pm(\gamma)].$$

The weight $w_\pm(\gamma)$ can be written as the product

$$\hat{\mathcal{T}}_\pm(\varnothing, \gamma_1)\hat{\mathcal{T}}_\pm(\gamma_1,\gamma_2)\cdots\hat{\mathcal{T}}_\pm(\gamma_{n-1},\varnothing), \tag{3.11}$$

where $\gamma_1,\dots,\gamma_{n-1}$ denote the ordered vertices visited by the path after leaving $\varnothing$ and before returning to $\varnothing$, and for simplicity we have omitted the explicit dependency on z from our notation.

We are going to prove that for any $k \in \mathbb{N}$,

$$m_{2k} \leqslant 2^{2k} C(z,k)\,, \qquad C(z,k) := \mathbb{E}\left[(|z|^2 + (1\vee\Psi))^k\right], \tag{3.12}$$

where $\Psi := \sum_{j=1}^\infty \omega_j^2$, and $\{\omega_j\}$ is defined in (2.15).

Once (3.12) is established we may write

$$\begin{aligned}\int_0^\infty \mathrm{e}^{\lambda t}\nu_{\alpha,z}(dt) &\leqslant 2\int_0^\infty \cosh(\lambda t)\nu_{\alpha,z}(dt)\\ &= 2\sum_{k=0}^\infty \frac{\lambda^{2k} m_{2k}}{(2k)!} \leqslant 2\sum_{k=0}^\infty \frac{(2\lambda)^{2k} C(z,k)}{(2k)!}.\end{aligned}$$

From the definition of $C(z,k)$ one has $C(z,k) \leqslant C(z,2k)$ and therefore

$$\sum_{k=0}^\infty \frac{(2\lambda)^{2k} C(z,k)}{(2k)!} \leqslant \mathbb{E}\left[\cosh(2\lambda(|z|^2 + (1\vee\Psi))\right] \leqslant \mathrm{e}^{2\lambda(1+|z|^2)}\mathbb{E}\left[\mathrm{e}^{2\lambda\sum_k \omega_k}\right],$$

where we use $\cosh(t) \leqslant \mathrm{e}^t$, $t \geqslant 0$, and $(1\vee\Psi) \leqslant 1 + \sum_k \omega_k$. On the other hand, from Campbell's formula we have, for any $\lambda \in \mathbb{R}$:

$$\mathbb{E}\left[\mathrm{e}^{2\lambda\sum_k \omega_k}\right] = \exp\left(\int_0^\infty \left(\mathrm{e}^{\frac{2\lambda t}{t+q}} - 1\right)\frac{\alpha\, dt}{t^{1+\alpha}}\right) < \infty,$$

which proves (3.9).

It remains to prove (3.12). To illustrate the computation, let us first consider the cases $n = 2, 4$ in detail. When $n = 2$,

$$\sum_{\gamma\in\Gamma_1} \mathbb{E}[w_+(\gamma)] = |z|^2 + \sum_{k=1}^\infty \mathbb{E}[\zeta_k^2]\,, \quad \sum_{\gamma\in\Gamma_1} \mathbb{E}[w_-(\gamma)] = |z|^2 + \sum_{k=1}^\infty \mathbb{E}[\omega_k^2],$$

where $\{\zeta_k\}, \{\omega_k\}$ are defined in (2.3) and (2.15) respectively. Since $0 \leqslant \zeta_k, \omega_k \leqslant 1$, $\sum_k \zeta_k = 1$ and $\mathbb{E}[\Psi] \leqslant \sum_k \mathbb{E}[\omega_k] = 1$, it follows that

$$\sum_{\gamma\in\Gamma_2} \mathbb{E}[w_\pm(\gamma)] \leqslant |z|^2 + \mathbb{E}[1\vee\Psi] = C(z,1).$$

If $n = 4$ we can partition the paths in Γ_2 into two families: the ones that reach distance 2 from the root, say Γ_2^2, and the ones that do not, say Γ_2^1. The first term gives

$$\sum_{\gamma\in\Gamma_2^2} \mathbb{E}[w_+(\gamma)] = |z|^2 \sum_{j=1}^\infty \mathbb{E}\left[\omega_j^2\right] + |z|^2 \sum_{k=1}^\infty \mathbb{E}\left[\zeta_k^2\right] + \sum_{k,j=1}^\infty \mathbb{E}\left[\zeta_k^2\omega_j^2\right],$$

where $\{\zeta_k\}, \{\omega_k\}$ are independent, and

$$\sum_{\gamma\in\Gamma_2^2} \mathbb{E}[w_-(\gamma)] = |z|^2 \sum_{j=1}^{\infty} \mathbb{E}\left[\zeta_j^2\right] + |z|^2 \sum_{k=1}^{\infty} \mathbb{E}\left[\omega_k^2\right] + \sum_{k,j=1}^{\infty} \mathbb{E}\left[\left(\tfrac{\xi_k}{\xi_k+S_k}\right)^2 \left(\tfrac{\xi_j^{(k)}}{\xi_k+S_k}\right)^2\right],$$

where $\{\xi_k\}, \{\xi_j^{(k)}\}$ are independent processes with laws PPP(α), and $S_k = \sum_{j=1}^{\infty} \xi_j^{(k)}$; see (2.25). In particular, using $\xi_j^{(k)}/(\xi_k + S_k) \leqslant \xi_j^{(k)}/S_k$, one has

$$\sum_{\gamma\in\Gamma_2^2} \mathbb{E}[w_\pm(\gamma)] \leqslant 1 + 2|z|^2 \leqslant C(z, 2).$$

The second family of paths contributes

$$\sum_{\gamma\in\Gamma_2^1} \mathbb{E}[w_+(\gamma)] = |z|^4 + 2|z|^2 \sum_{k}^{\infty} \mathbb{E}[\zeta_k^2] + \sum_{k,j=1}^{\infty} \mathbb{E}[\zeta_k^2\zeta_j^2],$$

$$\sum_{\gamma\in\Gamma_2^1} \mathbb{E}[w_-(\gamma)] = |z|^4 + 2|z|^2 \sum_{k}^{\infty} \mathbb{E}[\omega_k^2] + \sum_{k,j=1}^{\infty} \mathbb{E}[\omega_k^2\omega_j^2].$$

Using again $0 \leqslant \zeta_k, \omega_k \leqslant 1$,

$$\sum_{\gamma\in\Gamma_2^1} \mathbb{E}[w_\pm(\gamma)] \leqslant \mathbb{E}\left[(|z|^2 + (1 \vee \Psi))^2\right] = C(z, 2).$$

To estimate the general case $n = 2k$, let $\mathcal{D}_k$ denote the set of Dick paths of length $2k$, i.e. the set of all $\mathbb{Z}_+$ paths staring at 0 and ending at 0 after $2k$ steps. In particular, $|\mathcal{D}_k| = \frac{1}{k+1}\binom{2k}{k}$. The set Γ_k can be partitioned according to the distance from the root at each step. This is encoded by a single element of $\mathcal{D}_k$, with the natural correspondence: say Γ_k^η, where $\eta = (0, \eta_1, \ldots, \eta_{2k-1}, 0) \in \mathcal{D}_k$, is the set of all paths $\gamma \in \Gamma_k$ such that after j steps, γ has distance η_j from the root, for all $j = 1, \ldots, 2k-1$. Then

$$\sum_{\gamma\in\Gamma_k} \mathbb{E}[w_\pm(\gamma)] = \sum_{\eta\in\mathcal{D}_k} \sum_{\gamma\in\Gamma_k^\eta} \mathbb{E}[w_\pm(\gamma)].$$

For a fixed η, we can write the path $\gamma \in \Gamma_k^\eta$ as $\gamma = (\varnothing, \gamma_1, \ldots, \gamma_{2k-1}, \varnothing)$ where γ_i is the label of the vertex after i steps, so that γ_i is at distance η_i from the root. Now consider the set of indexes $i_1, \ldots, i_\ell$ such that $\eta_{i_m} = \max\{\eta_1, \ldots, \eta_{2k-1}\}$. Reasoning as above we can write

$$\sum_{\gamma_{i_1}\cdots,\gamma_{i_\ell}} \mathbb{E}[w_\pm(\gamma)] \leqslant \mathbb{E}[w_\pm(\gamma \smallsetminus (\gamma_{i_1}, \ldots, \gamma_{i_\ell}))] C(z, \ell),$$

where we use the notation $\gamma \smallsetminus (\gamma_{i_1}, \ldots, \gamma_{i_\ell})$ for the path γ once the vertices $\gamma_{i_1} \ldots, \gamma_{i_\ell}$ have been removed. Notice that this is now a path of length $2k - 2\ell$. Then we may proceed by recursion. Using $C(z, \ell)C(z, m) \leqslant C(z, \ell + m)$, which follows by Hölder inequality, we obtain

$$\sum_{\gamma\in\Gamma_k^\eta} \mathbb{E}[w_\pm(\gamma)] \leqslant C(z, k),$$

for every fixed $\eta \in \mathcal{D}_k$. In conclusion this proves

$$m_{2k} \leqslant |\mathcal{D}_k|\, C(z, k)\,. \tag{3.13}$$

This ends the proof of (3.9).

To show the last assertion in the lemma, notice that it suffices to show that $\bar{\mu}^{(z)} = \bar{\mu}^{(w)}$, for all $z, w \in \mathbb{C}$ with $|z| = |w|$. Clearly, odd moments of $\bar{\mu}^{(z)}$ are equal to zero. Since $\bar{\mu}^{(z)}$ has finite exponential moments, it is uniquely determined by the numbers m_n, n even, as in (3.10); see e.g. [17] for sufficient conditions in the classical moments problem. From

the expression of the path weights $w_\pm(\gamma)$ in (3.11), we see that each edge weight u of the tree appears only through the value $|u|^2$. Therefore, the moments m_n do not change if we replace z by w with $|z| = |w|$. □

Lemma 3.3 (Support)**.** *For any $z \in \mathbb{C}$, $\alpha \in (0,1)$, the support of $\nu_{\alpha,z}$ is unbounded.*

Proof. It suffices to show $m_{2k}^{1/k} \to \infty$, $k \to \infty$, where m_n is defined in (3.10). Notice that

$$m_{2k} \geqslant \frac{1}{2} \sum_{\gamma \in \Gamma_k} \mathbb{E}[w_-(\gamma)] \geqslant \frac{1}{2}\mathbb{E}[\Psi^k],$$

where the bound is obtained by restricting to paths that are always at distance at most 1 from the root and we write $\Psi = \sum_{j=1}^\infty \omega_j^2$ as before. Now, Ψ is an unbounded random variable, and $\mathbb{P}(\Psi > t) > 0$ for all $t > 0$. Indeed, a simple estimate on $\mathbb{P}(\Psi > t)$ can be obtained by observing: for any $n \geqslant 1$, $\Psi \geqslant n\omega_n^2$, and $\omega_n^2 \geqslant (1+q)^{-2}$ if $\xi_n = q\omega_n/(1-\omega_n) \geqslant 1$; from (2.1) one has that the event $\xi_n \geqslant 1$ has probability at least $\mathrm{e}^{-1}/n!$ so that for all $n \in \mathbb{N}$,

$$\mathbb{P}(\Psi \geqslant n(1-q)^{-2}) \geqslant \frac{1}{e\, n!}.$$

Therefore $m_{2k} \geqslant \frac{1}{2}\,\mathbb{P}(\Psi > t)\, t^k$, and $\liminf_{k\to\infty} m_{2k}^{1/k} \geqslant t$. Letting $t \to \infty$ proves the claim. □

3.4. **Recursive distributional equations.** Let us first recall some important formulas for resolvents on a tree. Let $\mathcal{A}$ denote a self-adjoint operator defined by a weighted tree via adjacency; see Section 2.3. In particular, we write $\mathcal{A}(\mathbf{u},\mathbf{v}) = \overline{\mathcal{A}(\mathbf{v},\mathbf{u})}$ for the complex valued weight of the directed edge $(\mathbf{u},\mathbf{v})$ if $\mathbf{v} = \mathbf{u}j$ or $\mathbf{u} = \mathbf{v}j$ for some $j \in \mathbb{N}$ and $\mathcal{A}(\mathbf{u},\mathbf{v}) = 0$ otherwise. For $\eta \in \mathbb{C}_+$, we let

$$R^\eta = \langle \delta_\varnothing, (\mathcal{A} - \eta)^{-1}\delta_\varnothing \rangle$$

denote the resolvent's diagonal entry at the root. Let also $u_j = \mathcal{A}(\varnothing, j)$ denote the weight from the root to the j-th child. If we remove the j-th edge from the root, i.e. the edge with weight u_j, we are left with two disconnected weighted trees, one rooted at $\varnothing$ and one rooted at j. Let $\bar{\mathcal{A}}_j$ and $\mathcal{A}_j$ denote the associated self-adjoint operators respectively. For any $\eta \in \mathbb{C}_+$, we let $\bar{R}_j^\eta$ and R_j^η denote the corresponding resolvent's diagonal entries at the respective roots. We refer to [5, Th. 4.1] for a proof of the following lemma.

Lemma 3.4 (Resolvent recursion)**.** *For any $\eta \in \mathbb{C}_+$ the resolvents entries $R^\eta, \bar{R}_j^\eta$ and R_j^η satisfy the following relations. For any $j \in \mathbb{N}$*

$$R^\eta = \frac{\bar{R}_j^\eta}{1 - |u_j|^2 \bar{R}_j^\eta R_j^\eta}. \tag{3.14}$$

Moreover,

$$R^\eta = -\left(\eta + \sum_{j=1}^\infty |u_j|^2 R_j^\eta\right)^{-1}. \tag{3.15}$$

We are interested in the law of the random variable

$$h^{(\pm)}(\eta, z) := \langle \delta_\varnothing, (\hat{\mathcal{T}}_\pm(z) - \eta)^{-1}\delta_\varnothing \rangle \tag{3.16}$$

It is convenient to introduce also the modified tree $\bar{\mathcal{T}}_\pm(z)$ defined as the connected component at the root obtained from the tree $\hat{\mathcal{T}}_\pm(z)$ after the removal of the edge with weight $-z$ emanating from the root. Note that this tree still depends on z since all remaining children of the root have an edge with weight $-z$ emanating from them, and so on; see the definition of $\hat{\mathcal{T}}_\pm(z)$ after (3.1). Set

$$\bar{h}^{(\pm)}(\eta, z) := \langle \delta_\varnothing, (\bar{\mathcal{T}}_\pm(z) - \eta)^{-1}\delta_\varnothing \rangle \tag{3.17}$$

Proposition 3.5 (Recursive distributional equations)**.** *For any $\eta \in \mathbb{C}_+$, $z \in \mathbb{C}$, the random variables $h^{(\pm)} := h^{(\pm)}(\eta, z)$ and $\bar{h}^{(\pm)} := \bar{h}^{(\pm)}(\eta, z)$ satisfy the following distributional equations:*

$$h^{(-)} \stackrel{d}{=} \frac{\bar{h}^{(-)}}{1 - |z|^2 \bar{h}^{(-)} \bar{h}^{(+)}}, \qquad h^{(+)} \stackrel{d}{=} \frac{\bar{h}^{(+)}}{1 - |z|^2 \bar{h}^{(-)} \bar{h}^{(+)}}, \tag{3.18}$$

where $\bar{h}^{(-)}, \bar{h}^{(+)}$ are taken independent. Moreover, if $\psi(x) := -(\eta + x)^{-1}$, then

$$\bar{h}^{(+)} \stackrel{d}{=} \psi \left(\sum_{k=1}^{\infty} \zeta_k^2 h_k^{(-)} \right), \tag{3.19}$$

and $(h^{(-)}, \bar{h}^{(-)})$ satisfy the system of distributional equations

$$\begin{aligned} \bar{h}^{(-)} &\stackrel{d}{=} \psi \left(\sum_{k=1}^{\infty} (\tfrac{\xi_k}{\xi_k + S_k})^2 \, \psi \left(|z|^2 \bar{h}^{(-)} + \sum_{j=1}^{\infty} (\tfrac{\xi_j^{(k)}}{\xi_k + S_k})^2 h_j^{(-)} \right) \right) \\ h^{(-)} &\stackrel{d}{=} \frac{\bar{h}^{(-)}}{1 - |z|^2 \bar{h}^{(-)} \psi \left(\sum_{k=1}^{\infty} \zeta_k^2 h_k^{(-)} \right)}, \end{aligned} \tag{3.20}$$

where $\{h_k^{(-)}\}, \bar{h}^{(-)}$ are independent such that $\{h_k^{(-)}\}$ are i.i.d. copies of $h^{(-)}$, and

$$\{\xi_k, k \in \mathbb{N}\}, \ \{\xi_j^{(k)}, j \in \mathbb{N}\}_{k \in \mathbb{N}}, \ \{\zeta_k, k \in \mathbb{N}\}$$

are independent processes with laws PPP(α), PPP(α), and PD(α), respectively.

Proof. In the setting of Lemma 3.4 above, with $\mathcal{A}$ given by $\hat{\mathcal{T}}_{\pm}(z)$, we may identify R^η with $h^{(\pm)}(\eta, z)$ and $\bar{R}_j^\eta$ with $\bar{h}^{(\pm)}(\eta, z)$, if j represents the child associated with the weight $-z$ emanating from the root. Notice also that in this case one has that R_j^η is an independent copy of the random variable $\bar{h}^{(\mp)}(\eta, z)$. Thus from (3.14) one finds the distributional identities (3.18).

Next, observe that from (3.15) one has

$$\bar{h}^{(+)}(\eta, z) = - \left(\eta + \sum_{k=1}^{\infty} \zeta_k^2 h_k^{(-)}(\eta, z) \right)^{-1} \tag{3.21}$$

where $\{\zeta_k\}$ has law PD(α), while $\{h_k^{(-)}(\eta, z), k \in \mathbb{N}\}$ is a set of i.i.d. copies of $h^{(-)}(\eta, z)$, independent of $\{\zeta_k\}$. This proves (3.19).

Finally, again by (3.15) one has

$$\bar{h}^{(-)}(\eta, z) = - \left(\eta + \sum_{k=1}^{\infty} (\tfrac{\xi_k}{\xi_k + S_k})^2 h_k^{(+)}(\xi_k, \eta, z) \right)^{-1} \tag{3.22}$$

where $\{\xi_k\}$ has law PPP(α), $S_k = \sum_{j=1}^{\infty} \xi_j^{(k)}$ where $\{\xi_j^{(k)}, \, j \in \mathbb{N}\}_{k \in \mathbb{N}}$ are all i.i.d. processes with law PPP(α), and $h_k^{(+)}(\xi_k, \eta, z)$ is the resolvent entry associated to the sub-tree rooted at vertex k obtained from the tree $\bar{\mathcal{T}}_-(z)$ by deleting the root $\varnothing$ and all its descendants except for k. We remark that in (3.22), given $\{\xi_k\}$, the variables $\{h_k^{(+)}(\xi_k, \eta, z), \, k \in \mathbb{N}\}$ are independent but their law depends on $\{\xi_k\}$ in a non trivial way. We can however apply one more time the tree recursion (3.15) and obtain, for each $k \in \mathbb{N}$:

$$h_k^{(+)}(\xi_k, \eta, z) = - \left(\eta + |z|^2 \bar{h}_0^{(-)}(\eta, z) + \sum_{j=1}^{\infty} (\tfrac{\xi_j^{(k)}}{\xi_k + S_k})^2 h_j^{(-)}(\eta, z) \right)^{-1}, \tag{3.23}$$

where $\bar{h}_0^{(-)}(\eta, z)$ and $h_j^{(-)}(\eta, z)$, $j \in \mathbb{N}$, are independent copies of $\bar{h}^{(-)}(\eta, z)$ and $h^{(-)}(\eta, z)$ respectively. Using (3.18)–(3.23), one obtains the system of distributional equations (3.20). □

In principle, the system (3.20) should determine the law of $h^{(-)}$ and, via (3.18)-(3.19), also the law of $h^{(+)}$. However, we do not have a proof of the uniqueness of the solutions to these distributional equations.

We remark that in the special case $z = 0$, one has $\bar{h}^{(-)}(\eta, 0) = h^{(-)}(\eta, 0)$ and the above equations reduce to the following recursive distributional equation:

$$h^{(-)} \stackrel{d}{=} \psi\left(\sum_{k=1}^{\infty}(\tfrac{\xi_k}{\xi_k+S_k})^2\,\psi\left(\sum_{j=1}^{\infty}(\tfrac{\xi_j^{(k)}}{\xi_k+S_k})^2 h_k^{(-)}\right)\right), \tag{3.24}$$

where $\{h_k^{(-)}\}$ and $\{\xi_k\}, \{\xi_j^{(k)}\}$ are as above. Once we have the law of $h^{(-)}(\eta, 0)$, then $h^{(+)}(\eta, 0)$ is determined by the equation (3.21) which, in the case $z = 0$, becomes

$$h^{(+)} \stackrel{d}{=} \psi\left(\sum_{k=1}^{\infty}\zeta_k^2 h_k^{(-)}\right). \tag{3.25}$$

4. Convergence of eigenvalues

Here we prove Theorem 1.2. Once the convergence of singular values in Theorem 1.1 has been obtained, we may follow the well known Hermitization strategy, see e.g. [19] and [10], which allows one to prove $\mu_M \rightsquigarrow \mu_\alpha$ as $n \to \infty$, for some probability measure μ_α by showing the uniform integrability of the function $\mathbb{R} \ni x \mapsto |\log(x)|$ with respect to the family of measures ν_{M-z}. More precisely, we shall establish the following lemma.

Lemma 4.1. *For any $a > 0$, $z \in \mathbb{C}$, there exists $b > 0$ such that for all $n \in \mathbb{N}$:*

$$\mathbb{P}\left(\mathrm{supp}(\nu_{M-z}) \not\subset [n^{-b}, n^b]\right) \leqslant n^{-a}. \tag{4.1}$$

Moreover, for any $\varepsilon > 0$, for a.a. $z \in \mathbb{C}$

$$\lim_{\delta\to 0}\limsup_n \mathbb{P}\left(\int_{K_\delta^c} |\log(x)|\nu_{M-z}(\mathrm{d}x) > \varepsilon\right) = 0, \tag{4.2}$$

where $K_\delta = [\delta, \delta^{-1}]$.

From (4.2) and Theorem 1.1, as shown in [10, Lem. 4.3], one has that $\mu_M \rightsquigarrow \mu_\alpha$ in probability, where μ_α is the probability measure on $\mathbb{C}$ that satisfies

$$U_{\mu_\alpha}(z) = -\int_0^\infty \log(x)\nu_{\alpha,z}(\mathrm{d}x), \tag{4.3}$$

for almost all $z \in \mathbb{C}$. Here, for any probability measure μ on $\mathbb{C}$, the function

$$U_\mu(z) = -\int_{\mathbb{C}} \log|w - z|\mu(dw)\,,$$

denotes the logarithmic potential.

To improve the above convergence to the desired statement that $\mu_M \rightsquigarrow \mu_\alpha$ a.s. (almost surely), we argue as follows. As in [10, Lem. 4.3], it suffices to show that $U_{\mu_M}(z)$ converges almost surely. In particular, since the above facts show that $U_{\mu_M}(z) \to U_{\mu_\alpha}(z)$ in probability, it is now sufficient to prove that for a.a. $z \in \mathbb{C}$ there exists a deterministic sequence L_n such that, almost surely,

$$\lim_{n\to\infty}(U_{\mu_M}(z) - L_n) = 0.$$

By (4.1), and the Borel-Cantelli lemma, there exists $b > 0$ such that, almost surely

$$\mathrm{supp}(\nu_{M-z}) \subset [n^{-b}, n^b],$$

if n is large enough. If $f_n(x) := \mathbf{1}_{x\in[n^{-b},n^b]}\log(x)$, then a.s. for every n large enough one has the identity

$$U_{\mu_M}(z) = -\frac{1}{n}\sum_{i=1}^{n}\log|\lambda_i - z| = -\int_0^\infty \log(s)\nu_{M-z}(\mathrm{d}s) = -\int_0^\infty f_n(s)\nu_{M-z}(\mathrm{d}s).$$

Since the total variation norm of f_n is $O(\log n)$ the concentration inequality of [6, Lem. C.2] shows that

$$\mathbb{P}\left(|U_{\mu_M}(z) - \mathbb{E}U_{\mu_M}(z)| > \varepsilon\right) \leqslant 2\,\mathrm{e}^{-c\,\varepsilon\,n\,(\log n)^{-2}},$$

for some constant $c > 0$. Therefore, the conclusion follows from the Borel-Cantelli lemma letting $L_n = \mathbb{E}[U_{\mu_M}(z)]$.

The above argument completes the proof of the almost sure converge $\mu_M \rightsquigarrow \mu_\alpha$ as $n \to \infty$, where μ_α is a probability measure on the unit disc of radius 1 that depends only on $\alpha \in (0,1)$. From the identity (4.3), and the fact that $\nu_{\alpha,z}$ depends radially on $z \in \mathbb{C}$ one infers that μ_α is a radially symmetric measure. Moreover, the finiteness of $U_{\mu_\alpha}(0)$ implies that μ_α is not a Dirac mass at the origin. Similarly, one can observe that μ_α is not concentrated on the unit circle $|z| = 1$. Indeed, by an inequality of Weyl, see e.g. [6, Lem. B.5], the eigenvalues of M satisfy

$$\int_{\mathbb{C}} |\lambda|^2 \mu_M(\mathrm{d}\lambda) \leqslant \int_0^\infty x^2 \nu_M(\mathrm{d}x).$$

Passing to the limit $n \to \infty$, one has

$$\int_{\mathbb{C}} |z|^2 \mu_\alpha(dz) \leqslant \int_0^\infty x^2 \nu_{\alpha,0}(\mathrm{d}x) = m_2 < 1,$$

where the strict bound $m_2 < 1$ follows by direct computations as in Lemma 3.2. This ends the proof of Theorem 1.2 assuming the validity of Lemma 4.1. The rest of this section is concerned with the proof of Lemma 4.1.

4.1. **Extreme singular values.** Here we prove the first part of Lemma 4.1. Let $s_{1,z} \geqslant \cdots \geqslant s_{n,z}$ denote the singular values of $M - z$. We start with a simple upper bound on $s_{1,z}$. Notice that

$$\begin{aligned}\int_0^\infty x^2 \nu_{M-z}(\mathrm{d}x) &= \frac{1}{n}\sum_{i,j=1}^{n} |M_{i,j} - z\delta_{i,j}|^2 \\ &\leqslant \frac{2}{n}\sum_{i,j=1}^{n} M_{i,j}^2 + 2|z|^2 \leqslant 2(1+|z|^2), \end{aligned} \tag{4.4}$$

where we use $\sum_j M_{i,j}^2 \leqslant \sum_j M_{i,j} = 1$. This implies the deterministic bound

$$s_{1,z}^2 \leqslant 2n(1+|z|^2).$$

Thus (4.1) follows from the lower bound on $s_{n,z}$ given in Theorem 4.2 below. Notice that $M - z$ is not invertible at $z = 1$, i.e. $s_{n,1} = 0$. Because of the different scaling, in [7, Th. 1.4] it was sufficient to prove a lower bound on $s_{n,z}$ for all $z = O(n^{-1/2})$. In our setting instead, we need to establish a similar bound for all $z \neq 1$.

Theorem 4.2. *Assume (H2). Let $s_{n,z}$ denote the smallest singular value of $M - z$. For any $\delta \in (0,1)$, and $a > 0$, there exists $b = b(a,\delta) > 0$ such that if $z \in \mathbb{C}$ satisfies $|z-1| \geqslant \delta$, and $|z| \leqslant \delta^{-1}$, then for $n \gg 1$ one has*

$$\mathbb{P}(s_{n,z} \leqslant n^{-b}) \leqslant n^{-a}. \tag{4.5}$$

Proof. Fix $a > 0$ and $z \in \mathbb{C}$ with $|1 - z| > \delta$, $|z| < \delta^{-1}$. We write $M = DX$, where $D = \mathrm{diag}(\rho_1^{-1}, ..., \rho_n^{-1})$. Thus, $M - z = DY$ where $Y = X - zD^{-1}$, and

$$s_{n,z} \geqslant s_n(Y)\Big(\max_{i=1,...,n}\rho_i\Big)^{-1}, \tag{4.6}$$

with $s_n(Y)$ denoting the smallest singular value of Y. Since $\rho_i \leqslant n \max_j X_{i,j}$ one has, for any $\beta > 0$

$$\mathbb{P}\Big(\max_{i=1,\dots,n} \rho_i > n^{\beta} \Big) \leqslant n^2 \mathbb{P}(\mathbf{x} > n^{\beta-1}) \leqslant n^{2-\alpha(\beta-2)},$$

for all n large enough. Taking β sufficiently large ensures that $n^{2-\alpha(\beta-2)} \leqslant n^{-a}$. From (4.6) then we see that it suffices to prove that there exists $b > 0$ such that

$$\mathbb{P}\left(s_n(Y) \leqslant n^{-b} \right) \leqslant n^{-a} \tag{4.7}$$

A repetition of the argument in [7, Th. 1.4] now shows that for all $u > 0$:

$$\mathbb{P}\left(s_n(Y) \leqslant u \right) \leqslant 3BK_n\, u\,, \tag{4.8}$$

where B is the uniform bound on the probability density of the random variable $\mathbf{x}$, which is available thanks to the assumption $(H2)$, and K_n is the inverse of the smallest singular value of the $n \times n$ matrix

$$A_z := \mathbf{1}_n - z \begin{pmatrix} 1 & 0 & \cdot\cdot & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdot\cdot & 0 \end{pmatrix}.$$

Direct calculations, see e.g. [7, Lem. C.3], show that K_n satisfies

$$K_n^2 = \frac{1 + (n-1)|z|^2 + |1-z|^2 + \sqrt{(1+(n-1)|z|^2 + |1-z|^2)^2 - 4|1-z|^2}}{2|1-z|^2}. \tag{4.9}$$

Equation (4.9) can be easily estimated to obtain e.g.

$$\begin{aligned} K_n^2 &\leqslant \frac{1 + (n-1)|z|^2 + |1-z|^2}{|1-z|^2} \\ &\leqslant 1 + \frac{1 + |z|^2(n-1)}{\delta^2} \leqslant 1 + \delta^{-2} + \delta^{-4} n. \end{aligned} \tag{4.10}$$

Using (4.10), (4.7) follows by taking $u = n^{-b}$ for $b = b(a, \delta)$ large enough in (4.8). □

4.2. **Moderately small singular values.** Here we prove (4.2), which will conclude the proof of Lemma 4.1. In view of the bound (4.4), and using $|\log(x)| \leqslant x^{-p}$ for all sufficiently small $x > 0$, if $p > 0$, we see that (4.2) follows if we prove

$$\limsup_n \mathbb{E}\left(\mathbf{1}_{G_n} \int_0^{\infty} x^{-p} \nu_{M-z}(\mathrm{d}x) \right) < \infty, \tag{4.11}$$

for some $p > 0$ and some sequence of events G_n such that $\mathbb{P}(G_n) \to 1$, $n \to \infty$. To prove (4.11), we shall follow very closely the strategy introduced in [6, Sec. 3]; see also [10, Sec. 6].

The following statement can be established with a straightforward adaptation of [6, Prop. 3.7]. Set $a_n = (cn)^{1/\alpha}$ as in Lemma 2.1.

Lemma 4.3 (Distance to sub-space)**.** *Assume (H1) and (H2) and take $0 < \gamma < \alpha/4$. Let R_1 be the first row of the matrix $a_n(M - z)$. There exists a constant $C > 0$ and an event E such that for any d-dimensional subspace W of $\mathbb{C}^n$ with $d \leqslant n - n^{1-\gamma}$, one has*

$$\mathbb{E}\left[\mathrm{dist}^{-2}(R_1, W) \mathbf{1}_E \right] \leqslant C(n-d)^{-2/\alpha} \quad \textit{and} \quad \mathbb{P}(E^c) \leqslant C n^{-(1-2\gamma)/\alpha}\,.$$

Next, we prove (4.11). Fix $i \geqslant 2n^{1-\gamma}$, and let A be the matrix of the first $n - i/2$ rows of $a_n(M - z)$ and let $\vartheta_1 \geqslant \cdots \geqslant \vartheta_{n-i/2}$ denote its singular values. By the negative second moment identity of Tao and Vu [20, Lem. A4],

$$\vartheta_1^{-2} + \cdots + \vartheta_{n-i/2}^{-2} = \delta_1^{-2} + \cdots \delta_{n-i/2}^{-2},$$

where δ_j is defined as the euclidean distance $\mathrm{dist}(R_j, R_{-j})$, where R_j is the j-th row of A and R_{-j} is the span of all other rows of A. On the other hand by Cauchy interlacing lemma, $s_{n-i,z} \geqslant \vartheta_{n-i}/a_n$, and therefore $s^{-2}_{n-i,z} \leqslant \frac{2a_n^2}{i} \sum_{j=n-i}^{n-i/2} \vartheta_j^{-2}$. This implies

$$i\, s^{-2}_{n-i,z} \leqslant 2a_n^2 \sum_{j=1}^{n-i/2} \delta_j^{-2}. \tag{4.12}$$

Reasoning as in [6, Prop. 3.3] one has that the event F that $\delta_j \geqslant n^{(1-2\gamma)/\alpha}$ for all $j = 1, \dots, n - i/2$ has probability at least $1 - \mathrm{e}^{-n^\delta}$ for some $\delta > 0$.

Taking expectation in (4.12), we get

$$\mathbb{E}\left[i s^{-2}_{n-i,z} \mathbf{1}_F\right] \leqslant 2a_n^2 n \mathbb{E}\left[\delta_1^{-2}\, \mathbf{1}_F\right], \tag{4.13}$$

Moreover, if E denotes the event from Lemma 4.3, then since R_{-1} has dimension $d < n - i/2 \leqslant n - n^{1-\gamma}$, we see that

$$\mathbb{E}[\delta_1^{-2}\, \mathbf{1}_E] \leqslant C\, i^{-2/\alpha}. \tag{4.14}$$

From (4.14) it follows that

$$\begin{aligned}\mathbb{E}\left[\delta_1^{-2}\, \mathbf{1}_F\right] &\leqslant C\, i^{-2/\alpha} + \mathbb{E}\left[\delta_1^{-2}\, \mathbf{1}_{F \cap E^c}\right] \\ &\leqslant C\, i^{-2/\alpha} + n^{-2(1-2\gamma)/\alpha} \mathbb{P}(E^c) \leqslant C\, i^{-2/\alpha} + n^{-3(1-2\gamma)/\alpha},\end{aligned}$$

where we use the bound $\delta_1 \geqslant n^{(1-2\gamma)/\alpha}$ on F and the bound on $\mathbb{P}(E^c)$ from Lemma 4.3. If $\gamma < 1/6$, then $3(1-2\gamma)/\alpha > 2/\alpha$ and therefore $n^{-3(1-2\gamma)/\alpha} \leqslant i^{-2/\alpha}$. From (4.13) and recalling that $a_n = (cn)^{1/\alpha}$ one then obtains

$$\mathbb{E}\left[s^{-2}_{n-i,z} \mathbf{1}_F\right] \leqslant C' \left(\frac{n}{i}\right)^{(1+2/\alpha)}, \tag{4.15}$$

for some new constant $C' > 0$.

From (4.15) one can prove (4.11) as follows. Define the event $G_n := F \cap \{s_{n,z} \geqslant n^{-b}\}$ for some $b > 0$. From the above facts and from Theorem 4.2, we can choose b so that $\mathbb{P}(G_n) \to 1$. One has, for $0 < p \leqslant 2$

$$\begin{aligned}\mathbb{E}\left(\mathbf{1}_{G_n} \int_0^\infty x^{-p} \nu_{M-z}(\mathrm{d}x)\right) &= \frac{1}{n} \sum_{i=0}^{\lfloor 2n^{1-\gamma} \rfloor} \mathbb{E}[s^{-p}_{n-i,z} \mathbf{1}_{G_n}] + \frac{1}{n} \sum_{i=\lfloor 2n^{1-\gamma} \rfloor + 1}^{n-1} \mathbb{E}[s^{-p}_{n-i,z} \mathbf{1}_{G_n}] \\ &\leqslant 2n^{bp} n^{-\gamma} + \frac{1}{n} \sum_{i=\lfloor 2n^{1-\gamma} \rfloor + 1}^{n-1} \mathbb{E}[s^{-2}_{n-i,z} \mathbf{1}_F]^{p/2} \leqslant 2n^{bp} n^{-\gamma} + \frac{C}{n} \sum_{i=1}^{n} \left(\frac{n}{i}\right)^{\frac{p}{2}(1+2/\alpha)},\end{aligned}$$

for some new constant $C > 0$. If $\gamma \in (0, 1/6)$ is fixed, and $b > 0$ is given, then we choose p such that $p < \gamma/b$ and $p < 2\alpha(2+\alpha)$ so that the above expression is uniformly bounded.

References

[1] Radosław Adamczak, Djalil Chafaï, and Paweł Wolff. Circular law for random matrices with exchangeable entries. *Random Structures Algorithms*, 48(3):454–479, 2016. 2

[2] David Aldous. Asymptotics in the random assignment problem. *Probab. Theory Related Fields*, 93(4):507–534, 1992. 5

[3] David Aldous and J. Michael Steele. The objective method: probabilistic combinatorial optimization and local weak convergence. In *Probability on discrete structures*, volume 110 of *Encyclopaedia Math. Sci.*, pages 1–72. Springer, Berlin, 2004. 2, 5

[4] Charles Bordenave, Pietro Caputo, and Djalil Chafaï. Spectrum of large random reversible Markov chains: two examples. *ALEA Lat. Am. J. Probab. Math. Stat.*, 7:41–64, 2010. 2

[5] Charles Bordenave, Pietro Caputo, and Djalil Chafaï. Spectrum of large random reversible Markov chains: Heavy-tailed weigths on the complete graph. *Annals of Probability*, 39(4):1544–1590, 2011. 2, 5, 10, 12, 13, 16, 19

[6] Charles Bordenave, Pietro Caputo, and Djalil Chafaï. Spectrum of non-Hermitian heavy tailed random matrices. *Comm. Math. Phys.*, 307(2):513–560, 2011. 3, 5, 7, 13, 14, 16, 22, 23, 24
[7] Charles Bordenave, Pietro Caputo, and Djalil Chafaï. Circular law theorem for random Markov matrices. *Probab. Theory Related Fields*, 152(3-4):751–779, 2012. 2, 3, 22, 23
[8] Charles Bordenave, Pietro Caputo, and Justin Salez. Random walk on sparse random digraph. preprint arXiv:1508.06600, 2015. 3
[9] Charles Bordenave, Pietro Caputo, and Justin Salez. Cutoff at the "entropic time" for sparse markov chains. in preparation, 2016. 3
[10] Charles Bordenave and Djalil Chafaï. Around the circular law. *Probab. Surveys*, 9(0):1–89, 2012. 3, 21, 23
[11] Charles Bordenave and Marc Lelarge. Resolvent of large random graphs. *Random Structures Algorithms*, 37(3):332–352, 2010. 5
[12] Sourav Chatterjee, Persi Diaconis, and Allan Sly. Properties of random doubly stochastic matrices. *Annales de l'Institut Henri Poincaré. To appear*, 2015. preprint arXiv:1010.6136. 2
[13] Vyacheslav L. Girko. The circular law. *Teor. Veroyatnost. i Primenen.*, 29(4):669–679, 1984. 3
[14] Hoi H. Nguyen. Random doubly stochastic matrices: the circular law. *Ann. Probab.*, 42(3):1161–1196, 2014. 2
[15] Hoi H. Nguyen and Van H. Vu. Circular law for random discrete matrices of given row sum. *J. Comb.*, 4(1):1–30, 2013. 2
[16] Jim Pitman and Marc Yor. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.*, 25(2):855–900, 1997. 2, 5
[17] Michael Reed and Barry Simon. *Methods of modern mathematical physics. Vol. I and II.* Academic Press Inc., New York, second edition, 1980. 10, 16, 18
[18] Sidney I. Resnick. *Heavy-tail phenomena*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2007. Probabilistic and statistical modeling. 5, 8
[19] Terence Tao and Van Vu. Random matrices: the circular law. *Commun. Contemp. Math.*, 10(2):261–307, 2008. 3, 21
[20] Terence Tao and Vu Vu. Random matrices: universality of ESDs and the circular law. *Ann. Probab.*, 38(5):2023–2065, 2010. With an appendix by Manjunath Krishnapur. 23

(Charles Bordenave) IMT UMR5219 CNRS Université de Toulouse, France
E-mail address: charles.bordenave(at)math.univ-toulouse.fr
URL: http://www.math.univ-toulouse.fr/~bordenave/

(Pietro Caputo) Dipartimento di Matematica, Università Roma Tre, Italy
E-mail address: caputo(at)mat.uniroma3.it
URL: http://www.mat.uniroma3.it/users/caputo/

(Djalil Chafaï) CEREMADE UMR7534 CNRS Université Paris-Dauphine PSL IUF, France
E-mail address: djalil(at)chafai.net
URL: http://djalil.chafai.net/

(Daniele Piras) Dipartimento di Matematica, Università Roma Tre, Italy
E-mail address: piras(at)mat.uniroma3.it