

Two Applications of Random Spanning Forests

L. Avena¹  · A. Gaudillière²

Received: 4 February 2016 / Revised: 30 May 2017 / Published online: 18 July 2017
© The Author(s) 2017

Abstract We use random spanning forests to find, for any Markov process on a finite set of size n and any positive integer $m \leq n$, a probability law on the subsets of size m such that the mean hitting time of a random target that is drawn from this law does not depend on the starting point of the process. We use the same random forests to give probabilistic insights into the proof of an algebraic result due to Micchelli and Willoughby and used by Fill and by Miclo to study absorption times and convergence to equilibrium of reversible Markov chains. We also introduce a related coalescence and fragmentation process that leads to a number of open questions.

Keywords Finite networks · Spectral analysis · Spanning forests · Determinantal processes · Random sets · Hitting times · Coalescence and fragmentation · Local equilibria

Mathematics Subject Classification (2010) Primary: 05C81 · 60J20 · 15A15 · Secondary: 15A18 · 05C85

✉ L. Avena
l.avena@math.leidenuniv.nl

A. Gaudillière
alexandre.gaudilliere@math.cnrs.fr

¹ MI, University of Leiden, Leiden, The Netherlands

² CNRS, Centrale Marseille, I2M, Aix Marseille University, Marseille, France

1 Well-Distributed Points, Local Equilibria and Coupled Spanning Forests

1.1 Well-Distributed Points and Random Spanning Forests

Let $X = (X(t) : t \geq 0)$ be an irreducible continuous-time Markov process on a finite set \mathcal{X} with size $|\mathcal{X}| = n$. It is known, see, for example, Lemma 10.8 in [9], that if $R \in \mathcal{X}$ is chosen according to the equilibrium measure μ of the process, then the mean value of the hitting time $\mathbb{E}[E_x[T_R]]$ —where $E_x[\cdot]$ stands for the mean value according to the law of the process started at x and $\mathbb{E}[\cdot]$ stands for the mean value according to the law of R —does not depend on the starting point $x \in \mathcal{X}$. More generally, if a random subset $R \subset \mathcal{X}$ of any (possibly random) size has such a property, then we say that the law of R provides *well-distributed points*. One of our motivations for building such random sets was to find appropriate subsampling points for signal processing on arbitrary networks, in connection with intertwining equations and metastability studies (cf. [2]). In this paper, we build such a law on the subsets of any given size $m \leq n$. This is a trivial problem for $m = n$, and for $m = 1$, this property actually characterizes the law of R : in this case, the singleton R has to be chosen according to the equilibrium law.

To solve this problem in the general case, we use random rooted spanning forests, a standard variation—introduced, for example, in [13]—on the well-known uniform spanning tree theme. Let us first denote by $w(x, y)$ the jump rate of X from x to y in \mathcal{X} and by $\mathcal{G} = (\mathcal{X}, w)$ the weighted and oriented graph for which

$$\mathcal{E} = \{(x, y) \in \mathcal{X} \times \mathcal{X} : x \neq y \text{ and } w(x, y) > 0\}$$

is the edge set and $w(e) = w(x, y)$ is the weight of $e = (x, y)$ in \mathcal{E} . A rooted spanning forest ϕ is a subgraph of \mathcal{G} without cycle, with \mathcal{X} as set of vertices and such that, for each $x \in \mathcal{X}$, there is at most one $y \in \mathcal{X}$ such that (x, y) is an edge of ϕ . The root set $\rho(\phi)$ of the forest ϕ is the set of points $x \in \mathcal{X}$ for which there is no edge (x, y) in ϕ ; the connected component of ϕ are trees, each of them having edges that are oriented towards its own root. We call \mathcal{F} the set of all rooted spanning forests, we see each forest ϕ in \mathcal{F} as a subset of \mathcal{E} , and we associate with it the weight

$$w(\phi) = \prod_{e \in \phi} w(e).$$

In particular, $\emptyset \in \mathcal{F}$ is the spanning forest made of n degenerate trees reduced to simple roots and $w(\emptyset) = 1$. We can now define our random forests: for each $q > 0$, the random spanning forest Φ_q is a random variable in \mathcal{F} with law

$$\mathbb{P}(\Phi_q = \phi) = \frac{w(\phi)q^{|\rho(\phi)|}}{Z(q)}, \quad \phi \in \mathcal{F},$$

where the normalizing partition function is

$$Z(q) = \sum_{\phi \in \mathcal{F}} w(\phi)q^{|\rho(\phi)|}.$$

We can include the case $q = +\infty$ in our definition by setting $\Phi_\infty = \emptyset \in \mathcal{F}$ in a deterministic way.

It turns out that both the law of $\rho(\Phi_q)$ and the law of $\rho(\Phi_q)$ conditioned on the event $\{|\rho(\Phi_q)| = m\}$, for any $1 \leq m \leq n$, provide well-distributed points. And we can compute the common value of the mean hitting time in both cases in terms of the eigenvalues of the generator L given by

$$(Lf)(x) = \sum_{y \in \mathcal{X}} w(x, y)[f(y) - f(x)], \quad f : \mathcal{X} \rightarrow \mathbb{C}, \quad x \in \mathcal{X},$$

To this end, let us denote by $\lambda_0, \lambda_1, \dots, \lambda_{n-1}$ the eigenvalues of $-L$ and $(a_k : 0 \leq k \leq n)$ the coefficients of the characteristic polynomial of L , which computed in q is

$$\det(q - L) = \prod_{j < n} (q + \lambda_j) = \sum_{k \leq n} a_k q^k.$$

In this formula and all along the paper, we identify scalars with the appropriate multiples of the identity matrix. Recalling that X is irreducible and ordering the eigenvalue by non-decreasing real part, λ_0 is the only one zero eigenvalue, we have $a_0 = 0$ and we can set $a_{n+1} = 0$.

Theorem 1 *For all $x \in \mathcal{X}$ and all positive integer $m \leq n$ it holds*

$$\mathbb{E}[E_x[T_{\rho(\Phi_q)}]] = \frac{1}{q} \left(1 - \prod_{j > 0} \frac{\lambda_j}{q + \lambda_j} \right) \quad \text{and} \quad \mathbb{E}[E_x[T_{\rho(\Phi_q)}] \mid |\rho(\Phi_q)| = m] = \frac{a_{m+1}}{a_m}.$$

We prove this theorem in Sect. 3, in which we also compute, in both cases, as a consequence of it and as needed in [2], mean return times to $\rho(\Phi_q)$ from a uniformly chosen point in $\rho(\Phi_q)$. In doing so, we will see that the problem of finding a distribution that provides exactly m well-distributed points has infinitely many solutions as soon as $2 \leq m \leq n - 2$ and Theorem 1 simply provides one of them. The only cases when the convex set of solutions reduces to a singleton are the known case $m = 1$, the easy case $m = n - 1$ and the trivial one $m = n$.

1.2 Local Equilibria and Random Forests in the Reversible Case

For $\mathcal{B} \subset \mathcal{X}$ we identify the signed measures on $\mathcal{X} \setminus \mathcal{B}$ with the row vectors $v = (v(x) : x \in \mathcal{X} \setminus \mathcal{B})$. For $\mathcal{A} \subset \mathcal{X}$ and any matrix $M = (M(x, y) : x, y \in \mathcal{X})$, we write $[M]_{\mathcal{A}}$ for the submatrix

$$[M]_{\mathcal{A}} = (M(x, y) : x, y \in \mathcal{A}).$$

We identify L with its matrix representation with diagonal coefficients $-w(x)$, where

$$w(x) = \sum_{y \neq x} w(x, y), \quad x \in \mathcal{X}, \tag{1}$$

and we set

$$\alpha = \max_{x \in \mathcal{X}} w(x). \tag{2}$$

The sub-Markovian generator $[L]_{\mathcal{X} \setminus \mathcal{B}}$ is associated with the process killed in “the boundary” \mathcal{B} . We will assume in this section that L is reversible with respect to μ and we write

$$\lambda_{0,\mathcal{B}} < \lambda_{1,\mathcal{B}} \leq \dots \leq \lambda_{l-1,\mathcal{B}}$$

for the ordered eigenvalues of $-[L]_{\mathcal{X} \setminus \mathcal{B}}$, with $l = |\mathcal{X} \setminus \mathcal{B}|$. We can then inductively define from any probability measure $\nu = \nu_{l-1}$ on $\mathcal{X} \setminus \mathcal{B}$ the *a priori* signed measures ν_k , with $k < l$, by

$$\nu_k[L]_{\mathcal{X} \setminus \mathcal{B}} = \lambda_{k,\mathcal{B}}[\nu_k - \nu_{k-1}]. \tag{3}$$

To avoid ambiguity, we set $\nu_{k-1} = 0$ if $\lambda_{k,\mathcal{B}} = 0$. The following result is due to Micchelli and Willoughby (see Theorem 3.2 in [11]).

Micchelli and Willoughby’s Theorem *If L is reversible, then ν_k is a non-negative measure for all non-negative $k < l$ and any probability measure ν on $\mathcal{X} \setminus \mathcal{B}$.*

Since Eq. (3) can also be written (for $k > 0$ or $\mathcal{B} \neq \emptyset$, so that $\lambda_{k,\mathcal{B}} > 0$) as

$$\nu_{k-1} = \nu_k \frac{[L]_{\mathcal{X} \setminus \mathcal{B}} + \lambda_{k,\mathcal{B}}}{\lambda_{k,\mathcal{B}}},$$

we have

$$\sum_{x \in \mathcal{B}} \nu_{k-1}(x) \leq \sum_{x \in \mathcal{B}} \nu_k(x) \leq \dots \leq \sum_{x \in \mathcal{B}} \nu(x) = 1$$

and we can identify each ν_k with a probability measure on the quotient space \mathcal{X}/\mathcal{B} for the equivalence relation $\sim_{\mathcal{B}}$ such that $x \sim_{\mathcal{B}} y$ if and only if $x = y$ or $x, y \in \mathcal{B}$: we simply put the missing mass on \mathcal{B} . Equation (3) has then the following probabilistic interpretation. Starting from ν_k , the system decays into ν_{k-1} after an exponential time of parameter $\lambda_{k,\mathcal{B}}$, and, more precisely, starting from $\nu_k(\cdot|\mathcal{B})$, the system remains in this state for an exponential time of parameter $\lambda_{k,\mathcal{B}}$ before decaying into $\nu_{k-1}(\cdot|\mathcal{B})$ or reaching \mathcal{B} . This is rigorously established by Fill and Miclo (see [7] and [12]) to control convergence to equilibrium and absorption times of reversible processes, and this is the reason why the ν_k s can be described as *local equilibria*.

The previous probabilistic interpretation makes sense only once the non-negativity of the ν_k s is guaranteed by Micchelli and Willoughby’s theorem, which is crucial in Fill’s and Miclo’s analysis. The fully algebraic proof by Micchelli and Willoughby describes the ν_k s in terms of some divided differences and uses Cauchy’s interlacement theorem in an inductive argument to conclude to positivity. We will show in Sect. 4, on the one hand, that computing the probability of certain events related to our random forests Φ_q leads naturally to the divided difference representation of the ν_k s, when one has in mind their local equilibria interpretation. This will be done by using Wilson’s algorithm, which gives an alternative description of our random forests (see Sect. 2).

On the other hand, our random forest original description will lead to the key formula of the inductive step: from the random forest point of view, this algebraic formula is nothing but a straightforward connection between the previous event probabilities. Section 4 contains the full derivation of Micchelli and Willoughby’s theorem.

1.3 Coupling Random Forests, Coalescence and Fragmentation

In dealing with practical sampling issues in the next section, we will couple all the Φ_q s together in such a way that we will obtain the following side result.

Theorem 2 *There exists a (non-homogeneous) continuous-time Markov process $(F(s) \in \mathcal{F} : s \geq 0)$ that couples together all our random forests Φ_q for $q > 0$ as follows: for all $s \geq 0$ and $\phi \in \mathcal{F}$, it holds*

$$\mathbb{P}(F(s) = \phi) = \mathbb{P}(\Phi_{1/t} = \phi) = \mathbb{P}(\Phi_q = \phi)$$

with $t = 1/q$, $s = \ln(1 + \alpha t)$ and α as in Eq. (2).

With each spanning forest ϕ , we can associate a partition $\mathcal{P}(\phi)$ of \mathcal{X} , for which x and y in \mathcal{X} belong to the same class when they are in the same tree. We will see in Sect. 2.3 that the coupling $t \mapsto \Phi_{1/t} = F(\ln(1 + \alpha t))$ is then associated with a fragmentation and coalescence process, for which coalescence is strongly predominant, and at each jump time, one component of the partition is fragmented into pieces that possibly coalesce with the other components. This coupling will lead to a number of open questions: (1) Is it possible to use this process to sample efficiently Φ_q with a prescribed number of roots? (2) Can we use it to estimate the spectrum of L ? (3) How to characterize the law of the associated partition process? (See Sect. 2.3 for more details.)

2 Preliminary Remarks and Sampling Issues

2.1 Wilson Algorithm, Partition Function and the Root Process

Let us first slightly extend our notion of random forests. For any $\mathcal{B} \subset \mathcal{X}$, we denote by $\Phi_{q,\mathcal{B}}$ a random variable in \mathcal{F} with the law of Φ_q conditioned on the event $\{\mathcal{B} \subset \rho(\Phi_q)\}$. We then have, for any ϕ in \mathcal{F} ,

$$\mathbb{P}(\Phi_{q,\mathcal{B}} = \phi) = \frac{w(\phi)q^{|\rho(\phi)|-|\mathcal{B}|}}{Z_{\mathcal{B}}(q)} \mathbb{1}_{\{\mathcal{B} \subset \rho(\phi)\}}$$

with

$$Z_{\mathcal{B}}(q) = \sum_{\phi: \rho(\phi) \supset \mathcal{B}} w(\phi)q^{|\rho(\phi)|-|\mathcal{B}|}. \tag{4}$$

This law is non-degenerate even for $q = 0$, provided that \mathcal{B} is non-empty. And if \mathcal{B} is a singleton $\{r\}$, then $\Phi_{0,\{r\}}$ is the usual random spanning tree with a prescribed root r ,

which can be sampled with Wilson’s algorithm (cf. [14]). For $q > 0$, $\Phi_q = \Phi_{q,\emptyset}$ itself is also a special case of the usual random spanning tree on an extended weighted graph $\bar{\mathcal{G}} = (\bar{\mathcal{X}}, \bar{w})$ obtained by addition of an extra point r to \mathcal{X} —to form $\bar{\mathcal{X}} = \mathcal{X} \cup \{r\}$ —and by setting $\bar{w}(x, r) = q$ and $\bar{w}(r, x) = 0$ for all x in \mathcal{X} . Indeed, to get Φ_q from the usual random spanning tree on $\bar{\mathcal{X}}$, with the root in r , one only needs to remove all the edges going from \mathcal{X} to r . Following Propp and Wilson (cf. [13]), we can then use Wilson’s algorithm to sample $\Phi_{q,\mathcal{B}}$ for $q > 0$ or $\mathcal{B} \neq \emptyset$:

- a. start from $\mathcal{B}_0 = \mathcal{B}$ and $\phi_0 = \emptyset$, choose x in $\mathcal{X} \setminus \mathcal{B}_0$ and set $i = 0$;
- b. run the Markov process starting at x up to time $T_q \wedge T_{\mathcal{B}_i}$ with T_q an independent exponential random variable with parameter q (so that $T_q = +\infty$ if $q = 0$) and $T_{\mathcal{B}_i}$ the hitting time of \mathcal{B}_i ;
- c. with

$$\Gamma_{q,\mathcal{B}_i}^x = (x_0, x_1, \dots, x_k) \in \{x\} \times (\mathcal{X} \setminus (\mathcal{B}_i \cup \{x\}))^{k-1} \times (\mathcal{X} \setminus \{x\})$$

the loop-erased trajectory obtained from $X : [0, T_q \wedge T_{\mathcal{B}_i}] \rightarrow \mathcal{X}$, set $\mathcal{B}_{i+1} = \mathcal{B}_i \cup \{x_0, x_1, \dots, x_k\}$ and $\phi_{i+1} = \phi_i \cup \{(x_0, x_1), (x_1, x_2), \dots, (x_{k-1}, x_k)\}$ (so that $\phi_{i+1} = \phi_i$ if $k = 0$);

- d. if $\mathcal{B}_{i+1} \neq \mathcal{X}$, choose x in $\mathcal{X} \setminus \mathcal{B}_{i+1}$ and repeat b–d with $i + 1$ in place of i , and, if $\mathcal{B}_{i+1} = \mathcal{X}$, set $\Phi_{q,\mathcal{B}} = \phi_{i+1}$.

This algorithm is not only a practical algorithm to sample Φ_q , but also a powerful tool to analyse its law, one of its main strength points being the fact that the order of the chosen starting points x does not matter.

There are at least two ways to prove that this algorithm indeed samples $\Phi_{q,\mathcal{B}}$ with the desired law, whatever the way in which the starting points x are chosen. One can, on the one hand, follow Wilson’s original proof in [14], which makes use of the so-called Diaconis–Fulton stack representation of Markov chains (see Sect. 2.3). One can, on the other hand, follow Marchal who first computes in [10] the law of the loop-erased trajectory $\Gamma_{q,\mathcal{B}}^x$ obtained from the random trajectory $X : [0, T_q \wedge T_{\mathcal{B}}] \rightarrow \mathcal{X}$ started at $x \in \mathcal{X} \setminus \mathcal{B}$ and stopped in \mathcal{B} or at an exponential time T_q if T_q is smaller than the hitting time $T_{\mathcal{B}}$. One has indeed:

Theorem [Marchal] *For any self-avoiding path $(x_0, x_1, \dots, x_k) \in \mathcal{X}^{k+1}$ such that $x_0 = x \in \mathcal{X} \setminus \mathcal{B}$, it holds*

$$\mathbb{P}(\Gamma_{q,\mathcal{B}}^x = (x_0, x_1, \dots, x_k)) = \begin{cases} \prod_{j < k} w(x_j, x_{j+1}) \frac{\det[q - L]_{\mathcal{X} \setminus (\mathcal{B} \cup \{x_0, \dots, x_{k-1}\})}}{\det[q - L]_{\mathcal{X} \setminus \mathcal{B}}} & \text{if } x_k \in \mathcal{B}, \\ q \prod_{j < k} w(x_j, x_{j+1}) \frac{\det[q - L]_{\mathcal{X} \setminus (\mathcal{B} \cup \{x_0, \dots, x_k\})}}{\det[q - L]_{\mathcal{X} \setminus \mathcal{B}}} & \text{if } x_k \notin \mathcal{B}. \end{cases}$$

From this result, one can compute the law of $\Phi_{q,\mathcal{B}}$ defined by the previous algorithm to find, after telescopic cancellation,

$$\mathbb{P}(\Phi_{q,\mathcal{B}} = \phi) = \frac{w(\phi)q^{|\rho(\phi)| - |\mathcal{B}|}}{\det[q - L]_{\mathcal{X} \setminus \mathcal{B}}} \mathbb{1}_{\{\mathcal{B} \subset \rho(\phi)\}}, \quad \phi \in \mathcal{F}. \tag{5}$$

We can then identify the law of $|\rho(\Phi_{q,\mathcal{B}})|$ in terms of the eigenvalues $\lambda_{0,\mathcal{B}}, \lambda_{1,\mathcal{B}}, \dots, \lambda_{l-1,\mathcal{B}}$, with $l = |\mathcal{X} \setminus \mathcal{B}|$, of $-[L]_{\mathcal{X} \setminus \mathcal{B}}$. Let us write to this end

$$J = \{0, 1, 2, \dots, l - 1\} = J_0 \cup J_+ \cup J_-$$

with

$$J_0 = \{j \in J : \lambda_{j,\mathcal{B}} \in \mathbb{R}\}, \quad J_+ = \{j \in J : \text{Im}(\lambda_{j,\mathcal{B}}) > 0\},$$

$$J_- = \{j \in J : \text{Im}(\lambda_{j,\mathcal{B}}) < 0\},$$

let us set

$$p_j = \frac{q}{q + \lambda_{j,\mathcal{B}}}, \quad j \in J,$$

and define independent random variables B_j , with j in J_0 , and C_j , with j in J_+ , such that the B_j s follow Bernoulli laws and the C_j s follow convolutions of conjugated “complex Bernoulli laws”:

$$\mathbb{P}(B_j = 1) = p_j, \quad \mathbb{P}(B_j = 0) = 1 - p_j, \quad j \in J_0$$

$$\mathbb{P}(C_j = 2) = p_j \bar{p}_j, \quad \mathbb{P}(C_j = 1) = p_j(1 - \bar{p}_j) + \bar{p}_j(1 - p_j),$$

$$\mathbb{P}(C_j = 0) = (1 - p_j)(1 - \bar{p}_j), \quad \text{i.e.}$$

$$\mathbb{P}(C_j = 2) = |p_j|^2, \quad \mathbb{P}(C_j = 1) = 2\text{Re}(p_j) - 2|p_j|^2,$$

$$\mathbb{P}(C_j = 0) = 1 - 2\text{Re}(p_j) + |p_j|^2, \quad j \in J_+.$$

Note that the previous equations indeed define probability laws for the C_j s as soon as $2\text{Re}(p_j) \geq 2|p_j|^2$ for j in J_+ . This is equivalent to

$$\frac{q}{q + \lambda_{j,\mathcal{B}}} + \frac{q}{q + \bar{\lambda}_{j,\mathcal{B}}} \geq \frac{2q^2}{|q + \lambda_{j,\mathcal{B}}|^2} \Leftrightarrow 2q^2 + 2q\text{Re}(\lambda_{j,\mathcal{B}}) \geq 2q^2$$

and ensured by the fact that the eigenvalues have non-negative real part. We eventually set

$$S_{q,\mathcal{B}} = |\mathcal{B}| + \sum_{j \in J_0} B_j + \sum_{j \in J_+} C_j. \tag{6}$$

Proposition 2.1 $Z_{\mathcal{B}}$ is the characteristic polynomial of $[L]_{\mathcal{X} \setminus \mathcal{B}}$, i.e.

$$\sum_{\phi: \rho(\phi) \supset \mathcal{B}} w(\phi) q^{|\rho(\phi)| - |\mathcal{B}|} = \det(q - [L]_{\mathcal{X} \setminus \mathcal{B}}) = \prod_{j \in J} (q + \lambda_{j,\mathcal{B}}), \quad q \in \mathbb{R}, \tag{7}$$

and $|\rho(\Phi_{q,\mathcal{B}})|$ has the same law as $S_{q,\mathcal{B}}$ for $q > 0$ or $\mathcal{B} \neq \emptyset$.

Proof Equation (5) allows to identify $Z_{\mathcal{B}}$ by summing on ϕ in \mathcal{F} . The identity in law is obtained by identifying monomials with equal degree in Eq. (7) and by dividing each identity by $Z_{\mathcal{B}}(q)$ to get, for any $k \leq l = |\mathcal{X} \setminus \mathcal{B}|$,

$$\mathbb{P}(|\rho(\Phi_{q,\mathcal{B}})| = |\mathcal{B}| + k) = \sum_{\substack{I \subset J: i \in I \\ |I|=k}} \prod_{i \in I} \frac{q}{q + \lambda_{i,\mathcal{B}}} \prod_{i \in J \setminus I} \frac{\lambda_{i,\mathcal{B}}}{q + \lambda_{i,\mathcal{B}}} = \sum_{\substack{I \subset J: i \in I \\ |I|=k}} \prod_{i \in I} p_j \prod_{i \in J \setminus I} (1 - p_j).$$

Since for each j in J_+ there is j' in J_- such that $\lambda_{j',\mathcal{B}} = \bar{\lambda}_{j,\mathcal{B}}$ and $p_{j'} = \bar{p}_j$, the proof is complete. \square

The fact that Φ_q is the usual random spanning tree on an extended graph implies, through the (non-reversible) transfer current theorem (cf. [4] and [5]), that $\Phi_{q,\mathcal{B}} \subset \mathcal{E}$ is a determinantal process and so is $\rho(\Phi_{q,\mathcal{B}}) \subset \mathcal{X}$. (In the reversible case at least, the fact that the law of $|\rho(\Phi_{q,\mathcal{B}})|$ is a convolution of Bernoulli laws is also a consequence of this determinantal property.) Let us give a direct and short proof of the fact that $\rho(\Phi_{q,\mathcal{B}})$ is a determinantal process associated with a remarkable kernel.

Proposition 2.2 *For any $\mathcal{A} \subset \mathcal{X}$*

$$\mathbb{P}(\mathcal{A} \subset \rho(\Phi_{q,\mathcal{B}})) = \det[K_{q,\mathcal{B}}]_{\mathcal{A}}$$

with

$$K_{q,\mathcal{B}}(x, y) = P_x(X(T_q \wedge T_{\mathcal{B}}) = y), \quad x, y \in \mathcal{X}.$$

Proof Since for any $x \in \mathcal{B}$ we have $K_{q,\mathcal{B}}(x, \cdot) = \delta_x$, it holds $\det[K_{q,\mathcal{B}}]_{\mathcal{A}} = \det[K_{q,\mathcal{B}}]_{\mathcal{A} \setminus \mathcal{B}}$ and we can assume without loss of generality that $\mathcal{A} \subset \mathcal{X} \setminus \mathcal{B}$. By sampling $\Phi_{q,\mathcal{B}}$ with Wilson’s algorithm and choosing as first starting points for the loop-erased random walks the elements of \mathcal{A} , we have by Marchal’s formula and after telescopic cancellation

$$\begin{aligned} \mathbb{P}(\mathcal{A} \subset \rho(\Phi_{q,\mathcal{B}})) &= q^{|\mathcal{A}|} \frac{\det[q - L]_{(\mathcal{X} \setminus \mathcal{B}) \setminus \mathcal{A}}}{\det[q - L]_{\mathcal{X} \setminus \mathcal{B}}} = q^{|\mathcal{A}|} \det[[q - L]^{-1}]_{\mathcal{X} \setminus \mathcal{B}} \\ &= \det[q[q - L]^{-1}]_{\mathcal{A}}. \end{aligned}$$

The last but one equality, sometimes referred as Jacobi’s equality, is obtained from standard manipulations of the Schur complement

$$S = A - BD^{-1}C$$

of the block D in a 2×2 block matrix

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} \mathbb{1} & BD^{-1} \\ 0 & \mathbb{1} \end{pmatrix} \begin{pmatrix} S & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} \mathbb{1} & 0 \\ D^{-1}C & \mathbb{1} \end{pmatrix}.$$

This formula gives $\det M = \det S \det D$ and identifies S^{-1} as one block of

$$M^{-1} = \begin{pmatrix} S^{-1} & -S^{-1}BD^{-1} \\ -D^{-1}CS^{-1} & D^{-1} + D^{-1}CS^{-1}BD^{-1} \end{pmatrix},$$

the determinant of which is $\det S^{-1} = \det D / \det M$. The previous equality was obtained by taking $M = [q - L]_{\mathcal{X} \setminus \mathcal{B}}$ and $D = [q - L]_{(\mathcal{X} \setminus \mathcal{B}) \setminus \mathcal{A}}$ (so that S is the sub-Markovian generator of the trace on \mathcal{A} of the original process killed in \mathcal{B} and at rate q outside \mathcal{B} , while S^{-1} is the associated Green’s kernel).

Let us express $K_{q,\mathcal{B}}$ in terms of L to conclude this proof. The trajectory of X can be built by updating at each time of a Poisson process of intensity α , defined in Eq. (2), the current position $x \in \mathcal{X}$ to $y \in \mathcal{X}$ with probability

$$P(x, y) = w(x, y) / \alpha, \tag{8}$$

with the convention

$$w(x, x) = \alpha - w(x). \tag{9}$$

Since the probability of reaching the time T_q between two successive updating is $q / (q + \alpha)$, we get for any x and y in $\mathcal{X} \setminus \mathcal{B}$

$$\begin{aligned} P_x(X(T_q \wedge T_{\mathcal{B}}) = y) &= \sum_{k \geq 0} \frac{q}{q + \alpha} \left(1 - \frac{q}{q + \alpha}\right)^k [P]_{\mathcal{X} \setminus \mathcal{B}}^k(x, y) \\ &= \frac{q}{q + \alpha} \left(1 - \frac{\alpha}{q + \alpha} [P]_{\mathcal{X} \setminus \mathcal{B}}\right)^{-1}(x, y) \\ &= q(q - \alpha [P - 1]_{\mathcal{X} \setminus \mathcal{B}})^{-1}(x, y) = q[q - L]_{\mathcal{X} \setminus \mathcal{B}}^{-1}(x, y). \end{aligned}$$

And for any $\mathcal{A} \subset \mathcal{X} \setminus \mathcal{B}$ it holds $[K_{q,\mathcal{B}}]_{\mathcal{A}} = [q[q - L]_{\mathcal{X} \setminus \mathcal{B}}^{-1}]_{\mathcal{A}}$. □

We conclude this section by observing that the Markov chain tree theorem (see, for example, [1]) allows to compute the root distribution when conditioning on $\mathcal{P}(\Phi_q)$, the partition of \mathcal{X} that is associated with Φ_q . We write $\mathcal{P}(\Phi_q) = [\mathcal{X}_1, \dots, \mathcal{X}_m]$ if Φ_q is made of m trees, each of them spanning one of the \mathcal{X}_i s. For each $i \leq m$, we denote by L_i the generator of X restricted to \mathcal{X}_i , which is defined by

$$(L_i f)(x) = \sum_{y \in \mathcal{X}_i} w(x, y)[f(y) - f(x)], \quad x \in \mathcal{X}_i, \quad f : \mathcal{X}_i \rightarrow \mathbb{C},$$

and by X_i this restricted process. Since, by construction, the root of the spanning tree of \mathcal{X}_i is reachable by X_i from any point in \mathcal{X}_i , this process admits only one invariant measure μ_i , which is equal to $\mu(\cdot | \mathcal{X}_i)$ (recall that μ is the invariant measure of X) when X is reversible. If $[\mathcal{X}_1, \dots, \mathcal{X}_m]$ is an admissible partition of \mathcal{X} , that is if $\mathcal{P}(\Phi_q) = [\mathcal{X}_1, \dots, \mathcal{X}_m]$ with nonzero probability, then, denoting by \mathcal{T}_i the set of

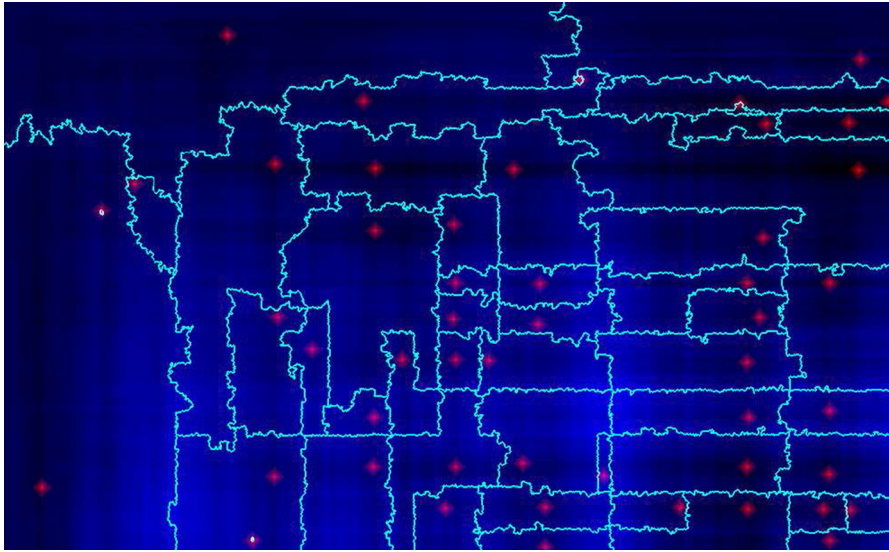


Fig. 1 A sample of $\mathcal{P}(\Phi_q)$ and $\rho(\Phi_q)$ with 50 roots on the 987×610 rectangular grid and for the Metropolis random walk at inverse temperature $\beta = .06$ in a Brownian sheet potential V , i.e. such that nearest-neighbours rates are given by $w(x, y) = \exp\{-\beta[V(y) - V(x)]_+\}$ with V being the grid restriction of a Brownian sheet with 0 value on the north and west sides of the box. This random walk is reversible with respect to $\exp\{-\beta V\}$. The cyan lines separate neighbouring trees, the roots are at the centre of the red diamonds, and blue levels depend on the potential: the darker the blue, the lower the potential (Color figure online)

spanning trees of \mathcal{X}_i and by $\rho(\tau_i)$ the root $x_i \in \mathcal{X}_i$ of $\tau_i \in \mathcal{T}_i$, we can compute for any (x_1, \dots, x_m) in $\mathcal{X}_1 \times \dots \times \mathcal{X}_m$

$$\begin{aligned} & \mathbb{P}\left(\rho(\Phi_q) = \{x_1, \dots, x_m\} \mid \mathcal{P}(\Phi_q) = [\mathcal{X}_1, \dots, \mathcal{X}_m]\right) \\ &= \frac{q^m \sum_{\tau_1 \in \mathcal{T}_1} \dots \sum_{\tau_m \in \mathcal{T}_m} \prod_{i=1}^m w(\tau_i) \mathbb{1}_{\{\rho(\tau_i)=x_i\}}}{q^m \sum_{\tau_1 \in \mathcal{T}_1} \dots \sum_{\tau_m \in \mathcal{T}_m} \prod_{i=1}^m w(\tau_i)} \\ &= \prod_{i=1}^m \frac{\sum_{\tau_i \in \mathcal{T}_i} w(\tau_i) \mathbb{1}_{\{\rho(\tau_i)=x_i\}}}{\sum_{\tau_i \in \mathcal{T}_i} w(\tau_i)}. \end{aligned}$$

The Markov chain tree theorem gives then

Proposition 2.3 *For any admissible partition $[\mathcal{X}_1, \dots, \mathcal{X}_m]$ and any (x_1, \dots, x_m) in $\mathcal{X}_1 \times \dots \times \mathcal{X}_m$, it holds*

$$\mathbb{P}\left(\rho(\Phi_q) = \{x_1, \dots, x_m\} \mid \mathcal{P}(\Phi_q) = [\mathcal{X}_1, \dots, \mathcal{X}_m]\right) = \prod_{i=1}^m \mu_i(x_i).$$

See Fig. 1 for an illustration with the two-dimensional nearest-neighbour random walk in a Brownian sheet potential, which is easy to sample and gives rise to a rich and anisotropic energy landscape.

2.2 Sampling Approximately m Roots

While Wilson’s algorithm provides a practical way to sample Φ_q , we do not have such an algorithm for Φ_q conditioned on $\{|\rho(\Phi_q)| = m\}$. In this section, we explain how to get Φ_q with approximately m roots, with an error of order \sqrt{m} at most. By Proposition 2.1, it suffices to choose q solution of

$$\sum_{j < n} \frac{q}{q + \lambda_j} = m. \tag{10}$$

Indeed,

$$\mathbb{E}[|\rho(\Phi_q)|] = \sum_{j \in J_0} \mathbb{E}[B_j] + \sum_{j \in J_+} \mathbb{E}[C_j] = \sum_{j \in J} \frac{q}{q + \lambda_j},$$

while

$$\begin{aligned} \text{Var}(B_j) &= p_j(1 - p_j) \leq p_j = \mathbb{E}[B_j], \quad j \in J_0 \\ \text{Var}(C_j) &= 2\text{Re}(p_j) + 2|p_j|^2 - 4\text{Re}(p_j)^2 \leq 4\text{Re}(p_j) = 2\mathbb{E}[C_j], \quad j \in J_+, \end{aligned}$$

so that $\text{Var}(|\rho(\Phi_q)|) \leq 2\mathbb{E}[|\rho(\Phi_q)|]$. But we do not want to solve Eq. (10) analytically, and we do not want to compute the eigenvalues λ_j . One way to find an approximate value of the solution q^* of Eq. (10) is to use, on the one hand, the fact that q^* is the only one stable attractor of the recursive sequence defined by $q_{k+1} = f(q_k)$ with

$$f : q > 0 \mapsto q \times \frac{m}{\sum_{j < n} \frac{q}{q + \lambda_j}} = \frac{m}{\sum_{j < n} \frac{1}{q + \lambda_j}},$$

on the other hand, the fact that $|\rho(\Phi_q)|$ and $\mathbb{E}[|\rho(\Phi_q)|] = \sum_{j < n} q/(q + \lambda_j)$ are typically of the same order, at least when $\mathbb{E}[|\rho(\Phi_q)|]$, i.e. q , is large enough, since $\text{Var}(|\rho(\Phi_q)|)/\mathbb{E}^2[|\rho(\Phi_q)|] \leq 2/\mathbb{E}[|\rho(\Phi_q)|]$. We then propose the following algorithm to sample Φ_q with $m \pm 2\sqrt{m}$ roots.

- a. Start from any $q_0 > 0$, for example $q_0 = \alpha = \max_{x \in \mathcal{X}} w(x)$, and set $i = 0$.
- b. Sample Φ_{q_i} with Wilson’s algorithm.
- c. If $|\rho(\Phi_{q_i})| \notin [m - 2\sqrt{m}, m + 2\sqrt{m}]$, set $q_{i+1} = mq_i/|\rho(\Phi_{q_i})|$ and repeat b–c with $i + 1$ instead of i , if $|\rho(\Phi_{q_i})| \in [m - 2\sqrt{m}, m + 2\sqrt{m}]$, then return Φ_{q_i} .

To see that this algorithm rapidly produces the desired result, it is convenient to write $\gamma = \ln q$ and introduce the global contraction

$$g : \gamma \in \mathbb{R} \mapsto \ln f(e^\gamma).$$

While f is a contraction in a neighbourhood of q^* only, let us show that g is indeed a global contraction. For all $\gamma \in \mathbb{R}$, it holds

$$g'(\gamma) = \frac{\sum_{j < n} \left(\frac{q}{q+\lambda_j}\right)^2}{\sum_{j < n} \frac{q}{q+\lambda_j}} = \frac{\sum_{j \in J_0} \left(\frac{q}{q+\lambda_j}\right)^2 + 2 \sum_{j \in J_+} \operatorname{Re}^2\left(\frac{q}{q+\lambda_j}\right) - \operatorname{Im}^2\left(\frac{q}{q+\lambda_j}\right)}{\sum_{j \in J_0} \frac{q}{q+\lambda_j} + 2 \sum_{j \in J_+} \operatorname{Re}\left(\frac{q}{q+\lambda_j}\right)}.$$

With $\alpha_j = \operatorname{Re}(\lambda_j) > 0$ and $\beta_j = \operatorname{Im}(\lambda_j) > 0$ for j in J_+ , we have

$$0 < \operatorname{Re}\left(\frac{q}{q+\lambda_j}\right) = \frac{q(q+\alpha_j)}{(q+\alpha_j)^2 + \beta_j^2} < 1 \quad \text{and} \quad \operatorname{Im}\left(\frac{q}{q+\lambda_j}\right) = \frac{-q\beta_j}{(q+\alpha_j)^2 + \beta_j^2},$$

so that

$$\operatorname{Re}^2\left(\frac{q}{q+\lambda_j}\right) \vee \operatorname{Im}^2\left(\frac{q}{q+\lambda_j}\right) < \operatorname{Re}\left(\frac{q}{q+\lambda_j}\right),$$

since $0 < \operatorname{Re}(q/(q+\lambda_j)) < 1$ and

$$\frac{\operatorname{Im}^2\left(\frac{q}{q+\lambda_j}\right)}{\operatorname{Re}\left(\frac{q}{q+\lambda_j}\right)} = \frac{q\beta_j^2}{(q+\alpha_j)((q+\alpha_j)^2 + \beta_j^2)} < 1.$$

Then, $g'(\gamma)$ being the difference between two non-negative terms that are strictly smaller than 1, we have $|g'(\gamma)| < 1$, for all γ in \mathbb{R} . Now, writing for $k \geq 0$

$$\epsilon_k = |\ln(1 + \delta_k)| = \left| \ln\left(\frac{|\rho(\Phi_{q_k})|}{\mathbb{E}[|\rho(\Phi_{q_k})|]}\right) \right| \quad \text{and} \quad \gamma_{k+1} = \ln q_{k+1} = g(\gamma_k) - \ln(1 + \delta_k),$$

there are some non-negative $\theta_k < 1$ such that, with $\gamma^* = \ln q^* = g(\gamma^*)$,

$$|\gamma_{k+1} - \gamma^*| \leq \theta_k |\gamma_k - \gamma^*| + \epsilon_k,$$

and, by induction on $k > 0$,

$$\begin{aligned} |\gamma_k - \gamma^*| &\leq \theta_{k-1} \cdots \theta_0 |\gamma_0 - \gamma^*| + \theta_{k-1} \cdots \theta_1 \epsilon_0 + \theta_{k-1} \cdots \theta_2 \epsilon_1 \\ &\quad + \cdots + \theta_{k-1} \epsilon_{k-2} + \epsilon_{k-1}. \end{aligned}$$

Since Chebyshev’s inequality gives for all $\delta > 0$

$$\mathbb{P}(|\delta_k| > \delta) \leq \frac{\operatorname{Var}(|\rho(\Phi_{q_k})|)}{\delta^2 \mathbb{E}^2[|\rho(\Phi_{q_k})|]} \leq \frac{2}{\delta^2 \mathbb{E}[|\rho(\Phi_{q_k})|]},$$

after “a few iterations” (we cannot be more precise in absence of extra information on the spectrum of L that would be needed to give uniform bounds on $|g'|$ and the θ_k s)

the approximation error for γ^* is of the same order as ϵ_{k-1} —itself of order $1/\sqrt{m}$ at most—and we get m roots for Φ_q within an error of order \sqrt{m} .

2.3 Coupled Forests

Instead of stopping the iterations of the previous algorithm when reaching a forest with $m \pm 2\sqrt{m}$ roots, one can proceed up to reaching exactly m roots. This typically requires order \sqrt{m} extra iterations at most, and this is what we have done to obtain the exactly 50 roots of Fig. 1. Starting with $q = q_0$ larger than the solution q^* of Eq. (10), it takes generally much more time to reach exactly m roots than to decrease q down to a good approximation of q^* according to the updating procedure $q \leftarrow q \times m/|\rho(\Phi_q)|$. For example, starting from $q = 4$ for the Metropolis random walk in Brownian sheet potential of Fig. 1, we got 361,782 roots at the first iteration, 51 roots and $q = 5.26 \times 10^{-6}$ at the tenth iteration, and we needed 55 extra iterations to get exactly 50 roots with $q = 4.92 \times 10^{-6}$, getting in the mean time root numbers oscillating between 43 and 59 for q between 3.96×10^{-6} and 6.07×10^{-6} . While decreasing q , we produce a number of forests with a larger root number than desired, and, sampling for large q being less time-consuming than sampling for small q , the total running time of the iterations to decrease q to the correct order is essentially of the same order as the running time of one iteration for q of this correct order. This suggests that if we could continuously decrease q in such a way that Φ_q would cross all the manifolds

$$\mathcal{F}_m = \{ \phi \in \mathcal{F} : |\rho(\phi)| = m \}, \quad m \leq n, \tag{11}$$

then we might be able to find a more efficient algorithm to sample Φ_q with a prescribed root number. It turns out that we are able to implement such a “continuously decreasing q algorithm,” building in this way the coupling of Theorem 2. But this is not sufficient to improve our sampling algorithm for a prescribed root number.

In this section, we prove Theorem 2, characterize the associated root process and describe the associated coalescence and fragmentation process, which leads to further open questions. This coupling is the natural extension of Wilson’s algorithm based on Diaconis and Fulton’s stack representation of random walk (cf. [6]) as used by Wilson and Propp in [14] and [13].

Stack representations Assume that an infinite list or collection of arrows is attached to each site of the graph, each arrow pointing towards one of its neighbour. Assume in addition that these arrows are distributed according to the probability kernel P of the discrete-time skeleton of X which is defined by Eqs. (8)–(9). Assume in other words that these arrows are independently distributed at each level of the stacks and that an arrow pointing towards the neighbour y of a given site x appears with probability $P(x, y)$, considering in this context x itself as one of its neighbours. Imagine finally that each list of arrows attached to any site is piled down in such a way that it makes sense to talk of an infinite stack with an arrow on the top of this stack. By using this representation, one can generate the Markov process as follows: at each jump time of a Poisson process with intensity α , our walker steps to the neighbour pointed by the

arrow at the top of the stack where it was sitting, and the top arrow is erased from this stack.

To describe Wilson's algorithm for sampling Φ_q , one has to introduce a further ingredient: pointers to an absorbing state r in each stack. Such a pointer should independently appear with probability $q/(q + \alpha)$ at each level in the different stacks. One way to introduce it is by generating independent uniform random variables U together with each original arrow in the stacks. We can then replace the latter by a pointer to the absorbing state whenever $U < q/(q + \alpha)$. A possible description of Wilson's algorithm is then the following.

- a. Start with a particle on each site. Both particles and sites will be declared either *active* or *frozen*. At the beginning, all sites and particles are declared to be active.
- b. Choose an arbitrary particle among all the active ones and look at the arrow at the top of the stack it is seated on. Call x the site where the particle is seated.
 - If the arrow is the pointer to r , declare the particle to be frozen and site x as well.
 - If the arrow points towards another site $y \neq x$, remove the particle and keep the arrow. We say that this arrow is *uncovered*.
 - If the arrow points to x itself, remove the arrow.
- c. Once again, choose an arbitrary particle among all the active ones, look at the arrow on the top of the stack it is seated on, and call x the site where the particle is seated.
 - If the arrow points to r , the particle is declared to be frozen, and so are declared x and all the sites eventually leading to x by following uncovered top pile arrow paths.
 - If the arrow points to a frozen site, remove the chosen particle at x , keep the (now uncovered) arrow, and freeze the site x as well as any site eventually leading to x by following uncovered top pile arrow paths.
 - If the arrow points to an active site, then there are two possibilities. By following from this site the uncovered arrows at the top of the stacks, we either reach a different active particle or run in a loop back to x . In the former case, remove the chosen particle from site x and keep the discovered arrow. In the latter case, erase all the arrows along the loop and put an active particle on each site of the loop. Note that this last case includes the possibility for the discovered arrow of pointing to x itself, in which case we just have to remove the discovered arrow.
- d. Repeat the previous step up to exhaustion of the active particles.

The crucial observation, which is due to Propp and Wilson, is that whatever the choice of active particles all along the algorithm, at the end of the day the same arrows are erased and *the same spanning forest of uncovered arrows, with a frozen particle at each root, is obtained*. In particular, by choosing at each step the last encountered active particle, or the same as in the previous step when we just erased a loop, we perform a simple loop-erased random walk up to freezing.

Proof of Theorem 2. Since Φ_q is sampled for any q by the previously described algorithm and the same uniform variables U can be used for each q , this provides a global

coupling for all the Φ_{q_1} . We first note that this coupling allows to sample Φ_{q_2} from a sampled Φ_{q_1} for $q_2 < q_1$. Indeed, by running this algorithm for sampling Φ_{q_2} , one can reach at some point the spanning forest of uncovered arrows Φ_{q_1} with this difference that the frozen particles of the final configuration obtained with parameter q_1 can be still active at this intermediate step of the algorithm run with q_2 : it suffices to choose the sequence of active particles in the same way with both parameters, and this is possible since each pointer to r in the stacks with parameter q_2 is associated with a pointer to r at the same level in the stacks with parameter q_1 . Thus, to sample Φ_{q_2} from a sampled Φ_{q_1} , we just have to replace some frozen particles in $\rho(\Phi_{q_1})$ and continue the algorithm with parameter q_2 . To decide which particle has to be unfrozen we can proceed as follows. With probability

$$p = \mathbb{P}\left(U < \frac{q_2}{q_2 + \alpha} \mid U < \frac{q_1}{q_1 + \alpha}\right) = \frac{q_2(q_1 + \alpha)}{q_1(q_2 + \alpha)} \tag{12}$$

each particle in $\rho(\Phi_{q_1})$, independently from each other, is kept frozen. With probability $1 - p$ a particle in a site x of $\rho(\Phi_{q_1})$ is declared active and we set at the top of the pile in x an arrow that points towards y with probability $P(x, y) = w(x, y)/\alpha$.

When $q = 1/t$ continuously decreases, we obtain a right-continuous process $t \mapsto \Phi_{1/t}$, for which we can practically sample not only the “finite dimensional distributions”—i.e. the law of $(\Phi_{1/t_1}, \dots, \Phi_{1/t_k})$ for any choice of $t_1 < \dots < t_k$ —but the whole trajectories $(\Phi_{1/t} : t \leq t^*)$ too, for any finite t^* . Indeed, at each time $t = 1/q$, the next frozen particle to become active is uniformly distributed among the m roots at time t , and the next jump time T when it will “wake up” is such that the random variable

$$V = \frac{1/T}{\alpha + 1/T} = \frac{1}{1 + \alpha T} \tag{13}$$

has the law of the maximum of m independent uniform variables on $[0, q/(q + \alpha)] = [0, 1/(1 + \alpha t)]$. Since

$$\mathbb{P}(V < v) = (v(1 + \alpha t))^m, \quad v < \frac{1}{1 + \alpha t},$$

V has the same law as $U^{1/m}/(1 + \alpha t)$ with U uniform on $[0, 1)$. Using Eq. (13), we can then sample the next jump time T by solving

$$\frac{1 + \alpha T}{1 + \alpha t} = U^{-1/m}. \tag{14}$$

Setting $s = \ln(1 + \alpha t)$ and $S = \ln(1 + \alpha T)$, the random variable $S - s$ has the same law as the minimum of m independent exponential random variables of rate 1.

Our Markov process $(F(s) \in \mathcal{F} : s \geq 0)$ is then built in the following way. We associate m independent exponential random clocks of rate 1 with the m roots of $F(s)$ at time s . At the first ring time $S \geq s$ at some root x , we define $F(S)$ by declaring active the particle at x , putting an arrow to y with probability $P(x, y) = w(x, y)/\alpha$ and restarting our algorithm with parameter $q = 1/T = \alpha/(e^S - 1)$. \square

A *determinantal formula for the associated root process*. Proposition 2.2, from which we recall the definition of the probability kernel $K_{q,\mathcal{B}}$, can be extended to characterize the law of the coupled root process $t \mapsto \rho(\Phi_{1/t})$.

Proposition 2.4 *For all $0 < t_1 < \dots < t_k < t_{k+1} = 1/q_{k+1}$ and all $\mathcal{A}_1, \dots, \mathcal{A}_k, \mathcal{A}_{k+1}$ contained in \mathcal{X} , it holds*

$$\begin{aligned} & \mathbb{P}(\mathcal{A}_{k+1} \subset \rho(\Phi_{1/t_{k+1}}) \mid \mathcal{A}_k \subset \rho(\Phi_{1/t_k}), \dots, \mathcal{A}_1 \subset \rho(\Phi_{1/t_1})) \\ &= \sum_{\mathcal{B}_k \subset \mathcal{A}'_k} \sum_{\mathcal{B}_{k-1} \subset \mathcal{A}'_{k-1}} \dots \sum_{\mathcal{B}_1 \subset \mathcal{A}'_1} \prod_{i=1}^k \left(\frac{t_i}{t_{k+1}}\right)^{|\mathcal{B}_i|} \left(1 - \frac{t_i}{t_{k+1}}\right)^{|\mathcal{A}'_i \setminus \mathcal{B}_i|} \det[K_{q_{k+1}, \mathcal{B}}]_{\mathcal{A}_{k+1}} \\ & \text{with } \mathcal{A}'_k = \mathcal{A}_k, \mathcal{A}'_{k-1} = \mathcal{A}_{k-1} \setminus \mathcal{A}_k, \dots, \mathcal{A}'_1 = \mathcal{A}_1 \setminus (\mathcal{A}_k \cup \mathcal{A}_{k-1} \cup \dots \cup \mathcal{A}_2) \\ & \text{and } \mathcal{B} = \bigcup_{i=1}^k \mathcal{B}_i. \end{aligned} \tag{15}$$

Proof Let us first consider the case $k = 1$, so that $\mathcal{A}'_1 = \mathcal{A}_1$. As far as $\Phi_{1/t}$ is concerned for $t > t_1$, conditioning on $\{\mathcal{A}_1 \subset \rho(\Phi_1)\}$ is nothing but a conditioning on the value of the uniform random variables at the top of the stacks in \mathcal{A}_1 . With $q_1 = 1/t_1$, we cannot sample Φ_{q_2} conditioned on $\{\mathcal{A}_1 \subset \rho(\Phi_1)\}$ by keeping “frozen” each site in \mathcal{A}_1 with probability p defined by Eq. (12), calling \mathcal{B} the set of the remaining frozen sites and sampling $\Phi_{q_2, \mathcal{B}}$ with this random $\mathcal{B} \subset \mathcal{X}$, so that the root set would be a determinantal process with kernel $K_{q_2, \mathcal{B}}$. The walking up procedure we defined after Eq. (12) indeed introduces a bias in the distribution at the top of the pile for the unfrozen sites: top pile arrows cannot be replaced by pointers to r . To recover a determinantal process with random kernel $K_{q_2, \mathcal{B}}$ for the conditional root process, the random set \mathcal{B} has to be built by keeping frozen each site in \mathcal{A}_1 with a smaller probability p' solving

$$p = p' + (1 - p') \frac{q_2}{q_2 + \alpha}.$$

Top pile arrows of unfrozen sites can then still be replaced by pointers to r with probability $q_2/(q_2 + \alpha)$, and this equation makes that we recover the correct biased probability. Solving it, we get $p' = q_2/q_1 = t_1/t_2$ and Eq. (15).

When k is larger than 1, the formula is simply obtained by keeping frozen each site x in $\bigcup_{i \leq k} \mathcal{A}_k$ with a probability that depends on the largest i such that $x \in \mathcal{A}_i$. This is the reason why we introduced the sets $\mathcal{A}'_i : i^*$ is the largest i such that $x \in \mathcal{A}_i$ if and only if $x \in \mathcal{A}'_{i^*}$. □

Fragmentation, coalescence and open questions. At each jump time $S = S_{k+1}$ of F and in the proof of Theorem 2, there is only one root x to “wake up,” which means that there is only one piece of the associated partition into m pieces at the previous jump time S_k that can be fragmented into different trees, the other pieces of the previous partition remaining contained in different pieces of the new partition at time S_{k+1} . At time S_{k+1} we can have both fragmentation, produced by the loop-erasure procedure, and coalescence: the trees covering the possibly fragmented piece can be eventually

grafted to the other $m - 1$ non-fragmented frozen trees, when their associated loop-erased random walk freezes by running into these frozen trees.

Fragmentation can increase the total number of pieces up to $m + k - 1$, with k the number of sites in the tree that is rooted at x : this happens when this tree is completely fragmented and no coalescence occurs. Coalescence can decrease the number of pieces by 1 at most: when each tree of the possibly fragmented piece is eventually grafted to the other pieces. But coalescence strongly dominates the process: as $q = 1/t$ decreases, so does $\mathbb{E}[|\rho(\Phi_q)|]$, with limited fluctuations, as a consequence of Proposition 2.1 (cf. Figs. 2, 3, 4). And the fact that when $|\rho(\Phi_q)|$ decreases, it does so by one unit at most, implies that the process $t \mapsto \Phi_{1/t}$ crosses all the manifolds \mathcal{F}_m defined by Eq. (11).

It is then easy to sample Φ_{1/T_m} with T_m the first time when $t \mapsto \Phi_{1/t}$ reaches \mathcal{F}_m . Unfortunately, Φ_{1/T_m} has not the same law as Φ_q conditioned on $\{|\rho(\Phi_q)| = m\}$. One has indeed a counterexample already for $n = 2$ and $m = 1$. The random set $\rho(\Phi_{T_m})$ is, generally, not even well distributed: for $m = n - 1$, there is only one distribution on the subsets of size m that produces well-distributed points. We then get to our first open question

Q1: Is there a way to use the process $t \mapsto \Phi_{1/t}$ to sample the measure $\mathbb{P}(\Phi_q \in \cdot \mid |\rho(\Phi_q)| = m)$?

One can also use this process to estimate $t \geq 0 \mapsto \sum_j 1/(1+t\lambda_j)$ since this sum is the expected value of $|\rho(\Phi_{1/t})|$, which presents limited fluctuations only (see Fig. 4). This leads to our second open question

Q2: Is there a way to use the process $t \mapsto \Phi_{1/t}$ to estimate in an efficient way the spectrum of $-L$, or its higher part at least?

Our third open question concerns the law of the “rooted partition” associated with the forest process $t \mapsto \Phi_{1/t}$. (We call it *rooted* since a special vertex, the root, is associated with each piece of the partition.) Figures 3 and 4 and Wilson’s algorithm show that as $q = 1/t$ decreases, the partition process naturally tends to break the space into larger and larger valleys in which the process is trapped on time scale $t = 1/q$ (note that the difference of $12 = 27 - 15$ between the extreme values of $s = \ln(1+\alpha t)$ in the right picture of Fig. 4 corresponds to a ratio of order 1.6×10^5 between the associated times t). But, while we could characterize in Proposition 2.4 the law of the associated root process, we are far from obtaining a similar result for the rooted partition.

Q3: Which characterization can be given of the rooted partition associated with $t \mapsto \Phi_{1/t}$?

We actually know very little beyond Proposition 2.3 on this partition for a fixed value of q and an easier question would be that of characterizing the law of the forest process itself. Even though Fig. 2 echoes Figure 5 of [3], which illustrates a coalescence process that is also associated with random spanning forests, the two processes are quite different and we do not know the scaling limit of our process, even for a fixed value of q . The process considered in [3] is a pure coalescence process, while fragmentation is also involved in our case; at a fixed time t , the tree number in that case of the

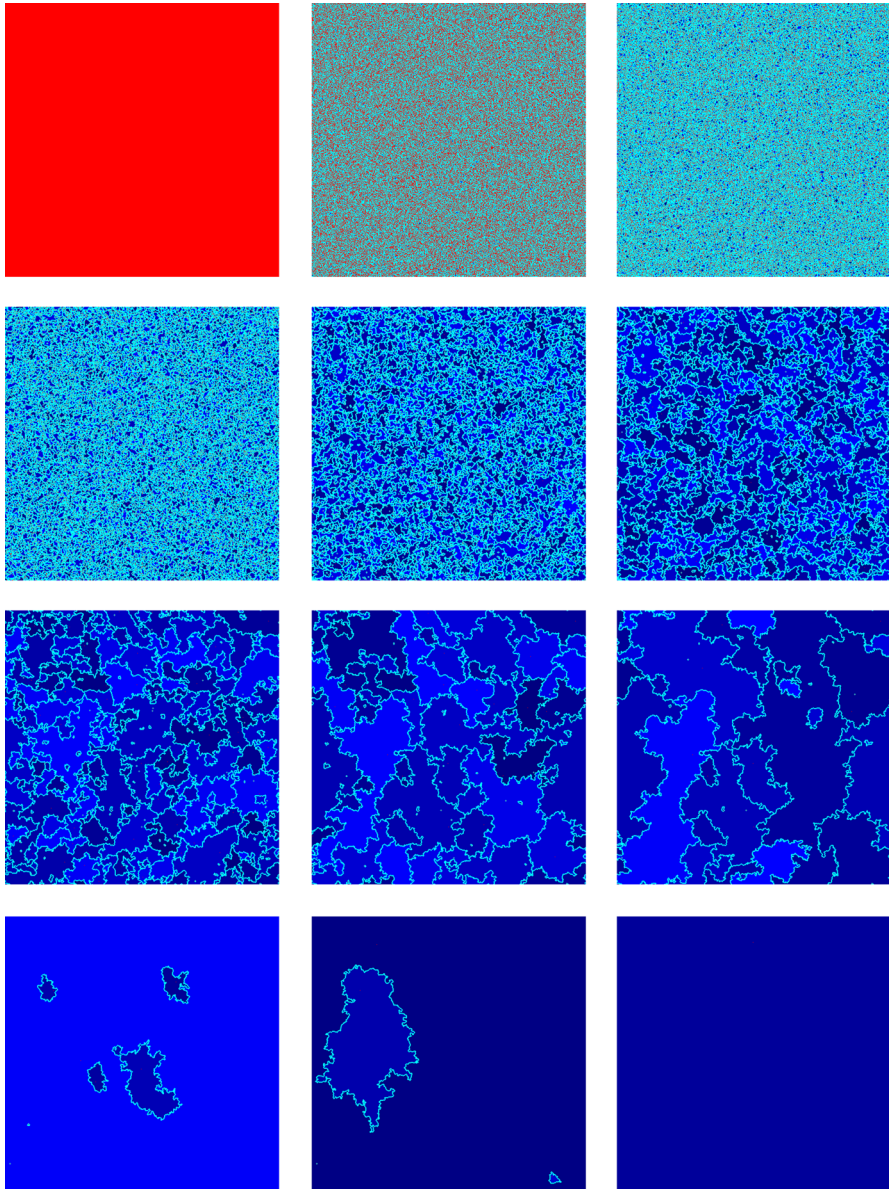


Fig. 2 Snapshots at times $s = \ln(1 + \alpha t)$ with $t = 1/q$ equal to 0, .5, 2, 8, 32, 128, 512, ..., 524,288 of the coalescence and fragmentation process $s \mapsto F(s)$ for the simple random walk on the torus with uniform nearest-neighbours rates $w(x, y) = 1$. Roots are *red*, non-root leaves are *cyan*, and other vertices are *blue*, different shades of *blue* being used for different trees (Color figure online)

uniformly cut uniform spanning tree follows a binomial distribution, while the tree number of our process is distributed as a sum of Bernoulli random variables with non-homogeneous parameters; and, even conditioned on a same tree number, if the weights

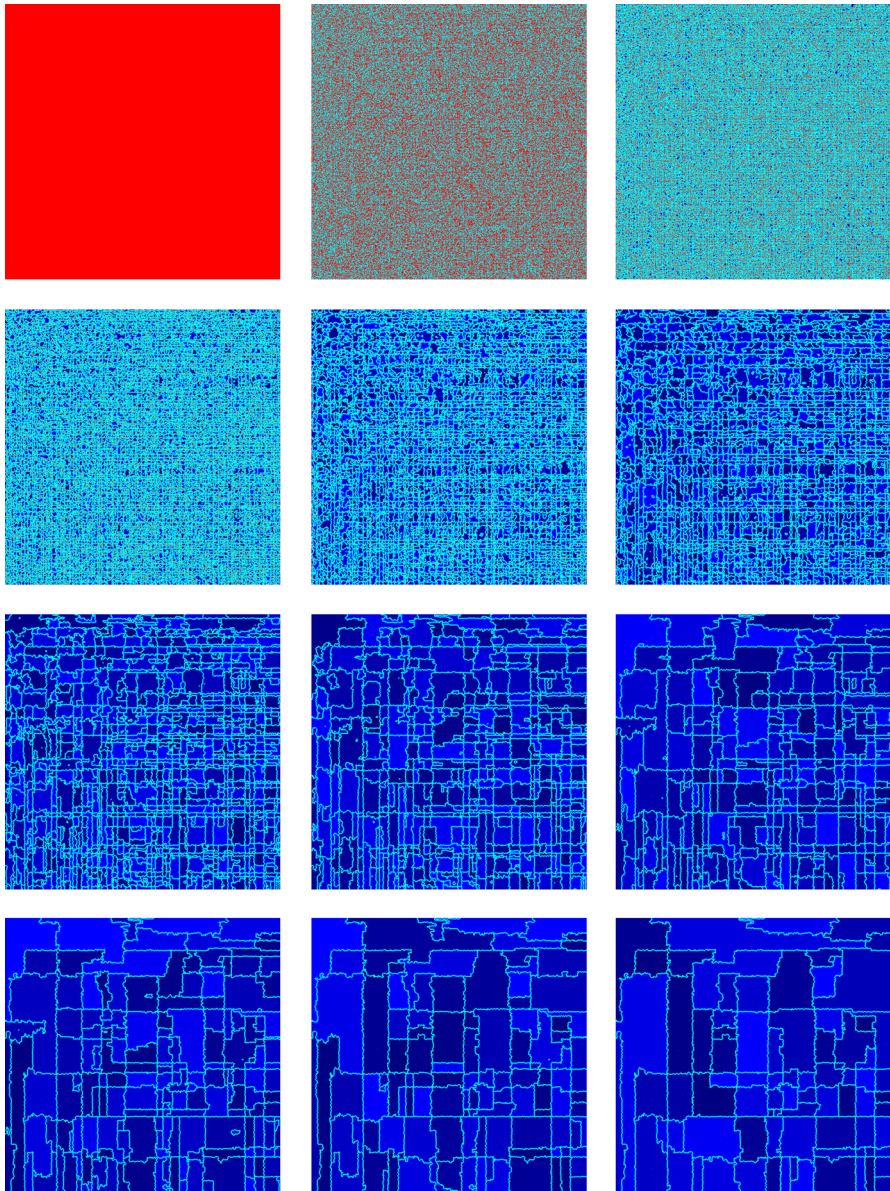


Fig. 3 Snapshots at times $s = \ln(1 + \alpha t)$ with $t = 1/q$ equal to 0, .5, 2, 8, 32, 128, 512, ..., 524,288 of the coalescence and fragmentation process $s \mapsto F(s)$ on the square grid for the random walk in a Brownian sheet potential with inverse temperature $\beta = .16$. The colour conventions are the same as in Fig. 2 (Color figure online)

of the associated partitions share the same product of unrooted spanning tree number for each piece of the partition, the extra entropic factor depends in that case of these pieces' boundaries, while in our case it is simply given by the product of their size.

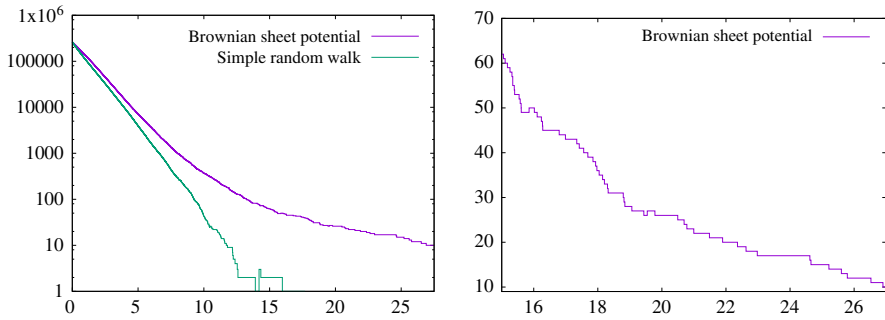


Fig. 4 The left picture shows the tree number as a function of time for the coalescence and fragmentation processes $s \mapsto F(s)$ of Figs. 2 and 3 in semi-logarithmic scale. The right picture shows the same quantity in natural scale from time 15 to 27 for the coalescence and fragmentation process of Fig. 3 only

3 Hitting Times

3.1 Forest Formulas for Hitting Distributions, Green’s Kernels and Mean Hitting Time

In order to prove Theorem 1, we first use Wilson’s algorithm to give forest representations of hitting distributions, Green’s kernels and mean hitting times. Two at least of these formulas, Formula (16) and Formula (17), already appeared in the work of Freidlin and Wentzell (see Lemma 3.2 and Lemma 3.3 in [8]).

We recall that $T_{\mathcal{B}}$ stands for the hitting time of $\mathcal{B} \subset \mathcal{X}$, we denote by $\tau_x(\phi)$ the unique maximal tree in $\phi \in \mathcal{F}$ that covers $x \in \mathcal{X}$, and we recall that $\rho(\tau_x(\phi)) \in \rho(\phi)$ is the root of $\tau_x(\phi)$ and recall Eq. (4). By considering Wilson’s algorithm for sampling $\Phi_{0,\mathcal{B}}$ and choosing x as first starting point for the loop-erased random walk, we first note that for all $y \in \mathcal{B}$, it holds

$$P_x(X(T_{\mathcal{B}}) = y) = \frac{1}{Z_{\mathcal{B}}(0)} \sum_{\substack{\phi: \rho(\phi) = \mathcal{B}, \\ \rho(\tau_x(\phi)) = y}} w(\phi). \tag{16}$$

We then get the following forest representation of the Green’s kernel

$$G_{\mathcal{B}}(x, z) = E_x \left[\int_0^{T_{\mathcal{B}}} \mathbb{1}_{\{X(t)=z\}} dt \right], \quad x, z \in \mathcal{X}.$$

Lemma 3.1 For any $\mathcal{B} \subset \mathcal{X}$, $x \in \mathcal{X}$ and $z \notin \mathcal{B}$, it holds

$$G_{\mathcal{B}}(x, z) = \frac{1}{Z_{\mathcal{B}}(0)} \sum_{\substack{\phi: \rho(\phi) = \mathcal{B} \cup \{z\}, \\ \rho(\tau_x(\phi)) = z}} w(\phi).$$

We finally get Freidlin and Wentzell’s forest representation for mean hitting times:

$$E_x[T_{\mathcal{B}}] = \sum_{z \notin \mathcal{B}} G_{\mathcal{B}}(x, z) = \frac{1}{Z_{\mathcal{B}}(0)} \sum_{z \notin \mathcal{B}} \sum_{\substack{\phi: \rho(\phi) = \mathcal{B} \cup \{z\} \\ \rho(\tau_x(\phi)) = z}} w(\phi). \tag{17}$$

Proof of Lemma 3.1: We introduce once again the discrete skeleton of X —that is the Markov chain \hat{X} with transition kernel P defined by Eqs. (8)–(9)—and we call $\hat{G}_{\mathcal{B}}$ the Green’s kernel of \hat{X} stopped in \mathcal{B} . Let us denote by \hat{T}_z , \hat{T}_z^+ and $\hat{T}_{\mathcal{B}}$ the hitting time of z , the return time to z and the hitting time of \mathcal{B} for the Markov chain \hat{X} . Since $\hat{G}_{\mathcal{B}}(x, z) = P_x(\hat{T}_z < \hat{T}_{\mathcal{B}})\hat{G}_{\mathcal{B}}(z, z) = P_x(\hat{T}_z < \hat{T}_{\mathcal{B}})/P_z(\hat{T}_z^+ > \hat{T}_{\mathcal{B}})$, it holds

$$\begin{aligned} G_{\mathcal{B}}(x, z) &= \frac{1}{\alpha} \hat{G}_{\mathcal{B}}(x, z) = \frac{P_x(\hat{T}_z < \hat{T}_{\mathcal{B}})}{\alpha P_z(\hat{T}_z^+ > \hat{T}_{\mathcal{B}})} = \frac{P_x(\hat{T}_z < \hat{T}_{\mathcal{B}})}{\alpha \sum_{y \neq z} P(z, y) P_y(\hat{T}_z > \hat{T}_{\mathcal{B}})} \\ &= \frac{P_x(T_z < T_{\mathcal{B}})}{\sum_{y \neq z} w(z, y) [1 - P_y(T_z < T_{\mathcal{B}})]}. \end{aligned}$$

Then, since $P_y(T_z < T_{\mathcal{B}}) = P_y(X(T_{\mathcal{B} \cup \{z\}}) = z)$ for $z \notin \mathcal{B}$, it holds, using Formula (16),

$$\begin{aligned} G_{\mathcal{B}}(x, z) &= \frac{P_x(X(T_{\mathcal{B} \cup \{z\}}) = z)}{\sum_{y \neq z} [1 - P_y(X(T_{\mathcal{B} \cup \{z\}}) = z)] w(z, y)} \\ &= \frac{\frac{1}{Z_{\mathcal{B} \cup \{z\}}(0)} \sum_{\phi} w(\phi) \mathbb{1}_{\{\rho(\phi) = \mathcal{B} \cup \{z\}, \rho(\tau_x(\phi)) = z\}}}{\sum_{y \neq z} \frac{1}{Z_{\mathcal{B} \cup \{z\}}(0)} \sum_{\phi} w(\phi) \mathbb{1}_{\{\rho(\phi) = \mathcal{B} \cup \{z\}, \rho(\tau_y(\phi)) \neq z\}} w(z, y)} \\ &= \frac{\sum_{\phi} w(\phi) \mathbb{1}_{\{\rho(\phi) = \mathcal{B} \cup \{z\}, \rho(\tau_x(\phi)) = z\}}}{\sum_{\phi} w(\phi) \mathbb{1}_{\{\rho(\phi) = \mathcal{B} \cup \{z\}\}} \sum_{y \neq z} \mathbb{1}_{\{\rho(\tau_y(\phi)) \neq z\}} w(z, y)}. \end{aligned}$$

If we associate with each forest ϕ' such that $\rho(\phi') = \mathcal{B}$ the forest $\phi = \phi' \setminus \{(z, y)\}$ with (z, y) the only edge in ϕ' that is issued from z , then we have $\rho(\phi) = \mathcal{B} \cup \{z\}$, $\rho(\tau_y(\phi)) \neq z$, $w(\phi') = w(\phi)w(z, y)$ and we recognize $\sum_{\phi'} w(\phi') \mathbb{1}_{\{\rho(\phi') = \mathcal{B}\}} = Z_{\mathcal{B}}(0)$ in the last denominator. \square

3.2 Well-Distributed Roots

Proof of Theorem 1. We first note that for any $\mathcal{B} \subset \mathcal{X}$, it holds [recall Eq. (4)]

$$\mathbb{P}(\rho(\Phi_q) = \mathcal{B}) = \frac{Z_{\mathcal{B}}(0)q^{|\mathcal{B}|}}{Z(q)}$$

and, for $0 < m \leq n$,

$$\mathbb{P}(|\rho(\Phi_q)| = m) = \frac{a_m q^m}{Z(q)}. \tag{18}$$

Then, by using Formula (17),

$$\begin{aligned} \mathbb{E}\left[E_x[T_{\rho(\Phi_q)}] \mid |\rho(\Phi_q)| = m\right] &= \sum_{\mathcal{B} \subset \mathcal{X}: |\mathcal{B}|=m} \frac{Z_{\mathcal{B}}(0)q^m}{a_m q^m} \frac{1}{Z_{\mathcal{B}}(0)} \sum_{z \notin \mathcal{B}} \sum_{\substack{\phi: \rho(\phi) = \mathcal{B} \cup \{z\} \\ \rho(\tau_x(\phi)) = z}} w(\phi) \\ &= \frac{1}{a_m q^m} \sum_{\phi: |\rho(\phi)|=m+1} \frac{w(\phi)q^{|\rho(\phi)|}}{q} \sum_{z \in \rho(\phi)} \mathbb{1}_{\{\rho(\tau_x(\phi))=z\}} \\ &= \frac{a_{m+1}q^{m+1}}{a_m q^{m+1}} = \frac{a_{m+1}}{a_m}. \end{aligned}$$

By using again Eq. (18) and Proposition 2.1 with $S_q = S_{q, \emptyset}$, it follows

$$\begin{aligned} \mathbb{E}\left[E_x[T_{\rho(\Phi_q)}]\right] &= \sum_{m>0} \frac{a_{m+1}q^m}{Z(q)} = \frac{\sum_{k \geq 2} a_k q^k}{q Z(q)} = \frac{\mathbb{P}(S_q \geq 2)}{q} = \frac{1 - \mathbb{P}(S_q = 1)}{q} \\ &= \frac{1}{q} \left(1 - \prod_{j>0} \frac{\lambda_j}{q + \lambda_j} \right). \end{aligned}$$

□

We conclude this section by computing the mean return time

$$T_R^{++} = \inf\{t \geq 0 : \exists s > 0, X(s) \neq X(0), s \leq t, X(t) \in R\}$$

to $R = \rho(\Phi_q)$ starting from a uniformly distributed point in R . The reason why we use this heavy double $+$ notation is that we will also consider the maybe less natural but often more useful *randomized* or *skeleton return time* T_R^+ , which is defined as follows. Assuming that X is built by updating its current position at each time of a Poisson process of intensity α according to the probability kernel P defined by Eqs. (8)–(9), the skeleton return time is

$$T_R^+ = \inf\{t \geq \tau_1 : X(t) \in R\},$$

with τ_1 the first updating time in the Poisson process. One always has $T_R \leq T_R^+ \leq T_R^{++}$ and, for any $x \in R$, it holds

$$E_x[T_R^{++}] = E_x[T_R^+] + P_x(X(\tau_1) = x) E_x[T_R^{++}],$$

so that

$$E_x[T_R^{++}] = \frac{\alpha}{w(x)} E_x[T_R^+],$$

with $w(x)$ defined by Eq. (1). Like in the previous proof, we write S_q for $S_{q,\emptyset}$ defined in Eq. (6), and we stress that its law depends on the spectrum of L only.

Proposition 3.2 *For all $m \leq n$ and all $q > 0$, it holds*

$$\mathbb{E} \left[E_{U(\rho(\Phi_q))} \left[T_{\rho(\Phi_q)}^+ \right] \mid |\rho(\Phi_q)| = m \right] = \frac{n/m}{\alpha}$$

and

$$\mathbb{E} \left[E_{U(\rho(\Phi_q))} \left[T_{\rho(\Phi_q)}^{++} \right] \mid |\rho(\Phi_q)| = m \right] = \left(\frac{1}{n} \sum_{x \in \mathcal{X}} \frac{1}{w(x)} \right) \frac{n}{m},$$

so that

$$\begin{aligned} \mathbb{E} \left[E_{U(\rho(\Phi_q))} \left[T_{\rho(\Phi_q)}^+ \right] \right] &= \frac{1}{\alpha} \mathbb{E} \left[\frac{n}{S_q} \right] \quad \text{and} \quad \mathbb{E} \left[E_{U(\rho(\Phi_q))} \left[T_{\rho(\Phi_q)}^{++} \right] \right] \\ &= \left(\frac{1}{n} \sum_{x \in \mathcal{X}} \frac{1}{w(x)} \right) \mathbb{E} \left[\frac{n}{S_q} \right], \end{aligned}$$

where $U(\rho(\Phi_q))$ stands for a random point uniformly distributed in $\rho(\Phi_q)$.

Proof We work in discrete time, and we use the same notation as in the proof of Lemma 3.1. For any $\mathcal{B} \subset \mathcal{X}$ and $x \in \mathcal{X}$, we set

$$h_{\mathcal{B}}(x) = E_x[\hat{T}_{\mathcal{B}}] = \alpha E_x[T_{\mathcal{B}}].$$

When x belongs to \mathcal{B} , it holds $h_{\mathcal{B}}(x) = 0$ and, when $x \notin \mathcal{B}$,

$$\begin{aligned} h_{\mathcal{B}}(x) &= \sum_{y \in \mathcal{X}} P(x, y) E_x[\hat{T}_{\mathcal{B}} \mid \hat{X}(1) = y] = \sum_{y \in \mathcal{X}} P(x, y)(1 + h_{\mathcal{B}}(y)) \\ &= 1 + (Ph_{\mathcal{B}})(x). \end{aligned}$$

Setting

$$g_{\mathcal{B}}(x) = \mathbb{1}_{\{x \in \mathcal{B}\}} E_x[\hat{T}_{\mathcal{B}}^+] = \mathbb{1}_{\{x \in \mathcal{B}\}} \alpha E_x[T_{\mathcal{B}}^+] = \mathbb{1}_{\{x \in \mathcal{B}\}} w(x) E_x[T_{\mathcal{B}}^{++}], \tag{19}$$

we also have, when $x \in \mathcal{B}$,

$$g_{\mathcal{B}}(x) = 1 + (Ph_{\mathcal{B}})(x).$$

Let us denote by ν any probability measure on the subsets of \mathcal{X} that produces well-distributed points. By setting $h(x) = \sum_{\mathcal{B} \subset \mathcal{X}} \nu(\mathcal{B}) h_{\mathcal{B}}(x)$ for all $x \in \mathcal{X}$, we then get

$$\begin{aligned}
 (Ph)(x) &= \sum_{\mathcal{B} \subset \mathcal{X}} \nu(\mathcal{B})(Ph_{\mathcal{B}})(x) = \sum_{\mathcal{B} \ni x} \nu(\mathcal{B})(h_{\mathcal{B}} - 1)(x) + \sum_{\mathcal{B} \ni x} \nu(\mathcal{B})(g_{\mathcal{B}} - 1)(x) \\
 &= - \sum_{\mathcal{B} \subset \mathcal{X}} \nu(\mathcal{B}) + \sum_{\mathcal{B} \ni x} \nu(\mathcal{B})h_{\mathcal{B}}(x) + \sum_{\mathcal{B} \ni x} \nu(\mathcal{B})g_{\mathcal{B}}(x) \\
 &= -1 + h(x) + \sum_{\mathcal{B} \subset \mathcal{X}} \nu(\mathcal{B})g_{\mathcal{B}}(x),
 \end{aligned}$$

where we used $\sum_{\mathcal{B}} \nu(\mathcal{B}) = 1$, $h_{\mathcal{B}}(x) = 0$ if $x \in \mathcal{B}$, and $g_{\mathcal{B}}(x) = 0$ if $x \notin \mathcal{B}$. Since ν produces well-distributed point, the function h is constant on \mathcal{X} , so that $(Ph)(x) = h(x)$ for all x in \mathcal{X} , which implies, together with the previous equality,

$$\sum_{\mathcal{B} \subset \mathcal{X}} \nu(\mathcal{B})g_{\mathcal{B}}(x) = 1, \quad x \in \mathcal{X}. \tag{20}$$

Since by Theorem 1 we can choose for ν the distribution of $\rho(\Phi_q)$ conditioned on $\{|\rho(\Phi_q)| = m\}$, this proves, together with Eq. (19), after dividing by αm or $w(x)m$ and summing on $x \in \mathcal{X}$, the first part of the proposition. The last two equalities simply follow from Proposition 2.1.

Since a function h is harmonic on \mathcal{X} if and only if it is constant, the previous arguments actually show that any distribution ν on the subsets \mathcal{B} of \mathcal{X} provides well-distributed points if and only if it satisfies Eq. (20), which is actually a list of $n = |\mathcal{X}|$ equations the $\nu(\mathcal{B})$ s have to satisfy, together with the positivity constraints $\nu(\mathcal{B}) \geq 0$ for all $\mathcal{B} \subset \mathcal{X}$ and the additional equation $\sum_{\mathcal{B}} \nu(\mathcal{B}) = 1$. If we restrict ourselves to distributions supported by sets of a fixed size m , these are $n + 1$ linear equations for $\binom{n}{m}$ unknown variables. Since $\binom{n}{m} > n + 1$ for $2 \leq m \leq n - 2$ and Theorem 1 provides a solution ν with positive mass $\nu(\mathcal{B}) > 0$ for each subset \mathcal{B} of size m , this shows that there are infinitely many solutions in this case. When $m = 1$ or $m = n - 1$, we have more equations than variables. If $m = 1$, then solving Eq. (20) is straightforward and we have a unique solution. If $m = n - 1$, then it is more convenient to solve directly the equation set

$$\sum_{|\mathcal{B}|=n-1} \nu(\mathcal{B}) = 1, \quad \sum_{|\mathcal{B}|=n-1} \nu(\mathcal{B})h_{\mathcal{B}}(x) = t, \quad x \in \mathcal{X},$$

with the additional unknown variable t , to see that the solution is unique.

4 Re-reading Micchelli and Willoughby’s Proof

Before following in Sects. 4.1–4.3 the three main steps of Micchelli and Willoughby’s proof, we give some heuristic on the divided difference representation of the $\nu_k = \nu_k^x$ in the case $\nu = \delta_x$ for any $x \in \mathcal{X} \setminus \mathcal{B}$. (In the general case, ν is a convex combination of such Dirac masses and we just need to prove the theorem in this special case of a generic Dirac mass.) For $0 \leq k < l$, we have

$$v_k^x = \delta_x \prod_{j>k} \frac{[L]_{\mathcal{X} \setminus \mathcal{B}} + \lambda_{j,\mathcal{B}}}{\lambda_{j,\mathcal{B}}},$$

and, if $\mathcal{B} \neq \emptyset$, the same formula gives $v_{-1}^x = 0$ by Hamilton–Cayley theorem, while, if $\mathcal{B} = \emptyset$, $\lambda_{0,\mathcal{B}} = 0$. Having in mind that each v_k should be interpreted as a local equilibrium from which the system decays into v_{k-1} at rate λ_k , this means that the system leaves the state v_0 only to be absorbed in \mathcal{B} when $\mathcal{B} \neq \emptyset$ (the probability measure on the quotient set \mathcal{X}/\mathcal{B} that is associated with v_{-1} is fully concentrated on \mathcal{B}), while v_0 is the non-decaying global equilibrium μ when $\mathcal{B} = \emptyset$. Now, by considering Wilson’s algorithm with a first loop-erased random walk started from x to sample $\Phi_{q,\mathcal{B}}$ and by comparing the successive decay times with the exponential random time T_q , we get for any y in $\mathcal{X} \setminus \mathcal{B}$, using the notation of Proposition 2.2 and the notation recalled before Eq. (16),

$$\begin{aligned} \mathbb{P}(\rho(\tau_x(\Phi_q)) = y) &= P_x(X(T_q \wedge T_{\mathcal{B}}) = y) = K_{q,\mathcal{B}}(x, y) \\ &= \frac{q}{q + \lambda_{l-1,\mathcal{B}}} v_{l-1}^x(y) + \frac{\lambda_{l-1,\mathcal{B}}}{q + \lambda_{l-1,\mathcal{B}}} \frac{q}{q + \lambda_{l-2,\mathcal{B}}} v_{l-2}^x(y) \\ &+ \dots + \frac{\lambda_{l-1,\mathcal{B}}}{q + \lambda_{l-1,\mathcal{B}}} \dots \frac{\lambda_{2,\mathcal{B}}}{q + \lambda_{2,\mathcal{B}}} \frac{q}{q + \lambda_{1,\mathcal{B}}} v_1^x(y) \\ &+ \frac{\lambda_{l-1,\mathcal{B}}}{q + \lambda_{l-1,\mathcal{B}}} \dots \frac{\lambda_{1,\mathcal{B}}}{q + \lambda_{1,\mathcal{B}}} \frac{q}{q + \lambda_{0,\mathcal{B}}} v_0^x(y). \end{aligned} \tag{21}$$

Let us divide by q and multiply by $Z_{\mathcal{B}}(q) = \det(q - [L]_{\mathcal{X} \setminus \mathcal{B}})$ both sides of this equation to have a simpler polynomial right-hand side. With $W_{\mathcal{B}}(q)$ the matrix defined by

$$W_{\mathcal{B}}(q)(x, y) = \frac{Z_{\mathcal{B}}(q)}{q} \mathbb{P}(\rho(\tau_x(\Phi_q)) = y) = \frac{1}{q} \sum_{\substack{\phi: \rho(\tau_x(\phi))=y, \\ \rho(\phi) \supset \mathcal{B}}} w(\phi) q^{|\rho(\phi) \setminus \mathcal{B}|}, \quad x, y \in \mathcal{X} \setminus \mathcal{B}, \tag{22}$$

or equivalently, since we have seen in the proof of Proposition 2.2 that $K_{q,\mathcal{B}} = q[q - L]_{\mathcal{X} \setminus \mathcal{B}}^{-1}$,

$$W_{\mathcal{B}}(q) = Z_{\mathcal{B}}(q)[q - L]_{\mathcal{X} \setminus \mathcal{B}}^{-1}, \tag{23}$$

Equation (21) now reads

$$\begin{aligned} W_{\mathcal{B}}(q)(x, y) &= (q + \lambda_{0,\mathcal{B}}) \dots (q + \lambda_{l-2,\mathcal{B}}) v_{l-1}^x(y) + (q + \lambda_{0,\mathcal{B}}) \dots (q + \lambda_{l-3,\mathcal{B}}) \lambda_{l-1,\mathcal{B}} v_{l-2}^x(y) \\ &+ \dots + (q + \lambda_{0,\mathcal{B}}) \lambda_{l-1,\mathcal{B}} \lambda_{l-2,\mathcal{B}} \dots \lambda_{2,\mathcal{B}} v_1^x(y) + \lambda_{l-1,\mathcal{B}} \dots \lambda_{1,\mathcal{B}} v_0^x(y). \end{aligned} \tag{24}$$

and suggests a divided difference representation for the v_k s according to the following definition.

Definition 4.1 For any function f defined on \mathbb{R} and with values in a real vector space, if x_0, x_1, \dots, x_{l-1} are distinct real numbers, the divided differences $f[x_0]$,

$f[x_0, x_1], \dots, f[x_0, \dots, x_{l-1}]$ are the coefficients of the unique polynomial Q of degree less than l such that

$$Q(x) = f[x_0] + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) + \dots + f[x_0, \dots, x_{l-1}](x - x_0) \cdots (x - x_{l-2})$$

for all x in \mathbb{R} and $Q(x_k) = f(x_k)$ for $k < l$.

Remark These divided differences $f[x_0, x_1, \dots, x_k]$ are given by Lagrange interpolation formula

$$f[x_0, x_1, \dots, x_k] = \sum_{j=0}^k \frac{f(x_j)}{\prod_{i \leq k, i \neq j} (x_j - x_i)}, \quad k < l, \tag{25}$$

which shows that divided differences are permutation invariant, and one can also compute them inductively with Newton interpolation formulas

$$f[x_0] = f(x_0),$$

$$f[x_0, \dots, x_k] = \frac{f[x_1, \dots, x_k] - f[x_0, \dots, x_{k-1}]}{x_k - x_0}, \tag{26}$$

which explain the terminology. See, for example, Chapter II of [15] for more details.

To prove that our v_k^x s are non-negative measures, we can assume without loss of generality that the eigenvalues $\lambda_{k,B}$ are all distinct (we can modify them slightly and use the continuity of the v_k^x s). Equation (24) suggests in this case that for all $k < l$ and x in \mathcal{X}

$$W_{\mathcal{B}}[-\lambda_{0,B}, \dots, -\lambda_{k,B}](x, \cdot) = \lambda_{l-1,B} \cdots \lambda_{k+1,B} v_k^x = \delta_x \prod_{j>k} ([L]_{\mathcal{X} \setminus \mathcal{B}} + \lambda_{j,B}),$$

i.e.

$$v_k^x = \frac{W_{\mathcal{B}}[-\lambda_{0,B}, \dots, -\lambda_{k,B}](x, \cdot)}{\lambda_{l-1,B} \cdots \lambda_{k+1,B}}(x, \cdot), \tag{27}$$

or, equivalently,

$$W_{\mathcal{B}}[-\lambda_{0,B}, \dots, -\lambda_{k,B}] = \prod_{j>k} ([L]_{\mathcal{X} \setminus \mathcal{B}} + \lambda_{j,B}). \tag{28}$$

It is worth noting at this point that Eq. (28) would be a *consequence* of the theorem by Micchelli and Willoughby (the local equilibrium interpretation of each v_k makes sense only once its non-negativity is established), but our goal is to *prove* this theorem. This is what we are ready to do now.

4.1 Checking Eq. (28)

We simply check that the right-hand side and the left-hand side of Eq. (28) act in the same way on each left-eigenvector μ_r of $[L]\mathcal{X}\setminus\mathcal{B}$, with $\mu_r[L]\mathcal{X}\setminus\mathcal{B} = -\lambda_{r,\mathcal{B}}\mu_r$. Recall that $W_{\mathcal{B}}(q)$ is equivalently defined by Eq. (22) or Eq. (23). By using the latter and Formula (25), it holds, for each $r < l$ and $k < l$,

$$\begin{aligned} \mu_r W_{\mathcal{B}}[-\lambda_{0,\mathcal{B}}, \dots, -\lambda_{k,\mathcal{B}}] &= \sum_{j=0}^k \frac{\mu_r W_{\mathcal{B}}(-\lambda_{j,\mathcal{B}})}{\prod_{i \leq k, i \neq j} (-\lambda_{j,\mathcal{B}} + \lambda_{i,\mathcal{B}})} \\ &= \sum_{j=0}^k \frac{\prod_{i < l, i \neq r} (-\lambda_{j,\mathcal{B}} + \lambda_{i,\mathcal{B}})}{\prod_{i \leq k, i \neq j} (-\lambda_{j,\mathcal{B}} + \lambda_{i,\mathcal{B}})} \mu_r. \end{aligned}$$

Since each numerator in the right-hand side is equal to 0 unless $j = r$, which can happen only if $r \leq k$, we get

$$\begin{aligned} \mu_r W_{\mathcal{B}}[-\lambda_{0,\mathcal{B}}, \dots, -\lambda_{k,\mathcal{B}}] &= \mathbb{1}_{\{r \leq k\}} \frac{\prod_{i < l, i \neq r} (-\lambda_{r,\mathcal{B}} + \lambda_{i,\mathcal{B}})}{\prod_{i \leq k, i \neq r} (-\lambda_{r,\mathcal{B}} + \lambda_{i,\mathcal{B}})} \mu_r \\ &= \mathbb{1}_{\{r \leq k\}} \left(\prod_{i > k} (-\lambda_{r,\mathcal{B}} + \lambda_{i,\mathcal{B}}) \right) \mu_r \\ &= \left(\prod_{i > k} (-\lambda_{r,\mathcal{B}} + \lambda_{i,\mathcal{B}}) \right) \mu_r = \mu_r \prod_{j > k} ([L]\mathcal{X}\setminus\mathcal{B} + \lambda_{j,\mathcal{B}}). \end{aligned}$$

4.2 A Combinatorial Identity

The key point of the proof lies in the following lemma.

Lemma 4.2 *For any $x \neq y$ in $\mathcal{X}\setminus\mathcal{B}$, it holds*

$$W_{\mathcal{B}}(q)(x, x) = Z_{\mathcal{B}\cup\{x\}}(q) \tag{29}$$

and

$$\begin{aligned} W_{\mathcal{B}}(q)(x, y) &= w(x, y) Z_{\mathcal{B}\cup\{x,y\}}(q) \\ &\quad + \sum_{z, z' \in \mathcal{X}\setminus(\mathcal{B}\cup\{x,y\})} w(x, z) W_{\mathcal{B}\cup\{x,y\}}(q)(z, z') w(z', y). \end{aligned} \tag{30}$$

Proof Equation (22) can be rewritten as

$$W_{\mathcal{B}}(q)(x, y) = \sum_{\substack{\phi: \rho(\tau_x(\phi))=y, \\ \rho(\phi) \supset \mathcal{B}}} w(\phi) q^{|\rho(\phi)|-1-|\mathcal{B}|}, \tag{31}$$

for $x \neq y$ we have

$$Z_{\mathcal{B} \cup \{x,y\}}(q) = \sum_{\phi': \rho(\phi') \supset \mathcal{B} \cup \{x,y\}} w(\phi') q^{|\rho(\phi')| - 2 - |\mathcal{B}|}, \tag{32}$$

and it holds

$$\begin{aligned} W_{\mathcal{B} \cup \{x,y\}}(q)(z, z') &= \frac{1}{q} \sum_{\substack{\phi'': \rho(\tau_z(\phi'')) = z', \\ \rho(\phi'') \supset \mathcal{B} \cup \{x,y\}}} w(\phi'') q^{|\rho(\phi'')| - 2 - |\mathcal{B}|} \\ &= \sum_{\substack{\phi'': \rho(\tau_z(\phi'')) = z', \\ \rho(\phi'') \supset \mathcal{B} \cup \{x,y\}}} w(\phi'') q^{|\rho(\phi'')| - 3 - |\mathcal{B}|}. \end{aligned} \tag{33}$$

Now we can define, for each ϕ appearing in Eq. (31), $\phi' = \phi \setminus \{(x, y)\}$ if (x, y) belongs to ϕ , and $\phi'' = \phi \setminus \{(x, z), (z', y)\}$ if x is connected in ϕ to y through $(x, z) \in \phi$ and $(z', y) \in \phi$, possibly with $z = z'$. Since $|\rho(\phi')| = |\rho(\phi)| + 1$ and $|\rho(\phi'')| = |\rho(\phi)| + 2$, Eq. (30) follows from Eqs. (31)–(33). Equation (29) is given by Eq. (31) with y in place of x . □

4.3 Conclusion with Cauchy Interlacement Theorem

We will use the following lemma from [11] and for which we give an alternative proof.

Lemma [Micchelli and Willoughby] *Let $f : x \in \mathbb{R} \mapsto \prod_{j < l} (x - \alpha_j) \in \mathbb{R}$ be a polynomial of degree l with l distinct zeros $\alpha_0 > \alpha_1 > \dots > \alpha_{l-1}$. Let $\beta_0 > \beta_1 > \dots > \beta_{L-1}$ be $L \geq l$ real numbers such that $\beta_j \geq \alpha_j$ for each $j < l$. Then, for any $k \leq L$, $f[\beta_0, \beta_1, \dots, \beta_k] \geq 0$.*

Proof We prove the lemma by induction on $r = l - k$. First, since f is a polynomial of degree l with a dominant coefficient equal to 1, Definition 4.1 gives $f[\beta_0, \dots, \beta_k] = 0$ if $k > l$ —that is $r < 0$ — $f[\beta_0, \dots, \beta_k] = 1$ if $k = l$ —that is $r = 0$ —and the claim is established for $r \leq 0$.

Now, for $r > 0$, we first show that $f[\beta_0, \alpha_1, \dots, \alpha_k] \geq 0$. In the case $\beta_0 = \alpha_0$, this is obvious by Formula (25), which gives $f[\alpha_0, \dots, \alpha_k] = 0$, while, if $\beta_0 > \alpha_0$, we have by permutation invariance, Formula (26), and induction hypothesis,

$$\begin{aligned} f[\beta_0, \alpha_1, \dots, \alpha_k] &= (\beta_0 - \alpha_0) \frac{f[\beta_0, \alpha_1, \dots, \alpha_k] - f[\alpha_0, \alpha_1, \dots, \alpha_k]}{\beta_0 - \alpha_0} \\ &= (\beta_0 - \alpha_0) \frac{f[\beta_0, \alpha_1, \dots, \alpha_k] - f[\alpha_1, \dots, \alpha_k, \alpha_0]}{\beta_0 - \alpha_0} \\ &= (\beta_0 - \alpha_0) f[\beta_0, \alpha_1, \dots, \alpha_k, \alpha_0] = (\beta_0 - \alpha_0) f[\beta_0, \alpha_0, \dots, \alpha_k] \\ &\geq 0. \end{aligned}$$

We then show that $f[\beta_0, \beta_1, \alpha_2, \dots, \alpha_k] \geq 0$. In the case $\beta_1 = \alpha_1$, this is what we have just proved, while, if $\beta_1 > \alpha_1$, we have in the same way

$$\begin{aligned} f[\beta_0, \beta_1, \alpha_2, \dots, \alpha_k] &= f[\beta_0, \alpha_1, \dots, \alpha_k] + (\beta_1 - \alpha_1) \\ &\quad \frac{f[\beta_0, \beta_1, \alpha_2, \dots, \alpha_k] - f[\beta_0, \alpha_1, \alpha_2, \dots, \alpha_k]}{\beta_1 - \alpha_1} \\ &= f[\beta_0, \alpha_1, \dots, \alpha_k] + (\beta_1 - \alpha_1)f[\beta_0, \beta_1, \alpha_1, \dots, \alpha_k] \geq 0. \end{aligned}$$

Proceeding similarly we eventually get to $f[\beta_0, \beta_1, \dots, \beta_k] \geq 0$.

The theorem is eventually proven by showing by induction on $l = |\mathcal{X} \setminus \mathcal{B}|$ the stronger statement (recall Formula (27)):

$$W_{\mathcal{B}}[\xi_0, \dots, \xi_k](x, y) \geq 0, \quad k \leq L, \quad x, y \in \mathcal{X} \setminus \mathcal{B},$$

for all $\xi_0 > \xi_1 > \dots > \xi_{L-1}$ with $L \geq l$ and such that $\xi_j \geq -\lambda_{j, \mathcal{B}}$ for $j < l$. The claim is obvious for $l = 1$, and we distinguish to cases for $l > 1$. If $x = y$, we do not need the inductive hypothesis: it follows from Formula (29) that

$$W_{\mathcal{B}}[\xi_0, \dots, \xi_k](x, x) = Z_{\mathcal{B} \cup \{x\}}[\xi_0, \dots, \xi_k],$$

by Cauchy interlacement theorem—this is where the reversibility hypothesis matters— $\xi_j \geq \lambda_{j, \mathcal{B}}$ implies $\xi_j \geq \lambda_{j, \mathcal{B} \cup \{x\}}$ and the non-negativity follows from the lemma. If $x \neq y$, the claim follows in the same way from Formula (30) and the inductive hypothesis.

Acknowledgements Part of this work was done during two long visits by A.G. to Leiden, supported by the ERC Advanced Grant 267356-VARIS of Frank den Hollander. L.A. was supported by NWO Gravitation Grant 024.002.003-NETWORKS.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Anantharam, V., Tsoucas, P.: A proof of the Markov chain tree theorem. *Stat. Probab. Lett.* **8**(2), 189–192 (1989)
2. Avena, L., Castell, F., Gaudillière, A., Mélot, C.: Approximate and exact solutions of intertwining equations through random spanning forests (2017). [arXiv:1702.05992](https://arxiv.org/abs/1702.05992)
3. Benoist, S., Dumaz, L., Werner, W.: Near critical spanning forests and renormalization (2015). [arXiv:1503.08093](https://arxiv.org/abs/1503.08093)
4. Burton, R., Pemantle, R.: Local characteristics, entropy and limit theorems for spanning trees and domino tilings via transfer-impedances. *Ann. Probab.* **21**, 1329–1371 (1993)
5. Chang, Y.: Contribution à l'étude des laçets Markoviens. thèse de doctorat de l'université Paris Sud—Paris XI, tel-00846462 (2013)
6. Diaconis, P., Fulton, W.: A growth model, a game, an algebra, Lagrange inversion, and characteristic classes. *Rend. Semin. Mat. Univ. Pol. Torino* **49**(1), 95–119 (1991)

7. Fill, J.: On hitting times and fastest strong stationary times for skip-free and more general chains. *J. Theor. Probab.* **22**(3), 558–586 (2009)
8. Freidlin, M.I., Wentzell, A.D.: *Random Perturbations of Dynamical Systems*, Grundlehren der Mathematischen Wissenschaften 260, 2nd edn. Springer, New York (1998)
9. Levin, D.A., Peres, Y., Wilmer, E.L.: *Markov Chains and Mixing Times*. American Mathematical Society, Providence (2008)
10. Marchal, P.: Loop-erased random walks, spanning trees and Hamiltonian cycles. *Electron. Commun. Probab.* **5**, 39–50 (2000)
11. Micchelli, C.A., Willoughby, R.A.: On functions which preserve the class of Stieltjes matrices. *Linear Algebra Appl.* **23**, 141–156 (1979)
12. Miclo, L.: On absorption times and Dirichlet eigenvalues. *ESAIM Probab. Stat.* **14**, 117–150 (2010)
13. Propp, J., Wilson, D.: How to get a perfectly random sample from a generic Markov chain and generate a random spanning tree of a directed graph. *J. Algorithms* **27**, 170–217 (1998)
14. Wilson, D.: Generating random spanning trees more quickly than the cover time. In: *Proceedings of the Twenty-Eight Annual ACM Symposium on the Theory of Computing*, pp. 296–303 (1996)
15. Whittaker, E.T., Robinson, G.: *The Calculus of Observations; A Treatise on Numerical Mathematics*. Blackie and Son Limited, London (1924)