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

Charles Bordenave, Pietro Caputo, Djalil Chafai

Spectrum of large random reversible Markov chains – heavy–tailed weights on the complete graph

Charles Bordenave, Pietro Caputo, Djalil Chafaï

Preprint – March 2009

Abstract

We consider the random reversible Markov kernel K on the complete graph with n vertices obtained by putting i.i.d. positive weights of law $\mathcal{L}$ on the $n(n+1)/2$ edges of the graph and normalizing each weight by the corresponding row sum. We have already shown in a previous work that if $\mathcal{L}$ has finite second moment then, as n goes to infinity, the limiting spectral distribution of $n^{1/2}K$ is Wigner's semi–circle law. In the present work, we consider the case where $\mathcal{L}$ belongs to the domain of attraction of an α–stable law, $\alpha \in (0,2)$. When $1 < \alpha < 2$, we show that for a suitable regularly varying sequence κ_n of index $1 - 1/\alpha$, the limiting spectral distribution of $\kappa_n K$ coincides with the one of the random symmetric matrix of the un-normalized weights (i.i.d. entries). In contrast, when $0 < \alpha < 1$, we show that the empirical spectral distribution of K converges, without any rescaling, to a non–trivial law supported on $[-1,1]$, whose moments are the return probabilities of the random walk on a suitable Poisson weighted infinite tree of Aldous. The limiting operator is naturally linked with the Poisson–Dirichlet distribution $\mathrm{PD}(\alpha, 0)$. The "critical" cases $\alpha = 1$ and $\alpha = 2$ are not solved here.

Keywords: random matrices; random walks; reversible Markov chains; random graphs; random environment; spectral analysis of operators; heavy–tailed distributions; α-stable laws; Poisson–Dirichlet laws.

AMS-MSC: 15A52; 60F99; 60G57.

Contents

1 Introduction

Following [11], we consider a finite connected undirected graph $G = (V, E)$ with vertex set V and edge set E. On each edge $\{i, j\} \in E$ we put a positive random weight (or conductance) $U_{i,j} = U_{j,i}$. We assume that the weights $\mathbf{U} = \{U_{i,j}; \{i, j\} \in E\}$ are i.i.d. with common law $\mathcal{L}$. The weighted graph $(G, \mathbf{U})$ is called a *network*. On this network we consider the random walk in random environment with state space V and transition probabilities

$$K_{i,j} = \frac{U_{i,j}}{\rho_i} \quad \text{with} \quad \rho_i = \sum_{j\,:\,\{i,j\}\in E} U_{i,j}. \tag{1}$$

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

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

for all $i, j \in V$. Positivity of the weights implies that K is *irreducible* and ρ is its unique invariant distribution, up to normalization. When the weights are all equal this is just the standard random walk on G, and $K - I$ is the associated Laplacian. If $n = |V|$ then K can be identified to a square $n \times n$ random Markov matrix by putting $K_{i,j} = 0$ if $\{i, j\} \notin E$. In general, the random matrix K has non–independent entries due the normalizing sums ρ_i. This matrix is in general non–symmetric, but by reversibility it is self–adjoint in $L^2(\rho)$ and its spectrum $\sigma(K)$ is real. Moreover, $\sigma(K) \subset [-1, +1]$, and $1 \in \sigma(K)$. We denote by

$$-1 \leqslant \lambda_n(K) \leqslant \cdots \leqslant \lambda_1(K) = 1$$

the eigenvalues of K in ascending order. Since K is irreducible, the eigenspace of the largest eigenvalue 1 is one–dimensional and thus $\lambda_2(K) < 1$. It is also known [24] that K has period 1 or 2 and that this last case is equivalent to $\lambda_n(K) = -1$ (the spectrum is then symmetric).

In Random Matrix Theory, the *global behavior* of the spectrum $\sigma(M)$ of a square $n \times n$ matrix M with eigenvalues $\lambda_1(M), \ldots, \lambda_n(M)$ is studied via the *Empirical Spectral Distribution* (ESD)

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

Since K is Markov, its ESD μ_K encodes a global pathwise information on the random walk on the network $(G, \mathbf{U})$. Namely, for any $\ell \in \mathbb{N}$, if $p_\ell(i)$ denotes the probability that the random walk on $(G, \mathbf{U})$ started at i returns to i after ℓ steps, then the ℓ^{th} moment of μ_K satisfies

$$\int_{-1}^{+1} x^\ell \mu_K(dx) = \frac{1}{n} \mathrm{tr}(K^\ell) = \frac{1}{n} \sum_{i\in V} p_\ell(i). \tag{2}$$

It the present work, we shall restrict to the simple case where G is the *complete graph* on the vertex set V. From now we set $V = \{1, \ldots, n\}$ and $E = \{\{i, j\}\,,\ 1 \leqslant i, j \leqslant n\}$. Note that we have a self–loop with weight $U_{i,i}$ at every vertex i. It is merely for convenience: all results presented here hold if instead $U_{i,i} = 0$ at every vertex i. The law $\mathcal{L}$ is supported on $(0, \infty)$ (with no atom at 0) and does not depend on n.

Our aim is to show that there exists a non–random probability distribution μ which may depend on $\mathcal{L}$, called the *Limiting Spectral Distribution* (LSD), such that $\mu_K \to \mu$ as $n \to \infty$ for some notion of convergence. As in the case of matrices with i.i.d. entries, the asymptotic behavior of μ_K as $n \to \infty$ depends mostly 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 $t \to t\,U_{i,j}$. If $\mathcal{L}$ has a finite second moment we write $\sigma^2 = \int_0^\infty (x-1)^2\,\mathcal{L}(dx)$.

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 $(U_{i,j})_{i,j\geqslant 1}$ of i.i.d. random variables of law $\mathcal{L}$. This gives a meaning to the almost sure (a.s.) convergence of $\mu_{\sqrt{n}K}$.

Theorem 1.1 (Wigner–like behavior)**.** *If $\mathcal{L}$ has finite positive variance $0 < \sigma^2 < \infty$ then a.s.*

$$\frac{1}{n}\sum_{k=1}^{n} \delta_{\sqrt{n}\lambda_k(K)} \xrightarrow[n\to\infty]{w} \mathcal{W}_{2\sigma}\,, \tag{3}$$

where "$\xrightarrow{w}$" denotes weak convergence of measures and $\mathcal{W}_{2\sigma}$ stands for the Wigner semi–circle law on $[-2\sigma, +2\sigma]$ with Lebesgue density

$$x \mapsto \frac{1}{2\pi\sigma^2}\sqrt{4\sigma^2 - x^2}\;\mathbb{1}_{[-2\sigma,+2\sigma]}(x). \tag{4}$$

Note that $\lambda_1(\sqrt{n}K) = \sqrt{n} \to \infty$. Moreover, it is shown in [11] that if $\mathcal{L}$ has finite fourth moment, then the remaining extremal eigenvalues $\lambda_2(\sqrt{n}K)$ and $\lambda_n(\sqrt{n}K)$ converge a.s. to the edge of the limiting support $[-2\sigma, +2\sigma]$. This allows to reinforce the weak convergence of $\mu_{\sqrt{n}K}$ into a convergence of all moments (after removal of the leading eigenvalue $\sqrt{n}$).

The 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, i.e. for all $t > 0$,

$$G(t) := \mathcal{L}((t, \infty)) = L(t)\,t^{-\alpha} \tag{5}$$

where L is a function with slow variation at ∞, i.e. for any $x > 0$,

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

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

$$n\,G(a_n t) \to t^{-\alpha} \quad \text{as} \quad n \to \infty \quad \text{for all } t > 0. \tag{6}$$

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

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

for some function ℓ with slow variation at ∞, see for instance Feller [14, VIII.8 and XVII.5] or Resnick [23, 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 $t \geqslant 1$, and $a_n = n^{1/\alpha}$.

Remarkable works have been devoted recently to random symmetric matrices with i.i.d. heavy–tailed (not necessarily positive) entries. The analysis of the LSD for $\alpha \in (0, 2]$

can be considerably harder than the finite second moment case $\alpha > 2$ (Wigner matrices). The LSD is usually non–explicit and the available results are weaker (in terms of the nature of the convergence). Theorem 1.2 below has been investigated by the physicists Bouchaud and Cizeau [12] and rigorously proved by Ben Arous and Guionnet [7], see also Zakharevich [28] for a different method. Recent work of Belinschi, Dembo, and Guionnet [5] proves a.s. convergence of the LSD.

Theorem 1.2 (Symmetric i.i.d. matrix, $\alpha \in (0,2)$)**.** *Let $X = (X_{i,j})_{1 \leqslant i,j \leqslant n}$ be the symmetric matrix such that $(X_{i,j})_{1 \leqslant i \leqslant j \leqslant n}$ are i.i.d. random variables, $U_{i,j} := |X_{i,j}|$ has law in $\mathbb{H}_\alpha$ with $\alpha \in (0,2)$, and*

$$\theta = \lim_{t \to \infty} \frac{\mathbb{P}(X_{i,j} > t)}{\mathbb{P}(|X_{i,j}| > t)} \in [0,1]\,. \tag{7}$$

If a_n is as in (6), let

$$\mu_{a_n^{-1}X} = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(a_n^{-1}X)}.$$

Then there exists a symmetric law μ_α on $\mathbb{R}$ depending only on α such that, in probability,

$$\mu_{a_n^{-1}X} \xrightarrow[n\to\infty]{w} \mu_\alpha.$$

The "weak convergence in probability" of $\mu_{a_n^{-1}X}$ to μ_α means that for every bounded continuous function $f : \mathbb{R} \to \mathbb{R}$, in probability (or in law since the limit is deterministic),

$$\int_{\mathbb{R}} f(x)\, \mu_{a_n^{-1}X}(dx) \underset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}} f(x)\, \mu_\alpha(dx).$$

It is well known that, for $\alpha \in (0,2)$, a random variable X is in the domain of attraction of an α-stable law iff the law of $|X|$ is in $\mathbb{H}_\alpha$ and the limit (7) exists, cf. [14, Theorem IX.8.1a]. In Section 3.2, we give a new independent proof of Theorem 1.2 and we connect the works of [7] and [28]. The key idea of the proof is to exhibit a limiting (self–adjoint) operator $\mathbf{T}$ for the sequence of matrices $a_n^{-1}X$ on a suitable Hilbert space, and then use well–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 this setting the measure μ_α arises as the expected value of the (random) spectral measure of $\mathbf{T}$ at the root. The PWIT and the limiting operator are defined in Section 2. Our method of proof can be seen as a new variant of the resolvent method, based on local convergence of operators.

Back to our random reversible Markov kernels on the complete graph, constructed via (1) from weights with law $\mathcal{L} \in \mathbb{H}_\alpha$, we obtain very 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 and is a special case of Theorem 1.1. We define the sequence

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

Theorem 1.3 (Reversible Markov matrix, $\alpha \in (1,2)$)**.** *Suppose that $\mathcal{L} \in \mathbb{H}_\alpha$ with $\alpha \in (1,2)$. If $\mu_{\kappa_n K}$ is the ESD of $\kappa_n K$, then, as $n \to \infty$, in probability,*

$$\mu_{\kappa_n K} \xrightarrow[n\to\infty]{w} \mu_\alpha\,,$$

where μ_α is the LSD which appears in the symmetric i.i.d. case (Theorem 1.2).

Theorem 1.4 (Reversible Markov matrix, $\alpha \in (0,1)$)**.** *Suppose that $\mathcal{L} \in \mathbb{H}_\alpha$ with $\alpha \in (0,1)$. If μ_K is the ESD of K then, as $n \to \infty$, in probability,*

$$\mu_K \xrightarrow[n\to\infty]{w} \widetilde{\mu}_\alpha,$$

where $\widetilde{\mu}_\alpha$ is a non–random symmetric law supported on $[-1,1]$ depending only on α.

The proof of Theorem 1.3 and Theorem 1.4 is given in Sections 3.3 and 3.1 respectively. Again the 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. Indeed, 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 best understood by observing 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. Since the support of μ_K is included in $[-1,1]$, Theorem 1.4 implies that for all $\ell \geqslant 1$, in probability,

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

The LSD $\widetilde{\mu}_\alpha$ is obtained in the sequel as the expectation of the (random) spectral measure of $\mathbf{K}$ at the root of the PWIT. It will follow that γ_ℓ (the ℓ^{th} moment of $\widetilde{\mu}_\alpha$) is the expected value of the (random) probability that the random walk in random environment returns to the root in ℓ-steps. In particular, the symmetry of $\widetilde{\mu}_\alpha$ follows from the bipartite nature of the PWIT. Further properties of the measures μ_α and $\widetilde{\mu}_\alpha$ are discussed below.

We believe that the weak convergence in probability provided by theorems 1.3 and 1.4 can be upgraded to an almost sure weak convergence, but this would require some extra input. In the simpler i.i.d. case, the almost sure version of Theorem 1.2 was obtained by Belinschi, Dembo, and Guionnet [5]. Additionally, we believe that Theorem 1.3 remains valid for $\alpha = 1$ (with κ_n replaced by $na_n^{-1}\int_0^{a_n} x\mathcal{L}(dx)$) and for $\alpha = 2$, and that μ_2 is Wigner's semi–circle law. 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, and by Belinschi, Dembo, and Guionnet [5, Remark 1.2 and Lemma 5.2] that μ_α tends to Wigner's semi–circle law as $\alpha \nearrow 2$.

Properties of the LSD

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

Theorem 1.5 (Properties of μ_α)**.** *Let μ_α be the symmetric LSD in Theorems 1.2-1.3.*

(i) μ_α is absolutely continuous on $\mathbb{R}$ with bounded density.

(ii) The density of μ_α at 0 is equal to

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

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

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

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

Theorem 1.6 (Properties of $\widetilde{\mu}_\alpha$)**.** *Let* $\widetilde{\mu}_\alpha$ *be the symmetric LSD in Theorem 1.4.*

(i) For $\alpha \in (0,1)$*, there exists* $\delta > 0$ *such that the moments of* $\widetilde{\mu}_\alpha$ *satisfy*

$$\gamma_{2n} \geqslant \delta\, n^{-\alpha} \quad \text{for all} \quad n \geqslant 1\,.$$

Moreover, $\liminf_{\alpha \nearrow 1} \gamma_2 > 0$ *(in particular,* $\widetilde{\mu}_\alpha \not\to \delta_0$ *as* $\alpha \nearrow 1$*).*

(ii) For the topology of the weak convergence, the mapping $\alpha \mapsto \tilde{\mu}_\alpha$ *is continuous in* $(0,1)$*.*

(iii) For the topology of the weak convergence,

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

It is delicate to provide liable numerical simulations of the ESDs due to the underlying heavy tails. Nevertheless, Figure 5 provides histograms for various values of α and a large value of n, illustrating theorems 1.3 and 1.4.

Invariant measure and edge–behavior

The invariant probability distribution $\hat{\rho}$ of K is obtained by normalizing the 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 $n \max_{1 \leqslant i \leqslant n} |\hat{\rho}_i - n^{-1}| \to 0$ 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. Note that if $\alpha > 1$ then by the strong law of large numbers, for every fixed i we have $n|\hat{\rho}_i - n^{-1}| \to 0$ as $n \to \infty$ a.s. but this almost sure convergence is not uniform over $1 \leqslant i \leqslant n$ when $\alpha \leqslant 2$, and one has to consider a weaker notion of convergence in order to identify the nature of $\hat{\rho}$ as $n \to \infty$.

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 $\widetilde{\rho}_1 \geqslant \widetilde{\rho}_2 \geqslant \cdots$ and their sum is 1. The symbol $\xrightarrow{d}$ denotes convergence in distribution. We refer to Subsection 2.4 for more details on weak convergence in the space of ranked sequences and for the definition of the Poisson–Dirichlet law $\mathrm{PD}(\alpha, 0)$.

Theorem 1.7 (Behavior of the invariant probability measure of K)**.** *Suppose that* $\mathcal{L} \in \mathbb{H}_\alpha$*.*

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

$$\widetilde{\rho} \xrightarrow[n\to\infty]{d} \frac{1}{2}\,(V_1, V_1, V_2, V_2, \ldots)\,, \tag{8}$$

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

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

$$\kappa_{n(n+1)/2}\,\widetilde{\rho} \xrightarrow[n\to\infty]{d} \frac{1}{2}\,(x_1, x_1, x_2, x_2, \dots)\ , \tag{9}$$

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

Theorem 1.7 is proved in Section 5. These results will be derived from the statistics of the ranked values of the weights $U_{i,j}$, $i<j$. The duplication in the sequences in (9) and (8) comes from the fact that each of the largest weights, on the scale $a_{n(n+1)/2}$, belongs to two distinct rows and determines alone the limiting value of the associated row sum.

Theorem 1.7 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 mean–field trap model introduced by Bouchaud [16], see also [6, 13, 15], here the particle 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 [26, 4]. 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 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 ∞

$$1-\lambda_2 = \begin{cases} O(n^{-1/\alpha}) & \alpha \in (0,1) \\ O(n^{-(2-\alpha)/\alpha}) & \alpha \in [1,2) \end{cases} \tag{10}$$

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 seems to be an interesting open problem and we hope to come back to this question in future work.

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 positive 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 $\mathbb{N}^f := \cup_{k\in\mathbb{N}}\mathbb{N}^k$ 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 $(\mathbf{v}1), (\mathbf{v}2), \cdots \in \mathbb{N}^{k+1}$ (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 n^{th} 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 $\Xi_\varnothing = \{y_1, y_2, \dots\}$ be ordered in such a way that $|y_1| \leqslant |y_2| \leqslant \cdots$, 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 $\Xi_{\mathbf{v}} = \{y_{\mathbf{v}1}, y_{\mathbf{v}2}, \dots\}$ satisfy $|y_{\mathbf{v}1}| \leqslant |y_{\mathbf{v}2}| \leqslant \cdots$

Note that $\Xi_{\mathbf{v}}$ has in average $\nu(\mathbb{R}) \in (0, \infty]$ elements. As a convention, if $\nu(\mathbb{R}) < \infty$, one sets the remaining marks equal to ∞. For example, if $\nu = \lambda\delta_1$ is proportional to a Dirac mass, then, neglecting infinite marks, PWIT(ν) is the tree obtained from a Yule Process (with all marks equal to 1). In the sequel we shall only consider cases where ν is not finite and each vertex has a.s. an infinite number of offsprings with finite and distinct marks. If ν is Lebesgue's measure on $[0, \infty)$ we obtain the original PWIT in [1].

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

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

where $\phi, \psi \in \mathbb{C}^V$ and δ_u denotes 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. 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 $z \in \mathbb{C}_+$,*

$$\langle \delta_u, (\mathbf{S}_n - zI)^{-1} \delta_u \rangle \to \langle \delta_v, (\mathbf{S} - zI)^{-1} \delta_v \rangle. \tag{11}$$

Proof of Theorem 2.2. It is a special case of Reed and Simon [22, 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, i.e. $(\widetilde{\mathbf{S}}_n - zI)^{-1}\psi \to (\mathbf{S} - zI)^{-1}\psi$ for any $z \in \mathbb{C}_+$, $\psi \in L^2(V)$. The conclusion follows by taking the scalar product $\langle \delta_v, (\widetilde{\mathbf{S}}_n - zI)^{-1} \delta_v \rangle = \langle \delta_u, (\mathbf{S}_n - zI)^{-1} \delta_u \rangle$. □

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 cases 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 (11) 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 our matrix X or K, 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

$$\sum_k |y_{\mathbf{v}k}|^{-2/\alpha} < \infty, \tag{12}$$

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 $\mathbf{v}, \mathbf{w} \in \mathbb{N}^f$,

$$\mathbf{T}(\mathbf{v},\mathbf{w}) = \langle \delta_{\mathbf{v}}, \mathbf{T}\delta_{\mathbf{w}}\rangle = \begin{cases} \operatorname{sign}(y_{\mathbf{w}})|y_{\mathbf{w}}|^{-1/\alpha} & \text{if } \mathbf{w} = \mathbf{v}k \text{ for some integer } k \\ \operatorname{sign}(y_{\mathbf{v}})|y_{\mathbf{v}}|^{-1/\alpha} & \text{if } \mathbf{v} = \mathbf{w}k \text{ for some integer } k \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

Note that if every edge e in the tree with mark y_e is given the "weight" $\operatorname{sign}(y_e)|y_e|^{-1/\alpha}$ 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 $\mathcal{D} \subset D(\mathbf{T}) \subset L^2(\mathbb{N}^f)$, see for example Reed and Simon [22, VIII.2]). With a slight abuse of notation, we identify $\mathbf{T}$ with its closed extension. As stated below, $\mathbf{T}$ is the natural local limit of the sequence of i.i.d. $n \times n$ matrices $a_n^{-1}X$. 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, 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, in the case of Markov matrices K, for $\alpha \in [1,2)$, $\mathbf{T}$ (now with $\theta = 1$) is the local limit operator of $\kappa_n K$. To work directly with symmetric operators we introduce the symmetric matrix

$$S_{i,j} = \frac{U_{i,j}}{\sqrt{\rho_i \rho_j}}, \tag{14}$$

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 PWIT(ℓ_1), where ℓ_1 is Lebesgue's measure on $[0,\infty)$. We define an operator corresponding to the random walk on this tree with conductance equal to the marks 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 $\mathbf{v}, \mathbf{w} \in \mathbb{N}^f$,

$$\mathbf{K}(\mathbf{v},\mathbf{w}) = \langle \delta_{\mathbf{v}}, \mathbf{K}\delta_{\mathbf{w}} \rangle = \begin{cases} \frac{y_{\mathbf{w}}^{-1/\alpha}}{\rho(\mathbf{v})} & \text{if } \mathbf{w} = \mathbf{v}k \text{ for some integer } k \\ \frac{y_{\mathbf{v}}^{-1/\alpha}}{\rho(\mathbf{v})} & \text{if } \mathbf{v} = \mathbf{w}k \text{ for some integer } k \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

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' inequality implies

$$\begin{aligned} \langle \mathbf{K}\phi, \mathbf{K}\phi \rangle_\rho^2 &= \sum_{\mathbf{u}} \rho(\mathbf{u}) \big| \sum_{\mathbf{v}} \mathbf{K}(\mathbf{u},\mathbf{v}) \phi_{\mathbf{v}} \big|^2 \\ &\leqslant \sum_{\mathbf{u}} \rho(\mathbf{u}) \sum_{\mathbf{v}} \mathbf{K}(\mathbf{u},\mathbf{v}) |\phi_{\mathbf{v}}|^2 = \sum_{\mathbf{v}} \rho(\mathbf{v}) |\phi_{\mathbf{v}}|^2 = \langle \phi, \phi \rangle_\rho^2 \end{aligned}$$

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

$$\mathbf{S}(\mathbf{v},\mathbf{w}) := \sqrt{\frac{\rho(\mathbf{v})}{\rho(\mathbf{w})}}\, \mathbf{K}(\mathbf{v},\mathbf{w}) = \frac{\mathbf{T}(\mathbf{v},\mathbf{w})}{\sqrt{\rho(\mathbf{v})\,\rho(\mathbf{w})}}\,. \tag{16}$$

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

$$\langle \phi, \mathbf{S}\psi \rangle = \langle \mathbf{D}^{-1}\phi, \mathbf{K}\mathbf{D}^{-1}\psi \rangle_\rho\,, \tag{17}$$

for all $\phi, \psi \in L^2(V)$. In this way, when $\alpha \in (0,1)$, $\mathbf{S}$ will be the limiting operator associated to the matrix S defined in (14). 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 $\alpha \in (0,2)$ 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, this in turn gives convergence of the expected values of the Stieltjes transform of the ESD of our matrices. To obtain convergence in probability we will need a little more than the above local convergence, see Theorem 2.8 below. The rest of this section is devoted to the proof of Theorem 2.3.

2.4 Image of a single vector

In this paragraph, we recall some facts about the ordered statistics of the entries of the vector

$$Xe_1 = (X_{1,1}, \dots, X_{1,n}) \quad \text{and} \quad Ke_1 = (U_{1,1}, \dots, U_{1,n})/\rho_1$$

where $\{e_i\}_{1 \leqslant i \leqslant n}$ is the canonical basis of $\mathbb{R}^n$. Let us denote by $V_1 \geqslant V_2 \geqslant \cdots \geqslant V_n$ the order statistics of the variables $U_{1,j} = |X_{1,j}|$ where $1 \leqslant j \leqslant n$. 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 $v_1 \geqslant v_2 \geqslant \cdots \geqslant 0$ 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

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

as element 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, §2.1.3].

Let $E_1, E_2, \dots$ 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 $\mathcal{A}$

$$Y = \left(\Gamma_1^{-\frac{1}{\alpha}}, \Gamma_2^{-\frac{1}{\alpha}}, \dots\right)$$

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 $\mathcal{A}_1$

$$Z = \left(\frac{\Gamma_1^{-\frac{1}{\alpha}}}{\sum_{n=1}^\infty \Gamma_n^{-\frac{1}{\alpha}}}, \frac{\Gamma_2^{-\frac{1}{\alpha}}}{\sum_{n=1}^\infty \Gamma_n^{-\frac{1}{\alpha}}}, \dots\right).$$

For $\alpha \in (0,1)$ the sum $\sum_n \Gamma_n^{-\frac{1}{\alpha}}$ 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 [20, 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 $\alpha \in (0,2)$, $(a_n^{-1}V_j)_{j\geqslant 1}$ is a.s. uniformly square integrable, i.e. $\lim_k \sup_{n>k} a_n^{-2}\,\Delta_{k,n}^2 = 0$.

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

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

As an example, from (i), we retrieve the well–known fact that for any $\alpha > 0$, the random variable $a_n^{-1}\max(U_{1,1}, \dots, U_{1,n})$ converges weakly as $n \to \infty$ to the law of $\Gamma_1^{-\frac{1}{\alpha}}$. This law, known as a Fréchet law on $(0,\infty)$, has density $\alpha x^{-\alpha-1}e^{-x^{-\alpha}}$.

Proof of Lemma 2.4. As in LePage, Woodroofe and Zinn [18] 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 (5) and write

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

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

$$\left(G^{-1}(\Gamma_1/\Gamma_{n+1}), \ldots, G^{-1}(\Gamma_n/\Gamma_{n+1})\right), \tag{18}$$

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

$$Y_n \stackrel{d}{=} \left(\frac{G^{-1}(\Gamma_1/\Gamma_{n+1})}{a_n}, \ldots, \frac{G^{-1}(\Gamma_n/\Gamma_{n+1})}{a_n}\right),$$

which follows from (18). 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 (6), almost surely, for every j:

$$a_n^{-1} G^{-1}(\Gamma_j/\Gamma_{n+1}) \to \Gamma_j^{-\frac{1}{\alpha}}, \tag{19}$$

and the convergence in distribution of Y_n to Y follows. Moreover, from (6), for any $\delta > 0$ we can find n_0 such that

$$a_n^{-1} V_j = a_n^{-1} G^{-1}(\Gamma_j/\Gamma_{n+1}) \leqslant (n\Gamma_j/(1+\delta)\Gamma_{n+1})^{-\frac{1}{\alpha}},$$

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

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 $\Gamma_j^{-\frac{1}{\alpha}}$. To prove the convergence of Z_n we use the distributional identity

$$Z_n \stackrel{d}{=} \left(\frac{G^{-1}(\Gamma_1/\Gamma_{n+1})}{\sum_{j=1}^{n} G^{-1}(\Gamma_j/\Gamma_{n+1})}, \ldots, \frac{G^{-1}(\Gamma_n/\Gamma_{n+1})}{\sum_{j=1}^{n} G^{-1}(\Gamma_j/\Gamma_{n+1})}\right).$$

As a consequence of (19), we then have almost surely

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

and (ii) follows. Finally, (iii) is an easy consequence of the exchangeability of the variable $(U_{1,i})$: $\mathbb{P}(V_k^I \neq V_k) \leqslant \mathbb{P}(\exists j \in I : U_{1,j} \geqslant V_k) \leqslant |I|\mathbb{P}(U_{1,1} \geqslant V_k) = |I|k/n$. □

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 $\xi_1^n, \xi_2^n, \ldots$ be sequences of i.i.d. random variables such that*

$$n\,\mathbb{P}(\xi_1^n \in \cdot) \xrightarrow[n\to\infty]{w} \nu, \tag{20}$$

where ν is a Radon measure on $\mathbb{R}$. Then, for any finite set $I \subset \mathbb{N}$ the random measure $\sum_{i\in\{1,\ldots,n\}\setminus I} \delta_{\xi_i^n}$ 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 Resnick [23, Theorem 5.3, p. 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).

2.5 Local weak convergence

In the above paragraph, we have considered the convergence of the vectors $(X_{1,1}, \cdots, X_{1,n})/a_n$ and $(U_{1,1}, \cdots, U_{1,n})/\rho_1$. In this paragraph, 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. The reader may skip this part and jump directly to Section 3.

Let G_n be the complete network on $\{1, \ldots, n\}$ whose mark on edge (i, j) equals $\xi^n_{i,j}$, for some collection $(\xi^n_{ij})_{1 \leqslant i \leqslant j \leqslant n}$ of i.i.d. random variables with values in $\mathbb{R}$, with $\xi^n_{j,i} = \xi^n_{i,j}$. 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^n_{ij}), and for any $B, H \in \mathbb{N}$, such that $(B^{H+1} - 1)/(B - 1) \leqslant n$, 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

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

the indexing being given by an injective map σ_n from $J_{B,H}$ to $V_n := \{1, \ldots, n\}$. 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 $v \in V_n \backslash I_\varnothing$ is given the index $(k) = \sigma_n^{-1}(v)$, $1 \leqslant k \leqslant B$, if $\xi^n_{(1,v)}$ has the k^{th} smallest absolute value among $\{\xi^n_{1,j},\ j \neq 1\}$, 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 $H \geqslant 2$, we repeat the indexing procedure for the vertex indexed by (1) (the first child) on the set $V_n \backslash I_1$. We obtain a new set $\{11, \cdots, 1B\}$ 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 $V_n \backslash I_2$ and obtain a new set $\{21, \cdots, 2B\}$, and so on. When we have constructed $\{B1, \cdots, BB\}$, 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 $\{\mathbf{u}1, \cdots, \mathbf{u}B\}$ 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 with all the self loops. 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 $\xi^n_{i,j}$ 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 same sorting procedure. 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 $(\xi^n_{i,j})_{1\leqslant i\leqslant j\leqslant n}$ be a collection of i.i.d. random variables with values in $\mathbb{R}$ and set $\xi^n_{j,i}=\xi^n_{i,j}$. Let ν be a Radon measure on $\mathbb{R}$ with no mass at 0 and assume that*

$$n\mathbb{P}(\xi^n_{12}\in\cdot)\xrightarrow[n\to\infty]{w}\nu \quad as \quad n\to\infty. \tag{21}$$

Let G_n be the complete network on $\{1,\dots,n\}$ whose mark on edge (i,j) equals ξ^n_{ij}. 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], Aldous and Steele [3], see also Aldous and Lyons [2]. Let us give some examples of application of this proposition. Consider the case where $\xi^n_{ij}=1$ with probability λ/n and $\xi^n_{i,j}=\infty$ otherwise. The network G_n is an Erdös-Rényi random graph with parameter λ/n. From the proposition, we retrieve the well–known result that it converges to the tree of a Yule process of intensity λ. If $\xi^n_{i,j}=nY_{i,j}$, where $Y_{i,j}$ is a non–negative 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

$$\xi^n_{i,j}=(a_n/X_{i,j})\,,\quad \nu_\alpha(dx):=\left[\theta\mathbb{1}_{\{x>0\}}+(1-\theta)\mathbb{1}_{\{x<0\}}\right]\alpha|x|^{\alpha-1}dx\,, \tag{22}$$

where $|X_{i,j}|\in\mathbb{H}_\alpha$. Note that the proposition applies to all $\alpha>0$, provided (7) holds.

Proof of Proposition 2.6. We order the elements of $J_{B,H}$ in the lexicographic order, i.e. $\varnothing\prec 1\prec 2\prec\cdots\prec B\prec 11\prec 12\prec\cdots\prec B\dots B$. 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 $I_{\mathbf{v}}=\sigma_n(\cup_{\mathbf{w}\prec\mathbf{v}}O_{\mathbf{w}})$. At every step of the indexing procedure, we sort the marks of the neighboring edges that have not been explored at an earlier step $\{1,\cdots,n\}\backslash I_1$, $\{1,\cdots,n\}\backslash I_2$, $\cdots$. Therefore, for all $\mathbf{u}$,

$$(\xi^n_{\sigma_n(\mathbf{u}),i})_{i\notin I_{\mathbf{u}}}\overset{d}{=}(\xi^n_{1,i})_{1\leqslant i\leqslant n-|I_{\mathbf{u}}|}. \tag{23}$$

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

$$x^n_{\mathbf{v},\mathbf{w}}=\xi^n_{\sigma_n(\mathbf{v}),\sigma_n(\mathbf{w})}.$$

Also, let $\{y^n_{\mathbf{v},\mathbf{w}}\,,\ \mathbf{v},\mathbf{w}\in J_{B,H}\}$ denote independent variables with common law $|\xi^n_{1,2}|$. 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 $k\in\{1,\dots,B\}$ such that $\mathbf{u}=\mathbf{v}k$ or $\mathbf{v}=\mathbf{u}k$). Lemma 2.7 below implies that the vector $\{|x^n_{\mathbf{v},\mathbf{w}}|\,,\ \{\mathbf{v},\mathbf{w}\}\in E^{B,H}\}$ stochastically dominates the vector $\mathcal{Y}^n:=$

$\{y^n_{\mathbf{v},\mathbf{w}}\,,\ \{\mathbf{v},\mathbf{w}\}\in E^{B,H}\}$, i.e. there exists a coupling of the two vectors such that almost surely $|x^n_{\mathbf{v},\mathbf{w}}|\geqslant y^n_{\mathbf{v},\mathbf{w}}$, for all $\{\mathbf{v},\mathbf{w}\}\in E^{B,H}$. Since $\mathcal{Y}^n$ contains a finite (independent of n) number of variables, (21) implies that the probability of the event $\{\min_{\{\mathbf{v},\mathbf{w}\}\in E^{B,H}}|x^n_{\mathbf{v},\mathbf{w}}|\leqslant t\}$ 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 $(x^n_e)_{e\in J_{B,H}\times J_{B,H}}$ 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 mappings $\sigma_{n,1}$, $\sigma_{n,2}$ from $J_{B,H}$ to $\{1,\cdots,n\}$. We need to show that in distribution,

$$((G_n,1)^{B,H},(G_n,2)^{B,H})\to((\mathcal{T}_1,\varnothing)^{B,H},(\mathcal{T}_2,\varnothing)^{B,H}). \tag{24}$$

Let $V^{B,H}_{n,i}=\sigma_{n,i}(J_{B,H})$ be the vertex set of $(G_n,i)^{B,H}$, $i=1,2$. There are $C:=(B^{H+1}-1)/(B-1)$ vertices in $V^{B,H}_{n,i}$, hence the exchangeability of the variables implies that

$$\mathbb{P}(2\in V^{B,H}_{n,1})\leqslant C/n.$$

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

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

Therefore in order to prove (24) it is sufficient to prove that with probability tending to 1,

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

Indeed, on the event $\{V^{B,H}_{n,1}\cap V^{B,H}_{n,2}=\varnothing\}$, $(G_n,2)^{B,H}$ 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 $I_{\mathbf{v},2}=\{2\}\bigcup\cup_{\mathbf{w}\prec\mathbf{v}}O_{\mathbf{w},2}$, and

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

For any $\mathbf{u},\mathbf{v}\in J_{B,H}$, if $V^{B,H}_{n,1}\cap I_{\mathbf{v},2}=\varnothing$, 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 $|\xi^n_{\sigma_{n,1}(\mathbf{u}),\sigma_{n,2}(\mathbf{v})}|$ given $V^{B,H}_{n,1}\cap I_{\mathbf{v},2}=\varnothing$ dominates stochastically $|\xi^n_{1,2}|$, and is independent of $(|\xi^n_{\sigma_{n,2}(\mathbf{v}),k}|)_{k\in\{1,\ldots,n\}\backslash(V^{B,H}_{n,1}\cup I_{\mathbf{v},2})}$, an i.i.d. vector with law $|\xi^n_{1,2}|$. It follows that

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

Therefore,

$$\begin{aligned}\mathbb{P}(O_{\mathbf{v},2}\cap V^{B,H}_{n,1}\neq\varnothing\,|\,V^{B,H}_{n,1}\cap I_{\mathbf{v},2}=\varnothing)&\leqslant\sum_{\mathbf{u}\in J_{B,H}}\mathbb{P}(\sigma_{n,1}(\mathbf{u})\in O_{\mathbf{v},2}\,|\,V^{B,H}_{n,1}\cap I_{\mathbf{v},2}=\varnothing)\\&\leqslant\frac{CB}{n-2C}.\end{aligned}$$

Finally,

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

which converges to 0 as $n \to \infty$. □

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 $\{1, \dots, n\}$, including self-loops, such that $\{\sigma_n^{-1}(i), \sigma_n^{-1}(j)\}$ is not an edge of the finite tree on $J_{B,H}$. Clearly, any self-loop $\{i, i\}$ belongs to $\mathcal{E}_n^{B,H}$. Also, for $\mathbf{u} \neq \varnothing$ 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 $(B^{H+1} - 1)/(B - 1) \leqslant n$, 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 $\varnothing \neq \mathbf{u} \in J_{B,H}$:

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

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

Proof of Lemma 2.7. 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 $Y_1, \dots, Y_m$ denote i.i.d. positive random variables. Suppose $m = n_1 + \cdots n_\ell$, for some positive integers ℓ, $n_1, \dots, n_\ell$, and partition the m variables in ℓ blocks $I^1, \dots, I^\ell$ of $n_1, \dots, n_\ell$ variables each. Fix some non–negative integers k_j such that $k_j \leqslant n_j$ and call $q_1^j, \dots, q_{k_j}^j$, 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 $Y_1, \dots, Y_{n_1}$, $Y_{q_2^1}$ is the second lowest of the $Y_1, \dots, Y_{n_1}$ and so on). Consider the random index sets of the k_j minimal values in the j^{th} block, $J^j := \cup_{i=1}^{k_j} \{q_i^j\}$, and set $J = \cup_{j=1}^{\ell} J^j$. If $k_j = 0$ we set $J^j = \varnothing$. Finally, let $\widetilde{Y}$ denote the vector $\{Y_i\,,\ i = 1, \dots, m\,;\ i \notin J\}$. 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 $y_1, \dots, y_m$ of the whole vector. Given this we isolate the index sets $J^1, \dots, J^\ell$ within each block. We then consider two vectors $\mathcal{Z}, \mathcal{V}$ obtained as follows. $\mathcal{Z}_1 = (z_1^1, \dots, z_{n_1-k_1}^1, z_1^2, \dots, z_{n_2-k_2}^2, \dots, z_{n_\ell-k_\ell}^\ell)$ is obtained by extracting the $n_1 - k_1$ values $z_1^1, \dots, z_{n_1-k_1}^1$ uniformly at random (without replacement) from the values $y_1, \dots, y_{n_1}$ (in the block I^1), the $n_2 - k_2$ variables $z_1^2, \dots, z_{n_2-k_2}^2$ in the same way from the values $y_{n_1+1}, \dots, y_{n_1+n_2}$ (in the block I^2), and so on. On the other hand, the vector $\mathcal{V} = (v_1^1, \dots, v_{n_1-k_1}^1, v_1^2, \dots, v_{n_2-k_2}^2, \dots, v_{n_\ell-k_\ell}^\ell)$ is obtained as follows. For the first block we take v_i^1, $i = 1, \dots, n_1 - k_1$ equal to z_i^1 whenever an index $i \in I^1 \setminus J^1$ was picked for the vector $z_1^1, \dots, z_{n_1-k_1}^1$ 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 $z_1^1, \dots, z_{n_1-k_1}^1$. We repeat this procedure for all other blocks to assign all values of $\mathcal{V}$. By construction, $\mathcal{V} \geqslant \mathcal{Z}$ 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 $\widetilde{Y}$. □

2.6 Proof of Theorem 2.3

Proof of Theorem 2.3(i). Let ν_α be as in (22) and let $(\mathcal{T}_\alpha, \varnothing)$ be a realization of the PWIT(ν_α). 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, $x_{(\mathbf{v},\mathbf{w})} = \infty$ if $\mathbf{v}$ and $\mathbf{w}$ are at graph-distance different from 1. In particular, if we set $y_\mathbf{v} = \mathrm{sign}(x_\mathbf{v})|x_\mathbf{v}|^\alpha$, then the point sets $\Xi_\mathbf{v} = \{y_{\mathbf{v}k}\}_{k\geqslant 1}$ are independent Poisson point process 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. (13). Next, we apply Proposition 2.6 with the setting described in (22). 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 to the network $(G_n, 1)^{B,H}$ (see the construction given before Proposition 2.6). From 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 $(B_n^{H_n+1}-1)/(B_n-1) \leqslant n$ and such that for any pair $\mathbf{u}, \mathbf{v} \in \mathbb{N}^f$ we have $\xi^n_{(\widetilde{\sigma}_n(\mathbf{u}),\widetilde{\sigma}_n(\mathbf{v}))} \to x_{(\mathbf{v},\mathbf{w})}$ 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.

$$\langle \delta_\mathbf{u}, \widetilde{\sigma}_n^{-1}(a_n^{-1}X)\widetilde{\sigma}_n\delta_\mathbf{v}\rangle = \frac{1}{\xi^n_{(\widetilde{\sigma}_n(\mathbf{u}),\widetilde{\sigma}_n(\mathbf{v}))}} \to \frac{1}{x_{(\mathbf{u},\mathbf{v})}} = \langle \delta_\mathbf{u}, \mathbf{T}\delta_\mathbf{v}\rangle\,. \tag{25}$$

For any $\mathbf{v}$, 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 for any $\mathbf{v} \in \mathbb{N}^f$, $\psi_n^\mathbf{v} \to \mathbf{T}\delta_\mathbf{v}$ in $L^2(\mathbb{N}^f)$ almost surely as $n \to \infty$, i.e.

$$\sum_\mathbf{u} (\langle \delta_\mathbf{u}, \psi_n^\mathbf{v}\rangle - \langle \delta_\mathbf{u}, \mathbf{T}\delta_\mathbf{v}\rangle)^2 \to 0\,. \tag{26}$$

Since from (25) 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 Lemma 2.7 and Lemma 2.4(i). The proof of Theorem 2.3(i) is complete. □

Proof of Theorem 2.3(ii). Next, we claim that, for all $\mathbf{u} \in \mathbb{N}^f$, a.s.

$$\lim_{n\to\infty} \frac{\rho_{\widetilde{\sigma}_n(\mathbf{u})}}{n} = 1. \tag{27}$$

To prove this we first observe that by Lemma 2.7 and the strong law of large numbers, we have a.s.

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

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 a.s. $\limsup_{n\to\infty} \rho_{\widetilde{\sigma}_n(\mathbf{u})}/n \leqslant 1$. Next, we can estimate

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

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 (23) and the strong law of large numbers imply, a.s. $\liminf_{n\to\infty} \rho_{\widetilde{\sigma}_n(\mathbf{u})}/n \geqslant 1$. This proves (27).

We need to show that for each $\mathbf{v} \in \mathbb{N}^f$, (26) holds with the new vector $\psi_n^\mathbf{v} := \widetilde{\sigma}_n^{-1}(\kappa_n S)\widetilde{\sigma}_n\delta_\mathbf{v}$,

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

Now,

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

The sum of the second term above converges to zero as in the proof of point (i). Moreover, as in that proof, the first term above is uniformly integrable so that from (27), the sum converges to zero. This proves point (ii). □

Proof of Theorem 2.3(iii). The setting is as in the proof of (ii) above, but now $\alpha\in(0,1)$. We build the operator $\mathbf{S}$ on the tree $\mathcal{T}_\alpha$ as in (16). We need to prove that for any $\mathbf{v}\in\mathbb{N}^f$, a.s.

$$\sum_{\mathbf{w}}(\langle\delta_{\mathbf{w}},\psi_n^{\mathbf{v}}\rangle-\langle\delta_{\mathbf{w}},\mathbf{S}\delta_{\mathbf{v}}\rangle)^2\to 0\,, \tag{28}$$

with $\psi_n^{\mathbf{v}}:=\widetilde{\sigma}_n^{-1}S\widetilde{\sigma}_n\delta_{\mathbf{v}}$, i.e.

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

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

$$\frac{U_{\widetilde{\sigma}_n(\mathbf{w}),\widetilde{\sigma}_n(\mathbf{v})}}{\sqrt{\rho_{\widetilde{\sigma}_n(\mathbf{v})}\rho_{\widetilde{\sigma}_n(\mathbf{w})}}}\to\frac{\langle\delta_{\mathbf{w}},\mathbf{T}\delta_{\mathbf{v}}\rangle}{\sqrt{\rho(\mathbf{v})\rho(\mathbf{w})}}=\langle\delta_{\mathbf{w}},\mathbf{S}\delta_{\mathbf{v}}\rangle\,. \tag{29}$$

Multiplying and dividing by a_n and using (25) with $\theta=1$, we see that (29) holds if

$$a_n^{-1}\rho_{\widetilde{\sigma}_n(\mathbf{v})}\to\rho(\mathbf{v})\,, \tag{30}$$

almost surely, for every $\mathbf{v}\in\mathbb{N}^f$. In turn, (30) 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 $\mathbf{v}$, (25) 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 Lemma 2.7 and Lemma 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 (30) and (29).

Once we have (29), to conclude the proof it is sufficient to show that a.s.

$$\lim_{k\to\infty}\sup_n\sum_{\mathbf{w}\notin J_{k,k}}(\langle\delta_{\mathbf{w}},\psi_n^{\mathbf{v}}\rangle)^2=0\,. \tag{31}$$

However, using (30) and the simple bound $(\langle\delta_{\mathbf{w}},\psi_n^{\mathbf{v}}\rangle)^2\leqslant\frac{U_{\widetilde{\sigma}_n(\mathbf{v}),\widetilde{\sigma}_n(\mathbf{w})}}{\rho_{\widetilde{\sigma}_n(\mathbf{v})}}$, we have that (31) again follows from an application of Lemma 2.7 and Lemma 2.4(ii). This completes the proof of Theorem 2.3(iii). □

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: $X \otimes X(\phi_1, \phi_2) = (X\phi_1, X\phi_2)$, 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 $\mathcal{T}^1_\alpha$, $\mathcal{T}^2_\alpha$ of the PWIT(ν_α), 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 $((\mathcal{T}^1_\alpha, \varnothing)^{B,H}, (\mathcal{T}^2_\alpha, \varnothing)^{B,H})$. 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 mapping 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 $(a_n^{-1} X \oplus X, (1, 2)) \to (\mathbf{T}_1 \oplus \mathbf{T}_2, (\varnothing, \varnothing))$,

(ii) if $\alpha \in (1, 2)$ and $\theta = 1$ then $(\kappa_n S \oplus S, (1, 2)) \to (\mathbf{T}_1 \oplus \mathbf{T}_2, (\varnothing, \varnothing))$,

(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 Proof of Theorem 1.4: Markov matrices with $\alpha \in (0, 1)$

Recall that $\mathbf{S}$ is a bounded self–adjoint operator on $L^2(V)$, whose spectrum is contained in $[-1, 1]$, cf. (17). 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

$$\gamma_\ell := \langle \delta_\varnothing, \mathbf{S}^\ell \delta_\varnothing \rangle \tag{32}$$

Note that $\gamma_\ell = \frac{1}{\rho(\varnothing)} \langle \delta_\varnothing, \mathbf{K}^\ell \delta_\varnothing \rangle_\rho$ is the probability that the random walk on the PWIT associated to the stochastic operator $\mathbf{K}$ comes back to the root (where it started) after ℓ steps. In particular, $\gamma_\ell = 0$ for ℓ odd. We set $\gamma_0 = 1$. Let $\mu_\varnothing$ denote the spectral measure of $\mathbf{S}$ associated to $\delta_\varnothing$ (see e.g. [22, Chapter VII]). Equivalently, $\mu_\varnothing$ is the spectral measure of $\mathbf{K}$ associated to the $L^2(V, \rho)$ normalized vector $\hat{\delta}_\varnothing := \delta_\varnothing / \sqrt{\rho(\varnothing)}$, cf. (17). In particular, $\mu_\varnothing$ is a probability measure supported on $[-1, 1]$ and such that $\gamma_\ell = \int_{-1}^1 x^\ell \mu_\varnothing(dx)$, for every ℓ. Since all odd moments vanish $\mu_\varnothing$ is symmetric. Moreover, for any $z \in \mathbb{C}_+$ we have

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

i.e. $\langle \delta_\varnothing, R(z) \delta_\varnothing \rangle$ is the Stieltjes transform of $\mu_\varnothing$. Recall that the 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 and $|m_\mu(z)| \leqslant (\Im z)^{-1}$ 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

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

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

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

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

$$\lim_{n\to\infty} m_{\mathbb{E}\mu_K}(z) = m_{\mathbb{E}\mu_\varnothing}(z). \tag{33}$$

We define

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

The proof of Theorem 1.4 will be complete if we prove that, for all $z \in \mathbb{C}_+$:

$$\lim_{n\to\infty} \mathbb{E}\left|m_{\mu_K}(z) - m_{\mathbb{E}\mu_\varnothing}(z)\right| = 0. \tag{34}$$

We write

$$\mathbb{E}|m_{\mu_K}(z) - m_{\mathbb{E}\mu_\varnothing}(z)| \leqslant \mathbb{E}\left|m_{\mu_K}(z) - \mathbb{E}m_{\mu_K}(z)\right| + \left|m_{\mathbb{E}\mu_K}(z) - m_{\mathbb{E}\mu_\varnothing}(z)\right|$$

On the right hand side, the second term converges to 0 by (33). The first term is equal to

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

By exchangeability, we note that

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{1}{n}\sum_{k=1}^{n}\left[R^n_{k,k}(z) - \mathbb{E}R^n_{k,k}(z)\right]\right)^2\right] \\
&\quad= \frac{1}{n}\mathbb{E}\left(R^n_{1,1} - \mathbb{E}R^n_{1,1}\right)^2 + \frac{n(n-1)}{n^2}\mathbb{E}\left[\left(R^n_{1,1} - \mathbb{E}R^n_{1,1}\right)\left(R^n_{2,2} - \mathbb{E}R^n_{2,2}\right)\right] \\
&\quad\leqslant \frac{1}{n(\Im z)^2} + \mathbb{E}\left[\left(R^n_{1,1} - \mathbb{E}R^n_{1,1}\right)\left(R^n_{2,2} - \mathbb{E}R^n_{2,2}\right)\right].
\end{aligned}$$

Finally, Theorem 2.2 and Theorem 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 (34) follows.

3.2 Proof of Theorem 1.2: i.i.d. symmetric matrices with $\alpha \in (0,2)$

Theorem 1.2 is a corollary of Theorem 3.1 below. As in [7], we will define the LSD μ_α by an expression for its Stietljes transform $m_{\mu_\alpha}(z)$. Let $\mathcal{H}$ be the set of analytic functions h from $\mathbb{C}_+$ to $\mathbb{C}_+$ such that $|h(z)| \leqslant (\Im z)^{-1}$. We define the distance on $\mathcal{H}$, $d(f,g) = \int_\Omega |f(z) - g(z)|dz$ where Ω is a non–empty bounded open set in $\mathbb{C}+$. With this distance, $\mathcal{H}$ is a compact separable metric space (compactness follows from Vitali's convergence Theorem). The symbol $\stackrel{d}{=}$ stands for equality in distribution.

Theorem 3.1 (Recursive Distributional Equation)**.** *With the assumption of Theorem 1.2.*

(i) There exists a probability distribution L on $\mathcal{H}$ such that L-a.s. for all $z \in \mathbb{C}_+$,

$$h(-\bar{z}) = -\bar{h}(z)$$

and

$$h(z) \stackrel{d}{=} -\left(z + \sum_{k\in\mathbb{N}} \xi_k h_k(z) \right)^{-1}, \tag{35}$$

where h and $(h_k)_{k\in\mathbb{N}}$ are i.i.d. with law L, independent of $\{\xi_k\}_{k\in\mathbb{N}}$, a Poisson point process of $\mathbb{R}_+$ with intensity $\frac{\alpha}{2}x^{-\frac{\alpha}{2}-1}dx$. Moreover, for all $z \in \mathbb{C}_+$, the law of $h(z)$ depends only on α.

(ii) For all $z \in \mathbb{C}_+$, $\mu_{a_n^{-1}X}$ converges in probability to a deterministic probability measure μ_α whose Stieltjes transform is given by $m_{\mu_\alpha}(z) = \mathbb{E}h(z)$.

Statement (ii) is equivalent to the main result of Ben Arous and Guionnet [7]. The Recursive Distributional Equation (RDE) (35) is new. The random variable h appears has the fixed point of a nice cascade. The analysis of the RDE (35) is performed in Section 4. We conjecture that the law L in (i) is unique but we cannot prove it. In any case, since the marginal law $h(z)$ depends only on α, $\mathbb{E}h(z)$ is the same for all probability measures solving the RDE (35). The key idea of the proof is the same as in the proof of Theorem 1.4: apply Theorem 2.2 in conjunction with Theorems 2.3 and 2.8. Unfortunately, this direct approach will fail since we could not manage to prove that $\mathbf{T}$ is a self–adjoint operator, and we cannot apply directly Theorem 2.2. We will instead use a truncation argument to overcome this difficulty which makes the problem more interesting per se.

Proof of Theorem 3.1(i). To simplify notation, we set $A_n = a_n^{-1}X$ and $\beta = \alpha/2$. Let h be a random variable in $\mathcal{H}$ satisfying Equation (35). For $z = it$, with $t > 0$, the identity, $h(-\bar{z}) = -\bar{h}(z)$ reads $\Re h(it) = 0$. Thus, the equation satisfied by $g(it) = \Im h(it) \geqslant 0$ is

$$g(it) \stackrel{d}{=} \left(t + \sum_{k\in\mathbb{N}} \xi_k g_k(it) \right)^{-1}. \tag{36}$$

Lemma 3.2 (Existence and Uniqueness of solution for the Random Distributional Equation)**.** *For each $t > 0$, there exists a unique probability measure L^{it} on $\mathbb{R}_+$, solution of* (36)*.*

Proof of Lemma 3.2. If (Y_k) is an i.i.d. sequence of non–negative random variables, independent of $\{\xi_k\}_{k\in\mathbb{N}}$, such that $\mathbb{E}[Y_1^\beta] < \infty$ then it is well known that $\sum_k \xi_k Y_k \stackrel{d}{=} \sum_k \xi_k (\mathbb{E}[Y_1^\beta])^{1/\beta}$ (see for example Lemma 6.5.15 in Talagrand [27] or (38) below). This implies the unicity for Equation (36) provided that the equation satisfied by $\mathbb{E}[g(it)^\beta]$ has a unique solution. Recall the formulas of Laplace transforms, for $y \geqslant 0$, $\eta > 0$ and $0 < \eta < 1$ respectively,

$$y^{-\eta} = \Gamma(\eta)^{-1} \int_0^\infty x^{\eta-1} e^{-xy} dx \quad \text{and} \quad y^\eta = \Gamma(1-\eta)^{-1} \eta \int_0^\infty x^{-\eta-1}(1 - e^{-xy}) dx. \tag{37}$$

From the exponential formula we deduce that, with $s \geqslant 0$,

$$\begin{aligned} \mathbb{E} \exp\left(-s \sum_k \xi_k Y_k \right) &= \exp\left(\mathbb{E} \int_0^\infty (e^{-xsY_1} - 1) \beta x^{-\beta-1} dx \right) \\ &= \exp\left(-\Gamma(1-\beta) s^\beta \mathbb{E}[Y_1^\beta] \right). \end{aligned} \tag{38}$$

From Equation (36), $\mathbb{E}[g(it)^\beta]$ is 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 $t \geqslant 0$. 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. □

We shall need the following observations. 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 $(\mathbf{v}, \mathbf{w})$ if $\langle \delta_\mathbf{v}, \mathbf{S}\delta_\mathbf{w} \rangle \neq 0$. The next lemma is classical.

Lemma 3.3 (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 of Lemma 3.3. Assume first that $\mathbf{S}$ is bounded: for all $\mathbf{w} \in V$, $\|\mathbf{S}\delta_\mathbf{w}\| \leqslant C$. For $|z| > C$, the series expansion of the resolvent gives

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

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 [22]. □

For $z \in \mathbb{C}_+$, we define the Stieltjes transform

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

where

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

is the resolvent of A_n. Note that, by exchangeability, $\mathbb{E} m_{A_n}(z) = \mathbb{E} R^n_{1,1}(z)$. The main part of the proof of Theorem 3.1 is devoted to the convergence of $\mathbb{E} m_{A_n}(z)$ as n goes to infinity. To this end, we will use truncation arguments and exhibit the limiting self–adjoint operator for the truncated versions of A_n.

Recall how the operator $\mathbf{T}$ was build on $(\mathcal{T}, \varnothing)$, a realization of the PWIT(ℓ_θ), cf. Section 2.3. For $\kappa > 0$, we define the truncated operator $\mathbf{T}_\kappa$ on $\mathcal{D}$, for $\mathbf{v}, \mathbf{w} \in \mathbb{N}^f$,

$$\langle \delta_\mathbf{v}, \mathbf{T}_\kappa \delta_\mathbf{w} \rangle = \langle \delta_\mathbf{v}, \mathbf{T} \delta_\mathbf{w} \rangle \, \mathbb{1}(|\langle \delta_\mathbf{v}, \mathbf{T} \delta_\mathbf{w} \rangle| < \kappa).$$

The operator $\mathbf{T}_\kappa$ is symmetric and as usual, we identify $\mathbf{T}_\kappa$ with its closed extension. with domain $D(\mathbf{T}_\kappa)$. By construction, $\mathcal{D}$ is a core for $\mathbf{T}_\kappa$.

Proposition 3.4.

(i) With probability one, $\mathbf{T}_\kappa$ is self–adjoint.

(ii) If $h^\kappa(z) = \langle \delta_\varnothing, (\mathbf{T}_\kappa - zI)^{-1}\delta_\varnothing\rangle$ then with probability one, for all $z \in \mathbb{C}_+$,

$$h^\kappa(-\bar{z}) = -\bar{h^\kappa}(z)$$

and

$$h^\kappa(z) \stackrel{d}{=} -\left(z + \sum_{k\in\mathbb{N}} \xi_k \mathbb{1}_{\{\xi_k<\kappa^2\}} h_k^\kappa(z)\right)^{-1}, \tag{39}$$

where $(h_k^\kappa)_{k\in\mathbb{N}}$ are i.i.d. with law h^κ, independent of $\{\xi_k\}_{k\in\mathbb{N}}$, a Poisson point process of $\mathbb{R}_+$ with intensity $\frac{\alpha}{2}x^{-\frac{\alpha}{2}-1}dx$.

(iii) Theorem 3.1(i) holds true, and as κ goes to infinity, any accumulation point of h^κ satisfies Theorem 3.1(i).

Proof of Proposition 3.4(i). We shall see in a moment that the proofs of Theorem 2 and Proposition 10 in Zakharevich [28] gives precisely that

$$\mathbb{E}\langle \delta_\varnothing, \mathbf{T}_\kappa^{2n}\delta_\varnothing\rangle \leqslant R^n n^n\,, \tag{40}$$

for some constant R depending on κ. Recall that $\mathbf{T}_\kappa$ is symmetric, hence

$$\|\mathbf{T}_\kappa^n \delta_\varnothing\|^2 = \langle \delta_\varnothing, (\mathbf{T}_\kappa)^{2n}\delta_\varnothing\rangle.$$

Then, by the Markov inequality we get $\mathbb{P}(\|\mathbf{T}_\kappa^n\delta_\varnothing\|^2 \geqslant (2R)^n n^n) \leqslant 2^{-n}$ and it follows from the Borel-Cantelli Lemma that a.s.

$$\|\mathbf{T}_\kappa^n \delta_\varnothing\|^2 < (2R)^n n^n$$

A similar estimate holds for $\|\mathbf{T}_\kappa^n\delta_{\mathbf{v}}\|$, for every $\mathbf{v} \in \mathbb{N}^f$. We may then conclude (i) by applying Nelson's *analytic vector theorem*, see e.g. Reed and Simon [21, Theorem X.39].

The bound (40) comes from a combinatoric computation which follows from [28]. By definition, $\langle \mathbf{T}_\kappa^{2n}\delta_\varnothing, \delta_\varnothing\rangle$ is the sum, on all paths of length $2n$ from $\varnothing$ to $\varnothing$, of the products of the weights $w^{-1/\alpha}$ on the path. Let Π_n denote the set of equivalence classes of such paths (for the equivalence relation obtained from isomorphisms on the vertices that preserve the root $\varnothing$). Since $\mathcal{T}$ is a tree, the graph formed by the edges along a path in $\pi \in \Pi_n$ is a rooted tree on $l+1$ vertices. A given edge e in this graph is visited $2b_e$ times by the path and

$$\sum_{i=1}^{l} b_{e_i} = n.$$

Recall the Campbell formula for a Poisson point process $\Xi = \{\zeta_k\}_{k\in\mathbb{N}}$ of intensity measure $d\lambda(x)$: if $f_1, \ldots, f_l$ non–negative functions and Ξ^{l*} is the set of l-tuples of distinct elements in Ξ, we have

$$\mathbb{E}\sum_{(\zeta_1,\ldots,\zeta_l)\in\Xi^{l*}} \prod_{i=1}^{l} f_i(\zeta_i) = \prod_{i=1}^{l}\int f_i(x)d\lambda(x) = \mathbb{E}\sum_{\zeta_1\in\Xi_1,\ldots,\zeta_l\in\Xi_l} \prod_{i=1}^{l} f_i(\zeta_i).$$

where $\Xi_1, \ldots, \Xi_l$ are independent copies of Ξ. In our present case, $\Xi = \{|w_k|^{-1/\alpha} \mathbb{1}_{\{|w_k^{-1/\alpha}|<\kappa\}}\}_{k\in\mathbb{N}}$ is a Poisson point process of intensity measure $\alpha x^{-1-\alpha}\mathbb{1}_{\{0<x<\kappa\}}dx$. Expanding all coefficients, we get easily that

$$\begin{aligned}
\mathbb{E}\langle \delta_\varnothing, \mathbf{T}_\kappa^{2n}\delta_\varnothing\rangle &= \sum_{\pi\in\Pi_n}\prod_{i=1}^{l(\pi)}\int_0^\kappa x^{2b_{e_i}(\pi)}\alpha x^{-1-\alpha}dx \\
&= \sum_{\pi\in\Pi_n}\prod_{i=1}^{l(\pi)}\frac{\alpha\kappa^{2b_{e_i}(\pi)-\alpha}}{2b_{e_i}(\pi)-\alpha} \\
&\leqslant \kappa^{2n}\max\left(1,\frac{\alpha}{2-\alpha}\right)^n|\Pi_n|.
\end{aligned}$$

In the proof of Theorem 2 in [28], Zakharevich finds a bijection from Π_n to a set of on rooted colored trees, then her Proposition 10 gives precisely (40). □

Proof of Proposition 3.4(ii). With probability one, $\mathbf{T}_\kappa$ is self–adjoint, so the property $h^\kappa(-\bar{z}) = -\bar{h^\kappa}(z)$ is a consequence of Lemma 3.3 (indeed trees are bipartite graphs). The recursive distributional equation (39) follows now from a classical operator version of Schur complement formula (see e.g. Proposition 2.1 in Klein [17] 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 $\mathbb{N}^f = \{\varnothing\}\bigcup\cup_k k\mathbb{N}^f$. We define $\mathbf{T}_\kappa^{(k)}$ as the projection of $\mathbf{T}_\kappa$ 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

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

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

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

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

$$\widetilde{R}(z)(\mathbf{T}_\kappa - \widetilde{\mathbf{T}}_\kappa)R(z) = \widetilde{R}(z) - R(z)\,. \tag{41}$$

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 $\mathbb{N}^f = \{\varnothing\}\bigcup\cup_k k\mathbb{N}^f$ 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 (41) we then obtain, for $k \in \mathbb{N}$:

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

It follows that

$$\langle \delta_\varnothing, \widetilde{R}(z)(\mathbf{T}_\kappa - \widetilde{\mathbf{T}}_\kappa)R(z)\delta_\varnothing\rangle = \sum_{k\in\mathbb{N}}\widetilde{R}_{\varnothing,\varnothing}(z)u_kR_{k,\varnothing}(z) = -\sum_{k\in\mathbb{N}}\widetilde{R}_{\varnothing,\varnothing}(z)\widetilde{R}_{k,k}(z)u_k^2R_{\varnothing,\varnothing}(z)\,.$$

From (41) 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 (ii) 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 $z \in \mathbb{C}_+$. □

Proof of Proposition 3.4(iii). We consider the sequence of random analytic functions $(h^\kappa)_{\kappa>0}$. Since $\mathcal{H}$ is compact, $(h^\kappa)_{\kappa>0}$ is tight, and any accumulation point, say h, is a random variable in $\mathcal{H}$ satisfying with probability one, for all $z \in \mathbb{C}_+$, $h(-\bar{z}) = -\bar{h}(z)$. Let $D = i[a,b]$ be a non–empty interval of $i\mathbb{R}_+ \subset \mathbb{C}_+$. By continuity, for each $z \in D$, $h(z)$ satisfies Equation (35) which has a unique pure imaginary solution by Lemma 3.2. Therefore, for each $z \in D$, the law of $h(z)$ is unique and does not depend on the accumulation point. By analyticity we may extend this result to all $\mathbb{C}_+$. □

This proves Theorem 3.1(i).

□

Proof of Theorem 3.1(ii). We shall use the next lemma, which is in fact a consequence of Lidskii's rank difference inequality.

Lemma 3.5 (Rank inequality on Stieltjes transforms)**.** *Let m_A and m_B be the Stieltjes transform of the spectral measure of A and B, two $n \times n$ hermitian matrices. Then for all $z \in \mathbb{C}_+$,*

$$|m_A(z) - m_B(z)| \leqslant \frac{\pi}{n\Im z}\mathrm{rank}(A-B)$$

Proof of Lemma 3.5. Let $F_A(x)$ and $F_B(x)$ be the partition functions of the spectral measures of A and B. From Lidskii's rank difference inequality we have

$$\|F_A - F_B\| = \sup_{x\in\mathbb{R}} |F_A(x) - F_B(x)| \leqslant n^{-1}\mathrm{rank}(A-B).$$

On the other hand, since $m_A(z) = \int F_A(x)/(x-z)^2 dx$, we get

$$|m_A(z) - m_B(z)| \leqslant \int \frac{|F_A(x) - F_B(x)|}{|x-z|^2} dx \leqslant \|F_A - F_B\| \int \frac{dx}{x^2 + (\Im z)^2}.$$

□

We define the $n \times n$ matrix A_n^κ with entries $(A_n^\kappa)_{ij} = (A_n)_{ij}\mathbb{1}(|(A_n)_{ij}| < \kappa)$. Since $\mathrm{rank}(A_n - A_n^\kappa) \leqslant \sum_{i=1}^n \mathbb{1}_{\{((A_n)_{i1},\ldots,(A_n)_{in}\} \neq ((A_n^\kappa)_{i1},\ldots,(A_n^\kappa)_{in}))}$, we get

$$\begin{aligned} n^{-1}\mathbb{E}\left[\mathrm{rank}(A_n - A_n^\kappa)\right] &\leqslant \mathbb{P}\left(((A_n)_{1,1},\ldots,(A_n)_{1,n}) \neq ((A_n^\kappa)_{1,1},\ldots,(A_n^\kappa)_{1n})\right) \\ &\leqslant \mathbb{P}\left(\exists\, 1 \leqslant i \leqslant n : U_{1,i} > \kappa a_n\right) \\ &\leqslant 1 - (1 - G(a_n\kappa))^n. \end{aligned}$$

From (6), $G(a_n\kappa)$ is equivalent to $(n\kappa^\alpha)^{-1}$ as $n \to \infty$. Now, since $(1-x/n)^n$ converges to e^{-x}, we deduce that for all $\epsilon > 0$, there exists κ_ϵ and n_ϵ, such that for all $\kappa > \kappa_\epsilon$ and $n > n_\epsilon$ we have

$$n^{-1}\mathbb{E}\left[\mathrm{rank}(A_n - A_n^\kappa)\right] < \epsilon, \tag{42}$$

(note that Lemma 2.4(1) in [7], is a refinement of this bound).

The operator $\mathbf{T}_\kappa$ is the weak local limit of the sequence of matrices (A_n^κ) as n goes to infinity. Indeed, we may restate Theorems 2.3, 2.8(i) as: for all $\kappa > 0$, in distribution,

$$(A_n^\kappa, 1) \to (\mathbf{T}_\kappa, \varnothing) \quad \text{and} \quad (A_n^\kappa \oplus A_n^\kappa, (1,2)) \to (\mathbf{T}_{\kappa,1} \oplus \mathbf{T}_{\kappa,2}, (\varnothing, \varnothing)), \tag{43}$$

where $\mathbf{T}_{\kappa,1}$ and $\mathbf{T}_{\kappa,2}$ are two independent copies of $\mathbf{T}_\kappa$. Since $\mathbf{T}_\kappa$ and A_n^κ are self adjoint operators with probability one, we can apply Theorem 2.2. If $h_n^\kappa(z) = ((A_n^\kappa - zI)^{-1})_{1,1}$, then for all $z \in \mathbb{C}_+$, in distribution,

$$h_n^\kappa(z) \to h^\kappa(z) \tag{44}$$

where h^κ was defined in Proposition 3.4.

By exchangeability, we have $\mathbb{E}m_{A_n^\kappa}(z) = n^{-1}\sum_{1\leqslant k\leqslant n}\mathbb{E}R_{k,k}^{n,\kappa} = \mathbb{E}h_n^\kappa(z)$, with $R^{n,\kappa} = ((A_n^\kappa - zI)^{-1})$. Hence, by Lemma 3.5,

$$\begin{aligned}|\mathbb{E}m_{A_n}(z) - \mathbb{E}h(z)| &\leqslant \mathbb{E}\left|m_{A_n}(z) - m_{A_n^\kappa}(z)\right| + |\mathbb{E}h_n^\kappa(z) - \mathbb{E}h^\kappa(z)| + |\mathbb{E}h^\kappa(z) - \mathbb{E}h(z)| \\ &\leqslant \frac{\pi}{n\Im z}\mathbb{E}\mathrm{rank}(A_n - A_n^\kappa) + |\mathbb{E}h_n^\kappa(z) - \mathbb{E}h^\kappa(z)| + |\mathbb{E}h^\kappa(z) - \mathbb{E}h(z)|\,.\end{aligned}$$

By Equation (44), $\mathbb{E}h_n^\kappa(z)$ converges to $\mathbb{E}h^\kappa(z)$ while Proposition 3.4 states that $\lim_\kappa \mathbb{E}h^\kappa(z) = \mathbb{E}h(z)$. Thus (42) implies that for all $z \in \mathbb{C}_+$,

$$\lim_{n\to\infty} \mathbb{E}m_{A_n}(z) = \mathbb{E}h(z)$$

The proof of Theorem 3.1(ii) will be complete if we manage to check

$$\lim_{n\to\infty} \mathbb{E}|m_{A_n}(z) - \mathbb{E}h(z)| = 0.$$

As above we write

$$\begin{aligned}\mathbb{E}|m_{A_n}(z) - \mathbb{E}h(z)| \quad &\leqslant \quad \mathbb{E}\left|m_{A_n}(z) - m_{A_n^\kappa}(z)\right| + \mathbb{E}\left|m_{A_n^\kappa}(z) - \mathbb{E}h_n^\kappa(z)\right| \\ &\qquad + |\mathbb{E}h_n^\kappa(z) - \mathbb{E}h^\kappa(z)| + |\mathbb{E}h^\kappa(z) - \mathbb{E}h(z)|\end{aligned}$$

On the right hand side, the only term that we have not treated yet is

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

The proof of the Theorem will be complete if we show that all $\kappa > 0$,

$$\lim_{n\to\infty} \mathbb{E}\frac{1}{n}\left|\sum_{k=1}^n R_{k,k}^{n,\kappa} - \mathbb{E}R_{k,k}^{n,\kappa}\right| = 0. \tag{45}$$

As in the proof of Theorem 1.4 we get

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

and we conclude the proof by using (43). This ends the proof of Theorem 3.1(ii). □

3.3 Proof of Theorem 1.3: Markov matrices with $\alpha \in (1,2)$

The proof given above applies to the matrix $A_n := \kappa_n S$, where $S_{i,j} = \frac{U_{i,j}}{\sqrt{\rho_i \rho_j}}$. In particular, we may restate Theorem 2.3(ii) and Theorem 2.8(iii) as in (43), using the same notation, with the truncated matrix now given by

$$(A_n^\kappa)_{i,j} = \frac{\kappa_n \, U_{i,j} \, \mathbb{1}_{\{U_{i,j} < \kappa\}}}{\sqrt{\rho_i \rho_j}} \, .$$

The proof of Theorem 3.1 works verbatim. Note, in particular, that the rank inequality leads exactly to the same inequality as in (42). □

4 Properties of the Limiting Spectral Distributions

4.1 Proof of Theorem 1.5: μ_α, $\alpha \in (0,2)$.

Before going into the proof of Theorem 1.5, we introduce some notation. Let $\beta = \alpha/2$ and let $\mathcal{K}_\alpha$ denote the set of probability measures on $(0,\infty)$ with finite β moment. We define the mapping Ψ on probability measures on $\mathbb{R}_+ \cup \{\infty\}$, where $\Psi(Q)$ is the law of

$$Z = \left(\sum_{k \in \mathbb{N}} \xi_k Y_k \right)^{-1} , \tag{46}$$

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 4.1. *Ψ satisfies the following*

(i) Ψ is a mapping from $\mathcal{K}_\alpha$ to $\mathcal{K}_\alpha$. Let $(P_n)_{n \in \mathbf{N}}$ 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 $\mathbb{E}\exp(-tS) = \exp\left(-t^\beta \sqrt{\frac{\Gamma(1+\beta)}{\Gamma(1-\beta)}}\right)$, $t \geqslant 0$.

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

Proof of Lemma 4.1. As in the proof of Lemma 3.2, we get

$$\begin{aligned}
\mathbb{E}Z^\beta &= \mathbb{E}\left(\sum_k \xi_k Y_k \right)^{-\beta} \\
&= \mathbb{E}\frac{1}{\Gamma(\beta)} \int_0^\infty x^{\beta-1} e^{-x\sum_k \xi_k Y_k} dx \\
&= \frac{1}{\Gamma(\beta)} \int_0^\infty x^{\beta-1} e^{-x^\beta \Gamma(1-\beta)\mathbb{E}Y_1^\beta} dx \\
&= \frac{1}{\beta\Gamma(\beta)} \int_0^\infty e^{-s\Gamma(1-\beta)\mathbb{E}Y_1^\beta} ds \\
&= (\Gamma(\beta+1)\Gamma(1-\beta)\mathbb{E}Y_1^\beta)^{-1},
\end{aligned}$$

(in the last line we have used the identity $z\Gamma(z) = \Gamma(z+1)$). Therefore, Ψ is a mapping from $\mathcal{K}_\alpha$ to $\mathcal{K}_\alpha$. Also as a consequence of (38):

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

Statement (i) follows from the continuity of the mapping $x \mapsto 1/x$ 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 (38) we obtain for all $t \geqslant 0$,

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

□

Proof of Theorem 1.5(i). From Theorem 3.1, for $z \in \mathbb{C}_+$,

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

where h solves RDE (35). 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

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

and

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

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

$$g(z) \leqslant_{st} \left(v + \sum_k \xi_k g_k(z)\right)^{-1} \leqslant_{st} \left(\sum_k \xi_k g_k(z)\right)^{-1}. \tag{47}$$

(In fact, we also have $|h(z)| \leqslant_{st} \left(\sum_k \xi_k g_k(z)\right)^{-1}$). Using the computation in Lemma 4.1, we obtain $\mathbb{E}g(z)^\beta \leqslant (\Gamma(\beta+1)\Gamma(1-\beta)\mathbb{E}g(z)^\beta)^{-1}$. Thus

$$\mathbb{E}g(z)^\beta \leqslant \frac{1}{\sqrt{\Gamma(\beta+1)\Gamma(1-\beta))}}. \tag{48}$$

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

$$\mathbb{E}\left[\left(\sum_k \xi_k g_k(z)\right)^{-\eta}\right] = \frac{1}{\Gamma(\eta)}\int_0^\infty x^{\eta-1}e^{-x^\beta\Gamma(1-\beta)\mathbb{E}g(z)^\beta}dx\,. \tag{49}$$

We now study the weak limit of $g(u+iv)$ when $v \downarrow 0$, $u \in \mathbb{R}$. Equation (48) implies tightness, so let $g(u+i0)$ be a weak limit. If this limit is non–zero then $\mathbb{E}g^\beta(u+i0) > 0$, and Equations (47)-(49) imply for all $\eta > 0$ and $u \in \mathbb{R}$,

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

Since $\mathbb{E}h(z)$ is the Stieltjes transform of μ_α, taking $\eta = 1$, we deduce that μ_α is absolutely continuous with bounded density, see for example [25, Theorem 11.6]. □

Proof of Theorem 1.5(ii). In view of [25, Theorem 11.6], it is sufficient to show that

$$\lim_{t\downarrow 0} \mathbb{E}g(it) = \Gamma\left(1+\frac{1}{\beta}\right)\left(\frac{\Gamma(1+\beta)}{\Gamma(1-\beta)}\right)^{\frac{1}{2\beta}}. \tag{50}$$

As above, (48) 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 4.1, $g(i0) \stackrel{d}{=} 1/S$, and (49) gives

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

Using the identity $z\Gamma(z) = \Gamma(z+1)$, we get (50). □

Proof of Theorem 1.5 (ii). We start with a Tauberian–type theorem for the Stieltjes transform of symmetric probability measures. As usual, let m_μ denote the 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 4.2 (Tauberian–like lemma)**.** *If L is slowly varying and $0<\alpha<2$, the following are equivalent: as t goes to $+\infty$*

$$\mu((t,\infty)) \sim L(t)t^{-\alpha} \tag{51}$$

$$\Im m_\mu(it) - t^{-1} \sim -\Delta(\alpha)L(t)t^{-\alpha-1} \tag{52}$$

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

Sketch of Proof of Lemma 4.2. The proof is an adaptation of the proof of Karamata's Tauberian Theorem in Bingham, Goldie and Teugels [10, p. 37–38]. 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 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 (52) holds, namely

$$\mathcal{S}\mu(t) \sim \Delta(\alpha)L(t)t^{-\alpha}. \tag{53}$$

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

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

We get precisely (51). The reciprocal implication can be proved similarly, see [10, p. 37–38] (it is straightforward for $L(t) = c$, the case that we will actually use). □

We now come back to the RDE (36) and define $Q(t) = \mathbb{E}[g(it)^\beta]$. From (36), we have a.s. $tg(it) \leqslant 1$. Note also, from a.s. $\sum_k \xi_k g_k(it) \leqslant t^{-1}\sum_k \xi_k$, that a.s. $\lim_{t\to+\infty} tg(it) = 1$. The dominated convergence Theorem leads to

$$\lim_{t\to\infty} t^\beta Q(t) = 1. \tag{54}$$

Moreover, as already pointed in Lemma 3.2,

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

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

$$\begin{aligned} \Im m_{\mu_\alpha}(it) = \mathbb{E}g(it) &= \mathbb{E}\frac{t}{t^2 + tQ(t)^{1/\beta}\sum_k \xi_k} \\ &= C(t)\mathbb{E}\frac{tC(t)}{(tC(t))^2 + \sum_k \xi_k} \\ &= C(t)\Im m_{\mathcal{L}(Y)}(iC(t)t), \end{aligned} \tag{55}$$

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 (38), 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

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

We get by Lemma 4.2, $\Im m_{\mathcal{L}(Y)}(it) - t^{-1} \sim -\frac{t^{-\alpha-1}}{2}\Delta(\alpha)$. Thus by (54)-(55), we have

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

The conclusion follows by a reverse application of Lemma 4.2. □

Remark 4.3. *In the proof of Lemma 3.2, 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*.*

4.2 Proof of Theorem 1.6: $\widetilde{\mu}_\alpha$, $\alpha \in (0,1)$

We denote by $\mathbf{p}_{2n}(\varnothing)$ the return probability after $2n$ steps starting from the root $\varnothing$, for the random walk on the PWIT with transition kernel $\mathbf{K}$ given by (15).

Proof of Theorem 1.6 (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:

$$\gamma_{2n} \geqslant \delta\, \varepsilon^\alpha\, (1-\varepsilon)^{2n}\,. \tag{56}$$

Theorem 1.6 (i) follows by choosing $\varepsilon = 1/2n$. From (32) we have $\gamma_{2n} = \mathbb{E}[\mathbf{p}_{2n}(\varnothing)]$. To prove (56) we use the simple bound $\mathbf{p}_{2n}(\varnothing) \geqslant (\mathbf{K}(\varnothing, 1)\mathbf{K}(1, \varnothing))^n$, 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. It follows that

$$\gamma_{2n} \geqslant \mathbb{E}\left[(\mathbf{K}(\varnothing, 1)\mathbf{K}(1, \varnothing))^n\right]. \tag{57}$$

Therefore (56) holds if the event

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

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 $\varnothing$, i.e. $x_1 > x_2 > \dots$ 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,

$$\begin{aligned}
\mathbb{P}(A_\varepsilon) &= \mathbb{P}\left(x_1 \geqslant (1-\varepsilon)\phi\,,\ x_1 \geqslant (1-\varepsilon)(x_1 + \phi')\right) \\
&= \mathbb{P}\left(x_1 \geqslant (1-\varepsilon)\phi\,,\ \phi' \leqslant \frac{\varepsilon\, x_1}{(1-\varepsilon)}\right) \\
&\geqslant \mathbb{P}\left(x_1 \geqslant (1-\varepsilon)\phi\,,\ x_1 \geqslant \varepsilon^{-1}\,,\ \phi' \leqslant 1\right).
\end{aligned}$$

Let $\delta_1 := \mathbb{P}(\phi \leqslant 1) = \int_0^1 f(t)\, dt > 0$, 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 [20, Proposition 10], or (38) with β replaced by α and $Y_k = 1$). Since ϕ' is independent of (x_i) we obtain

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

To estimate the last quantity we observe that if $\widetilde{x}$ is a size-biased pick from (x_i) then $x_1 \geqslant \widetilde{x}$. We recall that $\widetilde{x}$ is a random variable such that, given the sequence (x_i) the

probability that $\widetilde{x}$ equals x_i is x_i/ϕ. It is not hard to check (see e.g. [19, Lemma 2.2]) that the random variable $\widetilde{x}$ has a probability density on $(0,\infty)$ given by

$$\alpha\, x^{-\alpha-1} \int_0^\infty f(t)\, \frac{x}{x+t}\, dt\,, \tag{58}$$

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

$$\begin{aligned}
&\mathbb{P}\left(x_1 \geqslant (1-\varepsilon)\phi\,,\ x_1 \geqslant \varepsilon^{-1}\right) \geqslant \mathbb{P}\left(\widetilde{x} \geqslant (1-\varepsilon)\phi\,,\ \widetilde{x} \geqslant \varepsilon^{-1}\right) \\
&\quad = \alpha \int_0^\infty dt\, f(t) \int_0^\infty dx\, x^{-\alpha-1}\, \frac{x}{x+t}\, \mathbb{1}_{\{x \geqslant (1-\varepsilon)(x+t)\}}\, \mathbb{1}_{\{x \geqslant \varepsilon^{-1}\}} \\
&\quad \geqslant \alpha \int_0^1 dt\, f(t) \int_0^\infty dx\, x^{-\alpha-1}\, (1-\varepsilon)\, \mathbb{1}_{\{x \geqslant \varepsilon^{-1}\}} \\
&\quad = \delta_1\, (1-\varepsilon)\, \varepsilon^\alpha\,.
\end{aligned}$$

In conclusion, $\mathbb{P}(A_\varepsilon) \geqslant \delta_1^2\, (1-\varepsilon)\, \varepsilon^\alpha \geqslant \frac{1}{2}\, \delta_1^2\, \varepsilon^\alpha$, and the claim (56) follows.

It remains to show that $\liminf_{\alpha \nearrow 1} \gamma_2 > 0$. Actually, if (x_i), $\widetilde{x}$, 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

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

Now the identity

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

gives

$$\begin{aligned}
\gamma_2 &= \alpha\Gamma(2-\alpha) \int_0^\infty \int_0^\infty e^{-\Gamma(1-\alpha)(u^\alpha+v^\alpha)}\, (u+v)^{-2+\alpha}\, du\, dv \\
&= \frac{\alpha\Gamma(2-\alpha)}{\Gamma(1-\alpha)} \int_0^\infty \int_0^\infty e^{-t^\alpha - s^\alpha}\, (t+s)^{-2+\alpha}\, ds\, dt\,.
\end{aligned}$$

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

$$\int_0^\infty \int_0^\infty e^{-t^\alpha - s^\alpha}\, (t+s)^{-2+\alpha}\, ds\, dt \geqslant e^{-2} \int_0^1 \int_0^1 (t+s)^{-2+\alpha}\, ds dt \geqslant \frac{c_1}{1-\alpha}.$$

and

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

□

Proof of Theorem 1.6 (ii). It is convenient to make here the dependence over α explicit in all the notations. In particular, for every $\alpha \in (0,1)$, we denote by $\mathbf{S}_\alpha$ the operator $\mathbf{S}$ given by (16). These operators are defined on a common probability space, and are self-adjoint in $L^2(V)$. Moreover, it follows from Subsection 3.1 that

$$\tilde{\mu}_\alpha = \mathbb{E}\mu_{\alpha,\varnothing},$$

where $\mu_{\alpha,\varnothing}$ is the spectral measure of $\mathbf{S}_\alpha$ at vector $\delta_\varnothing$. By the dominated convergence Theorem, in order to prove that $\alpha \mapsto \tilde{\mu}_\alpha$ is continuous in $(0,1)$, it is sufficient to show that a.s. $\alpha \mapsto \mu_{\alpha,\varnothing}$ is continuous. From Theorem VIII.25(a) in Reed and Simon [22], 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 mapping from $(0,1)$ to $L^2(V)$. From (16), for all $\mathbf{u} \in V$, the mapping $\alpha \mapsto \mathbf{S}_\alpha(\mathbf{u},\mathbf{v})$ 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

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

Then notice that for all $\alpha \in (0, 1-\varepsilon)$, $y_{\mathbf{v}k}^{-1/\alpha} \leqslant \max(1, y_{\mathbf{v}k}^{-1/(1-\varepsilon)})$ and $\rho_\alpha(\mathbf{v}) \geqslant \min(1, y_{\mathbf{v}1}^{-1/(1-\varepsilon)})$. We may conclude on the uniform square integrability by recalling that a.s. $\lim_k y_{\mathbf{v}k}/k = 1$ and $y_{\mathbf{v}1} > 0$. □

Proof of Theorem 1.6 (iii). As in the proof of Theorem 1.6 (ii), we make here the dependence over α explicit in all the notations. It follows from Subsection 3.1

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

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

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

By construction $V_\alpha(\mathbf{v})$ is a $\mathrm{PD}(\alpha, 0)$ random variable. Thus, by Corollary 18 in Pitman and Yor [20], as $\alpha \downarrow 0$, $V_\alpha(\mathbf{v})$ converge weakly to the deterministic vector $(1, 0, \cdots)$. 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 i.e. $\mathbb{P}(U=0) = \mathbb{P}(U=1) = 1/2$. 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, we get

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

Since $\mu_{\alpha,\varnothing}$ is symmetric, we finally have

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

Taking expectation, we obtain the claimed statement on $\tilde{\mu}_\alpha$. □

5 Invariant Measure: Proof of Theorem 1.7

We start with a lemma. Let $(X_1, \ldots, X_n)$, $X_1 \geqslant \cdots \geqslant X_n$, 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 of Lemma 5.1. There are $m_n = n(n+1)/2$ edges, including self–loops. Let us denote by U_e the weight of edge $e \in \{1, \dots, m_n\}$. 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, i.e. edges of the form $e = \{i, j\}$ with $i \neq j$. Let $U_{e_1} \geqslant U_{e_2} \geqslant \cdots$ 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

$$b_n^{-1}(U_{e_1}, U_{e_2}, \dots) \xrightarrow[n\to\infty]{d} (x_1, x_2, \dots)\,. \tag{59}$$

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 $\varepsilon > 0$:

$$\lim_{n\to\infty} \mathbb{P}\left(\max_{1 \leqslant \ell \leqslant k} \Delta_\ell \geqslant \varepsilon\, b_n \right) = 0\,. \tag{60}$$

To prove this we use an estimate due to Soshnikov [26]. Let B_n denote the event that there exists no $i \in \{1, \dots, n\}$ such that $\{\rho_i > b_n^{\frac{3}{4}+\frac{\alpha}{8}}$ and $\rho_i - \max_j U_{i,j} > b_n^{\frac{3}{4}+\frac{\alpha}{8}}\}$. Then, see [26] and [4, Lemma 3], one has

$$\lim_{n\to\infty} \mathbb{P}(B_n) \to 1\,. \tag{61}$$

Clearly, on the event B_n, if $\max_{1 \leqslant \ell \leqslant k} \Delta_\ell \geqslant \varepsilon\, b_n$, then $U_{e_k} \leqslant b_n^{\frac{3}{4}+\frac{\alpha}{8}}$ which has vanishing probability in the limit by (59). This proves (60).

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 (60) and (59) prove that

$$b_n^{-1}(R_1, R_2, R_3, R_4 \dots) \xrightarrow[n\to\infty]{d} (x_1, x_1, x_2, x_2, \dots)\,. \tag{62}$$

It remains to show that for every fixed k:

$$\lim_{n\to\infty} \mathbb{P}\left(\cup_{1 \leqslant i \leqslant 2k} \{R_i \neq X_i\}\right) = 0\,. \tag{63}$$

By construction, we have $X_i \geqslant R_i$ 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 $e \neq e_1$ such that $U_e \geqslant U_{e_1} - b_n^{\frac{3}{4}+\frac{\alpha}{8}}$. However, this event has vanishing probability by (59) 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 (61) this shows that $\mathbb{P}(X_1 > R_1 \text{ or } X_2 > R_2) \to 0$. Recursively, the probability of $X_{2i+1} > R_{2i+1}$ or $X_{2i+2} > R_{2i+2}$ on the event $B_n \cap \{X_j = R_j,\ \forall j = 1, \dots, 2i\}$ 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 (63) as required. □

Proof of Theorem 1.7(i). Let us define $m_n = n(n+1)/2$. Observe that

$$\sum_{i=1}^{n} \rho_i = 2S_n + D_n \quad \text{where} \quad S_n := \sum_{e \in O_n} U_e \quad \text{and} \quad D_n := \sum_{i=1}^{n} U_{i,i}\,. \tag{64}$$

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

$$\lim_{n\to\infty} \frac{1}{m_n} \sum_{i=1}^{n} \rho_i = 2 \quad \text{in probability.} \tag{65}$$

Theorem 1.7(i) follows directly from Lemma 5.1 and (65). □

Proof of Theorem 1.7(ii). If $U_{e_1} \geqslant U_{e_2} \geqslant \cdots$ 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 (64) then by Lemma 2.4(ii), replacing n with m_n, we have

$$\left(\frac{U_{e_1}}{S_n}, \frac{U_{e_2}}{S_n}, \dots \right) \xrightarrow[n \to \infty]{d} \left(\frac{x_1}{\sum_{i=1}^{\infty} x_i}, \frac{x_2}{\sum_{i=1}^{\infty} x_i}, \dots \right) \tag{66}$$

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 $X_1, X_2, \dots$ 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 (64). Let

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

Thanks to (66) it is sufficient to prove that $\mathbb{P}(\max_{1 \leqslant i \leqslant 2k} |Y_i| > \varepsilon) \to 0$, as $n \to \infty$, for any fixed $\varepsilon > 0$ and $k \in \mathbb{N}$. This follows from the argument used in the proof of (60) and (63). □

Acknowledgements. P.C. was partially supported by NSF Grant DMS-0301795.

References

[1] D. Aldous, *Asymptotics in the random assignment problem*, Probab. Theory Related Fields **93** (1992), no. 4, 507–534. 4, 8, 13, 14

[2] D. Aldous and R. Lyons, *Processes on unimodular random networks*, Electron. J. Probab. **12** (2007), no. 54, 1454–1508 (electronic). 14

[3] D. Aldous and J. M. Steele, *The objective method: probabilistic combinatorial optimization and local weak convergence*, Probability on discrete structures, Encyclopaedia Math. Sci., vol. 110, Springer, Berlin, 2004, pp. 1–72. 4, 14

[4] A. Auffinger, G. Ben Arous, and S. Péché, *Poisson convergence for the largest eigenvalues of Heavy Tailed Random Matrices*, preprint `arXiv:0710.3132 [math.PR]` to appear in Annales de l'Institut Henri Poincaré – Probabilités et Statistique, 2007. 7, 34

[5] S. Belinschi, A. Dembo, and Guionnet A., *Spectral measure of heavy tailed band and covariance random matrices*, preprint `arXiv:0811.1587v2 [math.PR]` to appear in Communications in Mathematical Physics, 2009. 4, 5, 6

[6] G. Ben Arous and J. Černý, *The arcsine law as a universal aging scheme for trap models*, Comm. Pure Appl. Math. **61** (2008), no. 3, 289–329. 7

[7] G. Ben Arous and A. Guionnet, *The spectrum of heavy tailed random matrices*, Comm. Math. Phys. **278** (2008), no. 3, 715–751. 4, 5, 6, 20, 21, 25

[8] I. Benjamini and O. Schramm, *Recurrence of distributional limits of finite planar graphs*, Electron. J. Probab. **6** (2001), no. 23, 13 pp. (electronic). 14

[9] J. Bertoin, *Random fragmentation and coagulation processes*, Cambridge Studies in Advanced Mathematics, vol. 102, Cambridge University Press, Cambridge, 2006. 11

[10] N. H. Bingham, C. M. Goldie, and J. L. Teugels, *Regular variation*, Encyclopedia of Mathematics and its Applications, vol. 27, Cambridge University Press, Cambridge, 1989. 29, 30

[11] Ch. Bordenave, P. Caputo, and D. Chafaï, *Spectrum of large random reversible Markov chains*, preprint `arXiv:0811.1097 [math.PR]`, 2008. 2, 3, 6, 7, 9

[12] J. Bouchaud and P. Cizeau, *Theory of Lévy matrices*, Phys. Rev. E **3** (1994), 1810–1822. 4

[13] A. Bovier and A. Faggionato, *Spectral characterization of aging: the REM-like trap model*, Ann. Appl. Probab. **15** (2005), no. 3, 1997–2037. 7

[14] W. Feller, *An introduction to probability theory and its applications. Vol. II.*, Second edition, John Wiley & Sons Inc., New York, 1971. 3, 4

[15] L. R. G. Fontes and P. Mathieu, *K-processes, scaling limit and aging for the trap model in the complete graph*, Ann. Probab. **36** (2008), no. 4, 1322–1358. MR MR2435851 7

[16] Bouchaud J.-P., *Weak ergodicity breaking and aging in disordered systems*, J. Phys. I France **2** (1992), 1705–1713. 7

[17] A. Klein, *Extended states in the Anderson model on the Bethe lattice*, Adv. Math. **133** (1998), no. 1, 163–184. 24

[18] R. LePage, M. Woodroofe, and J. Zinn, *Convergence to a stable distribution via order statistics*, Ann. Probab. **9** (1981), no. 4, 624–632. 12

[19] M. Perman, J. Pitman, and M. Yor, *Size-biased sampling of Poisson point processes and excursions*, Probab. Theory Related Fields **92** (1992), no. 1, 21–39. 32

[20] J. Pitman and M. Yor, *The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator*, Ann. Probab. **25** (1997), no. 2, 855–900. MR MR1434129 (98f:60147) 11, 31, 33

[21] M. Reed and B. Simon, *Methods of modern mathematical physics. II. Fourier analysis, self-adjointness*, Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1975. 23

[22] ———, *Methods of modern mathematical physics. I*, second ed., Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1980, Functional analysis. 8, 9, 19, 22, 32

[23] S. I. Resnick, *Heavy-tail phenomena*, Springer Series in Operations Research and Financial Engineering, Springer, New York, 2007, Probabilistic and statistical modeling. 3, 12

[24] E. Seneta, *Non-negative matrices and Markov chains*, Springer Series in Statistics, Springer, New York, 2006, Revised reprint of the second (1981) edition [Springer-Verlag, New York; MR0719544]. MR MR2209438 2

[25] B. Simon, *Trace ideals and their applications*, second ed., Mathematical Surveys and Monographs, vol. 120, American Mathematical Society, Providence, RI, 2005. 28, 29

[26] A. Soshnikov, *Poisson statistics for the largest eigenvalues of Wigner random matrices with heavy tails*, Electron. Comm. Probab. **9** (2004), 82–91 (electronic). 7, 34

[27] M. Talagrand, *Spin glasses: a challenge for mathematicians*, vol. 46, Springer-Verlag, Berlin, 2003, Cavity and mean field models. 21

[28] I. Zakharevich, *A generalization of Wigner's law*, Comm. Math. Phys. **268** (2006), no. 2, 403–414. 4, 23, 24

Compiled March 20, 2009

Charles Bordenave, E-mail: charles.bordenave(AT)math.univ-toulouse.fr
CNRS UMR5219 & Institut de Mathématiques de Toulouse
Université Toulouse III, France.

Pietro Caputo, E-mail: caputo(AT)mat.uniroma3.it
Dipartimento di Matematica
Università Roma Tre, Italy.

Djalil Chafaï, E-mail: djalil(AT)chafai(DOT)net
INRA UMR181 ENVT & Institut de Mathématiques de Toulouse
Université de Toulouse, France.

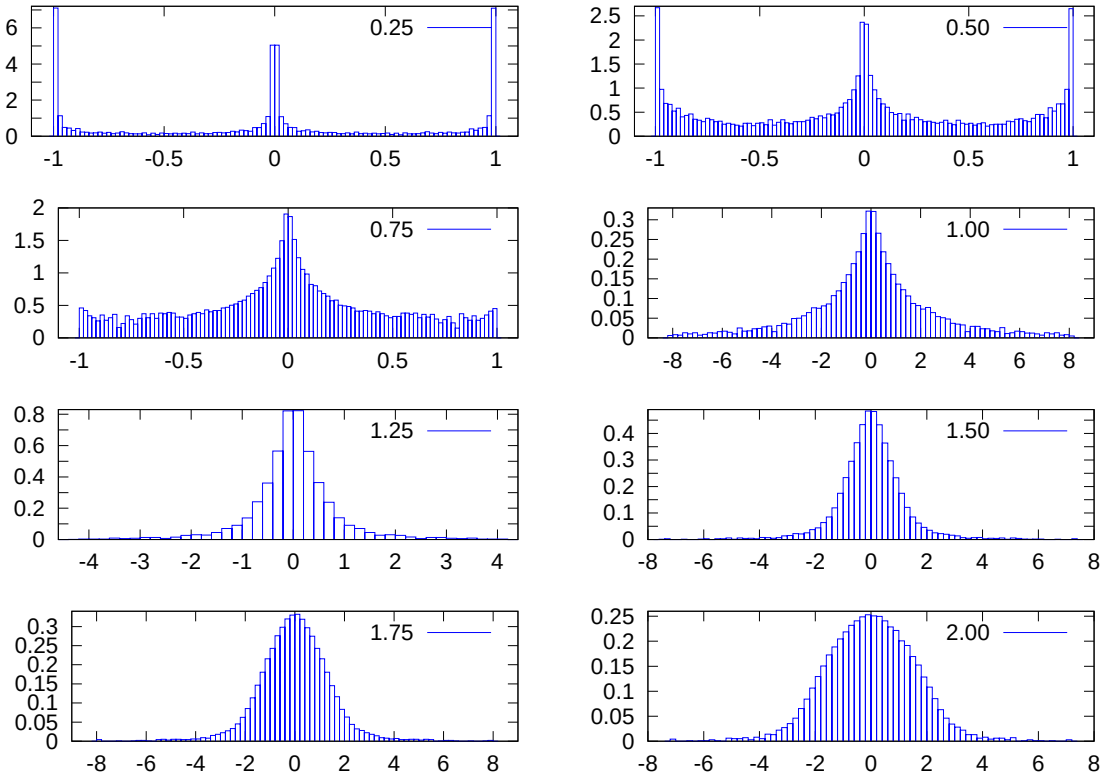

Figure 1: Histograms of scaled ESDs illustrating the convergence stated by theorems 1.3 and 1.4, for eight different values of α, namely 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00. Here $n = 5000$ and $\mathcal{L}$ is the law of $V^{-1/\alpha}$ where V is a uniform random variable on the interval $(0, 1)$. The first three plots are the histogram of the spectrum of a single realization of K. The fourth plot, which corresponds to $\alpha = 1$, 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 the eigenvalues $\lambda_{\lfloor \log(n) \rfloor}, \ldots, \lambda_{\lfloor n - \log(n) \rfloor}$ were used.