A new combinatorial representation of the additive coalescent
Jean-François Marckert, Minmin Wang

To cite this version:
Jean-François Marckert, Minmin Wang. A new combinatorial representation of the additive coalescent. 2016. <hal-01361053>

HAL Id: hal-01361053
https://hal.archives-ouvertes.fr/hal-01361053
Submitted on 6 Sep 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
A new combinatorial representation of the additive coalescent.

Jean-François Marckert
CNRS, LaBRI
Université de Bordeaux
351 cours de la Libération
33405 Talence cedex
France

Minmin Wang
Conicet-UBA
Universidad de Buenos Aires
Ciudad Universitaria
Capital Federal, C1428EGA
Argentina

Abstract

The standard additive coalescent starting with \( n \) particles is a Markov process which owns several combinatorial representations, one by Pitman as a process of coalescent forests, and one by Chassaing & Louchard as the block sizes in a parking scheme. In the coalescent forest representation, some edges are added successively between a random node and a random root. In this paper, we investigate an alternative construction by adding edges between the roots. This construction induces the same process at the level of cluster sizes, but allows one to make numerous connections with some combinatorial and probabilistic models that were not known to be connected with additive coalescent. The variety of the combinatorial objects involved here – size biased percolation, parking scheme in a tree, increasing trees, random cuts of trees – justifies our interests in this construction.

Mathematics Subject Classification (2000) 60C05, 60K35, 60J25, 60F05, 68R05

Key Words: additive coalescent, Cayley trees, increasing trees, parking, random walks on trees

Acknowledgement: The research has been supported by ANR-14-CE25-0014 (ANR GRAAL).

1 Introduction

The aim of this paper is to present a new combinatorial point of view on the additive coalescent starting with \( n \) particles with same masses. In his paper [29], Pitman proposes a combinatorial construction of this process. The idea of his construction is rooted in the combinatorics of Cayley trees. In this construction, the coalescence is realized by successively adding (random) edges in a forest, giving rise eventually to a uniform Cayley tree in the end. Here, we investigate an alternative construction which leads to, as we believe, a richer combinatorial environment in providing a link with several other processes defined on trees: a notion of size-biased percolation on a tree, a tree-shape parking scheme, a random walker model in the tree, a model of decreasing edge-labellings of a Cayley tree, as well as the random cutting problem which dates back to Meir & Moon [26]. Finally, we propose an enriched version of the parking scheme studied by Chassaing & Louchard [13] for the additive coalescence which unifies our construction and Pitman’s construction. We also study the asymptotics for the additive coalescence, based on the parking scheme.
Notation. Throughout the paper, for any integer \( n \geq 1 \), \([n]\) denotes the set \( \{1, \ldots, n\} \). We use both notation \( \#A \) and \(|A|\) for the cardinal of a finite set \( A \). If \( t \) is a tree or a forest, its size \(|t|\) is the number of its nodes. The number of edges of \( t \) is denoted by \(|t|_E\), and the set of edges by \( E(t)\).

In the paper, all random variables (r.v.) are assumed to be defined on a common probability space \((\Omega, \mathcal{A}, \mathbb{P})\). For any finite set \( E \), we write \( X \sim \text{Uniform}(E)\) to state that the r.v. \( X \) is uniform on \( E \).

1.1 Additive coalescent process

Throughout this work, \( n \geq 1 \) denotes a positive integer.

Let us first recall the Marcus–Lushnikov process. At time 0 the system contains \( n \) particles labelled from 1 to \( n \), each with mass 1. Consider these particles as the vertices of a complete graph with virtual edges, and equip the edges between vertices \( i \) and \( j \) with random exponential clocks of parameter \( K(x_i, x_j) = x_i + x_j \) where \( x_i \) and \( x_j \) are the masses of \( i \) and \( j \). When the clock between \( i \) and \( j \) rings, replace the masses \( x_i \) and \( x_j \) by \( x_i + x_j \) and 0, respectively, and update the parameters of the clocks involving \( i \) and \( j \) so that the rates remain given by the kernel. This is the so-called Marcus-Lushnikov [25, 24] process, which is a continuous-time Markov process. Marcus–Lushnikov processes may be defined even when the number of particles is infinite, provided that the total mass is finite (see e.g. [16]). The Marcus–Lushnikov process has been introduced in order to model coalescence systems, in which particles have a propensity (represented by the kernel \( K \)) to coalesce.

Another representation of this process relies on a combinatorial process taking its values in \( \mathcal{P}_n \) the set of partitions of \([n]\). Again, at time 0, \( n \) particles labelled from 1 to \( n \) are present in the system at time 0. The masses do not play an explicit role here, but it may help the reader to consider that the particles all have mass 1 at time 0. Denote by \( \mathcal{P}_n \) the set of partitions of \([n]\). The additive coalescent (of length \( n \)) is a Markov chain \( \{\Pi^+_k : 0 \leq k \leq n - 1\} \) taking values in \( \mathcal{P}_n \) such that:

1) \( \Pi^+_k \) is a partition of \([n]\) into \( n-k \) subsets (so that \( \Pi^+_0 \) is the partition of \([n]\) into \( n \) singletons);
2) given \( \Pi^+_k = \{\pi^+_1, \ldots, \pi^+_{n-k}\} \) with \( n_i = |\pi^+_i|, 1 \leq i \leq n-k \), \( \Pi^+_{k+1} \) is obtained from \( \Pi^+_k \) as follows: pick a pair of parts \( \{\pi^+_{k,i}, \pi^+_{k,j}\} \) with probability

\[
\sum_{1 \leq i' < j' \leq n-k} (n_{i'} + n_{j'}) = \frac{n_i + n_j}{n(n-k-1)}, \quad 1 \leq i < j \leq n-k;
\]

and merge this pair into one subset (and let the other parts unchanged).

At the level of component sizes, the Markov chain \( \{\Pi^+_k : 0 \leq k \leq n - 1\} \) has the same evolution as the Marcus–Lushnikov process (at its jumping times) as described above.

**Proposition 1** ([29], Proposition 6). For each \( 0 \leq k \leq n-1 \) and each partition \( \pi := \{\pi_1, \ldots, \pi_{n-k}\} \) of \([n]\) into \( n-k \) subsets with \(|\pi_i| = n_i, 1 \leq i \leq n-k\), we have

\[
\mathbb{P}(\Pi^+_k = \pi) = \frac{\Pi_{1 \leq i \leq n-k} n_i^{n_i-1}}{n^k \binom{n}{n-k-1}}.
\]
An idea to encode coalescence consists in adding some edges between the components that merge, creating in such a way forests and trees... and this works wonderfully, since the forests and trees that come into play are particularly simple, and this has numerous implications.

Let us briefly recall Pitman’s construction of \((\Pi_k^+)^+\) in \([29]\) which is based on a coalescent process of rooted forests. A rooted forest over \([n]\) here is a graph on \([n]\) whose connected components are trees, each tree \(t\) being rooted at one of its vertex \(r\). Here and below, edges are considered directed towards their roots. For \(1 \leq k \leq n\), denote by \(\mathcal{F}_{k,n}\) the set of rooted forests over \([n]\) which have \(k\) tree components. Notice that \(\mathcal{F}_{1,n}\) is the set of rooted labelled trees (Cayley trees) with \(n\) nodes. By Riordan \([31]\) or Pitman \([29, \text{Formula (5)}]\),

\[
\#\mathcal{F}_{k,n} = n^{n-k} \binom{n-1}{k-1}, \quad \text{for } 1 \leq k \leq n.
\] (2)

Let us define a sequence of random rooted forests \((F^\text{PF}_k, 0 \leq k \leq n-1)\) over \([n]\) as follows (the exponent “PF” refers to “Pitman’s forest”). First, \(F^\text{PF}_0\) is the only element of \(\mathcal{F}_{n,n}\), namely, the trivial forest of \(n\) trees reduced to their roots which are labelled from 1 to \(n\).

**Pitman’s Algorithm (PF Algo).**

For \(1 \leq k \leq n-1\), to construct \(F^\text{PF}_k\) from \(F^\text{PF}_{k-1}\), do the following three steps.

(a) Choose a uniform node \(\alpha^\text{PF}_k\) in \(F^\text{PF}_{k-1}\). Let \(t_k\) be the tree which contains \(\alpha^\text{PF}_k\).

(b) Choose a tree \(t^\text{PF}_k\) in \(F^\text{PF}_{k-1}\) uniformly among the trees different from \(t_k\). Denote by \(\beta^\text{PF}_k\) the root of \(t^\text{PF}_k\).

(c) Add the directed edge \(\beta^\text{PF}_k \to \alpha^\text{PF}_k\).

Clearly, \(F^\text{PF}_k \in \mathcal{F}_{n-k,n}\) and has exactly \(k\) edges, for \(0 \leq k \leq n-1\). Observe that the vertex sets of the tree components of \(F^\text{PF}_k\) form a partition of \([n]\); denote by \(\Pi^\text{PF}_k\) this partition.

**Theorem 2. \([29], \text{Theorem 4}\) (i) The following distribution equality holds**

\[
(\Pi^\text{PF}_k, 0 \leq k \leq n-1) \overset{(d)}{=} (\Pi_k^+, 0 \leq k \leq n-1).
\] (3)

(ii) Moreover, for \(0 \leq k \leq n-1\), \(F^\text{PF}_k \sim \text{Uniform}(\mathcal{F}_{n-k,n})\).

This representation has led to several interesting consequences:

– one gets a limit representation of the coalescent process when \(n \to +\infty\) (see Section 1.4 for more information);

– one shows that the time-reversal of the coalescence process, the so-called fragmentation process, is simple too, since it amounts to choosing at each step a uniform edge and removing it (see Section 3 for details and a related question);

– the genealogy of the fragments hence produced has a branching structure, studied per se \([10]\) (see also \([14, 12]\) and Section 3).

We wish to comment on two simple facts:

• It is not apparent at first glance that the evolution of tree sizes in Pitman’s construction coincides
with the additive coalescent, because of the asymmetry in the construction of the forest. On the other hand, two trees \( t_1 \) and \( t_2 \) will be connected anyway either we add an edge from \( t_1 \) to \( t_2 \) or from \( t_2 \) to \( t_1 \). Thus, the probability that these two trees get merged is proportional to \(|t_i| + |t_j|\).

- In above, an edge is added between a uniform random vertex \( u \) taken in the global forest, and a random root \( r \) taken among the roots of the trees which do not contain \( u \). This rule is somehow arbitrary. Another choice of the edge will yield the same partitions (thus the same distribution as given in Theorem 2 (i)), as long as it connects the same pairs of trees. And a natural question is to wonder what will become Theorem 2 (ii) if another rule is applied. The following paragraph answers partially to this question.

### 1.2 The sorted coalescence algorithm

We propose the following variant of Pitman’s Algo.

**Sorted coalescence algorithm (SC Algo).**

Let \( F_{0}^{SC} \) be the trivial forest of \( n \) trees. We construct \( F_{k}^{SC} \) from \( F_{k-1}^{SC} \) for \( 1 \leq k \leq n-1 \), as follows.

1. Choose a uniform node \( \alpha_{k}^{SC} \) in \( F_{k-1}^{SC} \). Let \( t_{k} \) be the tree which contains \( \alpha_{k}^{SC} \), and let \( r(\alpha_{k}^{SC}) \) be the root of \( t_{k} \).
2. Choose a tree \( t_{k}^{SC} \) in \( F_{k-1}^{SC} \) uniformly in \( F_{k-1}^{SC} \) among the trees different from \( t_{k} \). Denote by \( \beta_{k}^{SC} \) the root of \( t_{k}^{SC} \).
3. Add the directed edge \( e_{k}^{SC} := r(\alpha_{k}^{SC}) \rightarrow \beta_{k}^{SC} \).

In other words, the dynamic is almost the same as that of Pitman’s. Instead of adding the edge \( \beta_{k}^{PF} \rightarrow \alpha_{k}^{PF} \), we add the edge \( r(\alpha_{k}^{SC}) \rightarrow \beta_{k}^{SC} \), but the distribution of \((\alpha_{k}^{PF}, \beta_{k}^{PF})\) coincides with that of \((\alpha_{k}^{SC}, \beta_{k}^{SC})\). The new edge is then added between roots of trees in the forest, when it was between a random node and a root before. Since this “small” difference occurs for each \( k \), the successive adding of edges leads finally to much different forests.

However, in terms of the cluster size evolution, or in terms of the induced partitions, both constructions induce exactly the same distribution: at each time step, both construction consist in merging two components, the first one being chosen with probability proportional to its size, the second one being taken uniformly among the other parts.

For \( 0 \leq k \leq n - 1 \), let \( \Pi_{k}^{SC} \) be the partition of \([n]\) induced by the vertex sets of the tree components of \( F_{k}^{SC} \). We then have immediately

**Theorem 3.** The following distributional equality holds

\[(\Pi_{k}^{SC}, 0 \leq k \leq n-1) \overset{(d)}{=} (\Pi_{k}^{PF}, 0 \leq k \leq n-1) \overset{(d)}{=} (\Pi_{k}^{+}, 0 \leq k \leq n-1). \] (4)

On the other hand, by putting the label \( k \) on the \( k \)th created edge \( e_{k} \) (we denote \( L_{k}^{SC} = k \)), we obtain at each time step a forest of Cayley trees equipped with a decreasing edge-labelling along the branches (Theorem 6). We call such a tree a decreasing tree. Knowing the complete sequence
\((F^SC_k, 0 \leq k \leq n-1)\), the edge labelling can be recovered since between time \(k\) and \(k+1\) only one edge has been added... and it has label \(k+1\). On the other hand, the single tree \((F^SC_{n-1}, L^SC_{n-1})\) equipped with its edge labelling also encodes completely the history of the additive coalescent.

**Remark 4.** Visibly, there is a similarity in the constructions of the processes \((F^{PF}_k, 0 \leq k \leq n-1)\) and \((F^SC_k, 0 \leq k \leq n-1)\). We can construct a coupling between the two sequences of forests, but this is not the right moment... We postpone it to Section 2.8.

### 1.3 Content of the paper

In Section 2.1, we study the distribution of the process \(((F^SC_k, L^SC_k), 0 \leq k \leq n-1)\). In particular, we will see that conditional on \(F^SC_{n-1}\), the labelling \(L^SC_{n-1}\) is uniform among the decreasing edge labellings of \(F^SC_{n-1}\) (Theorem 6).

We then show \(((F^SC_k, L^SC_k), 0 \leq k \leq n-1)\) – our forest encoding of the additive coalescent – is connected to many new natural stochastic models:

- In Section 2.2, it is shown that, given a Cayley tree \(T\), a uniform decreasing edge-labelling may be constructed thanks to a simple Markov model that we call “random walker algorithm” in the tree (Lemma 7).
- In Section 2.3, Proposition 9 shows a link between the uniform decreasing edge-labelling and a percolation process in the tree that we call size-biased percolation: the infection starts from the root; at each step a neighboring node is infected proportionally to the size of the tree hanging out of this node.
- In Section 2.4, the model of decreasing trees is shown to be related with a parking scheme in the tree in which successive cars choose an edge \(e\) uniformly and park at the edge nearest to the root among the available edges on the branch from \(e\) to the root.

Each of the constructions mentioned above provides a representation of the additive coalescent conditional on the final tree.

- In Section 2.5, we recall the representation of the additive coalescent using a linear parking scheme introduced in Chassaing & Louchard [13]. Based on this representation, they study the asymptotic behaviors of the cluster sizes in the additive coalescent which has started with \(n\) particles and has experienced \(n - \lambda \sqrt{n}\) mergings (this is the critical window). In Section 2.6, we revisit and enrich this construction in order to make it encode faithfully the additive coalescent, and not only the cluster sizes. Using this enriched parking scheme, the study of the asymptotic is still possible and we recover the limit behavior proved by [13] (see Section 2.9). Moreover, the connection between this enriched parking problems, Pitman construction and our model of decreasing trees appears to be somehow trivial (Sections 2.6.2 and 2.6.3).

Finally in Section 3, we use the forests \((F^SC_k, 0 \leq k \leq n-1)\) and the coupling with \((F^{PF}_k, 0 \leq k \leq n-1)\) to give a new and simple proof on a well-known problem on random cutting of trees.
1.4 Related works on the additive coalescent

The aim of coalescent processes is to model physical/chemical systems containing a large numbers of weighted particles that have a propensity to merge. There are several types of models, but each of them relies on the so-called collision kernel $K$, which encodes the propensity of two particles with masses $x$ and $y$ to merge: they merge “at rate $K(x, y)$”.

Smoluchowski coagulation equations [32, 6] provide a description of the so-called mean-field behavior of a coalescent system with an infinite number of particles by specifying the evolution of the densities of particle masses. These equations form an infinite system of ordinary differential equations. Explicit solutions are available in some special cases, including the additive kernel, i.e., $K(x, y) = x + y$, the multiplicative one, i.e., $K(x, y) = xy$ (see [6, 8] and [18]), and the Kingman coalescent $K(x, y) = 1$ ([21]).

The probabilistic counterpart of the Smoluchowski equations is the Marcus–Lushnikov model [25, 24], intended to describe the continuous-time evolution of the masses – at the particles level – for a system with a finite number of particles. Between each pair of particles $(p_1, p_2)$, put a clock which rings according to independent exponential variables with parameter $K(\text{Mass}(p_1), \text{Mass}(p_2))$. When a clock rings, merge the corresponding pair of particles and update all the clocks. For any $K$ this is a time homogeneous continuous-time Markov process.

In the additive coalescent case, when starting from $n$ particles with initial mass 1, the time needed to pass from the $k$th merging to the $(k+1)$th in the Marcus–Lushnikov model is given by $\tau_k = \min\{\tau(p_1, p_2), \{p_1, p_2\} \in P_k\}$ where the $\tau(p, p')$ are independent, exponentially distributed with parameter $K(\text{Mass}(p), \text{Mass}(p'))$, and $P_k$ is the set of pairs of (different) particles in the system after $k$ merging. One checks that $\tau_k$ is exponential with parameter $n(n - k - 1)$, so that it depends only on the number of mergings already done, and not on the details of the current masses. For this reason, the discrete-time version studied here, which is parametrized by the number of mergings, can be related to the continuous-time version by a simple time change.

There is an extensive literature in general coalescent processes or in the additive coalescent (see Pitman [30, Section 9] for an overview). Let us cite a few works which are most related to the present one.

First of all, in his work [29] where he introduced the model of coalescent forests for the additive coalescence as recalled in Section 1.1, Pitman also observed the following things:

a) the tree obtained eventually from the construction is uniform among the set of Cayley trees with $n$ nodes;

b) if we rank the edges according to their appearance, given the final tree, this ordering of the edges is uniformly distributed among all the edge-orderings of the tree. This leads to a representation of the additive coalescent as a time reversal of a fragmentation process defined on the uniform tree.

We have discussed the case where initially there are $n$ masses. We can also talk about the limit case when $n \to \infty$. This “limit”, or the convergence of the additive coalescence processes, say, is well understood thanks to Evans & Pitman [16] who showed the following. There exists a Feller process $X^\infty = (X^\infty(t), t \in \mathbb{R}_+)$ taking values in

$$\ell_1 = \{(x_i, i \geq 1) : x_1 \geq \cdots \geq x_i \geq x_{i+1} \geq \cdots \geq 0, \sum x_i < +\infty\}$$

(5)
which has the dynamics of additive coalescent; moreover, if $X^n$ denotes the Markov process with the same kernel starting from $(1/n, \cdots, 1/n, 0, 0, \cdots)$ (this corresponds to the component sizes in Marcus–Lushnikov model with $n$ initial masses $1/n$), then

$$
(X^n(\cdot + (\log n)/2)) \xrightarrow{d}{n\to\infty} (X^\infty(\cdot)),
$$

where the convergence holds in $\mathbb{D}((-\infty, +\infty), \ell^1)$. The limit is called the standard additive coalescent (see Section 2.9 below for more information). Moreover, the coalescence-fragmentation duality mentioned above remains valid when $n$ tends to infinity, leading to the Aldous–Pitman's representation [5] of the standard additive coalescent on Aldous' continuum random tree.

Another combinatorial representation of the additive coalescent has been investigated by Chassaing & Louchard [13] (see Section 2.5). They observe that in a certain parking scheme, known to be related to a well-known data structure named “hashing with linear probing” in computer science, parking blocks evolve as the additive coalescent. This representation has the following advantage: it allows for a direct encoding of the additive coalescent (starting with $n$ particles with mass 1) by a Markov chain on $\mathbb{Z}$ such that the sizes of the parking blocks at each time correspond to the excursion lengths of the process above its running infimum (see Section 2.5). Moreover, it is shown in [13] that these excursions lengths converge to those of a Brownian excursion with a linear drift. See Bertoin [7] for a related work. Let us also mention that this is close in spirit in the way that Aldous encodes the multiplicative coalescent using another random walk which converges to the Brownian motion with a parabolic drift (see [4]).

Recently, Broutin and the first author [11] revisited the study of the asymptotic additive and multiplicative coalescents. They proved by a unified approach, by a passage to the limit, that the excursions of the Brownian motion with parabolic drift above the running minimum $(B_t - \lambda t^2/2, t \geq 0)$, and those of $(\varepsilon_t - \lambda t)$ are the limit of the cluster sizes in the multiplicative and additive coalescent as a process indexed by the time (which is $\lambda$ here).

2 A new point of view on the additive coalescent

2.1 Alternative to Pitman’s construction

We discuss here the properties of the sequence of forests $(F^\text{SC}_k, 0 \leq k \leq n - 1)$ produced by SC Algo, introduced in Section 1.2.

The terminology “sorted coalescence algorithm” is motivated by the following fact. We will see that in this construction, if we fix the tree that appears in the end, then the coalescence takes place in an ordered manner: along the branches of the tree, from the leaves towards the root.

Recall that the edges of the forest $F^\text{SC}_k$ have been labelled by their rank of appearance. We say that an edge-labelling $L = (L_e, e \in E(f))$ of a forest $f = (f_1, \cdots, f_k)$ is decreasing if it has the following properties:

- the set of labels is $\{1, \cdots, |f|\}$, each label being used exactly once,
- in each tree $f_j$, the labelling is decreasing along each simple path going out of the root $r$: if $(u_0 = r, u_1, \cdots, u_h)$ is a simple path in $f_j$ for some $h \geq 1$, then

$$
L(u_0, u_1) > \cdots > L(u_{h-1}, u_h).
$$

(6)
Let \( \text{DL}(f) \) be the set of decreasing edge-labellings of a forest \( f \). For a vertex \( u \in f \), let \( \text{Sub}_u(f) \) denote the subtree of \( f \) rooted at \( u \). Also recall that for a tree or a forest \( t \), \( |t| \) stands for the number of its nodes.

**Lemma 5.** For \( 1 \leq k \leq n \) and \( f \in \mathcal{F}_{k,n} \),

\[
\#\text{DL}(f) = (n-k)! \prod_{u \in f \setminus \{\rho_i, 1 \leq i \leq k\}} \frac{1}{|\text{Sub}_u(f)|},
\]

where \( \rho_i, 1 \leq i \leq k \), are the \( k \) roots of \( f \).

**Proof.** We first show that for a rooted tree \( t \),

\[
\#\text{DL}(t) = \prod_{u \in t} \frac{(|\text{Sub}_u(t)| - 1)!}{|\text{Sub}_u(t)|!}.
\]

Notice that after simplification, (8) leads to the formula (7) when \( k = 1 \).

Now, (8) can be checked by an induction on the size of \( t \). For \( |t| = 1 \), this is trivial. Suppose (8) holds for trees with sizes up to \( m - 1 \geq 1 \). For a rooted tree \( t \) of size \( m \) with root \( r \), denote by \( J \) the in-degree of \( r \). Let \( (E_j, 1 \leq j \leq J) \) be the partition of the edge set of \( t \) induced by the \( J \) subtrees above \( r \). Then a decreasing edge labelling of \( t \) is obtained by first choosing the subsets of the labels for \( E_j, 1 \leq j \leq J \), and then assigning the largest label of each subset to the edge adjacent to the root. Applying the induction hypothesis yields (8). To get to the enumeration for rooted forests, we first choose the subsets of labels for \( t_i, 1 \leq i \leq k \), then apply (8) for each \( t_i \).

Equation (7) easily follows. \( \square \)

Write \( T^{SC} = F_{n-1}^{SC} \), the tree which appears at the end of the construction. Recall that we defined an edge labelling by defining \( L^SC_e = k \) if \( e \) is the \( k \)th added edge. Let \( L^{SC} = (L_e, e \in E(T^{SC})) \) be the global edge labelling of \( T^{SC} \). For \( 0 \leq k \leq n - 1 \), let \( L^SC_k \) be the subset of \( L^{SC} \) obtained by removing those labels \( > k \). Then it is clear that \( L^SC_k \) is a decreasing edge-labelling of \( F_k^{SC} \). As said before, the pair \( (T^{SC}, L^{SC}) \) encodes the history of the coalescence so that we can recover from \( (T^{SC}, L^{SC}) \) the sequence of the pairs \( (r(\alpha^{SC}_k), \beta^{SC}_k), 1 \leq k \leq n - 1 \).

**Theorem 6.** For \( 0 \leq k \leq n - 1 \), we have

\[
\mathbb{P}\left((F_k^{SC}, L_k^{SC}) = (f, t)\right) = \frac{1_{\mathcal{F}_{n-k,n}}(f)}{\#\mathcal{F}_{n-k,n}} \cdot \frac{1_{\text{DL}(f)}(t)}{\#\text{DL}(f)}.
\]

Hence, \( F_k^{SC} \sim \text{Uniform}(\mathcal{F}_{n-k,n}) \) and conditional on \( F_k^{SC} = f \), \( L_k^{SC} \sim \text{Uniform}(\text{DL}(F_k^{SC})) \).

In contrast to Theorem 3, this theorem should appear a bit unexpected, since at the level of forests the sorted coalescence construction is rather different from the one of Pitman, and seems to “favor” the neighborhoods of the roots. On the other hand, notice that after the edge \( r(\alpha^{SC}_k) \rightarrow \beta^{SC}_k \) has been added, then \( r(\alpha^{SC}_k) \) is not a root anymore, so that in the sequel of the construction, no more edges will be added from or to \( r(\alpha^{SC}_k) \).
Proof. Let \( f \in \mathcal{F}_{n-k,n} \) and \( \ell \in \text{DL}(f) \). Notice that if \((F_k^{SC}, L_k^{SC}) = (f, \ell)\) we can recover the sequence \((F_j^{SC}, 0 \leq j \leq k)\) from \((f, \ell)\). In particular, if we write \( x_j \to y_j \) for the edge labelled \( j \) in the labelling \( \ell \) (recall that the edges are directed towards the root), then \( \text{Sub}_{x_j}(f) \) is the tree component of \( F_{j-1}^{SC} \) which contains \( \alpha_j^{SC}, 1 \leq j \leq k \). By our construction, the probability that \( F_j^{SC} \) is obtained from \( F_{j-1}^{SC} \) by adding the edge \( x_j \to y_j \) is \(|\text{Sub}_{x_j}(f)|/(n(n-j))\). This yields
\[
\mathbb{P}(F_k^{SC} = f; L_k^{SC} = \ell) = \frac{(n-k-1)!}{n^k(n-1)!} \prod_{1 \leq j \leq k} |\text{Sub}_{x_j}(f)|.
\]
Since each vertex of \( f \) except the \( n-k \) roots appears exactly once in \( \{x_j, 1 \leq j \leq k\} \), we obtain that
\[
\mathbb{P}(F_k^{SC} = f; L_k^{SC} = \ell) = \frac{(n-k-1)!}{n^k(n-1)!} \prod_{u \in f \setminus \{\rho_i, 1 \leq i \leq n-k\}} |\text{Sub}_u(f)|,
\]
where \( \rho_i, 1 \leq i \leq n-k \) are the \( n-k \) roots of \( f \). Comparing this with (2) and (7), we get (9). \( \square \)

2.2 Random walker algorithm

Here we discuss a model of random walkers in a Cayley tree \( T \) with \( n \) nodes. In this model, \( n-1 \) random walkers will climb successively in the tree, walking from one edge to an adjacent one. The \( k \)th walker will decide to stop at a certain edge \( e \), and will label this edge \( k \). The global labelling will be a uniform decreasing edge-labelling of \( T \).

At time 0 all the edges of \( T \) are white. Let \( r \) be the root of \( T \). Add to \( T \) an extra white edge \( e^* \) from an extra vertex \( u \) and \( r \) (see Fig. 1). The starting point for the random walkers will always be this edge \( e^* \). Consider \( T \cup \{u\} \) rooted in the extra vertex \( u \). The usual parent-children relation among the nodes can be extended to the edges in an obvious way. This leads to a notion of child-edges of an edge \( e \in E(T) \).

Random walker algorithm.

For \( k \) going from 1 to \( n-1 \), the \( k \)th walker starts from the root-edge \( e^* \). Its journey runs as follows. Assume that at some moment, she is at some edge \( e \). Let \( d \) be the number of white children-edges of \( e \).

- If \( d = 0 \), that is, if there are no white child-edge of \( e \), paint in black this last edge \( e \) and label it with the rank of the walker (put \( L_e^{\text{RW}}(T) = 1 \) for the first random walker, \( L_e^{\text{RW}}(T) = 2 \) for the second, etc).
- If \( d \geq 1 \), let \( e_1, e_2, \ldots, e_d \) be the white child-edges of \( e \). The walker goes to \( e_i \) with probability proportional to \( W_i, 1 \leq i \leq d \), where \( W_i \) is the number of the white edges contained in the subtree of \( T \) rooted at \( e_i \) (including \( e_i \)).

Hence, when \( k \) walkers have climbed, there are exactly \( k \) black edges, labelled from 1 to \( k \) (see Fig. 1). In the end, all the edges of \( T \) are painted black. Moreover, \( T \) is equipped with an edge-labelling \( L^{\text{RW}}(T) = (L_e^{\text{RW}}(T), e \in E(T)) \) induced by the ranks of the walkers.

Lemma 7. For each fixed \( T \in \mathcal{F}_{1,n} \), \( L^{\text{RW}}(T) \sim \text{Uniform}(\text{DL}(T)) \).

9
Proof. If $e_i$, $1 \leq i \leq k$, are the edges adjacent to the root, write $T_i^+$ for the subtrees rooted at $e_i$. By a decomposition at the root and an induction argument, we see that the proof reduces to identifying the joint distribution of $\{\ell(e), e \in T_i^+\}, 1 \leq i \leq k$, the sets of the labels attributed to the $T_i^+$’s. For this, only the first step of each walker needs to be followed. If the $w$th walker goes in $T_i^+$, let us set $c(w) = i$, $1 \leq w \leq n - 1$. When $t$ walkers have climbed, let $l_t(i) = \#\{w \in \{1, 2, \ldots, t\} : c(w) = i\}$ be the number of the walkers which have gone into $T_i^+$; then $\mathbb{P}(c(t + 1) = i) = (|T_i^+|_E - l_t(i)) / |T|_E - t)$. It follows that $\{\ell(e), e \in T_i^+\}, 1 \leq i \leq k$ forms a uniform partition of $\{1, \ldots, |T|_E\}$ into $k$ parts of respective sizes $|T_i^+|_E$, $1 \leq i \leq k$. By the previous arguments, this shows that $L^{RW}(T) \sim \text{Uniform}(\text{DL}(T))$. \hfill $\square$

For $0 \leq k \leq n - 1$, by discarding the edges $e$ whose labels $L_e^{RW}(T) > k$, we obtain a forest over $[n]$. Then the vertex sets of the tree components of this forest induce a partition of $[n]$, which we denote by $\Pi_k^{RW}$. Alternatively, $\Pi_k^{RW}$ can be defined by looking at the black components of $T$ defined as follows. When $k$ walkers have climbed, there are exactly $k$ black edges. Then we find in $T$ (see Fig. 1):

- the white edges constitute the edge set of a subtree of $T$ which contains the root,
- the black edges along with the $n$ vertices form a forest over $[n]$ with $n - k$ black tree components.

Here is another representation of the additive coalescent:

**Theorem 8.** For $T \sim \text{Uniform}(\mathcal{F}_{1,n})$,

$$\left( \Pi_k^{RW}, 0 \leq k \leq n - 1 \right) \overset{(d)}{=} \left( \Pi_k^+, 0 \leq k \leq n - 1 \right).$$

*Proof.* From Lemma 7 and Theorems 3 and 6, we readily see that the pair $(T, L^{RW}(T))$ produced by the random walkers model has the same distribution as $(T^{SC}, L^{SC})$. \hfill $\square$
2.3 Size-biased percolation

We introduce a model of percolation on a tree that we call size-biased percolation. It will be seen to be a time-reversal of the random walker process. Take a tree $T$ rooted at $r$ in $\mathcal{F}_{1,n}$. We define a sequence $(\text{Perc}_k(T), 0 \leq k \leq |T|_E)$ which will represent the percolated cluster at each time.

– At time $k = 0$, the cluster $\text{Perc}_0(T)$ is reduced to $r$.

– For $0 \leq k \leq |T|_E - 1$, add to the cluster $\text{Perc}_k(T)$ the edge $e$ adjacent to it with probability proportional to the size of the subtree hanging out of $e$, that is proportional to the number of nodes in the subtree of $T \setminus \text{Perc}_k(T)$ which is adjacent to $e$.

Observe that each $\text{Perc}_k(T)$ is a subtree of $T$ containing the root. Also, $k \mapsto \text{Perc}_k(T)$ is increasing for the inclusion order. For $0 \leq k \leq |T|_E$, let $T^\text{Perc}_k$ be the graph obtained from $T$ by removing the edges of $\text{Perc}_k(T)$ (the nodes are kept), and let $\Pi^\text{Perc}_k(T)$ be the partition of $[n]$ induced by the vertex sets of the components of $T^\text{Perc}_k$ (see Fig. 2).

![Figure 2: construction of $(\text{Perc}_k(T), 0 \leq k \leq |T|_E)$ by successively adding edges to the present cluster (drawn in red). Each time, choose a uniform white edge $e$ and color in red the white edge on the path from $e$ to the root which is nearest to the root.](image)

As before, we introduce an edge-labelling $L^{\text{Perc}}(T)$ of $T$ by putting a label $n - k$ to the unique edge $e$ in $\text{Perc}_k(T) \setminus \text{Perc}_{k-1}(T)$, $1 \leq k \leq |T|_E$. It is easy to see that $L^{\text{Perc}}(T)$ belongs to $\text{DL}(T)$, the set of decreasing edge-labellings of $T$.

**Proposition 9.** For any rooted tree $T$ we have $L^{\text{Perc}}(T) \sim \text{Uniform}(\text{DL}(T))$. If $T$ uniform in $\mathcal{F}_{1,n}$, then

$$\left(\Pi^{\text{Perc}}_{n-1-k}, 0 \leq k \leq |T|_E\right) \overset{(d)}{=} \left(\Pi^\text{SC}_k, 0 \leq k \leq |T|_E\right);$$

(10)

as a consequence, $\left(\Pi^{\text{Perc}}_{n-1-k}, 0 \leq k \leq n - 1\right)$ is the additive coalescent.

**Proof.** The proof is similar to the one of Lemma 7; we omit the details. 

2.4 A tree-shaped parking scheme

We claim that for any fixed $T \in \mathcal{F}_{1,n}$, the process $(\text{Perc}_k(T), 0 \leq k \leq n - 1)$ may be represented as a tree-shaped parking scheme (TSPS). In the usual “linear” parking scheme, the places are
arranged along the line (see Section 2.5). The cars successively choose a place \( c \) and then park at the first free place among \( c, c + 1, c + 2, \cdots \). In the tree-shaped parking scheme, the cars choose a node, and then parks at a free place in the direction to the root. We will need this free place to be the one nearest to the root. More precisely, there are two equivalent versions:

A) Consider a sequence \((V_i, i \geq 1)\) of i.i.d. random variables uniformly distributed among the nodes of \( T \). We will define the places chosen by the successive cars based on this sequence. In words: a car choose a place \( c \) uniformly in the tree and look for unoccupied nodes in the path from \( c \) to the root; if such nodes exist, it parks at the one nearest to the root. If no more place is available, it chooses a new uniform place \( c \), and proceeds as before.

Formally: assume that at time 0, the root \( r \) of \( T \) is the only occupied node. Set \( k_0 = 0 \) and \( \text{TPS}_0(T) = \{r\} \) (the tree reduced to the root); more generally for \( 1 \leq j \leq n - 1 \), let \( k_j = \min \{i > k_{j-1} : V_i \notin \text{TPS}_{j-1}(T)\} \) and denote by \( V'_j \) the node on the path \([r, V_{k_j}]\) which is nearest to \( r \) and not contained in \( \text{TPS}_{j-1}(T) \); then define \( \text{TPS}_{j+1}(T) \) as the smallest tree containing \( \text{TPS}_j(T) \) and \( V'_j \). Clearly, \( (\text{TPS}_j(T), 0 \leq j \leq n - 1) \) is an increasing sequence of subtrees of \( T \). In this version a place is a node.

B) In this second version, a place is an edge instead of being a node. Start from a tree \( T \) whose edges are painted in white. Each time, one picks uniformly at random a white edge \( e \) in \( T \), and color in red the last white edge on the path from \( e \) to the root (note that there is an obvious correspondence between choosing a uniform node distinct from the root and choosing an uniform edge). In this way, we obtain a sequence of subtrees formed by the red edges (see Fig. 2). It is clear that this sequence has the same distribution as \( (\text{TPS}_j(T), 0 \leq j \leq n - 1) \) as defined above.

Once again, one may attribute a labelling \( L^{\text{TPS}}(T) \) to the edges according to their entrance in \( \text{TPS}(T) \): put a label \( L^\text{TPS}_e(T) = n - k \) to the \( k \)th added edge. This labelling belongs to \( DL(T) \). Moreover, we have the following.

**Proposition 10.** For any rooted tree \( T \), the following identities in distribution hold

\[
L^{\text{TPS}}(T) \overset{(d)}{=} L^{\text{Perc}}(T) ; \tag{11}
\]

\[
(\text{TPS}_k(T), 0 \leq k \leq |T|_E) \overset{(d)}{=} (\text{Perc}_k(T), 0 \leq k \leq |T|_E) . \tag{12}
\]

Proof. To choose a subtree \( \tau \) of \( T \setminus \text{Perc}_k(T) \) with probability proportional to its size, one can take a sequence of i.i.d. uniform variables \( (U_i, i \geq 1) \) taking values in the vertex set of \( T \), and keep the first \( U_i \) that falls in \( T \setminus \text{Perc}_k(T) \). The subtree \( \tau' \) of \( T \setminus \text{Perc}_k(T) \) which contains this first \( U_i \) has the right distribution.

\[\square\]

### 2.5 Linear parking scheme representation revisited, and asymptotics

Parking scheme and its relation with additive coalescence is the essence of the paper [13] by Chassaing & Louchard. Here, in contrast with Section 2.4, parking scheme refers to linear/circular parking scheme.
Let us explain some of their considerations before proposing an enriched version of their construction, which will allow us to unify the parking representation with the model of coalescent forests of Pitman and our decreasing trees. The parking contains \( n \) places labelled from 1 to \( n \). The parking is circular so that place \( n+1 \) and 1 coincide: Here, for \( x \in \mathbb{Z}, \pi \) stands for the integer in \([n]\) equaling \( x \) modulo \( n \).

The parking scheme can be presented as an algorithm taking as entry some data, and producing a sequence of parking occupations. The “data” is a vector \((C_i, 1 \leq i \leq n-1)\) taking values in \([n]^{n-1}\). At the beginning all places are free, and a total of \( n - 1 \) cars will park, successively.

**Parking algorithm.**
- At time \( i \), the car \( i \) enters in the parking (and \( i \) varies from 1 to \( n - 1 \)). The variable \( C_i \) is the choice of car \( i \).
- Upon the arrival of the car \( i \), the places already occupied in the parking are \( P_1, \ldots, P_{i-1} \).
- The car \( i \) parks at \( P_i \), the first free place among \( C_i, C_i + 1, C_i + 2, \ldots \).

We may restate the previous condition as: car \( i \) parks at \( P_i := C_i + l_\ast \), where \( l_\ast := \min\{l \in \{0, 1, \ldots, n-1\} : C_i + 1 \notin \{P_1, \ldots, P_{i-1}\}\} \).

In the sequel we will consider the case that the \( C_1, \ldots, C_{n-1} \) are random variables, i.i.d. and uniform in \([n]\).

**Note 11.** It will be useful to consider that the \( C_i \)'s are random variables defined on the probability space \((\Omega^{\text{Par}}, \mathcal{A}^{\text{Par}}, \mathbb{P}^{\text{Par}})\) where \( \Omega^{\text{Par}} = [n]^{n-1}, \mathcal{A}^{\text{Par}} = \text{Powerset}(\Omega^{\text{Par}}) \), and \( \mathbb{P}^{\text{Par}} \), the uniform distribution on \( \Omega^{\text{Par}} \). On this space, the functions \( C_i \) are defined by \( C_i(\omega) = \pi_i(\omega) \) (the extraction of the \( i \)th coordinate). The parking process \((P_i, 1 \leq i \leq n-1)\), as well as the quantities that will be defined in this section, are random variables on this space. For sake of shortness, we will mix considerations for which the \((C_i, 1 \leq i \leq n)\)'s are random variables or just elements of \( \Omega^{\text{Par}} = [n]^{n-1} \).

For any \( i, j \in [n] \), denote by \([i, j]\) the circular interval formed by \(\overline{i, i+1, \ldots, i+k}\) where \( k = j - i \).

Define a block in the parking as a circular interval \([x, x+k]\), maximal for the inclusion partial order, so that \( x+i \) is occupied for \( 0 \leq i < k \), and \( x+k \) is not occupied. With this definition, we have the following description of the parking.

- In the beginning, the blocks are the \( n \) trivial blocks \( I[0] := (I_0^j, 1 \leq j \leq n) \) where \( I_0^j = [j, j] \);
- At all times, the number of blocks is given by the number of empty places. Thus, for \( 0 \leq m \leq n-1 \), when \( m \) cars parked, there are exactly \( n-m \) blocks which we denote by \( I[m] := (I_m^j, 1 \leq j \leq n-m) \).

Notice that \( I[m] \) forms a partition of \([1, n] \): the total block sizes sum to \( n \) for any \( m \), since at each step an empty place is replaced by an occupied one. Write now \(|I[m]|\) for the multiset \(|I_m^j|, 1 \leq j \leq n-m\).

There is some slight difference between the above definition and the original one in [13]. Namely, here the block size is 1 plus the number of the cars in it. Thanks to this small modification, we have an exact correspondence between the parking scheme and the additive coalescent (this point is also discussed in [13, Section 8].) To state the correspondence, we need the following notation. For a partition \( \pi = (\pi_1, \cdots, \pi_k) \) of \([n]\), denote by \(|\pi|\) the multiset \(|\pi_j|, 1 \leq j \leq k\).
Proposition 12. [Chassaing & Louchard [13]] The block sizes in the parking scheme evolve in the same way as the cluster sizes in the additive coalescence:

\[(|I[k]|, 0 \leq k \leq n - 1) \equiv (|\Pi_k^k|, 0 \leq k \leq n - 1).\]  

Let us mention that the proof we propose for Proposition 17 also uses ideas from Chassaing & Louchard [13].

Remark 13. An alternative description for the evolution of the block sizes can be given as follows. For \(m\) varying from 0 to \(n - 2\),

- let \(j \in \{1, \ldots, n - m\}\) be the index such that \(C_{m+1} \in I_j^m\) and let \(I^*\) be the interval which is at the right of \(I_j^m\) on the circle;
- merge \(I_j^m\) and \(I^*\) and relabel the intervals from 1 to \(n - m - 1\).

Chassaing & Louchard remarked a connection between their parking scheme with Pavlov forests. Let us briefly explain this connection. For \(1 \leq m \leq n - 1\) and \(1 \leq i \leq n\), let

\[C_i^m = \{j : C_j = i, 1 \leq j \leq m\}\]

be the set of cars among the first \(m\) ones that have chosen the place \(i\) (and then parked at some position after \(i\) on the circle). By construction, the distribution of \((\#C_i^m, 1 \leq i \leq n)\) is multinomial: for a vector of non negative integers \((c_i, 1 \leq i \leq n)\) satisfying \(\sum_{1 \leq i \leq n} c_i = m\),

\[P(\#C_i^m = c_i, 1 \leq i \leq n) = \binom{m}{c_1, \ldots, c_n} / n^m.\]

Set

\[S_k^m := \sum_{i=1}^{k} (\#C_i^m - 1), \quad \text{for } 0 \leq k \leq n,\]

which is a “bridge” process with exchangeable increments. This is reminiscent of the Lukasiewicz walk for Poisson Galton-Watson forests. Indeed, under the “excursion condition”

\[(\text{ExcCond}) : S_k^{m-1} \geq 0, \quad \text{for } 0 \leq k \leq n - 1 \quad \text{and} \quad S_n^{m-1} = -1,\]

\((S_k^{m-1}, 0 \leq k \leq n)\) has the same distribution as the Lukasiewicz walk of a Poisson Galton–Watson tree conditioned to have size \(n\).

On the other hand, by employing a cyclic permutation (see Remark 16 below for more details), we have the following.

Proposition 14 (Chassaing & Louchard [13]). (13) also holds true conditionally on (ExcCond).

One may associate bijectively a tree with the state of the parking at time \(n - 1\), as follows. The root of the tree is a node labelled 0. For \(1 \leq i \leq n\), \(C_i^{n-1}\) is the set of the labels for the children of the \(i\)th node (in depth-first order); so in particular, the \(i\)th node has \(\#C_i^{n-1}\) children. See Fig. 3 for an example. In a similar way, when only \(m\) cars have parked, the sequence \((C_i, 1 \leq i \leq m)\)
encodes a Pavlov forest with \( n - m \) components (which is a forest of \( n - m \) trees whose roots are labelled by \( r_1, \ldots, r_{n-m} \) and whose non-root nodes are labelled by \([m]\)), \( 1 \leq m \leq n - 1 \) (see also Fig. 3).

Even if the combinatorial correspondence explained above is exact, one may say it is not really satisfactory in the sense that in the additive coalescent, what we add as time passes are edges, and not nodes! For this reason, one might expect \((C_i)\) to encode rather edges than nodes. In what follows we explain how to realize this combinatorial encoding.

### 2.6 A new connection via an enriched parking process

The number of different parking histories is given by the number of vectors \((C_i, 1 \leq i \leq n)\), which is \(n^{n-1}\). A simple counting argument shows that standard parkings can not faithfully encode the process \((F^{\text{PP}}_k, 1 \leq k \leq n - 1)\), nor \(((F^{\text{SC}}_k, L^{\text{SC}}_k), 1 \leq k \leq n - 1)\). Indeed, the number of different forest histories is:

- \(n^{n-1}(n-1)!\) for \((F^{\text{PP}}_k, 1 \leq k \leq n - 1)\) since this is the number of different choices for the directed added edges,
- \(n!(n-1)!\) for \(((F^{\text{SC}}_k, L^{\text{SC}}_k), 1 \leq k \leq n - 1)\) for the same reason,
- to encode the additive coalescent, the pairs of blocks that merge at each moment need to be encoded, so that the number of histories is \(n!(n-1)!/(2^{n-1})\).

Therefore, we need to enrich the parking structure in order to encode the forests processes or
2.6.1 Encoding of additive coalescent by an enriched parking process

To produce an enriched parking process, we take here as data a pair \((\sigma, (C_i, 1 \leq i \leq n-1))\) where as previously \((C_i, 1 \leq i \leq n-1)\) is a vector taken in \([n]^{n-1}\) and \(\sigma\) is a permutation taken in \(S[n]\), the symmetric group over \([n]\).

<table>
<thead>
<tr>
<th>Enriched parking algorithm.</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Run the parking algorithm defined in Section 2.5 using the sequence ((C_i, 1 \leq i \leq n-1)) and obtain a sequence ((I[m], 0 \leq m \leq n-1)) of partitions of ([n]).</td>
</tr>
<tr>
<td>- For (0 \leq m \leq n-1), define</td>
</tr>
</tbody>
</table>
| \[CC^m_j := \{\sigma_i, i \in I^m_j\} \quad \text{and} \quad CC[m] = \{CC^m_j, 0 \leq j \leq n-m\}, \]
| the connected components at time \(m\). For each \(m\), \(CC[m]\) is a partition of \([n]\). |

In the enriched parking, \(I[m]\) encodes the connected components but are not themselves connected components, this role being taken by \(CC[m]\): the parking places are still labelled from 1 to \(n\), but the \(i\)th place of the parking corresponds to the element/node \(\sigma_i\). For those \(CC^m_j\) defined in (18), we will sometimes say that \(CC^m_j\) is above \(I^m_j\).

In the sequel we will consider the case that the \(C_1, \ldots, C_{n-1}\) are i.i.d. random variables uniformly distributed in \([n]\), and that \(\sigma \sim \text{Uniform}(S[n])\) and independent from the \(C_i\)'s. In particular, this will ensure the exchangeability of the parking blocks in our construction.

**Note 15.** As in Note 11, it is convenient to consider that \((\sigma, (C_1, \ldots, C_{n-1}))\) are defined on a probability space \((\Omega^{\text{EPar}}, \mathcal{A}^{\text{EPar}}, \mathbb{P}^{\text{EPar}})\) where \(\Omega^{\text{EPar}} = [n]^{n-1} \times S[n]\), \(\mathcal{A}^{\text{EPar}} = \text{Powerset}(\Omega^{\text{EPar}})\), and \(\mathbb{P}^{\text{EPar}}\) is the uniform distribution on \(\Omega^{\text{EPar}}\). On this space, the functions \(C_i\) are defined by \(C_i(\omega \times \omega') = \pi_i(\omega)\) (the extraction of the \(i\)th coordinate “of the \([n]^{n-1}\) component”), \(\sigma(\omega \times \omega') = \omega'\). Hence, the \(C_i\)'s and \(\sigma\) are random variables on this space, and the other quantities defined in the algorithms (for example \(CC[m]\)) are also random variables defined on this probability space.

**Remark 16.** [Equivalence classes of enriched parkings] Let us say that two pairs of data \((\sigma, (C_i, 1 \leq i \leq n-1))\) and \((\tilde{\sigma}, (\tilde{C}_i, 1 \leq i \leq n-1))\) are equivalent if there exists some \(k \in [n]\) such that

\[
\begin{align*}
\sigma(k + j) &= \tilde{\sigma}(j), & \text{for any } j \in [n], \\
\frac{k + C_j}{k + C_j'} &= \tilde{C}_j, & \text{for all } 1 \leq j \leq n - 1.
\end{align*}
\]
Then each equivalence class contains exactly \( n \) elements. To distinguish, we add \( \tilde{\sigma} \) to the variables \( CC[m], C_i^m \) and \( S_i^m \) which are defined using \( (\tilde{\sigma}, (C_j, 1 \leq j \leq n-1)) \). It is clear that

\[
CC[m] = \tilde{CC}[m], \quad \text{for } 0 \leq m \leq n-1.
\]

Namely, equivalent data produces same connected component processes, since their enriched parking processes are equal up to a rotation. Moreover, by taking \( k = \min\{j \in \{1, \cdots, n\} : S_j^{n-1} = \min S^{n-1}\} \), we find exactly one pair of \( (\sigma, (C_j, 1 \leq j \leq n-1)) \) in each equivalence class for which \((\text{ExcCond})\) is obeyed. This trick is classical and sometimes referred to as the rotation principle or the cyclic lemma in the literature (see Dvoretzky & Motzkin [15]). In particular, conditioning on \((\text{ExcCond})\) does not change the distribution of \((CC[m], 0 \leq m \leq n-1)\) by the previous argument (a parking under condition \((\text{ExcCond})\) is called “confined parking sequence” in [13].)

Notice that the block sizes \(|I_m|, 0 \leq m \leq n-1\) are the same for the parking algorithm and for the enriched version, so that Propositions 12 and 14, as well as the representation provided by Remark 13 hold for the enriched version. In contrast to \(|I_m|\), which are just some interval lengths, \(CC[m]\) are partitions, with exchangeable elements. As a consequence, we have the following, which reinforces Propositions 12 and 14.

**Proposition 17.** The following distributional equality holds

\[
(CC[m], 0 \leq m \leq n-1) \overset{\text{(d)}}{=} (\Pi_m^+, 0 \leq m \leq n-1).
\]

Further, (20) also holds conditionally on \((\text{ExcCond})\).

**Proof.** For \( m \in [n-2] \), let us show that conditional on \(CC[j], j \leq m\), two blocks \(CC_j^m\) and \(CC_i^m\) of \(CC[m]\) merge at time \( m \) with probability \((|CC_j^m| + |CC_i^m|)/(n(n-m-1))\). By the definition of \(\Pi^+\), this will entail (20). Intuitively, the proof is the same as that of Proposition 12 since its relies on an exchangeability property in the parking. In [13], a counting argument in provided. Here, we will instead, detail a symmetry argument.

First, notice that the sequence of connected components \((CC_j^m, 1 \leq j \leq n-m)\) may be equipped with a total order inherited from the order between their minimal elements: \(CC_j^m < CC_i^m \iff \min(CC_j^m) < \min(CC_i^m)\). Up to a relabelling of the \(CC[m]\), we may and will assume that the index \( j \) in \((CC_j^m, 1 \leq j \leq n-m)\) corresponds to the rank in this order. Recall that the \(CC_j^m\)'s correspond to some intervals \(I_j^m\)'s (we said the \(CC_j^m\) were above them).

For any \(s_1, \cdots, s_k\) positive integers summing to \(n\), let \(X[s_1, \cdots, s_k]\) be the set of \(k\)-tuples \((I_1, \cdots, I_k)\) of intervals that forms a partition of \([1, n]\), and such that \(|I_j| = s_j\) for any \(j\). The formula,

\[
\#X[s_1, \cdots, s_k] = nk!/k = n(k-1)!
\]

comes from the fact that given the sizes of the intervals, their respective order around the circle is given by a permutation considered up to a rotation (this provides the factor \((k-1)!\)), and then
their position along the circle is specified by fixing the position of the first one, which gives the additional factor $n$.

Now, we claim that the distribution of $I[m] := (I_j^m, 1 \leq j \leq m)$ conditionally on $CC[m]$ is uniform in $X[\lfloor CC_1^m \rfloor, \ldots, \lfloor CC_n^m \rfloor]$: for any element $I_{[n-m]}^* := (I^*_j, 1 \leq j \leq n - m)$ in this set
\[
\mathbb{P}(I[m] = I_{[n-m]}^* \mid CC[m]) = \frac{1}{n(n - m - 1)!}.
\tag{22}
\]
This can be proved by computing the number of pairs $(\sigma, (C_j, 1 \leq j \leq m))$ for which the partitions above a sequence of intervals $I_{[n-m]}^* \in X[\lfloor CC_1^m \rfloor, \ldots, \lfloor CC_{n-m}^m \rfloor]$ are exactly the $CC_j^m$'s, and to see that this number is the same for any $I_{[n-m]}^*$ in this set.

But since $(\sigma, (C_j, 1 \leq j \leq m))$ is uniformly distributed on $S[n] \times [n]^m$, to prove (22), we can just show that the number of pairs $(\sigma, (C_j, 1 \leq j \leq m)) \in S[n] \times [n]^m$ which produce $CC[m]$ above $I_{[n-m]}^*$ are above $I_{[n-m]}'$ are the same for any $I_{[n-m]}^*, I_{[n-m]}' \in X[\lfloor CC_1^m \rfloor, \ldots, \lfloor CC_{n-m}^m \rfloor]$. A symmetry argument suffices:

- first, $\sigma$ and $(C_j, 1 \leq j \leq m)$ are both invariant by translation:
  \[
  ((\sigma_i, 1 \leq i \leq n), (C_i, 1 \leq i \leq n))^{(d)} = ((1 + \sigma_i, 1 \leq i \leq n), (1 + C_i, 1 \leq i \leq n)).
  \]

- second, the probability of an interval configuration is unchanged if we permutate two adjacent intervals: consider two consecutive intervals $I = [a, b - 1], I' = [b, c]$ and the map $f$ defined by $f(x) = x$ if $x \not\in I \cup I'$, $f(x) = a + x - b$ if $x \in I'$, $f(x) = x + c - b$ if $x \in I$, then again
  \[
  ((\sigma_i, 1 \leq i \leq n), (C_i, 1 \leq i \leq n))^{(d)} = ((\sigma_{f(i)}, 1 \leq i \leq n), (C_{f(i)}, 1 \leq i \leq n)).
  \]
These two properties ensure that the intervals below the $CC[m]$ are exchangeable, and invariant by translation.

Next, in the enriched parking algorithm, two blocks $CC_j^m$ and $CC_i^m$ will be merged if the corresponding intervals $I_j^m$ and $I_i^m$ are neighbors and if $C_{m+1}$ falls in the first one (around the circle). Since (22) implies that $I_j^m$ being next to $I_i^m$ takes place with probability $1/(n - m - 1)$. Taking into account that $I_j^m$ can be to the left or to the right of $I_i^m$, we find that $C_{m+1}$ will produce a merging of $I_j^m$ and $I_i^m$ with probability $(|I_j^m| + |I_i^m|)/(n(n - m - 1))$, that is, proportional to $(|I_j^m| + |I_i^m|)$. Finally, (20) also holds conditionally on (ExcCond), since we have seen in Remark 16 that the conditioning does not change the distribution of $(CC[m], 0 \leq m \leq n - 1)$.

Now, we introduce two constructions of graphs where $CC[m]$ will appear as vertex sets of the connected components there. In consequence, we obtain, on the probability space where $(C_i, 1 \leq i \leq n) \leq n - 1$ and $\sigma$ are defined, a coupling between the enriched parking process, Pitman’s forests $(F_k^{PF}, 0 \leq k \leq n - 1)$ and the forests with decreasing labelling $(F_k^{SC}, L_k^{SC}), 0 \leq k \leq n - 1$.

### 2.6.2 Encoding of Pitman forest

For $0 \leq m \leq n - 1$, we associate with the partition $CC[m]$ defined in (18) a forest belonging to $F_{n-m,n}$, whose tree components induce the same partition of $[n]$ as $CC[m]$. We then show that the sequence of the forests is distributed as $(F_k^{PF}, 0 \leq k \leq n - 1)$.
Recall that $\sigma \sim \text{Uniform}(S[n])$ and $C_i, 1 \leq i \leq n - 1$, are i.i.d. uniformly distributed in $[n]$. For a circular interval $I$, set $L(I)$ to be its leftmost element of $I$, which is the element $\ell \in I$ such that $\ell - 1 \notin I$. At the beginning the intervals $I^0_j$'s are the trivial intervals $[j, j]$, for $1 \leq j \leq n$.

**Algo 1.**

For $m$ varying from $0$ to $n - 2$, do the following:

- Let $j \in \{1, \ldots, n - m\}$ be the index such that $C_{m+1} \in I^m_j$ and let $I^* \text{ be the next one on the circle}$.
- Add an edge from $\sigma_{L(I^*)}$ to $\sigma_{C_{m+1}}$ and label it $m + 1$ (see the second line in Fig. 4 for an example).
- Relabel the intervals from $1$ to $n - m - 1$.

![Image of Algo 1](https://via.placeholder.com/150)

**Figure 4:** In the first line, the numbers above $\sigma_k$ are the labels of cars which have chosen the place $k$. In this example, $C_1 = C_2 = C_8 = 4, C_2 = 3, C_4 = 5, C_5 = C_6 = 1, C_7 = 7$. In the second line, an illustration for constructing Pitman’s forest: add an edge between $\sigma_{C_{m+1}}$ and the left endpoint of the next interval. Turn the picture clockwise by $\pi/2$ degrees and obtain the final tree of Pitman’s forest. Discarding the edges with labels $> m$ gives the forest after $m$ coalescences. In the third line, the illustration for the construction of $T^{SC}$: add an edge between the right endpoint of the connected component containing $\sigma_{C_{m+1}}$ and the right endpoint of the next interval. Turn the picture clockwise by $\pi/2$ degrees and obtain the canonical planar embedding of a rooted tree equipped with decreasing edge-labels. Discarding the edges with labels $> m$ gives the corresponding forest after $m$ coalescences. The left-depth first traversal of the tree is given by $(\sigma_9, \sigma_8, \ldots, \sigma_1)$. Observe that the connected components of both constructions induce the same partitions of the nodes.

It is easy to see that **Algo 1** produces an increasing sequence of forests, since the edges go from right to the left and only at most one outgoing edge comes out at each position.
Proposition 18. Algo 1 produces a sequence of labelled forests which has the same distribution as \((F_k^{PF}, 0 \leq k \leq n - 1)\) introduced just above Theorem 2\([29] , \text{Theorem 4}\). Denote this sequence of forests also by \((F_k^{PF}, 0 \leq k \leq n - 1)\). If we write \(\text{CC}(F_k^{PF})\) for the partitions induced by the connected components of \(F_k^{PF}\), we get \((\text{CC}(F_k^{PF}), 0 \leq k \leq n - 1) = (\text{CC}[k], 0 \leq k \leq n - 1)\).

Proof. This is an immediate consequence of the fact that Algo 1 encodes exactly Pitman construction of \((F_k^{PF}, 0 \leq k \leq n - 1)\) using a linear parking. To figure out this fact, observe Fig. 4. At time 0, no edge is present, each node \(\sigma_j\) is considered alone. The addition of edges produces block mergings, but at any time \(m\), each block \(b\) corresponds to a tree \(t\), and the leftmost node \(u\) of \(b\) is the root of \(t\). Step 2 of Algo 1 preserves this property all along the construction. Now let us work only on the set of trees \(f_m = \{t_1, \ldots, t_{n-m}\}\) present at time \(m\), and consider an edge addition. The node \(\sigma_{C_{m+1}}\) is uniform on the set of nodes. It corresponds then to the choice of the uniform node \(\alpha_{m+1}\) in the Pitman forest. Now, (22) in the proof of Proposition 17 implies that \(\sigma_{L(I^*)}\), which is the tail of the edge to be added, is distributed as the root of a uniform tree in the remaining forest, namely, it has the same distribution as \(\beta_{m+1}^{PF}\). \(\square\)

2.6.3 Encoding of the forests with decreasing labellings

Recall that \((\text{CC}[k], 0 \leq k \leq n - 1)\) has been defined in Section 2.6.1 using \(\sigma \sim \text{Uniform}(\mathcal{S}[n])\) and \(C_i, 1 \leq i \leq n - 1, \text{i.i.d. and uniformly distributed in } [n]\). We construct a sequence of forests such that \(\text{CC}[k]\) appear as the vertex sets of the connected components in the \(k\)th forest. And we will see that the sequence coincides in distribution with \(((F_k^{SC}, L_k^{SC}), 0 \leq k \leq n - 1)\) as defined in Section 2.1.

For a circular interval \(I\), let \(R(I)\) be its rightmost element, i.e. the element \(r \in I\) such that \(r + 1 \not\in I\). At the beginning the intervals \(I_j^0\) are the trivial intervals \([j, j]\) for \(1 \leq j \leq n\).

\[
\text{Algo 2.}
\]

For \(m\) varying from 0 to \(n - 2\) do the following:
- Let \(j \in \{1, \ldots, n - m\}\) be the index such that \(C_{m+1} \in I_j^m\), and let \(I^*\) be the next one around the circle.
- Add an edge from \(\sigma_{R(I_j^m)}\) to \(\sigma_{R(I^*)}\) and label the edge \(m + 1\).
- Relabel the intervals from 1 to \(n - m - 1\).

Again, it is easy to see that Algo 2 produces an increasing sequence of forests, since the edges go from left to the right and only at most one outgoing edge comes out at each position. Further, since added edges are labelled from the first one 1 to the last one \(n - 1\), the forests constructed are at each moment equipped with a decreasing edge-labelling.

Proposition 19. Algo 2 produces a sequence of labelled forests which has the same distribution as \(((F_k^{SC}, L_k^{SC}), 0 \leq k \leq n - 1)\) described in Section 2.1. Denote this sequence of forests also by \(((F_k^{SC}, L_k^{SC}), 0 \leq k \leq n - 1)\). If we write \(\text{CC}(F_k^{SC})\) for the partitions induced by the connected components of \(F_k^{SC}\), we get \((\text{CC}(F_k^{SC}), 0 \leq k \leq n - 1) = (\text{CC}[k], 1 \leq k \leq n - 1)\).
Proof. Again this is a consequence of the fact that Algo 2 is just a linear encoding of the construction of \((F^\text{SC}_k, 0 \leq k \leq n - 1)\). As in the proof of Proposition 18, observe Fig. 4. All along the construction, each block \(I\) in the parking corresponds to a tree. The set \(\{\sigma_i, i \in I\}\) coincides with the vertex sets of the corresponding tree, and moreover, \(\sigma_{R(I)}\), “the right most node of each block” is the root of the corresponding tree (this property is guarantied by the second point of Algo 2). The rest of the proof can be easily adapted from that of Proposition 18. \qed

2.7 Trees with decreasing labellings and canonical orderings

We may define several orders on a Cayley tree equipped with a decreasing edge-labelling \((T, L)\). Using the edge labelling, we may order the edge-children of each nod. This induces an order on the nodes, that we call, canonical order. Using this order, we can “draw” a tree in the plane, drawing the in-going edges of each node in such a way that the edge labels are increasing from left to right. Call this the canonical embedding of \(T\) and denote it by \(\text{Can}(T, L)\). Then the depth-first search of \(\text{Can}(T, L)\) defines a linear order on the nodes of \(T\) which we call the left-depth first traversal of \(T\). Algo 2 produces a decreasing tree with a given left-depth first traversal. Indeed, using the fact that in Algo 2 edges are always added between the right endpoints of two neighboring intervals, one can readily check that the left-depth first traversal of the tree \(T^{\text{SC}}\) defined by Algo 2 is given by \((\sigma_{j_0 + 1 - j}, 1 \leq j \leq n)\), where \(\sigma_{j_0}\) is the root of \(T^{\text{SC}}\) (see the last line in Fig. 4). This leads to the following.

Proposition 20. Let \(T \in F_{1, n}\) and \(L \in D\ell(T)\). The number of elements \((\sigma, (C_j, 1 \leq j \leq n - 1)) \in \Omega^{\text{Epar}}\) (see Note 15) which produces \((T, L)\) in Algo 2 is given by \(n \prod_{u \in T \setminus \{r(T)\}} |\text{Sub}_u(T)|\).

Notice that the arguments in Remark 16 also entail that elements in the same rotation class produce the same sequence \(((F^\text{SC}_k, L^\text{SC}_k), 0 \leq k \leq n - 1)\) in Algo 2. Thus, the number of rotation classes that are sent onto \((T, L)\) by Algo 2 is \(\prod_{u \in T \setminus \{r(T)\}} |\text{Sub}_u(T)|\). This is consistent with Lemma 5 which says that \#\(D\ell(T) = (n - 1)!/\prod_{u \in T \setminus \{r(T)\}} |\text{Sub}_u(T)|\) so that the total number of rotation classes is indeed \((n - 1)!n^{n - 1}\) as expected.

Proof. Since elements in the same rotation class produce the same \((T^{\text{SC}}, L^{\text{SC}})\) by Algo 2, consider the unique element \((\sigma, (C_i, 1 \leq i \leq n))\) of the rotation class such that the root of \(T^{\text{SC}}\) is \(\sigma_n\). In this case, Algo 2 produce \((T, L)\) from \((\sigma, (C_i, 1 \leq i \leq n))\) if and only if \(\sigma_{C_j} \in \text{Sub}_{u_j}(T)\) where \(u_j\) is the node from which the edge labelled \(j\) is directed, for \(1 \leq j \leq n - 1\). As each non-root node of \(T\) appears once in \((u_j, 1 \leq j \leq n - 1)\), we easily conclude. \qed

We may explain a bit further, the link between \((T^{\text{SC}}, L^{\text{SC}})\) and \((\sigma, (C_j, 1 \leq j \leq n))\), in describing precisely where the \((m + 1)\)th edge is added. Recall that \(C^m_j = \{i : C_j = i, 1 \leq j \leq m\}\) and that \(S^m_k = \sum_{i=1}^k (\#C^m_i - 1)\). We further suppose \((\text{ExcCond})\). We recognize \((S^m_k, 0 \leq k \leq n)\) as the Lukasiewicz walk for Pitman’s forest \(F^\text{PF}_s\). Under the condition \((\text{ExcCond})\), the partition \(I[m]\), which encodes the connected components of \(F^\text{PF}_m\), corresponds to the excursion intervals of \(S^m - \Sigma^m\), where \(\Sigma^m_j := \min\{S^m_i : i \leq j\}\). But \(F^\text{PF}_m\) and \(F^\text{SC}_m\) share the same connected components
by construction. It follows that the rightmost node of the block which contains \( C_{m+1} \) is \( \sigma_{R_{m+1}} \), where
\[
R_{m+1} := \min\{ j \geq C_{m+1} : S_j^m = S_j^m \}.
\]
Similarly, the rightmost node of the next block is given by \( \sigma_{R'_{m+1}} \), where
\[
R'_{m+1} := \min\{ j \geq R_{m+1} + 1 : S_j^m = S_j^m \}.
\]
Then the edge labelled \( m + 1 \) is from \( \sigma_{R_{m+1}} \) to \( \sigma_{R'_{m+1}} \).

### 2.8 Coupling between \((F_k^{PF}, 0 \leq k \leq n - 1)\) and \(((F_k^{SC}, L_k^{SC}), 0 \leq k \leq n - 1)\)

Everything has already been said on this point. We collect in this section some simple facts spread throughout the paper.

In Sections 1.1 and 1.2, the two processes \( F_k^{PF}[n - 1] = (F_k^{PF}, 0 \leq k \leq n - 1) \) and \( F_k^{SC}[n - 1] = ((F_k^{SC}, L_k^{SC}), 0 \leq k \leq n - 1) \) have been defined using respectively \( (\alpha_k^{SC}, \beta_k^{SC}) \) and \( (\alpha_k^{PF}, \beta_k^{PF}) \). In Sections 2.6.2 and 2.6.3, relying on the enriched parking scheme and using as data \( (\sigma, (C_i)_{1 \leq i \leq n-1}) \sim \text{Uniform}(S[n] \times [n]^{n-1}) \), we obtain an alternative construction of \( F_k^{PF}[n - 1] \) (resp. of \( F_k^{SC}[n - 1] \)). It follows that both \( F_k^{PF}[n - 1] \) and \( F_k^{SC}[n - 1] \) are defined on the same probability space where \( (\sigma, (C_i)_{1 \leq i \leq n-1}) \) is defined. Observe that this coupling between \( F_k^{PF}[n - 1] \) and \( F_k^{SC}[n - 1] \) amounts to
- taking \( \alpha_k^{SC} = \alpha_k^{PF} \),
- taking \( \beta_k^{SC} \) to be the root of the tree in \( F_k^{SC} \) which contains the node with label \( \beta_k^{PF} \).

We assume from now on that \( F_k^{PF}[n - 1] \) and \( F_k^{SC}[n - 1] \) are coupled in this way. Observe that at each instant, the partitions induced by the both forests coincide, and new edges are added between the same pairs of connected components.

The information contained in \( F_k^{SC}[n - 1] \) is not sufficient to reconstitute the rotation class of \( (\sigma, (C_j, 1 \leq j \leq n - 1)) \). The reason is that since we add the edge between the roots of trees, then the \( \alpha_k^{SC} \)'s (which corresponds to the \( C_i \)'s) can not be recovered from \( F_k^{SC}[n - 1] \). Notice that \( F_k^{PF}[n - 1] \) is a sequence of forests without edge-labelling. On the other hand, since the difference between \( F_k^{PF} \) and \( F_k^{PF'} \) is the added edge at time \( k \), this process allows one to recover the sequence \( ((\alpha_k^{SC}, \beta_k^{SC}), 1 \leq k \leq n - 1) \). Hence, \( F_k^{PF}[n - 1] \) characterizes \( F_k^{SC}[n - 1] \), but the converse is false.

Nevertheless, the distribution of \( F_k^{PF}[n - 1] \) knowing \( F_k^{SC}[n - 1] \) can be described: assume that we have been able to construct a sequence of forests \( (F_k^{PF'}, 0 \leq k \leq m) \) thanks to \( ((F_k^{SC}, L_k^{SC}), 0 \leq k \leq m) \) for some \( m \), such that both forests have the expected distribution. Assume moreover that the partition induced on \([n]\) by these forests coincides at time \( m \) (at time 0, the property is clear).

Now, observe in \((F_{m+1}^{PF'}, L_{m+1}^{PF'})\) the edge labelled \( m + 1 \). Its goes from the root \( u_k \) of a tree to the root \( v_k \) of another tree. By construction, we know that \( u_k = r(\alpha_k^{SC}) \), and \( v_k = \beta_k^{SC} \). Consider in \( F_k^{PF'} \) the trees \( t_k \) which contains \( u_k \), and \( t'_{k} \) which contains \( v_k \) (recall that the nodes are labelled).

To construct \( F_k^{PF'} \), add an edge from the root of \( t'_{k} \) to the node \( w_k \) chosen uniformly in \( t_k \).

### 2.9 Asymptotics

The asymptotic of additive coalescent processes is known to be described by the Brownian excursions as follows. Let \( e = (e(x), 0 \leq x \leq 1) \) be the normalized Brownian excursion of length 1,
and consider the operator
\[ \Psi : C[0,1] \rightarrow C[0,1], \quad f \mapsto \Psi(f) : [0,1] \rightarrow \mathbb{R}^+ \quad x \mapsto f(x) - \min\{f(y), y \leq x\}. \] (23)

For any \( \lambda \geq 0 \), let \( e_\lambda \) be the process
\[ e_\lambda(x) = e(x) - \lambda x, \quad x \in [0,1]. \] (24)

Let \( f \in C[0,1] \). An interval \( I = [a,b] \) with \( b > a \) is said to be an excursion of \( f \) above its minimum if \( b = \inf\{t > a, f(t) = f(a)\} \) and if \( f(a) = f(b) = \min\{f(t), t \leq b\} \). Let \( EL^+(f) \) be the sequence of excursion lengths of \( f \) above its minimum, sorted in decreasing order, and completed by an infinite list of zeros (if the number of excursions is finite).

Denote by \( \gamma^+(\lambda) \) the sequence of lengths of excursions of \( \Psi(e_\lambda) \) sorted in decreasing order so that \( \gamma^+(\lambda) \in \ell_1^\downarrow \), defined in (5). Consider for \( n \) fixed and for any \( \lambda > 0 \), the sequence \( \gamma^n(\lambda) \) defined as the rearrangement of \( \left|C_j^{[n-\lambda \sqrt{n}]}/n, 1 \leq j \leq \lfloor \lambda \sqrt{n} \rfloor \right| \) in decreasing order. This finite sequence completed by a sequence of zeros, is considered as an element of \( \ell_1^\downarrow \).

**Theorem 21.** The following convergence holds in \( \mathbb{D} \left([0, +\infty), \ell_1^\downarrow \right) \)
\[ (\gamma^n(\lambda), \lambda \geq 0) \xrightarrow{(d)}_{n \rightarrow \infty} (\gamma^+(\lambda), \lambda \geq 0). \] (25)

Several proofs of this theorem already exist in the literature. Evans & Pitman [16] showed that the additive coalescent process is a Feller Markov process. Aldous & Pitman [5] gave a description of the fragmentation associated with the additive coalescent by studying the fragment sizes of a cutting process on the Brownian CRT (however, no connection was made yet with \( e_\lambda \)). The above representation of the standard additive coalescent in terms of the excursion lengths of \( f \) above its minimum, sorted in decreasing order, and completed by an infinite list of zeros (if the number of excursions is finite).

In Broutin & Marckert [11], the edges \( (e \in E(T)) \) of a uniform Cayley tree \( T \) are labelled by some i.i.d. uniform random variables \( U[T] = (U_e, e \in E(T)) \) on \([0,1]\). Denote the forest \( F_n(t) \) obtained by discarding the edges with label greater than \( t \). The process \( (F_n(t), t \in [0,1]) \) coincides with the additive coalescent, up to a time change. In [11], it is explained that the Prim order on the tree \( T \) equipped with \( U[T] \) can be used to defined a traversal of \( T \), and further a parking process. The block sizes in the parking is then a faithful encoding of the tree sizes in the forest process \( F_n(t) \), and the coalescence coincides with the coalescence of adjacent parking blocks.

In Propositions 18 and 19, we observed that each of **PF Algo** and **SC Algo** provides a coupling between a parking process and some process of coalescing forests. The three processes, in turn, encode the additive coalescent. However, the fact that the parking/forest constructions do not need a pre-labelling of the edges of \( T \) makes of our present construction a structure a bit more easy to handle than that in [11], at least for the asymptotic behavior. For sake of completeness, but also
to illustrate the advantage obtained in terms of argument concision, we provide a short proof of
Theorem 21 below.

Consider again the random walk \((S_{k}^{n-1}, 0 \leq k \leq n)\) defined in (16), conditional on \((\text{ExcCond})\).

**Remark 22.** In fact, the condition \((\text{ExcCond})\) is needed here contrary to most of the places in
the paper, because under this condition, the excursions length of \(S^{m}\) above its current minimum
coincide with the connected component sizes, while it would be false – because of the first and last
excursion – without this condition.

First, recall the following classical result due originally to Kaigh [20]. Conditional on \((\text{ExcCond})\):
\[
\left( \frac{1}{\sqrt{n}} S_{n}^{m-1} \right)_{t \in [0,1]} \xrightarrow{(d) n \to \infty} e \quad \text{in } C([0,1], \mathbb{R}),
\]
where \(S_{n}^{m-1}\), as well as the other processes \(S^{m}\) that will be considered below, is interpolated between
integer points.

Recall that \(C_{j}^{m} = \{ i : C_{i} = j, 1 \leq i \leq m \}\). Clearly, \(#C_{j}^{m}\) is non decreasing in \(m\) for any \(j\). Let
\[
D_{j}^{m} = \sum_{i=1}^{j} \left( \#C_{i}^{n-1} - \#C_{i}^{m} \right), \quad \text{for any } j \in \{0, \cdots, n\},
\]
which counts the number of cars which have chosen a place left to \(j\) after time \(m\). Then,
\[
S_{j}^{m} = S_{j}^{n-1} - D_{j}^{m} \quad \text{for any } j \in \{0, \cdots, n\}.
\] (27)

Notice that \(D_{j}^{m}\), conditional on \(\{C_{i}^{n-1}, 1 \leq i \leq n\}\) is hypergeometrically distributed :
\[
\mathbb{P} \left( D_{j}^{m} = k \mid \#C_{i}^{n-1}, 1 \leq i \leq n \right) = \frac{\binom{j+S_{j}^{n-1}}{k} \binom{n-1-j-S_{j}^{n-1}}{n-m-k}}{\binom{n}{n-m}}.
\] (28)
The formula comes from the fact that \(j + S_{j}^{n-1} = \sum_{i=1}^{j} \#C_{i}^{n-1}\) and \(n-1-j-S_{j}^{n-1} = \sum_{i=j+1}^{n} \#C_{i}^{n-1}\).

Set
\[
s_{n,\lambda}(t) := \frac{1}{\sqrt{n}} S_{n t}^{n-\lambda \sqrt{n}}, \quad \lambda \geq 0, \quad t \in [0,1].
\] (29)

where the discrete process \((S_{y}^{x}, 0 \leq x \leq m, 0 \leq y \leq n - 1)\) is obtained by interpolating \(S_{k}^{m}\), first
“in \(k\)” between integer points, and then in \(m\) between integer points, in order to get a process
continuous at the end.

**Theorem 23.** Take some \(a \in (0, \infty)\) and \(\Lambda = [0, a]\). Conditionally on \((\text{ExcCond})\),
\[
(s_{n,\lambda}(t))_{(t,\lambda) \in [0,1] \times \Lambda} \xrightarrow{(d) n \to \infty} (e_{\lambda}(t))_{(t,\lambda) \in [0,1] \times \Lambda},
\]
in \(C([0,1] \times \Lambda, \mathbb{R})\).

24
Proof. Let us work on a probability space where the convergence in (26) holds almost surely, in other words on this space, \( s_{n,0} \xrightarrow{\text{as.}} e \). This space exists thanks to the Skorokhod representation theorem. We claim that, for \( \lambda > 0 \) fixed, in \( C[0,1] \),

\[
\left( \frac{D_{nt}^{[n-\lambda\sqrt{n}]}\lambda}{\sqrt{n}} \right)_{t \in [0,1]} \xrightarrow{n \to \infty} (\lambda t)_{t \in [0,1]}.
\] (31)

There are two arguments:

- First, since \( t \mapsto D_{nt}^{[n-\lambda\sqrt{n}]} \) is non decreasing, the convergence in the uniform topology will follow from the pointwise convergence (this is a consequence of Dini’s theorem).
- For fixed \( t \), the convergence (in distribution) of \( D_{nt}^{[n-\lambda\sqrt{n}]} \) to \( \lambda t \) is a combined consequence of the hypergeometric representation (28) and the tightness of \( S_{nt}^{n-1}/\sqrt{n} \) (implied by (26)).

By (26), (27) and (31), we deduce that for a fixed \( \lambda \), in \( C[0,1] \),

\[
(s_{n,\lambda}(t))_{t \in [0,1]} \xrightarrow{n \to \infty} (e_{\lambda}(t))_{t \in [0,1]}.
\] (32)

On the probability space we are working in, the limit depends only on the limit \( e \), the a.s. limit of \( s_{n,0} \). In other words, conditional to this limit, the limit of \( s_{n,\lambda} \) is deterministic, so that the uni-dimensional a.s. convergence stated in (32) implies the following joint convergence: for any \( 0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_k \) one has

\[
\left[ (s_{n,\lambda_j}(t))_{t \in [0,1]}, 0 \leq j \leq k \right] \xrightarrow{\text{as.}} \left[ (e_{\lambda}(t))_{t \in [0,1]}, 0 \leq j \leq k \right].
\] (33)

This time, using the continuity and monotonicity of the limit in \( \lambda \), we conclude. \( \square \)

Proof of Theorem 21. The sizes of the excursions of \( s_{n,\lambda}(t) \) above its running minimum do not exactly correspond to the component sizes of \( [CC_j^{[n-\lambda\sqrt{n}]}]/n, 1 \leq j \leq \lambda\sqrt{n}] \). A slight adjustment has to be made: we say that \( I = [a,b] \) is an excursion of \( s_n \) above its minimum if \( a,b \in \{0,1/n, \cdots, n/n\} \), and if \( b = \inf\{t > a, s_n(t) = s_n(a) - 1/\sqrt{n}\} \). We denote by \( EL_{\downarrow}^j(f) \) be the list of excursion lengths of \( f \) above its minimum, for this definition, sorted in decreasing order.

Consider again the convergence stated in (30) and assume that we are still working on a probability space on which this convergence holds a.s.. For any \( \lambda \geq 0 \), let

\[
\text{locm}(\lambda) = \{ x \in [0,1], e_{\lambda} \text{ has a local minima at } x \}. \] (34)

If the values of \( e_{\lambda} \) on the set of local minima are different (for \( x \neq x' \in \text{locm}(\lambda) \), \( e_{\lambda}(x) \neq e_{\lambda}(x') \)), then one can prove using the arguments of Aldous [4, Section 2.3] that

\[
\left( s_{n,\lambda} \xrightarrow{n \to \infty} e_{\lambda} \right) \Rightarrow \left( EL_{\downarrow}^j(s_{n,\lambda}) \xrightarrow{n \to \infty} EL_{\downarrow}^j(e_{\lambda}) \right). \] (35)

This convergence is a bit weaker than the conclusion of Theorem 21. It is somehow similar to the convergence of finite dimensional distributions, even if the set of \( \lambda \in [0,a] \) for which the condition
stated above on the local minima, has a.s. Lebesgue measure \( a \). As discussed in Broutin & Marckert [11, Section 7], to end the proof we may proceed as follows: denote by \( F^k_n(\lambda) \) the sum of the \( k \) largest excursion sizes in \( s_{n,\lambda} \) and \( F^k(\lambda) \) the sum of the \( k \) largest excursion sizes of \( \xi_\lambda \). The process \( F^k_n(\lambda) \) converges simply almost everywhere to \( F^k(\lambda) \), and the process \( \lambda \mapsto F^k(\lambda) \) is non increasing, càdlàg, which is also the case for \( \lambda \mapsto F^k_n(\lambda) \). For this class of processes, simple convergence implies convergence in \( \mathbb{D}([0,1],\mathbb{R}) \), so that

\[
\left( (F^k_n(\lambda), k \leq K)_{\lambda \in [0,a]} \right) \xrightarrow{d_{n \to \infty}} \left( (F^k(\lambda), k \leq K)_{\lambda \in [0,a]} \right) \quad (36)
\]
in \( \mathbb{D}([0,a],\mathbb{R}^K) \). The property stated in (35) allows us to conclude. \( \square \)

3 Dual fragmentation processes and numbers of cuts

In this part, we are interested in the time-reversal of the two forest coalescent processes: \( (F^\text{PF}_{n-1-k}, 0 \leq k \leq n-1) \) and \( (F^\text{SC}_{n-1-k}, 0 \leq k \leq n-1) \). As pointed out by Pitman, \( (F^\text{PF}_{n-1-k}, 0 \leq k \leq n-1) \) is a Markov chain whose evolution can be described as follows: \( F^\text{PF}_{n-1} \) is a uniform element of \( \mathcal{F}_{1,n} \); for \( 1 \leq k \leq n-1 \), \( F^\text{PF}_{n-1-k} \) is obtained from \( F^\text{PF}_{n-k} \) by choosing a uniform edge among the \((n-k)\) edges of \( F^\text{PF}_{n-k} \) and remove it. This process also comes under the name random cutting problem of uniform Cayley trees, whose study is initiated by Meir & Moon [26]. In this terminology, each edge removal is referred to as a cut on the initial tree \( F^\text{PF}_{n-1} \), which reduces \( F^\text{PF}_{n-1} \) into finer and finer fragments. In [26], Meir & Moon are interested in the number of cuts necessary to isolate a vertex \( v \). More precisely, recall that \( \Pi^\text{PF}_k \) is the partition of \([n]\) induced by the vertex sets of the tree components of \( F^\text{PF}_k \), \( 0 \leq k \leq n-1 \). For \( v \in [n] \), which is a vertex of \( F^\text{PF}_{n-1} \), denote by \( \pi^\text{PF}_k(v) \) the element of \( \Pi^\text{PF}_k \) which contains \( v \). Then the number of cuts necessary to isolate \( v \) is the number of cuts which have reduced the tree components containing \( v \) during the process \( (F^\text{PF}_{n-1-k}, 0 \leq k \leq n-1) \), namely

\[
X^\text{PF}(v) := \#\{ k \in \{1, \ldots, n-1\} : \pi^\text{PF}_{n-1-k}(v) \neq \pi^\text{PF}_{n-k}(v) \}. \quad (37)
\]

Nowadays, it is well-known from the works [28, 19, 9, 1] that if we take a random vertex \( V_n \sim \text{Uniform}([n]) \), then

\[
n^{-\frac{1}{2}} \cdot X^\text{PF}(V_n) \xrightarrow{d_{n \to \infty}} R, \quad (38)
\]

where \( R \) is a Rayleigh variable, i.e. \( \mathbb{P}(R > r) = \exp(-r^2/2) \), for \( r \in (0,\infty) \). Here, we propose a simple proof of this fact based on the process \( (F^\text{SC}_{n-1-k}, 0 \leq k \leq n-1) \) and its coupling with \( (F^\text{PF}_{n-1-k}, 0 \leq k \leq n-1) \) as explained in Section 2.8.

Recall from Section 2.1 the sequence of forests \( (F^\text{SC}_k, 0 \leq k \leq n-1) \) and the (random) decreasing edge-labelling \( L^\text{SC} \) which encodes the coalescent history. By Theorem 6, the time-reversal process \( (F^\text{SC}_{n-1-k}, 0 \leq k \leq n-1) \) can be described as follows: take \( T^\text{SC} \sim \text{Uniform}(\mathcal{F}_{1,n}) \) and a decreasing edge-labelling \( L^\text{SC} \sim \text{Uniform}(\text{DL}(T^\text{SC})) \); for \( 0 \leq k \leq n-1 \), the forest \( F^\text{SC}_k \) is obtained from \( T^\text{SC} \) by
removing all the edges with labels $> k$. Let us mention that the construction in Section 2.3 implies that $(F_{n-1-k}^\text{SC}, 0 \leq k \leq n-1)$ is a Markov chain. However, this fact is not used in the rest of this part. What is useful here is the following simple observation. During the fragmentation process $(F_{n-1-k}^\text{SC}, 0 \leq k \leq n-1)$, along each branch $B$ of $T^\text{SC}$, edges are removed in the same order, that is, independently of the choice of $L^\text{SC}$: among the set of edges of $B$, the first removed edge is incident to the root, then the edge incident to this edge, etc. In analogy to the random cutting problem, let us define for each vertex $v \in [n]$,

$$X^\text{SC}(v) := \#\{k \in \{1, \ldots, n-1\} : \pi_{n-1-k}^\text{SC}(v) \neq \pi_n^\text{SC}(v)\},$$

(39)

where $\pi_k^\text{SC}(v)$ is the vertex set of the tree component of $F_k^\text{SC}(v)$ which contains $v$, $0 \leq k \leq n-1$. We have the following.

**Lemma 24.** For the coalescent forests $(F_{n-1-k}^\text{SC}, 0 \leq k \leq n-1)$ defined by the SC Algo and $X^\text{SC}(v)$ defined in (39), we have for all $v \in [n]$,

$$X^\text{SC}(v) = H^\text{SC}(v) + D^\text{SC}(v),$$

(40)

where $H^\text{SC}(v)$ denotes the height of $v$ in $T^\text{SC}$, namely, the number of edges on the unique path of $T^\text{SC}$ from the root to $v$, and $D^\text{SC}(v)$ denotes the in-degree of $v$ in $T^\text{SC}$. Moreover, under the coupling in Section 2.8, we have

$$X^\text{PF}(v) = X^\text{SC}(v), \quad \text{for all } v \in [n].$$

(41)

**Proof.** For the first statement, observe that in the sequence $(F_{n-1-k}^\text{SC}, 0 \leq k \leq n-1)$, the vertex $v$ is isolated after each edge on the path between the root and $v$ is removed (in the order from the root towards $v$) and also all the in-going edges of $v$ are removed (in the order determined by $L^\text{SC}$). This easily leads to (40). Next, in the coupling of Section 2.8, we have $(\Pi_k^\text{PF}, 0 \leq k \leq n-1) = (\Pi_k^\text{SC}, 0 \leq k \leq n-1)$. Then comparing the definitions (37) and (39), we get (41).

From Lemma 24, we deduce the convergence of the number of cuts.

**Proposition 25.** For the coalescent forests $(F_{n-1-k}^\text{SC}, 0 \leq k \leq n-1)$ defined by the SC Algo and $X^\text{SC}(v)$ defined in (39), we have

$$\max_{v \in T^\text{SC}} n^{-\frac{1}{2}} |X^\text{SC}(v) - H^\text{SC}(v)| \xrightarrow{\mathbb{P}} n \to \infty 0.$$  

(42)

Moreover, for the coalescent forests $(F_{n-1-k}^\text{PF}, 0 \leq k \leq n-1)$ defined by PF Algo and $X^\text{PF}(v)$ defined in (37), let $V_n \sim \text{Uniform}([n])$ for each $n \geq 1$, we have

$$n^{-\frac{1}{2}} \cdot X^\text{PF}(V_n) \xrightarrow{\mathbb{P}} R,$$

(43)

where $R$ is a Rayleigh variable defined by $\mathbb{P}(R > r) = \exp(-r^2/2)$, for $r \in (0, \infty)$.
Theorem 26. For each $n \geq 1$, let the pair $(T^{PF}, T^{SC})$ be defined using the coupling in Section 2.8. Let $(T, \text{cut}(T))$ be the pair of the Brownian CRT and its cut tree. Then,

\[
\left( \left( \frac{1}{\sqrt{n}} T^{PF}, \frac{1}{\sqrt{n}} T^{SC} \right), n \geq 1 \right) \xrightarrow{(d)}_{n \to \infty} (T, \text{cut}(T)),
\]

where the convergence is with respect to the Gromov–Hausdorff topology.

**Sketch of proof.** For each $n \geq 1$, let $V^n_k, k \geq 1$ be an i.i.d. sequence of r.v. distributed as $\text{Uniform}([n])$. Similarly, let $V_k, k \geq 1$ be an i.i.d. sequence of uniform points of $T$ sampled according to the mass measure of $T$. Write $H_n(v)$ (resp. $H(v)$ and $H^{\text{cut}}(v)$) for the height of a vertex $v$ in $T^{PF}$ (resp. height of a point $v$ in $T$ and the height of a point $v$ in cut($T$).) Then by [10], we have for each $k \geq 1$,

\[
\left( \left( \frac{1}{\sqrt{n}} H_n(V^n_i), \frac{1}{\sqrt{n}} X^{PF}(V^n_i) \right), 1 \leq i \leq k \right) \xrightarrow{n \to \infty (d)} \left( (H(V_i), H^{\text{cut}}(V_i), 1 \leq i \leq k \right).
\]
Combined with (42), this then entails that for all \( k \geq 1 \),
\[
\left( \left( \frac{1}{\sqrt{n}} H_n(V_{i}^n), \frac{1}{\sqrt{n}} H^\text{SC}(V_{i}^n) \right), 1 \leq i \leq k \right) \xrightarrow{n \to \infty} \left( \left( H(V_i), H^\text{cut}(V_i) \right), 1 \leq i \leq k \right),
\]
where \( H^\text{SC}(V_{i}^n) \) stands for the height of the vertex \( V_{i}^n \) in \( T^\text{SC} \). Following the arguments in [10], we can deduce from this the following convergence:
\[
\left( \left( \frac{1}{\sqrt{n}} T^\text{PF}, \frac{1}{\sqrt{n}} T^\text{SC} \right), n \geq 1 \right) \xrightarrow{\text{(d)}} \left( T^\text{cut}(T) \right),
\]
where the convergence is with respect to the Gromov–Prokhorov convergence (see for example Section 1.4 in [10] for a definition). Next, we observe that both \( T^\text{PF} \) and \( T^\text{SC} \) are uniformly distributed in \( \mathcal{F}_{1,n} \). By Aldous’ Theorem [2], both \( \left( \frac{1}{\sqrt{n}} T^\text{PF}, n \geq 1 \right) \) and \( \left( \frac{1}{\sqrt{n}} T^\text{PF}, n \geq 1 \right) \) are tight with respect to the Gromov–Hausdorff topology. It follows that the pairs \( \left( \frac{1}{\sqrt{n}} T^\text{PF}, \frac{1}{\sqrt{n}} T^\text{SC}, n \geq 1 \right) \) also form a tight sequence. Combined with (47), this then entails (46) (we refer to [23] for more details).

References


