# Tree Walks and the Spectrum of Random Graphs

**Eva-Maria Hainzl**
TU Wien, Austria

**Élie de Panafieu**
Nokia Bell Labs, Nozay, France

## Abstract

It is a classic result in spectral theory that the limit distribution of the spectral measure of random graphs $G(n,p)$ converges to the semicircle law in case $np$ tends to infinity with $n$. The spectral measure for random graphs $G(n,c/n)$ however is less understood. In this work, we combine and extend two combinatorial approaches by Bauer and Golinelli (2001) and Enriquez and Menard (2016) and approximate the moments of the spectral measure by counting walks that span trees.

**2012 ACM Subject Classification** Mathematics of computing → Generating functions; Mathematics of computing → Spectra of graphs

**Keywords and phrases** Spectrum of random matrices, generating functions

**Digital Object Identifier** 10.4230/LIPIcs.AofA.2024.11

**Related Version** *Full Version*: `http://arxiv.org/abs/2405.08347` [12]

**Funding** This work was supported by the RandNET project, MSCA-RISE - Marie Skłodowska-Curie Research and Innovation Staff Exchange Programme (RISE), Grant agreement 101007705.
*Eva-Maria Hainzl*: This work was partially founded by the Austrian Science Foundation FWF, projects F50-02, F55-02.

**Acknowledgements** We thank Nicolas Curien and Laurent Ménard for encouraging us to work on this topic, their availability and their insights on the spectrum of random graphs. We also thank the Institut de Recherche en Informatique Fondamentale (IRIF), Université Paris Cité, for hosting us.

## 1 Introduction

Random matrix theory studies the spectrum of random matrices and has found many applications, including in physics [22], wireless communication [19] and numerical analysis [7]. A fundamental result of this field is that the limit distribution of the spectral measure of so-called Wigner matrices converges to the semicircle law [20, 21] and it is worth mentioning that a common proof of this theorem by the moment method relies on counting closed walks on trees (*e.g.* [10]). This universal law has been extended to several other classes, such as adjacency matrices of random regular graphs [13, 18] and Erdős-Rényi random graphs $G(n,p)$ when $pn \to \infty$. In particular, Bauer and Golinelli [1] pointed out the importance of the spectral measure of adjacency matrices of random graphs and explained how to compute the moments by counting walks on trees. Zakharevich [24] picked up on the approach and showed further that the spectral distribution of $G(n,c/n)$ converges to a limit distribution $\mu^c$ which has infinite support. However, for $p = c/n$, several technical conditions of classic theorems in probability theory are not met such that one could apply standard techniques and despite recent progress [3, 15, 4, 16, 6], $\mu^c$ remains an enigma. In [8], Enriquez and Ménard returned to combinatorial methods and computed several terms of the asymptotic expansion, as $c$ tends to infinity, of the moments of the normalized spectral measures

$$\mu_n^c = \frac{1}{n} \sum_{\lambda \in Sp(c^{-1/2} A(G(n,c/n)))} \delta_\lambda$$


35th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2024).
Editors: Cécile Mailler and Sebastian Wild; Article No. 11; pp. 11:1–11:15
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

where $A(G(n, c/n))$ is the adjacency matrix of a random graph $G(n, c/n)$. We go along the steps in the computation of moments of this measure for clarity and start with

$$m_\ell(\mu_n^c) = \sum_G \mathbb{P}\left[G\left(n, \frac{c}{n}\right) = G\right] \cdot \frac{1}{n} \sum_{\lambda \in Sp\left(c^{-1/2}A(G)\right)} \lambda^\ell.$$

This formulation reduces to counting closed walks in $G$, since the sum of the eigenvalues to the power $\ell$ is just the trace of the matrix to the power $\ell$, and a value $\left(A^\ell\right)_{i,i}$ on the diagonal of this matrix is the number of closed walks of length $\ell$ starting at the vertex $i$. That is,

$$\sum_{\lambda \in Sp\left(c^{-1/2}A(G)\right)} \lambda^\ell = \left(\frac{1}{c}\right)^{\ell/2} \operatorname{tr}\left(A(G)^\ell\right) = \sum_{\text{closed walk } (v_1, v_2, \dots v_\ell, v_1) \in G} \left(\frac{1}{c}\right)^{\ell/2}.$$

Thus, the moment equals

$$m_\ell(\mu_n^c) = \frac{1}{n}\frac{1}{c^{\ell/2}} \sum_{(v_1, v_2, \dots v_\ell) \in [n]^\ell} \mathbb{E}\left[X_{v_1,v_2} X_{v_2,v_3} \cdots X_{v_\ell,v_1}\right],$$

where $X_{v_i,v_j}$ is the random variable taking the value 1 if the edge $(v_i, v_j)$ is in the graph and 0 otherwise. Observe that if a closed walk $(v_1, \dots, v_\ell, v_1)$ contains $e$ distinct edges, then $\mathbb{E}\left[X_{v_1,v_2} X_{v_2,v_3} \cdots X_{v_\ell,v_1}\right] = (c/n)^e$. The number of closed walks on $[n]$ of length $\ell$ with $m$ vertices is bounded by $n^m m^\ell$. Since the total number of vertices is bounded by the length, we have $n^m m^\ell \le n^m \ell^\ell$. The contribution to the moment of all such closed walks containing $e$ distinct edges is bounded by

$$\frac{1}{n}\frac{1}{c^{\ell/2}} n^m \ell^\ell \left(\frac{c}{n}\right)^e = c^{e-\ell/2} \ell^\ell n^{m-e-1}.$$

We are considering a fixed moment $\ell$, so this tends to 0 with $n$ whenever $m < e + 1$, that is, whenever the graph (necessarily connected) induced by the closed walk is not a tree. In particular, when $\ell$ is odd, the induced graph cannot be a tree, so the moment of order $\ell$ tends to 0.

Let $w_{m,2\ell}$ denote the number of closed walks of length $2\ell$ spanning a tree with $m$ vertices. We now consider the even moment of order $2\ell$ and split the sum according to the number $m$ of distinct vertices in the closed walk

$$m_{2\ell}(\mu_n^c) = \frac{1}{n}\frac{1}{c^\ell} \sum_{m=1}^{\ell+1} \binom{n}{m} \left(\frac{c}{n}\right)^{m-1} w_{m,2\ell}.$$

Let us define the limit distribution $\mu^c = \lim_{n \to +\infty} \mu_n^c$. Then its odd moments are zero and its moment of order $2\ell$ is

$$m_{2\ell}(\mu^c) = \lim_{n \to +\infty} \frac{1}{n}\frac{1}{c^\ell} \sum_{m=1}^{\ell+1} \binom{n}{m} \left(\frac{c}{n}\right)^{m-1} w_{m,2\ell} = \sum_{m=1}^{\ell+1} \frac{1}{c^{\ell-m+1}} \frac{w_{m,2\ell}}{m!}.$$

By identifying the generating functions of $(w_{m,2\ell})_{\ell \ge 0}$, for $m = \ell+1$ and $m = \ell$, as the Stieltjes transform of a specific measure, Enriquez and Ménard were able to derive an approximation of the moments of the limit law and computational experiments showed that even the density of this measure approximated the shape of the histograms of eigenvalues of sampled matrices quite well. An extension of this approximation to the order $c^{-2}$ took considerable effort on several sides, including the combinatorics of closed walks on trees.

The aim of this paper is to provide further insight into what we call *tree walks*, and consequently an efficient way to compute the numbers $w_{m,2\ell}$, for all $2\ell \geq 0$ and $0 \leq m \leq \ell+1$, and their generating functions. But as we delve further into their connection with the spectral measure, we come across surprising and beautiful identities involving the generating function of the Catalan numbers.

Section 2 presents the formal definition of various tree walk families and our main results, which are Theorem 3 and Theorem 4. Theorem 3 expresses the generating function of tree walks as a rational function of the Catalan generating function. Theorem 4 gives several error terms for an asymptotic approximation of $\mu^c$ as $c$ tends to infinity. We also presents Conjecture 5, which states that this asymptotic approximation could be extended to an arbitrary order, turning it into a form of asymptotic expansion. This paper contains only the main steps of the proofs, a complete version being available on arxiv [12]. The main steps of the proof of Theorem 3 and Theorem 4 are given respectively in Sections 3 and 4. Numerical experiments are provided in Section 5.

## 2 Main results

Before we state our main results, let us clarify some definitions.

▶ **Definition 1** (Tree walks). *A* tree walk *of size $m$ is a walk on the complete labeled graph of size $m$ that visits every node, starts and ends at the same node, and induces a tree. More formally, a tree walk $W = (v_1, v_2, \dots, v_\ell)$ is a sequence of $v_i \in [m]$ such that*

$$V := \bigcup_{j \in [\ell]} \{v_j\}, \quad E := \bigcup_{j \in [\ell-1]} \{(v_j, v_{j+1})\} \cup \{(v_\ell, v_1)\}$$

*define a labelled tree $T(W)$ with vertex set $V = [m]$ and edge set $E$. Further, we define $v_1$ to be the root of the induced tree $T(W)$. Thus, we talk freely about the root and leaves of $W$, when referring to the root and leaves of $T(W)$. We further stick to the convention that if the root has degree 1 it is also a leaf of $T(W)$.*
*In the following, we study the number $w_{m,2\ell}$ of tree walks of length $2\ell$ that span a tree of size $m$ and the generating function*

$$W(v, z) = \sum_{\ell, m \geq 0} w_{m,2\ell} \frac{v^m}{m!} z^\ell.$$

*Since a walk of length $2\ell$ spans a tree with at most $\ell+1$ vertices, we have $w_{m,2\ell} = 0$ if $\ell < m-1$ and for $\ell \geq 1$, we define $w_{0,2\ell} = w_{1,2\ell} = 0$ and $w_{1,0} = 1$, $w_{0,0} = 0$.*

The ordinary generating function of the moments of $\mu^c$ is therefore given by

$$M_{\mu^c}(z) = \sum_{\ell \geq 0} m_{2\ell}(\mu^c) z^{2\ell} = \sum_{\ell \geq 0} \sum_{m=0}^{\ell+1} w_{m,2\ell} \frac{c^m}{m!} c^{-\ell-1} z^{2\ell} = \frac{1}{c} W\left(c, \frac{z^2}{c}\right), \tag{1}$$

where $m_0(\mu^c) = 1$ as always. However, we could have restructured $M_{\mu^c}(z)$ like Enriquez and Ménard in [8] as well. We just sum over the negative exponent $\xi = \ell + 1 - m$ of $c$ such that

$$M_{\mu^c}(z) = \sum_{\ell \geq 0} m_{2\ell}(\mu^c) z^{2\ell} = \sum_{\ell \geq 0} \sum_{m \geq 0} w_{m,2\ell} \frac{c^m}{m!} c^{-\ell-1} z^{2\ell} = \sum_{\ell \geq 0} \sum_{\xi \geq 0} \frac{w_{\ell-\xi+1,2\ell}}{(\ell-\xi+1)!} \frac{z^{2\ell}}{c^\xi}. \tag{2}$$

This expansion in turn motivates the following definition.

▶ **Definition 2** (Excess of a tree walk). *If an edge is traversed $2k$ times in a tree walk, then the* excess *of the edge $e$ is defined as $\xi(e) = k - 1$. The* excess *of a tree walk $W$ is the sum over the excess of all edges in the induced tree $T(W) = (V, E)$. Hence, it is half its length minus the number of edges of the tree, $\xi(W) = \ell - |E|$. An edge with positive excess is called an* excess edge *and an edge without excess a* simple edge. *We denote the generating function of tree walks with excess $\xi$ by*

$$W_\xi(z) = \sum_{\ell \geq 0} \frac{w_{\ell-\xi+1,2\ell}}{(\ell - \xi + 1)!} z^\ell,$$

*where $w_{m,2\ell}$ is the number of tree walks of length $2\ell$ that span a tree of size $m$.*

Thus, the relation between the generating functions we have defined so far is

$$M_{\mu^c}(z) = \frac{1}{c} W\left(c, \frac{z^2}{c}\right) = \sum_{\xi \geq 0} \frac{1}{c^\xi} W_\xi\left(z^2\right).$$

Bauer and Golinelli [1] introduced in the sequence $w_{m,2\ell}$ an additional parameter $d$ counting the number of times the walk leaves the root. This approach allowed them to compute the values $w_{m,2\ell}$ for $2\ell$ and $m$ up to 120 [17], and they conjectured a particular form for $w_{m,2\ell}$ that we prove in the next theorem. When we translate this decomposition in generating functions, an equation for $W(x, v, z)$ is obtained, where $x$ marks the parameter $d$. Unfortunately, this equation is not particularly amenable to classic analysis with complex analytic methods as it involves a Laplace transform. Our approach on the other hand is reminiscent of the decomposition of graphs with given excess by Wright [23]. Not only do we prove a well founded recursion in $z$ and $v$, but we provide more insight into the structure of tree walks and their generating function. Most importantly, we compute closed expressions for $w_{m,2\ell}, \ell \geq 0$ for fixed (small) $m$ and prove a conjecture from [1].

▶ **Theorem 3.** *Let $C(z) = \frac{1-\sqrt{1-4z}}{2z}$ denote the generating function of the Catalan numbers, and $W_\xi(z)$ denote the generating function of tree walks of excess $\xi$ from Definition 2. Then $W_0(z) = C(z)$ and for any $\xi \geq 1$, there are polynomials $(K_{\xi,s}(x))_{0 \leq s \leq 2\xi-2}$ with non-negative coefficients of degree $2\xi + s$ such that*

$$W_\xi(z) = C(z) \sum_{s=0}^{2\xi-2} \frac{K_{\xi,s}\left(zC(z)^2\right)}{\left(1 - zC(z)^2\right)^{s+1}}.$$

*In particular, denoting by $\operatorname{Cat}(n)$ the $n$-th Catalan number, we have*

$$K_{\xi,2\xi-2}(x) = \operatorname{Cat}(\xi - 1) x^{4\xi-2} \qquad \text{and} \qquad K_{\xi,2\xi-3}(x) = (3\xi - 1) \operatorname{Cat}(\xi - 1) x^{4\xi-3}.$$

We establish a recursion for the polynomials $K_{\xi,s}(x)$ in Section 3. This enables the successive computation of three quantities. First, the generating function $W_\xi(z)$ for any $\xi$, then the series $M_{\mu^c}(z)$ up to an arbitrary degree in $c$, given sufficient computational power, and finally the moments $m_{2\ell}$.

Our next theorem significantly extends Theorem 3 from [8]. It approximates $\mu^c$ for large $c$. There are many notions of convergence for measures. The one we consider here is the convergence of all moments (restricting to the even ones since the odd ones vanish). Further, when looking at the limit of a sequence of random variables, it is common to rescale them by their mean and standard deviation. Here, the rescaling takes the form of a dilation operator $\Lambda_\alpha$, for $\alpha > 0$. This operator transforms a measure $\mu$ into the measure $\Lambda_\alpha(\mu)$ satisfying for every Borel set $A$

$$\Lambda_\alpha(\mu)(A) = \mu(A/\alpha).$$

▶ **Theorem 4.** *Let $m_\ell(\mu)$ denote the $\ell$-th moment of a measure $\mu$ and $\Lambda_\alpha$ the dilation operator defined above. Then as $c \to \infty$, it holds for all $\ell \geq 0$ that*

$$m_{2\ell}(\mu^c) = m_{2\ell}\left(\Lambda_{f(1/c)}\left(\sigma + \sum_{i=1}^{5} \frac{1}{c^i}\sigma_i\right)\right) + \mathcal{O}\left(\frac{1}{c^6}\right)$$

*where $f(1/c) = 1 + \frac{1}{2c} + \frac{3}{8c^2} + \frac{29}{16c^3} + \frac{1987}{128c^4} + \frac{47247}{256c^5}$, $\sigma$ is the semicircle law and all $\sigma_1, \sigma_2, \ldots \sigma_5$ are signed measures explicitly given in Section 5, with total mass 0.*

This approximation entails some curious identities concerning the generating functions $W_\xi(z)$ and prompts us to state the following conjecture which we discuss in more detail in Section 4.

▶ **Conjecture 5.** *Let $M_{\mu^c}(z)$ be the ordinary moment generating function of $\mu^c$ as defined in (1). Then there exists a unique power series $P(x)$ with non-negative integer coefficients such that all $V_i(z)$ which are given by*

$$V_i(z) := \left[c^{-i}\right] M_{\mu^c}\left(\sqrt{\frac{z}{P(1/c)}}\right), \quad i \geq 0$$

*are the product of $C\left(z\right)$ and a polynomial in $zC\left(z\right)^2$.*

Let us denote by $f_k(x)$ the truncation of order $k$ of $\sqrt{P(x)}$. If the previous conjecture holds, there exist signed measures $\sigma_1, \ldots, \sigma_k$ explicitly computable from $V_1(z), \ldots, V_k(z)$ such that for any $\ell$, the moment of order $2\ell$ of $\mu^c$ is

$$m_{2\ell}(\mu^c) = m_{2\ell}\left(\Lambda_{f_k(1/c)}\left(\sigma + \sum_{i=1}^{k} c^{-i}\sigma_i\right)\right) + \mathcal{O}(c^{-k-1}).$$

Thus, Conjecture 5 provides a form of asymptotic expansion for $\mu^c$ as $c$ tends to infinity.

## 3 Decomposition of tree walks

Our proof of Theorem 3 involves reducing a tree walk with excess $\xi$ by most of its simple edges to its *kernel walk* and subsequently reversing the contraction by blowing it up to an arbitrary tree walk with excess $\xi$. The following subsection is focused on this decomposition process and the subsequent subsection on the proof of Theorem 3 and a recursion enumerating kernel walks.

### 3.1 Kernel walks

Recall that an edge of a tree walk $W$ is *simple* if it is traversed exactly twice, and is an *excess edge* otherwise.

▶ **Definition 6** (Kernel walks). *Given a tree walk $W$, we define the* kernel *of the tree walk or simply the* kernel walk *$W_K$ as the tree walk we obtain by the following procedure.*

1. *Set $W' = W$ and let $T(W')$ be its induced tree.*
2. *While there exists a simple edge $e$ incident to a leaf in $T(W')$ which is not the root, delete both occurrences of $e$ in $W'$.*
3. *While the root $u$ of the tree is a leaf and incident to a simple edge $\{u, v\}$, delete this edge in $W'$ and choose $v$ as the root of $T(W')$.*

4. *While there exists a vertex $v$ in $T(W')$ that is not the root and only incident to two simple edges $e_i = e_{j+1} = \{u, v\}$ and $e_{i+1} = e_j = \{v, w\}$, replace both consecutive pairs $e_i, e_{i+1}$ and $e_j, e_{j+1}$ with $\{u, w\}$ in $W'$.*
5. *Set $W_K = W'$.*

*Naturally, a tree walk $W$ with kernel $W_K = W$ is itself called a kernel walk. Further, we define $k_{\xi,s,2\ell}$ to be the number of kernel walks of length $2\ell$ with excess $\xi$, where the induced tree has $s$ simple edges and we define the corresponding generating function*

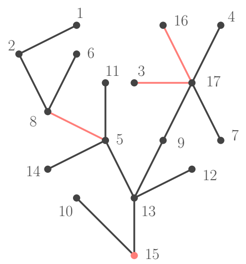
$$K_\xi(u, v, z) = \sum_{s,\ell \geq 0} k_{\xi,s,2\ell}\, u^s \frac{v^{\ell-\xi+1}}{(\ell - \xi + 1)!} z^\ell,$$

*where $u$ counts the number of simple edges, $v$ the number of vertices in the induced tree and $z$ the half-length of the walk.*

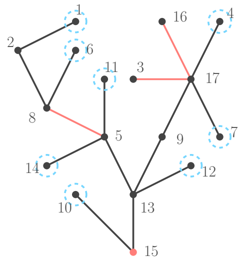
This procedure is illustrated below. Note that the variable $v$ in the generating function of kernel walks is superfluous since its exponent is fully determined by the length and the excess of the walk. However, we choose to keep it to explain the factorial in the denominator. If we consider the generating function $K(u, v, z) = \sum_{\xi \geq 0} K_\xi(u, v, z)$, we can reconstruct the individual generating functions by

$$K_\xi(u, v, z) = \left[y^{\xi-1}\right] K\left(u, \frac{v}{y}, yz\right).$$

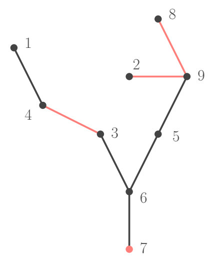
**Example.** Reducing a tree walk $W$ and its induced tree $T(W)$ to its kernel.



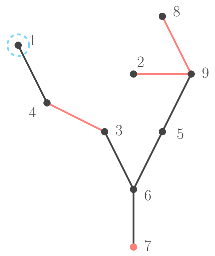

**(a)** Step 1: Set $T(W')$. Excess edges and the root in $T(W')$ are marked red.



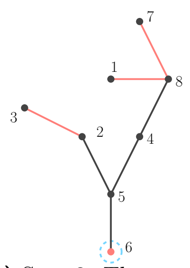

**(b)** Step 2: Identify all leaves which are not incident to an excess edge.



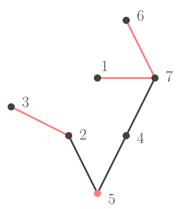

**(c)** Step 2: Remove blue vertices and update $T(W')$ by relabeling the vertices.



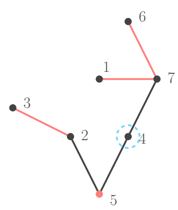

**(d)** Repeat Step 2.



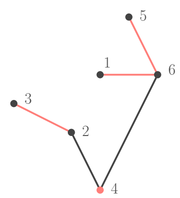

**(e)** Step 3: The root is a leaf.




**(f)** Step 3: Choose new root and relabel vertices.




**(g)** Step 4: Find adjacent simple edges.




**(h)** Step 4: Update $T(W')$ by deleting vertex 4 and relabeling the vertices.

Tree walks of a given excess $\xi$ can be arbitrarily large. However, our next result establishes that there are only finitely many kernel walks of excess $\xi$. This is reminiscent of the result of Wright [23] on the enumeration of connected graphs.

▶ **Lemma 7.** *Let $W_K$ be a kernel walk with excess $\xi$. Then its induced tree $T(W_K) = (V, E)$ satisfies $|V| \leq 3\xi - 1$ and the number of its simple edges is at most $2\xi - 2$. These bounds are tight. Thus, $K_\xi(u, v, z)$ is a polynomial of degree $2\xi - 2$ in $u$, $3\xi - 1$ in $v$ and $4\xi - 2$ in $z$.*

**Proof.** Consider a kernel walk $W_K$ of excess $\xi$, with $m$ vertices, $\ell_1$ leaves, $\ell_2$ vertices of degree 2 that are not the root, and outdegree sequence $(d_1, \dots, d_m)$. Each leaf is incident to an excess edge, so $\ell_1 \leq \xi$. Each vertex of degree 2 is incident to an excess edge, so $\ell_2 \leq 2\xi - \ell_1$. The sum of the outdegrees is $m - 1$, so

$$m - 1 = \sum_j d_j \geq 2\,(m - \ell_1 - \ell_2) + \ell_2$$

which implies $m \leq 3\xi - 1$. The number of simple edges is bounded by $m - 1 - \xi \leq 2\xi - 2$. The kernel walk has at most $2\xi - 2$ half-steps along the simple edges, and $2\xi$ half-steps along the excess edges, so the half-length is bounded by $4\xi - 2$. Any binary tree on $2\xi - 1$ vertices, with additional edges of excess 1 attached to each leaf, reaches those bounds. ◀

Although expressing $K_\xi(u, v, z)$ directly is challenging, some subfamilies of kernel walks have a simple expression. A kernel walk of excess $\xi$ is said to be *optimal* if it contains $2\xi - 2$ simple edges, and *near-optimal* if it contains $2\xi - 3$ simple edges.

▶ **Lemma 8.** *Let $\mathrm{Cat}(n)$ denote the $n$-th Catalan number. There are $(3\xi - 1)!\,\mathrm{Cat}(\xi - 1)$ optimal kernels of excess $\xi$ for $\xi \geq 1$, and $(3\xi - 1)!\,\mathrm{Cat}(\xi - 1)$ near-optimal kernels of excess $\xi$ for $\xi \geq 2$. Let $K_{\xi,s}(z)$ denote the generating function of kernel walks with excess $\xi$ and $s$ simple edges in the induced tree, where $z$ marks the half-length of the walk. This implies*

**(a)** $K_{\xi,2\xi-2}(z) = \mathrm{Cat}(\xi - 1) z^{4\xi-2}$, *for* $\xi \geq 1$

**(b)** $K_{\xi,2\xi-3}(z) = (3\xi - 1)\,\mathrm{Cat}(\xi - 1) z^{4\xi-3}$, *for* $\xi \geq 2$.

Given a kernel walk $W_K$ with excess $\xi$, we reconstruct a tree walk $W$ by substituting every simple edge by a sequence of back and forth steps, adding a sequence of steps at the root of $T(W_K)$, moving the root to the leaf of this attached path and adding a tree walk without excess at the beginning and after each step in this extension of $W_K$.

▶ **Lemma 9.** *Let $W_\xi(z)$ be the generating function of the number of tree walks with excess $\xi \geq 1$ and $K_\xi(u, v, z)$ the generating function of kernel walks with excess $\xi$ and where $u$ marks the number of simple edges, $v$ the number of vertices and $z$ the half-length of the walk. Then*

$$W_\xi(z) = \frac{C(z)}{1 - zC(z)^2} K_\xi\left(\frac{1}{1 - zC(z)^2}, 1, zC(z)^2\right).$$

The proof of Theorem 3 is now straightforward.

## 3.2 A recursion for the generating function of tree walks of excess $\xi$

Theorem 3 raises the question of the computation of the generating function $K_{\xi,s}(z)$ of kernel walks of excess $\xi$ with $s$ simple edges, where $z$ marks the half-length. There exists a recurrence, but we prefer to decompose the tree walks further and enumerate simpler objects. This path also alleviates the work of the computer algebra system when computing $K_{\xi,s}(z)$.

▶ **Definition 10.** *A superreduced walk is a tree walk where no edge is simple. Denoting by $s_{m,2\ell}$ the number of such walks of length $2\ell$ on $m$ vertices, their generating function is*

$$S(v, z) = \sum_{\substack{\ell \geq 0 \\ m \geq 0}} s_{m,2\ell} \frac{v^m}{m!} z^\ell.$$

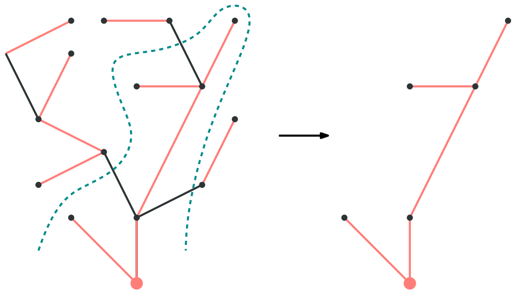


**Figure 2** Decomposing kernel walks by isolating the superreduced component including the root (red).

The following lemma reduces the enumeration of kernels to the enumeration of superreduced walks. The main idea is to consider the induced tree of a kernel walk and isolate the component which contains the root after deleting all simple edges (see Figure 2). The restriction of the kernel walk to this component is a superreduced walk and the restriction of the kernel walk to all of the other components are kernel walks again.

▶ **Lemma 11.** *Let $S(v,z)$ be the generating function of superreduced walks, that is, kernel walks without simple edges, where $v$ counts the number of vertices in the induced tree and $z$ the half-length of the walk. Then for the generating function of kernel walks $K(u,v,z) = \sum_{\xi\geq 0} K_\xi(u,v,z)$ it holds that*

$$K(u,v,z) = \frac{1}{(1-uz(K(u,v,z)-v))} S\left(v, \frac{z}{(1-uz(K(u,v,z)-v))^2}\right) - uvz(K(u,v,z)-v).$$

Once given the generating function $S(v,z)$ of superreduced walks, we compute $K(u,v,z)$ by Lagrange inversion (see *e.g.* [11]). Our next lemma provides an equation characterizing $S(v,z)$. The proof relies on the idea from [1] to mark the number of times the walk leaves the root (see Figure 3). Applying the symbolic method [2, 9] to translate it into generating functions results in a series $S(x,v,z)$ for superreduced walks, where a new auxiliary variable $x$ marks how often the walk leaves the root.

▶ **Lemma 12.** *Let $s_{j,m,2\ell}$ denote the number of superreduced walks on $m$ vertices, length $2\ell$ and leaving the root $j$ times. Let*

$$S(x,v,z) = \sum_{j,m,\ell\geq 0} s_{j,m,2\ell} \frac{x^j}{j!} \frac{v^m}{m!} z^\ell$$

*denote the generating function of superreduced kernel walks, where $z$ marks the half-length of the walk, $v$ the number of vertices in the induced tree and $x$ how often the walk leaves the root. Then*

$$S(x,v,z) = v \exp\Big(\mathcal{L}_{t=1}\Big(D\,(t,xz)\,S(t,v,z)\Big)\Big)$$

*where $D(t,x) = \sum_{k\geq 1} \frac{x^{k+1}}{(k+1)!}\frac{t^k}{k!}$ and $\mathcal{L}_{t=1}\,(A(t)) = \sum_{k\geq 0} k!\,[t^k]A(t)$.*

By implementing this well founded recursion in $v$ *and* $z$ it is easy to compute $S(v,z)$ up to order $\xi+1$ in $v$ and $2\xi$ in $z$, then we compute $[u^s]K(u,v,z)$ for $s \in [1,\xi]$ by Lagrange

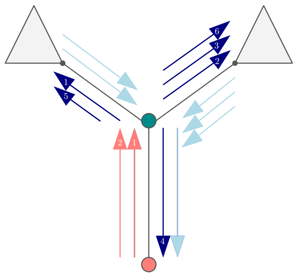


**Figure 3** Decomposition of a superreduced walk.

inversion from Lemma 11, and finally $W_\xi(z)$. For example for $\xi = 1, 2, 3$, we obtain the generating functions

$$W_1(z) = \frac{z^2 C(z)^5}{1 - zC(z)^2}, \quad W_2(z) = C(z) \frac{z^3 C(z)^6 + 4z^4 C(z)^8 - 6z^5 C(z)^{10} + 2z^6 C(z)^{12}}{\left(1 - zC(z)^2\right)^3}$$

$$W_3(z) = z^4 C(z)^9 \frac{1+16zC(z)^2+11z^6C(z)^{12}+95z^4C(z)^8-54z^5C(z)^{10}-62z^3C(z)^6-5z^2C(z)^4}{\left(1-zC(z)^2\right)^5},$$

recovering and extending the results of [1] and of [8] (except for $W_2(z)$ where our calculation differs from [8] and agree with [1]).

## 4 A refined normalisation of the spectral measure and some curious identities

In this section, we return to our initial motivation to describe the moments of the spectral measure $\mu^c$ by the identity

$$M_{\mu^c}(z) = \frac{1}{c} W\left(c, \frac{z^2}{c}\right) = \sum_{\xi \geq 0} \frac{1}{c^\xi} W_\xi\left(z^2\right).$$

As Zacharevich [24] pointed out, $\mu^c$ is fully determined by its moments and if $\mu^c$ were a continuous measure, we could compute its density by the inversion formula of Stieltjes-Perron. This is not the case ($\mu^c$ has a dense set of atoms [5, 4]), but nonetheless a better understanding of the Stieltjes transform of $\mu^c$ would entail a better understanding of the measure itself.

In combinatorial terms, the Stieltjes transform $S_\mu(z)$ of a measure $\mu$ with finite moments is simply the ordinary generating function of moments evaluated at $z^{-1}$ multiplied by $z^{-1}$. That is,

$$S_\mu(z) = \sum_{\ell \geq 0} m_\ell(\mu) z^{-(\ell+1)}.$$

In turn, under some conditions, the Stieltjes-Perron formula expresses the density $\rho$ of the measure $\mu$ by

$$\rho(z) = \lim_{\varepsilon \to 0} -\frac{1}{\pi} \mathrm{Im}\left(S_\mu(z + i\varepsilon)\right). \tag{3}$$

For example, the Stieltjes transform of the limit law $\mu$ of the normalized spectral measure of $G(n, p)$ with $p$ constant, and its density, are respectively

$$S_\mu(z) = \frac{1}{z} C\left(\frac{1}{z^2}\right), \qquad \lim_{\varepsilon \to 0} -\frac{1}{\pi} \mathrm{Im}\left(S_\mu(z + i\varepsilon)\right) = \frac{\sqrt{4 - z^2}}{2\pi} \mathbb{1}_{(-2,2)}(z).$$

The distribution given by this density is called after its shape, the *semicircle distribution.* The Stieltjes transform of $\mu^c$ equals

$$S_{\mu^c}(z) = \frac{1}{z} M_{\mu^c}\left(\frac{1}{z}\right) = \sum_{\xi \geq 0} \frac{1}{z c^\xi} W_\xi \left(\frac{1}{z^2}\right).$$

Given the structure of $W_\xi(z)$ from Theorem 3, $S_{\mu^c}(z)$ is a sum of rational functions in $S_\mu(z)$

$$S_{\mu^c}(z) = S_\mu(z) + \frac{1}{c} \cdot \frac{S_\mu(z)^5}{1 - S_\mu(z)^2} + \frac{1}{c^2} \cdot \frac{S_\mu(z)^7 + 4S_\mu(z)^9 - 6S_\mu(z)^{11} + 2S_\mu(z)^{13}}{\left(1 - S_\mu(z)^2\right)^3} + \dots$$

Now one could hope that the inversion formula applied to each of the $z^{-1}W_\xi(z^{-1})$ would yield a density of a measure and the density of $\mu_c$ would turn out to be a weighted sum of them. This hope is certainly too far fetched, as $\mu^c$ has a dense set of atoms. But Enriquez and Ménard [8] found a way to still make use of this expansion by using a dilation operator in their Theorem 3. The main idea is to scale the spectral measure and evaluate $M_{\mu^c}(z)$ at $z/(1 + \frac{1}{2c})$ instead. This scaling entails a perturbation on the level of coefficients of $1/c$. In particular, $\sum_{\ell \geq 0} m_{2\ell} \left(\frac{z}{1+\frac{1}{2c}}\right)^{2\ell}$ is equal to

$$\sum_{\ell \geq 0} \left( w_{0,2\ell} z^{2\ell} + \tfrac{1}{c}(w_{1,2\ell} - \ell w_{0,2\ell}) z^{2\ell} + \tfrac{1}{c^2} \left( w_{2,2\ell} - \ell w_{1,2\ell} + \left(\tfrac{\ell^2}{2} + \tfrac{\ell}{4}\right) w_{0,2\ell} \right) z^{2\ell} + \dots \right).$$

Now the generating functions at $c^{-1}$ and $c^{-2}$ are polynomials in $z^2 C(z^2)^2$ multiplied by $C(z^2)$, and the corresponding densities can be computed with the inversion formula. We expand their calculation to order 5 instead of 2.

Instead of using the dilation operator, we can rescale the adjacency matrix $A(G(n, c/n))$ of $G(n, c/n)$ by $\frac{1}{\sqrt{c\,p(1/c)}}$ instead of $\frac{1}{\sqrt{c}}$. We define

$$\mu_n^p = \frac{1}{n} \sum_{\lambda \in Sp\left((c\,p(1/c))^{-1/2} A(G(n,c/n))\right)} \delta_\lambda,$$

where $p(x)$ is a polynomial in $x$ with constant term 1 which is yet to be determined, and $\mu^p$ for the limit as $n$ tends to infinity. This implies

$$\mu_n^p = \Lambda_{p(1/c)^{-1/2}}(\mu_n^c) \quad \text{and} \quad M_{\mu^p}(z) = M_{\mu^c}\left(\frac{z}{\sqrt{p(1/c)}}\right).$$

The original scaling factor $1/\sqrt{c}$ derives from the classical scaling of Wigner matrices, where one scales the matrix by $1/\sqrt{n\mathbb{V}(X)}$, where $X$ is distributed as the individual matrix entries. In the case of adjacency matrices of $G(n, c/n)$ the variance of Bernoulli variables determining the entries of the matrix is of course $c/n(1 - c/n)$ such that we obtain the scaling factor $1/\sqrt{c}$ in the limit. We do not have a similar interpretation for our proposed alternative scaling.

► **Proposition 13.** *Let $p_5(x) = 1 + x + x^2 + 4x^3 + 33x^4 + 386x^5$ and $M_{\mu^c}(z)$ be the ordinary moment generating function of $\mu^c$ as defined in (1). Then for $V_i(z) = \left[c^{-i}\right] M_{\mu^p}\left(\sqrt{z}\right)$ we have*

$$V_i(z) = C(z) Q_i \left(z C(z)^2\right), \quad i = 0, 1, 2, \dots, 5,$$

*where*

$$\begin{aligned}
&Q_0(x) = 1, \qquad Q_1(x) = -x, \qquad Q_2(x) = -2x^3,\\
&Q_3(x) = -\left(11x^5 + x^4 - 2x^3 + 2x^2 + 3x\right),\\
&Q_4(x) = -\left(90x^7 + 27x^6 - 19x^5 + 17x^4 + 23x^3 + 20x^2 + 26x\right),\\
&Q_5(x) = -\left(931x^9 + 529x^8 - 163x^7 + 166x^6 + 301x^5 + 239x^4 + 249x^3 + 266x^2 + 324x\right).
\end{aligned}$$

The proof consists in computing the coefficients of $p_5(x)$ one by one, starting with $[x^0]p(x) = 1$. Then, for any $k$, let us assume the first $k-1$ coefficients have been computed and set $[x^k]p(x)$ as a variable. We observe in our computations that for the first few values of $k$,

$$[c^{-k}]M_{\mu^c}\left(\sqrt{\frac{z}{\sum_{j=0}^{k}[x^j]p(x)c^{-j}}}\right)$$

is a fraction with denominator a power of $1 - zC(z)^2$, and the coefficient $[x^k]p(x)$ can be chosen so that this fraction reduces to a polynomial.

**Proof of Theorem 4.** The generating function of the moments of $\Lambda_{f(1/c)}(\mu^c)$ is given by $M_{\mu^c}\left(\frac{z}{f(1/c)}\right)$. Note that

$$f(x)^2 = 1 + x + x^2 + 4x^3 + 33x^4 + 386x^5 + O(x^6)$$

such that we can expand

$$M_{\mu^c}\left(\frac{z}{f(1/c)}\right) = M_{\mu^{f^2}}(z) = \sum_{i=0}^{5}\frac{1}{c^i}V_i\left(z^2\right) + \sum_{i\geq 6}\frac{1}{c^i}[c^{-i}]M_{\mu^{f^2}}(z)$$

where the $V_i(z)$ are given by Proposition 13. Applying the inversion formula to these functions yield densities of signed measures with null mass. ◀

To illustrate why the existence of $p_5(x)$ is surprising, we observe that if

$$\tilde{W}_2(z) := W_2(z) + \frac{z^3C(z)^7}{(1 - zC(z)^2)^3}$$

is given instead of $W_2(z)$, then there is no choice for $[x^2]p(x)$ allowing this magical simplification between numerator and denominator and the reduction to a polynomial.

This example highlights the difficulty of proving the existence of $P(x)$ in Conjecture 5. A combinatorial approach seems reasonable, but we are not aware of any combinatorial meaning of the generating functions $V_i(z)$, nor do we have a combinatorial interpretation of the differential equations which are satisfied by the generating functions $W_\xi(z)$, except for the equation of $V_1(z)$. Nevertheless the next theorem sheds partial light on *why* the scaling by $p_5(x)$ results in Proposition 13. It shows that keeping the same first two coefficients as in $p_5(x)$ but changing the others gives fractions, in the expansion in $c^{-1}$, with denominators that are powers of $1 - zC(z)^2$ that are two less than expected.

▶ **Theorem 14.** *Let $p(x) = \sum_{i\geq 0} x^i$. Then $\hat{V}_i(z) := [c^{-i}]M_{\mu^p}\left(\sqrt{z}\right)$ is a polynomial in $zC(z)^2$ multiplied by $C(z)$ for $i = 0, 1, 2$ and for $i \geq 3$ there exist polynomials $\hat{Q}_i(x)$ such that*

$$\hat{V}_i(z) = C(z)\frac{\hat{Q}_i(zC(z)^2)}{(1 - zC(z)^2)^{2i-3}}.$$

# 5 Computational experiments

As curious as Conjecture 5 is from a purely mathematical perspective, the alternative scaling of the matrices of the spectral measure seems to have advantages in the approximation of the limit measure $\bar{\mu}^c$. There are certain important details to take into account though.

Since the $V_i(z)$ in Corollary 13 are polynomials in $zC(z)^2$ multiplied by $C(z)$, the evaluation $\frac{1}{z}V_i\left(\frac{1}{z}\right)$ is a polynomial in the Stieltjes transform of the semicircle law. The inversion formula (3) therefore always yields densities of signed measures with zero mass for these Stieltjes transforms. In particular, we obtain a sequence of densities $f_i(z)$ from the Stieltjes transforms $\frac{1}{z}V_i\left(\frac{1}{z}\right)$ for $1 \le i \le 5$ which are given by

$$\begin{aligned}
f_0(z) &= \frac{1}{2\pi}\sqrt{4-z^2}\,\mathbb{1}_{(-2,2)}(z),\\
f_1(z) &= \frac{1}{2\pi}\left(1-z^2\right)\sqrt{4-z^2}\,\mathbb{1}_{(-2,2)}(z),\\
f_2(z) &= \frac{1}{2\pi}\left(1-6z^2+5z^4-z^6\right)\sqrt{4-z^2}\,\mathbb{1}_{(-2,2)}(z),\\
f_3(z) &= \frac{1}{2\pi}\left(9-140z^2+358z^4-299z^6+98z^8-11z^{10}\right)\sqrt{4-z^2}\,\mathbb{1}_{(-2,2)}(z),\\
f_4(z) &= \frac{1}{2\pi}\big(56+1602z^2-8625z^4+16004z^6\\
&\qquad -13447z^8+5624z^{10}-1143z^{12}+90z^{14}\big)\sqrt{4-z^2}\,\mathbb{1}_{(-2,2)}(z),\\
f_5(z) &= \frac{1}{2\pi}\big(442-17946z^2+171911z^4-574676z^6+904447z^8\\
&\qquad -768354z^{10}+373181z^{12}-103622z^{14}+15298z^{16}-931z^{18}\big)\sqrt{4-z^2}\,\mathbb{1}_{(-2,2)}(z).
\end{aligned}$$

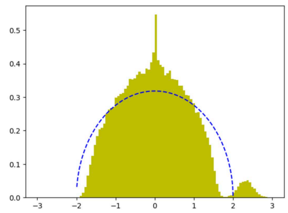
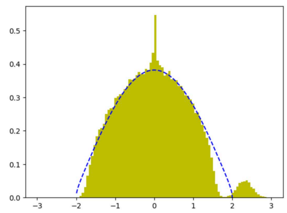
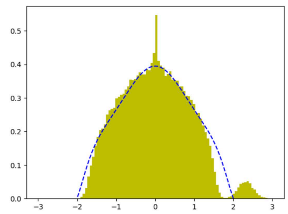
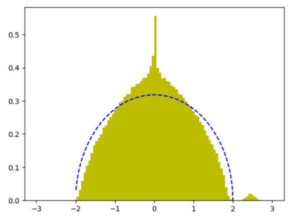
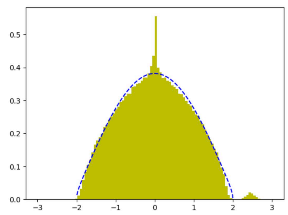
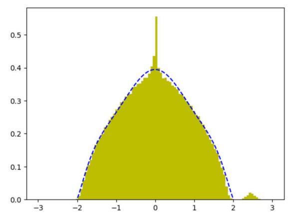
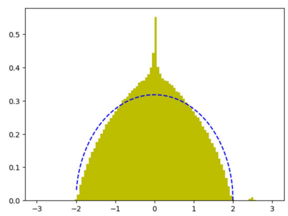
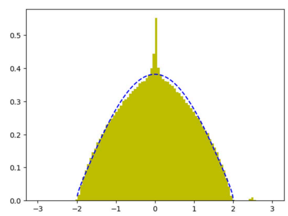
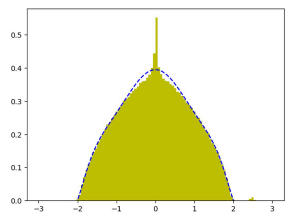
Now, it is easy to see that the coefficients of the polynomial factors of the $f_i(z)$ grow rapidly and that these functions oscillate quite heavily. Hence, there exists a largest integer $t(c)$ depending on $c$ such that

$$\sum_{\xi=0}^{t(c)} \frac{1}{c^\xi} f_\xi(z)$$

takes non-negative values on the interval $(-2,2)$ and is therefore the density of a probability measure. Experiments for $c = 5, 10, 20$ show that this $t(c)$ seems to be the right scaling for $\bar{\mu}^c$ such that most of the eigenvalues are exactly in the interval $(-2,2)$. This is reminiscent of divergent asymptotic expansions (see *e.g.* the introduction of [14]). For example, consider Stirling's asymptotic expansion $n! \approx n^n e^{-n}\sqrt{2\pi n}(s_0 + s_1 n^{-1} + s_2 n^{-2} + \cdots)$ where $(s_0, s_1, s_2, \ldots) = (1, \frac{1}{12}, \frac{1}{288}, \ldots)$. For any $n$, there exists $t(n)$ such that the accuracy of the approximation of order $k$ improves for $k$ from 0 to $t(n)$, then decreases with $k$.

Further, the densities seem to approximate the histograms of eigenvalues of sampled matrices quite well. In Table 1, we can see histograms of random matrices with $p = 5/n$. In each row, we sampled $N$ matrices of size $n \times n$ such that we always obtained 100000 eigenvalues. They were scaled by $\sqrt{c(1+1/c)}$ such that we would expect a reasonable approximation by the density $f_0(z) + 1/cf_1(z) + 1/c^2 f_2(z)$. Indeed, in the columns we see the histograms of the eigenvalues in green and the densities given by the approximations of $f_0(z)$, $f_0(z) + c^{-1}f_1(z)$ and $f_0(z) + c^{-1}f_1(z) + c^{-2}f_2(z)$. As $n$ grows, the curve of the latter fits the histogram best. Another example is illustrated in Table 2. In this case, $c = 10$ and $t(c) = 3$ such that we consider the densities $f_0(z) + c^{-1}f_1(z)$, $f_0(z) + c^{-1}f_1(z) + c^{-2}f_2(z)$ and $f_0(z) + c^{-1}f_1(z) + c^{-2}f_2(z) + c^{-3}f_3(z)$.

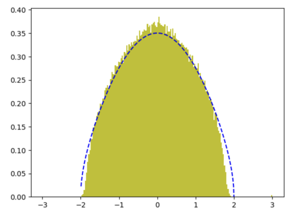
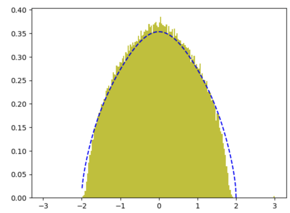
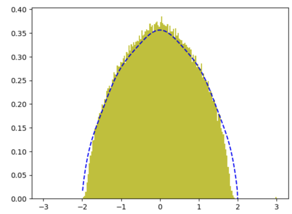
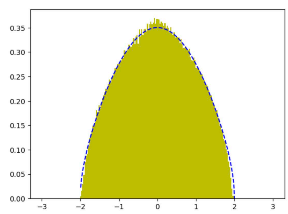
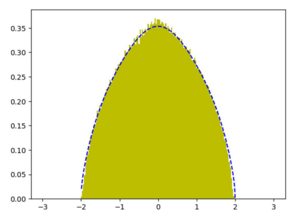
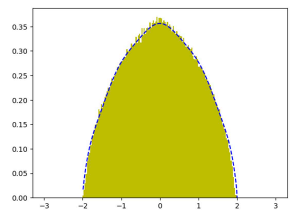
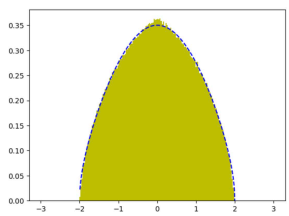
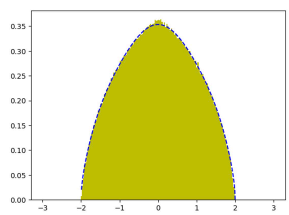
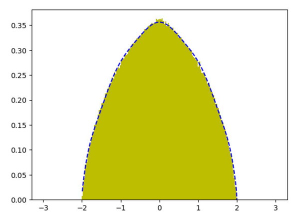
**Table 1** Histograms (100 bins) of eigenvalues of $N$ random adjacency matrices of $G(n, 5/n)$ compared to the densities $f_0(z)$, $f_1(z)$ and $f_2(z)$.

| sample size | $f_0(z)$ | $f_0(z) + \frac{1}{5} f_1(z)$ | $f_0(z) + \frac{1}{5} f_1(z) + \frac{1}{25} f_2(z)$ |
|---|---|---|---|
| n=40 N=2500 |  |  |  |
| n=200 N=500 |  |  |  |
| n=1000 N=100 |  |  |  |

## References


1. Michel Bauer and Olivier Golinelli. Random incidence matrices: Moments of the spectral density. *Journal of Statistical Physics*, 103:301–337, 2001. doi:10.1023/A:1004879905284.
2. François Bergeron, Gilbert Labelle, and Pierre Leroux. *Combinatorial Species and Tree-like Structures*. Cambridge University Press, 1997.
3. Charles Bordenave, Marc Lelarge, and Justin Salez. The rank of diluted random graphs. *Annals of Probability*, 39(3):1097–1121, 2011. doi:10.1214/10-AOP567.
4. Charles Bordenave, Arnab Sen, and Bálint Virág. Mean quantum percolation. *Journal of the European Mathematical Society*, 19(12):3679–3707, 2017. doi:10.4171/JEMS/750.
5. JT Chayes, L Chayes, Judy R Franz, James P Sethna, and SA Trugman. On the density of state for the quantum percolation problem. *Journal of Physics A: Mathematical and General*, 19(18):L1173, 1986. doi:10.1088/0305-4470/19/18/011.
6. Simon Coste and Justin Salez. Emergence of extended states at zero in the spectrum of sparse random graphs. *Annals of Probability*, 49(4), 2021. doi:10.1214/20-AOP1499.
7. Alan Edelman and N Raj Rao. Random matrix theory. *Acta numerica*, 14:233–297, 2005. doi:10.1017/S0962492904000236.
8. Nathanaël Enriquez and Laurent Ménard. Spectra of large diluted but bushy random graphs. *Random Structures & Algorithms*, 49(1):160–184, 2016. doi:10.1002/rsa.20618.
9. Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009. URL: http://www.cambridge.org/uk/catalogue/catalogue.asp?isbn=9780521898065.

**Table 2** Histograms (200 bins) of eigenvalues of $N$ random adjacency matrices of $G(n, 10/n)$ compared to the densities $\tilde{f}_1(z)$, $\tilde{f}_2(z)$ and $\tilde{f}_3(z)$.

| sample size | $\tilde{f}_1(z) = f_0(z) + \frac{1}{10} f_1(z)$ | $\tilde{f}_2(z) = \tilde{f}_1(z) + \frac{1}{100} f_2(z)$ | $\tilde{f}_3(z) = \tilde{f}_2(z) + \frac{1}{1000} f_3(z)$ |
|---|---|---|---|
| n=125 N=800 |  |  |  |
| n=500 N=200 |  |  |  |
| n=2000 N=50 |  |  |  |


10 Zoltán Füredi and János Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1:233–241, 1981. doi:10.1007/BF02579329.

11 Ira M Gessel. Lagrange inversion. *Journal of Combinatorial Theory, Series A*, 144:212–249, 2016. doi:10.1016/j.jcta.2016.06.018.

12 Eva-Maria Hainzl and Élie de Panafieu. Tree walks and the spectrum of random graphs. *arXiv*, arXiv:2405.08347:26, 2024. doi:10.48550/arXiv.2405.08347.

13 Brendan D McKay. The expected eigenvalue distribution of a large regular graph. *Linear Algebra and its applications*, 40:203–216, 1981. doi:10.1016/0024-3795(81)90150-6.

14 Frank Olver. *Asymptotics and special functions*. CRC Press, 1997. doi:10.1201/9781439864548.

15 Justin Salez. Every totally real algebraic integer is a tree eigenvalue. *Journal of Combinatorial Theory, Series B*, 111:249–256, 2015. doi:10.1016/j.jctb.2014.09.001.

16 Justin Salez. Spectral atoms of unimodular random trees. *Journal of the European Mathematical Society (EMS Publishing)*, 22(2), 2020. doi:10.4171/JEMS/923.

17 N.J.A. Sloane. sequence a294439. *The on-line encyclopedia of integer sequences*, 2017.

18 Linh V Tran, Van H Vu, and Ke Wang. Sparse random graphs: Eigenvalues and eigenvectors. *Random Structures & Algorithms*, 42(1):110–134, 2013. doi:10.1002/rsa.20406.

19 Antonia M Tulino, Sergio Verdú, et al. Random matrix theory and wireless communications. *Foundations and Trends® in Communications and Information Theory*, 1(1):1–182, 2004. doi:10.1561/0100000001.

**20** Eugene P Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, pages 548–564, 1955. `doi:10.2307/1970079`.

**21** Eugene P Wigner. On the distribution of the roots of certain symmetric matrices. *Annals of Mathematics*, 67(2):325–327, 1958. `doi:10.2307/1970008`.

**22** Eugene P Wigner. Random matrices in physics. *SIAM review*, 9(1):1–23, 1967. `doi:10.1137/1009001`.

**23** Edward M Wright. The number of connected sparsely edged graphs I-III. *Journal of Graph Theory*, 1977-80.

**24** Inna Zakharevich. A generalization of wigner's law. *Communications in Mathematical Physics*, 268:403–414, 2006. `doi:10.1007/s00220-006-0074-5`.