Spectrum Estimation through Kirchhoff Random Forests

Simon Barthelme, Fabienne Castell, Alexandre Gaudillière, Clothilde Mélot, Matteo Quattropani, Nicolas Tremblay

Spectrum Estimation through Kirchhoff Random Forests

S. Barthelmé* F. Castell† A. Gaudillière‡ C. Mélot§ M. Quattropani¶
N. Tremblay‖

August 28, 2025

Abstract

Given a non-oriented edge-weighted graph, we show how to make some estimation of the associated Laplacian eigenvalues through Monte Carlo evaluation of spectral quantities computed along Kirchhoff random rooted spanning forest trajectories. The sampling cost of this estimation is only linear in the node number, up to a logarithmic factor. By associating a double cover of such a graph with any symmetric real matrix, we can then perform spectral estimation in the same way for the latter.

1 Introduction

Exact eigendecomposition of very large matrices is infeasible, due to its high computational cost, which scales as $O(n^3)$, for a $n \times n$ matrix. In many applications, matrices are indeed very large, but exact spectral information is not required. For instance, in applications to physics [WWAF06], it can be enough to estimate the distribution function of eigenvalues rather than each exact eigenvalue. This problem of estimating the *spectral cumulative distribution* is the topic of this paper.

We introduce a method for spectral estimation of large matrices, which consists in exploiting a stochastic process to gain information on the spectrum of a *graph Laplacian*. Graph Laplacians are a subset of symmetric matrices, but we shall see that the method can be extended to general matrices via an embedding. The stochastic process we use is a *Kirchhoff forest*. Kirchhoff forests are a type of random spanning forests that have been used successfully in various applications which involve linear-algebraic computations in large graphs [ACGM20, JFA+23, PABT21].

Briefly, our approach follows the same pattern used in classical methods: we estimate (generalised) moments of the spectral density, and combine these estimates in order to reconstruct the cumulative spectral density.

Classical methods like the kernel polynomial method [WWAF06] estimate standard (polynomial) moments of the spectral distribution. Here, we construct a set of observables which can be used to estimate the expectation of certain *rational* functions under the spectral density. We show how many such expectations can be estimated efficiently by coupling Kirchhoff forests at different intensities. In addition, we show that our estimators have *relative* error in $O(\frac{1}{\sqrt{n}})$, and cost linear in n.

Since we do not estimate classical but rational moments, we require a bespoke reconstruction method. This method relies on (pointwise) transformation to a classical truncated moment problem. We explain how to get bounds on the cumulative spectral density, and how to perform a maximum-entropy reconstruction.

Our method performs encouragingly, as shown in section 6. A comparison to state-of-the-art methods can be found in a related conference paper [BCG+25]. Here the focus is on introducing and studying the relevant observables, on implementing the coupling process introduced in [AG18], and on formulating a reconstruction process. The end result is a spectral estimation method with complexity linear in n, and weak dependence on the sparsity level of the matrix, unlike all classical methods.

*CNRS, Univ. Grenoble Alpes, Grenoble INP, GIPSA-lab Grenoble, France, e-mail: simon.barthelme@grenoble-inp.fr
†Aix-Marseille Univ., CNRS, Centrale Marseille, I2M, Marseille, France, e-mail: fabienne.castell@univ-amu.fr
‡Aix-Marseille Univ., CNRS, Centrale Marseille, I2M, Marseille, France, e-mail: alexandre.gaudilliere@math.cnrs.fr
§Aix-Marseille Univ., CNRS, Centrale Marseille, I2M, Marseille, France, e-mail: clothilde.melot@univ-amu.fr
¶Sapienza Università di Roma, Italy, e-mail: matteo.quattropani@uniroma1.it
‖1) Department of Mathematics & Statistics, UiT the Arctic University of Norway, Tromsø, Norway
and 2) CNRS, Univ. Grenoble Alpes, Grenoble INP, GIPSA-lab, Grenoble, France,
e-mail: nicolas.tremblay@cnrs.fr

1.1 Results

To give a high-level overview of our results, we need to introduce some notation. We focus on a specific set of symmetric matrices called *graph Laplacians* (more general matrices are considered in appendix A). Consider a non-oriented edge-weighted graph without self-loops, with vertex set $\mathcal{X}$ of finite size n and with non negative weights

$$w(x,y) = w(y,x) \geq 0, \qquad x,y \in \mathcal{X},$$

using the convention that an edge connects x and y if and only if $w(x,y) > 0$. With such an edge we associate two distinct oriented *arrows* (x,y) and (y,x). We say that (x,y) *leaves* x and *reaches* (or *points to*) y. We define the total weight of a node as

$$w(x) = \sum_{y \neq x} w(x,y) \tag{1}$$

which corresponds to the degree of x in the unweighted case. We note m the total number of edges.

The graph Laplacian is a symmetric matrix L with entries

$$L_{x,y} = \begin{cases} w(x,y) & x \neq y \\ -w(x) & x = y \end{cases} \tag{2}$$

Graph Laplacians are diagonally-dominant matrices, and are known to have favourable computational properties with respect to (approximate) inversion [ST04]. Note that in our definition L is negative-definite, for consistency with the definition of the classical Laplace operator, but many sources in spectral graph theory use the opposite convention.

The *spectral measure* of L is defined as:

$$\sigma = \frac{1}{n} \sum_{j<n} \delta_{\lambda_j}. \tag{3}$$

where $0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1}$ are the eigenvalues of $-L$, counted with multiplicity. We will seek to estimate the cumulative spectral distribution

$$F(q) = \sigma([0,q]) = \frac{1}{n} \left|\left\{ j < n : \lambda_j \leq q \right\}\right| \tag{4}$$

for different values of q. Defining α as

$$\alpha = \max_x w(x)$$

a straightforward application of the Gershgorin circle theorem gives the well-known bound $\lambda_{n-1} \leq 2\alpha$. The spectral measure has support on $[0, 2\alpha]$.

We will do so by estimating certain expectations under σ, specifically:

$$m_k(q) = \frac{1}{n} \sum_{j<n} \left(\frac{q}{q+\lambda_j} \right)^k, \qquad q > 0, \tag{5}$$

which up to trivial rescalings correspond to the Stietjes transform of the spectral measure (for $k = 1$) and its $(k-1)$-th derivatives, evaluated on the positive real axis.

Our main result is that expectations of this form can be estimated efficiently, with a control on the relative error. Since our estimators involve running random walks, some preprocessing may be required in order to ensure that a random neighbour can be chosen with probability $\frac{w(x,y)}{w(x)}$ in time $O(1)$. In unweighted graphs this is trivial. In weighted graphs, if the weights are not too variable rejection sampling can be used, otherwise we may need to perform the preprocessing necessary to run Walker's alias algorithm at each node [Wal77], which costs $O(m)$ in total. This is no more than the cost of reading the matrix.

Combining Theorems 3 and 2, we have:

Theorem 1 (Informal)**.** *Given a preprocessing step of cost at most $O(m)$, we can obtain an unbiased estimator of $m_k(q)$ for $k \leq l$ and any r values of $q \in [q_{\min}, q_{\max}]$ at cost $O\left(nl\left(\frac{\bar{\lambda}}{q_{\min}} + \log\frac{\alpha}{q_{\min}} + r\right)\right)$ where $\bar{\lambda} = \frac{1}{n}\sum_{i<n}\lambda_i = \frac{1}{n}\mathrm{Tr}(-L)$. This estimator has pointwise* relative *error in $O(\frac{1}{\sqrt{n}})$.*

In large graphs we obtain in practice very reliable estimates, which can be used for effective reconstruction, as we show.

1.2 Outline of the paper

The remainder of the paper is organized as follows. In Section 2 we describe the main object underlying our analysis, i.e., *Kirchhoff forests*, and provide some insight on the way they are related to the spectral distribution of the graph they live in. In Section 3 we recall the well-known *Wilson's algorithm* and present some modification of the classical algorithm which will turn out to be useful in sampling efficiently a whole continuum of forests. Section 4 is devoted to a description of the techniques we develop to extract information about the spectral distribution from the statistics collected by the randomized algorithm. Section 5 gathers the proofs of Theorems 2 and 3. In Section 6 we present the outcome of our algorithm on some benchmark examples and comment on related numerical issues. Appendix A explains how coupled Kirchhoff forest replicas can also be used to perform spectral estimation with any symmetric real matrix. We eventually give a pseudocode version of our coupled forest algorithm in Appendix B, before collecting in Appendix C some basics, sometimes in a different form, of the fundamental work of Krein and Nudel'man [KN77] after Markov's work on the moment problem. These basics are needed to get bounds on the spectral cumulative distribution function in Section 4.

2 Monte-Carlo estimates using Kirchhoff forests

In this section, we define the Kirchhoff forest, and describe the observables we use to obtain Monte-Carlo estimates of the Stieltjes transform of the spectral measure σ, and of its derivatives. The algorithms we use to sample the Kirchhoff forest, and the sampling computational costs will be discussed in section 3.

2.1 Kirchhoff forest

A **rooted spanning forest** (r.s.f.) is a collection ϕ of arrows such that:

i) for each node in $\mathcal{X}$, there is one arrow at most that leaves it;

ii) one cannot close any loop by following arrows from the node they leave to the node they reach.

When no arrow in ϕ leaves a given node x in $\mathcal{X}$, we say that x is a **root** of the forest ϕ. The set of roots of ϕ is denoted by $\rho(\phi)$.

For any $q > 0$, we call **Kirchhoff forest with rate** q a random rooted spanning forest Φ_q such that for all rooted spanning forest ϕ,

$$\mathbb{P}\big(\Phi_q = \phi\big) = \frac{q^{|\rho(\phi)|} \prod_{(x,y)\in\phi} w(x,y)}{Z(q)}$$

where $|\rho(\phi)|$ is the cardinality of $\rho(\phi)$ and $Z(q)$ is the normalizing constant

$$Z(q) = \sum_{\phi \,\mathrm{r.s.f}} q^{|\rho(\phi)|} \prod_{(x,y)\in\phi} w(x,y).$$

When $q = 1$ and in the case of an unweighted graph —i.e. when the $w(x, y)$ can only take the values 0 and 1— Kirchhoff forest is the uniformly distributed rooted spanning forest. For smaller or larger values of q and non-uniform weights $w(x, y)$ the random forest Φ_q is sometimes still referred as the "uniform" rooted spanning forest. Because of its connection with the Gaussian free field with positive mass it is also referred in [DEM24] as the "massive spanning forest". We prefered to give it another name and we called it Kirchhoff forest in view of his 1847's theorem that identifies the combinatorial partition function Z with the characteristic polynomial of the graph Laplacian L.

Kirchhoff's theorem [Kir47]. *For all $q > 0$, it holds**

$$Z(q) = \det\big(q\mathrm{Id} - L\big).$$

Kirchhoff's theorem reads

$$\ln Z(q) = \sum_{j<n} \ln(q + \lambda_j), \qquad q > 0. \tag{6}$$

*See [AG18] for a probabilistic proof that follows [Mar00] with the notation of the present paper.

Computing the first derivative in q of both sides of (6) leads to

$$\mathbb{E}\big[|\rho(\Phi_q)|\big] = \sum_{j<n} \frac{q}{q+\lambda_j}.$$

And computing the second derivative in q of both sides of (6), one obtains

$$\mathrm{Var}\big(|\rho(\Phi_q)|\big) \leq \mathbb{E}\big[|\rho(\Phi_q)|\big]. \tag{7}$$

For $q_0 > 0$ given, we describe in section 3 a way to sample the coupled forest process $q \geq q_0 \mapsto \Phi_q$ whithin a cost which is almost linear in the number of nodes. Therefore, we have access to Monte Carlo estimations of the function

$$q \geq q_0 \mapsto \mathbb{E}\big[|\rho(\Phi_q)|\big] = \sum_{j<n} \frac{q}{q+\lambda_j}. \tag{8}$$

Up to sign conventions and an extra factor qn, (8) is the Stieltjes transform

$$\mathcal{S} : q \geq q_0 \mapsto \mathcal{S}(q) = \frac{1}{n}\sum_{j<n} \frac{1}{q+\lambda_j} = \frac{1}{qn}\mathbb{E}\big[|\rho(\Phi_q)|\big] \tag{9}$$

of the Laplacian spectral measure σ. Moreover, this estimate is quite sharp since by (7), its variance is bounded by its first, rather than second, moment.

However, inverting the Stieltjes transform over the positive real domain is extremely ill-conditioned. Therefore, small estimation errors on the Stieltjes transform will produce huge errors after inversion for the eigenvalues we want to estimate.

2.2 Replicas

A first tentative strategy to overcome the aforementioned problem would then consist in relating the *complex* Stieltjes transform to some Kirchhoff forest observables. It holds for example, for all complex number $z = qe^{i\theta}$ with $q > 0$, $\theta \in [0, 2\pi[$ and $z \neq -\lambda_0, -\lambda_1, \ldots$

$$\frac{1}{n}\sum_{j<n} \frac{1}{z+\lambda_j} = \frac{\mathbb{E}\Big[|\rho(\Phi_q)|e^{i\theta|\rho(\Phi_q)|}\Big]}{nz\mathbb{E}\big[e^{i\theta|\rho(\Phi_q)|}\big]}.$$

But the relative precision needed by this formula for these expectations of signed or complex variables is completely out of reach for direct Monte Carlo estimations. We will rather define new observables closely related to the first derivatives of the Stieltjes transform in the positive domain, and we will later deal with the conditioning issue using a statistical point of view.

Consider l replicas $\Phi_{k,q}$, $k = 0, \cdots, l-1$, i.e. l independent copies, of the Kirchhoff forest Φ_q. The reader should think of very modest l (we will take $l = 4$ in our applications). For any x in $\mathcal{X}$ and any rooted spanning forest ϕ, x is covered by a single tree in ϕ. By writing $\rho_x(\phi)$ for the root of this tree and setting $R^0(x) = x$, we inductively define

$$R^{k+1}(x) = \rho_{R^k(x)}(\Phi_{k,q}), \qquad k < l.$$

We set

$$\xi_q^k = \big\{x \in \mathcal{X} : R^k(x) = x\big\} = \big\{x \in R^k(\mathcal{X}) : R^k(x) = x\big\}, \qquad k = 0, \ldots, l, \tag{10}$$

so that $\xi_q^0 = \mathcal{X}$, $\xi_q^1 = \rho(\Phi_{0,q})$ and $\xi_q^{k+1} \subset \rho(\Phi_{k,q})$ for all $k < l$. We will prove in Section 5

Theorem 2. *For all $k \leq l$, it holds*

$$\mathbb{E}\big[|\xi_q^k|\big] = \sum_{j<n} \left(\frac{q}{q+\lambda_j}\right)^k.$$

In addition,

$$\mathbb{P}\big(x, y \in \xi_q^k\big) \leq \mathbb{P}\big(x \in \xi_q^k\big)\mathbb{P}\big(y \in \xi_q^k\big), \qquad x \neq y \in \mathcal{X},$$

so that

$$\mathrm{Var}(|\xi_q^k|) \leq \mathbb{E}[|\xi_q^k|]. \tag{11}$$

Using s replicated forests, i.e. ls independent copies of Φ_q, one can therefore estimate $\mathbb{E}\big[|\xi_q^k|\big]$ by

$$\hat{\xi}_q^k = \frac{1}{s}\sum_{j=1}^{s}\Big|\xi_q^{k,(j)}\Big|, \tag{12}$$

where $(\xi_q^{k,(j)}, j = 1, \cdots, s)$ are s i.i.d. copies of ξ_q^k, each one obtained from l independent copies of the Kirchhoff forest. This estimate has relative error $(\hat{\xi}_q^k - \mathbb{E}\big[|\xi_q^k|\big])/\mathbb{E}\big[|\xi_q^k|\big]$ of order

$$\frac{\sqrt{\mathrm{Var}(|\xi_q^k|)}}{\sqrt{s}\mathbb{E}\big[|\xi_q^k|\big]} \le \frac{1}{\sqrt{s\mathbb{E}\big[|\xi_q^k|\big]}} \le \frac{C}{\sqrt{sn}},$$

as long as $\mathbb{E}\big[|\xi_q^k|\big]$ is of order n. This will be the case as soon as q_0 is chosen to be of the order of $\bar{\lambda} = \frac{1}{n}\sum_{j<n}\lambda_j$ since

$$n \ge \mathbb{E}\big[|\xi_q^k|\big] \ge n\left(\frac{q}{q+\bar{\lambda}}\right)^k \ge n\left(\frac{q_0}{q_0+\bar{\lambda}}\right)^k.$$

Up to simple and explicit multiplicative factors, $(\mathbb{E}\big[|\xi_q^k|\big], k < l)$ are the Stieltjes transform of the spectral measure σ and its first derivatives:

$$\mathcal{S}^{(k)}(q) = \frac{(-1)^k k!}{n}\sum_{j<n}\frac{1}{(q+\lambda_j)^{k+1}} = \frac{(-1)^k k!}{nq^k}\mathbb{E}\big[|\xi_q^{k+1}|\big].$$

We close this section with a last comment on the bounds (7) and (11). Kirchhoff's theorem actually characterizes the law or $|\rho(\Phi_q)|$, as the law of a sum of independent 0-1 Bernoulli random variables with mean $q/(q+\lambda_j)$, $j < n$. This identity in law can in turn be seen as a consequence of the determinantality[†] of the point process $\rho(\Phi_q)$, which derives from that of the arrow process Φ_q proved in [BP93]. Now, the determinantality of the point process $\rho(\Phi_q)$ leads directly to its negative correlations: for all $x \neq y$ in $\mathcal{X}$,

$$\mathbb{P}\big(x, y \in \rho(\Phi_q)\big) \le \mathbb{P}\big(x \in \rho(\Phi_q)\big)\mathbb{P}\big(y \in \rho(\Phi_q)\big).$$

And such negative correlations *by themselves* imply the bound (7). The point processes ξ_q^k are *not* determinantal but *have* negative correlations, which is sufficient to get (11) .

3 Sampling Kirchhoff forests

In this section, we address several sampling issues:

- We describe Wilson's algorithm whose output is a r.s.f. Φ_q for some given $q > 0$.
- We explain next how to couple all the $(\Phi_q, q > 0)$ using the stack representation of Diaconis & Fullton.
- We describe the process $t > 0 \mapsto \Phi_{1/t}$ as a non homogeneous Markov jump process on the space of rooted spanning forest. This description involves an auxiliary Markov process on the r.s.f. with active and frozen trees. At each jump of $\Phi_{1/t}$, the new state will be some final configuration of the auxiliary Markov process.
- We describe the actual algorithm we use to sample trajectories $t > 0 \mapsto \Phi_{1/t}$.
- We bound the computational cost of this algorithm.

Apart from the last point, all the material of this section is essentially a rewriting of [Wil96, AG18].

[†]This means that it exists a matrix $K_q = \big(K_q(x,y)\big)_{x,y\in\mathcal{X}}$ such that for all $A \subset \mathcal{X}$,

$$\mathbb{P}\big(A \subset \rho(\Phi_q)\big) = \det\big(K_q(x,y)\big)_{x,y\in A}.$$

Here we have $K_q = q(q\mathrm{Id} - L)^{-1}$ (see [AG18]).

3.1 Wilson's algorithm

For any given $q > 0$, Wilson's algorithm builds a random rooted spanning forest in the following way:

i) starting from any node x, it runs the continuous time random walk with generator L up to an independent random exponential time T_q of rate q, and erases all the possibly created loops as soon as they appear. It produces in this way a first path Γ_q^x, which is oriented and self-avoiding, joining x to the node where the random walk was stopped. Call A the collection of arrows along this path and B the set of the $|A| + 1$ nodes along Γ_q^x.

ii) Starting from another node x outside B, if any, the algorithm runs another loop-erased random walk up to the minimum of another independent exponential time of rate q and the hitting time of B. The algorithm produces in this way a new self-avoiding path $\Gamma_{q,B}^x$. It adds to A the arrows along that path, and adds to B the new nodes along it.

iii) If $B = \mathcal{X}$ the algorithm returns the rooted spanning forest $\Phi_q = A$; if not it repeats steps ii)–iii).

Wilson's theorem [Wil96]. *The output Φ_q of the algorithm is a Kirchhoff forest with rate q.*

It is useful to attribute a status to each node x during the algorithm. A node is said to be

- **frozen** if it belongs to an edge of A;
- **active** otherwise.

The current node of the loop-erased random walk is by definition active. The algorithm is initialized by setting all the nodes to active and by choosing a first current node.

3.2 Stack representation and coupled forests.

We stress the crucial fact that Wilson's theorem implies that the way the starting nodes are chosen in steps i) and ii) is irrelevant to the law of Φ_q as constructed by the algorithm. This feature of Wilson's algorithm is referred to as its **abelianity**.

This is made clear by Wilson's proof that relies on the stack representation of random walks introduced in [DF91]. The latter associates with each node x an infinite stack of arrows (x, y) or stops that appear independently and at each level with probability $w(x, y)/(q + w(x))$ and $q/(q + w(x))$, respectively. We can then build the continuous time loop-erased random walk that is stopped at rate q by reading and removing at rate $q+w(x)$ the first stop or arrow from the stack associated with the current node x before processing this stop or arrow, i.e., stopping after reading a stop or following it after reading an arrow.

With such a construction, whatever the chosen starting nodes, the same loops will be erased by Wilson's algorithm, the same arrows will eventually form Φ_q, and its roots will lie at each node where a stop has been read. Actually, as explained in [Wil96], the forest obtained from the algorithm is the forest one would obtain by removing the cycles when looking at the stacks from the top, as illustrated in Figure 1 in the case of a triangle. Therefore, it is not even mandatory to follow the path of the random walk to read the arrows in the stacks. We can as well choose as current node any active node which is a root of a tree. We will speak of **Wilson's order** when the arrows in the stacks are read according to the original loop-erased random walk paths. Figure 2 illustrates the forest construction when the stacks are read according to Wilson's order, while Figure 3 the current node is chosen at random among the active roots.

As noted in [AG18], such a stack representation allows for a coupling of all Φ_q, $q > 0$, together. Instead of sampling stops *or* arrows (x, y) at all levels of the stack associated with a node x, we can independently sample marks U in $[0, 1]$ *and* arrows (x, y), the former being uniformly distributed on $[0, 1]$ and the latter appearing now with probability $w(x, y)/w(x)$. This alea does not depend on q. To build Φ_q it suffices to interpret (or replace) any arrow by a stop each time its associated mark U is smaller that $q/(q + w(x))$.

It is explained in [AG18] and further develop in sections 3.3 and 3.4 that this coupling leads to piecewise-constant random forest trajectories

$$t \geq 0 \mapsto \Phi_{1/t}$$

that start from the deterministic empty forest Φ_∞ made of n trees reduced to a simple root. This random forests process has the property that all arrows that are read to compute a Kirchhoff forest Φ_q at time

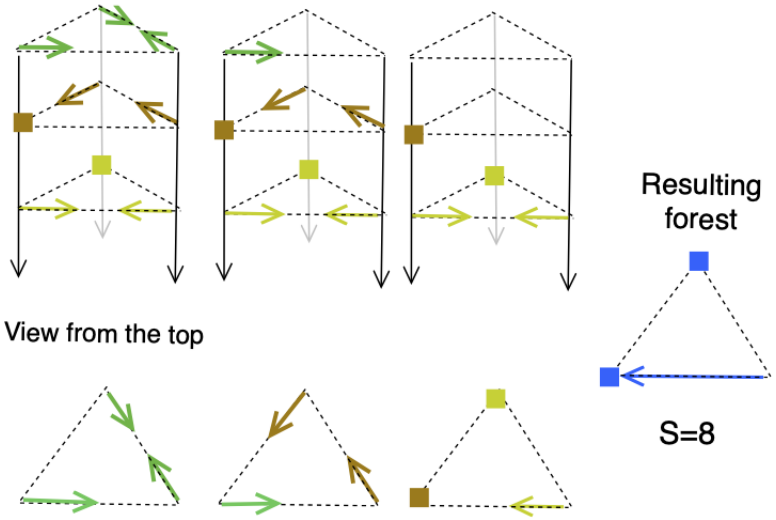


Figure 1: Cycle erasure in the stacks viewed from the top, in a triangle. The squares in the stacks corresponds to stops. S is the number of arrows or stops that have to be read from the stacks to obtain a rooted forest.

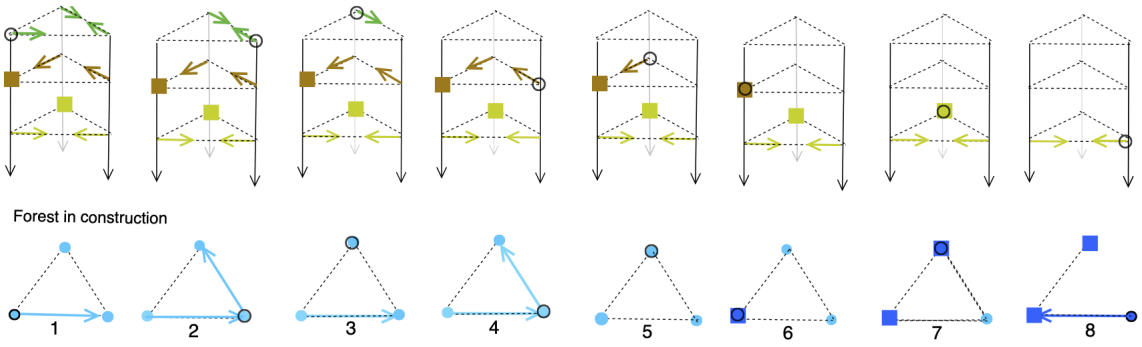


Figure 2: Constructing the forest by reading the stacks according to Wilson's order. The light blue nodes are active, while the dark blue ones are frozen. The current node is marked by a black circle. S is now the number of steps, i.e. reading a stop or an arrow, before the rooted forest is entirely constructed.

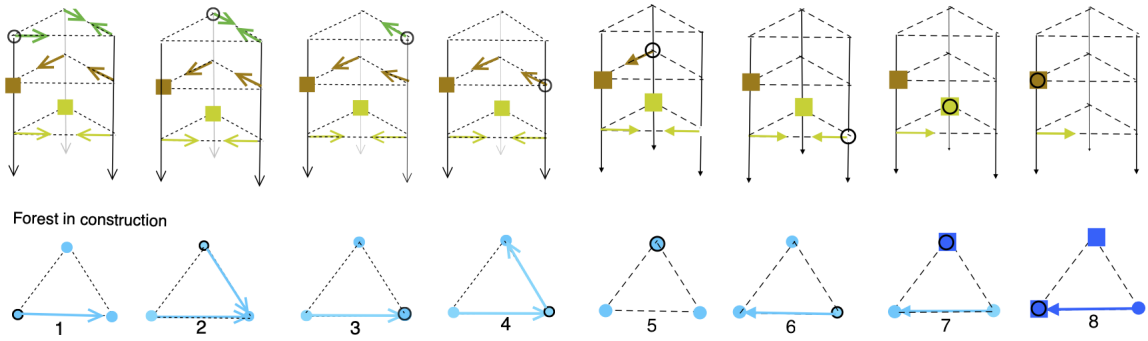


Figure 3: Constructing the forest by reading the stacks in random order. Unlike the Wilson's order, the current node is chosen randomly among the active roots.

$t = 1/q$, will also be read to compute the Kirchhoff forest $\Phi_{q'}$ with $q' < q$, at a later time $t' = 1/q'$. The jumps of the random forest trajectories can be described through an auxiliary random process that we introduce in the next section. The coupled forest process will be described as a non-homogeneous Markov process in section 3.4 and the sampling algorithm we actually implement is given in section 3.5.

3.3 An auxiliary Markov process: frozen and active trees.

For fixed $q > 0$, we can see the Kirchhoff forest Φ_q produced by Wilson's algorithm as the final configuration of a continuous time Markov process $(\Psi^q_s)_{s\geq 0}$ with a family of absorbing states or frozen configurations. As we will discuss later, this is *not* convenient from a practical, algorithmic, point of view. But this helps to describe the coupled forest process.

Consider a rooted spanning forest and declare each tree to be either *active* or *frozen*. Such forests ϕ with active and frozen trees will be the configurations of our Markov process. Each root is also declared to be active or frozen, in accordance with the status of the tree it belongs to. We will denote by (ϕ, A) the configuration with active root set $A \subset \rho(\phi)$ in the forest ϕ, and by $\Psi^{q;\phi,A}$ the process Ψ^q started at (ϕ, A).

Here are the evolution rules of Ψ^q:

- frozen roots will remain such;
- each active root freezes at rate q, and so does the tree it belongs to;
- each active root in $x \in \mathcal{X}$ also adds at rate $w(x, y)$ the arrow (x, y) to the current forest ϕ with two possible outcomes:
  - a cycle appears in $\phi \cup \{(x, y)\}$, i.e. if (x, y) joins x to some node y in its tree. In this case, each arrow in the cycle is erased, the tree is fragmented, new roots appear at each site along the erased cycle, and ϕ is updated to this smaller forest (with less edges).
  - otherwise $\phi' = \phi \cup \{(x, y)\}$ is a rooted spanning forest and ϕ is updated to ϕ', so that the root x disappears, and two trees coalesce.

  In the fragmentation case, new trees are declared active trees. In the coalescence case, the status of the new tree is inherited from that of its root.

Ψ^q can then evolve in four different manners only:

- by freezing a tree;
- by fragmenting an active tree into smaller active trees;
- by grafting an active tree on another one to get a larger active tree;
- by grafting an active tree on a frozen tree to get a larger frozen tree.

Building Φ_q from $\Psi^{q;\emptyset,\mathcal{X}}$— By starting Ψ^q from the configuration with n active roots, (so that the initial forest ϕ is the empty one), running Ψ^q amounts to discovering marks and arrows of the stack representation in a random order determined by the successive times when the active roots add a new arrow. Indeed, the difference with Wilson's order, where only the current node can add an edge or freeze, is that here all active roots can add edges and freeze. This is illustrated in Figure 3. The abelianity implies that Φ_q is the final configuration of $\Psi^{q;\emptyset,\mathcal{X}}$.

Building $\Phi_{q'}$ from $\Psi^{q';\Phi_q,W}$ with $q > q'$ and W a random subset of $\rho(\Phi_q)$— By conditioning on Φ_q, the stack distribution of the non-erased arrows and marks is described with the same product measure as the original one with an additional bias: the first mark at each root $x \in \rho(\Phi_q)$ is uniformly distributed on $[0, q/(q + w(x))]$ instead of $[0, 1]$. Hence, in order to build $\Phi_{q'}$ from Φ_q, we only have to sample a set W of roots to unfreeze by independently choosing each root of $x \in \rho(\Phi_q)$ with a suitable probability p_x, before running $\Psi^{q';\Phi_q,W}$ up to freezing. According to our original coupling, the probability p'_x that a frozen root x of Φ_q remains such for our smaller q' without reading any further arrow in x, is the probability that a uniform variable on $[0, q/(q + w(x))]$ is smaller than $q'/(q' + w(x))$:

$$p'_x = \frac{q'/(q' + w(x))}{q/(q + w(x))}.$$

By running $\Psi^{q';\Phi_q,W}$ up to freezing, this is achieved in two ways only: by keeping x out of W, or by unfreezing it before discovering that the first *unbiased* mark in x is below $q'/(q'+w(x))$. The probability p_x must then solve

$$(1-p_x)+p_x\frac{q'}{q'+w(x)}=p'_x,$$

so that

$$p_x=1-\frac{q'}{q}. \tag{13}$$

Note that p_x does not depend on x.

3.4 Coupled forests as a non homogeneous Markov process.

After observing that $(\Phi_\infty,\emptyset)$ is the deterministic final frozen configuration of $\Psi^{\infty;\emptyset,\mathcal{X}}$, we can finally describe $\Phi_{1/\cdot}$ as a non-homogeneous Markov process on the frozen forests, started at $\Phi_\infty=\emptyset$, and indexed by the time $t=1/q$.

In the limit $q\to q'_+$ and using

$$-\frac{dq}{q}=\frac{dt}{t},$$

Equation (13) gives that each root x in $\rho(\Phi_{1/t_-})$ wakes up at rate $1/t$, and $\Phi_{1/t}$ is obtained as the final forest of $\Psi^{1/t}$ started in $(\Phi_{1/t_-},\{x\})$. The difficulty raised by the fact that rates $1/t$ diverge in the neighbourhood of 0 is removed by observing that any wake-up of a given root x at a given time t is ineffective if x is frozen again before putting any arrow to some neighbour $y\neq x$. This occurs with probability

$$\frac{q}{q+w(x)}=\frac{1}{1+tw(x)}=1-\frac{tw(x)}{1+tw(x)}.$$

It follows that the effective wake-up rate in x at time t is $w(x)/(1+tw(x))$, which is bounded.

By considering only effective wake-up events, we obtain the following description of the frozen forest process $\Phi_{1/\cdot}$, which starts from the empty forest with n frozen roots:

- each root is unfrozen at rate $w(x)/(1+tw(x))$;
- when a root x is unfrozen at time $t=1/q$, an arrow (x,y) is chosen with probability $w(x,y)/w(x)$ to be added to Φ_{1/t_-} before being processed with grafting or pruning;
- in the pruning case —when y is rooted in x in Φ_{1/t_-}— $\Phi_{1/t}$ is the final configuration of $\Psi^{1/t;\phi,A}$, with ϕ the pruned forest obtained by removing from $\Phi_{1/t_-}\cup\{(x,y)\}$ all the arrows that form its only one cycle, and with A the foot set of those arrows;
- in the grafting case —when y is not rooted in x in Φ_{1/t_-}— $\Phi_{1/t}$ is equal to $\phi=\Phi_{1/t_-}\cup\{(x,y)\}$, final and initial configuration of $\Psi^{1/t;\phi,\emptyset}$.

3.5 Wilson's order

While relevant to describe $(\Phi_{1/t})_{t\geq 0}$ as a non-homogeneous Markov process, the update rule account at unfreezing times $t=1/q$ through the Markov process $(\Psi^q_s)_{s\geq 0}$ is not convenient for practical sampling. Indeed, exploring the stacks in this random order, with many active trees evolving *in parallel* with a same continuous time s, would *a priori* imply to read many times the already sampled arrows to decide between fragmentation and coalescence at each new arrow occurrence. By running *successive* loop-erased random walks, Wilson's exploring order deals in contrast with only one non-trivial active tree at each time, actually reduced to a single branch. In this situation, no extra read is needed to decide between coalescence or fragmentation (either the new arrow points to an active node and the active tree is fragmented, or the new arrow points to a frozen node and the tree coalesces to another frozen one). Extra reads are possibly required for cycle erasure only, in the fragmentation case[‡].

For this reason we will stick to Wilson's order in building $\Phi_{1/t}$ from the configuration $(\Phi_{1/t_-},\{x\})$ with one active tree only, when some root x wakes up at time t. After adding and processing a new

[‡]When sampling a single forest Φ_q for a given q, the last version of the algorithm in [Wil96] actually avoids any cycle erasure. We reproduce it in Appendix B.

arrow (x, y), we will always freeze or add and process arrows from the root of the tree that covers x in the current forest, until freezing or grafting on a frozen tree. We will then choose another active root x, which will necessarily belong to the nodes of the unfrozen tree at time t, and proceed in the same way. In repeating this procedure up to complete freezing, we may have to read again some previously sampled arrows, used to build Φ_{1/t_-}, while building $\Phi_{1/t}$. However, by proceeding however in this order, any read again or re-sampled arrow that might be followed to decide between coalescence and fragmentation would be part a cycle to be immediately erased. As a consequence and as detailed in the pseudo-code appendix B, we can tag each node where an arrow as been read again or re-sampled, before re-sampling its mark and arrow each time we are led to this node again. Each needed extra read, which is associated with a previously sampled arrow used to build Φ_{1/t_-}, is performed at most once for each node initially covered by the active tree we started with. This is the key for the control in Theorem 3 of the number of extra reads required, and is illustrated in Figure 4.

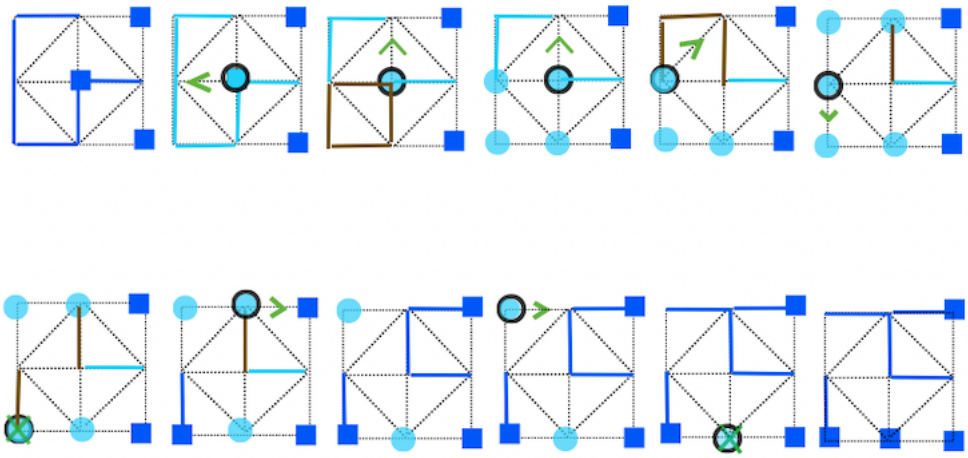


Figure 4: Computing the extra reads in the coupled forest algorithm following Wilson's order. The dark blue trees are frozen. The light blue ones are active. The root of the current tree is marked by a black circle, and the next arrow or stop in the corresponding stack is indicated in green, a cross corresponding to a stop. New sampled arrows are in brown, as are the arrows that we have to read again in order to decide between a cycle or a grafting to an active tree. Once in brown, the decision between "cycle or grafting" does not need more extra reads: if the current root points to a brown arrow, there is a cycle; if it points to a light blue one, the current tree coalesces. The number of reread arrows R is the number of times one light blue arrow becomes brown ($R = 5$ in our example).

In Appendix B we give a pseudocode description of this procedure to build $(\Phi_q)_{q_{\max} \geq q \geq q_{\min}}$ for $q_{\max} > q_{\min} > 0$ in terms of two algorithms. The first one follows Wilson's algorithm to sample $\Phi_{q_{\max}}$ in terms of a list of all node successors together with a priority queue Q that orders the roots x of $\Phi_{q_{\max}}$ according to their unfreezing time $1/q_x$. The second one extends this priority queue Q to describe all unfreezing times $t = 1/q$ between $1/q_{\max}$ and $1/q_{\min}$ together with the successor lists that describe the associated forests Φ_q.

3.6 Sampling costs.

We turn now to the cost of the algorithm described in section 3.5 to sample one trajectory $\Phi_{1/t}$ on the whole time interval $[0, 1/q_0]$. This cost is essentially made of two ingredients:

- the sampling cost of the number S of arrows to be sampled, in order to build this trajectory;
- the total number R of extra reads of already sampled arrows, which are possibly kept in the new forest built at each unfreezing time.

Theorem 3 gives estimates on the mean number $\mathbb{E}(S)$ and $\mathbb{E}(R)$. It will be proved in Section 5.

Theorem 3. *Let W denote the diagonal matrix whose entries are $(w(x), x \in \mathcal{X})$. For all $q_0 > 0$, the sampled arrows number S and the extra reads number R needed to build one random forest trajectory*

$$t \in [0, 1/q_0] \mapsto \Phi_{1/t}$$

satisfy

$$\mathbb{E}[S] = \mathrm{Tr}\big((q_0\mathrm{Id} - L)^{-1}(q_0\mathrm{Id} + W)\big) \leq n\left(1 + \frac{\bar{\lambda}}{q_0}\right)$$

and

$$\mathbb{E}[R] \leq n \ln\left(1 + \frac{\alpha}{q_0}\right).$$

Using Walker's alias algorithm to sample arrows at a given vertex at cost $O(1)$, we get a total sampling cost for one trajectory in $O(n(\bar{\lambda}q_0^{-1} + \ln(\alpha q_0^{-1}))$ for large n and small $q_0/\bar{\lambda}$. In the case when we start from the weight matrix, this of course amounts to assuming that each non-zero entry has been previously read once. The numerical cost of this operation is of the order of the edge number. However as soon as we go to Monte Carlo estimation of the expectation of different observables along our random trajectories, i.e., empirical mean computations for these observables, the repeated sampling cost is almost linear in the number of nodes and not the number of edges, as long as q_0 is a fixed small fraction ϵ of the mean eigenvalue $\bar{\lambda}$, that is $\epsilon = q_0/\bar{\lambda} \leq nq_0/\alpha$.

4 Estimating the spectral cumulative distribution function.

We turn now to the inverse problem of getting estimates of the cumulative distribution function of σ

$$F : q \in [0, 2\alpha] \mapsto \frac{1}{n}\left|\{j < n : \lambda_j \leq q\}\right|$$

from the estimates of the Stieltjes transform and its derivatives defined in section 2. We look at this problem through a moment problem. Recall that σ is indeed supported on $[0, 2\alpha]$ since there exists a unique stochastic matrix P satisfying

$$L = \alpha(P - \mathrm{Id});$$

the spectrum of any symmetric stochastic matrix P being between -1 and 1, we get

$$0 \leq \lambda_j \leq 2\alpha, \qquad j < n.$$

4.1 A moment problem

Take a positive $q \leq 2\alpha$ and consider the random variable

$$Y_q = \frac{q}{q + \lambda_J} \in \left[\frac{q}{q + 2\alpha}, 1\right]$$

with J uniformly distributed on $\{0, \cdots, n-1\}$. After rescaling, an estimate $\hat{\xi}_q^k$ of

$$\mathbb{E}\big[|\xi_q^k|\big] = \sum_{j<n}\left(\frac{q}{q + \lambda_j}\right)^k,$$

for $k \geq 1$, is an estimate $\hat{m}_k$ of the k^{th} non-trivial moment of Y_q

$$m_k = E\big[Y_q^k\big] = \frac{1}{n}\sum_{j<n}\left(\frac{q}{q + \lambda_j}\right)^k.$$

Since

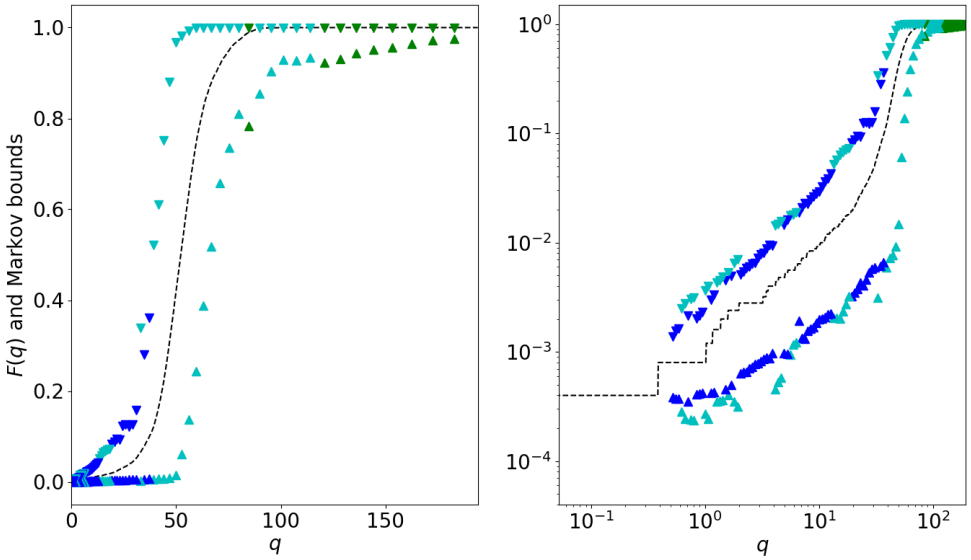
$$F(q) = \frac{1}{n}\left|\{j < n : \lambda_j \leq q\}\right| = P\left(Y_q \geq \frac{1}{2}\right)$$

we are led to the problem of estimating the tail distribution function G of a bounded-support random variable Y_q with values in $[a, b] = [q/(q + 2\alpha), 1]$ at one precise point $y = 1/2$, given its first l non-trivial

moments $m_1, \ldots, m_l$. This problem was solved by Markov who gave the sharpest possible lower and upper bounds on $G(y)$ given the constraints. We refer to [KN77] and Appendix C in the present paper for details.

In the particular case of the unweighted ($w(x,y) \in \{0,1\}$) Stanford bunny graph —a subsampled surface of a ceramic bunny figuring, in which each point is connected to a few dozens of its nearest neighbours— Figure 5 shows Markov bounds, computed from Monte Carlo estimation $\hat{m}_k$ of m_k with $k \leq l = 4$, along $s = 400$ forest trajectories with l replicas (a total of 1600 forest trajectories), on $q \in [q_0, 2\alpha] \mapsto F(q)$ with $q_0 = \bar{\lambda}/100$. Colors show in each value of q the number of "valid" moment estimates: not all finite sequences $\hat{m}_1, \ldots, \hat{m}_k$ are indeed admissible moment sequences for a probability distribution on a given interval $[a,b]$ (see section C for a precise definition) . In first approximation, we say that the number of valid moment estimates is the largest k for which any sequence $\hat{m}_1, \ldots, \hat{m}_{k-1}, \tilde{m}_k$ with $\tilde{m}_k$ in the k^{th} 95% confidence interval associated with our Monte Carlo estimates, is an admissible moment sequence. We refer to Section 6 for more details.

Figure 5: Cumulative distribution function (dashed line) in natural (left) and log-log (right) scales of the spectral measure for the bunny graph with 2053 nodes and mean degree 52.33 together with lower (upward triangles) and upper (downward triangles) bounds computed from Monte Carlo estimation of $m_1, \ldots, m_4$ after sampling 400 replicated forest trajectories up to time $1/q_0$ with $q_0 = \bar{\lambda}/100$ and with 1 (yellow), 2 (green), 3 (cyan) or 4 (blue) valid moment estimates.



These are very bad, or quite bad, bounds on the cumulative distribution function F of the spectral measure σ. In this example $F(50)$ is essentially estimated by the trivial bounds 0 and 1. Only the median of the spectral measure or the spectrum bulk location can be roughly identified with Markov bounds of Figure 1: these are around $q = 45$. An explanation of this poor performance is that Markov bounds are reached by very singular distributions, supported by three atoms at most for $l = 4$, while Y_q can generically take n distinct values. We will go beyond Markov bounds by using a standard equivalence between canonical and microcanonical ensembles to make a prediction on F, instead of bounding it.

4.2 Maximal entropy predictions

Going from our Monte Carlo noisy estimates made from s samples $\xi^k_{q,i} \subset \mathcal{X}$, $i < s$,

$$\hat{m}_k : q \geq q_0 \mapsto \hat{m}_k(q) = \frac{1}{sn} \sum_{i<s} |\xi^k_{q,i}|, \qquad k \leq l,$$

of the moment functions

$$q \geq q_0 \mapsto \mathbb{E}[Y_q^k] = \frac{1}{n} \sum_{j<n} \left(\frac{q}{q + \lambda_j} \right)^k, \qquad k \leq l,$$

to an estimate of the cumulative distribution function

$$F : q \geq 0 \mapsto \frac{1}{n} \left| \{ j < n : \lambda_j \leq q \} \right|$$

can be made in many different ways with many different designs for a statistical approach. In this paper we simply point out that for each value of $q \geq q_0$ and for many different graphs, the few estimates $\hat{m}_1(q), \ldots, \hat{m}_l(q)$ (we take $l = 4$ in our applications) can already lead through low cost maximal entropy principle to sound predictions on $F(q)$ by themselves, i.e., disregarding all other information contained in $\hat{m}_k$, $k \leq l$, that are all the values of $\hat{m}_k(q')$ for $q' \neq q$.

To see this, fix again a positive $q \leq 2\alpha$. From s samples of l replicas, we get estimates $\hat{m}_1(q), \ldots, \hat{m}_l(q)$ of the first l non-trivial moments m_k of Y_q with a relative Monte Carlo error of order $1/\sqrt{sn}$. This means that we can build small $\epsilon_1, \ldots, \epsilon_l$ in the large n regime such that

$$m_k \in [\hat{m}_k(q)(1 - \epsilon_k), \hat{m}_k(q)(1 + \epsilon_k)], \qquad 1 \leq k \leq l,$$

with probability 0.95 at least. We claim that in a weak sense and as a consequence of an equivalence of ensembles —see Section 4.3 for a precise statement— if $\hat{m}_1(q), \ldots, \hat{m}_l(q)$ form an admissible moment sequence, then for all

$$y \in [a, b] = \left[\frac{q}{q + 2\alpha}, 1 \right]$$

the overwhelming majority in this large n regime of the atomic distributions

$$\nu = \frac{1}{n} \sum_{j<n} \delta_{y_j},$$

with $y_j \in [a, b]$ for all $j < n$ and

$$\int_a^b y^k \nu(dy) = \frac{1}{n} \sum_{j<n} y_j^k \in [\hat{m}_k(q)(1 - \epsilon_k), \hat{m}_k(q)(1 + \epsilon_k)], \qquad 1 \leq k \leq l, \tag{14}$$

have a tail probability $\nu([y, b])$ that is close to the tail probability $\nu^*([y, b])$ of the continuous probability distribution ν^* on $[a, b]$ that maximizes the entropy

$$h(\nu) = -\int_a^b \nu(dx) \log \frac{d\nu}{dx}$$

under the constraints

$$\int_a^b x^k \nu(dx) = \hat{m}_k(q), \quad k \leq l. \tag{15}$$

Defining for any $\beta = (\beta_1, \ldots, \beta_l) \in \mathbb{R}^l$

$$\nu_\beta(dy) = \frac{1}{\Xi_\beta} \exp\left\{ -\sum_{k=1}^{l} \beta_k y^k \right\} dy, \qquad \Xi_\beta = \int_a^b \exp\left\{ -\sum_{k=1}^{l} \beta_k y^k \right\} dy, \tag{16}$$

it holds $\nu^* = \nu_{\beta^*}$ with β^* the unique minimizer of the smooth and convex function

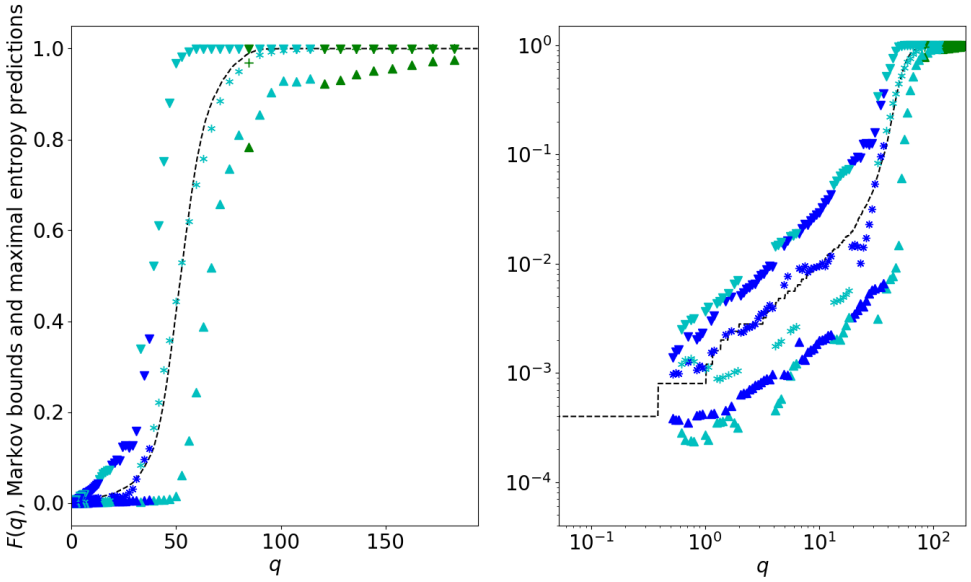
$$\beta \in \mathbb{R}^l \mapsto \ln \Xi_\beta + \sum_{k=1}^{l} \beta_k \hat{m}_k(q),$$

which can be easily computed by Newton's method in small dimension l (see for example [MP84]) except when β^* is large, i.e., ν_{β^*} is strongly concentrated around a few points. But in this case Markov bounds can already provides a good estimation for $F(q)$. And we can take $\nu_{\beta^*}([1/2, 1])$ as predicted value for $F(q)$ in the generic case.

Figure 6 adds to Figure 5 the predicted values $\nu_{\beta^*}([1/2, 1])$ for $F(q)$. We found them much less disappointing than the previous Markov bounds. More numerical experiments are presented in Section 6.

We will argue in Section 6 that, to be consistent with the relative errors ϵ_k of order $1/\sqrt{sn}$ for s sampled trajectories, the computational cost of these predictions is of order $\sqrt{sn}$ only. It is then negligible with respect to the numerical cost for estimating $m_1, \ldots, m_l$.

Figure 6: Cumulative distribution function (dashed line) in natural (left) and log-log (right) scales of the spectral measure for the bunny graph with 2053 nodes and mean degree 52.33 together with its lower (upward triangles) and upper (downward triangles) bounds and its predictions (yellow vertical segment, green cross, cyan or blue asterisks) computed from Monte Carlo estimation of $m_1, \ldots, m_4$ after sampling 400 replicated forest trajectories up to time $1/q_0$ with $q_0 = \bar{\lambda}/100$ and with 1 (yellow), 2 (green), 3 (cyan) or 4 (blue) valid moment estimates.



4.3 Equivalence of ensembles

The material of this section is rather standard. It is intended to make precise our previous claim on the "overwhelming majority" of the atomic measures ν that satisfy the constraint (14). Let $\bar{V}_n$ be the set of atomic measures on a interval $[a, b]$

$$\nu = \frac{1}{n} \sum_{j<n} \delta_{y_j}$$

that satisfy the constraint

$$\int_a^b y^k \nu(dy) \in [m_k(1-\epsilon_k), m_k(1+\epsilon_k)], \qquad k = 1, \ldots, l, \tag{17}$$

for an admissible moment sequence $m_1, \ldots, m_l$ and small $\epsilon_1, \ldots, \epsilon_l > 0$.

For making sense of the notion of "overwhelming majority of elements of $\bar{V}_n$", we first have to define a probability distribution on the set $\bar{V}_n$. i.e., to define a random atomic measure $\tilde{\zeta}$ that satisfies the moment constraints (17). We will then be able to check that, for all $\epsilon > 0$ and all $y \in [a, b]$,

$$\lim_{n\to\infty} P\big(|\tilde{\zeta}([y, b]) - \nu^*([y, b])| > \epsilon\big) = 0. \tag{18}$$

Building $\tilde{\zeta}$ and checking (18). Let μ_{β^*} be the canonical Gibbs measure on $[a, b]^n$ defined by

$$\mu_{\beta^*}(dz) = \frac{1}{\Xi^n_{\beta^*}} \exp\left\{ -\sum_{k=1}^{l} \beta^*_k \sum_{j<n} z_j^k \right\} dz,$$

with Ξ_{β^*} the partition function of ν_{β^*} defined in (16). We denote by $\zeta = (\zeta_j)_{j<n}$ the random variable in $[a, b]^n$ with law μ_{β^*}. The coordinates ζ_j, $j < n$, are independent random variables with the same law ν_{β^*} on $[a, b]$. With

$$V_n = \left\{ z \in [a, b]^n : \forall 1 \le k \le l, \ \frac{1}{n} \sum_{j<n} z_j^k \in [m_k(1-\epsilon_k), m_k(1+\epsilon_k)] \right\},$$

it holds then

$$\lim_{n\to\infty} P(\zeta \in V_n) = 1$$

by the weak law of large numbers. By associating with each $z \in [a,b]^n$ a probability

$$\bar{z} = \frac{1}{n}\sum_{j<n} \delta_{z_j}.$$

on $[a,b]$ we build the random atomic measure $\tilde{\zeta}$ by conditioning $\bar{\zeta}$ to

$$\{\bar{\zeta} \in \bar{V}_n\} = \{\zeta \in V_n\}.$$

While ζ conditioned to $\{\zeta \in V_n\}$ is not distributed according to the Lebesgue measure on V_n, the microcanonical ensemble obtained by a further conditioning on $\sum_{j<n} \zeta_j^k$, $k \le l$, can be seen as a "uniform distribution" on our atomic measures with moments $\sum_{j<n} \zeta_j^k/n$, $k \le l$. Together with the fact that we are working with small ϵ_k, the law of $\tilde{\zeta}$ does then provide a reasonable statistical model on $\bar{V}_n$. Within this model, Equation (18) immediately follows from

$$P\big(|\tilde{\zeta}([y,b]) - \nu_{\beta^*}([y,b])| > \epsilon\big) = \frac{P\big(|\bar{\zeta}([y,b]) - \nu_{\beta^*}([y,b])| > \epsilon, \zeta \in V_n\big)}{P\big(\zeta \in V_n\big)} \le \frac{P\big(|\bar{\zeta}([y,b]) - \nu_{\beta^*}([y,b])| > \epsilon\big)}{P\big(\zeta \in V_n\big)}$$

and the weak law of large number again. □

5 Proofs of Theorems 2 and 3

5.1 Proof of Theorem 2

Recall that, for a given node x and forest ϕ, $\rho_x(\phi)$ is the root of the tree that covers x in ϕ. By using again the abelianity of Wilson's algorithm, we can assume x to be our first starting point for the loop-erased random walks. Consider the discrete time random walk $(\tilde{X}_k)_{k\ge 0}$ on $\mathcal{X} \cup \{\Delta\}$, where Δ is a cemetery state, with transition probabilities

$$\tilde{P}(x,y) = \begin{cases} \frac{w(x,y)}{q+w(x)} & \text{if } x \neq y,\ x,y \neq \Delta \\ \frac{q}{q+w(x)} & \text{if } x \neq \Delta, y = \Delta \\ 0 & \text{if } x = y \neq \Delta \\ 1 & \text{if } x = y = \Delta. \end{cases} \tag{19}$$

Call T_Δ the absorption time of the chain in Δ. For all x and y in $\mathcal{X}$,

$$\begin{aligned} \mathbb{P}\big(\rho_x(\Phi_q) = y\big) &= \mathbb{P}_x\big(\tilde{X}(T_\Delta - 1) = y\big) = \sum_{k\ge 1} \mathbb{P}_x\big(\tilde{X}_{k-1} = y, \tilde{X}_k = \Delta\big) \\ &= \frac{q}{q+w(y)} \sum_{k\ge 1} \tilde{P}_{\mathcal{X}}^{k-1}(x,y) = \frac{q}{q+w(y)} \Big(\mathrm{Id} - \tilde{P}_{\mathcal{X}}\Big)^{-1}(x,y) \\ &= q\Big(\mathrm{Id} - \tilde{P}_{\mathcal{X}}\Big)^{-1}\Big(q\mathrm{Id} + W\Big)^{-1}(x,y), \end{aligned}$$

where $\tilde{P}_{\mathcal{X}}$ is the square submatrix of $\tilde{P}$ restricted to $\mathcal{X}$. Note that $\tilde{P}_{\mathcal{X}} = \Big(q\mathrm{Id} + W\Big)^{-1}(L+W)$, so that

$$\Big(q\mathrm{Id} + W\Big)\Big(\mathrm{Id} - \tilde{P}_{\mathcal{X}}\Big) = q\mathrm{Id} + W - L - W = q\mathrm{Id} - L.$$

Setting $K_q = q(q\mathrm{Id} - L)^{-1}$, this gives $\mathbb{P}\big(\rho_x(\Phi_q) = y\big) = K_q(x,y)$ and therefore, for all $k \ge 0$,

$$\mathbb{P}\big(R^k(x) = y\big) = K_q^k(x,y).$$

Hence,

$$\mathbb{E}\big[|\xi_q^k|\big] = \sum_{x\in\mathcal{X}} \mathbb{P}\big(x \in \xi_q^k\big) = \sum_{x\in\mathcal{X}} \mathbb{P}\big(R^k(x) = x\big) = \sum_{x\in\mathcal{X}} K_q^k(x,x) = \mathrm{Tr}\big(K_q^k\big) = \sum_{j<n} \Big(\frac{q}{q+\lambda_j}\Big)^k.$$

Let us now think of the forests $\Phi_{k,q}$, $k < l$, as forests on different layers that are l distinct copies of $\mathcal{X}$. Since negative correlations are relevant for positive k only, we choose $k \geq 1$ and observe that for checking the occurrence of $\{x, y \in \xi_q^k\}$ with $x \neq y$, we can first sample the multilayer path leading from x to $R_q^k(x)$, with one tree branch in each of the first k layers, before sampling the multilayer path leading from y to $R_q^k(y)$. Negative correlations simply follow from the fact that conditionally on $\{x \in \xi_q^k\}$, in order to have $\{y \in \xi_q^k\}$, the second path must avoid the first one, since otherwise we would have

$$R_q^k(y) = R_q^k(x) = x \neq y.$$

Using the negative correlations property, the variance bound is standard:

$$\begin{aligned}\mathrm{Var}\big(|\xi_q^k|\big) &= \sum_x \mathbb{P}\big(x \in \xi_q^k\big) + \sum_{x \neq y} \mathbb{P}\big(x \in \xi_q^k, y \in \xi_q^k\big) - \sum_x \mathbb{P}\big(x \in \xi_q^k\big)^2 - \sum_{x \neq y} \mathbb{P}\big(x \in \xi_q^k\big)\mathbb{P}\big(y \in \xi_q^k\big) \\ &\leq \sum_x \mathbb{P}\big(x \in \xi_q^k\big) = \mathbb{E}\big[|\xi_q^k|\big]\,.\end{aligned}$$

□

5.2 Proof of Theorem 3

Using the stack representation, the arrows used in the stack of x in order to construct Φ_q correspond to marks $U > \frac{q}{q+w(x)}$. Since for $q \geq q'$, $\frac{q}{q+w(x)} \geq \frac{q'}{q'+w(x)}$, the same arrows are used to construct Φ'_q. Hence, the number S of sampled arrows necessary to build the whole trajectory

$$t \in [0, 1/q_0] \mapsto \Phi_{1/t},$$

is exactly the number of sampled arrows necessary to build Φ_{q_0} with Wilson's algorithm. To get an upper bound on its expected value, we simply observe that the number of arrows and marks to be sampled at each node x is stochastically dominated by the depth at which the first mark below $q_0/(q_0 + w(x))$ appears in the associated stack. This is a geometric random variable with mean $1 + w(x)/q_0$. It follows that

$$\mathbb{E}[S] \leq \sum_{x \in \mathcal{X}} 1 + \frac{w(x)}{q_0} = n + \frac{1}{q_0}\mathrm{Tr}(-L) = n\left(1 + \frac{\bar{\lambda}}{q_0}\right).$$

To prove the identity

$$\mathbb{E}[S] = \mathrm{Tr}\big((q_0\mathrm{Id} - L)^{-1}(W + q_0\mathrm{Id})\big) \tag{20}$$

we consider again the discrete time random walk $(\tilde{X}_k)_{k \geq 0}$, whose transition probabilities are given in (19). For every $x \in \mathcal{X}$, $\ell(x) = \sum_{k=0}^{T_\Delta - 1} \mathbb{1}_{\{\tilde{X}_k = x\}}$, is the local time spent at x before absorption. We denote by $M(x, y)$ the expectation of $\ell(y)$ when the starting node is x, i.e.

$$M(x, y) = \mathbb{E}_x[\ell(y)].$$

If we start Wilson's algorithm at x, the expected number of arrows in the stack of x that we need to read is given by $\mathbb{E}_x[\ell(x)]$. Therefore, by linearity of expectation, what we need to compute is

$$\mathbb{E}[S] = \mathrm{Tr}(M).$$

In order to prove (20) it is then sufficient to show that

$$M = (q_0\mathrm{Id} - L)^{-1}(W + q_0\mathrm{Id}).$$

Let us notice that, for every couple $x, y \in \mathcal{X}$ we have

$$M(x, y) = \mathbb{1}_{\{x=y\}} + \sum_{z \in \mathcal{X} \setminus \{x\}} \frac{w(x, z)}{q_0 + w(x)} M(z, y),$$

thanks to Markov property at time 1. Multiplying by $(q_0 + w(x))$ both sides of the equality we get

$$\begin{aligned}q_0 M(x, y) &= \mathbb{1}_{\{x=y\}}(q_0 + w(x)) - w(x)M(x, y) + \sum_{z \neq x} L(x, z)M(z, y) \\ &= \mathbb{1}_{\{x=y\}}(q_0 + w(x)) + LM(x, y),\end{aligned}$$

which reads

$$(q_0\mathrm{Id} - L)M = (q_0\mathrm{Id} + W),$$

from which (20) immediately follows.

We note that our upper bound on $\mathbb{E}[S] = \mathrm{Tr}M$ can also be deduced from (20) by computing the trace of M in an orthonormal basis of eigenvectors of L. This shows that this upper bound is sharp for graphs whose spectral measure is concentrated around $\lambda_0 = 0$.

We turn now to R. Since each root x of $\Phi_{1/t}$ is unfrozen at rate $w(x)/(1+w(x)t)$, and when unfreezing x a time t the number extra reads is bounded by the size $|\tau_x(\Phi_{1/t})|$ of the tree that covers x in $\Phi_{1/t}$, it holds

$$\begin{aligned}
\mathbb{E}[R] &\leq \mathbb{E}\left[\int_0^{1/q_0} \sum_{x\in\rho(\Phi_{1/t})} \left|\tau_x(\Phi_{1/t})\right| \frac{w(x)}{1+w(x)t}\mathrm{dt}\right] \\
&\leq \mathbb{E}\left[\int_0^{1/q_0} \frac{\alpha}{1+\alpha t} \sum_{x\in\rho(\Phi_{1/t})} \left|\tau_x(\Phi_{1/t})\right| \mathrm{dt}\right] \\
&= n\int_0^{1/q_0} \frac{\alpha}{1+\alpha t}\mathrm{dt} = n\ln\left(1+\frac{\alpha}{q_0}\right).
\end{aligned}$$

□

6 Numerical issues and results

6.1 Subsampling the Stieltjes transform and its first derivatives

Our coupled forest algorithm defined and described in Section 3.2, Section 3.4 and Appendix B allows to build a single piecewise constant random forest trajectory $(\Phi_q)_{2\alpha\geq q\geq\epsilon_0\bar{\lambda}}$ at a sampling cost in $O\left(n\left(\frac{1}{\epsilon_0}+\ln\left(\frac{\alpha}{\epsilon_0\bar{\lambda}}\right)\right)\right)$ for large n and small ϵ_0.

For each value of q in $[\epsilon_0\bar{\lambda}, 2\alpha]$, the estimates of the Stieltjes transform and its derivatives given by (12) are based on s samples of l replicas of such trajectories. l is typically small ($l = 4$ in our experiments). For each sample, (12) requires the computation of the size of $\xi_q^1, \ldots, \xi_q^l$ defined in (10). This counting cost is in $O(n)$, for each sample and each q.

As a consequence we decided to make these estimates in $\lceil 1/\epsilon_0\rceil$ values of q, naturally organized in a geometric progression $q_i = q_0 r^i$, $i < 1/\epsilon_0$, with

$$r = \exp\left\{\epsilon_0 \ln\frac{2\alpha}{q_0}\right\},$$

so that $q_0 r^{1/\epsilon_0} = 2\alpha$. The total sampling and counting cost for all these values of q is then in $O\left(n\left(\frac{1}{\epsilon_0}+\ln\left(\frac{\alpha}{\epsilon_0\bar{\lambda}}\right)\right)\right)$ for each replicated coupled forest trajectory. Since $\alpha/n \leq \bar{\lambda} \leq \alpha$ (recall that $n\bar{\lambda} = \mathrm{Tr}(-L) = \sum_{x\in\mathcal{X}} w(x)$), one obtains a sampling cost which is at most $O\left(\frac{n\ln n}{\epsilon_0}\right)$ for each replicated coupled forest trajectory.

6.2 Newton's descent algorithm, numerical integration and valid moment estimates

At a total computational cost in $O(s\epsilon_0^{-1} n\ln n)$ and for all $q = q_i$, $i \leq 1/\epsilon_0$, we get then estimates $\hat{m}_k(q)$ of the moments

$$m_k = \frac{1}{n}\sum_{j<n}\left(\frac{q}{q+\lambda_j}\right)^k, \qquad k \leq l,$$

with a relative error of order $1/\sqrt{sn}$ associated with a Monte Carlo confidence interval. In order to find the maximal entropy estimator of $F(q)$ at a given q in $[q_0, 2\alpha]$, we have to perform a numerical integration to compute the low dimensional Hessian to be inverted in Newton's method. This can then be made with the same relative precision and has therefore a computational cost in $O(\sqrt{sn})$, provided that we can ensure fast convergence of Newton's method, which requires to address two related issues.

First, the constraint (15) must be satisfiable by some measure ν on $[a,b]$. By using Lemma C.1 from Appendix C.4 we can check recursively that $\hat{m}_1(q)$, ..., $\hat{m}_k(q)$, $k \leq l$, form an admissible moment sequence : $\hat{m}_1(q)$ must lie in $[a,b]$ and $\hat{m}_k(q)$ has to be in an explicit interval I_k computed from $\hat{m}_1(q)$, ..., $\hat{m}_{k-1}(q)$. Since each estimate $\hat{m}_k(q)$ comes with its confidence interval, our prediction will use only the longest sequence $\hat{m}_1(q)$, ..., $\hat{m}_k(q)$ for which $\hat{m}_1(q)$, ..., $\hat{m}_{k-1}(q)$, $\tilde{m}_k$ form an admissible moment sequence for all $\tilde{m}_k$ in the Monte Carlo 95% confidence interval associated with $\hat{m}_k(q)$.

Second, for such an admissible moment sequence $\hat{m}_1(q)$, ..., $\hat{m}_k(q)$, despite the fact that our maximal entropy estimator with k non-trivial estimated moment is well-defined, we should expect to run into numerical difficulties in the case when our estimator would be associated with a large β^*, solution of

$$\int_a^b y^j \nu_{\beta^*}(dy) = \hat{m}_j(q), \qquad j \leq k. \tag{21}$$

But in this case Markov's bounds on $F(q)$, which are given in Appendix C.6 and explicitly computed from $\hat{m}_1(q)$, ..., $\hat{m}_k(q)$, can already provide a good estimate of $F(q)$. In practice, we will make use of Newton's method only in the generic case when Markov's bounds does not provide an estimate on $F(q)$ with a relative error of 1% or less.

In order to minimize the iteration number of each Newton's descent algorithm to be run for a given value q_i and with k non-trivial moment estimated, we will start the descent from the parameter β^* computed at the previous value $q_{i+1} > q_i$ with the same k, if this parameter is available. When this is not the case, we use the parameter β^* computed with one moment less at the same q_i. We will also fix a maximal iteration number equal to 50 to deal with the exceptional cases when Markov's bounds are not sufficient for providing a good estimate of $F(q)$ but Newton's algorithm does not reach an approximate solution of (21) in a few iterations only. This ensures that the extra computational cost for giving Markov bounds and computing the maximal entropy estimator of $F(q_i)$, for all $i < 1/\epsilon_0$, is in $O(\epsilon_0^{-1}\sqrt{sn})$ which is negligible with respect to the sampling cost $O(\epsilon_0^{-1} sn \ln n)$.

We can finally precisely define the number of valid moment estimates that appears for each value of q through the different colors in Figure 5 and Figure 6 as in the experimental results presented below. For all $k \leq l$, we say that $\hat{m}_1(q)$, ..., $\hat{m}_k(q)$ is a valid moment estimate

- if so is $\hat{m}_1(q)$, ..., $\hat{m}_{k-1}(q)$;
- if $\hat{m}_1(q)$, ..., $\hat{m}_{k-1}(q)$, $\tilde{m}_k$ form an admissible moment sequence for all $\tilde{m}_k$ in the Monte Carlo 95% confidence interval associated with $\hat{m}_k(q)$;
- and if $\hat{m}_1(q)$, ..., $\hat{m}_k(q)$ either provides an estimate of $F(q)$ within a relative error of less than 1% through Markov bounds or leads to the computation of its maximal entropy estimator in less than 50 iterations of Newton's algorithm.

6.3 Experimental results

In all pictures of this section, we follow the same conventions as in Figure 6 to show our experimental results. Except when otherwise specified, we considered unweighted graphs with $n = 5000$ nodes, we took $\epsilon_0 = q_0/\bar{\lambda} = .01$ and we sampled for each graph 400 coupled replicated forest trajectories with $l = 4$ replicas, i.e., 1600 coupled forest trajectories $(\Phi_q)_{2\alpha \geq q \geq q_0}$.

6.3.1 What went right

Figure 7 shows our results for two different Erdős-Rényi graphs of size $n = 5000$ with small density $p = (3 \ln n)/n$.

Figure 8 shows our results for two different Erdős-Rényi graphs with mean degree 100.

Figure 9 shows our results for two different Barabási-Albert preferential attachment random networks with $m = 5$ new edges for each new node.

Figure 10 shows our results for two different realisations of a stochastic bloc model with 20 quite well formed communities of the equal sizes, with intra and inter connection probabilities equal to .049 and .0008, respectively, resulting in an average degree equal to 16.

Figure 11 shows our results for two different sensor graphs, for which each node in a random set of points is connected with its 5 nearest neighbours.

Figure 12 shows our results for two deterministic graph. On the left the comet graph, which is a rooted tree with 499 branches of length 1 and 1 branch of length 4500. On the right the 71×71 torus.

Figure 7: Two sparse Erdős-Rényi graphs.

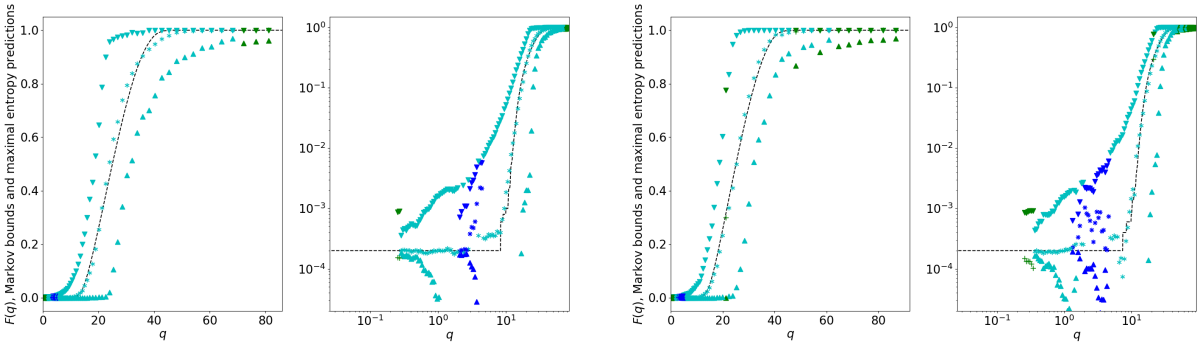


Figure 8: Two denser Erdős-Rényi graphs.

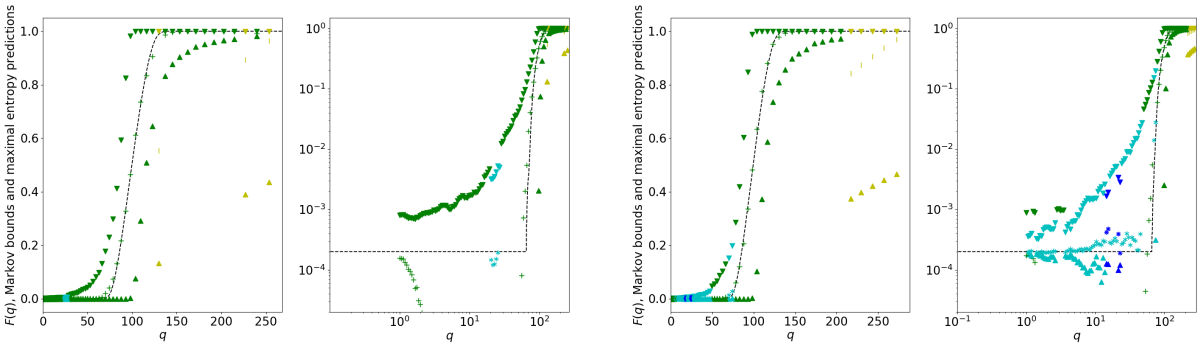


Figure 9: Two preferential attachment networks.

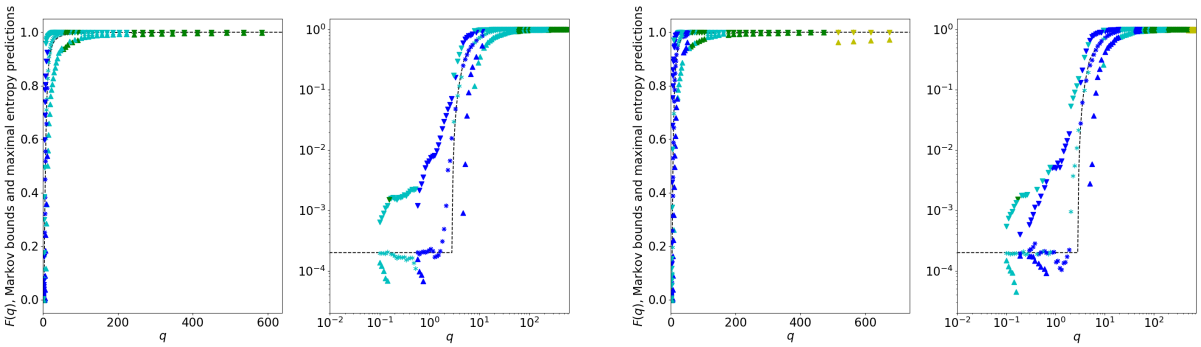


Figure 10: Two stochastic bloc models.

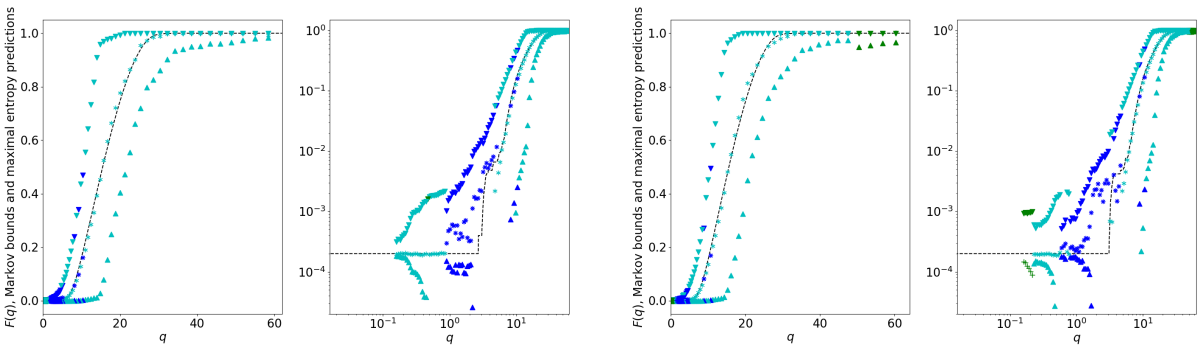


Figure 11: Two sensor graphs.

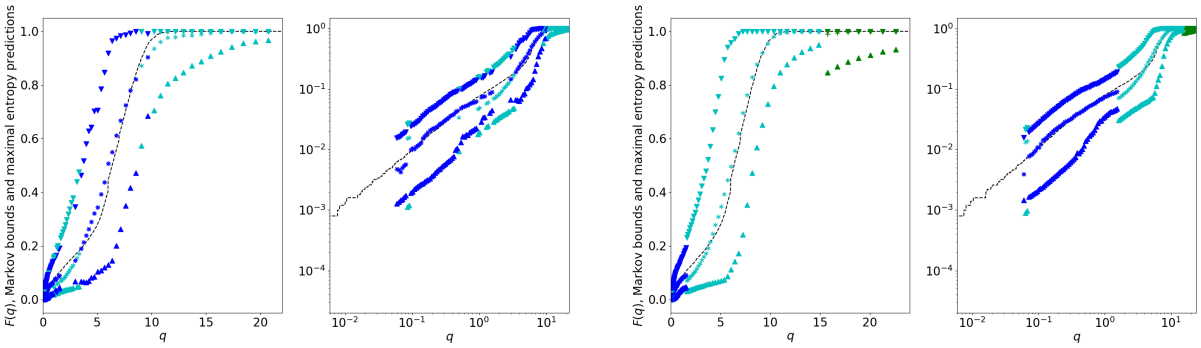


Figure 12: Comet and torus graphs.

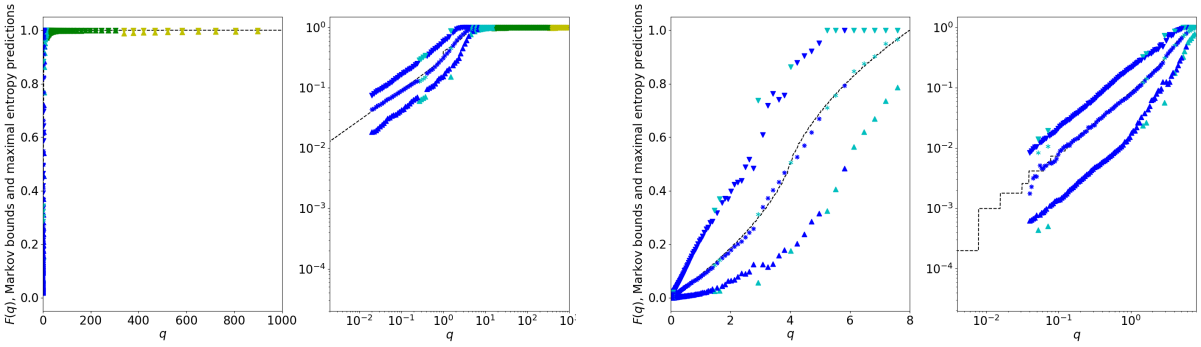
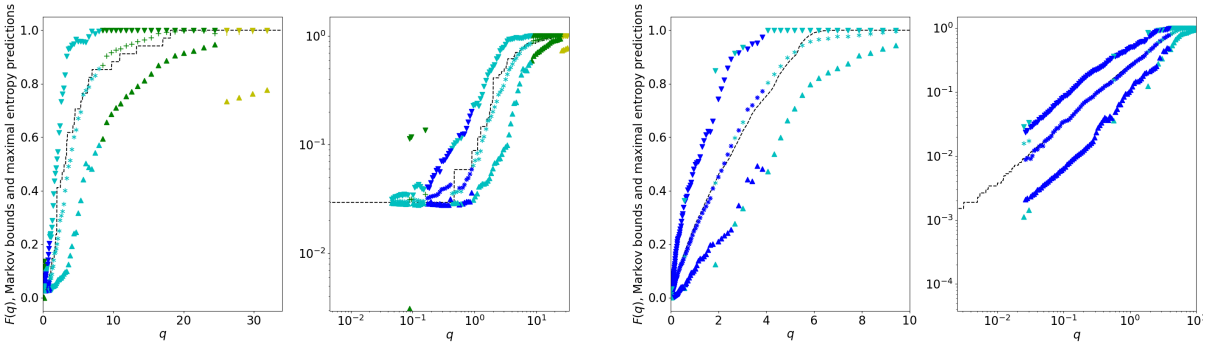


Figure 13 shows our results for the so-called karate club network with 34 vertices and 78 edges on the left, and for the for the Minnesota road network with 2642 nodes and 3304 links on the right.

Figure 13: Karate club and Minnesota road networks.

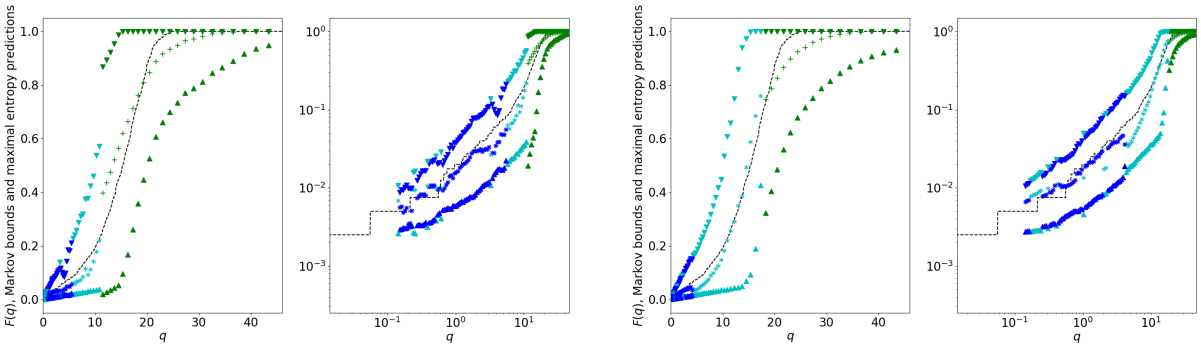


6.3.2 What went wrong

Figures 6–13 already show that reaching a larger number of valid moment estimates will not always improve the maximal entropy estimator. It strengthens Markov's bound, but these give a large window that goes quite slowly to zero. Getting more than a few valid moment estimates is actually out of reach such a Monte Carlo method. When l increases, the width $m_{l+1}^{+} - m_{l+1}^{-}$ given by Lemma C.1 for the interval where m_{l+1} must lie in goes dramatically to zero, indeed. Hence, increasing the forest number when our microcanonical model gives a wrong prediction with a few moments does not help. Figure 14 illustrates this for the upper part of the spectrum of $-L$. It shows our results with 400 replicated coupled forest on the left and 1600 replicated coupled forest on the right for a "swiss roll" graph with 400 vertices and 2883 edges.

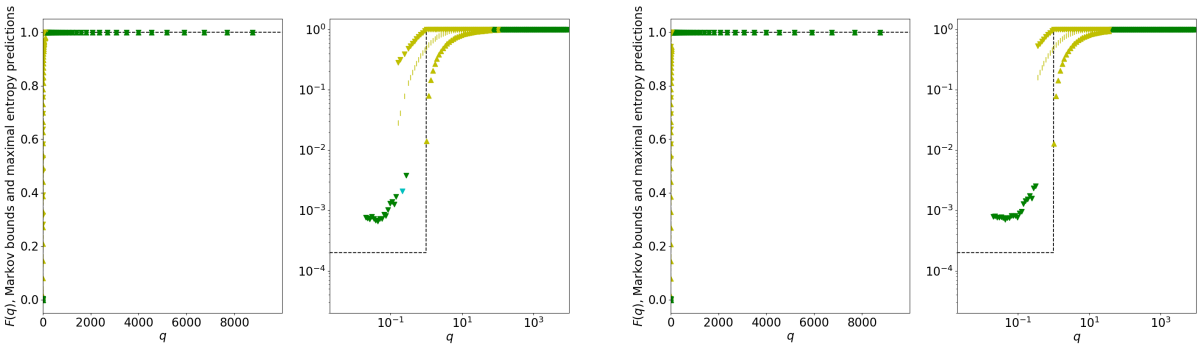
This problem can also occur when only one moment is available, which is the case for the star graph, rooted tree with $n - 1 = 4999$ branches of length 1. As shown by Figure 15, we cannot rely on

Figure 14: 400 and 1600 replicated coupled forests for a swiss roll graph.



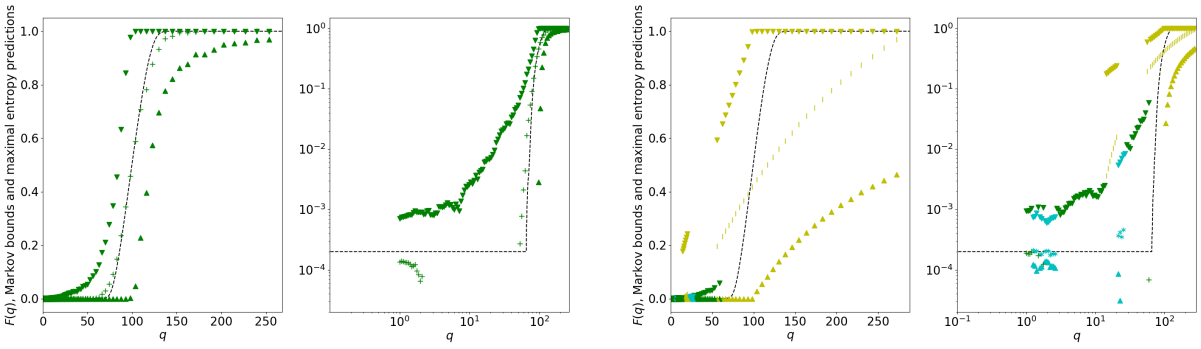
the microcanonical ensemble with only one moment and simply ignore higher (non-admissible) moment estimates.

Figure 15: 400 and 1600 replicated coupled forests for the star graph.



This is an extreme case and we can usually go beyond one valid moment estimate by sampling more forests. However, when reaching one more valid moment estimate, our maximum entropy estimator changes drastically in this case. This is a source of instability as illustrated by Figure 16, which shows our results with only 100 replicated forests for the previous two denser Erdős-Rényi graphs: in one case we got a second valid moment estimate for the relevant values of q, in one case we did not.

Figure 16: 100 replicated coupled forests for our two denser Erdős-Rényi graphs.



6.4 Room for improvement

6.4.1 Monte Carlo measures

We decided to make moment estimates and compute Markov bounds as well as maximal entropy estimators in only $1/\epsilon_0$ values of q because each measurement of $|\xi_q^k|$, $k \leq l$, for a given value of q has a computational cost in $O(n)$. But we could instead measure the difference $|\xi_{q_{i+1}}^k| - |\xi_{q_i}^k|$ between two

successive values of q along the coupled forest algorithm. Since each forest update is only local, we might be able to compute this difference at a lower numerical cost. We could then perform spectral estimation on a finer scale.

We could also rely on other observables along coupled forest trajectories, still related with the Laplacian spectrum. For example, by adding self loops into the arrow stacks to get space-homogeneous unfreezing rates $1/(1+\alpha t)$, $t \geq 0$, instead of $1/(1+w(x)t)$, $x \in \mathcal{X}$, the total number of arrows one has to read to update the current forest at an unfreezing time t is a mixture of geometric random variables with n mixture weights and success probabilities, each of them depending on a single eigenvalue λ_j of $-L$. Working with a mixture rather than a sum of random variables —like $|\rho(\Phi_q)|$, which is a sum of $0-1$ Bernoulli random variables with mean $q/(q+\lambda_j)$, $j < n$— can give better insights into the spectrum. Those geometric laws could also be related with better scale separations than these Bernoulli laws.

We finally note that in our experiments we did not take advantage of the fact that each permutation of the l forests needed to build ξ_q^k, $k \leq l$, for a given value of q lead to a different set with the same mean size. This could be used to implement some variance reduction in our Monte Carlo measurements.

6.4.2 Using more information for Stieltjes transform inversion

Our last observation in Section 6.3.2 strongly suggests that we should not simply get rid of higher moment estimates $\hat{m}_{k+1}(q)$, $\hat{m}_{k+2}(q)$, ... when we only have k valid moment estimates at a given q in $[q_0, 2\alpha]$. In the comparison with state of the art spectrum estimation methods that we presented in [BCG+25], we already included some projection on the set of admissible moment sequences. It is however clear that in order to estimate $F(q)$ there is still much more information to be used: that contained in $\hat{m}_k(q')$, $k \leq l$, for $q' \neq q$. How to use it is outside the scope of this paper. Here we simply wanted to point out that our low-cost pointwise inversion, which relates the law of Φ_q for a given q with the number of eigenvalues below q, already provides a sound estimation of this number. Together with the coupled forest process that covers all values of $q \geq \epsilon_0 \bar{\lambda}$ at a sampling cost in $O(\epsilon_0^{-1} n \ln n)$, this seems to us quite promising for spectrum estimation in large networks. By using the double cover trick that we recall in Appendix A, we can extend this perspective to spectrum estimation of all real symmetric matrices.

A Working with any symmetrical real matrix

A symmetric real matrix $M = (M(x,y))_{x,y\in\mathcal{X}}$ has to satisfy two constraints for being interpreted as the Laplacian L of a graph with vertex set $\mathcal{X}$. First, its off-diagonal entries must be non-negative,

$$M(x,y) \geq 0, \qquad x \neq y, \tag{22}$$

second, each diagonal entry on a line x in $\mathcal{X}$ must be the opposite of the sum all off-diagonal entries on this line,

$$M(x,x) = -\sum_{y\neq x} M(x,y), \qquad x \in \mathcal{X}.$$

As described in Section A.1, our spectrum estimation method immediately extend to sub-Laplacians $L = M$, which satisfy (22) but only

$$M(x,x) \leq -\sum_{y\neq x} M(x,y), \qquad x \in \mathcal{X}.$$

In Section A.2 we recall the double cover trick, which appears in [Gre96]. It reduces to the sub-Laplacian case that of any diagonally dominant symmetric matrix, which satisfies

$$M(x,x) \leq -\sum_{y\neq x} |M(x,y)|, \qquad x \in \mathcal{X}. \tag{23}$$

Since by subtracting a large enough multiple of the identity one can always reduce the case of any real symmetric matrix to this diagonally dominant case with a shifted spectrum, our spectrum estimation method extends to all real symmetric matrices.

A.1 Sub-Markovian generators

In the case of a sub-Laplacian $L = M$ we only have to consider loop-erased random walks killed in each x in $\mathcal{X}$ at rate $q + \delta(x)$ instead of q, with

$$\delta(x) = w(x) - \sum_{y\neq x} w(x,y) \geq 0, \qquad x \in \mathcal{X},$$

where $w(x) = -M(x,x)$ and $w(x,y) = M(x,y)$ for all $x \neq y$. As far as stacks of marks and arrows are concerned, this amounts to sample at each level associated with a vertex x a uniform mark as previously and an arrow (x,y), with $x \neq y$, with the same probability $w(x,y)/w(x)$ or a "killing arrow" with the remaining probability $\delta(x)/w(x)$. To build the coupled forest process $(\Phi_q)_{q>0}$, we just have to replace each arrow by a stop whenever the associated mark is below $q/(q + w(x))$. We then have two kinds of roots in each Φ_q, those associated with such a low mark, which can be unfrozen, and those associated with a killing arrow, which cannot. For replicated forests $\Phi_{k,q}$, $k < l$, we denote by $\tilde{\rho}(\Phi_{k,q})$ the set of roots of the first kind in $\Phi_{k,q}$, we define (recall (10))

$$\tilde{\xi}_q^k = \left\{x \in \mathcal{X} : R^1(x) \in \tilde{\rho}(\Phi_{0,q}),\, \ldots,\, R^k(x) \in \tilde{\rho}(\Phi_{k-1,q}),\, R^k(x) = x\right\}$$

we get

$$\mathbb{E}\left[\,|\tilde{\xi}_q^k|\right] = \sum_{j<n} \left(\frac{q}{q+\lambda_j}\right)^k$$

where the λ_j, $j < n$, are the eigenvalues of $-L$, and the whole analysis can be repeated.

A.2 The double cover trick

For any diagonally dominant real symmetric matrix $M = (M(x,y))_{x,y\in\mathcal{X}}$ satisfying (23), we call $\mathcal{X}_1 = \mathcal{X}$ a vertex set, and we associate with each vertex x in $\mathcal{X}$ another vertex $-x$ in a one-to-one correspondence with another and disjoint vertex set $-\mathcal{X}$. We define $\mathcal{X}_2 = \mathcal{X} \cup -\mathcal{X}$ in order to build two sub-Laplacians $L_1 = (L_1(x,y))_{x,y\in\mathcal{X}_1}$ and $L_2 = (L_2(x,y))_{x,y\in\mathcal{X}_2}$ such that

$$2\sigma_2 = \sigma_1 + \sigma \tag{24}$$

with σ_2, σ_1 and σ the spectral measures[§] associated with L_2, L_1 and M respectively. With F_2, F_1 and F the associated cumulative distribution functions, it will then hold

$$F = 2F_2 - F_1,$$

so that an estimates of F_1 and F_2 provides an estimate on F.

To this end we set for all *distinct* x and y in $\mathcal{X}$

$$L_2(x,y) = L_2(-x,-y) = [M(x,y)]_+ \geq 0,$$

$$L_2(x,-y) = L_2(-x,y) = [M(x,y)]_- \geq 0, \qquad L_2(x,-x) = L_2(-x,x) = 0$$

and

$$L_2(x,x) = L_2(-x,-x) = M(x,x) \leq -\sum_{z\neq x} [M(x,z)]_+ + [M(x,z)]_-$$

so that L_2 is a sub-Laplacian, indeed. We then defined L_1 as the sub-Markovian generator that describes the projection on $\mathcal{X} = \mathcal{X}_1$ of the killed random walk described by L_2. For distinct x and y in $\mathcal{X}$, it holds

$$L_1(x,y) = L_2(x,y) + L_2(x,-y) = L_2(-x,y) + L_2(-x,-y) = [M(x,z)]_+ + [M(x,z)]_- = |M(x,y)|$$

and

$$-L_1(x,x) = -L_2(x,x) = -L_2(-x,-x) = -M(x,x) \geq \sum_{z\neq x} L_1(x,z).$$

To prove (24) it suffices then to observe that the orthogonal vector spaces of even and odd functions f in $\mathbb{R}^{\mathcal{X}_2}$ —such that, for all $x \in \mathcal{X}$, $f(-x)$ is equal to $f(x)$ and $-f(x)$, respectively— are stable by L_2, which acts on them as L_1 and M, respectively. L_2 has indeed a block matrix representation

$$L_2 = \begin{pmatrix} M_+ & M_- \\ M_- & M_+ \end{pmatrix}$$

such that for all row vector U it holds

$$L_2 \begin{pmatrix} U \\ U \end{pmatrix} = \begin{pmatrix} M_+U + M_-U \\ M_-U + M_+U \end{pmatrix} = \begin{pmatrix} L_1U \\ L_1U \end{pmatrix}, \qquad L_2 \begin{pmatrix} U \\ -U \end{pmatrix} = \begin{pmatrix} M_+U - M_-U \\ M_-U - M_+U \end{pmatrix} = \begin{pmatrix} MU \\ -MU \end{pmatrix}.$$

[§]The factor 2 accounts for the normalisations of spectral measures.

B Pseudocodes

Algorithm 1 (Next, $\mathcal{Q}$) = `RandomForest`$(\mathcal{G}, q)$

Inputs:
  $\mathcal{G} = (\mathcal{X}, w)$
  $q > 0$
Initialize:
  # Initially, the forest is empty
  $\forall i \in \mathcal{X}$, InForest$[i] \leftarrow$ `false`
  $\forall i \in \mathcal{X}$, Next[i] $\leftarrow$ nil
  $\forall i \in \mathcal{X}$, $w[i] \leftarrow \sum_j w(i,j)$ *# Weighted degrees*
  $\mathcal{Q} \leftarrow \emptyset$ *# Priority queue*
for $i \in \mathcal{X}$ **do**
  $u \leftarrow i$
  # Start a random walk to create a forest branch
  while not InForest$[u]$ **do** *# Stop if u is in the forest*
    $\mathcal{U} \leftarrow Unif([0,1])$
    if $\mathcal{U} \leq \frac{q}{q+w[u]}$ **then** *#If true, u becomes a root*
      InForest$[u] \leftarrow$ `true` *# Add u to the forest*
      Next$[u] \leftarrow 0$ *# Set next of u to null*
      Add $\left(q' = \frac{\mathcal{U}w[u]}{1-\mathcal{U}}, u\right)$ to the priority queue $\mathcal{Q}$
    else *# If false, continue the random walk*
      Next[u] $\leftarrow$ `RandomSuccessor`$(u, \mathcal{G})$
      $u \leftarrow$ Next$[u]$
    end if
  end while
  $u \leftarrow i$ *# Go back to the initial node*
  # Add the newly created branch to the forest
  while not InForest$[u]$ **do**
    InForest$[u] \leftarrow$ `true`
    $u \leftarrow$ Next$[u]$
  end while
end for
`return` Next, $\mathcal{Q}$

Algorithm 2 $(Q, \{\text{Next}_q\}_{q\in Q}) = \texttt{CoupledRandomForest}(\mathcal{G}, q_{\min}, q_{\max})$

Inputs:
  $\mathcal{G} = (\mathcal{X}, w)$
  $q_{\max} \geq q_{\min} > 0$
Initialize:
  # start by sampling a random forest at q_{max}
  $Q = [q_{\max}]$ *# Initialize the list Q with q_{max}*
  $(\text{Next}, \mathcal{Q}) = \texttt{RandomForest}(\mathcal{G}, q_{\max})$
  $\text{Next}_{q_{\max}} \leftarrow \text{Next}$
  $(q, u) \leftarrow \texttt{find\_largest}(\mathcal{Q})$ *# finds the couple (q, u) in $\mathcal{Q}$ with maximal q*
  Remove (q, u) from $\mathcal{Q}$
  $\forall i \in \mathcal{X}, \quad \text{InForest}[i] \leftarrow \texttt{true}$
while $q > q_{\min}$ **do**
  Find $\mathcal{T}$, the set of nodes of the tree rooted in u that is re-activated
  $\forall i \in \mathcal{T}, \quad \text{InForest}[i] \leftarrow \texttt{false}$ *# Next[i] can be resampled*
  $\forall i \in \mathcal{T}, \quad \text{next\_is\_already\_sampled}[i] \leftarrow \texttt{true}$ *# Next[i] can also be kept and read again instead*
  Next[u] $\leftarrow \texttt{RandomSuccessor}(u, \mathcal{G})$ *# draw next of u: this is the reactivation*
  for $i \in \mathcal{T}$ **do**
    $u \leftarrow i$
    # Start a random walk to create a forest branch
    while not InForest[u] **do** *# Stop if u is in the forest*
      if next_is_already_sampled[u] **then** *# Don't redraw: follow previously drawn arrow*
        next_is_already_sampled[u] $\leftarrow \texttt{false}$ *# Next time, its next will be redrawn*
        $u \leftarrow \text{Next}[u]$
      else
        $\mathcal{U} \leftarrow Unif([0, 1])$
        if $\mathcal{U} \leq \frac{q}{q+w[u]}$ **then** *#If true, u becomes a root*
          InForest[u] $\leftarrow \texttt{true}$ *# Add u to the forest*
          Next[u] $\leftarrow 0$ *# Set next of u to null*
          Add $\left(q' = \frac{\mathcal{U} w[u]}{1-\mathcal{U}}, u\right)$ to the priority queue $\mathcal{Q}$
        else *# If false, continue the random walk*
          Next[u] $\leftarrow \texttt{RandomSuccessor}(u, \mathcal{G})$
          $u \leftarrow \text{Next}[u]$
        end if
      end if
    end while
    $u \leftarrow i$ *# Go back to the initial node*
    # Add the newly created branch to the forest
    while not InForest[u] **do**
      InForest[u] $\leftarrow \texttt{true}$
      $u \leftarrow \text{Next}[u]$
    end while
  end for
  Append q to Q and $\text{Next}_q \leftarrow \text{Next}$
  $(q, u) \leftarrow \texttt{find\_largest}(\mathcal{Q})$ *# finds the couple (q, u) in $\mathcal{Q}$ with maximal q*
  Remove (q, u) from $\mathcal{Q}$
end while
$\texttt{return}\ Q, \{\text{Next}_q\}_{q\in Q}$

C Canonical representations and Markov's bounds for the truncated moment problem on a bounded interval

We collect here some basics of truncated moments problems, which provide our needed Markov's bounds and that we essentially learned from [KN77]. Although they rely on well-know quadrature formulas, one important difficulty we faced in making this work was to recover these bounds, that we felt to have been at least partially forgotten in our community.

C.1 Admissible moment sequences and quadrature formulas

Let $[a,b] \subset \mathbb{R}$ and $m_0, m_1, \ldots, m_l$ be $n = l+1$ real numbers. We say that $m_0, \ldots, m_l$ is an **admissible** moment sequence if there exists a non-negative measure μ on $[a,b]$ such that

$$m_k = \int_a^b x^k \mu(dx), \qquad 0 \le k \le l.$$

In this case we say that μ is a *representation* of $m_0, \ldots, m_l$. We denote by $\mathcal{D}_{a,b}^l \subset \mathbb{R}^n$ the set of all such admissible moment sequences and we set $\mathcal{D}_{a,b} = \cup_{l \ge 0} \mathcal{D}_{a,b}^l$

For $m \in \mathcal{D}_{a,b}^l$ with representation μ, the linear form on the n-dimensional space $\mathbb{R}_l[X]$ of the polynomials with real coefficients and degree less than of equal to l

$$L_m : P = \sum_{k=0}^{l} p_k X^k \in \mathbb{R}_l[X] \mapsto \int_a^b P(x)\mu(dx) = \sum_{k=0}^{l} p_k m_k$$

is completely determined by m only. In particular, if an atomic measure $\nu = \sum_{i \in I} w_i \delta_{x_i}$ with atoms x_i, $i \in I$, in $[a,b]$ provides another representation of m, i.e,

$$\sum_{i \in I} w_i x_i^k = m_k, \qquad 0 \le k \le l, \tag{25}$$

then we have the quadrature formula

$$L_m(P) = \int_a^b P(x)\mu(dx) = \sum_{i \in I} w_i P(x_i), \qquad P \in \mathbb{R}_l[X]. \tag{26}$$

Equations (25) and (26) are actually equivalent.

C.2 The first orthogonal polynomials

Let μ be a measure on $[a,b]$ with $n = l+1$ first moments $m_0, m_1, \ldots, m_l$. This moment sequence m does in general *not* determine completely the bilinear form on $\mathbb{R}_l[X]$

$$\langle P, Q \rangle_\mu = \int_a^b P(x)Q(x)\mu(dx), \qquad P, Q \in \mathbb{R}_l[X],$$

since such quantities usually depend on moments of larger order $l+1, \ldots, 2l$. As a consequence it does not determine the unitary and normalized orthogonal polynomials U_k and $V_k = U_k / \langle U_k, U_k \rangle_\mu^{1/2}$, $k \ge 0$, associated with a regular enough measure μ. The moments $m_0, \ldots, m_l$ do however determine the first of them, which can be computed inductively with the formulas

$$U_0 = 1, \qquad V_0 = \frac{1}{\langle U_0, U_0 \rangle_\mu^{1/2}} U_0;$$

$$\beta_0 = \frac{\langle XU_0, U_0 \rangle_\mu}{\langle U_0, U_0 \rangle_\mu}, \qquad U_1 = (X - \beta_0)U_0, \qquad V_1 = \frac{1}{\langle U_1, U_1 \rangle_\mu^{1/2}} U_1;$$

and, for $k \ge 1$,

$$\beta_k = \frac{\langle XU_k, U_k \rangle_\mu}{\langle U_k, U_k \rangle_\mu}, \qquad \gamma_k = \frac{\langle XU_k, U_{k-1} \rangle_\mu}{\langle U_{k-1}, U_{k-1} \rangle_\mu},$$

$$U_{k+1} = (X - \beta_k)U_k - \gamma_k U_{k-1}, \qquad V_{k+1} = \frac{1}{\langle U_{k+1}, U_{k+1}\rangle_\mu^{1/2}} U_{k+1};$$

provided that none of these denominators vanishes. If some of them vanishes we say that m is a **singular** moment sequence, otherwise we say that m is **regular** —so that for any admissible moment sequence $m_0, \ldots, m_l$ there is $0 \leq k \leq l$ such that $m_0, \ldots, m_j$ is regular for all $j < k$ and singular for all $j \geq k$.

These formulas show that all moments up to order $2k+1 = 2(k+1) - 1$ have to be known to compute U_{k+1}, and all moments up to order $2k+2 = 2(k+1)$ have to be known to compute V_{k+1}. As a consequence :

- if $l = 2r$ is even, then m determines the first $r+1$ unitary and normalized orthogonal polynomials $U_0, \ldots, U_r$ and $V_0, \ldots, V_r$;
- if $l = 2r+1$ is odd, then m determines the first $r+2$ unitary orthogonal polynomials $U_0, \ldots, U_{r+1}$, but only the first $r+1$ normalized orthogonal polynomials $V_0, \ldots, V_r$.

We recall that orthogonal polynomials of degree $k \geq 0$ have k distinct roots in $]a, b[$.

C.3 Principal representations and quadratures

When looking for a quadrature formula on $[a, b]$ or an atomic representation ν of $n = l + 1$ moments of a measure μ, we will distinguish four different cases depending whether neither a nor b are atoms of ν, both a and b are atoms of ν, only a is an atom of ν or only b is an atom of ν. Solving Equation (25) with k atoms outside $\{a, b\}$ amounts then to solve a problem with n equations and $2k$, $2k+2$, $2k+1$ or $2k+1$ degrees of freedom, respectively. For an atomic distribution ν on $[a, b]$ with j atoms in $\{a, b\}$ and k atoms outside $\{a, b\}$ we then define the **index** of ν as

$$\mathrm{ind}(\nu) = j + 2k.$$

We will treat separately the "odd case" $l = 2r + 1$ and the "even case" $l = 2r$. For a regular moment sequence, in both cases we will get two uniquely defined atomic representations of index n and known as the **upper** and **lower principal representations** depending whether b is an atom or not. In all cases this is nothing more than rebuilding some well-known quadrature formulas.

C.3.1 Gauss and Lobatto quadratures for the odd case

Lower principal representation. For $l = 2r + 1$ and a regular sequence m of $n = l + 1 = 2(r + 1)$ moments $m_0, \ldots, m_l$, let us first look for an atomic representation of m with index n and without mass in $\{a, b\}$, i.e., for $x_0, \ldots, x_r$ in $]a, b[$ and $w_0, \ldots, w_r > 0$, such that

$$\sum_{i=0}^{r} w_i x_i^k = m_k, \qquad 0 \leq k \leq l. \tag{27}$$

Let $Q = \prod_{i=0}^{r}(X - x_i)$. Since Q is of degree $r + 1$, Equation (26) gives then

$$\langle Q, X^j\rangle_\mu = 0, \qquad 0 \leq j \leq l - (r + 1) = r,$$

which identifies Q as the unitary orthogonal polynomial U_{r+1}, and the x_i, $0 \leq i \leq r$, as its roots. We can then solve the Vandermonde system given by the first $r + 1$ equations in (27) to identify the w_i, $0 \leq i \leq r$.

To check that these x_i and w_i, $0 \leq i \leq r$, form a solution indeed, we first note that the quadrature formula (26) holds by construction for

$$P = 1, X, \ldots, X^r, Q, XQ, \ldots, X^r Q,$$

that is on a whole basis of $\mathbb{R}_l[X]$. It only remains to check that our weights w_i, $0 \leq i \leq r$, are positive. To this end we consider for any $0 \leq j \leq r$, the Legendre polynomial L_j such that $L_j(x_j) = 1$, and $L_j(x_i) = L_j'(x_i) = 0$ for $i \neq j$. This is a positive polynomial of degree $2r = l - 1$ and our quadrature formula gives

$$w_j = \int_a^b L_j(x)\mu(dx) \geq 0$$

for any representation μ of m. And had some w_j been equal to zero, we would have get an atomic representation of m with index less than n, so that m would have been singular.

A regular moment sequence of $n = 2(r+1)$ moments admits then a unique atomic representation with $r+1$ atoms in $]a,b[$, which gives rise to the standard Gaussian quadrature. We call it its **lower principal representation**.

Upper principal representation. For $l = 2r+1$ and a regular sequence m of $n = l+1 = 2r+2$ moments $m_0, \ldots, m_l$, we now look for an atomic representation of m with index n and mass both in a and b, i.e., for r atoms x_i, $i < r$, in $]a,b[$ as well as w_a, w_b and r weights w_i, $i < r$, in $\mathbb{R}_+ \setminus \{0\}$ such that

$$w_a a^k + w_b b^k + \sum_{i<r} w_i x_i^k = m_k, \qquad 0 \le k \le l. \tag{28}$$

For μ a representation of m, let $\mu^{a,b}$ be the measure on $[a,b]$ defined by

$$\mu^{a,b}(dx) = (b-x)(x-a)\mu(dx), \qquad x \in [a,b], \tag{29}$$

and let $m_0^{a,b}, \ldots, m_{l^{a,b}}^{a,b}$, with $l^{a,b} = l-2 = 2r-1 = 2(r-1)+1$, be its first $1+l^{a,b}$ moments, which depend on m only. The polynomial $Q = \prod_{i<r}(X-x_i)$ has degree r and Equation (26) gives

$$\langle Q, X^j\rangle_{\mu^{a,b}} = \langle (b-X)(X-a)Q, X^j\rangle_\mu = 0, \qquad 0 \le j \le l-(2+r) = r-1,$$

which identifies Q with the unitary polynomial $U_r^{a,b}$ associated with $m_0^{a,b}, \ldots, m_{l^{a,b}}^{a,b}$. As previously this identifies the x_i, $i < r$, as the roots of $U_r^{a,b}$ and the first $r+2$ equations in (28) identify w_a, w_b and w_i, $i < r$. The quadrature formula (26) holds for

$$P = 1,\ X,\ \ldots,\ X^{r+1},\ (b-X)(X-a)Q,\ X(b-X)(X-a)Q,\ \ldots,\ X^{r-1}(b-X)(X-a)Q,$$

again a whole basis of $\mathbb{R}_l[X]$. And we can conclude in the same way by considering the Legendre polynomials L_j, $j < r$, such that $L_j(x_j) = 1$ and $L_j(a) = L_j(b) = L_j(x_i) = L_j'(x_i) = 0$ for $i \neq j$ as well as L_a and L_b such that $L_a(a) = L_b(b) = 1$ and $L_a(b) = L_b(a) = L_a(x_i) = L_b(x_i) = L_a'(x_i) = L_b'(x_i) = 0$, $i < r$, which are non-negative on $[a,b]$ and have degree $2(r-1)+2 = 2r = l-1$, $2r+1 = l$ and $2r+1 = l$, respectively.

A regular moment sequence of $n = 2(r+1)$ moments admits then a unique atomic representation with r atoms in $]a,b[$, one in a and one in b, which gives rise to the Gauss-Lobatto quadrature. We call it its **upper principal representation**.

C.3.2 Gauss-Radau quadratures for the even case

For $l = 2r$ and a regular sequence m of $n = l+1 = 2r+1$ moments $m_0, \ldots, m_l$, we can look for an atomic representation of m with index n and some mass either in a or b. As previously, this leads to the Gauss-Radau quadratures associated with the **lower** or **upper principal representations**, respectively.

In the first case we look for x_i, $i < r$, in $]a,b[$ as well as w_a and w_i, $i < r$, in $\mathbb{R}_+ \setminus \{0\}$ such that

$$w_a a^k + \sum_{i<r} w_i x_i^k = m_k, \qquad 0 \le k \le l. \tag{30}$$

The x_i, $i < r$, are the roots of the unitary orthogonal polynomial U_r^a with degree r and associated with μ^a defined by

$$\mu^a(dx) = (x-a)\mu(dx), \qquad x \in [a,b]. \tag{31}$$

The last known moment or μ^a is of order $l^a = l-1 = 2r-1 = 2(r-1)+1$. Then, w_a as well as w_i, $i < r$, are identified by inverting the Vandermonde system given by the first $r+1$ equations in (30).

In the second case we solve in the same way

$$w_b b^k + \sum_{i<r} w_i x_i^k = m_k, \qquad 0 \le k \le l.$$

The x_i, $i < r$, are the roots of the unitary orthogonal polynomial U_r^b with degree r and associated with μ^b defined by

$$\mu^b(dx) = (b-x)\mu(dx), \qquad x \in [a,b]. \tag{32}$$

C.4 The truncated moment problem

The truncated moment problem is that of identifying $\mathcal{D}_{a,b}$. After observing that $\mathcal{D}^0_{a,b} = \mathbb{R}_+$ and that $m_0 = 0$ constitutes the only singular sequence in $\mathcal{D}^0_{a,b}$, the problem can be solved with the following

Lemma C.1. *If $m = (m_0, \ldots, m_l)$ is a regular moment sequence with lower and upper principal representations ν^- and ν^+, then their respective next moments m^-_{l+1} and m^+_{l+1} satisfy*

$$m^-_{l+1} = \int_a^b x^{l+1}\nu^-(dx) < m^+_{l+1} = \int_a^b x^{l+1}\nu^+(dx)$$

and it holds

$$m^-_{l+1} \leq \int_a^b x^{l+1}\mu(dx) \leq m^+_{l+1}$$

for any representation μ of m.

In addition, if μ has a moment m_{l+1} of order $l+1$ that coincides with m^-_{l+1} or m^+_{l+1}, then μ coincides with ν^- or ν^+, respectively. In this case $m_0, \ldots, m_{l+1}$ is a singular moment sequence. Conversely, any sequence $m_0, \ldots, m_{l+1}$ is a regular moment sequence in $\mathcal{D}^{l+1}_{a,b}$ provided that $m^-_{l+1} < m_{l+1} < m^+_{l+1}$.

Proof: Consider in the odd case $l = 2r+1$, with $x_0, \ldots, x_r$ the atoms of ν^-, the positive polynomial

$$P = \prod_{i=0}^{r}(X - x_i)^2 = X^{2r+2} + Q.$$

Writing m_{l+1} for the moment of order $l+1$ of a representation μ of m, it holds

$$0 \leq \int_a^b P(x)\mu(dx) = m_{l+1} + \int_a^b Q(x)\mu(dx) = m_{l+1} + \int_a^b Q(x)\nu^-(dx),$$

since $Q \in \mathbb{R}_l[X]$. The last term in the previous equation is $-m^-_{l+1}$, since it also holds

$$0 = \int_a^b P(x)\nu^-(dx) = m^-_{l+1} + \int_a^b Q(x)\nu^-(dx).$$

This proves $m^-_{l+1} \leq m_{l+1}$. And in case of equality, it holds

$$\int_a^b P(x)\mu(dx) = 0,$$

which implies that the support of μ is contained in that of ν^-. Writing $\mu = \sum_{i=0}^r w_i\delta_{x_i}$ the weights w_i are then identified by the Vandermonde equations

$$\sum_{i=0}^{r} w_i x_i^k = m_k, \qquad 0 \leq k \leq r,$$

which implies that μ and ν^- coincide, indeed.

The upper bound $m_{l+1} \leq m^+_{l+1}$ is proved similarly by considering the positive polynomial on $[a, b]$

$$P = (b - X)(X - a)\prod_{i<r}(X - x_i)^2 = -X^{2r+2} + Q,$$

with a, b and x_i, $i < r$, the atoms of ν^+. And we can deal with the case of equality in the same way.

Since lower and upper principal representations are distinct representations of a regular moment sequence, this also proves that $m^-_{l+1} < m^+_{l+1}$. Finally, if $m_{l+1} = \alpha m^-_{l+1} + (1-\alpha)m^+_{l+1}$ for some $\alpha \in]0, 1[$, then $(m_0, \ldots, m_{l+1})$ admits $\alpha\nu^- + (1-\alpha)\nu^+$ as a representation: it forms a regular moment sequence.

We can deal with the even case $l = 2r$ in the same way by considering the polynomial

$$P = (X - a)\prod_{i<r}(X - x_i)^2$$

and

$$P = (b - X)\prod_{i<r}(X - x_i)^2,$$

with x_i, $i < r$, the atoms of the lower and upper principal representation, respectively. □

We note, as the consequence of this lemma, that a sequence of $n = l+1$ moments $m_0, \ldots, m_l$ forming a singular sequence has a unique representation ν, which is atomic and satisfies $\mathrm{ind}(\nu) < n$.

C.5 Canonical representations

Consider $n = l + 1$ moments $m_0, \ldots, m_l$ forming a regular sequence m. For each $\xi \in]a, b[$ there is a unique atomic measure ν^ξ with an atom at ξ and of index strictly less than $n+2$ —actually n or $n+1$— known as the **canonical representation** of m associated with ξ.

To see that such a representation exists, we can consider

$$t = \sup\big\{s \geq 0 : m - (s, s\xi, s\xi^2, \ldots, s\xi^l) \in \mathcal{D}^l_{a,b}\big\} \leq m_0.$$

The previous lemma implies that $t > 0$ and $m - (t, \ldots, t\xi^l)$ is singular, i.e., has an atomic representation ν with index strictly less that n. So that $\nu + t\delta_\xi$ is an atomic representation of m with index $n + 1$ at most and with an atom in ξ. We show how to build it in this section.

C.5.1 The odd case

In the odd case $l = 2r + 1$, i.e.,

$$n + 1 = l + 2 = 2r + 1 + 2,$$

we are looking either for $x_i \in [a, b] \setminus \{\xi\}$ and $w_i \geq 0$, $i < r$, as well as $w_a \geq 0$ and $t > 0$ such that

$$t\xi^k + w_a a^k + \sum_{i<r} w_i x_i^k = m_k, \qquad k < 2r + 2, \tag{33}$$

or for $x_i \in [a, b] \setminus \{\xi\}$ and $w_i \geq 0$, $i < r$, as well as $w_b \geq 0$ and $t > 0$ such that

$$t\xi^k + w_b b^k + \sum_{i<r} w_i x_i^k = m_k, \qquad k < 2r + 2. \tag{34}$$

In the first case, consider some representation μ of m and the measure μ^a defined by (31). The first $l+1$ moments $m_0, \ldots, m_l$ of μ define the first moments $m^a_0, \ldots, m^a_{l^a}$ of μ^a, with $l^a = l - 1 = 2r$, hence the first $r + 1$ associated normalized orthogonal polynomials $V^a_0, \ldots, V^a_r$. With $P = \prod_{i<r}(X - x_i)$, Equation (26) gives then

$$\langle V^a_k, P\rangle_{\mu^a} = \langle (X - a)V^a_k, P\rangle_\mu = t(\xi - a)P(\xi)V^a_k(\xi), \qquad 0 \leq k \leq r.$$

Since $P \in \mathbb{R}_r[X]$, this shows that P and

$$Q^a = \sum_{i=0}^{r} V^a_i(\xi)V^a_i$$

are proportional, namely

$$P = t(\xi - a)P(\xi)Q^a,$$

and gives

$$t = \frac{1}{(\xi - a)Q^a(\xi)},$$

since $P(\xi)$ and $Q^a(\xi)$ are not zero. Provided that (33) does have a solution in the prescribed domain, this solution is unique, since the singular moment sequence $m - (t, \ldots, t\xi^l)$ admits one representation only. Our x_i, $i < r$, are in practice the roots of Q^a; w_a and w_i, $i < r$, can then be found by solving the Vandermonde system given by the first equations in (33).

In the second case we get in the same way

$$t = \frac{1}{(b - \xi)Q^b(\xi)},$$

with

$$Q^b = \sum_{i=0}^{r} V^b_k(\xi)V^b_k,$$

where $V^b_0, \ldots, V^b_r$ are the first $r+1$ normalized orthogonal polynomials associated with the first $l^b + 1 = 2r + 1$ moments $m^b_0, \ldots, m^b_{l^b}$ of μ^b defined by (32). Again, our x_i, $i < r$, are the roots of Q^b, and the

weights w_i, $i < r$, as well as w_b can be found by solving the Vandermonde system given by the first equations in (34).

This proves that ν^ξ must be either the unique solution of (33), if it exists, or that of (34), if it exists —we already know that at least one of them exists. This does not completely settle the uniqueness issue, except when $(\xi - a)Q^a(\xi) = (b - \xi)Q^b(\xi)$, i.e., when (33) and (34) implies a same value for t, so that $m - (t, \dots, t\xi^l)$ is a singular moment sequence with only one representation. We will see later that when $(\xi - a)Q^a(\xi) \neq (b - \xi)Q^b(\xi)$, only one of the two equations (33) and (34) can lead to a solution inside the desired domain.

C.5.2 The even case

For $l = 2r$ we have

$$n + 1 = l + 2 = 2r + 2$$

and we are looking either for $x_i \in [a, b] \setminus \{\xi\}$ and $w_i \geq 0$, $i < r$, as well as $t > 0$ such that

$$t\xi^k + \sum_{i<r} w_i x_i^k = m_k, \qquad k < 2r + 1, \tag{35}$$

or for $x_i \in [a, b] \setminus \{\xi\}$ and $w_i \geq 0$, $i < r - 1$, as well as $w_a \geq 0$, $w_b \geq 0$ and $t > 0$, such that

$$t\xi^k + w_a a^k + w_b b^k + \sum_{i<r-1} w_i x_i^k = m_k, \qquad k < 2r + 1 = 2(r - 1) + 2 + 1, \tag{36}$$

Similarly to the odd case, given a representation μ of m, its first $l + 1$ moments $m_0, \dots, m_l$ define the first $r + 1$ normalized orthogonal polynomials $V_0, \dots, V_r$ associated with μ, as well as the first $l^{a,b} + 1 = l - 2 + 1 = 2(r - 1) + 1$ moments $m_0^{a,b}, \dots, m_{l^{a,b}}^{a,b}$ and the first r normalized orthogonal polynomials $V_0^{a,b}, \dots, V_{r-1}^{a,b}$ associated with $\mu^{a,b}$ defined by (29). We can then define

$$Q = \sum_{k=0}^{r} V_k(\xi) V_k$$

and

$$Q^{a,b} = \sum_{k<r} V_k^{a,b}(\xi) V_k^{a,b}$$

to get that (35) leads to

$$t = \frac{1}{Q(\xi)}$$

and identifies the x_i, $i < r$, as the roots of Q, then the weights w_i, $i < r$, as the solution of the Vandermonde system given by the first equations in (35), while (36) leads to

$$t = \frac{1}{(b - \xi)(\xi - a)Q^{a,b}(\xi)}$$

and identifies x_i, $i < r - 1$, as the roots of $Q^{a,b}$, then the weights w_i, $i < r$, as well as w_a and w_b as the solution of the Vandermonde system given by the first equations in (36).

Again, we know that one at least of (35) and (36) gives a solution in the prescribed domain, which shows that the canonical representation ν^ξ is uniquely defined in the special case where

$$Q(\xi) = (b - \xi)(\xi - a)Q^{a,b}(\xi).$$

We will see in the next section that equations (35) and (36) cannot both lead to a solution in the desired domain when

$$Q(\xi) \neq (b - \xi)(\xi - a)Q^{a,b}(\xi),$$

so that the canonical representation ν^ξ is always uniquely defined, indeed.

C.6 Markov's bounds

Markov's bounds are tight lower an upper bounds on $\mu(]\xi,b])$ and $\mu([\xi,b])$ for a given $\xi \in]a,b[$ and any measure μ the first $n = l+1$ moments of which form a given regular moment sequence. These bounds are derived from an associated canonical representation

$$\nu^\xi = t\delta_\xi + \nu'$$

with $t > 0$ and $\mathrm{ind}(\nu') \leq n+1-2 = l$. They read

$$\nu^\xi(]\xi,b]) \leq \mu(]\xi,b]) \leq \mu([\xi,b]) \leq \nu^\xi([\xi,b]) \tag{37}$$

and

$$\mu(\{\xi\}) = \mu([\xi,b]) - \mu(]\xi,b]) \leq t. \tag{38}$$

When $\mathrm{ind}(\nu') = 2k+1$ is odd and c is the only one atom of ν' in $\{a,b\}$, call x_i, $i<k$, the interior atoms of ν' and consider the Legendre polynomials L_- and L_+ defined by

$$L_-(\xi) = 0, \qquad L_+(\xi) = 1,$$

$$L_-(c) = L_+(c) = \mathbb{1}_{\{c>\xi\}}, \qquad L_-(x_i) = L_+(x_i) = \mathbb{1}_{\{x_i>\xi\}}, \qquad i<k,$$

and

$$L'_-(x_i) = L'_+(x_i) = 0, \qquad i<k.$$

L_- and L_+ satisfy

$$L_- \leq \mathbb{1}_{]\xi,b]} \leq \mathbb{1}_{[\xi,b]} \leq L_+$$

and have degree $2k+1 \leq l$. It follows that

$$\nu'(]\xi,b]) = \nu^\xi(]\xi,b]) = \int_a^b L_-(x)\nu^\xi(dx) = \int_a^b L_-(x)\mu(dx) \leq \int_a^b \mathbb{1}_{]\xi,b]}(x)\mu(dx) = \mu(]\xi,b])$$

and

$$\mu([\xi,b]) = \int_a^b \mathbb{1}_{[\xi,b]}(x)\mu(dx) \leq \int_a^b L_+(x)\mu(dx) = \int_a^b L_+(x)\nu^\xi(dx) = \nu^\xi([\xi,b]) = t + \nu'(]\xi,b]),$$

which proves (37) and (38).

When $\mathrm{ind}(\nu') = 2k$ and $x_i \in]a,b[$, $i<k$, are all the atoms of ν', consider the Legendre polynomials L_- and L_+ defined by

$$L_-(\xi) = 0, \qquad L_+(\xi) = 1,$$

$$L_-(x_i) = L_+(x_i) = \mathbb{1}_{\{x_i>\xi\}}, \qquad i<k,$$

and

$$L'_-(x_i) = L'_+(x_i) = 0, \qquad i<k.$$

Since these have degree $2k \leq l$, (37) and (38) follow in the same way.

Finally, when $\mathrm{ind}(\nu') = 2k$ and $x_i \in]a,b[$, $i<k-1$, as well as a and b are the atoms of ν', the same result is obtained by considering the Legendre polynomials of degree $2k \leq l$ defined by

$$L_-(\xi) = 0, \qquad L_+(\xi) = 1,$$

$$L_-(a) = L_+(a) = 0, \qquad L_-(b) = L_+(b) = 1, \qquad L_-(x_i) = L_+(x_i) = \mathbb{1}_{\{x_i>\xi\}}, \qquad i<k-1,$$

and

$$L'_-(x_i) = L'_+(x_i) = 0, \qquad i<k-1.$$

Now, Equation (38) shows that there cannot be two canonical representations associated with a same $\xi \in]a,b[$ and two different weights in ξ: we would get a contradiction by taking that with the larger weight in place of μ and that with the smaller weight t in place of ν^ξ. This concludes the proof of uniqueness for canonical representations.

Krein and Nudel'man even show in [KN77] that the weight in any $\xi \in]a, b[$ of the canonical representation ν^ξ is always, with the notation of the previous section,

$$t = \frac{1}{\max\big((\xi - a)Q^a(\xi),\, (b - \xi)Q^b(\xi)\big)}$$

in the odd case or

$$t = \frac{1}{\max\big(Q(\xi),\, (b - \xi)(\xi - a)Q^{a,b}(\xi)\big)}$$

in the even case. Here, we simply conclude:

Theorem. (Markov bounds) *For any $\xi \in]a, b[$, each regular moment sequence $m = (m_0, \ldots, m_l)$ admits a unique canonical representation ν^ξ with an atom in ξ. Then, for any representation μ of m, it holds*

$$\nu^\xi(]\xi, b]) \leq \mu(]\xi, b]) \leq \mu([\xi, b]) \leq \nu^\xi([\xi, b]).$$

Ackowledgments. This work was partially funded by the Région Auvergne-Rhône-Alpes project TrustGNN, the ANR project GRANOLA (ANR-21-CE48-0009) and the European Union (Marie Curie project Rand4TrustPool, 101148828)

References

[ACGM20] Luca Avena, Fabienne Castell, Alexandre Gaudillière, and Clothilde Mélot. Intertwining wavelets or multiresolution analysis on graphs through random forests. *Applied and Computational Harmonic Analysis*, 48(3):949–992, 2020.

[AG18] Luca Avena and Alexandre Gaudillière. Two Applications of Random Spanning Forests. *Journal of Theoretical Probability*, 31(4):1975–2004, 2018.

[BCG+25] Simon Barthelmé, Fabienne Castell, Alexandre Gaudillière, Clothilde Melot, Matteo Quattropani, and Nicolas Tremblay. Estimating a graph's spectrum via random kirchhoff forests, 2025. arXiv:2503.24236.

[BP93] Robert Burton and Robin Pemantle. Local Characteristics, Entropy and Limit Theorems for Spanning Trees and Domino Tilings Via Transfer-Impedances. *The Annals of Probability*, 21(3):1329–1371, 1993.

[DEM24] Matteo D'Achille, Nathanaël Enriquez, and Paul Melotti. Local limit of massive spanning forests on the complete graph, 2024. arXiv:2403.11740.

[DF91] Persi Diaconis and William Fulton. A growth model, a game, an algebra, Lagrange inversion, and characteristic classesi. *Rendiconti del Seminario Matematico Università e Politecnico di Torino*, 49(1):95–119, 1991.

[Gre96] Keith D. Gremban. *Combinatorial Preconditioners for Sparse, Symmetric, Diagonally Dominant Linear Systems*. Phd thesis, Carnegie Mellon University, Pittsburgh, October 1996. Tech Report CMU-CS-96-12.

[JFA+23] Hugo Jaquard, Michaël Fanuel, Pierre-Olivier Amblard, Rémi Bardenet, Simon Barthelmé, and Nicolas Tremblay. Smoothing complex-valued signals on graphs with monte-carlo. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2023.

[Kir47] Gustav Kirchhoff. Ueber die auflösung der gleichungen, auf welche man bei der untersuchung der linearen vertheilung galvanischer ströme geführt wird. *Annalen der Physik*, 148(12):497–508, 1847.

[KN77] M. G. Krein and A. A. Nudel'man. *The Markov Moment Problem and Extremal Problems*, volume 50 of *Translations of Mathematical Monographs*. American Mathematical Society, 1977.

[Mar00] Philippe Marchal. Loop-Erased Random Walks, Spanning Trees and Hamiltonian Cycles. *Electronic Communications in Probability*, 5:39–50, 2000.

[MP84] Lawrence R. Mead and N. Papanicolaou. Maximum entropy in the problem of moments. *Journal of Mathematical Physics*, 25(8):2404–2417, 1984.

[PABT21] Yusuf Yiğit Pilavcı, Pierre-Olivier Amblard, Simon Barthelmé, and Nicolas Tremblay. Graph tikhonov regularization and interpolation via random spanning forests. *IEEE transactions on Signal and Information Processing over Networks*, 7:359–374, 2021.

[ST04] Daniel A Spielman and Shang-Hua Teng. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 81–90, 2004.

[Wal77] Alastair J Walker. An efficient method for generating discrete random variables with general distributions. *ACM Transactions on Mathematical Software (TOMS)*, 3(3):253–256, 1977.

[Wil96] David Bruce Wilson. Generating random spanning trees more quickly than the cover time. In *STOC '96*, 1996.

[WWAF06] Alexander Weiße, Gerhard Wellein, Andreas Alvermann, and Holger Fehske. The kernel polynomial method. *Reviews of modern physics*, 78(1):275–306, 2006.