Probabilistic graph inspections through forests

Koperberg, V.T.

CHAPTER 4

Coupled Kirchhoff forests: dynamic loop-ensemble, Laplacian spectrum & DGFF

This chapter is based on joint work in progress with L. Avena and A. Gaudillière.

4.1 Introduction

This chapter is a continuation of chapter 3. As in chapter 3 we consider the *coupled forest process*, defined in section 3.2, that is constructed by coupling together a continuum of realizations of Wilson's algorithm. Where in chapter 3 the observable of interest was the associated occupation field, in the current chapter the focus is on the configurations of colored cycles that are 'popped' (i.e. deleted) during Wilson's procedure. In this chapter, three results will be presented that are related to these cycle configurations.

(1) *Poissonian loop-ensemble.* In [55] Le Jan introduced a Poisson point process on the set of closed walk, called the Poissonian loop-ensemble, and showed how such a loop-ensemble can be obtained from a single application of Wilson's algorithm by using Poisson-Dirichlet random variables to decompose the closed walks generated by the algorithm.

The first result of this chapter, theorem 4.1, shows that using the dynamic coupled version of Wilson's algorithm circumvents the need for Poisson-Dirichlet decompositions, allowing for the construction of the loop-ensemble directly from the cycle popping procedure, without requiring any additional randomness.

An important role in this construction is being filled by a bijection between the set of closed walks and a subset of 'rooted' configurations of colored cycles. This bijection, which is defined in section 4.5.1 below, allows us to associate a closed walk to each time of the coupled forest process at which new cycles are popped. Incidentally, in proving theorem 4.1 we will provide an alternative proof of theorem 3.6, which is more probabilistic in flavor than the generating function based proof provided in the previous chapter.

(2) *Spectral decomposition.* The second result, provided in theorem 4.2 and its direct corollary corollary 4.2.1, shows that under some assumptions on the Laplacian, the occupation field process admits a spectral decomposition, in the sense that its distribution can be written as a mixture of which each term corresponds to a single Laplacian eigenvalue.

Recently, in [9], the coupled forest process was used to devise an estimation procedure for the Laplacian spectrum of a graph. The procedure devised there utilizes an observable obtained solely from the forests, that is distributed as a sum of random variables, that are each associated with one of the Laplacian eigenvalues. In contrast, the spectral decomposition provided in theorem 4.2 shows that, by considering the cycle configurations produced by Wilson's algorithm rather than the forests, we can construct observables that have a mixture distribution, with each component of the mixture depending on a single eigenvalue. As mixtures are better suited than sums for estimating the parameters of the components, theorem 4.2 could possibly be used to improve on the estimation procedure in [9].

(3) *Gaussian free field with mass.* Our third result concerns the relation between the occupation field of Wilson's algorithm and the Gaussian free field. Lupu exhibited in [59] a remarkable coupling between the discrete Gaussian free field and the Poissonian loop-ensemble. This coupling fits very well with the dynamic framework of the coupled

forest process. Adjusting Lupu's coupling to this framework, allows us to couple together in a continuum of Gaussian free fields, parametrized by their distinct masses, as is shown in proposition 4.3.

4.1.1 Colored cycle configurations

In this section we start by providing a definition of the random colored cycle configurations produced by Wilson's algorithm. As this chapter is a continuation of chapter 3, the setting in the this chapter will be identical to the setting presented in section 3.1.1, which therefore will not be repeated here.

Recall from section 3.1.2.4 that for fixed $t > 0$ cycle popping of the DF-stacks $\{(A_i(x), B_i^t(x))_{i\in\mathbb{N}_0} : \ x \in \mathcal{X}\}$ produces a sequence of 'popped' cycles $(\Gamma_0^\circ, \Gamma_1^\circ, \dots, \Gamma_{i^*-1}^\circ)$, where, conditionally on the DF-stacks, the exact sequence that is produced depends on some arbitrary choice of ordering on the set of cycles. This dependence of the sequence of cycles on an arbitrary ordering is inconvenient. Therefore, we will instead consider the set of colored cycles produced by cycle popping, since, as was shown by Wilson [77, Thm. 4], the set of colored cycles produced is the same for any choice of ordering of the cycles used in the cycle popping procedure.

A *colored cycle*[1] is a pair (γ°, c), where γ° is a cycle and $c : s(\gamma^\circ) \to \mathbb{N}_0$ is a map that assigns an integer 'color' to each vertex in the cycle.

By giving each vertex in a popped cycle Γ_i° a color value equal the number of deleted arrows from its stack, we see that the cycle popping procedure produces a set of colored cycles $\{(\Gamma_i^\circ, c_i) : \ i \in [i^* - 1]_0\}$, with the coloring map $c_i : s(\gamma^\circ) \to \mathbb{N}_0$ given by $c_i(x) := \mathbf{d}_i(x)$, where we recall from section 3.1.2.4 that $\mathbf{d}_i(x)$ denotes the number of deleted arrows. So, an arrow's color in a colored cycle corresponds to its 'level' in the DF-stack.

For $t \geq 0$ we denote by $\mathfrak{C}_t$ the set of colored cycles produced by cycle popping the DF-stacks $\{(A_i(x), B_i^t(x))_{i\in\mathbb{N}_0} : \ x \in \mathcal{X}\}$. We note that $\mathfrak{C}_s \subseteq \mathfrak{C}_t$ for $s < t$, since as time progresses an increasing amount of cycles are popped from the DF-stacks. Recalling from section 3.2.1 that $(N_t)_{t\geq 0}$ denotes the occupation field process, we note that the occupation field N_t can be expressed using the colored cycles $\mathfrak{C}_t$ as

$$N_t = \underline{1} + \sum_{(\gamma^\circ, c)\in\mathfrak{C}_t} \mathbf{1}_{s(\gamma^\circ)}. \tag{4.1}$$

That is, N_t is one more than the sum, taken over all cycles occurring in $\mathfrak{C}_t$, of the indicators of their support.

[1] This terminology was introduced by Wilson and is adopted here. We assume no confusion arises due to the red and green colors of the arrows introduced in section 3.1.2.4.

4.2 Constructing the Poissonian loop-ensemble

The Poissonian loop-ensemble is associated to a given sub-Markovian transition matrix. Here we give the definition of the loop-ensemble for the specific sub-Markovian transition matrix $(I - \frac{1}{1/t+\delta}L)$ of the killed random walk $(X_k)_{0\le k\le T_t}$ introduced in section 3.1.1. We remark that, in principle, the coupled forest process can be constructed using other choices of random walks, and can accommodate inhomogeneous killing rates among vertices as well. Hence, our specific choice of random walk should not be seen as a restriction.

As was done in section 3.1.2.1, we denote by $\mathcal{P}^{\mathrm{cl}}$ the set of closed walks.

Definition 4.2.1. For $\alpha > 0$ a Poisson point process on $\mathcal{P}^{\mathrm{cl}}$ with intensity measure

$$\{\gamma\} \mapsto \frac{\alpha}{l}\frac{w(e(\gamma))}{(1/t+\delta)^l}, \qquad \text{for all } \gamma = (x_0, x_1, \dots, x_l) \in \mathcal{P}^{\mathrm{cl}} \tag{4.2}$$

is called a *loop-ensemble* of intensity α with killing rate $1/t$. ■

This definition differs from that of Le Jan, in that it omits the infinitely many length 0 walks (trivial loops) that occur in the loop-ensemble according to the definition by Le Jan.

The theorem below introduces a dynamic generalization $\mathcal{L}$ of the loop-ensemble, and shows that it can be constructed with the Kirchhoff forest coupling introduced in section 3.2.

Theorem 4.1. *There exists a Poisson point process $\mathcal{L}$ on the product space $\mathcal{P}^{\mathrm{cl}} \times (0,\infty)$ such that for any $t > 0$ the random atomic measure on $\mathcal{P}^{\mathrm{cl}}$ defined by*

$$\{\gamma\} \mapsto \mathcal{L}(\{\gamma\} \times (0,t]) \tag{4.3}$$

is a loop-ensemble of intensity 1 with killing rate $1/t$, that is measurable with respect to the σ-algebra generated by $(\mathfrak{C}_s, \Phi_{1/s})_{0\le s\le t}$.

The measurabiltity statement in theorem 4.1 can be made explicit. The process $\mathcal{L}$ is obtained from $(\mathfrak{C}_s, \Phi_{1/s})_{0\le s\le t}$ by applying a specific bijection, defined in lemma 4.4, between cycle configurations and closed walks, as will be detailed in the proof of theorem 4.1 below.

4.3 Spectral decomposition

Recall from eq. (3.28) the definition of the set of jump times

$$\mathcal{T} := \{t \in (0,\infty) \colon \quad N_t \neq \lim_{s\uparrow t} N_s\}. \tag{4.4}$$

The jump rate κ_t in eq. (3.27) can be written as a sum with each term depending only on a single eigenvalue of L,

$$\kappa_t := \mathrm{Tr}\left[\tfrac{1}{t}(K_{1/t} - \tfrac{1}{1+\delta t}I)\right] = \sum_{j<n} \frac{\delta - \lambda_j}{(1+\delta t)(1+\lambda_j t)}. \tag{4.5}$$

This formulation of κ_t suggest that $\mathcal{T}$ can be interpreted as a union of independent Poisson point processes where each Poisson point process is associated to an eigenvalue.

We require two additional assumptions on the Laplacian matrix L and the random walk parameter δ.

ASSUMPTION 1a:

The random walk X is irreducible and reversible. (AS1a)

From the irreducibility assumption follows that X has a unique stationary distribution, denoted by μ. The assumption (AS1a) on the Laplacian further ensures that the Laplacian spectrum is real valued, and non-negative. The maximal Laplacian eigenvalue is denoted by $\lambda_{\max}$. Moreover, with this assumption there exists a basis $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{n-1}$ of $\mathbb{R}^{\mathcal{X}}$ consisting of eigenvectors of L that are orthonormal with respect to the inner product $\langle f, g\rangle_\mu := \sum_{x\in\mathcal{X}} f(x)g(x)\mu(x)$. Naturally, these eigenvectors are indexed in such a manner that $\mathbf{v}_j$ has corresponding eigenvalue λ_j.

Our second assumption adds a restriction to the parameter δ.

ASSUMPTION 2:

The random walk parameter satisfies $\delta \geq \lambda_{\max}$. (AS2)

By Gershgorin's circle theorem it holds that $\lambda_{\max} \leq \max_{x\in\mathcal{X}} 2L(x,x)$, so that a sufficient condition for assumption (AS2) to hold is that $\delta \geq \max_{x\in\mathcal{X}} 2L(x,x)$.

The two assumptions (AS1a) and (AS2) are chosen to ensure that for all $j < n$ the individual terms $\frac{\delta-\lambda_j}{(1+\delta t)(1+\lambda_j t)}$ of the Poisson rate function in eq. (4.5) are real valued and non-negative. This allows us to define for all $j < n$ and $t \in [0,\infty)$ the random geometric killing times T_j^t, the random vertices Y_j, and for all $x \in \mathcal{X}$, $l \in \mathbb{N}$ the random closed walks $\Gamma_{l,x}$, such that all these random variables are independent with joint law P given by

$$\begin{aligned} P(\Gamma_{l,x} = \gamma) &:= \mathbf{P}_x((X_k)_{0\leq k\leq l} = \gamma \mid X_l = x) = \frac{w(e(\gamma))}{\delta^l(I - \frac{1}{\delta}L)^l(x,x)}, \\ P(T_j^t = k) &:= \frac{1+\lambda_j t}{1+\delta t}\left(1 - \frac{1+\lambda_j t}{1+\delta t}\right)^k, \qquad P(Y_j = x) := \mathbf{v}_j(x)^2\mu(x), \end{aligned} \tag{4.6}$$

where l denotes the length of γ.

Theorem 4.2. *The Poisson point process $\mathcal{L}$ on $\mathcal{P}^{\mathrm{cl}} \times (0,\infty)$ has intensity measure*

$$\{\gamma\} \times (a,b] \mapsto \sum_{j<n} \int_a^b \tfrac{\delta-\lambda_j}{(1+\delta t)(1+\lambda_j t)}\; P(Y_j = x_0)P(T_j^t + 1 = l)P(\Gamma_{l,x_0} = \gamma)\; \mathrm{d}t,$$

with $\gamma = (x_0, x_1, \ldots, x_l) \in \mathcal{P}^{\mathrm{cl}}$.

So, at each jump time the closed walk that is generated by the cycle popping procedure by applying the bijection of lemma 4.4, has the same law as a closed walk obtained as follows. First sample a random eigenvalue with probability proportional to $\frac{\delta-\lambda_j}{(1+\delta t)(1+\lambda_j t)}$, then conditionally on the eigenvalue sample independently a random starting point and a random length, and finally sample a closed walk with the given starting point and length.

A remarkable aspect of theorem 4.2 is that, although the increment of the running time $\sum_{x\in\mathcal{X}}\ell[\mathfrak{C}_t^{\Delta}](x)$ and the waking root V_t are themselves not independent of each other, by conditioning on the 'artificial' random eigenvalue their counterparts T_j^t and Y_j do become independent.

As a consequence of theorem 4.2, we obtain, in corollary 4.2.1 below, a rephrasing of the result of the previous chapter.

On the same probability space where we defined the random variables Y_j and T_j^t of eq. (4.6), we further define for all $j < n$ Poisson point processes Ψ_j with intensity measures

$$\mu_j((a,b]) := \int_a^b \frac{\delta-\lambda_j}{(1+\delta t)(1+\lambda_j t)}\,\mathrm{d}t, \tag{4.7}$$

and for all $t > 0$ we let $\Gamma_{l,x,t}$ be copies of $\Gamma_{l,x}$, such that all these random variables are independent.

Corollary 4.2.1. *The distributions of the running time process and of the occupation field process are, respectively, given by*

$$(M_t)_{t\geq 0} \overset{d}{=} \left(1+\sum_{j<n}\int_0^t (T_j^s+1)\ \Psi_j(\mathrm{d}s)\right)_{t\geq 0} \tag{4.8}$$

and

$$(N_t)_{t\geq 0} \overset{d}{=} \left(\underline{1}+\sum_{j<n}\int_0^t \ell[\Gamma^-_{T_j^s+1,Y_j,s}]\ \Psi_j(\mathrm{d}s)\right)_{t\geq 0}. \tag{4.9}$$

4.4 Coupling DGFFs of different masses

In this section we can drop the assumption (AS2) above. Instead, we require a strengthening of assumption (AS1a).

ASSUMPTION 1b:

The Laplacian matrix L is symmetric. (AS1b)

Assumption (AS1b) ensures that the Gaussian free field, defined in definition 4.4.1 below, is well-defined.

Definition 4.4.1 (Discrete Gaussian free field with mass). For $q > 0$, an n-dimensional centered Gaussian random variable ϕ with covariance matrix

$$\Sigma_q^2 := \tfrac{1}{2}(q+\delta)(qI+L)^{-1} \tag{4.10}$$

is called a *discrete Gaussian free field* (abbrv. DGFF) on $\mathcal{G}$ with mass q. ■

The above definition is unconventional, as customarily the scaling factor $\frac{1}{2}(q+\delta)$ is omitted from the defining covariance matrix. However, the use of this scaling factor is better suited to our purposes.

As was shown by Le Jan in [56], the occupation field of Wilson's algorithm is related to the Gaussian free field. To explore this connection, we add exponential waiting times to the random walks used in Wilson's algorithm.

To the DF-stacks used to define the coupled forest process we add stacks of exponential waiting times, thus obtaining the collection $\{(A_i(x), U_i(x), \eta_i(x))_{i\in\mathbb{N}_0} :\ x \in \mathcal{X}\}$ where for each $k \in \mathbb{N}_0$, $x \in \mathcal{X}$ the random variables $\eta_k(x)$ are Exp(1) distributed and independent of each other and of all $A_i(x)$ and $U_i(x)$. We define the *continuous occupation field process* $(\Xi_t)_{t\geq 0}$ as the $(0,\infty)^{\mathcal{X}}$-valued process

$$\Xi_t(x) := \sum_{k=0}^{N_t(x)-1} \eta_k(x). \tag{4.11}$$

The scaled continuous occupation field $\frac{t}{1+\delta t}\Xi_t$ corresponds to the occupation field of Wilson's algorithm at intensity $1/t$, when we use continuous-time random walks with infinitesimal generator $-L$, killed at random exponential times of rate $1/t$. While this scaled field might be a more natural object than Ξ_t, the process $(\Xi_t)_{t\geq 0}$ has the convenient property of having piece-wise constant trajectories.

From definition (4.11) it follows that $\Xi_0(x) = \eta_0(x)$. By the Box-Muller transform, we can assume for every $x \in \mathcal{X}$ that $\eta_0(x)$ is coupled to two independent standard Gaussians $Z(x)$ and $\tilde{Z}(x)$ such that $\eta_0(x) = \frac{1}{2}Z^2(x) + \frac{1}{2}\tilde{Z}^2(x)$, where we write $Z^2 := (Z(x)^2)_{x\in\mathcal{X}}$. Thus allowing us to decompose Ξ_0 as a sum of squares two i.i.d Gaussians.

For each $t > 0$ let $Y_t \sim \mathrm{Ber}(\frac{1}{2})$ be Bernoulli random variables, that are independent of each other and of the DF-stacks $\{(A_i(x), U_i(x), \eta_i(x))_{i\in\mathbb{N}_0} :\ x \in \mathcal{X}\}$, and denote by $\Xi_t^\Delta := \Xi_t - \lim_{s\uparrow t}\Xi_s$ the increment of the continuous occupation field at time t.

Denote by $\Psi := \sum_{\tau\in\mathcal{T}}\delta_\tau$ the random atomic measure supported on $\mathcal{T}$, the set of jump times of the occupation field process.

We then define a *thinned occupation field process* $(\xi_t)_{t\geq 0}$ by

$$\xi_t := \sqrt{\tfrac{1}{2}Z^2 + \int_0^t \Xi_s^\Delta Y_s \ \Psi(\mathrm{d}s)}. \tag{4.12}$$

We remark that this construction provides us with a second process $(\tilde{\xi}_t)_{t\geq 0}$, that is defined by

$$\tilde{\xi}_t := \sqrt{\tfrac{1}{2}\tilde{Z}^2 + \int_0^t \Xi_s^\Delta (1-Y_s) \ \Psi(\mathrm{d}s)}. \tag{4.13}$$

Since the processes $(\xi_t)_{t\geq 0}$ and $(\tilde{\xi}_t)_{t\geq 0}$ are obtained by thinning the Poisson point process Ψ, they are independent of each other.

Le Jan showed [56, Thm. 2] that the thinned occupation field is distributed as the absolute value of a DGFF with mass $1/t$, i.e.

$$\xi_t \overset{d}{=} |\phi|_t, \tag{4.14}$$

where ϕ_t is a DGFF with mass $1/t$, and $|\phi|_t$ denotes its term-wise absolute value, $|\phi|_t(x) := |\phi_t(x)|$ for all $x \in \mathcal{X}$.

In addition to the thinned occupation field $(\xi_t)_{t\geq 0}$, we further require for each *undirected* edge $e = \{x, y\}$ independent uniform random variables $U_e \sim \mathrm{Unif}(0, 1)$. We recall that L is symmetric, so that the set of undirected edges is given by $\bar{\mathcal{E}} := \{\{x, y\} \in \binom{\mathcal{X}}{2} \colon\ (x, y) \in \mathcal{E}\}$. The variable U_e is used to couple together for all $t > 0$ the Bernoulli random variables O_e^t defined by

$$O_e^t := \mathbf{1}\left\{U_e > \exp\left(-\tfrac{2tw(x,y)}{1+\delta t}\sqrt{\xi_t(x)\xi_t(y)}\right)\right\}.$$

These Bernoulli variables were introduced in Lupu's construction of the DGFF [59, Thm. 1], and are adapted here to the dynamic setting by coupling them together for distinct t.

Recall that $\xi_0 := \frac{1}{\sqrt{2}}|Z|$, where Z is a vector of i.i.d. standard Gaussians, and write $S_0 := \mathrm{sign}(Z)$, i.e. $S_0(x) := \mathrm{sign}(Z(x))$ for all $x \in \mathcal{X}$.

Define the filtration $(\mathcal{F}_t)_{t\geq 0}$ by $\mathcal{F}_t := \sigma(\xi_s, (O_e^s)_{e\in\bar{\mathcal{E}}}, S_0 \colon\ s \leq t)$.

Proposition 4.3. *There exists an $\mathbb{R}^{\mathcal{X}}$-valued stochastic process $(\phi_t)_{t\geq 0}$ with the following properties:*

- *(i) for each $t > 0$ the marginal ϕ_t is a DGFF with mass $1/t$;*
- *(ii) trajectories of $(\phi_t)_{t\geq 0}$ are piece-wise constant and càdlàg;*
- *(iii) the process $(\phi_t)_{t\geq 0}$ satisfies the Markov property;*
- *(iv) the same sign components, i.e. connected components of the spanning subgraph with edges $\{(x, y) \in \mathcal{E} \colon\ \mathrm{sign}(\phi_t(x)) = \mathrm{sign}(\phi_t(y))\}$, are unions of coalescing clusters;*
- *(v) for each $x \in \mathcal{X}$ the map $t \mapsto |\phi_t(x)|$ is non-decreasing;*
- *(vi) the process $(\phi_t)_{t\geq 0}$ is adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$.*

Proof. For each $t \geq 0$, we will set the absolute value of ϕ_t equal to ξ_t, hence we only require a procedure to determine $\mathrm{sign}(\phi_t)$.

For each $t \geq 0$ we define a spanning subgraph $\mathcal{G}_t = (\mathcal{X}, \mathcal{E}_t)$, whose edges are given by

$$\mathcal{E}_t := \bigcup_{(\gamma^\circ,\, c)\in\mathfrak{C}_t} e(\gamma^\circ) \cup \{(x, y) \in \mathcal{E} \colon\ O^t_{\{x,y\}} = 1\},$$

and denote by $\mathcal{W}_t$ the set of connected components of $\mathcal{G}_t$. Note that the process $(\mathcal{W}_t)_{t\geq 0}$ is a process of coalescing partitions, such that at time $t = 0$ the starting

partition $\mathcal{W}_0$ consists of isolated vertices.

We let $\mathcal{S} := \{t \in (0,\infty)\colon\ \mathcal{W}_t \neq \lim_{s\uparrow t} \mathcal{W}_s\}$ be the set of jump times of the process $(\mathcal{W}_t)_{t\geq 0}$. Write $\tau_0 := 0$ and $\mathcal{S} = \{\tau_1, \tau_2, \tau_3, \ldots\}$ with $\tau_i < \tau_{i+1}$ for all $i \in \mathbb{N}_0$.

For each $\tau_i \in \mathcal{S}$ we will construct a $\{-1,1\}^{\mathcal{X}}$-valued sign vector S_{τ_i} at time τ_i, that assigns the same sign to vertices belonging to the same block of the partition $\mathcal{W}_{\tau_i}$. For this we require some arbitrary decision procedure that can be used to iteratively determine S_{τ_i}, given the partition $\mathcal{W}_{\tau_i}$, the previous signs $S_{\tau_{i-1}}$, and the absolute values $\xi_{\tau_{i-1}}$. An example of such a procedure would be, to assign to a block in $\mathcal{W}_{\tau_i}$ that is obtained by the coalescing of several blocks of $\mathcal{W}_{\tau_{i-1}}$, the sign of the largest of these coalescing blocks.

Denoting by $\Pi(\mathcal{X})$ the set of partitions of $\mathcal{X}$, we can represent this decision procedure by a map $g : \Pi(\mathcal{X}) \times (0,\infty)^{\mathcal{X}} \times \{-1,1\}^{\mathcal{X}} \times (0,\infty) \to \{-1,1\}^{\mathcal{X}}$ such that $g(\pi, \mathbf{x}, \sigma, t)(x) = g(\pi, \mathbf{x}, \sigma, t)(y)$ for any x, y that belong to the same block of the partition π. We then set

$$S_{\tau_i} := g(\mathcal{W}_{\tau_i}, \xi_{\tau_{i-1}}, S_{\tau_{i-1}}, \tau_i),$$

where we recall that $S_{\tau_0} = \operatorname{sign}(Z)$.

Having constructed the signs S_{τ_i} at all times in $\mathcal{S}$, we construct S_t for any $t > 0$ by setting

$$S_t := S_{\tau_{i_t}}, \qquad \text{where } i_t := \max\{i \in \mathbb{N}_0\colon\ \tau_{i_t} \leq t\}.$$

We then define the process $(\phi_t)_{t\geq 0}$ by

$$\phi_t(x) := S_t(x)\xi_t(x).$$

That ϕ_t is a DGFF with mass $1/t$ follows directly from Lupu's coupling [59, Thm. 1]. Properties $(ii) - (vi)$ follow by construction. □

4.5 Proofs

4.5.1 Bijection between closed walks and popped cycles

We call a finite set of colored cycles *admissible* if it could be produced by cycle popping some realization of the DF-stacks. Equivalently, a set C of colored cycles is admissible if it holds that:

- (i) for all distinct $(\gamma_1^\circ, c_1), (\gamma_2^\circ, c_2) \in C$ either $c_1(x) < c_2(x)$ for all $x \in s(\gamma_1^\circ) \cap s(\gamma_2^\circ)$ or $c_1(x) > c_2(x)$ for all $x \in s(\gamma_1^\circ) \cap s(\gamma_2^\circ)$;
- (ii) for all $(\gamma^\circ, c) \in C$ and any $x \in s(\gamma^\circ)$ with $c(x) \geq 1$, there exists $(\tilde{\gamma}^\circ, \tilde{c}) \in C$ with $x \in s(\tilde{\gamma}^\circ)$ and $\tilde{c}(x) = c(x) - 1$;
- (iii) if for any $k \in \mathbb{N}_{\geq 3}$ and $(\gamma_1^\circ, c_1), \ldots, (\gamma_k^\circ, c_k) \in C$ it holds that $c_i(x) < c_{i+1}(x)$ for all $i \in [k-1]$ and all $x \in s(\gamma_i^\circ) \cap s(\gamma_{i+1}^\circ)$, then it holds that $c_1(x) < c_k(x)$ for all $x \in s(\gamma_1^\circ) \cap s(\gamma_k^\circ)$.

The set of all admissible sets of colored cycles is denoted by $\mathcal{A}$.

Wilson observed that for a fixed multi-set of cycles there is a bijection between its admissible colorings and the partial orderings on that multi-set. From an admissible set of colored cycles C we obtain a partial ordering $\preccurlyeq$ by writing $\tilde{\gamma}^\circ \preccurlyeq \gamma^\circ$ for any $(\gamma^\circ, c), (\tilde{\gamma}^\circ, \tilde{c}) \in C$ if (γ°, c) has to be popped before $(\tilde{\gamma}^\circ, \tilde{c})$ can be popped during the cycle popping procedure. That is, if there exist $k \in \mathbb{N}_0$ and $(\gamma_0^\circ, c_0), \dots, (\gamma_k^\circ, c_k) \in C$ with $(\gamma_0^\circ, c_0) = (\gamma^\circ, c)$ and $(\gamma_k^\circ, c_k) = (\tilde{\gamma}^\circ, \tilde{c})$ such that for all $i \in [k]$ it holds that $s(\gamma_{i-1}^\circ) \cap s(\gamma_i^\circ) \neq \varnothing$ and $c_{i-1}(x) < c_i(x)$ for all $x \in s(\gamma_{i-1}^\circ) \cap s(\gamma_i^\circ)$.

The inverse of the above bijection is as follows. For each cycle γ° in partially ordered multi-set of cycles let its coloring map c be such that $c(x)$ equals the number of cycles with x in their support that are smaller than γ°.

So, cycles that are colored using larger values will be smaller in the resulting partial order. This reflects that cycles with large colors appear lower in the DF-stack.

We call an admissible set of colored cycles C a *cycle clump* if it contains a unique monochromatic cycle with color 0, i.e. if

$$|\{(\gamma^\circ, c) \in C\colon\ c(x) = 0 \text{ for all } x \in s(\gamma^\circ)\}| = 1. \tag{4.15}$$

The cycle with color 0 we denote by $\gamma_{\max}^\circ$. This notation refers to the fact that the partially ordered multi-set of cycles associated to a cycle clump has $\gamma_{\max}^\circ$ as its maximum. A pair (C, v) consisting of a cycle clump C and a vertex $v \in \mathcal{X}$ is called a *rooted cycle clump* if $v \in s(\gamma_{\max}^\circ)$.

Lemma 4.4. *There exists an explicit bijection f from the set of rooted cycle clumps to the set of closed walks $\mathcal{P}^{\mathrm{cl}}$ such that for any rooted cycle clump (C, v) it holds for $f(C, v) = (x_0, x_1, \dots, x_l)$ that*

(i) $x_0 = x_l = v$;

(ii) $\sum_{(\gamma^\circ, c) \in C} \mathbf{1}_{e(\gamma^\circ)} = \sum_{i \in [l]} \mathbf{1}_{\{(x_{i-1}, x_i)\}}$.

That is, v is the starting point of the closed walk $f(C, v)$, and the edges contained in the cycle clump C are the same as the edges traversed by the closed walk.

Proof. We use the above defined partial ordering to define the inverse $\overleftrightarrow{C}$ of an admissible set of colored cycles C. If $\preccurlyeq$ is the partial ordering associated with C, then we let $\overleftrightarrow{C}$ be the admissible set of colored cycles associated with the reversed partial ordering $\overleftrightarrow{\preccurlyeq}$ given by $\gamma_1^\circ \overleftrightarrow{\preccurlyeq} \gamma_2^\circ$ iff $\gamma_2^\circ \preccurlyeq \gamma_1^\circ$.

Although less insightful, it is possible to give an equivalent definition of $\overleftrightarrow{C}$ without referencing the partial order. For any $x \in \cup_{(\gamma^\circ, c) \in C} s(\gamma^\circ)$ let $\check{c}(x) := \max\{c(x)\colon\ (\gamma^\circ, c) \in C\}$ denote the maximal color used for x by a coloring in C. Then, denoting $\overleftrightarrow{c}(x) := \check{c}(x) - c(x)$ for each coloring c, the inverse of C is the set

$$\overleftrightarrow{C} := \{(\gamma^\circ, \overleftrightarrow{c})\colon\ (\gamma^\circ, c) \in C\}.$$

Given a cycle clump C we define the (deterministic) DF-stacks of colored arrows[2] $\{(a_i(x), b_i(x))_{i \in \mathbb{N}_0}\colon\ x \in \mathcal{X}\}$ as follows. If there exists a colored cycle $(\gamma^\circ, \overleftrightarrow{c}) \in \overleftrightarrow{C}$

[2]The arrows $a_i(x)$ are colored either red or green, as indicated by $b_i(x)$, see section 3.1.2.4.

with $\overleftrightarrow{c}(x) = i$, then we set $b_i(x) = 0$ and $a_i(x) = y_1$, where $(y_0, y_1, \dots, y_l) \in \gamma^\circ$ is the unique representative of γ° with $y_0 = x$. Otherwise, if no such colored cycle exists in $\overleftrightarrow{C}$, then we set $b_i(x) = 1$ and $a_i(x) = x$.

So, all green arrows in the stacks correspond to edges that are traversed by cycles in $\overleftrightarrow{C}$.

As is explained in section 3.1.2.4, we can obtain a finite length walk $f(C, v) := (x_0, x_1, \dots, x_l)$ from the vertex r and these DF-stacks, by setting $x_0 := r$ and for $k \in \mathbb{N}_0$ iteratively setting

$$x_{k+1} := a_{i_k}(x_k), \quad \text{with } i_k := \ell[(x_j)_{0 \le j \le k-1}](x_k), \tag{4.16}$$

and $l = \min\{k \in \mathbb{N}_0 \colon \ b_{i_k}(x_k) = 1\}$.

For a closed walk $\gamma = (x_0, \dots, x_l) \in \mathcal{P}^{\mathrm{cl}}$ the inverse $f^{-1}(\gamma)$ can be described using the loop-erasure procedure, and the sequence of self-avoiding walks $(\gamma_i)_{0 \le i \le l}$ produced by this procedure, see section 3.1.2.1.

For each $i \in [l]$ with $x_i \in s(\gamma_{i-1})$, we define the cycle γ_i° to be the equivalence class of the closed walk $(y_{k_i}, y_{k_i+1}, \dots, y_m, x_i)$, which is the cycle that is erased by the loop-erasure procedure in iteration i. The coloring $c_i : s(\gamma_i^\circ) \to \mathbb{N}_0$ is given by $c_i(x) := \sum_{j<i} \mathbf{1}_{s(\gamma_j^\circ)}(x)$, i.e. $c_i(x)$ is equal to the number of times x occurs in the support of the previous cycles. We can then iteratively define the collection of colored cycles C_{i+1} by setting $C_0 = \varnothing$ and

$$C_i := \begin{cases} C_{i-1} & \text{if } x_i \notin s(\gamma_{i-1}) \\ C_{i-1} \cup \{(\gamma_i^\circ,\, c_i)\} & \text{if } x_i \in s(\gamma_{i-1}). \end{cases} \tag{4.17}$$

It then holds that $f^{-1}(\gamma) = \left(\overleftrightarrow{C}_l, x_0\right)$.

By construction it holds that $f^{-1}(f(C, v)) = (C, v)$, which concludes the proof. □

4.5.2 Constructing the Poissonian loop-ensemble

For a set of colored cycles C, we introduce its upward color shift $C_\uparrow$, which is the set of colored cycles defined as

$$C_\uparrow := \{(\gamma^\circ,\, c - \hat{c}|_{s(\gamma^\circ)}) \colon \ (\gamma^\circ,\, c) \in C\}, \tag{4.18}$$

where $\hat{c} : \cup_{(\gamma^\circ,\, c) \in C} \ s(\gamma^\circ) \to \mathbb{N}_0$ is given by $\hat{c}(x) = \min\{c(x) \colon \ (\gamma^\circ,\, c) \in C\}$, and $\hat{c}|_{s(\gamma^\circ)}$ denotes its restriction to the support of γ°.

Note that if C is admissible, then it holds for its up-shift that $C_\uparrow = C$.

Rather than considering the total set of colored cycles produced until time t, we are concerned with the colored cycles produced at a single time point

$$\mathfrak{C}_t^\Delta := \mathfrak{C}_t \setminus \bigcup_{s<t} \mathfrak{C}_s. \tag{4.19}$$

While $\mathfrak{C}_t^\Delta$ is itself not necessarily admissible, its up-shift $(\mathfrak{C}_t^\Delta)_\uparrow$ is. We remark that for the up-shift of $\mathfrak{C}_t^\Delta$ it holds that

$$(\mathfrak{C}_t^\Delta)_\uparrow = \{(\gamma^\circ,\, c - (N_t - \underline{1})|_{s(\gamma^\circ)})\colon\ (\gamma^\circ,\, c) \in \mathfrak{C}_t^\Delta\}.$$

For all jump times $\tau \in \mathcal{T}$ the DF-stacks $\{(A_i(x), U_i(x))_{i\in\mathbb{N}_0}\colon\ x \in \mathcal{X}\}$ can be used to define a vertex V_τ, which will be called the *waking root* at time τ. The waking root denotes the root of the coupled forest process $(\Phi_{1/t})_{t\geq 0}$ at which an arrow in the DF-stack changes its dynamic color from red to green at time τ, i.e. V_τ is the unique vertex v such that $\lim_{s\uparrow\tau} U_{N_s(v)-1}(v) = \frac{1}{1+\delta\tau}$.

Importantly, the waking root V_τ is measurable with respect to the σ-algebra generated by $(\mathfrak{C}_s, \Phi_{1/s})_{0\leq s\leq\tau}$, since V_τ is the only root of the forest $\Phi_{\frac{1}{\tau-\mathrm{d}t}}$ just before time τ at which the occupation field increases. In fact, as will become clear from the definition in eq. (4.22), the Poisson point process $\mathcal{L}$ of theorem 4.1 is measurable with respect to the σ-algebra generated by both $(\mathfrak{C}_t)_{t\geq 0}$ and the collection of waking roots.

Definition 4.5.1. The *cycle clump process* $\mathcal{C}$ is the random atomic measure on the product space $\mathcal{A}\times\mathcal{X}\times(0,\infty)$ defined by

$$\mathcal{C}(\{C\}\times\{x\}\times(a,b]) := |\{\tau\in\mathcal{T}\cap(a,b]\colon\ (\mathfrak{C}_\tau^\Delta)_\uparrow = C,\ V_\tau = x\}|. \tag{4.20}$$

■

An important observation is that the random atomic measure on $\mathcal{A}\times\mathcal{X}$ given by

$$\{C\}\times\{x\}\mapsto \mathcal{C}(\{C\}\times\{x\}\times(0,\infty)) \tag{4.21}$$

is supported on the set of rooted cycle clumps. Together, the above observation and the bijection f from lemma 4.4 allow us to define a random atomic measure $\mathcal{L}$ on $\mathcal{P}^{\mathrm{cl}}\times(0,\infty)$ by

$$\mathcal{L}(\{\gamma\}\times(a,b]) := \mathcal{C}(\{f^{-1}(\gamma)\}\times(a,b]). \tag{4.22}$$

As the notation suggests, $\mathcal{L}$ is indeed the sought Poisson point process of theorem 4.1, as will become clear in the remainder of this section.

For a collection of colored cycles C write $l(C) := \sum_{(\gamma^\circ,\, c)\in C} |e(\gamma^\circ)|$ to denote the total number of traversed edges by all cycles in C.

Lemma 4.5. *The the cycle clump process $\mathcal{C}$ is a Poisson point process on $\mathcal{A}\times\mathcal{X}\times(0,\infty)$ with intensity measure*

$$\{C\}\times\{v\}\times(a,b]\mapsto \mathbf{1}\{(C,v)\in \mathrm{RCC}\}\int_a^b \tfrac{1}{(1+\delta t)^2}\left(\tfrac{t}{1+\delta t}\right)^{l(C)-1}\prod_{(\gamma^\circ,\, c)\in C} w(e(\gamma^\circ))\ \mathrm{d}t,$$

where RCC *denotes the set of rooted cycle clumps.*

Proof. For any admissible set of colored cycles $C\in\mathcal{A}$ and any rooted forest F, it holds by Wilson's formula that

$$\mathbb{P}(\mathfrak{C}_t = C,\ \Phi_{1/t} = F) = \left(\tfrac{1}{1+\delta t}\right)^{r(F)}\left(\tfrac{t}{1+\delta t}\right)^{|F|+l(C)} w(F)\prod_{(\gamma^\circ,\, c)\in C} w(e(\gamma^\circ)), \tag{4.23}$$

where we recall that $r(F)$ and $|F|$ denote the number of roots and edges of F, respectively. In eq. (4.24), we extend this formula to the dynamic setting with DF-stacks $\{(A_i(x), U_i(x))_{i\in\mathbb{N}_0}\colon\ x \in \mathcal{X}\}$ that contain uniform killing marks. For an admissible set of colored cycles C and $x \in \mathcal{X}$ we write

$$c^*_C(x) := \begin{cases} 1 + \max\{c(x)\colon\ (\gamma^\circ,\, c) \in C,\ x \in s(\gamma^\circ)\} & \text{if } x \in \bigcup_{(\gamma^\circ,\, c)\in C} s(\gamma^\circ) \\ 0 & \text{otherwise.} \end{cases}$$

to denote the smallest color that is not used for vertex x by a colored cycle in C.

Let $\mathbf{u}^* : \mathcal{X} \to [0,1)$ be such that $\mathbf{u}^*(x) < \frac{1}{1+\delta t}$ for all $x \in \rho(F)$ and $\mathbf{u}^*(x) \geq \frac{1}{1+\delta t}$ for all $x \notin \rho(F)$. Write $k := |C|$ and $C = \{(\gamma_1^\circ, c_1), (\gamma_2^\circ, c_2), \dots, (\gamma_k^\circ, c_k)\}$, and for all $i \in [k]$ let $\mathbf{u}_i : s(\gamma_1^\circ) \to [\frac{1}{1+\delta t}, 1)$ be given. It holds that

$$\begin{aligned} &\mathbb{P}\left(\bigcap_{i\in[k]} \bigcap_{x\in s(\gamma_i^\circ)} U_{c_i(x)}(x) \in \mathrm{d}\mathbf{u}_i(x),\ \bigcap_{x\in\mathcal{X}} U_{c^*_C(x)}(x) \in \mathrm{d}\mathbf{u}^*(x) \,\middle|\, \mathfrak{C}_t = C,\ \Phi_{1/t} = F \right) \\ &= \left(\prod_{i\in[k]} \prod_{x\in s(\gamma_i^\circ)} \tfrac{1+\delta t}{\delta t}\ \mathrm{d}\mathbf{u}_i(x) \right) \left(\prod_{x\notin\rho(F)} \tfrac{1+\delta t}{\delta t}\ \mathrm{d}\mathbf{u}^*(x) \right) \left(\prod_{x\in\rho(F)} (1+\delta t)\ \mathrm{d}\mathbf{u}^*(x) \right), \end{aligned} \tag{4.24}$$

where, by abuse of notation, $\mathrm{d}\mathbf{u}_i(x)$ denotes both a small enough real number and the interval $(\mathbf{u}_i(x), \mathbf{u}_i(x) + \mathrm{d}\mathbf{u}_i(x)]$.[3]

We will use eq. (4.24) to compute the probability

$$\mathbb{P}\left(\mathcal{C}\left(\{C'\} \times \{v\} \times \mathrm{d}t\right) = 1 \,\middle|\, \mathfrak{C}_{t+\mathrm{d}t} = C \cup C'_{\downarrow C},\ \Phi_{\frac{1}{t+\mathrm{d}t}} = F \right), \tag{4.25}$$

where (C', v) is a rooted cycle clump, and we denote by

$$C'_{\downarrow C} := \{(\gamma^\circ,\, c + c^*_C)\colon\ (\gamma^\circ,\, c) \in C'\} \tag{4.26}$$

the set of colored cycles obtained by downwards shifting the colors of cycles in C' to fit underneath the colored cycles in C.

Conditionally on the event that both $\mathfrak{C}_{t+\mathrm{d}t} = C \cup C'_{\downarrow C}$ and $\Phi_{\frac{1}{t+\mathrm{d}t}} = F$, the event that both $\mathfrak{C}_t = C$ and $\mathcal{C}\left(\{C'\} \times \{v\} \times \mathrm{d}t\right) = 1$ can be expressed in terms of the uniform killing marks as the intersection of the following events:

(1) for all $(\gamma^\circ,\, c) \in C$ and all $x \in s(\gamma^\circ)$ it holds that $U_{c(x)}(x) \geq \frac{1}{1+\delta t}$;

(2) $U_{c^*_C(v)}(v) \in [\frac{1}{1+\delta(t+\mathrm{d}t)}, \frac{1}{1+\delta t})$;

(3) for all $x \in s(\gamma^\circ_{\max}) \setminus v$ it holds that $U_{c^*_C(x)}(x) \geq \frac{1}{1+\delta t}$

(4a) for all $(\gamma^\circ,\, c) \in C'_{\downarrow C} \setminus \{(\gamma^\circ_{\max}, c^*_C|_{s(\gamma^\circ_{\max})})\}$ and all $x \in s(\gamma^\circ)$ it holds that $U_{c(x)}(x) > U_{c^*_C(v)}(v)$,

[3] In this context, 'small enough' means that $\mathbf{u}_i(x) + \mathrm{d}\mathbf{u}_i(x) \leq 1$ for $i \in [k]$, $\mathbf{u}^*(x) + \mathrm{d}\mathbf{u}^*(x) \leq 1$ for $x \notin \rho(F)$, and $\mathbf{u}^*(x) + \mathrm{d}\mathbf{u}^*(x) < \frac{1}{1+\delta t}$ for $x \in \rho(F)$.

where we recall that $\gamma^\circ_{\max}$ denotes the maximal cycle in the cycle clump C', and $c^*_C|_{s(\gamma^\circ_{\max})}$ denotes the restriction of c^*_C to the support of γ°, which is the coloring of the maximal colored cycle in $C'_{\downarrow C}$. That is,

$$\begin{aligned}&\mathbb{P}\left(\mathfrak{C}_t = C,\ \mathcal{C}\left(\{C'\}\times\{v\}\times \mathrm{d}t\right)=1 \;\middle|\; \mathfrak{C}_{t+\mathrm{d}t} = C\cup C'_{\downarrow C},\ \Phi_{\frac{1}{t+\mathrm{d}t}} = F\right)\\ &= \mathbb{P}\left((1)\cap(2)\cap(3)\cap(4a) \;\middle|\; \mathfrak{C}_{t+\mathrm{d}t} = C\cup C'_{\downarrow C},\ \Phi_{\frac{1}{t+\mathrm{d}t}} = F\right).\end{aligned} \tag{4.27}$$

By defining the event

(4*b*) for all $(\gamma^\circ,\, c)\in C'_{\downarrow C}\backslash\{(\gamma^\circ_{\max}, c^*_C|_{s(\gamma^\circ_{\max})})\}$ and all $x\in s(\gamma^\circ)$ it holds that $U_{c(x)}(x)\geq \frac{1}{1+\delta t}$,

in the limit as $\mathrm{d}t\downarrow 0$ we have by eq. (4.24) that

$$\begin{aligned}&\mathbb{P}\left(\mathfrak{C}_t = C,\ \mathcal{C}\left(\{C'\}\times\{v\}\times \mathrm{d}t\right)=1 \;\middle|\; \mathfrak{C}_{t+\mathrm{d}t} = C\cup C'_{\downarrow C},\ \Phi_{\frac{1}{t+\mathrm{d}t}} = F\right)\\ &\geq \mathbb{P}\left((1)\cap(2)\cap(3)\cap(4b) \;\middle|\; \mathfrak{C}_{t+\mathrm{d}t} = C\cup C'_{\downarrow C},\ \Phi_{\frac{1}{t+\mathrm{d}t}} = F\right)\\ &= \frac{1+\delta(t+\mathrm{d}t)}{\delta(t+\mathrm{d}t)}\left(\frac{1}{1+\delta t}-\frac{1}{1+\delta(t+\mathrm{d}t)}\right)\left(\frac{1+\delta(t+\mathrm{d}t)}{\delta(t+\mathrm{d}t)}\frac{\delta t}{1+\delta t}\right)^{l(C\cup C'_{\downarrow C})-1}\\ &= \frac{1}{1+\delta t}\mathrm{d}t + o(\mathrm{d}t)\end{aligned}$$

and that

$$\begin{aligned}&\mathbb{P}\left(\mathfrak{C}_t = C,\ \mathcal{C}\left(\{C'\}\times\{v\}\times \mathrm{d}t\right)=1 \;\middle|\; \mathfrak{C}_{t+\mathrm{d}t} = C\cup C'_{\downarrow C},\ \Phi_{\frac{1}{t+\mathrm{d}t}} = F\right)\\ &\leq \mathbb{P}\left((2) \;\middle|\; \mathfrak{C}_{t+\mathrm{d}t} = C\cup C'_{\downarrow C},\ \Phi_{\frac{1}{t+\mathrm{d}t}} = F\right) = \frac{1+\delta(t+\mathrm{d}t)}{\delta(t+\mathrm{d}t)}\left(\frac{1}{1+\delta t}-\frac{1}{1+\delta(t+\mathrm{d}t)}\right)\\ &= \frac{1}{1+\delta t}\mathrm{d}t + o(\mathrm{d}t).\end{aligned}$$

Upper and lower bounding the probability in eq. (4.25) by

$$\begin{aligned}&\mathbb{P}\left(\mathfrak{C}_t = C,\ \mathcal{C}\left(\{C'\}\times\{v\}\times \mathrm{d}t\right)=1 \;\middle|\; \mathfrak{C}_{t+\mathrm{d}t} = C\cup C'_{\downarrow C},\ \Phi_{\frac{1}{t+\mathrm{d}t}} = F\right)\\ &\leq \mathbb{P}\left(\mathcal{C}\left(\{C'\}\times\{v\}\times \mathrm{d}t\right)=1 \;\middle|\; \mathfrak{C}_{t+\mathrm{d}t} = C\cup C'_{\downarrow C},\ \Phi_{\frac{1}{t+\mathrm{d}t}} = F\right)\\ &\leq \mathbb{P}\left(\mathfrak{C}_t = C,\ \mathcal{C}\left(\{C'\}\times\{v\}\times \mathrm{d}t\right)=1 \;\middle|\; \mathfrak{C}_{t+\mathrm{d}t} = C\cup C'_{\downarrow C},\ \Phi_{\frac{1}{t+\mathrm{d}t}} = F\right)\\ &\quad+\mathbb{P}\left(\mathcal{C}\left(\mathcal{A}\times\mathcal{X}\times \mathrm{d}t\right)\geq 2 \;\middle|\; \mathfrak{C}_{t+\mathrm{d}t} = C\cup C'_{\downarrow C},\ \Phi_{\frac{1}{t+\mathrm{d}t}} = F\right)\\ &= \mathbb{P}\left(\mathfrak{C}_t = C,\ \mathcal{C}\left(\{C'\}\times\{v\}\times \mathrm{d}t\right)=1 \;\middle|\; \mathfrak{C}_{t+\mathrm{d}t} = C\cup C'_{\downarrow C},\ \Phi_{\frac{1}{t+\mathrm{d}t}} = F\right)+o(\mathrm{d}t),\end{aligned} \tag{4.28}$$

gives us that

$$\mathbb{P}\left(\mathcal{C}\left(\{C'\}\times\{v\}\times \mathrm{d}t\right)=1 \;\middle|\; \mathfrak{C}_{t+\mathrm{d}t} = C\cup C'_{\downarrow C},\ \Phi_{\frac{1}{t+\mathrm{d}t}} = F\right) = \frac{1}{1+\delta t}\mathrm{d}t + o(\mathrm{d}t).$$

Therefore, by eq. (4.23), summing over all admissible sets of colored cycles and all rooted forests gives us that

$$\begin{aligned}
&\mathbb{P}\left(\mathcal{C}\left(\{C'\}\times\{v\}\times \mathrm{d}t\right)=1\right)\\
&=\sum_{C}\sum_{F}\mathbb{P}\left(\mathcal{C}\left(\{C'\}\times\{v\}\times \mathrm{d}t\right)=1,\ \mathfrak{C}_{t+\mathrm{d}t}=C\cup C'_{\downarrow C},\ \Phi_{\frac{1}{t+\mathrm{d}t}}=F\right)\\
&=\left(\frac{\mathrm{d}t}{1+\delta t}+o(\mathrm{d}t)\right)\sum_{C}\sum_{F}\mathbb{P}(\mathfrak{C}_{t+\mathrm{d}t}=C\cup C'_{\downarrow C},\ \Phi_{\frac{1}{t+\mathrm{d}t}}=F)\\
&=\left(\tfrac{\mathrm{d}t}{1+\delta t}+o(\mathrm{d}t)\right)\left(\tfrac{t+\mathrm{d}t}{1+\delta(t+\mathrm{d}t)}\right)^{l(C')}\left(\prod_{(\gamma^\circ,\,c)\in C'}w(e(\gamma^\circ))\right)\sum_{C}\sum_{F}\mathbb{P}(\mathfrak{C}_{t+\mathrm{d}t}=C,\ \Phi_{\frac{1}{t+\mathrm{d}t}}=F)\\
&=\frac{\mathrm{d}t}{1+\delta t}\left(\frac{t}{1+\delta t}\right)^{l(C')}\left(\prod_{(\gamma^\circ,\,c)\in C'}w(e(\gamma^\circ))\right)+o(\mathrm{d}t).
\end{aligned}\tag{4.29}$$

In a similar manner, we compute for $0<s<t$ and any two rooted cycle clumps (C'_1,v_1) and (C'_2,v_2) the probability

$$\mathbb{P}\left(\mathcal{C}\left(\{C'_1\}\times\{v_1\}\times \mathrm{d}s\right)=1,\ \mathcal{C}\left(\{C'_2\}\times\{v_2\}\times \mathrm{d}t\right)=1\right),$$

to show that $\mathcal{C}\left(\{C'_1\}\times\{v_1\}\times \mathrm{d}s\right)$ and $\mathcal{C}\left(\{C'_2\}\times\{v_2\}\times \mathrm{d}t\right)$ are independent. Let C_1 and C_2 be any two admissible sets of colored cycles. It is notationally convenient to define the following four admissible sets of colored cycles

$$\begin{aligned}
C_s &:= C_1\\
C_{s+\mathrm{d}s} &:= C_s\cup (C'_1)_{\downarrow C_s}\\
C_t &:= C_{s+\mathrm{d}s}\cup (C_2)_{\downarrow C_{s+\mathrm{d}s}}\\
C_{t+\mathrm{d}t} &:= C_t\cup (C'_2)_{\downarrow C_t}.
\end{aligned}$$

As in eq. (4.28), we have that

$$\begin{aligned}
&\mathbb{P}\left(\mathcal{C}\left(\{C'_1\}\times\{v_1\}\times \mathrm{d}s\right)=1,\ \mathcal{C}\left(\{C'_2\}\times\{v_2\}\times \mathrm{d}t\right)=1\right)\\
&=\mathbb{P}\begin{pmatrix}\mathfrak{C}_s=C_s, & \mathcal{C}\left(\{C'_1\}\times\{v_1\}\times \mathrm{d}s\right)=1\\ \mathfrak{C}_t=C_t, & \mathcal{C}\left(\{C'_2\}\times\{v_2\}\times \mathrm{d}t\right)=1\end{pmatrix}+o(\mathrm{d}s)+o(\mathrm{d}t).
\end{aligned}$$

It further holds for any rooted forest F that

$$\begin{aligned}
&\mathbb{P}\left(\begin{matrix}\mathfrak{C}_s=C_s, & \mathcal{C}\left(\{C'_1\}\times\{v_1\}\times \mathrm{d}s\right)=1\\ \mathfrak{C}_t=C_t, & \mathcal{C}\left(\{C'_2\}\times\{v_2\}\times \mathrm{d}t\right)=1\end{matrix}\ \middle|\ \begin{matrix}\mathfrak{C}_{s+\mathrm{d}s}=C_{s+\mathrm{d}s},\\ \mathfrak{C}_{t+\mathrm{d}t}=C_{t+\mathrm{d}t},\end{matrix}\quad \Phi_{\frac{1}{t+\mathrm{d}t}}=F\right)\\
&=\frac{\mathrm{d}s\ \mathrm{d}t}{(1+\delta s)(1+\delta t)}+o(\mathrm{d}s)+o(\mathrm{d}t).
\end{aligned}$$

We therefore have that

$$\begin{aligned}
&\mathbb{P}\left(\mathcal{C}\left(\{C_1'\}\times\{v_1\}\times \mathrm{d}s\right)=1,\ \mathcal{C}\left(\{C_2'\}\times\{v_2\}\times \mathrm{d}t\right)=1\right)\\
&=\left(\tfrac{\mathrm{d}s\ \mathrm{d}t}{(1+\delta s)(1+\delta t)}+o(\mathrm{d}s)+o(\mathrm{d}t)\right)\sum_{C_1,C_2}\sum_{F}\mathbb{P}\left(\mathfrak{C}_{s+\mathrm{d}s}=C_{s+\mathrm{d}s},\mathfrak{C}_{t+\mathrm{d}t}=C_{t+\mathrm{d}t},\ \Phi_{\frac{1}{t+\mathrm{d}t}}=F\right)\\
&=\left(\frac{\mathrm{d}s\ \mathrm{d}t}{(1+\delta s)(1+\delta t)}+o(\mathrm{d}s)+o(\mathrm{d}t)\right)\left(\frac{s+\mathrm{d}s}{1+\delta(s+\mathrm{d}s)}\right)^{l(C_1')}\left(\frac{t+\mathrm{d}t}{1+\delta(t+\mathrm{d}t)}\right)^{l(C_2')}\\
&\quad\times\sum_{C_1,C_2}\sum_{F}\mathbb{P}(\mathfrak{C}_{s+\mathrm{d}s}=C_1,\ \mathfrak{C}_{t+\mathrm{d}t}=C_1\cup(C_2)_{\downarrow C_1},\ \Phi_{\frac{1}{t+\mathrm{d}t}}=F)\\
&=\mathbb{P}\left(\mathcal{C}\left(\{C_1'\}\times\{v_1\}\times \mathrm{d}s\right)=1\right)\mathbb{P}\left(\mathcal{C}\left(\{C_2'\}\times\{v_2\}\times \mathrm{d}t\right)=1\right)+o(\mathrm{d}s)+o(\mathrm{d}t).
\end{aligned}$$

The above computations show that the random measures of any two disjoint subsets are independent. Independence for any k subsets is shown identically.

By the above independence, it follows from Kingman's representation theorem that $\mathcal{C}$ is a Poisson point process [46]. Since the rate function has been computed in eq. (4.29) above, this completes the proof. □

Corollary 4.5.1. *The the random atomic measure $\mathcal{L}$ is a Poisson point process on $\mathcal{P}^{\mathrm{cl}}\times(0,\infty)$ with intensity measure*

$$\{\gamma\}\times(a,b]\mapsto\int_a^b \kappa_t\ \mathbf{P}_{\mathrm{unif}}((X_k)_{0\le k\le T_t+1}=\gamma\mid X_{T_t+1}=X_0)\ \mathrm{d}t.$$

Proof of theorem 4.1. Theorem 4.1 follows from corollary 4.5.1 and the computations in the proof of corollary 3.6.1. □

4.5.3 Spectral decomposition

The result of theorem 4.2 is a direct consequence of corollary 4.5.1.

Proof of theorem 4.2. By corollary 4.5.1 the intensity measure of $\mathcal{L}$ is given by

$$\{\gamma\}\times(a,b]\mapsto\int_a^b \kappa_t\ \mathbf{P}_{\mathrm{unif}}((X_k)_{0\le k\le T_t+1}=\gamma\mid X_{T_t+1}=X_0)\ \mathrm{d}t.$$

Writing $\gamma=(x_0,x_1\ldots,x_l)$ gives us by the independence of X and T_t that

$$\begin{aligned}
&\mathbf{P}_{\mathrm{unif}}((X_k)_{0\le k\le T_t+1}=\gamma\mid X_{T_t+1}=X_0)\\
&\quad=P(\Gamma_{l,x_0}=\gamma)\mathbf{P}_{\mathrm{unif}}(T_t+1=l,\ X_0=x_0\mid X_{T_t+1}=X_0).
\end{aligned}$$

Using eq. (3.57), we have for any $k\in\mathbb{N}$ and $x\in\mathcal{X}$ that

$$\begin{aligned}
&\mathbf{P}_{\mathrm{unif}}(T_t+1=k,\ X_0=x\mid X_{T_t+1}=X_0)\\
&\quad=\frac{\mathbf{P}_{\mathrm{unif}}\left(X_0=x,\ X_k=x\mid T_t+1=k\right)\mathbf{P}_{\mathrm{unif}}(T_t+1=k)}{\mathbf{P}_{\mathrm{unif}}\left(X_{T_t+1}=X_0\right)}\\
&\quad=\frac{\mathbf{P}_{\mathrm{unif}}\left(X_0=x,\ X_k=x\right)}{\frac{1+\delta t}{n\delta}\kappa_t}\left(\tfrac{\delta t}{1+\delta t}\right)^{k-1}\tfrac{1}{1+\delta t}\\
&\quad=\tfrac{1}{t\kappa_t(1+\delta t)}\left(\tfrac{\delta t}{1+\delta t}\right)^{k}\mathbf{P}_x\left(X_k=x\right)=\tfrac{1}{t\kappa_t(1+\delta t)}\left(\tfrac{\delta t}{1+\delta t}\right)^{k}\left(I-\tfrac{1}{\delta}L\right)^k(x,x).
\end{aligned}$$

For two measures ν_f and ν_g with Radon-Nikodym derivatives $f := \frac{d\nu_f}{d\mu}$ and $g := \frac{d\nu_g}{d\mu}$ we write $\langle \nu_f, \nu_g \rangle_\mu^* := \langle f, g \rangle_\mu = \sum_{x \in \mathcal{X}} \frac{\nu_f(x)\nu_g(x)}{\mu(x)}$. Denote by $\nu_0, \ldots, \nu_{n-1}$ the left eigenmeasures of L, so that $\frac{d\nu_j}{d\mu} = \mathbf{v}_j$. Note that each eigenvector $\mathbf{v}_j$ of L is also an eigenvector of $I - \frac{1}{\delta}L$ with eigenvalue $1 - \frac{\lambda_j}{\delta}$. Hence, by letting $\mathbf{1}_x$ denote the indicator of x and δ_x the Dirac-measure at x, we have that

$$\begin{aligned}
\left(I - \tfrac{1}{\delta}L\right)^k (x,x) = \delta_x \left(I - \tfrac{1}{\delta}L\right)^k \mathbf{1}_x &= \left(\sum_{i<n} \langle \nu_i, \delta_x \rangle_\mu^* \, \nu_i\right) \left(\sum_{j<n} \langle \mathbf{v}_j, \mathbf{1}_x \rangle_\mu \left(1 - \tfrac{\lambda_j}{\delta}\right)^k \mathbf{v}_j\right) \\
&= \sum_{i<n} \sum_{j<n} \left(1 - \tfrac{\lambda_j}{\delta}\right)^k \langle \nu_i, \delta_x \rangle_\mu^* \langle \mathbf{v}_j, \mathbf{1}_x \rangle_\mu \, \nu_i \mathbf{v}_j \\
&= \sum_{j<n} \left(1 - \tfrac{\lambda_j}{\delta}\right)^k \langle \nu_j, \delta_x \rangle_\mu^* \langle \mathbf{v}_j, \mathbf{1}_x \rangle_\mu = \frac{1}{\delta^k} \sum_{j<n} (\delta - \lambda_j)^k \, \mathbf{v}_j(x)^2 \mu(x).
\end{aligned}$$

It follows that

$$\begin{aligned}
&\mathbf{P}_{\text{unif}}(X_0 = x,\ T_t + 1 = k \mid X_{T_t+1} = X_0) \\
&= \frac{1}{t\kappa_t(1+\delta t)} \left(\frac{t}{1+\delta t}\right)^k \sum_{j<n} (\delta - \lambda_j)^k \, \mathbf{v}_j(x)^2 \mu(x) \\
&= \frac{1}{\kappa_t} \sum_{j<n} \frac{\delta - \lambda_j}{(1+\delta t)(1+\lambda_j t)} P(T_j + 1 = k) P(Y_j = x),
\end{aligned}$$

which concludes the proof. □