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

BY CHARLES BORDENAVE, PIETRO CAPUTO[1] AND DJALIL CHAFAÏ

*Université de Toulouse III, Università Roma Tre
and Université Paris-Est Marne-la-Vallée*

We consider the random reversible Markov kernel K obtained by assigning i.i.d. nonnegative weights to the edges of the complete graph over n vertices and normalizing by the corresponding row sum. The weights are assumed to be in the domain of attraction of an α-stable law, $\alpha \in (0, 2)$. When $1 \leq \alpha < 2$, we show that for a suitable regularly varying sequence κ_n of index $1 - 1/\alpha$, the limiting spectral distribution μ_α of $\kappa_n K$ coincides with the one of the random symmetric matrix of the un-normalized weights (Lévy matrix with i.i.d. entries). In contrast, when $0 < \alpha < 1$, we show that the empirical spectral distribution of K converges without rescaling to a nontrivial law $\tilde{\mu}_\alpha$ supported on $[-1, 1]$, whose moments are the return probabilities of the random walk on the Poisson weighted infinite tree (PWIT) introduced by Aldous. The limiting spectral distributions are given by the expected value of the random spectral measure at the root of suitable self-adjoint operators defined on the PWIT. This characterization is used together with recursive relations on the tree to derive some properties of μ_α and $\tilde{\mu}_\alpha$. We also study the limiting behavior of the invariant probability measure of K.

1. Introduction. Let $G_n = (V_n, E_n)$ denote the complete graph with vertex set $V_n = \{1, \ldots, n\}$, and edge set $E_n = \{\{i, j\}, 1 \leq i, j \leq n\}$, including loops $\{i, i\}$, $1 \leq i \leq n$. Assign a nonnegative random weight (or conductance) $U_{i,j} = U_{j,i}$ to each edge $\{i, j\} \in E_n$, and assume that the symmetric weights $\mathbf{U} = \{U_{i,j}; \{i, j\} \in E_n\}$ are i.i.d. with common law $\mathcal{L}$ independent of n. This defines a random network, or weighted graph, denoted $(G_n, \mathbf{U})$. Next, consider the random walk on $(G_n, \mathbf{U})$ defined by the transition probabilities

$$(1.1) \qquad K_{i,j} := \frac{U_{i,j}}{\rho_i} \qquad \text{with } \rho_i := \sum_{j=1}^{n} U_{i,j}.$$

1544

The Markov kernel K is *reversible* with respect to the measure $\rho = \sum_{i \in V_n} \rho_i \delta_i$ in that

$$\rho_i K_{i,j} = \rho_j K_{j,i}$$

for all $i, j \in V_n$. Note that we have not assumed that $\mathcal{L}$ has no atom at 0. If $\rho_i = 0$ for some i, then for that index i we set $K_{i,j} = \delta_{i,j}$, $1 \leq j \leq n$. However, as soon as $\mathcal{L}$ is not concentrated at 0 then almost surely, for all n sufficiently large, $\rho_i > 0$ for all $1 \leq i \leq n$, K is irreducible and aperiodic and ρ is its unique invariant measure, up to normalization (see, e.g., [11]).

For any square $n \times n$ matrix M with eigenvalues $\lambda_1(M), \ldots, \lambda_n(M)$, the *Empirical Spectral Distribution* (ESD) is the discrete probability measure with at most n atoms defined by

$$\mu_M := \frac{1}{n} \sum_{j=1}^{n} \delta_{\lambda_j(M)}.$$

All matrices M to be considered in this work have real spectrum, and the eigenvalues will be labeled in such a way that $\lambda_n(M) \leq \cdots \leq \lambda_1(M)$.

Note that K defines a square $n \times n$ random Markov matrix whose entries are not independent due the normalizing sums ρ_i. By reversibility, K is self-adjoint in $L^2(\rho)$ and its spectrum $\sigma(K)$ is real. Moreover, $\sigma(K) \subset [-1, +1]$, and $1 \in \sigma(K)$. Since K is Markov, its ESD μ_K carries further probabilistic content. Namely, for any $\ell \in \mathbb{N}$, if $p_\ell(i)$ denotes the probability that the random walk on $(G_n, \mathbf{U})$ started at i returns to i after ℓ steps, then the ℓth moment of μ_K satisfies

$$(1.2) \qquad \int_{-1}^{+1} x^\ell \mu_K(dx) = \frac{1}{n} \operatorname{tr}(K^\ell) = \frac{1}{n} \sum_{i \in V} p_\ell(i).$$

Convergence of the ESD. The asymptotic behavior of μ_K as $n \to \infty$ depends strongly on the tail of $\mathcal{L}$ at infinity. When $\mathcal{L}$ has finite mean $\int_0^\infty x\mathcal{L}(dx) = m$ we set $m = 1$. This is no loss of generality since K is invariant under the dilation $t \to tU_{i,j}$. If $\mathcal{L}$ has a finite second moment we write $\sigma^2 = \int_0^\infty (x-1)^2 \mathcal{L}(dx)$ for the variance.

The following result, from [11], states that if $0 < \sigma^2 < \infty$, then the bulk of the spectrum of $\sqrt{n}K$ behaves, when $n \to \infty$, as if we had truly i.i.d. entries (Wigner matrix). Without loss of generality, we assume that the weights $\mathbf{U}$ come from the truncation of a unique infinite table $(U_{i,j})_{i,j \geq 1}$ of i.i.d. random variables of law $\mathcal{L}$. This gives a meaning to the almost sure (a.s.) convergence of $\mu_{\sqrt{n}K}$. The symbol $\xrightarrow{w}$ denotes weak convergence of measures with respect to continuous bounded functions. Note that $\lambda_1(\sqrt{n}K) = \sqrt{n} \to \infty$.

THEOREM 1.1 (Wigner-like behavior). *If $\mathcal{L}$ has variance $0 < \sigma^2 < \infty$, then a.s.*

$$(1.3) \qquad \mu_{\sqrt{n}K} := \frac{1}{n} \sum_{k=1}^{n} \delta_{\sqrt{n}\lambda_k(K)} \xrightarrow[n \to \infty]{w} \mathcal{W}_{2\sigma},$$

where $\mathcal{W}_{2\sigma}$ is the Wigner semi-circle law on $[-2\sigma, +2\sigma]$. Moreover, if $\mathcal{L}$ has finite fourth moment, then $\lambda_2(\sqrt{n}K)$ and $\lambda_n(\sqrt{n}K)$ converge a.s. to the edge of the limiting support $[-2\sigma, +2\sigma]$.

This Wigner-like scenario can be dramatically altered if we allow $\mathcal{L}$ to have a heavy tail at infinity. For any $\alpha \in (0, \infty)$, we say that $\mathcal{L}$ belongs to the class $\mathbb{H}_\alpha$ if $\mathcal{L}$ is supported in $[0, \infty)$ and has a regularly varying tail of index α, that is, for all $t > 0$,

$$(1.4) \qquad G(t) := \mathcal{L}((t, \infty)) = L(t)t^{-\alpha},$$

where L is a function with slow variation at ∞; that is, for any $x > 0$,

$$\lim_{t \to \infty} \frac{L(xt)}{L(t)} = 1.$$

Set $a_n = \inf\{a > 0 : nG(a) \le 1\}$. Then $nG(a_n) = nL(a_n)a_n^{-\alpha} \to 1$ as $n \to \infty$, and

$$(1.5) \qquad nG(a_n t) \to t^{-\alpha} \qquad \text{as } n \to \infty \text{ for all } t > 0.$$

It is well known that a_n has regular variation at ∞ with index $1/\alpha$, that is,

$$a_n = n^{1/\alpha}\ell(n)$$

for some function ℓ with slow variation at ∞ (see, e.g., Resnick [24], Section 2.2.1). As an example, if V is uniformly distributed on the interval $[0, 1]$, then for every $\alpha \in (0, \infty)$, the law of $V^{-1/\alpha}$, supported in $[1, \infty)$, belongs to $\mathbb{H}_\alpha$. In this case, $L(t) = 1$ for $t \ge 1$, and $a_n = n^{1/\alpha}$.

To understand the limiting behavior of the spectrum of K in the heavy-tailed case it is important to consider first the symmetric i.i.d. matrix corresponding to the un-normalized weights $U_{i,j}$. More generally, we introduce the random $n \times n$ symmetric matrix X defined by

$$(1.6) \qquad X = (X_{i,j})_{1 \le i, j \le n},$$

where $(X_{i,j})_{1 \le i \le j \le n}$ are i.i.d. such that $U_{i,j} := |X_{i,j}|$ has law in $\mathbb{H}_\alpha$ with $\alpha \in (0, 2)$, and

$$(1.7) \qquad \theta = \lim_{t \to \infty} \frac{\mathbb{P}(X_{i,j} > t)}{\mathbb{P}(|X_{i,j}| > t)} \in [0, 1].$$

It is well known that, for $\alpha \in (0, 2)$, a random variable Y is in the domain of attraction of an α-stable law iff the law of $|Y|$ is in $\mathbb{H}_\alpha$ and the limit (1.7) exists (cf. [17], Theorem IX.8.1a). It will be useful to view the entries $X_{i,j}$ in (1.6) as the marks across edge $\{i, j\} \in E_n$ of a random network $(G_n, \mathbf{X})$, just as the marks $U_{i,j}$ defined the network $(G_n, \mathbf{U})$ introduced above.

Remarkable works have been devoted recently to the asymptotic behavior of the ESD of matrices X defined by (1.6), sometimes called Lévy matrices. The analysis

of the *Limiting Spectral Distribution* (LSD) for $\alpha \in (0, 2)$ is considerably harder than the finite second moment case (Wigner matrices), and the LSD is nonexplicit. Theorem 1.2 below has been investigated by the physicists Bouchaud and Cizeau [15] and rigorously proved by Ben Arous and Guionnet [7], and Belinschi, Dembo and Guionnet [5] (see also Zakharevich [28] for related results).

THEOREM 1.2 [Symmetric i.i.d. matrix, $\alpha \in (0, 2)$]. *For every* $\alpha \in (0, 2)$, *there exists a symmetric probability distribution* μ_α *on* $\mathbb{R}$ *depending only on* α *such that* [*with the notation of* (1.5) *and* (1.6)] *a.s.*

$$\mu_{a_n^{-1}X} := \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(a_n^{-1}X)} \xrightarrow[n\to\infty]{w} \mu_\alpha.$$

In Section 3.2, we give a new independent proof of Theorem 1.2. The key idea of our proof is to exhibit a limiting self-adjoint operator **T** for the sequence of matrices $a_n^{-1}X$, defined on a suitable Hilbert space, and then use known spectral convergence theorems of operators. The limiting operator will be defined as the "adjacency matrix" of an infinite rooted tree with random edge weights, the so-called Poisson weighted infinite tree (PWIT) introduced by Aldous [1] (see also [3]). In other words, the PWIT will be shown to be the local weak limit of the random network $(G_n, \mathbf{X})$ when the edge marks $X_{i,j}$ are rescaled by a_n. In this setting the LSD μ_α arises as the expected value of the (random) spectral measure of the operator **T** at the root of the tree. The PWIT and the limiting operator **T** are defined in Section 2. Our method of proof can be seen as a variant of the resolvent method, based on local convergence of operators. It is also well suited to investigate properties of the LSD μ_α (cf. Theorem 1.6 below).

Let us now come back to our random reversible Markov kernel K defined by (1.1) from weights with law $\mathcal{L} \in \mathbb{H}_\alpha$. We obtain different limiting behavior in the two regimes $\alpha \in (0, 1)$ and $\alpha \in (1, 2)$. The case $\alpha > 2$ corresponds to a Wigner-type behavior (special case of Theorem 1.1). We set

$$\kappa_n = n a_n^{-1}.$$

THEOREM 1.3 [Reversible Markov matrix, $\alpha \in (1, 2)$]. *Let* μ_α *be the probability distribution which appears as the LSD in the symmetric i.i.d. case* (*Theorem* 1.2). *If* $\mathcal{L} \in \mathbb{H}_\alpha$ *with* $\alpha \in (1, 2)$ *then a.s.*

$$\mu_{\kappa_n K} := \frac{1}{n} \sum_{k=1}^{n} \delta_{\lambda_k(\kappa_n K)} \xrightarrow[n\to\infty]{w} \mu_\alpha.$$

THEOREM 1.4 [Reversible Markov matrix, $\alpha \in (0, 1)$]. *For every* $\alpha \in (0, 1)$, *there exists a symmetric probability distribution* $\widetilde{\mu}_\alpha$ *supported on* $[-1, 1]$ *depending only on* α *such that a.s.*

$$\mu_K := \frac{1}{n} \sum_{k=1}^{n} \delta_{\lambda_k(K)} \xrightarrow[n\to\infty]{w} \widetilde{\mu}_\alpha.$$

The proofs of Theorems 1.3 and 1.4 are given in Sections 3.3 and 3.1, respectively. As in the proof of Theorem 1.2, the main idea is to exploit convergence of our matrices to suitable operators defined on the PWIT. To understand the scaling in Theorem 1.3, we recall that if $\alpha > 1$, then by the strong law of large numbers, we have $n^{-1}\rho_i \to 1$ a.s. for every row sum ρ_i, and this is shown to remove, in the limit $n \to \infty$, all dependencies in the matrix $na_n^{-1}K$, so that we obtain the same behavior of the i.i.d. matrix of Theorem 1.2. On the other hand, when $\alpha \in (0, 1)$, both the sum ρ_i and the maximum of its elements are on scale a_n. The proof of Theorem 1.4 shows that the matrix K converges (without rescaling) to a random stochastic self-adjoint operator $\mathbf{K}$ defined on the PWIT. The operator $\mathbf{K}$ can be described as the transition matrix of the simple random walk on the PWIT and is naturally linked to Poisson–Dirichlet random variables. This is based on the observation that the order statistics of any given row of the matrix K converges weakly to the Poisson–Dirichlet law $\mathrm{PD}(\alpha, 0)$ (see Lemma 2.4 below for the details). In fact, the operator $\mathbf{K}$ provides an interesting generalization of the Poisson–Dirichlet law.

Since μ_K is supported in $[-1, 1]$, (1.2) and Theorem 1.4 imply that for all $\ell \geq 1$, a.s.

$$(1.8) \qquad \frac{1}{n}\sum_{i=1}^{n} p_\ell(i) = \int_{\mathbb{R}} x^\ell \mu_K(dx) \underset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}} x^\ell \widetilde{\mu}_\alpha(dx) =: \gamma_\ell.$$

The LSD $\widetilde{\mu}_\alpha$ will be obtained as the expectation of the (random) spectral measure of $\mathbf{K}$ at the root of the PWIT. It will follow that γ_ℓ (the ℓth moment of $\widetilde{\mu}_\alpha$) is the expected value of the (random) probability that the random walk returns to the root in ℓ-steps. In particular, the symmetry of $\widetilde{\mu}_\alpha$ follows from the bipartite nature of the PWIT.

It was proved by Ben Arous and Guionnet [7], Remark 1.5, that $\alpha \in (0, 2) \mapsto \mu_\alpha$ is continuous with respect to weak convergence of probability measures, and by Belinschi, Dembo and Guionnet [5], Remark 1.2 and Lemma 5.2, that μ_α tends to the Wigner semi-circle law as $\alpha \nearrow 2$. We believe that Theorem 1.3 should remain valid for $\alpha = 2$ with LSD given by the Wigner semi-circle law. Further properties of the measures μ_α and $\widetilde{\mu}_\alpha$ are discussed below.

The case $\alpha = 1$ is qualitatively similar to the case $\alpha \in (1, 2)$ with the difference that the sequence κ_n in Theorem 1.3 has to be replaced by $\kappa_n = na_n^{-1}w_n$ where

$$(1.9) \qquad w_n = \int_0^{a_n} x\mathcal{L}(dx).$$

Indeed, here the mean of $U_{i,j}$ may be infinite and the closest one gets to a law of large numbers is the statement that $\rho_i/nw_n \to 1$ in probability (see Section 3.4). The sequence w_n (and therefore κ_n) is known to be slowly varying at ∞ for $\alpha = 1$ (see, e.g., Feller [17], VIII.8). The following mild condition will be assumed: There exists $0 < \varepsilon < 1/2$ such that

$$(1.10) \qquad \liminf_{n\to\infty} \frac{w_{\lfloor n^\varepsilon \rfloor}}{w_n} > 0.$$

For example, if $U_{i,j}^{-1}$ is uniform on $[0, 1]$, then $\kappa_n = w_n = \log n$ and $\lim_{n \to \infty} w_{\lfloor n^\varepsilon \rfloor} / w_n = \varepsilon$. In the next theorem μ_1 stands for the LSD μ_α from Theorem 1.2, at $\alpha = 1$.

THEOREM 1.5 (Reversible Markov matrix, $\alpha = 1$).   *Suppose that $\mathcal{L} \in \mathbb{H}_\alpha$ with $\alpha = 1$ and assume (1.10). If $\mu_{\kappa_n K}$ is the ESD of $\kappa_n K$, with $\kappa_n = na_n^{-1} w_n$, then, as $n \to \infty$, a.s. $\mu_{\kappa_n K} \xrightarrow[n \to \infty]{w} \mu_1$.*

Properties of the LSD.   In Section 4 we prove some properties of the LSDs μ_α and $\widetilde{\mu}_\alpha$.

THEOREM 1.6 (Properties of μ_α).   *Let μ_α be the symmetric LSD in Theorems 1.2 and 1.3.*

 (i) *μ_α is absolutely continuous on $\mathbb{R}$.*
 (ii) *The density of μ_α at 0 is equal to*

$$\frac{1}{\pi} \Gamma\left(1 + \frac{2}{\alpha}\right) \left(\frac{\Gamma(1 - \alpha/2)}{\Gamma(1 + \alpha/2)}\right)^{1/\alpha}.$$

(iii) *μ_α is heavy tailed, and as t goes to $+\infty$,*

$$\mu_\alpha((t, +\infty)) \sim \tfrac{1}{2} t^{-\alpha}.$$

Statements (i) and (ii) answer some questions raised in [5, 7]. Statement (iii) is already contained in [5], Theorem 1.7, but we provide a new proof based on a Tauberian theorem for the Cauchy–Stieltjes transform that may be of independent interest.

THEOREM 1.7 (Properties of $\widetilde{\mu}_\alpha$).   *Let $\widetilde{\mu}_\alpha$ be the symmetric LSD in Theorem 1.4, with moments γ_ℓ as in (1.8). Then the following statements hold true.*

 (i) *For $\alpha \in (0, 1)$, there exists $\delta > 0$ such that*

$$\gamma_{2n} \geq \delta n^{-\alpha} \qquad \text{for all } n \geq 1.$$

Moreover, we have $\liminf_{\alpha \nearrow 1} \gamma_2 > 0$.
 (ii) *For the topology of the weak convergence, the map $\alpha \mapsto \widetilde{\mu}_\alpha$ is continuous in $(0, 1)$.*
(iii) *For the topology of the weak convergence,*

$$\lim_{\alpha \searrow 0} \widetilde{\mu}_\alpha = \frac{1}{4} \delta_{-1} + \frac{1}{2} \delta_0 + \frac{1}{4} \delta_1.$$

It is delicate to provide liable numerical simulations of the ESDs. Nevertheless, Figure 1 provides histograms for various values of α and a large value of n, illustrating Theorems 1.3–1.7.
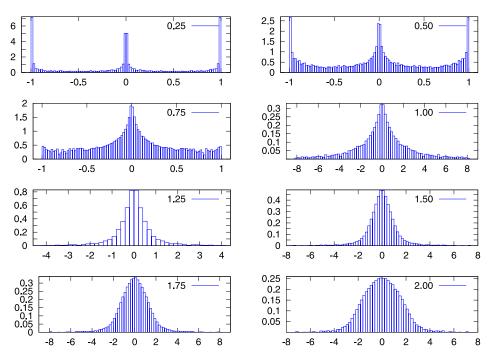
FIG. 1.   *Histograms of scaled ESDs illustrating the convergence stated by Theorems* 1.3 *and* 1.4, *for the following values of* α : 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00. *Here* $n = 5000$ *and* $\mathcal{L}$ *is the law of* $V^{-1/\alpha}$ *where* V *is a uniform random variable on* $(0, 1)$. *The first three plots are the histogram of the spectrum of a single realization of* K. *The fourth plot corresponds to* $\alpha = 1$ *and is a histogram of the spectrum of a single realization of* $\log(n)K$. *The four last plots are the histogram of the spectrum of a single realization of* $\kappa_n K$. *In order to avoid scaling problems, an asymptotically negligible portion of the spectrum edge was discarded: only* $\lambda_{\lfloor \log(n) \rfloor}, \ldots, \lambda_{\lfloor n - \log(n) \rfloor}$ *were used.*

Invariant measure and edge behavior.   Finally, we turn to the analysis of the invariant probability distribution $\hat{\rho}$ for the random walk on $(G, \mathbf{U})$. This is obtained by normalizing the vector of row sums ρ

$$\hat{\rho} = (\rho_1 + \cdots + \rho_n)^{-1}(\rho_1, \ldots, \rho_n).$$

Following [11], Lemma 2.2, if $\alpha > 2$, then $n \max_{1 \le i \le n} |\hat{\rho}_i - n^{-1}| \to 0$ as $n \to \infty$ a.s. This uniform strong law of large numbers does not hold in the heavy-tailed case $\alpha \in (0, 2)$: the large n behavior of $\hat{\rho}$ is then dictated by the largest weights in the system.

Below we use the notation $\widetilde{\rho} = (\widetilde{\rho}_1, \ldots, \widetilde{\rho}_n)$ for the ranked values of $\hat{\rho}_1, \ldots, \hat{\rho}_n$, so that $\widetilde{\rho}_1 \ge \widetilde{\rho}_2 \ge \cdots$ and their sum is 1. The symbol $\xrightarrow{d}$ denotes convergence in distribution. We refer to Section 2.4 for more details on weak convergence in the space of ranked sequences and for the definition of the Poisson–Dirichlet law $PD(\alpha, 0)$.

THEOREM 1.8 (Invariant probability measure).   *Suppose that* $\mathcal{L} \in \mathbb{H}_\alpha$.

(i) *If $\alpha \in (0, 1)$, then*

$$(1.11) \qquad \widetilde{\rho} \xrightarrow[n \to \infty]{d} \tfrac{1}{2}(V_1, V_1, V_2, V_2, \ldots),$$

where $V_1 > V_2 > \cdots$ *stands for a Poisson–Dirichlet* PD$(\alpha, 0)$ *random vector.*

(ii) *If $\alpha \in (1, 2)$, then*

$$(1.12) \qquad \kappa_{n(n+1)/2} \widetilde{\rho} \xrightarrow[n \to \infty]{d} \tfrac{1}{2}(x_1, x_1, x_2, x_2, \ldots),$$

where $x_1 > x_2 > \cdots$ *denote the ranked points of the Poisson point process on* $(0, \infty)$ *with intensity measure* $\alpha x^{-\alpha-1} dx$. *Moreover, the same convergence holds for $\alpha = 1$ provided the sequence κ_n is replaced by $n a_n^{-1} w_n$, with w_n as in* (1.9).

Theorem 1.8 is proved in Section 5. These results will be derived from the statistics of the ranked values of the weights $U_{i,j}$, $i < j$, on the scale $a_{n(n+1)/2}$ (diagonal weights $U_{i,i}$ are easily seen to give negligible contributions). The duplication in the sequences in (1.12) and (1.11) then comes from the fact that each of the largest weights belongs to two distinct rows and determines alone the limiting value of the associated row sum.

Theorem 1.8 is another indication that the random walk with transition matrix K shares the features of a *trap model*. Loosely speaking, instead of being trapped at a vertex, as in the usual mean field trap models (see [6, 14, 16, 18]) here the walker is trapped at an edge.

Large edge weights are responsible for the large eigenvalues of K. This phenomenon is well understood in the case of symmetric random matrices with i.i.d. entries, where it is known that, for $\alpha \in (0, 4)$, the edge of the spectrum gives rise to a Poisson statistics (see [4, 26]). The behavior of the extremal eigenvalues of K when $\mathcal{L}$ has finite fourth moment has been studied in [11]. In particular, it is shown there that the spectral gap $1 - \lambda_2$ is $1 - O(n^{-1/2})$. In the present case of heavy-tailed weights, in contrast, by localization on the largest edge weight it is possible to prove that, a.s. and up to corrections with slow variation at ∞,

$$(1.13) \qquad 1 - \lambda_2 = \begin{cases} O(n^{-1/\alpha}), & \alpha \in (0, 1), \\ O(n^{-(2-\alpha)/\alpha}), & \alpha \in [1, 2). \end{cases}$$

Similarly, for $\alpha \in (2, 4)$ one has that λ_2 is bounded below by $n^{-(\alpha-2)/\alpha}$. Understanding the statistics of the extremal eigenvalues remains an interesting open problem.

2. Convergence to the Poisson weighted infinite tree. The aim of this section is to prove that the matrices X and K appearing in Theorems 1.2, 1.3 and 1.4, when properly rescaled, converge "locally" to a limiting operator defined on the Poisson weighted infinite tree (PWIT). The concept of local convergence of operators is defined below. We first recall the standard construction of the PWIT.

2.1. *The PWIT.* Given a Radon measure ν on $\mathbb{R}$, PWIT(ν) is the random rooted tree defined as follows. The vertex set of the tree is identified with $\mathbb{N}^f :=$ $\bigcup_{k \in \mathbb{N}} \mathbb{N}^k$ by indexing the root as $\mathbb{N}^0 = \varnothing$, the offsprings of the root as $\mathbb{N}$ and, more generally, the offsprings of some $\mathbf{v} \in \mathbb{N}^k$ as $(\mathbf{v}1), (\mathbf{v}2), \ldots \in \mathbb{N}^{k+1}$ [for short notation, we write $(\mathbf{v}1)$ in place of $(\mathbf{v}, 1)$]. In this way the set of $\mathbf{v} \in \mathbb{N}^n$ identifies the nth generation.

We now assign marks to the edges of the tree according to a collection $\{\Xi_{\mathbf{v}}\}_{\mathbf{v} \in \mathbb{N}^f}$ of independent realizations of the Poisson point process with intensity measure ν on $\mathbb{R}$. Namely, starting from the root $\varnothing$, let $\Xi_\varnothing = \{y_1, y_2, \ldots\}$ be ordered in such a way that $|y_1| \leq |y_2| \leq \cdots$, and assign the mark y_i to the offspring of the root labeled i. Now, recursively, at each vertex $\mathbf{v}$ of generation k, assign the mark $y_{\mathbf{v}i}$ to the offspring labeled $\mathbf{v}i$, where $\Xi_{\mathbf{v}} = \{y_{\mathbf{v}1}, y_{\mathbf{v}2}, \ldots\}$ satisfy $|y_{\mathbf{v}1}| \leq |y_{\mathbf{v}2}| \leq \cdots$.

2.2. *Local operator convergence.* We give a general formulation and later specialize to our setting. Let V be a countable set, and let $L^2(V)$ denote the Hilbert space defined by the scalar product

$$\langle \phi, \psi \rangle := \sum_{u \in V} \bar{\phi}_u \psi_u, \qquad \phi_u = \langle \delta_u, \phi \rangle,$$

where $\phi, \psi \in \mathbb{C}^V$ and δ_u denote the unit vector with support u. Let $\mathcal{D}$ denote the dense subset of $L^2(V)$ of vectors with finite support.

DEFINITION 2.1 (Local convergence). Suppose $\mathbf{S}_n$ is a sequence of bounded operators on $L^2(V)$, and $\mathbf{S}$ is a closed linear operator on $L^2(V)$ with dense domain $D(\mathbf{S}) \supset \mathcal{D}$. Suppose further that $\mathcal{D}$ is a core for $\mathbf{S}$ (i.e., the closure of $\mathbf{S}$ restricted to $\mathcal{D}$ equals $\mathbf{S}$). For any $u, v \in V$ we say that $(\mathbf{S}_n, u)$ converges locally to $(\mathbf{S}, v)$ and write

$$(\mathbf{S}_n, u) \to (\mathbf{S}, v),$$

if there exists a sequence of bijections $\sigma_n : V \to V$ such that $\sigma_n(v) = u$ and, for all $\phi \in \mathcal{D}$,

$$\sigma_n^{-1} \mathbf{S}_n \sigma_n \phi \to \mathbf{S}\phi,$$

in $L^2(V)$, as $n \to \infty$.

In other words, this is the standard strong convergence of operators up to a re-indexing of V which preserves a distinguished element. With a slight abuse of notation we have used the same symbol σ_n for the linear isometry $\sigma_n : L^2(V) \to L^2(V)$ induced in the obvious way, that is, such that $\sigma_n \delta_v = \delta_{\sigma_n(v)}$ for all $v \in V$. The point for introducing Definition 2.1 lies in the following theorem on strong resolvent convergence. Recall that if $\mathbf{S}$ is a self-adjoint operator its spectrum is real, and for all $z \in \mathbb{C}_+ := \{z \in \mathbb{C} : \Im z > 0\}$, the operator $\mathbf{S} - zI$ is invertible with bounded inverse. The operator-valued function $z \mapsto (\mathbf{S} - zI)^{-1}$ is the resolvent of $\mathbf{S}$.

THEOREM 2.2 (From local convergence to resolvents). *If $\mathbf{S}_n$ and $\mathbf{S}$ are self-adjoint operators that satisfy the conditions of Definition* 2.1 *and* $(\mathbf{S}_n, u) \to (\mathbf{S}, v)$ *for some* $u, v \in V$, *then, for all* $z \in \mathbb{C}_+$,

$$(2.1) \qquad \langle \delta_u, (\mathbf{S}_n - zI)^{-1} \delta_u \rangle \to \langle \delta_v, (\mathbf{S} - zI)^{-1} \delta_v \rangle.$$

PROOF. It is a special case of [23], Theorem VIII.25(a). Indeed, if we define $\widetilde{\mathbf{S}}_n = \sigma_n^{-1} \mathbf{S}_n \sigma_n$, then $\widetilde{\mathbf{S}}_n \phi \to \mathbf{S} \phi$ for all ϕ in a common core of the self-adjoint operators $\widetilde{\mathbf{S}}_n, \mathbf{S}$. This implies the strong resolvent convergence, that is, $(\widetilde{\mathbf{S}}_n - zI)^{-1} \psi \to (\mathbf{S} - zI)^{-1} \psi$ for any $z \in \mathbb{C}_+$, $\psi \in L^2(V)$. The conclusion follows by taking the scalar product

$$\langle \delta_v, (\widetilde{\mathbf{S}}_n - zI)^{-1} \delta_v \rangle = \langle \delta_u, (\mathbf{S}_n - zI)^{-1} \delta_u \rangle. \qquad \square$$

We shall apply the above theorem in cases where the operators $\mathbf{S}_n$ and $\mathbf{S}$ are random operators on $L^2(V)$, which satisfy with probability one the conditions of Definition 2.1. In this case we say that $(\mathbf{S}_n, u) \to (\mathbf{S}, v)$ *in distribution* if there exists a random bijection σ_n as in Definition 2.1 such that $\sigma_n^{-1} \mathbf{S}_n \sigma_n \phi$ converges in distribution to $\mathbf{S}\phi$, for all $\phi \in \mathcal{D}$ [where a random vector $\psi_n \in L^2(V)$ converges in distribution to ψ if

$$\lim_{n \to \infty} \mathbb{E} f(\psi_n) = \mathbb{E} f(\psi)$$

for all bounded continuous functions $f : L^2(V) \to \mathbb{R}$]. Under these assumptions then (2.1) becomes convergence in distribution of (bounded) complex random variables. In our setting the Hilbert space will be $L^2(V)$, with $V = \mathbb{N}^f$, the vertex set of the PWIT, the operator $\mathbf{S}_n$ will be a rescaled version of the matrix X defined by (1.6) or the matrix K defined by (1.1). The operator $\mathbf{S}$ will be the corresponding limiting operator defined below.

2.3. *Limiting operators.* Let θ be as in Theorem 1.2, and let ℓ_θ be the positive Borel measure on the real line defined by $d\ell_\theta(x) = \theta \mathbb{1}_{\{x > 0\}} dx + (1 - \theta) \mathbb{1}_{\{x < 0\}} dx$. Consider a realization of PWIT(ℓ_θ). As before the mark from vertex $\mathbf{v} \in \mathbb{N}^k$ to $\mathbf{v}k \in \mathbb{N}^{k+1}$ is denoted by $y_{\mathbf{v}k}$. We note that almost surely

$$(2.2) \qquad \sum_k |y_{\mathbf{v}k}|^{-2/\alpha} < \infty,$$

since a.s. $\lim_k |y_{\mathbf{v}k}|/k = 1$ and $\sum_k k^{-2/\alpha}$ converges for $\alpha \in (0, 2)$. Recall that for $V = \mathbb{N}^f$, $\mathcal{D}$ is the dense set of $L^2(V)$ of vectors with finite support. We may a.s. define a linear operator $\mathbf{T} : \mathcal{D} \to L^2(V)$ by letting, for $\mathbf{v}, \mathbf{w} \in \mathbb{N}^f$,

$$(2.3) \quad \begin{aligned} \mathbf{T}(\mathbf{v}, \mathbf{w}) &= \langle \delta_{\mathbf{v}}, \mathbf{T} \delta_{\mathbf{w}} \rangle \\ &= \begin{cases} \operatorname{sign}(y_{\mathbf{w}}) |y_{\mathbf{w}}|^{-1/\alpha}, & \text{if } \mathbf{w} = \mathbf{v}k \text{ for some integer } k, \\ \operatorname{sign}(y_{\mathbf{v}}) |y_{\mathbf{v}}|^{-1/\alpha}, & \text{if } \mathbf{v} = \mathbf{w}k \text{ for some integer } k, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Note that if every edge e in the tree with mark y_e is given the "weight" $\text{sign}(y_e)|y_e|^{-1/\alpha}$ then we may look at the operator $\mathbf{T}$ as the "adjacency matrix" of the weighted tree. Clearly, $\mathbf{T}$ is symmetric, and therefore it has a closed extension with domain $D(\mathbf{T}) \subset L^2(\mathbb{N}^f)$ such that $\mathcal{D} \subset D(\mathbf{T})$ (see, e.g., [23], Chapter VIII, Section 2). We will prove in Proposition A.2 below that $\mathbf{T}$ is essentially self-adjoint, that is, the closure of $\mathbf{T}$ is self-adjoint. With a slight abuse of notation, we identify $\mathbf{T}$ with its closed extension. As stated below, $\mathbf{T}$ is the weak local limit of the sequence of $n \times n$ i.i.d. matrices $a_n^{-1} X$, where X is defined by (1.6). To this end we view the matrix X as an operator in $L^2(V)$ by setting $\langle \delta_i, X\delta_j \rangle = X_{i,j}$, where $i, j \in \mathbb{N}$ denote the labels of the offsprings of the root (the first generation), with the convention that $X_{i,j} = 0$ when either $i > n$ or $j > n$, and by setting $\langle \delta_\mathbf{u}, X\delta_\mathbf{v} \rangle = 0$ when either $\mathbf{u}$ or $\mathbf{v}$ does not belong to the first generation.

Similarly, taking now $\theta = 1$, in the case of Markov matrices K defined by (1.1), for $\alpha \in [1, 2)$, $\mathbf{T}$ is the local limit operator of $\kappa_n K$. To work directly with symmetric operators we introduce the symmetric matrix

$$(2.4) \qquad\qquad S_{i,j} = \frac{U_{i,j}}{\sqrt{\rho_i \rho_j}},$$

which is easily seen to have the same spectrum of K (see, e.g., [11], Lemma 2.1). Again the matrix S can be embedded in the infinite tree as described above for X.

In the case $\alpha \in (0, 1)$ the Markov matrix K has a different limiting object that is defined as follows. Consider a realization of $\text{PWIT}(\ell_1)$, where ℓ_1 is the Lebesgue measure on $[0, \infty)$. We define an operator corresponding to the random walk on this tree with conductance equal to the mark to the power $-1/\alpha$. More precisely, for $\mathbf{v} \in \mathbb{N}^f$, let

$$\rho(\mathbf{v}) = y_\mathbf{v}^{-1/\alpha} + \sum_{k \in \mathbb{N}} y_{\mathbf{v}k}^{-1/\alpha}$$

with the convention that $y_\varnothing^{-1/\alpha} = 0$. Since a.s. $\lim_k |y_{\mathbf{v}k}|/k = 1$, $\rho(\mathbf{v})$ is almost surely finite for $\alpha \in (0, 1)$. We define the linear operator $\mathbf{K}$ on $\mathcal{D}$, by letting, for $\mathbf{v}, \mathbf{w} \in \mathbb{N}^f$,

$$(2.5) \quad \mathbf{K}(\mathbf{v}, \mathbf{w}) = \langle \delta_\mathbf{v}, \mathbf{K}\delta_\mathbf{w} \rangle = \begin{cases} \dfrac{y_\mathbf{w}^{-1/\alpha}}{\rho(\mathbf{v})}, & \text{if } \mathbf{w} = \mathbf{v}k \text{ for some integer } k, \\ \dfrac{y_\mathbf{v}^{-1/\alpha}}{\rho(\mathbf{v})}, & \text{if } \mathbf{v} = \mathbf{w}k \text{ for some integer } k, \\ 0, & \text{otherwise.} \end{cases}$$

Note that $\mathbf{K}$ is not symmetric, but it becomes symmetric in the weighted Hilbert space $L^2(V, \rho)$ defined by the scalar product

$$\langle \phi, \psi \rangle_\rho := \sum_{\mathbf{u} \in V} \rho(\mathbf{u}) \bar{\phi}_\mathbf{u} \psi_\mathbf{u}.$$

Moreover, on $L^2(V, \rho)$, $\mathbf{K}$ is a bounded self-adjoint operator since Schwarz's inequality implies

$$\langle \mathbf{K}\phi, \mathbf{K}\phi \rangle_\rho^2 = \sum_{\mathbf{u}} \rho(\mathbf{u}) \left| \sum_{\mathbf{v}} \mathbf{K}(\mathbf{u}, \mathbf{v}) \phi_{\mathbf{v}} \right|^2$$

$$\leq \sum_{\mathbf{u}} \rho(\mathbf{u}) \sum_{\mathbf{v}} \mathbf{K}(\mathbf{u}, \mathbf{v}) |\phi_{\mathbf{v}}|^2$$

$$= \sum_{\mathbf{v}} \rho(\mathbf{v}) |\phi_{\mathbf{v}}|^2 = \langle \phi, \phi \rangle_\rho^2.$$

so that the operator norm of $\mathbf{K}$ is less than or equal to 1. To work with self-adjoint operators in the unweighted Hilbert space $L^2(V)$ we shall actually consider the operator $\mathbf{S}$ defined by

$$(2.6) \qquad \mathbf{S}(\mathbf{v}, \mathbf{w}) := \sqrt{\frac{\rho(\mathbf{v})}{\rho(\mathbf{w})}} \mathbf{K}(\mathbf{v}, \mathbf{w}) = \frac{\mathbf{T}(\mathbf{v}, \mathbf{w})}{\sqrt{\rho(\mathbf{v})\rho(\mathbf{w})}}.$$

This defines a bounded self-adjoint operator in $L^2(V)$. Indeed, the map $\delta_{\mathbf{v}} \to \sqrt{\rho(\mathbf{v})}\delta_{\mathbf{v}}$ induces a linear isometry $\mathbf{D}: L^2(V, \rho) \to L^2(V)$ such that

$$(2.7) \qquad \langle \phi, \mathbf{S}\psi \rangle = \langle \mathbf{D}^{-1}\phi, \mathbf{K}\mathbf{D}^{-1}\psi \rangle_\rho,$$

for all $\phi, \psi \in L^2(V)$. In this way, when $\alpha \in (0, 1)$, $\mathbf{S}$ will be the limiting operator associated with the matrix S defined in (2.4). Note that no rescaling is needed here. The main result of this section is the following.

THEOREM 2.3 (Limiting operators). *As n goes to infinity, in distribution*:

(i) *if $\alpha \in (0, 2)$ and $\theta \in [0, 1]$, then $(a_n^{-1}X, 1) \to (\mathbf{T}, \varnothing)$;*
(ii) *if $\alpha \in (1, 2)$ and $\theta = 1$, then $(\kappa_n S, 1) \to (\mathbf{T}, \varnothing)$;*
(iii) *if $\alpha \in (0, 1)$, then $(S, 1) \to (\mathbf{S}, \varnothing)$.*

From the remark after Theorem 2.2 we see that Theorem 2.3 implies convergence in distribution of the resolvent at the root. As we shall see in Section 3, this in turn gives convergence of the expected values of the Cauchy–Stieltjes transform of the ESD of our matrices. The rest of this section is devoted to the proof of Theorem 2.3.

2.4. *Weak convergence of a single row.* In this paragraph, we recall some facts about the order statistics of the first row of the matrix X and K, that is,

$$(X_{1,1}, \ldots, X_{1,n}) \quad \text{and} \quad (U_{1,1}, \ldots, U_{1,n})/\rho_1,$$

where $U_{1,j} = |X_{1,j}|$ has law $\mathbb{H}_\alpha$. Let us denote by $V_1 \geq V_2 \geq \cdots \geq V_n$ the order statistics of the variables $U_{1,j}$, $1 \leq j \leq n$. Recall that $\rho_1 = \sum_{j=1}^n V_j$. Let us define

$\Delta_{k,n} = \sum_{j=k+1}^{n} V_j$ for $k < n$ and $\Delta_{k,n}^2 = \sum_{j=k+1}^{n} V_j^2$. Call $\mathcal{A}$ the set of sequences $\{v_j\} \in [0, \infty)^{\mathbb{N}}$ with $v_1 \geq v_2 \geq \cdots \geq 0$ such that $\lim_{j \to \infty} v_j = 0$, and let $\mathcal{A}_1 \subset \mathcal{A}$ be the subset of sequences satisfying $\sum_j v_j = 1$. We shall view

$$Y_n = \left(\frac{V_1}{a_n}, \ldots, \frac{V_n}{a_n} \right) \quad \text{and} \quad Z_n = \left(\frac{V_1}{\rho_1}, \ldots, \frac{V_n}{\rho_1} \right)$$

as elements of $\mathcal{A}$ and $\mathcal{A}_1$, respectively, simply by adding zeros to the right of V_n/a_n and V_n/ρ_1. Equipped with the standard product metric, $\mathcal{A}$ and $\mathcal{A}_1$ are complete separable metric spaces ($\mathcal{A}_1$ is compact), and convergence in distribution for $\mathcal{A}$, $\mathcal{A}_1$-valued random variables is equivalent to finite-dimensional convergence (cf., e.g., Bertoin [9]).

Let $E_1, E_2, \ldots$ denote i.i.d. exponential variables with mean 1 and write $\gamma_k = \sum_{j=1}^{k} E_j$. We define the random variable in $\mathcal{A}$

$$Y = (\gamma_1^{-1/\alpha}, \gamma_2^{-1/\alpha}, \ldots).$$

The law of Y is the law of the ordered points of a Poisson process on $(0, \infty)$ with intensity measure $\alpha x^{-\alpha-1} \, dx$. For $\alpha \in (0, 1)$ we define the variable in $\mathcal{A}_1$

$$Z = \left(\frac{\gamma_1^{-1/\alpha}}{\sum_{n=1}^{\infty} \gamma_n^{-1/\alpha}}, \frac{\gamma_2^{-1/\alpha}}{\sum_{n=1}^{\infty} \gamma_n^{-1/\alpha}}, \cdots \right).$$

For $\alpha \in (0, 1)$ the sum $\sum_n \gamma_n^{-1/\alpha}$ is a.s. finite. The law of Z in $\mathcal{A}_1$ is called the *Poisson–Dirichlet* law $\mathrm{PD}(\alpha, 0)$ (see Pitman and Yor [22], Proposition 10). The next result is rather standard but we give a simple proof for convenience.

LEMMA 2.4 (Poisson–Dirichlet laws and Poisson point processes).

(i) *For all $\alpha > 0$, Y_n converges in distribution to Y. Moreover, for $\alpha \in (0, 2)$, $(a_n^{-1} V_j)_{j \geq 1}$ is a.s. uniformly square integrable, that is, a.s. $\lim_k \sup_{n>k} a_n^{-2} \times \Delta_{k,n}^2 = 0$.*

(ii) *If $\alpha \in (0, 1)$, Z_n converges in distribution to Z. Moreover, $(a_n^{-1} V_j)_{j \geq 1}$ is a.s. uniformly integrable, that is, a.s. $\lim_k \sup_{n>k} a_n^{-1} \Delta_{k,n} = 0$.*

(iii) *If $I \subset \mathbb{N}$ is a finite set and $V_1^I \geq V_2^I \geq \cdots$ denote the order statistics of $\{U_{1,j}\}_{j \in \{1,\ldots,n\} \setminus I}$ then (i) and (ii) hold with $Y_n^I = (V_1^I/a_n, V_2^I/a_n, \ldots)$ and $Z_n^I = (V_1^I/\rho_1, V_2^I/\rho_1, \ldots)$.*

As an example, from (i), we retrieve the well-known fact that for any $\alpha > 0$, the random variable $a_n^{-1} \max(U_{1,1}, \ldots, U_{1,n})$ converges weakly as $n \to \infty$ to the law of $\gamma_1^{-1/\alpha}$. This law, known as a Fréchet law, has density $\alpha x^{-\alpha-1} e^{-x^{-\alpha}}$ on $(0, \infty)$.

PROOF OF LEMMA 2.4. As in LePage, Woodroofe and Zinn [20] we take advantage of the following well-known representation for the order statistics of

i.i.d. random variables. Let G be the function in (1.4) and write

$$G^{-1}(u) = \inf\{y > 0 : G(y) \le u\},$$

$u \in (0, 1)$. We have that $(V_1, \dots, V_n)$ equals in distribution the vector

(2.8)
$$\left(G^{-1}(\gamma_1/\gamma_{n+1}), \dots, G^{-1}(\gamma_n/\gamma_{n+1})\right),$$

where γ_j has been defined above. To prove (i) we start from the distributional identity

$$Y_n \stackrel{d}{=} \left(\frac{G^{-1}(\gamma_1/\gamma_{n+1})}{a_n}, \dots, \frac{G^{-1}(\gamma_n/\gamma_{n+1})}{a_n}\right),$$

which follows from (2.8). It suffices to prove that for every k, almost surely the first k terms above converge to the first k terms in Y. Thanks to (1.5), almost surely, for every j,

(2.9)
$$a_n^{-1} G^{-1}(\gamma_j/\gamma_{n+1}) \to \gamma_j^{-1/\alpha},$$

and the convergence in distribution of Y_n to Y follows. Moreover, from (1.5), for any $\delta > 0$ we can find n_0 such that

$$a_n^{-1} V_j = a_n^{-1} G^{-1}(\gamma_j/\gamma_{n+1}) \le \left(n\gamma_j/(1+\delta)\gamma_{n+1}\right)^{-1/\alpha},$$

for $n \ge n_0$, $j \in \mathbb{N}$. Since $n/\gamma_{n+1} \to 1$, a.s. we see that the expression above is a.s. bounded by $2(1+\delta)^{1/\alpha}\gamma_j^{-1/\alpha}$, for n sufficiently large, and the second part of (i) follows from a.s. summability of $\gamma_j^{-2/\alpha}$.

Similarly, if $\alpha \in (0, 1)$, $\Delta_{k,n}$ has the same law of

$$\sum_{j=k+1}^{n} G^{-1}(\gamma_j/\gamma_{n+1}),$$

and the second part of (ii) follows from a.s. summability of $\gamma_j^{-1/\alpha}$. To prove the convergence of Z_n we use the distributional identity

$$Z_n \stackrel{d}{=} \left(\frac{G^{-1}(\gamma_1/\gamma_{n+1})}{\sum_{j=1}^{n} G^{-1}(\gamma_j/\gamma_{n+1})}, \dots, \frac{G^{-1}(\gamma_n/\gamma_{n+1})}{\sum_{j=1}^{n} G^{-1}(\gamma_j/\gamma_{n+1})}\right).$$

As a consequence of (2.9), we then have almost surely

$$a_n^{-1} \sum_{j=1}^{n} G^{-1}(\gamma_j/\gamma_{n+1}) \to \sum_{j=1}^{\infty} \gamma_j^{-1/\alpha},$$

and (ii) follows. Finally, (iii) is an easy consequence of the exchangeability of the variable $(U_{1,i})$

$$\mathbb{P}(V_k^I \ne V_k) \le \mathbb{P}(\exists j \in I : U_{1,j} \ge V_k) \le |I|\mathbb{P}(U_{1,1} \ge V_k) = |I|\frac{k}{n}. \qquad \square$$

The intensity measure $\alpha x^{-\alpha-1}\, dx$ on $(0, \infty)$ is not locally finite at 0. It will be more convenient to work with Radon (i.e., locally finite) intensity measures.

LEMMA 2.5 (Poisson point processes with Radon intensity measures). *Let* $\xi_1^n, \xi_2^n, \dots$ *be sequences of i.i.d. random variables on* $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ *such that*

$$(2.10) \qquad n\mathbb{P}(\xi_1^n \in \cdot) \xrightarrow[n\to\infty]{w} \nu,$$

where ν *is a Radon measure on* $\mathbb{R}$. *Then, for any finite set* $I \subset \mathbb{N}$ *the random measure*

$$\sum_{i \in \{1,\dots,n\} \setminus I} \delta_{\xi_i^n}$$

converges weakly as $n \to \infty$ *to* $\mathrm{PPP}(\nu)$, *the Poisson point process on* $\mathbb{R}$ *with intensity law* ν, *for the usual vague topology on Radon measures.*

We refer to [24], Theorem 5.3, page 138, for a proof of Lemma 2.5. Note that for $\xi_j^{(n)} = a_n / U_{1,j}$ it is a consequence of Lemma 2.4(iii). In the case $\xi_j^{(n)} = a_n / X_{1,j}$, where $X_{i,j}$ is as in (1.6) and (1.7), the above lemma yields convergence to $\mathrm{PPP}(\nu_{\alpha,\theta})$, where

$$(2.11) \qquad \nu_{\alpha,\theta}(dx) = \left[\theta \mathbb{1}_{\{x>0\}} + (1-\theta)\mathbb{1}_{\{x<0\}}\right]\alpha|x|^{\alpha-1}\,dx.$$

2.5. *Local weak convergence to PWIT.* In the previous paragraph we have considered the convergence of the first row of the matrix $a_n^{-1}X$. Here we generalize this by characterizing the limiting local structure of the complete graph with marks $a_n / X_{i,j}$. Our argument is based on a technical generalization of an argument borrowed from Aldous [1]. This will lead us to Theorems 2.3 and 2.8 below.

Let G_n be the complete network on $\{1,\dots,n\}$ whose mark on edge (i,j) equals $\xi_{i,j}^n$, for some collection $(\xi_{ij}^n)_{1\le i \le j \le n}$ of i.i.d. random variables with values in $\mathbb{R}$, with $\xi_{j,i}^n = \xi_{i,j}^n$. We consider the rooted network $(G_n, 1)$ obtained by distinguishing the vertex labeled 1.

We follow Aldous [1], Section 3. For every fixed realization of the marks (ξ_{ij}^n), and for any $B, H \in \mathbb{N}$, such that $(B^{H+1} - 1)/(B-1) \le n$, we define a finite rooted subnetwork $(G_n, 1)^{B,H}$ of $(G_n, 1)$, whose vertex set coincides with a B-ary tree of depth H with root at 1.

To this end we partially index the vertices of $(G_n, 1)$ as elements in

$$J_{B,H} = \bigcup_{\ell=0}^{H} \{1,\dots,B\}^\ell \subset \mathbb{N}^f,$$

the indexing being given by an injective map σ_n from $J_{B,H}$ to $V_n := \{1,\dots,n\}$. The map σ_n can be extended to a bijection from a subset of $\mathbb{N}^f$ to V_n. We set $I_\varnothing = \{1\}$ and the index of the root 1 is $\sigma_n^{-1}(1) = \varnothing$. The vertex $v \in V_n \setminus I_\varnothing$ is given the index $(k) = \sigma_n^{-1}(v)$, $1 \le k \le B$, if $\xi_{(1,v)}^n$ has the kth smallest absolute

value among $\{\xi_{1,j}^n, j \neq 1\}$, the marks of edges emanating from the root 1. We break ties by using the lexicographic order. This defines the first generation. Now let I_1 be the union of $I_\varnothing$ and the B vertices that have been selected. If $H \geq 2$, we repeat the indexing procedure for the vertex indexed by (1) (the first child) on the set $V_n \setminus I_1$. We obtain a new set $\{11, \ldots, 1B\}$ of vertices sorted by their weights as before [for short notation, we concatenate the vector $(1, 1)$ into 11]. Then we define I_2 as the union of I_1 and this new collection. We repeat the procedure for (2) on $V_n \setminus I_2$ and obtain a new set $\{21, \ldots, 2B\}$, and so on. When we have constructed $\{B1, \ldots, BB\}$, we have finished the second generation (depth 2) and we have indexed $(B^3 - 1)/(B - 1)$ vertices. The indexing procedure is then repeated until depth H so that $(B^{H+1} - 1)/(B - 1)$ vertices are sorted. Call this set of vertices $V_n^{B,H} = \sigma_n J_{B,H}$. The subnetwork of G_n generated by $V_n^{B,H}$ is denoted $(G_n, 1)^{B,H}$ (it can be identified with the original network G_n where any edge e touching the complement of $V_n^{B,H}$ is given a mark $x_e = \infty$). In $(G_n, 1)^{B,H}$, the set $\{\mathbf{u}1, \ldots, \mathbf{u}B\}$ is called the set of children or offsprings of the vertex $\mathbf{u}$. Note that while the vertex set has been given a tree structure, $(G_n, 1)^{B,H}$ is still a complete network. The next proposition shows that it nevertheless converges to a tree (i.e., all circuits vanish, or equivalently, the extra marks diverge to ∞) if the $\xi_{i,j}^n$ satisfy a suitable scaling assumption.

Let $(\mathcal{T}, \varnothing)$ denote the infinite random rooted network with distribution PWIT(ν). We call $(\mathcal{T}, \varnothing)^{B,H}$ the finite random network obtained by the sorting procedure described in the previous paragraph. Namely, $(\mathcal{T}, \varnothing)^{B,H}$ consists of the sub-tree with vertices of the form $\mathbf{u} \in J_{B,H}$, with the marks inherited from the infinite tree. If an edge is not present in $(\mathcal{T}, \varnothing)^{B,H}$, we assign to it the mark $+\infty$.

We say that the sequence of random finite networks $(G_n, 1)^{B,H}$ converges in distribution (as $n \to \infty$) to the random finite network $(\mathcal{T}, \varnothing)^{B,H}$ if the joint distributions of the marks converge weakly. To make this precise we have to add the points $\{\pm\infty\}$ as possible values for each mark, and continuous functions on the space of marks have to be understood as functions such that the limit as any one of the marks diverges to $+\infty$ exists and coincides with the limit as the same mark diverges to $-\infty$. The next proposition generalizes [1], Section 3.

PROPOSITION 2.6 (Local weak convergence to a tree). *Let $(\xi_{i,j}^n)_{1 \leq i \leq j \leq n}$ be a collection of i.i.d. random variables with values in $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ and set $\xi_{j,i}^n = \xi_{i,j}^n$. Let ν be a Radon measure on $\mathbb{R}$ with no mass at 0 and assume that*

$$(2.12) \qquad n\mathbb{P}(\xi_{12}^n \in \cdot) \xrightarrow[n\to\infty]{w} \nu \qquad as\ n \to \infty.$$

Let G_n be the complete network on $\{1, \ldots, n\}$ whose mark on edge (i, j) equals ξ_{ij}^n. Then, for all integers B, H, as n goes to infinity, in distribution,

$$(G_n, 1)^{B,H} \longrightarrow (\mathcal{T}, \varnothing)^{B,H}.$$

Moreover, if $\mathcal{T}_1, \mathcal{T}_2$ are independent with common law PWIT(ν), *then, in distribution,*

$$((G_n, 1)^{B,H}, (G_n, 2)^{B,H}) \longrightarrow ((\mathcal{T}_1, \varnothing)^{B,H}, (\mathcal{T}_2, \varnothing)^{B,H}).$$

The second statement is the convergence of the joint law of the finite networks, where $(G_n, 2)^{B,H}$ is obtained with the same procedure as for $(G_n, 1)^{B,H}$, by starting from the vertex 2 instead of 1. In particular, the second statement implies the first.

This type of convergence is often referred to as *local weak convergence*, a notion introduced by Benjamini and Schramm [8] and Aldous and Steele [3] (see also Aldous and Lyons [2]). Let us give some examples of application of this proposition. Consider the case where $\xi_{ij}^n = 1$ with probability λ/n and $\xi_{i,j}^n = \infty$ otherwise. The network G_n is an Erdős–Rényi random graph with parameter λ/n. From the proposition, we retrieve the well-known fact that it locally converges to the tree of a Yule process of intensity λ. If $\xi_{i,j}^n = nY_{i,j}$, where $Y_{i,j}$ is any nonnegative continuous random variable with density 1 at 0+, then the network converges to PWIT(ℓ_1), where ℓ_1 is the Lebesgue measure on $[0, \infty)$. The relevant application for our purpose is given by the choice $\xi_{i,j}^n = (a_n/X_{i,j})$, and $\nu = \nu_{\alpha,\theta}$, where $X_{i,j}$ are such that $|X_{i,j}| \in \mathbb{H}_\alpha$ and (1.7) is satisfied, and $\nu_{\alpha,\theta}$ is defined by (2.11). Note that the proposition applies to all $\alpha > 0$ in this setting.

PROOF OF PROPOSITION 2.6.   We order the elements of $J_{B,H}$ in the lexicographic order, that is, $\varnothing \prec 1 \prec 2 \prec \cdots \prec B \prec 11 \prec 12 \prec \cdots \prec B \cdots B$. For $\mathbf{v} \in J_{B,H}$, let $O_\mathbf{v}$ denote the set of offsprings of $\mathbf{v}$ in $(G_n, 1)^{B,H}$. By construction, we have $I_\varnothing = \{1\}$ and $I_\mathbf{v} = \sigma_n(\bigcup_{\mathbf{w} \prec \mathbf{v}} O_\mathbf{w})$. At every step of the indexing procedure, we sort the marks of the neighboring edges that have not been explored at an earlier step $\{1, \ldots, n\} \setminus I_1, \{1, \ldots, n\} \setminus I_2, \ldots$. Therefore, for all $\mathbf{u}$,

$$(2.13) \qquad\qquad (\xi_{\sigma_n(\mathbf{u}),i}^n)_{i \notin I_\mathbf{u}} \overset{d}{=} (\xi_{1,i}^n)_{1 \le i \le n - |I_\mathbf{u}|}.$$

Thus, from Lemma 2.5 and the independence of the variables ξ^n, we infer that the marks from a parent to its offsprings in $(G_n, 1)^{B,H}$ converge weakly to those in $(\mathcal{T}, \varnothing)^{B,H}$. We now check that all other marks diverge to infinity. For $\mathbf{v}, \mathbf{w} \in J_{B,H}$, we define

$$x_{\mathbf{v},\mathbf{w}}^n = \xi_{\sigma_n(\mathbf{v}),\sigma_n(\mathbf{w})}^n.$$

Also, let $\{y_{\mathbf{v},\mathbf{w}}^n, \mathbf{v}, \mathbf{w} \in J_{B,H}\}$ denote independent variables distributed as $|\xi_{1,2}^n|$. Let $E^{B,H}$ denote the set of edges $\{\mathbf{u}, \mathbf{v}\} \in J_{B,H} \times J_{B,H}$ that do not belong to the finite tree (i.e., there is no $k \in \{1, \ldots, B\}$ such that $\mathbf{u} = \mathbf{v}k$ or $\mathbf{v} = \mathbf{u}k$). Lemma 2.7 below implies that the vector $\{|x_{\mathbf{v},\mathbf{w}}^n|, \{\mathbf{v}, \mathbf{w}\} \in E^{B,H}\}$ stochastically dominates the vector $\mathcal{Y}^n := \{y_{\mathbf{v},\mathbf{w}}^n, \{\mathbf{v}, \mathbf{w}\} \in E^{B,H}\}$, that is, there exists a coupling of the two vectors such that almost surely $|x_{\mathbf{v},\mathbf{w}}^n| \ge y_{\mathbf{v},\mathbf{w}}^n$, for all $\{\mathbf{v}, \mathbf{w}\} \in E^{B,H}$. Since $J_{B,H}$ is finite

(independent of n), $\mathcal{Y}^n$ contains a finite number of variables and (2.12) implies that the probability of the event $\{\min_{\{\mathbf{v},\mathbf{w}\}\in E_{B,H}} |x_{\mathbf{v},\mathbf{w}}^n| \leq t\}$ goes to 0 as $n \to \infty$, for any $t > 0$. Therefore it is now standard to obtain that if x_e denote the mark of edge e in $\mathcal{T}^{B,H}$, the finite collection of marks $(x_e^n)_{e \in J_{B,H} \times J_{B,H}}$ converges in distribution to $(x_e)_{e \in J_{B,H} \times J_{B,H}}$ as $n \to \infty$. In other words, $(G_n, 1)^{B,H}$ converges in distribution to $(\mathcal{T}, \varnothing)^{B,H}$.

It remains to prove the second statement. It is an extension of the above argument. We consider the two subnetworks $(G_n, 1)^{B,H}$ and $(G_n, 2)^{B,H}$ obtained from $(G_n, 1)$ and $(G_n, 2)$. This gives rise to two increasing sequences of sets of vertices $I_{\mathbf{v},1}$ and $I_{\mathbf{v},2}$ with $\mathbf{v} \in J_{B,H}$ and two injective maps $\sigma_{n,1}, \sigma_{n,2}$ from $J_{B,H}$ to $\{1, \ldots, n\}$. We need to show that, in distribution,

$$(2.14) \qquad ((G_n, 1)^{B,H}, (G_n, 2)^{B,H}) \longrightarrow ((\mathcal{T}_1, \varnothing)^{B,H}, (\mathcal{T}_2, \varnothing)^{B,H}).$$

Let $V_{n,i}^{B,H} = \sigma_{n,i}(J_{B,H})$ be the vertex set of $(G_n, i)^{B,H}$, $i = 1, 2$. There are

$$C := \frac{B^{H+1} - 1}{B - 1}$$

vertices in $V_{n,i}^{B,H}$, hence the exchangeability of the variables implies that

$$\mathbb{P}(2 \in V_{n,1}^{B,H}) \leq \frac{C}{n}.$$

Let $\widetilde{G}_n = G_n \setminus V_{n,1}^{B,H}$, the subnetwork of G_n spanned by the vertex set $V \setminus V_{n,1}^{B,H}$. Assuming that $2(B^{H+1} - 1)/(B - 1) < n$ and $2 \notin V_{n,1}^{B,H}$, we may then define $(\widetilde{G}_n, 2)^{B,H}$. If $2 \in V_{n,1}^{B,H}$, $(\widetilde{G}_n, 2)^{B,H}$ is defined arbitrarily. The above analysis shows that, in distribution,

$$((G_n, 1)^{B,H}, (\widetilde{G}_n, 2)^{B,H}) \longrightarrow ((\mathcal{T}_1, \varnothing)^{B,H}, (\mathcal{T}_2, \varnothing)^{B,H}).$$

Therefore in order to prove (2.14) it is sufficient to prove that with probability tending to 1,

$$V_{n,1}^{B,H} \cap V_{n,2}^{B,H} = \varnothing.$$

Indeed, on the event $\{V_{n,1}^{B,H} \cap V_{n,2}^{B,H} = \varnothing\}$, $(G_n, 2)^{B,H}$ and $(\widetilde{G}_n, 2)^{B,H}$ are equal. For $\mathbf{v} \in J_{B,H}$, let $O_{\mathbf{v},2}$ denote the set of offsprings of $\mathbf{v}$ in $(G_n, 2)^{B,H}$. We have

$$I_{\mathbf{v},2} = \{2\} \cup \bigcup_{\mathbf{w} \prec \mathbf{v}} O_{\mathbf{w},2}$$

and

$$\mathbb{P}(V_{n,1}^{B,H} \cap V_{n,2}^{B,H} \neq \varnothing)$$

$$\leq \mathbb{P}(2 \in V_{n,1}^{B,H}) + \sum_{\mathbf{v}=\varnothing}^{B\cdots B} \mathbb{P}(O_{\mathbf{v},2} \cap V_{n,1}^{B,H} \neq \varnothing | V_{n,1}^{B,H} \cap I_{\mathbf{v},2} = \varnothing).$$

For any $\mathbf{u}, \mathbf{v} \in J_{B,H}$, if $V_{n,1}^{B,H} \cap I_{\mathbf{v},2} = \varnothing$, then $\sigma_{n,2}(\mathbf{v})$ is neither the ancestor of $\sigma_{n,1}(\mathbf{u})$, nor an offspring of $\sigma_{n,1}(\mathbf{u})$. From Lemma 2.7 below we deduce that $|\xi_{\sigma_{n,1}(\mathbf{u}),\sigma_{n,2}(\mathbf{v})}^n|$ given $V_{n,1}^{B,H} \cap I_{\mathbf{v},2} = \varnothing$ dominates stochastically $|\xi_{1,2}^n|$, and is independent of the i.i.d. vector $(|\xi_{\sigma_{n,2}(\mathbf{v}),k}^n|)_{k \in \{1,\ldots,n\} \setminus (V_{n,1}^{B,H} \cup I_{\mathbf{v},2})}$, with law $|\xi_{1,2}^n|$. It follows that

$$\mathbb{P}(\sigma_{n,1}(\mathbf{u}) \in O_{\mathbf{v},2} | V_{n,1}^{B,H} \cap I_{\mathbf{v},2} = \varnothing) \le \frac{B}{n - C - |I_{\mathbf{v},2}|}.$$

Therefore,

$$\mathbb{P}(O_{\mathbf{v},2} \cap V_{n,1}^{B,H} \ne \varnothing | V_{n,1}^{B,H} \cap I_{\mathbf{v},2} = \varnothing)$$

$$\le \sum_{\mathbf{u} \in J_{B,H}} \mathbb{P}(\sigma_{n,1}(\mathbf{u}) \in O_{\mathbf{v},2} | V_{n,1}^{B,H} \cap I_{\mathbf{v},2} = \varnothing)$$

$$\le \frac{CB}{n - 2C}.$$

Finally,

$$\mathbb{P}(V_{n,1}^{B,H} \cap V_{n,2}^{B,H} \ne \varnothing) \le \frac{C}{n} + \frac{C^2 B}{n - 2C},$$

which converges to 0 as $n \to \infty$. $\quad\square$

We have used the following stochastic domination lemma. For any B, H and n let $\mathcal{E}_n^{H,B}$ denote the (random) set of edges $\{i, j\}$ of the complete graph on $\{1, \ldots, n\}$, such that $\{\sigma_n^{-1}(i), \sigma_n^{-1}(j)\}$ is not an edge of the finite tree on $J_{B,H}$. By construction, any loop $\{i, i\}$ belongs to $\mathcal{E}_n^{B,H}$. Also, for $\mathbf{u} \ne \varnothing$ on the finite tree, let $g(\mathbf{u})$ denote the parent of $\mathbf{u}$.

LEMMA 2.7 (Stochastic domination). *For any $n \in \mathbb{N}$, and $B, H \in \mathbb{N}$ such that*

$$\frac{B^{H+1} - 1}{B - 1} \le n,$$

the random variables

$$\{|\xi_{i,j}^n|, \{i, j\} \in \mathcal{E}_n^{B,H}\}$$

stochastically dominate i.i.d. random variables with the same law as law $|\xi_{1,2}^n|$. Moreover, for every $\varnothing \ne \mathbf{u} \in J_{B,H}$, the random variables

$$\{|\xi_{\sigma_n(\mathbf{u}),i}^n|, i \in \{1, \ldots, n\} \setminus \sigma_n(g(\mathbf{u}))\},$$

stochastically dominate i.i.d. random variables with the same law as law $|\xi_{1,2}^n|$.

PROOF. The censoring process which deletes the edges that belong to the tree on $J_{B,H}$ has the property that at each step the B lowest absolute values are deleted from some *fresh* (previously unexplored) subset of edge marks. Using this and the fact that the edge marks $\xi_{i,j}^n$ are i.i.d. we see that both claims in the lemma are implied by the following simple statement.

Let $Y_1, \ldots, Y_m$ denote i.i.d. positive random variables. Suppose $m = n_1 + \cdots + n_\ell$, for some positive integers ℓ, $n_1, \ldots, n_\ell$, and partition the m variables in ℓ blocks $I^1, \ldots, I^\ell$ of $n_1, \ldots, n_\ell$ variables each. Fix some nonnegative integers k_j such that $k_j \leq n_j$ and call $q_1^j, \ldots, q_{k_j}^j$, the (random) indexes of the k_j lowest values of the variables in the block I^j (so that $Y_{q_1^1}$ is the lowest of the $Y_1, \ldots, Y_{n_1}$, $Y_{q_2^1}$ is the second lowest of the $Y_1, \ldots, Y_{n_1}$ and so on). Consider the random index sets of the k_j minimal values in the jth block, $J^j := \bigcup_{i=1}^{k_j}\{q_i^j\}$, and set $J = \bigcup_{j=1}^\ell J^j$. If $k_j = 0$ we set $J^j = \varnothing$. Finally, let $\widetilde{Y}$ denote the vector $\{Y_i, i = 1, \ldots, m; i \notin J\}$. Then we claim that $\widetilde{Y}$ stochastically dominates $m - \sum_{j=1}^\ell k_j$ i.i.d. copies of Y_1.

Indeed, the coupling can be constructed as follows. We first extract a realization $y_1, \ldots, y_m$ of the whole vector. Given this we isolate the index sets $J^1, \ldots, J^\ell$ within each block. We then consider two vectors $\mathcal{Z}, \mathcal{V}$ obtained as follows. The vector $\mathcal{Z}_1 = (z_1^1, \ldots, z_{n_1-k_1}^1, z_1^2, \ldots, z_{n_2-k_2}^2, \ldots, z_{n_\ell-k_\ell}^\ell)$ is obtained by extracting the $n_1 - k_1$ values $z_1^1, \ldots, z_{n_1-k_1}^1$ uniformly at random (without replacement) from the values $y_1, \ldots, y_{n_1}$ (in the block I^1), the $n_2 - k_2$ variables $z_1^2, \ldots, z_{n_2-k_2}^2$ in the same way from the values $y_{n_1+1}, \ldots, y_{n_1+n_2}$ (in the block I^2), and so on. On the other hand, the vector $\mathcal{V} = (v_1^1, \ldots, v_{n_1-k_1}^1, v_1^2, \ldots, v_{n_2-k_2}^2, \ldots, v_{n_\ell-k_\ell}^\ell)$ is obtained as follows. For the first block we take $v_i^1, i = 1, \ldots, n_1 - k_1$ equal to z_i^1 whenever an index $i \in I^1 \setminus J^1$ was picked for the vector $z_1^1, \ldots, z_{n_1-k_1}^1$, and we assign the remaining values (if any) through an independent uniform permutation of those variables $y_i, i \in I^1 \setminus J^1$ which were not picked for the vector $z_1^1, \ldots, z_{n_1-k_1}^1$. We repeat this procedure for all other blocks to assign all values of $\mathcal{V}$. By construction, $\mathcal{V} \geq \mathcal{Z}$ coordinate-wise. The conclusion follows from the observation that $\mathcal{Z}$ is distributed like a vector of $m - \sum_{j=1}^\ell k_j$ i.i.d. copies of Y_1, while $\mathcal{V}$ is distributed like our vector $\widetilde{Y}$. $\quad \square$

2.6. *Proof of Theorem* 2.3.

PROOF OF THEOREM 2.3(i). Let $\nu = \nu_{\alpha,\theta}$ be as in (2.11), and let $(\mathcal{T}_\alpha, \varnothing)$ be a realization of the PWIT(ν). The mark on edge $(\mathbf{v}, \mathbf{v}k)$ in $\mathcal{T}_\alpha$ is denoted by $x_{(\mathbf{v}, \mathbf{v}k)}$ or simply $x_{\mathbf{v}k}$. By definition, we have $x_{(\mathbf{v}, \mathbf{w})} = \infty$ if $\mathbf{v}$ and $\mathbf{w}$ are at graph-distance different from 1. In particular, if we set $y_{\mathbf{v}} = \text{sign}(x_{\mathbf{v}})|x_{\mathbf{v}}|^\alpha$, then the point sets $\Xi_{\mathbf{v}} = \{y_{\mathbf{v}k}\}_{k \geq 1}$ are independent Poisson point processes of intensity $\ell_\theta = \theta \mathbb{1}_{\{x>0\}}\,dx + (1-\theta)\mathbb{1}_{\{x<0\}}\,dx$. We may thus build a realization of the operator $\mathbf{T}$ on $\mathcal{T}_\alpha$ [cf. (2.3)]. Let G_n be the complete network on $\{1, \ldots, n\}$ whose mark

on edge (i, j) is $\xi_{i,j}^n := a_n/X_{i,j}$. Next, we apply Proposition 2.6. For all B, H, $(G_n, 1)^{B,H}$ converges weakly to $(\mathcal{T}_\alpha, \varnothing)^{B,H}$. Let $\sigma_n^{B,H}$ be the map σ_n associated with the network $(G_n, 1)^{B,H}$ (see the construction given before Proposition 2.6). From the Skorokhod representation theorem we may assume that $(G_n, 1)^{B,H}$ converges a.s. to $(\mathcal{T}_\alpha, \varnothing)^{B,H}$ for all B, H. Thus we may find sequences B_n, H_n tending to infinity, such that $(B_n^{H_n+1} - 1)/(B_n - 1) \leq n$ and such that for any pair $\mathbf{u}, \mathbf{v} \in \mathbb{N}^f$ we have $\xi_{(\widetilde{\sigma}_n(\mathbf{u}), \widetilde{\sigma}_n(\mathbf{v}))}^n \to x_{(\mathbf{u},\mathbf{v})}$ a.s. as $n \to \infty$, where $\widetilde{\sigma}_n := \sigma_n^{B_n, H_n}$. The map $\widetilde{\sigma}_n$ can be extended to a bijection $\mathbb{N}^f \to \mathbb{N}^f$. It follows that a.s.

$$(2.15) \qquad \langle \delta_{\mathbf{u}}, \widetilde{\sigma}_n^{-1}(a_n^{-1}X)\widetilde{\sigma}_n \delta_{\mathbf{v}} \rangle = \frac{1}{\xi_{(\widetilde{\sigma}_n(\mathbf{u}), \widetilde{\sigma}_n(\mathbf{v}))}^n} \to \frac{1}{x_{(\mathbf{u},\mathbf{v})}} = \langle \delta_{\mathbf{u}}, \mathbf{T}\delta_{\mathbf{v}} \rangle.$$

Fix $\mathbf{v} \in \mathbb{N}^f$, and set $\psi_n^{\mathbf{v}} := \widetilde{\sigma}_n^{-1}(a_n^{-1}X)\widetilde{\sigma}_n \delta_{\mathbf{v}}$. To prove Theorem 2.3(i) it is sufficient to show that $\psi_n^{\mathbf{v}} \to \mathbf{T}\delta_{\mathbf{v}}$ in $L^2(\mathbb{N}^f)$ almost surely as $n \to \infty$, that is,

$$(2.16) \qquad \sum_{\mathbf{u}}(\langle \delta_{\mathbf{u}}, \psi_n^{\mathbf{v}} \rangle - \langle \delta_{\mathbf{u}}, \mathbf{T}\delta_{\mathbf{v}} \rangle)^2 \to 0.$$

Since from (2.15) we know that $\langle \delta_{\mathbf{u}}, \psi_n^{\mathbf{v}} \rangle \to \langle \delta_{\mathbf{u}}, \mathbf{T}\delta_{\mathbf{v}} \rangle$ for every $\mathbf{u}$, the claim follows if we have (almost surely) uniform (in n) square-integrability of $(\langle \delta_{\mathbf{u}}, \psi_n^{\mathbf{v}} \rangle)_{\mathbf{u}}$. This in turn follows from Lemmas 2.7 and 2.4(i). The proof of Theorem 2.3(i) is complete. $\square$

PROOF OF THEOREM 2.3(ii).   We need the following two facts:

$$(2.17) \qquad \lim_{n\to\infty} \frac{\rho_1}{n} = 1 \qquad \text{in probability,}$$

and there exists $\delta > 0$ such that

$$(2.18) \qquad \liminf_{n\to\infty} \min_{1\leq i \leq n} \frac{\rho_i}{n} > \delta \qquad \text{a.s.}$$

Clearly, (2.17) is a law of large numbers and holds actually a.s. (recall that for $\alpha > 1$ we assume the mean of $U_{i,j}$ to be 1). Let us establish the a.s. uniform bound (2.18). For every $\epsilon > 0$, there exists $R > 0$ such that $\mathbb{E}(U_{i,j}\mathbb{1}_{\{U_{i,j}<R\}}) \geq 1 - \epsilon$. If we define $\rho_i^R = \sum_{j=1}^n U_{i,j}\mathbb{1}_{\{U_{i,j}<R\}}$, then

$$\liminf_{n\to\infty} \min_{1\leq i \leq n} \frac{\rho_i}{n} \geq \liminf_{n\to\infty} \min_{1\leq i \leq n} \frac{\rho_i^R}{n}.$$

Therefore (2.18) is implied by the uniform law of large numbers in [11], Lemma 2.2, applied to the bounded variables $U_{i,j}\mathbb{1}_{\{U_{i,j}<R\}}$.

Next, we claim that for all $\mathbf{u} \in \mathbb{N}^f$, in probability

$$(2.19) \qquad \lim_{n\to\infty} \frac{\rho_{\widetilde{\sigma}_n(\mathbf{u})}}{n} = 1.$$

To prove this we first observe that by Lemma 2.7 and (2.17) we have in probability

$$\limsup_{n\to\infty}\frac{(\rho_{\widetilde{\sigma}_n(\mathbf{u})}-U_{\widetilde{\sigma}_n(\mathbf{u}),\widetilde{\sigma}_n(g(\mathbf{u}))})}{n}\le 1.$$

On the other hand $U_{\widetilde{\sigma}_n(\mathbf{u}),\widetilde{\sigma}_n(g(\mathbf{u}))}$ is stochastically dominated by the maximum of n i.i.d. variables with law $U_{i,j}$. The latter converges in distribution on the scale a_n [cf. Lemma 2.4(i)], and we know that $a_n/n \to 0$. It follows that in probability $\limsup_{n\to\infty}\rho_{\widetilde{\sigma}_n(\mathbf{u})}/n \le 1$. Next, we can estimate

$$\rho_{\widetilde{\sigma}_n(\mathbf{u})}\ge\sum_{i\in\{1,\dots,n\}\setminus I_{\mathbf{u}}}U_{\widetilde{\sigma}_n(\mathbf{u}),i}.$$

Now, observe that if $\mathbf{u}\in\mathbb{N}^f$ belongs to generation h, then the set $I_{\mathbf{u}}$ contains at most $O(B_n^h)$ elements, while n is at least of order $B_n^{H_n}$, where B_n, H_n are the sequences used in the proof of Theorem 2.3(i). In particular, it follows that $|I_{\mathbf{u}}|=o(n)$ and therefore (2.13) and (2.17) imply that $\liminf_{n\to\infty}\rho_{\widetilde{\sigma}_n(\mathbf{u})}/n\ge 1$ in probability. This proves (2.19).

Thanks to (2.19), from the Slutsky lemma and the Skorokhod representation theorem, we may also assume that for each $\mathbf{v}\in\mathbb{N}^f$, $\rho_{\widetilde{\sigma}_n(\mathbf{v})}/n$ converges a.s. to 1. We need to show that for each $\mathbf{v}\in\mathbb{N}^f$, (2.16) holds with the new vector $\psi_n^{\mathbf{v}}:=\widetilde{\sigma}_n^{-1}(\kappa_n S)\widetilde{\sigma}_n\delta_{\mathbf{v}}$,

$$\langle\delta_{\mathbf{w}},\psi_n^{\mathbf{v}}\rangle=\kappa_n\frac{U_{\widetilde{\sigma}_n(\mathbf{w}),\widetilde{\sigma}_n(\mathbf{v})}}{\sqrt{\rho_{\widetilde{\sigma}_n(\mathbf{v})}\rho_{\widetilde{\sigma}_n(\mathbf{w})}}}.$$

Thanks to (2.18), $(\langle\delta_{\mathbf{w}},\psi_n^{\mathbf{v}}\rangle)_{\mathbf{w}}$ is uniformly square-integrable [cf. the proof of (2.16)], and all we have to check is that $(\langle\delta_{\mathbf{w}},\psi_n^{\mathbf{v}}\rangle-\langle\delta_{\mathbf{w}},\mathbf{T}\delta_{\mathbf{v}}\rangle)^2\to 0$ for fixed $\mathbf{w}$. Here $\mathbf{T}$ is the operator appearing in the proof of Theorem 2.3(i) above, now with the choice $\theta=1$. We have

$$(\langle\delta_{\mathbf{w}},\psi_n^{\mathbf{v}}\rangle-\langle\delta_{\mathbf{w}},\mathbf{T}\delta_{\mathbf{v}}\rangle)^2$$
$$\le 2\big(a_n^{-1}U_{\widetilde{\sigma}_n(\mathbf{w}),\widetilde{\sigma}_n(\mathbf{v})}\big(1-n/\sqrt{\rho_{\widetilde{\sigma}_n(\mathbf{v})}\rho_{\widetilde{\sigma}_n(\mathbf{w})}}\big)\big)^2$$
$$+2\big(a_n^{-1}U_{\widetilde{\sigma}_n(\mathbf{w}),\widetilde{\sigma}_n(\mathbf{v})}-\langle\delta_{\mathbf{w}},\mathbf{T}\delta_{\mathbf{v}}\rangle\big)^2.$$

The second term above converges to zero as in the proof of point (i). For the first term we use $\rho_{\widetilde{\sigma}_n(\mathbf{v})}/n\to 1$ and $\rho_{\widetilde{\sigma}_n(\mathbf{w})}/n\to 1$. This proves point (ii). $\square$

PROOF OF THEOREM 2.3(iii).   The setting is as in the proof of point (ii) above, but now $\alpha\in(0,1)$. We build the operator $\mathbf{S}$ on the tree $\mathcal{T}_\alpha$ as in (2.6). We need to prove that for any $\mathbf{v}\in\mathbb{N}^f$, a.s.

$$(2.20)\qquad\qquad\sum_{\mathbf{w}}(\langle\delta_{\mathbf{w}},\psi_n^{\mathbf{v}}\rangle-\langle\delta_{\mathbf{w}},\mathbf{S}\delta_{\mathbf{v}}\rangle)^2\to 0,$$

with $\psi_n^{\mathbf{v}} := \widetilde{\sigma}_n^{-1} S \widetilde{\sigma}_n \delta_{\mathbf{v}}$, that is,

$$\langle \delta_{\mathbf{w}}, \psi_n^{\mathbf{v}} \rangle = \frac{U_{\widetilde{\sigma}_n(\mathbf{w}), \widetilde{\sigma}_n(\mathbf{v})}}{\sqrt{\rho_{\widetilde{\sigma}_n(\mathbf{v})} \rho_{\widetilde{\sigma}_n(\mathbf{w})}}}.$$

Let us first show that for any $\mathbf{v}, \mathbf{w} \in \mathbb{N}^f$ we have a.s.

$$(2.21) \qquad \frac{U_{\widetilde{\sigma}_n(\mathbf{w}), \widetilde{\sigma}_n(\mathbf{v})}}{\sqrt{\rho_{\widetilde{\sigma}_n(\mathbf{v})} \rho_{\widetilde{\sigma}_n(\mathbf{w})}}} \to \frac{\langle \delta_{\mathbf{w}}, \mathbf{T} \delta_{\mathbf{v}} \rangle}{\sqrt{\rho(\mathbf{v}) \rho(\mathbf{w})}} = \langle \delta_{\mathbf{w}}, \mathbf{S} \delta_{\mathbf{v}} \rangle.$$

Multiplying and dividing by a_n and using (2.15) with $\theta = 1$, we see that (2.21) holds if

$$(2.22) \qquad a_n^{-1} \rho_{\widetilde{\sigma}_n(\mathbf{v})} \to \rho(\mathbf{v}),$$

almost surely, for every $\mathbf{v} \in \mathbb{N}^f$. In turn, (2.22) can be proved as follows. Let $k \in \mathbb{N}$, and consider the tree with vertex set $J_{k,k}$, obtained as in Proposition 2.6 with $B = H = k$. Since $J_{k,k}$ is a finite set, for any $\mathbf{v}$, (2.15) implies that a.s.

$$a_n^{-1} \sum_{\mathbf{u} \in J_{k,k}} U_{\widetilde{\sigma}_n(\mathbf{v}), \widetilde{\sigma}_n(\mathbf{u})} \to \sum_{\mathbf{u} \in J_{k,k}} x_{\mathbf{v}, \mathbf{u}}^{-1}.$$

By Lemmas 2.7 and 2.4(ii), $\sum_{\mathbf{u} \notin J_{k,k}} a_n^{-1} U_{\widetilde{\sigma}_n(\mathbf{v}), \widetilde{\sigma}_n(\mathbf{u})}$ a.s. converges uniformly (in n) to 0 as k goes to infinity. This proves (2.21) and (2.22).

Once we have (2.21), to conclude the proof it is sufficient to show that a.s.

$$(2.23) \qquad \lim_{k \to \infty} \sup_n \sum_{\mathbf{w} \notin J_{k,k}} (\langle \delta_{\mathbf{w}}, \psi_n^{\mathbf{v}} \rangle)^2 = 0.$$

However, using (2.22) and the simple bound $(\langle \delta_{\mathbf{w}}, \psi_n^{\mathbf{v}} \rangle)^2 \leq \frac{U_{\widetilde{\sigma}_n(\mathbf{v}), \widetilde{\sigma}_n(\mathbf{w})}}{\rho_{\widetilde{\sigma}_n(\mathbf{v})}}$, we have that (2.23) again follows from an application of Lemmas 2.7 and 2.4(ii). This completes the proof of Theorem 2.3(iii). $\square$

2.7. Two-points local operator convergence.

In the proof of the main theorems, we will need a stronger version of Theorem 2.3. Define the $2n \times 2n$ matrices

$$X \oplus X \quad \text{and} \quad S \oplus S,$$

where "$\oplus$" denotes the usual direct sum decomposition, $X \oplus X(\phi_1, \phi_2) = (X\phi_1, X\phi_2)$, for n-dimensional vectors ϕ_1, ϕ_2. As for the limiting operators, we realize them on the Hilbert space $L^2(V) \oplus L^2(V)$ with $V = \mathbb{N}^f$. We consider two independent realizations $\mathcal{T}_\alpha^1, \mathcal{T}_\alpha^2$ of the PWIT(ℓ_θ), and call $\mathbf{T}_1, \mathbf{S}_1, \mathbf{T}_2, \mathbf{S}_2$ the associated operators as in Section 2.3. We may then define

$$\mathbf{T}_1 \oplus \mathbf{T}_2 \quad \text{and} \quad \mathbf{S}_1 \oplus \mathbf{S}_2.$$

By Proposition 2.6, $((G_n, 1))^{B,H}, (G_n, 2)^{B,H})$ converges weakly to $((\mathcal{T}_\alpha^1, \varnothing)^{B,H}, (\mathcal{T}_\alpha^2, \varnothing)^{B,H})$. As before we can view the matrices $X \oplus X$ and $S \oplus S$ as bounded

self-adjoint operators on $L^2(V) \oplus L^2(V)$. Therefore, arguing as in the proof of Theorem 2.3, it follows that, in distribution, for all $(\phi_1, \phi_2) \in \mathcal{D} \times \mathcal{D}$,

$$\sigma_n^{-1} a_n^{-1} X \oplus X \sigma_n (\phi_1, \phi_2) \to \mathbf{T}_1 \oplus \mathbf{T}_2 (\phi_1, \phi_2),$$

where, $\sigma_n = \sigma_n^1 \oplus \sigma_n^2$, and, as above, for $i \in \{1, 2\}$, σ_n^i is a bijection on $\mathbb{N}^f$, extension of the injective indexing map from $\mathbb{N}^f$ to $\{1, \ldots, n\}$, such that $\sigma_n^i(\varnothing) = i$. Analogous convergence results hold for the matrix $S \oplus S$. We can thus extend the statement of Theorem 2.3 to the following local convergence of operators in $L^2(V) \oplus L^2(V)$. To avoid lengthy repetitions we omit the details of the proof.

THEOREM 2.8.   *As n goes to infinity, in distribution*:

 (i) *if $\alpha \in (0, 2)$, then $(a_n^{-1} X \oplus a_n^{-1} X, (1, 2)) \to (\mathbf{T}_1 \oplus \mathbf{T}_2, (\varnothing, \varnothing))$;*
 (ii) *if $\alpha \in (1, 2)$ and $\theta = 1$, then $(\kappa_n S \oplus \kappa_n S, (1, 2)) \to (\mathbf{T}_1 \oplus \mathbf{T}_2, (\varnothing, \varnothing))$;*
(iii) *if $\alpha \in (0, 1)$, then $(S \oplus S, (1, 2)) \to (\mathbf{S}_1 \oplus \mathbf{S}_2, (\varnothing, \varnothing))$.*

3. Convergence of the empirical spectral distributions.

3.1. *Markov matrix, $\alpha \in (0, 1)$: Proof of Theorem* 1.4.   Recall that $\mathbf{S}$ is a bounded self-adjoint operator on $L^2(V)$, whose spectrum is contained in $[-1, 1]$ [cf. (2.7)]. The resolvents of S and $\mathbf{S}$ are the functions on $\mathbb{C}_+ = \{z \in \mathbb{C} : \Im z > 0\}$:

$$R^{(n)}(z) = (S - zI)^{-1} \quad \text{and} \quad R(z) = (\mathbf{S} - zI)^{-1}.$$

For $\ell \in \mathbb{N}$, set

(3.1) $$\mathbf{p}_\ell := \langle \delta_\varnothing, \mathbf{S}^\ell \delta_\varnothing \rangle.$$

Note that $\mathbf{p}_\ell = \frac{1}{\rho(\varnothing)} \langle \delta_\varnothing, \mathbf{K}^\ell \delta_\varnothing \rangle_\rho$ is the probability that the random walk on the PWIT associated with the stochastic operator $\mathbf{K}$ comes back to the root (where it started) after ℓ steps. In particular, $\mathbf{p}_\ell = 0$ for ℓ odd. We set $\mathbf{p}_0 = 1$. Let $\mu_\varnothing$ denote the spectral measure of $\mathbf{S}$ associated with $\delta_\varnothing$ (see e.g., [23], Chapter VII). Equivalently, $\mu_\varnothing$ is the spectral measure of $\mathbf{K}$ associated with the $L^2(V, \rho)$ normalized vector $\hat{\delta}_\varnothing := \delta_\varnothing / \sqrt{\rho(\varnothing)}$ [cf. (2.7)]. In particular, $\mu_\varnothing$ is a probability measure supported on $[-1, 1]$ and such that $\mathbf{p}_\ell = \int_{-1}^1 x^\ell \mu_\varnothing(dx)$, for every ℓ. Since all odd moments vanish $\mu_\varnothing$ is symmetric. Moreover, for any $z \in \mathbb{C}_+$ we have

$$\langle \delta_\varnothing, R(z) \delta_\varnothing \rangle = \int_{-1}^1 \frac{\mu_\varnothing(dx)}{x - z},$$

that is, $\langle \delta_\varnothing, R(z) \delta_\varnothing \rangle$ is the Cauchy–Stieltjes transform of $\mu_\varnothing$. Recall that the Cauchy–Stieltjes transform of a probability measure μ on $\mathbb{R}$ is the analytic function on $\mathbb{C}_+$ given by

$$m_\mu(z) = \int_{\mathbb{R}} \frac{\mu(dx)}{x - z}.$$

The function m_μ characterizes the measure μ, $|m_\mu(z)| \le (\Im z)^{-1}$, and weak convergence of μ_n to μ is equivalent to the convergence $m_{\mu_n}(z) \to m_\mu(z)$ for all $z \in \mathbb{C}_+$. By construction

$$\frac{1}{n} \operatorname{tr} R^{(n)}(z) = \int_{-1}^{1} \frac{\mu_K(dx)}{x - z} = m_{\mu_K}(z),$$

where μ_K is the ESD of K, which coincides with the ESD of S. Using exchangeability and linearity, we get

$$\mathbb{E} R_{1,1}^{(n)}(z) = \mathbb{E} m_{\mu_K}(z) = m_{\mathbb{E}\mu_K}(z).$$

Since $R^{(n)}(z)_{1,1} \le (\Im z)^{-1}$ is bounded, we may apply Theorems 2.2 and 2.3, and obtain, for all $z \in \mathbb{C}_+$,

(3.2)
$$\lim_{n \to \infty} m_{\mathbb{E}\mu_K}(z) = m_{\mathbb{E}\mu_\varnothing}(z).$$

We define

$$\widetilde{\mu}_\alpha = \mathbb{E}\mu_\varnothing.$$

Next, we shall prove that, for all $z \in \mathbb{C}_+$,

(3.3)
$$\lim_{n \to \infty} \mathbb{E}|m_{\mu_K}(z) - m_{\mathbb{E}\mu_\varnothing}(z)| = 0.$$

We have

$$\mathbb{E}|m_{\mu_K}(z) - m_{\mathbb{E}\mu_\varnothing}(z)| \le \mathbb{E}|m_{\mu_K}(z) - \mathbb{E}m_{\mu_K}(z)| + |m_{\mathbb{E}\mu_K}(z) - m_{\mathbb{E}\mu_\varnothing}(z)|.$$

On the right-hand side, the second term converges to 0 by (3.2). The first term is equal to

$$\mathbb{E}\left| \frac{1}{n} \sum_{k=1}^{n} [R_{k,k}^{(n)}(z) - \mathbb{E}R_{k,k}^{(n)}(z)] \right|.$$

By exchangeability, we note that

$$\mathbb{E}\left[\left(\frac{1}{n} \sum_{k=1}^{n} [R_{k,k}^{(n)}(z) - \mathbb{E}R_{k,k}^{(n)}(z)] \right)^2 \right]$$

$$= \frac{1}{n} \mathbb{E}(R_{1,1}^{(n)} - \mathbb{E}R_{1,1}^{(n)})^2 + \frac{n(n-1)}{n^2} \mathbb{E}[(R_{1,1}^{(n)} - \mathbb{E}R_{1,1}^{(n)})(R_{2,2}^{(n)} - \mathbb{E}R_{2,2}^{(n)})]$$

$$\le \frac{1}{n(\Im z)^2} + \mathbb{E}[(R_{1,1}^{(n)} - \mathbb{E}R_{1,1}^{(n)})(R_{2,2}^{(n)} - \mathbb{E}R_{2,2}^{(n)})].$$

Theorems 2.2 and 2.8 imply that $(R_{1,1}(z), R_{2,2}(z))$ are asymptotically independent. Since these variables are bounded, they are also asymptotically uncorrelated, and (3.3) follows.

Finally, observe that the sequence of measures μ_K is a.s. tight. Therefore the convergence (3.3) is sufficient to establish a.s. convergence of μ_K to $\widetilde{\mu}_\alpha$. This completes the proof of Theorem 1.4.

3.2. *I.i.d. matrix, $\alpha \in (0, 2)$: Proof of Theorem* 1.2. Set $A_n = a_n^{-1} X$. For $z \in \mathbb{C}_+$, we define the Cauchy–Stieltjes transform,

$$m_{A_n}(z) = \int \frac{d\mu_{A_n}(x)}{x - z} = \frac{1}{n} \sum_{k=1}^{n} R_{k,k}^{(n)}(z),$$

where

$$R^{(n)}(z) = (A_n - zI)^{-1},$$

is the resolvent of A_n. By exchangeability, $\mathbb{E}m_{A_n}(z) = \mathbb{E}R_{1,1}^{(n)}(z)$. From Proposition A.2 we know that $\mathbf{T}$ is self-adjoint. Therefore from Theorems 2.2 and 2.3 we infer

(3.4) $\mathbb{E}m_{A_n}(z) \to \mathbb{E}h(z), \qquad h(z) := \langle \delta_{\varnothing}, (\mathbf{T} - zI)^{-1} \delta_{\varnothing} \rangle.$

As in the proof of Theorem 1.4 we may write $\mathbb{E}h(z) = \mathbb{E}m_{\mu_{\varnothing}} = m_{\mathbb{E}\mu_{\varnothing}}$, that is the Cauchy–Stieltjes transform of the expected value of the random spectral measure $\mu_{\varnothing}$ associated to $\mathbf{T}$ at the root vector $\delta_{\varnothing}$. From (3.4) we obtain the weak convergence of $\mathbb{E}\mu_{A_n}$ to $\mu_{\alpha} := \mathbb{E}\mu_{\varnothing}$. To obtain a.s. weak convergence of μ_{A_n} to μ_{α}, from Lemma B.1 it suffices to prove the L^1 convergence of Cauchy–Stieltjes transforms as in (3.3). This in turn is obtained by repeating word by word the argument in the proof of Theorem 1.4.

Thus, we have obtained $\mu_{A_n} \to \mu_{\alpha}$ almost surely. Since the operator $\mathbf{T}$ only depends on the two parameters α and θ, where the latter is defined by (1.7), the LSD μ_{α} might still depend on the parameter θ. However, the fact that μ_{α} is independent of θ follows from Lemma 4.2 below, which implies in particular that the values $m_{\mu_{\alpha}}(it) = \mathbb{E}[h(it)]$, $t > 0$, are uniquely determined by α, and therefore by analyticity, all values $m_{\mu_{\alpha}}(z)$, $z \in \mathbb{C}_+$ are uniquely determined by α. This ends the proof of Theorem 1.2.

We remark that in the proof of Theorem 1.2 one can avoid establishing (3.3) plus almost sure tightness [Lemma B.1(i)] as we do above. Namely, the convergence of expected values $\mathbb{E}\mu_{a_n^{-1} X} \to \mu_{\alpha}$ is sufficient. This follows from an a priori concentration estimate (see [12]). However, we did that piece of extra work here since we need it anyway in the case of Markov matrices, where the mentioned concentration estimate is not available.

3.3. *Markov matrix, $\alpha \in (1, 2)$: Proof of Theorem* 1.3. The proof given above for the matrix $A_n = a_n^{-1} X$ applies without modifications to the new matrix $A_n := \kappa_n S$, where $S_{i,j} = \frac{U_{i,j}}{\sqrt{\rho_i \rho_j}}$. In particular, we use Theorems 2.3(ii), 2.8(ii) and Lemma B.1(ii) to obtain the a.s. weak convergence of μ_{A_n} to $\mu_{\alpha} = \mathbb{E}\mu_{\varnothing}$, where $\mu_{\varnothing}$ is the random spectral measure of $\mathbf{T}$ at the root. This ends the proof of Theorem 1.3.

3.4. *Markov matrix, $\alpha = 1$: Proof of Theorem* 1.5.   Suppose now that $\alpha = 1$ and set $w_n = \int_0^{a_n} x\mathcal{L}(dx)$ and $\kappa_n = na_n^{-1}w_n$. A close inspection of the proof of Theorem 2.3(ii) and Theorem 1.3 reveals that all arguments used for $\alpha \in (1,2)$ can be applied to the case $\alpha = 1$ without modifications except for the two estimates (2.17) and (2.18), which have to be replaced by (3.5) and (3.6) below, respectively. For (3.6) we shall use the hypothesis (1.10) on w_n. Let us start by proving that, in probability

$$(3.5) \qquad \lim_{n\to\infty} \frac{\rho_1}{nw_n} = 1.$$

We recall that, for fixed i, $a_n^{-1}(\rho_i - nw_n)$ converges in distribution to a 1-stable law (see, e.g., [20], Theorem 1). Therefore it suffices to show that $\kappa_n = a_n^{-1}nw_n \to \infty$. To see this we may argue as follows. Observe that, for any $\varepsilon > 0$

$$\kappa_n = \mathbb{E}\sum_{i=1}^n a_n^{-1}V_i \mathbb{1}_{\{a_n^{-1}V_i \leq 1\}} \geq \mathbb{E}\sum_{i=1}^n a_n^{-1}V_i \mathbb{1}_{\{\varepsilon \leq a_n^{-1}V_i \leq 1\}},$$

where $V_1 \geq V_2 \geq \cdots$ are the ranked values of $U_{1,j}$, $j = 1, \ldots, n$. From Lemma 2.4(i) the right-hand side above, for any $\varepsilon > 0$, converges to $\mathbb{E}\sum_i x_i \mathbb{1}_{\{\varepsilon \leq x_i \leq 1\}}$, where the x_i are distributed according to the PPP with intensity $x^{-2}dx$ on $(0,\infty)$. While this sum is finite for every $\varepsilon > 0$ it is easily seen to diverge (logarithmically) for $\varepsilon \to 0$. This achieves the proof of (3.5).

Next, we claim that if w_n satisfies (1.10), then there exists $\delta > 0$ such that, a.s.

$$(3.6) \qquad \liminf_{n\to\infty} \min_{1\leq i\leq n} \frac{\rho_i}{nw_n} \geq \delta.$$

To establish (3.6), let us define $b_n = a_{\lfloor n^\varepsilon \rfloor}$ so that $\mathbb{E}(U_{1,i}\mathbb{1}_{\{U_{1,i}\leq b_n\}}) = w_{\lfloor n^\varepsilon \rfloor}$ and

$$\rho_1 \geq S_n := \sum_{i=1}^n U_{1,i}\mathbb{1}_{\{U_{1,i}\leq b_n\}}.$$

From the union bound,

$$\mathbb{P}\left(\min_{1\leq i\leq n} \frac{\rho_i}{nw_n} < \delta\right) \leq n\mathbb{P}\left(\frac{\rho_1}{nw_n} < \delta\right).$$

From the Borel–Cantelli lemma, it is thus sufficient to prove that for some $\delta > 0$

$$(3.7) \qquad \sum_{n\geq 1} n\mathbb{P}(S_n < \delta nw_n) < \infty.$$

By assumption, there exists $\delta > 0$ such that for all n large enough, $w_{\lfloor n^\varepsilon \rfloor} \geq 2\delta w_n$. We define

$$V_i = U_{i,1}\mathbb{1}_{\{U_{1,i}\leq b_n\}} - w_{\lfloor n^\varepsilon \rfloor} \quad \text{and} \quad \overline{S}_n = \sum_{i=1}^n V_i.$$

Note that $\mathbb{E}V_i = \mathbb{E}\overline{S}_n = 0$. We get for all n large enough

$$(3.8) \qquad \mathbb{P}(S_n < \delta n w_n) = \mathbb{P}(\overline{S}_n < \delta n w_n - n w_{\lfloor n^\varepsilon \rfloor}) \leq \mathbb{P}(\overline{S}_n < -\delta n w_n).$$

By construction, w_n is slowly varying and $a_n = L(n)n$ where $L(n)$ is slowly varying. Hence $|V_i| \leq \max(w_{\lfloor n^\varepsilon \rfloor}, b_n) = L(n)n^\varepsilon$ where $L(n)$ is another slowly varying sequence. By the Hoeffding inequality, we get from (3.8)

$$\mathbb{P}(\overline{S}_n < -\delta n w_n) \leq \exp\left(-\frac{\delta^2 n^2 w_n^2}{n L(n)^2 n^{2\varepsilon}}\right) = \exp(-\widetilde{L}(n)n^{1-2\varepsilon}),$$

where $\widetilde{L}(n)$ is a slowly varying sequence. Since $\varepsilon < 1/2$ we obtain (3.7) and thus (3.6).

4. Properties of the limiting spectral distributions.

Recall that μ_α is characterized by the Cauchy–Stieltjes transform $m_{\mu_\alpha}(z) = \mathbb{E}h(z)$, $z \in \mathbb{C}_+$, where $h(z)$ is the random variable $h(z) = \langle \delta_\varnothing, (\mathbf{T} - zI)^{-1}\delta_\varnothing \rangle$ [cf. (3.4)]. The main novelty in our analysis of the LSD μ_α with respect to previous works [5, 7] is that we can work here with the distribution of $h(z)$ rather than only with its expectation.

4.1. *Recursive distributional equation.* The symbol $\stackrel{d}{=}$ stands for equality in distribution. The following result is at the heart of our analysis of the LSD μ_α.

THEOREM 4.1 (Recursive distributional equation). *For all $z \in \mathbb{C}_+$, the random variable*

$$h(z) = \langle \delta_\varnothing, (\mathbf{T} - zI)^{-1}\delta_\varnothing \rangle$$

satisfies to $h(-\bar{z}) = -\bar{h}(z)$ and

$$(4.1) \qquad h(z) \stackrel{d}{=} -\left(z + \sum_{k \in \mathbb{N}} \xi_k h_k(z)\right)^{-1},$$

where $(h_k)_{k \in \mathbb{N}}(z)$ are i.i.d. with the same law of $h(z)$, and $\{\xi_k\}_{k \in \mathbb{N}}$ is an independent Poisson point process with intensity $\frac{\alpha}{2}x^{-\alpha/2-1}\,dx$ on $(0, \infty)$.

PROOF. Since the PWIT is bipartite, the property $h(-\bar{z}) = -\bar{h}(z)$ is a consequence of Lemma A.1. We are left with the RDE (4.1). This can be interpreted as an operator version of the Schur complement formula (see, e.g., Proposition 2.1 in Klein [19] for a similar argument). Denote, as usual, by $k \in \mathbb{N}$ the descendants of the root $\varnothing$, and let $\mathcal{T}^{(k)}$ denote the subtree rooted at k (the set of vertices of $\mathcal{T}^{(k)}$ is then $k\mathbb{N}^f$). We have the direct sum decomposition $\mathbb{N}^f = \{\varnothing\} \cup \bigcup_k k\mathbb{N}^f$. We define $\mathbf{T}^{(k)}$ as the projection of $\mathbf{T}$ on $k\mathbb{N}^f$. Its skeleton is thus $\mathcal{T}^{(k)}$. Finally, define the operator $\mathbf{U}$ on $\mathcal{D}$ by its matrix elements

$$u_k := \langle \delta_\varnothing, \mathbf{U}\delta_k \rangle = \langle \delta_k, \mathbf{U}\delta_\varnothing \rangle = \langle \delta_\varnothing, \mathbf{T}\delta_k \rangle.$$

for all $k \in \mathbb{N}$ (offsprings of $\varnothing$) and $\langle \delta_{\mathbf{u}}, \mathbf{U}\delta_{\mathbf{v}} \rangle = 0$ otherwise. In this way we have

$$\mathbf{T} = \mathbf{U} + \widetilde{\mathbf{T}} \qquad \text{with } \widetilde{\mathbf{T}} = \bigoplus_{k \in \mathbb{N}} \mathbf{T}^{(k)}.$$

As $\mathbf{T}$, each $\mathbf{T}^{(k)}$ can be extended to a self-adjoint operator, which we denote again by $\mathbf{T}^{(k)}$. Therefore $\widetilde{\mathbf{T}}$ is self-adjoint. We shall write $R(z) = (\mathbf{T} - zI)^{-1}$ and $\widetilde{R}(z) = (\widetilde{\mathbf{T}} - zI)^{-1}$ for the associated resolvents, $z \in \mathbb{C}_+$. These operators satisfy the resolvent identity

$$(4.2) \qquad \widetilde{R}(z)(\mathbf{T} - \widetilde{\mathbf{T}})R(z) = \widetilde{R}(z) - R(z).$$

Set $\widetilde{R}_{\mathbf{u},\mathbf{v}}(z) := \langle \delta_{\mathbf{u}}, \widetilde{R}(z)\delta_{\mathbf{v}} \rangle$ and $R_{\mathbf{u},\mathbf{v}}(z) := \langle \delta_{\mathbf{u}}, R(z)\delta_{\mathbf{v}} \rangle$. Observe that $\widetilde{R}_{\varnothing,\varnothing}(z) = -z^{-1}$ and that the direct sum decomposition $\mathbb{N}^f = \{\varnothing\} \cup \bigcup_k k\mathbb{N}^f$ implies $\widetilde{R}_{k,l}(z) = 0$ for $k \neq l$. Similarly we have that $\widetilde{R}_{\varnothing,k}(z) = 0 = \widetilde{R}_{k,\varnothing}(z)$ for every $k \in \mathbb{N}$. From (4.2) we then obtain, for $k \in \mathbb{N}$,

$$\widetilde{R}_{k,k}(z)u_k R_{\varnothing,\varnothing}(z) = -R_{k,\varnothing}(z).$$

It follows that

$$\langle \delta_\varnothing, \widetilde{R}(z)(\mathbf{T} - \widetilde{\mathbf{T}})R(z)\delta_\varnothing \rangle = \sum_{k \in \mathbb{N}} \widetilde{R}_{\varnothing,\varnothing}(z)u_k R_{k,\varnothing}(z)$$

$$= -\sum_{k \in \mathbb{N}} \widetilde{R}_{\varnothing,\varnothing}(z)\widetilde{R}_{k,k}(z)u_k^2 R_{\varnothing,\varnothing}(z).$$

From (4.2) we then conclude that

$$R_{\varnothing,\varnothing}(z) = \frac{\widetilde{R}_{\varnothing,\varnothing}(z)}{1 - \widetilde{R}_{\varnothing,\varnothing}(z) \sum_{k \in \mathbb{N}} \widetilde{R}_{k,k}(z)u_k^2}.$$

Or, using $\widetilde{R}_{\varnothing,\varnothing}(z) = -z^{-1}$,

$$R_{\varnothing,\varnothing}(z) = -\left(z + \sum_{k \in \mathbb{N}} \widetilde{R}_{k,k}(z)u_k^2 \right)^{-1}.$$

Then (4.1) follows from the recursive construction of the PWIT: $\mathcal{T}^{(k)}$ are i.i.d. with distribution $\mathcal{T}$ and therefore $\widetilde{R}_{k,k}(z)$ are i.i.d. with the same law of $R_{\varnothing,\varnothing}(z)$, for every $z \in \mathbb{C}_+$. $\square$

Concerning the uniqueness of the solution to the RDE (4.1) we can establish the following useful result. For $z = it$, with $t > 0$, the identity, $h(-\bar{z}) = -\bar{h}(z)$ reads $\Re h(it) = 0$. Thus, the equation satisfied by $g(it) = \Im h(it) \geq 0$ is

$$(4.3) \qquad g(it) \overset{d}{=} \left(t + \sum_{k \in \mathbb{N}} \xi_k g_k(it) \right)^{-1}.$$

LEMMA 4.2 (Uniqueness of solution for the RDE). *For each $t > 0$, there exists a unique probability measure L^{it} on $\mathbb{R}_+$, solution of* (4.3).

PROOF. Set $\beta = \alpha/2$. If (Y_k) is an i.i.d. sequence of nonnegative variables, independent of $\{\xi_k\}_{k \in \mathbb{N}}$, such that $\mathbb{E}[Y_1^\beta] < \infty$ then it is well known that

$$\sum_k \xi_k Y_k \overset{d}{=} \sum_k \xi_k (\mathbb{E}[Y_1^\beta])^{1/\beta}$$

(see, e.g., [27], Lemma 6.5.1, or (4.5) below). This implies the unicity for (4.3) provided that the equation satisfied by $\mathbb{E}[g(it)^\beta]$ has a unique solution. Recall the formulas of Laplace transforms, for $y \geq 0$, $\eta > 0$ and $0 < \eta < 1$, respectively,

(4.4)
$$y^{-\eta} = \Gamma(\eta)^{-1} \int_0^\infty x^{\eta - 1} e^{-xy}\, dx \quad \text{and}$$
$$y^\eta = \Gamma(1 - \eta)^{-1} \eta \int_0^\infty x^{-\eta - 1} (1 - e^{-xy})\, dx.$$

From the Lévy–Khinchine formula we deduce that, with $s \geq 0$,

(4.5)
$$\mathbb{E} \exp\left(-s \sum_k \xi_k Y_k\right) = \exp\left(\mathbb{E} \int_0^\infty (e^{-xsY_1} - 1)\beta x^{-\beta - 1}\, dx\right)$$
$$= \exp(-\Gamma(1 - \beta) s^\beta \mathbb{E}[Y_1^\beta]).$$

From (4.3), $\mathbb{E}[g(it)^\beta]$ is the solution of the equation in y:

$$y = \frac{1}{\Gamma(\beta)} \int_0^\infty x^{\beta - 1} e^{-tx} e^{-x^\beta \Gamma(1 - \beta) y}\, dx.$$

The last equation has a unique solution for any $t \geq 0$. Indeed, the function from $\mathbb{R}_+$ to $\mathbb{R}_+$

$$\varphi : y \mapsto \frac{1}{\Gamma(\beta)} \int_0^\infty x^{\beta - 1} e^{-tx} e^{-x^\beta \Gamma(1 - \beta) y}\, dx$$

tends to 0 as $y \to \infty$, and it is decreasing since

$$\varphi'(y) = -\frac{\Gamma(1 - \beta)}{\Gamma(\beta)} \int_0^\infty x^{2\beta - 1} e^{-tx} e^{-x^\beta \Gamma(1 - \beta) y}\, dx.$$

Thus φ has a unique fixed point. $\square$

Before going into the proof of Theorem 1.6, we introduce some notation. Let $\beta = \alpha/2$ as above, and let $\mathcal{K}_\alpha$ denote the set of probability measures on $(0, \infty)$ with finite β moment. We define the map Ψ on probability measures on $\mathbb{R}_+ \cup \{\infty\}$, where $\Psi(Q)$ is the law of

(4.6)
$$Z = \left(\sum_{k \in \mathbb{N}} \xi_k Y_k\right)^{-1},$$

with $(Y_k, k \in \mathbb{N})$ i.i.d. with law Q independent of $\Xi = \{\xi_k\}_{k \in \mathbb{N}}$ a Poisson point process on $\mathbb{R}_+$ of intensity $\beta x^{-\beta-1}\, dx$.

LEMMA 4.3. Ψ *satisfies the following*:

(i) Ψ *is a map from $\mathcal{K}_\alpha$ to $\mathcal{K}_\alpha$. Let $(P_n)_{n \in \mathbb{N}}$ and P in $\mathcal{K}_\alpha$, if $\lim_{n \to \infty} \int x^\beta\, dP_n = \int x^\beta\, dP$ then $\Psi(P_n)$ converges weakly to $\Psi(P)$ and $\lim_{n \to \infty} \int x^\beta\, d\Psi(P_n) = \int x^\beta\, d\Psi(P)$.*

(ii) *The unique fixed point of Ψ in $\mathcal{K}_\alpha$ is the law of $1/S$ where S is the one-sided β-stable law with Laplace transform $\mathbb{E}\exp(-tS) = \exp(-t^\beta \times \sqrt{\Gamma(1+\beta)/\Gamma(1-\beta)})$, $t \geq 0$.*

(iii) $\mathbb{E}S^{-\beta} = (\Gamma(\beta+1)\Gamma(1-\beta))^{-1/2}$.

PROOF.   As in the proof of Lemma 4.2, we get

$$\mathbb{E}Z^\beta = \mathbb{E}\left(\sum_k \xi_k Y_k\right)^{-\beta}$$

$$= \mathbb{E}\frac{1}{\Gamma(\beta)} \int_0^\infty x^{\beta-1} e^{-x\sum_k \xi_k Y_k}\, dx$$

$$= \frac{1}{\Gamma(\beta)} \int_0^\infty x^{\beta-1} e^{-x^\beta \Gamma(1-\beta)\mathbb{E}Y_1^\beta}\, dx$$

$$= \frac{1}{\beta\Gamma(\beta)} \int_0^\infty e^{-s\Gamma(1-\beta)\mathbb{E}Y_1^\beta}\, ds$$

$$= \left(\Gamma(\beta+1)\Gamma(1-\beta)\mathbb{E}Y_1^\beta\right)^{-1},$$

[in the last line we have used the identity $z\Gamma(z) = \Gamma(z+1)$]. Therefore, Ψ is a map from $\mathcal{K}_\alpha$ to $\mathcal{K}_\alpha$. Also as a consequence of (4.5)

$$\mathbb{E}\exp(-tZ^{-1}) = \exp(-t^\beta \Gamma(1-\beta)\mathbb{E}Y_1^\beta).$$

Statement (i) follows from the continuity of the map $x \mapsto 1/x$ in $(0, \infty)$. If Z is a fixed point of Ψ then from the computation above $\mathbb{E}Z^\beta = (\Gamma(\beta+1)\Gamma(1-\beta))^{-1/2}$. Finally, from (4.5) we obtain for all $t \geq 0$,

$$\mathbb{E}\exp(-tZ^{-1}) = \exp(-t^\beta \Gamma(1-\beta)\mathbb{E}Z^\beta) = \exp\left(-t^\beta \sqrt{\frac{\Gamma(1+\beta)}{\Gamma(1-\beta)}}\right). \qquad \square$$

4.2. *Proof of Theorem* 1.6(i).   From Theorem 4.1, for $z \in \mathbb{C}_+$,

$$m_{\mu_\alpha}(z) = \mathbb{E}h(z),$$

where h solves RDE (4.1). Set $f(z) = \Re h(z)$ and $g(z) = \Im h(z)$. For $z = u + iv \in \mathbb{C}_+$, f and g satisfy the RDE

$$f(z) \overset{d}{=} -\frac{u + \sum_k \xi_k f_k(z)}{(u + \sum_k \xi_k f_k(z))^2 + (v + \sum_k \xi_k g_k(z))^2}$$

and

$$g(z) \stackrel{d}{=} \frac{v + \sum_k \xi_k g_k(z)}{(u + \sum_k \xi_k f_k(z))^2 + (v + \sum_k \xi_k g_k(z))^2}.$$

By construction, $0 \leq g(z) \leq 1/v$, thus the law of $g(z)$ is in $\mathcal{K}_\alpha$. If the stochastic domination of P by Q is denoted by $P \leq_{st} Q$, we have

$$(4.7) \qquad g(z) \leq_{st} \left(v + \sum_k \xi_k g_k(z) \right)^{-1} \leq_{st} \left(\sum_k \xi_k g_k(z) \right)^{-1}.$$

[In fact, we also have $|h(z)| \leq_{st} (\sum_k \xi_k g_k(z))^{-1}$.] Using the computation in Lemma 4.3, we obtain $\mathbb{E}g(z)^\beta \leq (\Gamma(\beta + 1)\Gamma(1 - \beta)\mathbb{E}g(z)^\beta)^{-1}$. Thus

$$(4.8) \qquad \mathbb{E}g(z)^\beta \leq \frac{1}{\sqrt{\Gamma(\beta + 1)\Gamma(1 - \beta))}}.$$

Again, the formula $y^{-\eta} = \Gamma(\eta)^{-1} \int_0^\infty x^{\eta-1} e^{-xy} dx$, for $y \geq 0$, $\eta > 0$, gives

$$(4.9) \qquad \mathbb{E}\left[\left(\sum_k \xi_k g_k(z) \right)^{-\eta} \right] = \frac{1}{\Gamma(\eta)} \int_0^\infty x^{\eta-1} e^{-x^\beta \Gamma(1-\beta)\mathbb{E}g(z)^\beta} dx.$$

We now study the weak limit of $g(u + iv)$ when $v \downarrow 0$, $u \in \mathbb{R}$. Equation (4.8) implies tightness, so let $g(u + i0)$ be a weak limit. If this limit is nonzero then $\mathbb{E}g^\beta(u + i0) > 0$, and equations (4.7)–(4.9) imply for all $\eta > 0$ and $u \in \mathbb{R}$,

$$\limsup_{u+iv \colon v \downarrow 0} \mathbb{E}g^\eta(u + iv) < \infty.$$

Since $\mathbb{E}h(z)$ is the Cauchy–Stieltjes transform of μ_α, taking $\eta = 1$, we deduce that μ_α is absolutely continuous (see, e.g., [25], Theorem 11.6).

4.3. *Proof of Theorem* 1.6(ii). In view of [25], Theorem 11.6, it is sufficient to show that

$$(4.10) \qquad \lim_{t \downarrow 0} \mathbb{E}g(it) = \Gamma\left(1 + \frac{1}{\beta}\right) \left(\frac{\Gamma(1 + \beta)}{\Gamma(1 - \beta)} \right)^{1/(2\beta)}.$$

As above, (4.8) implies the tightness of $(g(it), t > 0)$. So let $g(i0)$ be a weak limit. It is in $\mathcal{K}_\alpha$ and, by continuity, $g(i0)$ is solution of the RDE

$$g(i0) \stackrel{d}{=} \left(\sum_k \xi_k g_k(i0) \right)^{-1}.$$

By Lemma 4.3, $g(i0) \stackrel{d}{=} 1/S$, and (4.9) gives

$$\mathbb{E}g(i0) = \int_0^\infty e^{-x^\beta \sqrt{\Gamma(1-\beta)/\Gamma(1+\beta)}} dx = \frac{1}{\beta} \Gamma\left(\frac{1}{\beta}\right) \left(\frac{\Gamma(1 + \beta)}{\Gamma(1 - \beta)} \right)^{1/(2\beta)}.$$

Using the identity $z\Gamma(z) = \Gamma(z + 1)$, we get (4.10).

4.4. *Proof of Theorem* 1.6(iii).   We start with a Tauberian-type theorem for the Cauchy–Stieltjes transform of symmetric probability measures. As usual, let m_μ denote the Cauchy–Stieltjes transform of a symmetric probability measure μ on $\mathbb{R}$. Then, for all $t > 0$, $m_\mu(it) \in i\mathbb{R}_+$ and

$$\Im m_\mu(it) = \int_{-\infty}^{\infty} \frac{t}{t^2 + x^2} \mu(dx) = 2 \int_0^{\infty} \frac{t}{t^2 + x^2} \mu(dx).$$

LEMMA 4.4 (Tauberian-like lemma).   *If L is slowly varying and $0 < \alpha < 2$, the following are equivalent*: *as t goes to $+\infty$*

(4.11) $$\mu((t, \infty)) \sim L(t)t^{-\alpha},$$

(4.12) $$\Im m_\mu(it) - t^{-1} \sim -\Delta(\alpha)L(t)t^{-\alpha-1}$$

with $\Delta(\alpha) = 2\alpha \int_0^{\infty} \frac{x^{1-\alpha}}{1+x^2} dx$.

SKETCH OF PROOF OF LEMMA 4.4.   The proof is an adaptation of the proof of the Karamata's Tauberian theorem in [10], pages 37 and 38. Let $\mathcal{M}$ denote the set of symmetric measures on $\mathbb{R}$ such that $\int_0^{\infty} \min(1, x^2)\mu(dx) < +\infty$. On $\mathcal{M}$, define the transform

$$\mathcal{S}\mu : t \mapsto \int_0^{\infty} \frac{2x^2}{t^2 + x^2} \mu(dx).$$

Note that $\mathcal{S}\mu(t) = 1 - t\Im m_\mu(it) = 1 + itm_\mu(it)$. Recall that the Cauchy–Stieltjes transform characterizes the measure. Thus if for all $t > 0$, $(\mathcal{S}\mu_n(t))_{n\in\mathbb{N}}$ converges to $\mathcal{S}\mu$, then $(\mu_n)_{n\in\mathbb{N}}$ converges to μ over all bounded continuous function with 0 outside the support. Now, assume that (4.12) holds, namely

(4.13) $$\mathcal{S}\mu(t) \sim \Delta(\alpha)L(t)t^{-\alpha}.$$

Since $\lim_{x\to\infty} L(tx)/L(t) = 1$, we deduce that for all $t > 0$, as $x \to \infty$

$$\frac{\mathcal{S}\mu(xt)}{L(x)x^{-\alpha}} \to \Delta(\alpha)t^{-\alpha}.$$

The left-hand side is the $\mathcal{S}$ transform of the measure $\mu_x(dy) = \mu(x\,dy)/(L(x)x^{-\alpha})$ while the right-hand side is the $\mathcal{S}$ transform of $\mu_\infty(dy) = \alpha|y|^{-\alpha-1}\,dy$, thus

$$\frac{\mu((x, \infty))}{L(x)x^{-\alpha}} = \mu_x((1, \infty)) \quad \to \quad \mu_\infty(1, \infty) = 1.$$

We get precisely (4.11). The reciprocal implication can be proved similarly (see [10], pages 37 and 38) [it is straightforward for $L(t) = c$, the case that we will actually use].   □

We now come back to the RDE (4.3) and define $Q(t) = \mathbb{E}[g(it)^{\beta}]$. From (4.3), we have a.s. $tg(it) \leq 1$. Note also, from a.s. $\sum_k \xi_k g_k(it) \leq t^{-1} \sum_k \xi_k$, that a.s. $\lim_{t \to +\infty} tg(it) = 1$. The dominated convergence theorem leads to

$$(4.14) \qquad \lim_{t \to \infty} t^{\beta} Q(t) = 1.$$

Moreover, as already pointed in Lemma 4.2,

$$\sum_k \xi_k g_k(it) \overset{d}{=} Q(t)^{1/\beta} \sum_k \xi_k.$$

We deduce, with $C(t) = (t Q(t)^{1/\beta})^{-1/2}$, that

$$\Im m_{\mu_\alpha}(it) = \mathbb{E} g(it) = \mathbb{E} \frac{t}{t^2 + t Q(t)^{1/\beta} \sum_k \xi_k}$$

$$(4.15) \qquad\qquad = C(t) \mathbb{E} \frac{tC(t)}{(tC(t))^2 + \sum_k \xi_k}$$

$$= C(t) \Im m_{\mathcal{L}(Y)}(iC(t)t),$$

where $\mathcal{L}(Y)$ is the law of

$$Y = \varepsilon \sqrt{\sum_k \xi_k},$$

and ε is independent of $\{\xi_k\}_k$, $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = 1/2$. We have

$$\mathbb{P}(Y > t) = \frac{1}{2} \mathbb{P}\left(\sum_k \xi_k > t^2\right).$$

By (4.5), as $s \downarrow 0$, $\mathbb{E} \exp(-s \sum_k \xi_k) = \exp(-s^{\beta} \Gamma(1-\beta)) \sim 1 - s^{\beta} \Gamma(1-\beta)$. Using [10], Corollary 8.7.1, we obtain $\mathbb{P}(\sum_k \xi_k > t) \sim t^{-\beta}$ and

$$\mathbb{P}(Y > t) \sim \frac{t^{-\alpha}}{2}.$$

By Lemma 4.4, $\Im m_{\mathcal{L}(Y)}(it) - t^{-1} \sim -\frac{t^{-\alpha-1}}{2} \Delta(\alpha)$. Thus by (4.14) and (4.15),

$$\Im m_{\mu_\alpha}(it) - t^{-1} \sim -\frac{t^{-\alpha-1}}{2} \Delta(\alpha).$$

Theorem 1.6(iii) now follows from Lemma 4.4.

REMARK 4.5.   In the proof of Lemma 4.2, we have seen that the distribution of $g(it) = \Im h(it)$ was function of $Q(t) = \mathbb{E}[g^{\beta}(it)]$ which satisfies the equation

$$Q(t) = \frac{1}{\Gamma(\beta)} \int_0^\infty x^{\beta-1} e^{-tx} e^{-x^{\beta} \Gamma(1-\beta) Q(t)} \, dx = f_{\beta}(t, Q(t)).$$

We could push further our investigation at $t = 0$ and compute the derivative of Q at $t = 0$: $Q'(0) = -f_{\beta+1}(0, Q(0)) - \Gamma(1 - \beta) f_{2\beta}(0, Q(0)) Q'(0)$, with $Q(0) = (\Gamma(\beta + 1)\Gamma(1 - \beta))^{-1/2}$. There should be no obstacle for computing by recursion the successive derivatives of $Q(t)$ at $t = 0$. We would then obtain a series expansion of the partition function $\mu_\alpha((-\infty, t))$ in a neighborhood of 0.

4.5. *Proof of Theorem* 1.7: $\tilde{\mu}_\alpha$, $\alpha \in (0, 1)$. As in (3.1), let $\mathbf{p}_\ell$ denote the return probability after ℓ steps starting from the root $\varnothing$, for the random walk on the PWIT with transition kernel $\mathbf{K}$ given by (2.5). In particular, $\gamma_\ell = \mathbb{E}\mathbf{p}_\ell$ is the ℓth moment of the LSD $\tilde{\mu}_\alpha$.

PROOF OF THEOREM 1.7(i). For the first part, we shall show that there exists $\delta > 0$ such that for any $\varepsilon \in (0, 1/2]$ and any n

$$(4.16) \qquad \gamma_{2n} \geq \delta \varepsilon^\alpha (1 - \varepsilon)^{2n}.$$

Theorem 1.7 (i) follows by choosing $\varepsilon = 1/2n$. To prove (4.16) we use the simple bound $\mathbf{p}_{2n} \geq (\mathbf{K}(\varnothing, 1)\mathbf{K}(1, \varnothing))^n$, which states that to come back to the root in $2n$ steps the walk can move to the child with the highest weight, with probability $\mathbf{K}(\varnothing, 1)$, go back to the root, with probability $\mathbf{K}(1, \varnothing)$, and repeat this n times. Taking expectation, it follows that

$$(4.17) \qquad \gamma_{2n} \geq \mathbb{E}[(\mathbf{K}(\varnothing, 1)\mathbf{K}(1, \varnothing))^n].$$

Therefore (4.16) holds if the event

$$A_\varepsilon = \{\mathbf{K}(\varnothing, 1) \geq (1 - \varepsilon) \text{ and } \mathbf{K}(1, \varnothing) \geq (1 - \varepsilon)\}$$

has probability at least $\delta \varepsilon^\alpha$, for some $\delta > 0$ and for any $\varepsilon \in (0, 1/2]$.

Let $(x_i)_i$ denote the realization of the PPP at the root $\varnothing$, that is, $x_1 > x_2 > \cdots$ are the points of a PPP on $(0, \infty)$ with intensity measure $\alpha x^{-\alpha-1} dx$. We set $\phi := \sum_{i=1}^\infty x_i$ and let ϕ' denote an independent copy of ϕ. We can use the representation $\mathbf{K}(\varnothing, 1) = x_1/\phi$ and $\mathbf{K}(1, \varnothing) = x_1/(x_1 + \phi')$. Therefore,

$$\mathbb{P}(A_\varepsilon) = \mathbb{P}(x_1 \geq (1 - \varepsilon)\phi, x_1 \geq (1 - \varepsilon)(x_1 + \phi'))$$

$$= \mathbb{P}\left(x_1 \geq (1 - \varepsilon)\phi, \phi' \leq \frac{\varepsilon x_1}{(1 - \varepsilon)}\right)$$

$$\geq \mathbb{P}(x_1 \geq (1 - \varepsilon)\phi, x_1 \geq \varepsilon^{-1}, \phi' \leq 1).$$

Let $\delta_1 := \mathbb{P}(\phi \leq 1) = \int_0^1 f(t) \, dt > 0$, where $f(t)$ denotes the density of ϕ. The function $f(t)$ can be obtained from its Laplace transform, which is given by the known identity $\mathbb{E}[e^{-u\phi}] = e^{-\Gamma(1-\alpha)u^\alpha}$, $u > 0$ (see [22], Proposition 10, or (4.5) with β replaced by α and $Y_k = 1$). Since ϕ' is independent of (x_i) we obtain

$$\mathbb{P}(A_\varepsilon) \geq \delta_1 \mathbb{P}(x_1 \geq (1 - \varepsilon)\phi, x_1 \geq \varepsilon^{-1}).$$

To estimate the last quantity we observe that if $\tilde{x}$ is a size-biased pick from (x_i), then $x_1 \geq \tilde{x}$. We recall that $\tilde{x}$ is a random variable such that, given the sequence (x_i) the probability that $\tilde{x}$ equals x_i is x_i / ϕ. It is not hard to check (see, e.g., [21], Lemma 2.2) that the random variable $\tilde{x}$ has a probability density on $(0, \infty)$ given by

$$(4.18) \qquad \alpha x^{-\alpha-1} \int_0^\infty f(t) \frac{x}{x+t} \, dt,$$

where $f(t)$ is the density of the variable ϕ. Therefore,

$$\mathbb{P}\big(x_1 \geq (1-\varepsilon)\phi, x_1 \geq \varepsilon^{-1}\big)$$

$$\geq \mathbb{P}\big(\tilde{x} \geq (1-\varepsilon)\phi, \tilde{x} \geq \varepsilon^{-1}\big)$$

$$= \alpha \int_0^\infty dt \, f(t) \int_0^\infty dx \, x^{-\alpha-1} \frac{x}{x+t} \mathbb{1}_{\{x \geq (1-\varepsilon)(x+t)\}} \mathbb{1}_{\{x \geq \varepsilon^{-1}\}}$$

$$\geq \alpha \int_0^1 dt \, f(t) \int_0^\infty dx \, x^{-\alpha-1} (1-\varepsilon) \mathbb{1}_{\{x \geq \varepsilon^{-1}\}}$$

$$= \delta_1 (1-\varepsilon) \varepsilon^\alpha.$$

In conclusion, $\mathbb{P}(A_\varepsilon) \geq \delta_1^2 (1-\varepsilon) \varepsilon^\alpha \geq \frac{1}{2} \delta_1^2 \varepsilon^\alpha$, and the claim (4.16) follows.

It remains to show that $\liminf_{\alpha \nearrow 1} \gamma_2 > 0$. If (x_i), $\tilde{x}$, and ϕ are as above and if ϕ' is independent of the sequence (x_i) and identical in law to the random variable ϕ, then

$$\gamma_2 = \mathbb{E}\bigg[\sum_i \frac{x_i}{\phi} \frac{x_i}{x_i + \phi'}\bigg] = \mathbb{E}\bigg[\frac{\tilde{x}}{\tilde{x} + \phi'}\bigg] = \int_0^\infty \alpha x^{1-\alpha} \bigg(\int_0^\infty \frac{f(t)}{x+t} \, dt\bigg)^2 dx.$$

Now, from the Laplace transform $\mathbb{E}[e^{-u\phi}] = e^{-\Gamma(1-\alpha)u^\alpha}$ we have the identity

$$\int_0^\infty \frac{f(t)}{x+t} \, dt = \int_0^\infty e^{-\Gamma(1-\alpha)u^\alpha - ux} \, du.$$

This gives

$$\gamma_2 = \alpha\Gamma(2-\alpha) \int_0^\infty \int_0^\infty e^{-\Gamma(1-\alpha)(u^\alpha + v^\alpha)} (u+v)^{-2+\alpha} \, du \, dv$$

$$= \frac{\alpha\Gamma(2-\alpha)}{\Gamma(1-\alpha)} \int_0^\infty \int_0^\infty e^{-t^\alpha - s^\alpha} (t+s)^{-2+\alpha} \, ds \, dt.$$

Finally, the desired result follows from the bounds (for absolute constants c_1, $c_2 > 0$)

$$\int_0^\infty \int_0^\infty e^{-t^\alpha - s^\alpha} (t+s)^{-2+\alpha} \, ds \, dt \geq e^{-2} \int_0^1 \int_0^1 (t+s)^{-2+\alpha} \, ds \, dt$$

$$\geq \frac{c_1}{1-\alpha}$$

and

$$\Gamma(1-\alpha) = \int_0^\infty t^{-\alpha} e^{-t} \, dt \le \int_0^1 t^{-\alpha} \, dt + \int_1^\infty e^{-t} \, dt \le \frac{c_2}{1-\alpha}. \qquad \square$$

PROOF OF THEOREM 1.7(ii). It is convenient to make here the dependence over α explicit in all the notation. In particular, for every $\alpha \in (0, 1)$, we denote by $\mathbf{S}_\alpha$ the operator $\mathbf{S}$ given by (2.6). These operators are defined on a common probability space, and are self-adjoint in $L^2(V)$. Moreover, it follows from Section 3.1 that $\widetilde{\mu}_\alpha = \mathbb{E}\mu_{\alpha,\varnothing}$, where $\mu_{\alpha,\varnothing}$ is the spectral measure of $\mathbf{S}_\alpha$ at the vector $\delta_\varnothing$. By the dominated convergence theorem, in order to prove that $\alpha \mapsto \widetilde{\mu}_\alpha$ is continuous in $(0, 1)$, it is sufficient to show that a.s. $\alpha \mapsto \mu_{\alpha,\varnothing}$ is continuous. From [23], Theorem VIII.25(a), it is in turn sufficient to prove that for all $\mathbf{v} \in V$, $\alpha \mapsto \mathbf{S}_\alpha \delta_\mathbf{v}$ is a continuous map from $(0, 1)$ to $L^2(V)$. From (2.6), for all $\mathbf{u} \in V$, the map $\alpha \mapsto \mathbf{S}_\alpha(\mathbf{u}, \mathbf{v})$ is continuous. It thus remains to check the uniform square integrability of $(\mathbf{S}_\alpha(\mathbf{v}, \mathbf{u}))_{\mathbf{u} \in V}$. We start with the upper bound

$$(\mathbf{S}_\alpha(\mathbf{v}, \mathbf{v}k))^2 = \frac{y_{\mathbf{v}k}^{-1/\alpha}}{\rho_\alpha(\mathbf{v})} \frac{y_{\mathbf{v}k}^{-1/\alpha}}{\rho_\alpha(\mathbf{v}k)} \le \frac{y_{\mathbf{v}k}^{-1/\alpha}}{\rho_\alpha(\mathbf{v})}.$$

Then, notice that for all $\alpha \in (0, 1 - \varepsilon)$, one has $y_{\mathbf{v}k}^{-1/\alpha} \le \max(1, y_{\mathbf{v}k}^{-1/(1-\varepsilon)})$ and $\rho_\alpha(\mathbf{v}) \ge \min(1, y_{\mathbf{v}1}^{-1/(1-\varepsilon)})$. We may conclude by recalling that a.s. $\lim_k y_{\mathbf{v}k}/k = 1$ and $y_{\mathbf{v}1} > 0$. $\square$

PROOF OF THEOREM 1.7(iii). As in the proof of Theorem 1.7(ii), we make here the dependence over α explicit in all the notation. It follows from Section 3.1

$$\int x^{2\ell} \widetilde{\mu}_\alpha(dx) = \mathbb{E} \int x^{2\ell} \mu_{\alpha,\varnothing}(dx) = \mathbb{E}\mathbf{p}_{\alpha,2\ell},$$

where the expectation is over the randomness of the PWIT. We introduce for $\mathbf{v} \in V$,

$$V_\alpha(\mathbf{v}) = \left(\frac{y_{\mathbf{v}1}^{-1/\alpha}}{\sum_{k \ge 1} y_{\mathbf{v}k}^{-1/\alpha}}, \frac{y_{\mathbf{v}2}^{-1/\alpha}}{\sum_{k \ge 1} y_{\mathbf{v}k}^{-1/\alpha}}, \dots \right).$$

By construction $V_\alpha(\mathbf{v})$ is a $\mathrm{PD}(\alpha, 0)$ random variable. Thus, by [22], Corollary 18, as $\alpha \downarrow 0$, $V_\alpha(\mathbf{v})$ converge weakly to the deterministic vector $(1, 0, \dots)$. We may thus write

$$\mathbf{K}_\alpha(1, \varnothing) = \frac{y_1^{-1/\alpha}}{y_1^{-1/\alpha} + y_{11}^{-1/\alpha}(1 + \varepsilon_\alpha)},$$

where as α goes to 0, ε_α goes in probability to 0. We define $U = \mathbb{1}_{\{y_{11} > y_1\}}$, so that U is a symmetric Bernoulli, that is, $\mathbb{P}(U = 0) = \mathbb{P}(U = 1) = 1/2$. We have proved that in probability

$$\lim_{\alpha \downarrow 0} \mathbf{K}_\alpha(\varnothing, 1) = 1 \quad \text{and} \quad \lim_{\alpha \downarrow 0} \mathbf{K}_\alpha(1, \varnothing) = U.$$

In particular,

$$\lim_{\alpha \downarrow 0} \int x^{2\ell} \mu_{\alpha,\varnothing}(dx) = U.$$

Since $\mu_{\alpha,\varnothing}$ is symmetric,

$$\lim_{\alpha \downarrow 0} \mu_{\alpha,\varnothing} = \frac{U}{2}\delta_{-1} + (1-U)\delta_0 + \frac{U}{2}\delta_1.$$

Taking expectation, we obtain the claimed statement on $\widetilde{\mu}_\alpha$. $\quad\square$

5. Invariant measure: Proof of Theorem 1.8. We start with a lemma. Let $(X_1, \ldots, X_n)$, $X_1 \geq \cdots \geq X_n$, denote the ranked values of $\rho_1, \ldots, \rho_n$ and recall the notion of convergence in the space $\mathcal{A}$, cf. Section 2.4. We use the notation $b_n := a_{m_n}$, where $m_n = n(n+1)/2$.

LEMMA 5.1. *For any $\alpha \in (0,2)$, the sequence $b_n^{-1}(X_1, X_2, \ldots)$ converges in distribution to $(x_1, x_1, x_2, x_2, \ldots)$, where $x_1 > x_2 > \cdots$ denote the ranked points of the Poisson point process on $(0, \infty)$ with intensity $\alpha x^{-\alpha-1}\, dx$.*

PROOF. There are $m_n = n(n+1)/2$ edges, including self-loops. Let us denote by U_e the weight of edge $e \in \{1, \ldots, m_n\}$. The row sums are given by $\rho_i = \sum_{e : e \ni i} U_e$. We write O_n for the set of off-diagonal edges e, that is, edges of the form $e = \{i, j\}$ with $i \neq j$. Let $U_{e_1} \geq U_{e_2} \geq \cdots$ denote the ranked values of the i.i.d. random vector $(U_e)_{e \in O_n}$. Since there are $m_n - n$ edges in O_n, an application of Lemma 2.4(i) yields convergence in distribution

$$(5.1) \qquad\qquad b_n^{-1}(U_{e_1}, U_{e_2}, \ldots) \xrightarrow[n \to \infty]{d} (x_1, x_2, \ldots).$$

Each $e_i = \{u_i, v_i\} \in O_n$ identifies two row sums ρ_{u_i} and ρ_{v_i}. Set $\Delta_i = \max\{\rho_{u_i} - U_{e_i}, \rho_{v_i} - U_{e_i}\}$. Then, for every $k \in \mathbb{N}$ and $\varepsilon > 0$,

$$(5.2) \qquad\qquad \lim_{n \to \infty} \mathbb{P}\Big(\max_{1 \leq \ell \leq k} \Delta_\ell \geq \varepsilon b_n\Big) = 0.$$

To prove this we use an estimate due to Soshnikov [26]. Let B_n denote the event that there exists no $i \in \{1, \ldots, n\}$ such that

$$\Big\{\rho_i > b_n^{3/4+\alpha/8} \text{ and } \rho_i - \max_j U_{i,j} > b_n^{3/4+\alpha/8}\Big\}.$$

Then, from [26] and [4], Lemma 3, one has

$$(5.3) \qquad\qquad \lim_{n \to \infty} \mathbb{P}(B_n) \to 1.$$

Clearly, on the event B_n, if $\max_{1 \leq \ell \leq k} \Delta_\ell \geq \varepsilon b_n$, then $U_{e_k} \leq b_n^{3/4+\alpha/8}$ which has vanishing probability in the limit by (5.1). This proves (5.2).

For simplicity, we introduce the notation $R_{2\ell-1} = \max\{\rho_{u_\ell}, \rho_{v_\ell}\}$, $R_{2\ell} = \min\{\rho_{u_\ell}, \rho_{v_\ell}\}$. Therefore (5.2) and (5.1) prove that

$$(5.4) \qquad b_n^{-1}(R_1, R_2, R_3, R_4, \ldots) \xrightarrow[n\to\infty]{d} (x_1, x_1, x_2, x_2, \ldots).$$

It remains to show that for every fixed k

$$(5.5) \qquad \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{1\le i\le 2k} \{R_i \ne X_i\} \right) = 0.$$

By construction, we have $X_i \ge R_i$ for $i = 1, 2$. On the event B_n described above, to have $X_1 > R_1$ or $X_2 > R_2$ implies that there exists an edge $e \ne e_1$ such that $U_e \ge U_{e_1} - b_n^{3/4+\alpha/8}$. However, this event has vanishing probability by (5.1) and the fact that $b_n^{\delta-1} \max_i U_{i,i} \to 0$ in probability for all sufficiently small $\delta > 0$ (indeed by Lemma 2.4, $a_n^{-1} \max_i U_{i,i}$ converges weakly to the Fréchet distribution, see first comment after Lemma 2.4). Thanks to (5.3) this shows that $\mathbb{P}(X_1 > R_1$ or $X_2 > R_2) \to 0$. Recursively, the probability of $X_{2i+1} > R_{2i+1}$ or $X_{2i+2} > R_{2i+2}$ on the event $B_n \cap \{X_j = R_j, \forall j = 1, \ldots, 2i\}$ vanishes as $n \to \infty$. Indeed, at each step we have removed a row and a column corresponding to the largest off-diagonal weight and we may repeat the same reasoning as above. This proves (5.5) as required. □

PROOF OF THEOREM 1.8(ii). Let us define $m_n = n(n+1)/2$. Observe that

$$(5.6) \quad \sum_{i=1}^n \rho_i = 2S_n + D_n \qquad \text{where } S_n := \sum_{e\in O_n} U_e \quad \text{and} \quad D_n := \sum_{i=1}^n U_{i,i}.$$

Here, as in the previous proof O_n denotes the set of off-diagonal edges. For $\alpha \in (1,2)$, we have by the weak law of large numbers $S_n/m_n \to 1$ and $D_n/n \to 1$ in probability. Therefore

$$(5.7) \qquad \lim_{n\to\infty} \frac{1}{m_n} \sum_{i=1}^n \rho_i = 2 \qquad \text{in probability.}$$

Theorem 1.8(ii) thus follows directly from Lemma 5.1 and (5.7). The same reasoning applies in the case $\alpha = 1$ replacing the law of large numbers by the statement (3.5) which now gives (5.7) with m_n replaced by $m_n w_{m_n}$. □

PROOF OF THEOREM 1.8(i). If $U_{e_1} \ge U_{e_2} \ge \cdots$ are the ranked values of the i.i.d. random vector $(U_e)_{e\in O_n}$ and S_n is their sum as in (5.6), then by Lemma 2.4(ii), replacing n with m_n, we have

$$(5.8) \qquad \left(\frac{U_{e_1}}{S_n}, \frac{U_{e_2}}{S_n}, \ldots \right) \xrightarrow[n\to\infty]{d} \left(\frac{x_1}{\sum_{i=1}^\infty x_i}, \frac{x_2}{\sum_{i=1}^\infty x_i}, \ldots \right),$$

where $x_1 > x_2 > \cdots$ denote the ranked points of the Poisson point process on $(0, \infty)$ with intensity $\alpha x^{-\alpha-1}$.

Write $X_1, X_2, \ldots$ for the ranked values of row sums as in Lemma 5.1, so that $\widetilde{\rho}_i = X_i/(2S_n + D_n)$, where D_n, S_n are as in (5.6). Let

$$Y_{2\ell-1} = \frac{X_{2\ell-1}}{2S_n + D_n} - \frac{U_{e_\ell}}{2S_n}, \qquad Y_{2\ell} = \frac{X_{2\ell}}{2S_n + D_n} - \frac{U_{e_\ell}}{2S_n}.$$

Thanks to (5.8) it is sufficient to prove that $\mathbb{P}(\max_{1 \le i \le 2k} |Y_i| > \varepsilon) \to 0$, as $n \to \infty$, for any fixed $\varepsilon > 0$ and $k \in \mathbb{N}$. This follows from the argument used in the proof of (5.2) and (5.5). $\quad\square$

APPENDIX A: SELF-ADJOINT OPERATORS ON PWIT

The following classical lemma was used in Section 3. If $\mathbf{S}$ is a self-adjoint operator on $D(\mathbf{S}) \subset L^2(V)$ with V countable, the skeleton of $\mathbf{S}$ is the graph on V obtained by putting an edge between two vertices $(\mathbf{v}, \mathbf{w})$ iff $\langle \delta_{\mathbf{v}}, \mathbf{S}\delta_{\mathbf{w}} \rangle \ne 0$.

LEMMA A.1 (Resolvent of self-adjoint operators on bipartite graphs). *Let* $\mathbf{S}$ *be a self-adjoint operator on* $D(\mathbf{S}) \subset L^2(V)$ *with* V *countable. If the skeleton is a bipartite graph then for* $\mathbf{v} \in V$, $h(z) = \langle \delta_{\mathbf{v}}, (\mathbf{S} - zI)^{-1}\delta_{\mathbf{v}} \rangle$ *satisfies for all* $z \in \mathbb{C}_+$, $h(-\bar{z}) = -\bar{h}(z)$.

PROOF. Assume first that $\mathbf{S}$ is bounded: for all $\mathbf{w} \in V$, $\|\mathbf{S}\delta_{\mathbf{w}}\| \le C$. For $|z| > C$, the series expansion of the resolvent gives

$$h(z) = -\sum_{\ell \ge 0} \frac{\langle \delta_{\mathbf{v}}, \mathbf{S}^\ell \delta_{\mathbf{v}} \rangle}{z^{\ell+1}}.$$

However, since the skeleton is a bipartite graph, all cycles have an even length, and for ℓ odd $\langle \delta_{\mathbf{v}}, \mathbf{S}^\ell \delta_{\mathbf{v}} \rangle = 0$. We deduce that for $|z| > C$, $h(-\bar{z}) = -\bar{h}(z)$. We may then extend to $\mathbb{C}_+$ this last identity by analyticity.

If $\mathbf{S}$ is not bounded, then $\mathbf{S}$ is limit of a sequence of bounded operators, and we conclude by invoking Theorem VIII.25(a) in [23]. $\quad\square$

The arguments of Section 3 were crucially based on the following fact.

PROPOSITION A.2. *The operator* $\mathbf{T}$ *defined by* (2.3) *is essentially self-adjoint.*

To prove the proposition, we start with a deterministic lemma. Let $V = \mathbb{N}^f$ denote the vertex set of the PWIT, and let $\mathcal{D}$ be the space of finitely supported vectors. We write $\mathbf{u} \sim \mathbf{v}$ if $\mathbf{u} = \mathbf{v}k$ or $\mathbf{v} = \mathbf{u}k$ for some $k \in \mathbb{N}$ (i.e., if $\mathbf{u}, \mathbf{v}$ are neighbors) and $\mathbf{u} \nsim \mathbf{v}$ otherwise. Let $A : \mathcal{D} \to L^2(V)$ denote the symmetric linear operator defined by

(A.1) $$\langle \delta_{\mathbf{v}}, A\delta_{\mathbf{w}} \rangle = w_{\mathbf{u}, \mathbf{v}} = \overline{w}_{\mathbf{v}, \mathbf{u}},$$

and such that $w_{\mathbf{u}, \mathbf{v}} = 0$ whenever $\mathbf{u} \nsim \mathbf{v}$.

LEMMA A.3 (Criterion of self-adjointness). *Suppose that there exists a constant $\kappa > 0$ and a sequence of connected finite subsets $(S_n)_{n \geq 1}$ in V, such that $S_n \subset S_{n+1}$, $\bigcup_n S_n = V$, and for every n and $\mathbf{v} \in S_n$,*

$$\sum_{\mathbf{u} \notin S_n : \mathbf{u} \sim \mathbf{v}} |w_{\mathbf{u},\mathbf{v}}|^2 \leq \kappa.$$

Then the operator $\mathbf{A}$ defined by (A.1) is essentially self-adjoint.

PROOF.   It is sufficient to check that the only function $\varphi \in D(\mathbf{A}^*) \subset L^2(V)$ such that

$$\mathbf{A}^* \varphi = \pm i \varphi$$

is $\varphi = 0$ (see, e.g., [23], Theorem VIII.3). A similar argument is used in [13], Proposition 3. We deal with the case $\mathbf{A}^* \varphi = i \varphi$, that is, for all $\mathbf{u} \in V$,

$$i \varphi(\mathbf{u}) = \sum_{\mathbf{v} \sim \mathbf{u}} w_{\mathbf{u},\mathbf{v}} \varphi(\mathbf{v}).$$

Here we use the notation $\varphi(\mathbf{u}) = \langle \delta_u, \varphi \rangle$. Taking conjugate, we also have for all $\mathbf{u} \in V$

$$-i \overline{\varphi}(\mathbf{u}) = \sum_{\mathbf{v} \sim \mathbf{u}} \overline{w}_{\mathbf{u},\mathbf{v}} \overline{\varphi}(\mathbf{v}) = \sum_{\mathbf{v} \sim \mathbf{u}} w_{\mathbf{v},\mathbf{u}} \overline{\varphi}(\mathbf{v}).$$

For any finite set $S \subset V$, we deduce

$$\begin{aligned}
i \sum_{\mathbf{v} \in S} |\varphi(\mathbf{v})|^2 &= \sum_{\mathbf{v} \in S} \overline{\varphi}(\mathbf{v}) (\mathbf{A}^* \varphi)(\mathbf{v}) = \sum_{\mathbf{v} \in S} \overline{\varphi}(\mathbf{v}) \sum_{\mathbf{u} \sim \mathbf{v}} w_{\mathbf{v},\mathbf{u}} \varphi(\mathbf{u}) \\
&= \sum_{\mathbf{u} \in S} \varphi(\mathbf{u}) \sum_{\mathbf{v} \sim \mathbf{u}} w_{\mathbf{v},\mathbf{u}} \overline{\varphi}(\mathbf{v}) + \sum_{\mathbf{v} \in S} \overline{\varphi}(\mathbf{v}) \sum_{\mathbf{u} \sim \mathbf{v} : \mathbf{u} \notin S} w_{\mathbf{v},\mathbf{u}} \varphi(\mathbf{u}) \\
&\quad - \sum_{\mathbf{u} \in S} \varphi(\mathbf{u}) \sum_{\mathbf{v} \sim \mathbf{u} : \mathbf{v} \notin S} w_{\mathbf{v},\mathbf{u}} \overline{\varphi}(\mathbf{v}) \\
&= -i \sum_{\mathbf{u} \in S} |\varphi(\mathbf{u})|^2 + \sum_{\mathbf{v} \in S} \overline{\varphi}(\mathbf{v}) \sum_{\mathbf{u} \sim \mathbf{v} : \mathbf{u} \notin S} w_{\mathbf{v},\mathbf{u}} \varphi(\mathbf{u}) \\
&\quad - \sum_{\mathbf{u} \in S} \varphi(\mathbf{u}) \sum_{\mathbf{v} \sim \mathbf{u} : \mathbf{v} \notin S} w_{\mathbf{v},\mathbf{u}} \overline{\varphi}(\mathbf{v}).
\end{aligned}$$

We obtain a Green formula,

$$2i \sum_{\mathbf{v} \in S} |\varphi(\mathbf{v})|^2 = \sum_{\mathbf{v} \in S} \overline{\varphi}(\mathbf{v}) \sum_{\mathbf{u} \sim \mathbf{v} : \mathbf{u} \notin S} w_{\mathbf{v},\mathbf{u}} \varphi(\mathbf{u}) - \sum_{\mathbf{v} \in S} \varphi(\mathbf{v}) \sum_{\mathbf{u} \sim \mathbf{v} : \mathbf{u} \notin S} \overline{w}_{\mathbf{v},\mathbf{u}} \overline{\varphi}(\mathbf{u}).$$

From Cauchy–Schwarz's inequality,

$$\begin{aligned}
\sum_{\mathbf{v} \in S} |\varphi(\mathbf{v})|^2 &\leq \sum_{\mathbf{v} \in S} |\varphi(\mathbf{v})| \sum_{\mathbf{u} \sim \mathbf{v} : \mathbf{u} \notin S} |w_{\mathbf{v},\mathbf{u}}| |\varphi(\mathbf{u})| \\
&\leq \left(\sum_{\mathbf{v} \in S} |\varphi(\mathbf{v})|^2 \right)^{1/2} \left(\sum_{\mathbf{v} \in S} \left(\sum_{\mathbf{u} \sim \mathbf{v} : \mathbf{u} \notin S} |w_{\mathbf{v},\mathbf{u}}| |\varphi(\mathbf{u})| \right)^2 \right)^{1/2}.
\end{aligned}$$

Now take $S = S_n$. From the assumption of the lemma, using again Cauchy–Schwarz's inequality,

$$\left(\sum_{\mathbf{u} \sim \mathbf{v} \,:\, \mathbf{u} \notin S_n} |w_{\mathbf{v},\mathbf{u}}| |\varphi(\mathbf{u})| \right)^2 \leq \kappa \sum_{\mathbf{u} \sim \mathbf{v} \,:\, \mathbf{u} \notin S_n} |\varphi(\mathbf{u})|^2.$$

Since S_n is connected and the graph is a tree, if $\mathbf{u} \notin S_n$ and $\mathbf{u} \sim \mathbf{v}$ then for any $\mathbf{v}' \in S_n \setminus \mathbf{v}$, then $\mathbf{u} \nsim \mathbf{v}'$. It follows that

$$\sum_{\mathbf{v} \in S_n} |\varphi(\mathbf{v})|^2 \leq \sqrt{\kappa} \left(\sum_{\mathbf{v} \in S_n} |\varphi(\mathbf{v})|^2 \right)^{1/2} \left(\sum_{\mathbf{u} \in S_n^c} |\varphi(\mathbf{u})|^2 \right)^{1/2}.$$

Therefore,

$$\sum_{\mathbf{v} \in S_n} |\varphi(\mathbf{v})|^2 \leq \kappa \sum_{\mathbf{v} \notin S_n} |\varphi(\mathbf{v})|^2.$$

Since $\lim_n S_n = V$, as n grows, the right-hand side goes to 0, while the left-hand side goes to $\|\varphi\|_2^2$. We obtain $\varphi = 0$. $\square$

Next, we need a technical lemma.

LEMMA A.4.   *Let $\kappa > 0$, $0 < \alpha < 2$, and let $0 < x_1 < x_2 < \cdots$ be a Poisson process of intensity 1 on $\mathbb{R}_+$. Define $\tau_\kappa = \inf\{t \in \mathbb{N} : \sum_{k=t+1}^{\infty} x_k^{-2/\alpha} \leq \kappa\}$. Then $\mathbb{E}\tau_\kappa$ is finite and goes to 0 as κ goes to infinity.*

PROOF.   First of all, the fact that τ_κ is a.s. finite follows from the a.s. summability of $\sum_{k=1}^{\infty} x_k^{-2/\alpha}$. We deduce also that a.s. there exists $\kappa > 0$ such that $\tau_\kappa = 0$. From monotone convergence, it remains to check that $\mathbb{E}\tau_\kappa < \infty$. Let $n \geq 1$ and $S_n = \sum_{k=1}^{\infty} x_k^{-2/\alpha} \mathbb{1}_{\{x_k \geq n\}}$. From the Lévy–Khinchin formula, for $\theta > 0$,

$$\mathbb{E}\exp(\theta S_n) = \exp\left(\int_n^{\infty} (e^{\theta x^{-2/\alpha}} - 1)\, dx \right).$$

As n goes to infinity, if $\theta = o(n^{2/\alpha})$,

$$\int_n^{\infty} (e^{\theta x^{-2/\alpha}} - 1)\, dx \sim \frac{\theta}{2/\alpha - 1} n^{-2/\alpha + 1}.$$

Hence, taking $\theta = (2/\alpha - 1) n^{2/\alpha - 1}$, we deduce from the Chernov bound, that for any integer $n \geq n_0$,

$$\mathbb{P}(S_n > \kappa) \leq e^{-\theta \kappa} \mathbb{E}\exp(\theta S_n) \leq 3 e^{-c n^{2/\alpha - 1}},$$

where $n_0 \geq 1$ and $c = (2/\alpha - 1)\kappa$. Also recall (from the Chernov bound) that if N is a Poisson random variable with mean n, then for all $t > 0$,

$$\mathbb{P}(N \geq t) \leq \exp\left(-t \log \frac{t}{ne} \right).$$

Now if the event $\{\tau_\kappa > t\}$ holds, then either the number of points of the Poisson process $(x_k)_{k\geq 1}$ in $[0, n]$ is larger than t or $S_n > \kappa$. We get for any integer $n \geq n_0$,

$$\mathbb{P}(\tau > t) \leq e^{-t \ln(t/(ne))} + 3e^{-cn^{2/\alpha-1}}.$$

We conclude by taking $n = \max(n_0, t/(2e))$. $\square$

PROOF OF PROPOSITION A.2.   We apply Lemma A.3 with $\mathbf{A}$ given by $\mathbf{T}$, the operator defined by (2.3). For $\kappa > 0$ and $\mathbf{v} \in \mathbb{N}^f$, we define the integer

$$\tau_\kappa(\mathbf{v}) = \inf\left\{t \geq 0 : \sum_{k=t+1}^\infty |y_{\mathbf{v}k}|^{-2/\alpha} \leq \kappa\right\}.$$

The variables $(\tau_\kappa(\mathbf{v}))_\mathbf{v}$ are i.i.d., and by Lemma A.4, there exists $\kappa > 0$ such that $\mathbb{E}\tau_\kappa(\mathbf{v}) < 1$. We fix such κ. Next, we give a green color to all vertices $\mathbf{v}$ such that $\tau_\kappa(\mathbf{v}) \geq 1$ and a red color otherwise. We consider an exploration procedure starting from the root which stops at red vertices and goes on at green vertices. More formally, define the sub-forest $\mathcal{T}^g$ of the PWIT where we put an edge between green vertices $\mathbf{v}$ and $\mathbf{v}k$ iff $1 \leq k \leq \tau_\kappa(\mathbf{v})$.

The sets S_n appearing in Lemma A.3 are defined as follows. If the root $\varnothing$ is red, we set $S_1 = \{\varnothing\}$. If the root is green, we consider $T_\varnothing^g$, the maximal subtree of $\mathcal{T}^g$ that contains the root. It is a Galton–Watson tree with offspring distribution $\tau_\kappa(\mathbf{v})$. Thanks to our choice of κ, $T_\varnothing^g$ is almost surely finite. Let $V_\varnothing^g$ denote the set of vertices of $T_\varnothing^g$, and consider the set $L_\varnothing^g$ of the leaves of $T_\varnothing^g$. Note that $L_\varnothing^g$ is the set of vertices $\mathbf{v} \in V_\varnothing^g$ such that for all $1 \leq k \leq \tau_\kappa(\mathbf{v})$, $\mathbf{v}k$ is red. Thus, when the root is green, we set $S_1 = V_\varnothing^g \bigcup_{\mathbf{v}\in L_\varnothing^g}\{\mathbf{v}k : 1 \leq k \leq \tau_\kappa(\mathbf{v})\}$. By construction, the set S_1 satisfies the condition of Lemma A.3.

Next, define the outer boundary of the root as $\{\varnothing\}$ as $\partial\{\varnothing\} = \{1, \ldots, \tau_\kappa(\varnothing)\}$, and for $\mathbf{v} \neq \varnothing$, $\mathbf{v} = (i_1, \ldots, i_k)$, set

$$\partial\{\mathbf{v}\} = \{(i_1, \ldots, i_{k-1}, i_k + 1)\} \cup \{(i_1, \ldots, i_k, 1), \ldots, (i_1, \ldots, i_k, \tau_\kappa(\mathbf{v}))\}.$$

For a finite connected set S, its outer boundary is defined by

$$\partial S = \left(\bigcup_{\mathbf{v}\in S} \partial\{\mathbf{v}\}\right)\Big\backslash S.$$

To define the set S_2, suppose that $\partial S_1 = \{u_1, \ldots, u_n\}$. The above procedure defining S_1 for the PWIT rooted at $\varnothing$ can be now repeated for the subtrees rooted at $u_1, \ldots, u_n$ to obtain sets $S_1(u_1), \ldots, S_1(u_n)$. We can then define $S_2 = S_1 \cup \bigcup_{1\leq i\leq n} S_1(u_i)$. Iterating this procedure, we may thus almost surely define an increasing connected sequence (S_n) of vertices with the properties required in Lemma A.3. $\square$

APPENDIX B: TIGHTNESS ESTIMATES

Let X and K be the matrices defined by (1.6) and (1.1), respectively. Recall that, when $\alpha \geq 1$ we set $\kappa_n = n w_n a_n^{-1}$, where $w_n = 1$ if $\alpha > 1$ and $w_n = \int_0^{a_n} x \mathcal{L}(dx)$ if $\alpha = 1$.

LEMMA B.1.

(i) *For every $\alpha \in (0, 2)$, the sequence $\mu_{a_n^{-1} X}$ is a.s. tight.*
(ii) *For every $\alpha \in [1, 2)$, the sequence $\mu_{\kappa_n K}$ is a.s. tight.*

We first recall a classical lemma on truncated moments and a lemma on the eigenvalues.

LEMMA B.2 (Truncated moments [17], Theorem VIII.9.2). *For every $p > \alpha$,*

$$\mathbb{E}\big[|X_{1,1}|^p \mathbb{1}_{\{|X_{1,1}| \leq t\}}\big] \sim c(p) L(t) t^{p-\alpha},$$

where $c(p) := \alpha/(p - \alpha)$. In particular, $\mathbb{E}[|X_{1,1}|^p \mathbb{1}_{\{|X_{1,1}| \leq a_n\}}] \sim c(p) a_n^p / n$.

LEMMA B.3 (Schatten bound [29], proof of Theorem 3.32). *If A is an $n \times n$ complex Hermitian matrix then for every $0 < r \leq 2$,*

$$(B.1) \qquad \sum_{k=1}^{n} |\lambda_k(A)|^r \leq \sum_{i=1}^{n} \left(\sum_{j=1}^{n} |A_{i,j}|^2 \right)^{r/2}.$$

PROOF OF LEMMA B.1.
Proof of (i). Let us fix $r > 0$. By definition of μ_X we have

$$\int_0^\infty |t|^r \mu_{a_n^{-1} X}(dt) = \frac{1}{n} \sum_{k=1}^{n} |\lambda_k(a_n^{-1} X)|^r.$$

By using (B.1) we get for any $0 \leq r \leq 2$,

$$\int_0^\infty |t|^r \mu_{a_n^{-1} X}(dt) \leq Z_n := \frac{1}{n} \sum_{i=1}^{n} Y_{n,i} \qquad \text{where } Y_{n,i} := \left(\sum_{j=1}^{n} a_n^{-2} |X_{i,j}|^2 \right)^{r/2}.$$

We need to show that $(Z_n)_{n \geq 1}$ is a.s. bounded. Assume for the moment that

$$(B.2) \qquad \qquad \sup_{n \geq 1} \mathbb{E}(Y_{n,1}^4) < \infty$$

for some choice of r. Since $Y_{n,1}, \ldots, Y_{n,n}$ are i.i.d. for every $n \geq 1$, we get from (B.2) that

$$\mathbb{E}((Z_n - \mathbb{E} Z_n)^4) = n^{-4} \mathbb{E}\left(\left(\sum_{i=1}^{n} Y_{n,i} - \mathbb{E} Y_{n,i} \right)^4 \right) = O(n^{-2}).$$

Therefore, by the monotone convergence theorem, we get $\mathbb{E}(\sum_{n\geq 1}(Z_n - \mathbb{E}Z_n)^4) < \infty$, which gives $\sum_{n\geq 1}(Z_n - \mathbb{E}Z_n)^4 < \infty$ a.s. and thus $Z_n - \mathbb{E}Z_n \to 0$ a.s. Now the sequence $(\mathbb{E}Z_n)_{n\geq 1} = (\mathbb{E}Y_{n,1})_{n\geq 1}$ is bounded by (B.2), and it follows that $(Z_n)_{n\geq 1}$ is a.s. bounded.

It remains to show that (B.2) holds, say if $0 < 4r < \alpha$. To this end, let us define

$$S_{n,a,b} := \sum_{j=1}^{n} a_n^{-2}|X_{1,j}|^2 \mathbb{1}_{\{a_n^{-2}|X_{1,j}|^2 \in [a,b)\}} \qquad \text{for every } a < b.$$

Now $Y_{n,1}^4 = (S_{n,0,\infty})^{2r} = (S_{n,0,1} + S_{n,1,\infty})^{2r}$ and thus,

(B.3) $$\mathbb{E}(Y_{n,1}^4) \leq 2^{2r-1}\{\mathbb{E}(S_{n,0,1}^{2r}) + \mathbb{E}(S_{n,1,\infty}^{2r})\}.$$

We have $\sup_n \mathbb{E}(S_{n,0,1}^{2r}) < \infty$. Indeed, since $2r < 1$, from the Jensen inequality,

$$\mathbb{E}(S_{n,0,1}^{2r}) \leq (\mathbb{E}S_{n,0,1})^{2r}$$

and, by Lemma B.2, $\mathbb{E}S_{n,0,1} \sim_n \alpha/(2 - \alpha)$.

To deal with the second term of the right-hand side of (B.3), we define

$$M_n := \max_{1 \leq j \leq n} a_n^{-1}|X_{1,j}|\mathbb{1}_{\{a_n^{-1}|X_{1,j}|>1\}}$$

and

$$N_n := \#\{1 \leq j \leq n \text{ s.t. } a_n^{-1}|X_{1,j}| > 1\}.$$

From the Hölder inequality, if $1/p + 1/q = 1$, we have

(B.4) $$\mathbb{E}(S_{n,1,\infty}^{2r}) \leq \mathbb{E}(N_n^{2r}M_n^{4r}) \leq (\mathbb{E}N_n^{2rp})^{1/p}(\mathbb{E}M_n^{4rq})^{1/q}.$$

Recall that $\mathbb{P}(|X_{1,2}| > a_n) = (1 + o(1))/n \leq 2/n$ for large enough n. Using the union bound, for large enough n,

$$\mathbb{P}(N_n \geq k) \leq \binom{n}{k}\mathbb{P}(|X_{1,2}| > a_n)^k \leq \frac{n^k}{k!}\frac{2^k}{n^k} = \frac{2^k}{k!}.$$

In particular for any $\eta > 0$, $\sup_n \mathbb{E}N_n^\eta < \infty$. Similarly, since L is slowly varying, for large enough n and all $t \geq 1$,

$$\mathbb{P}(M_n \geq t) \leq n\mathbb{P}(|X_{1,2}| > ta_n) = na_n^{-\alpha}t^{-\alpha}L(a_nt) \leq 2t^{-\alpha}.$$

It follows that if $\gamma < \alpha$, $\sup_n \mathbb{E}M_n^\gamma < \infty$. Taking p and q so that $4rq < \alpha$, we thus conclude from (B.4) that $\sup_n \mathbb{E}(S_{n,1,\infty}^{2r}) < \infty$.

Proof of (ii). Recall that for any $\alpha \in [1, 2)$, $\kappa_n = w_n n a_n^{-1}$. Then, by using (B.1) we get for any $0 \leq r \leq 2$,

$$\int_0^\infty |t|^r \mu_{\kappa_n K}(dt) \leq Z_n' := \frac{1}{n}\sum_{i=1}^{n}\left(\frac{nw_n}{\rho_i}\right)^r Y_{n,i},$$

where

$$Y_{n,i} := \left(\sum_{j=1}^{n} a_n^{-2} |X_{i,j}|^2 \right)^{r/2}.$$

From (2.18) (for $1 < \alpha < 2$) and (3.6) (for $\alpha = 1$), there exists $c > 0$ such that a.s.,

$$\limsup_{n \to \infty} \max_{1 \le i \le n} \left(\frac{n w_n}{\rho_i} \right)^r < c.$$

Hence for all n large enough,

$$Z_n' \le \frac{c}{n} \sum_{i=1}^{n} Y_{n,i},$$

and we conclude by using the same argument as in the proof of (i). $\square$

REFERENCES

[1] ALDOUS, D. (1992). Asymptotics in the random assignment problem. *Probab. Theory Related Fields* **93** 507–534. MR1183889

[2] ALDOUS, D. and LYONS, R. (2007). Processes on unimodular random networks. *Electron. J. Probab.* **12** 1454–1508. MR2354165

[3] ALDOUS, D. and STEELE, J. M. (2004). The objective method: Probabilistic combinatorial optimization and local weak convergence. In *Probability on Discrete Structures. Encyclopaedia of Mathematical Sciences* **110** 1–72. Springer, Berlin. MR2023650

[4] AUFFINGER, A., BEN AROUS, G. and PÉCHÉ, S. (2009). Poisson convergence for the largest eigenvalues of heavy tailed random matrices. *Ann. Inst. H. Poincaré Probab. Statist.* **45** 589–610. MR2548495

[5] BELINSCHI, S., DEMBO, A. and GUIONNET, A. (2009). Spectral measure of heavy tailed band and covariance random matrices. *Comm. Math. Phys.* **289** 1023–1055. MR2511659

[6] BEN AROUS, G. and ČERNÝ, J. (2008). The arcsine law as a universal aging scheme for trap models. *Comm. Pure Appl. Math.* **61** 289–329. MR2376843

[7] BEN AROUS, G. and GUIONNET, A. (2008). The spectrum of heavy tailed random matrices. *Comm. Math. Phys.* **278** 715–751. MR2373441

[8] BENJAMINI, I. and SCHRAMM, O. (2001). Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.* **6** 13 pp. (electronic). MR1873300

[9] BERTOIN, J. (2006). *Random Fragmentation and Coagulation Processes. Cambridge Studies in Advanced Mathematics* **102**. Cambridge Univ. Press, Cambridge. MR2253162

[10] BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1989). *Regular Variation. Encyclopedia of Mathematics and Its Applications* **27**. Cambridge Univ. Press, Cambridge. MR1015093

[11] BORDENAVE, C., CAPUTO, P. and CHAFAÏ, D. (2010). Spectrum of large random reversible Markov chains: Two examples. *ALEA Lat. Am. J. Probab. Math. Stat.* **7** 41–64. MR2644041

[12] BORDENAVE, C., CAPUTO, P. and CHAFAÏ, D. (2010). Spectrum of non-Hermitian heavy tailed random matrices. Preprint. Available at http://arxiv.org/abs/1006.1713.

[13] BORDENAVE, C., LELARGE, M. and SALEZ, J. (2011). The rank of diluted random graphs. *Ann. Probab.* **39** 1097–1121.

[14] BOUCHAUD, J. P. (1992). Weak ergodicity breaking and aging in disordered systems. *J. Phys. I France* **2** 1705–1713.

[15] BOUCHAUD, J. P. and CIZEAU, P. (1994). Theory of Lévy matrices. *Phys. Rev. E* **3** 1810–1822.

[16] BOVIER, A. and FAGGIONATO, A. (2005). Spectral characterization of aging: The REM-like trap model. *Ann. Appl. Probab.* **15** 1997–2037. MR2152251

[17] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications. Vol. II*, 2nd ed. Wiley, New York. MR0270403

[18] FONTES, L. R. G. and MATHIEU, P. (2008). *K*-processes, scaling limit and aging for the trap model in the complete graph. *Ann. Probab.* **36** 1322–1358. MR2435851

[19] KLEIN, A. (1998). Extended states in the Anderson model on the Bethe lattice. *Adv. Math.* **133** 163–184. MR1492789

[20] LEPAGE, R., WOODROOFE, M. and ZINN, J. (1981). Convergence to a stable distribution via order statistics. *Ann. Probab.* **9** 624–632. MR0624688

[21] PERMAN, M., PITMAN, J. and YOR, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Probab. Theory Related Fields* **92** 21–39. MR1156448

[22] PITMAN, J. and YOR, M. (1997). The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25** 855–900. MR1434129

[23] REED, M. and SIMON, B. (1980). *Methods of Modern Mathematical Physics. I: Functional Analysis*, 2nd ed. Academic Press, New York. MR0751959

[24] RESNICK, S. I. (2007). *Heavy-tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York. MR2271424

[25] SIMON, B. (2005). *Trace Ideals and Their Applications*, 2nd ed. *Mathematical Surveys and Monographs* **120**. Amer. Math. Soc., Providence, RI. MR2154153

[26] SOSHNIKOV, A. (2004). Poisson statistics for the largest eigenvalues of Wigner random matrices with heavy tails. *Electron. Comm. Probab.* **9** 82–91 (electronic). MR2081462

[27] TALAGRAND, M. (2003). *Spin Glasses: A Challenge for Mathematicians: Cavity and Mean Field Models. Ergebnisse der Mathematik und Ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]* **46**. Springer, Berlin. MR1993891

[28] ZAKHAREVICH, I. (2006). A generalization of Wigner's law. *Comm. Math. Phys.* **268** 403–414. MR2259200

[29] ZHAN, X. (2002). *Matrix Inequalities. Lecture Notes in Math.* **1790**. Springer, Berlin. MR1927396

C. BORDENAVE
IMT UMR5219 CNRS
UNIVERSITÉ DE TOULOUSE III
FRANCE
E-MAIL: charles.bordenave(at)math.univ-toulouse.fr
URL: http://www.math.univ-toulouse.fr/~bordenave/

P. CAPUTO
DIPARTIMENTO DI MATEMATICA
UNIVERSITÀ ROMA TRE
ITALY
E-MAIL: caputo(at)mat.uniroma3.it
URL: http://www.mat.uniroma3.it/users/caputo/

D. CHAFAÏ
LAMA UMR8050 CNRS
UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE
FRANCE
E-MAIL: djalil(at)chafai.net
URL: http://djalil.chafai.net/