Probabilistic insights in matrix tree theorems

Loïc Chaumont

October 31, 2018

Contents

1 Introduction **2**

2 Some notation **2**

3 Finite state Markov chains and spanning trees **3**

3.1 Discrete time Markov chains 3

3.2 Continuous time Markov chains 5

3.3 The Markov chain tree theorem 6

3.4 Lifting the Markov chain to its spanning trees 10

3.5 The matrix tree theorem. 12

4 Enumeration of multitype forests. **16**

5 Branching trees and forests. **20**

5.1 Definitions 20

5.2 Total progeny of multitype branching forests 22

5.3 Coding trees and forests 23

5.4 Application to multitype branching forests 26

1 Introduction

Kirchhoff's matrix tree theorem is a combinatorial formula giving the number of spanning trees in a finite graph. These lectures aim at proving this result and presenting some important applications in probability theory and combinatorics. One of the most famous of these applications we will present here is an expression for the invariant distribution of irreducible finite state Markov chains. Direct applications of the matrix tree theorem also allow us to derive explicit formulas for the number of rooted multitype forests which can be obtained from a given set of labeled vertices, with a given sequence of degrees. These expressions are extensions of Cayley's formulas. As a direct consequence of these enumeration formulas will compute the distribution of the total progeny of a multitype Bienaymé-Galton-Watson (BGW) forest, by distinguishing types. This result will naturally lead us to introduce a coding of multitype Bienaymé-Galton-Watson forests through multi-indexed sequences of matrix valued random walks which extends the well known Lukasiewicz-Harris coding of Bienaymé-Galton-Watson single type forests. Such coding multi-indexed sequences can be seen as extensions of downward skip free random walks whose first passage times have the same distribution as the total progeny of multitype BGW forests. We will then see how coding sequences can be used to compute the distribution of some functionals attached to multitype BGW forests such as the number of vertices of a given degree.

Acknowledgements: I would like to warmly thank Maria Emilia Caballero and Geronimo Uribe for giving me the opportunity to do this lecture in the Instituto de Matemáticas of the UNAM, for the 9*th school of probability and stochastic processes*. I am also much indebted to my PhD student Marine Marolleau for her careful reading of the first proofs of these notes.

2 Some notation

This section aims at presenting some basic notions on graphs. Other definitions we will be given along the lines. Henceforth the elements of E will be called the *vertices* and will be thought of as integers from 1 to d, that is $E = [d]$, where $[d] := \{1, \dots, d\}$. A *directed graph* or a *digraph* on $[d]$ is obtained by associating to $[d]$ a set $A \subset [d] \times [d]$ of directed edges which we will call the *arcs*, each arc joining some pair of vertices. Since we only consider digraphs in these lectures, we will simply refer to them as graphs. A graph on $[d]$ with set of directed edges A will be denoted by $G = ([d], A)$. For two vertices i, j, the arc from i to j, if it exists, will be denoted by (i, j). In a graph, there exits at most one arc from i to j. If there exists more than one directed edge between two vertices, then G will be called a multigraph. We assume that there are no loops in graphs and multigraphs, that is no arc directed from a vertex to itself. We will say that there is a *path* from i to j if either the arc (i, j) exists or there are

$n \geq 1$ vertices $i_1, \ldots, i_n$ such that the arcs (i, i_1), (i_1, i_2),..., (i_{n-1}, i_n), (i_n, j) exist. In the latter case if $i = j$, then the path is called a *cycle*. A graph is said to be *connected* if for all $i, j \in E$, with $i \neq j$, there is a path from i to j. In the example below, one has $E = [6]$ and $A = \{(1,2); (1,4); (2,3); (3,4); (3,5); (4,5); (4,6); (5,3); (6,1); (6,4)\}$.

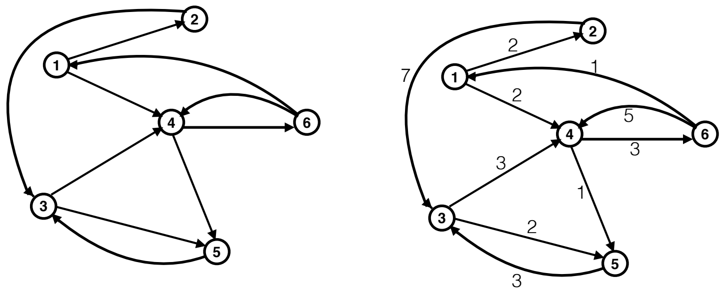

Figure 1: A digraph and its weighted version

We will often consider *weighted graphs* by associating to each arc $(i, j) \in A$ some quantity x_{ij}. Then the *Laplacian matrix* of a weighted graph $G = ([d], A)$ is the matrix $L = (x_{ij})_{i,j \in [d]}$, where $x_{ii} = -\sum_{j \in [d]} x_{ij}$.

By tree we mean a graph with no cycle. Such a tree is rooted if all edges are directed toward a particular vertex called the root. If (i, j) is an arc in a tree, then we say that j is the parent of i or that i is the child of j. A rooted plane tree is a rooted tree which is embedded in the plane. It is equivalent to say that at each generation, an order can be given to its vertices. A rooted plane forest is a finite set of rooted plane trees whose roots have been ordered. In these notes we will only consider rooted trees and rooted forests and we will often simply call them trees and forests.

3 Finite state Markov chains and spanning trees

3.1 Discrete time Markov chains

Let E be a finite set with cardinality $|E| = d \geq 1$ and $(\Omega, \mathcal{F}, \mathbb{P})$ be some probability space. A discrete time Markov chain with state space E is a sequence of E-valued random variables $Y := (Y_n)_{n \geq 0}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ such that for all $n \geq 0$ and all elements $i_0, i_1, \ldots, i_{n+1}$ of E satisfying $\mathbb{P}(Y_0 = i_0, \ldots, Y_n = i_n) > 0$,

$$\mathbb{P}(Y_{n+1} = i_{n+1} \,|\, Y_0 = i_0, \ldots, Y_n = i_n) = \mathbb{P}(Y_{n+1} = i_{n+1} \,|\, Y_n = i_n)\,. \qquad (3.1)$$

When the conditional probabilities $\mathbb{P}(Y_{n+1} = j | Y_n = i)$, $i, j \in E$ do not depend on n, the Markov chain is said to be *homogeneous*. We will always assume that this is actually the case and we will denote by $P = (p_{ij})_{i,j\in E}$ the *transition matrix*, that is

$$p_{ij} = \mathbb{P}(Y_{n+1} = j \,|\, Y_n = i)\,, \quad i, j \in E\,.$$

The law of Y_0 under $\mathbb{P}$ is called the *initial law* of Y. We will often denote it by μ. Note that the law of a Markov chain is characterized by both its initial law and its transition matrix. Indeed, it follows from (3.4) that for all $i_0, i_1, \dots, i_n \in E$,

$$\mathbb{P}(Y_0 = i_0, \dots, Y_n = i_n) = \mu(i_0)p_{i_0 i_1}p_{i_1 i_2}\dots p_{i_{n-1}i_n}\,. \tag{3.2}$$

The chain Y is said to be *irreducible* if for all $i, j \in E$, $i \neq j$, there are $n \geq 1$ and $i_1, \dots, i_{n-1}$ in E such that for $i_0 = i$ and $i_n = j$, $\mathbb{P}(Y_0 = i_0, \dots, Y_n = i_n) > 0$. For $n \geq 1$, let us denote by $p_{ij}^{(n)}$ the entries of the matrix P^n and set $P^0 = I$, where I is the identity matrix of dimension d. Then we can prove by using (3.2) that Y is irreducible if and only if for all $i, j \in E$, there is $n \geq 0$ such that $p_{ij}^{(n)} > 0$.

For $i \in E$, we will denote by $\mathbb{P}_i$ the probability measure on $(\Omega, \mathcal{F})$ under which the chain Y is issued from i, that is

$$\mathbb{P}_i(A) := \mathbb{P}(A \,|\, Y_0 = i)\,, \quad A \in \mathcal{F}\,.$$

Then for any probability measure μ on E, we denote by $\mathbb{P}_\mu$ the probability on $(\Omega, \mathcal{F})$ under which the chain Y has initial distribution μ, that is

$$\mathbb{P}_\mu(A) := \sum_{i\in E} \mu(i)\mathbb{P}_i(A)\,, \quad A \in \mathcal{F}\,.$$

In particular one has $\mathbb{P}_i = \mathbb{P}_{\delta_i}$ and $\mathbb{P}_\mu(Y_0 = i) = \mu(i)$. Moreover, for all $n \geq 0$, the law of Y_n under $\mathbb{P}_\mu$ is given by

$$\mathbb{P}_\mu(Y_n = i) = \mu P^n(i)\,, \quad i \in E\,.$$

A probability μ on E is said to be invariant if the law of Y_1 under $\mathbb{P}_\mu$ is μ or equivalently if μ is a left eigenvector of P associated to the eigenvalue 1, that is $\mu P = \mu$. Note that in this case, one has $\mu P^n = \mu$ for all $n \geq 0$ and the law of Y_n under $\mathbb{P}_\mu$ is μ.

An irreducible Markov chain with invariant distribution μ is said to be time-reversible or simply reversible if for all n, under $\mathbb{P}_\mu$, the sequence $(Y_0, \dots, Y_n)$ has the same law as the sequence $(Y_n, \dots, Y_0)$. This property is equivalent to the following:

$$\mu_i p_{ij} = \mu_j p_{ji}, \quad \text{for all } i, j \in E.$$

Irreducible Markov chains with finite state space satisfy the following property.

Theorem 3.1. *Let $Y := (Y_n)_{n\geq 0}$ be a Markov chain with finite state space. Assume that Y is irreducible. Then there exists a unique invariant probability measure μ for Y. Moreover, this distribution is given by*

$$\mu(j) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} p_{ij}^{(k)}\,, \quad \textit{for all } i, j \in E\,. \tag{3.3}$$

We say that P is a *primitive* matrix if there exists $n \geq 1$ such that $p_{ij}^{(n)} > 0$, for all $i, j \in E$. Note that if P is primitive, then it is irreducible. When P is primitive, Theorem 3.1 is a direct consequence of the Perron-Frobenius theorem.

Expression (3.3) means that starting from i, the mean number of visits to any state j by Y up to time $n-1$ converges toward $\mu(j)$. Note that if moreover P is *aperiodic* then the sequence Y under $\mathbb{P}_i$ converges in law towards μ. This shows that Markovian models satisfying those assumptions converge toward some equilibrium. Then it is very important to determine an explicit form of this distribution in terms of the transition matrix of the chain. This is what will be explained in the next sections but before we need to recall the notion of continuous time Markov chain which provides a more convenient framework for our purpose.

3.2 Continuous time Markov chains

A *q-matrix* on the set E is a squared d-dimensional matrix which we often denote by $Q = (q_{ij})_{i,j \in E}$ and satisfying:

1. $0 \leq q_{ij} < \infty$, for all $i, j \in E$ such that $i \neq j$,
2. $\sum_{j=1}^{d} q_{ij} = 0$, for all $i \in E$.

If $d = 1$, then by convention we set $Q = 0$. The *transition probability function* associated to Q is the matrix valued function, defined on $[0, \infty)$, $t \mapsto P(t)$ given by

$$P(t) = e^{tQ} = \sum_{n=0}^{\infty} \frac{(tQ)^n}{n!}, \quad t \geq 0\,.$$

The entries of $P(t)$ are denoted by $p_{ij}(t)$ and are called the transition probabilities. Then an E-valued *continuous time Markov chain* $X = (X_t)_{t \geq 0}$ with transition probability function $(P(t))_{t \geq 0}$ is a càdlàg continuous time stochastic process defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that for all $n \geq 0$, all elements $i_0, \dots, i_{n+1}$ of E and all real values $0 \leq t_0 < \dots < t_{n+1}$,

$$\mathbb{P}(X_{t_{n+1}} = i_{n+1} \,|\, X_{t_0} = i_0, \dots, X_{t_n} = i_n) = p_{i_n i_{n+1}}(t_{n+1} - t_n)\,. \tag{3.4}$$

In particular, X satisfies the *time homogeneous Markov property*,

$$\mathbb{P}(X_{t_{n+1}} = i_{n+1} \,|\, X_{t_0} = i_0, \dots, X_{t_n} = i_n) = \mathbb{P}(X_{t_{n+1}} = i_{n+1} \,|\, X_{t_n} = i_n)\,.$$

The initial law of X is the law of X_0 under $\mathbb{P}$. A continuous time Markov chain is actually a jump process and we can derive from the Markov property that the times elapsed in between the jumps are independent and exponentially distributed. Let us denote by $(T_n)_{n \geq 0}$ the ordered sequence of jump times of X, that is $T_0 = 0$ and for

$n \geq 1$, $T_n = \inf\{t \geq T_{n-1} : X_t \neq X_{T_{n-1}}\}$. For all $i \in E$, set $q_i = \sum_{j \neq i} q_{ij}$. Then the random sequence

$$Y_n = X_{T_n}, \quad n \geq 0,$$

is an E-valued Markov chain whose transition matrix P is given by

$$p_{ij} = \begin{cases} q_{ij}/q_i & \text{if } i \neq j \text{ and } q_i \neq 0 \\ 0 & \text{if } i \neq j \text{ and } q_i = 0 \end{cases} \quad \text{and} \quad p_{ii} = \begin{cases} 0 & \text{if } q_i \neq 0 \\ 1 & \text{if } q_i = 0 \end{cases} . \tag{3.5}$$

The discrete time Markov chain (Y_n) is sometimes called the *skeleton* of the Markov chain X. The Markov chain X (or equivalently the Q-matrix Q) is said to be irreducible if for all $i, j \in E$ such that $i \neq j$, there is $n \geq 1$ such that $q_{ij}^{(n)} > 0$. It is clear from (3.5) that Q is irreducible if and only if P is irreducible.

With a slight abuse of notation, we will also denote by $\mathbb{P}_i$ the probability measure on $(\Omega, \mathcal{F})$ such that X starts from i at time $t = 0$ and for any probability measure on E, we will denote by $\mathbb{P}_\lambda$ the probability measure under which X has initial law λ. In particular, we have $\mathbb{P}_i(X_0 = i) = 1$, $\mathbb{P}_\lambda(X_0 = i) = \lambda_i$ and the law of X_t under $\mathbb{P}_\lambda$ is given by $\mathbb{P}_\lambda(X_t = i) = \lambda P(t)_i$, that is the i-th coordinate of the row vector $\lambda P(t)$. An initial law is said to be invariant for X if $\mathbb{P}_\lambda(X_t = i) = \lambda P(t)_i = \lambda(i)$, for all $t \geq 0$. It follows from the definition of the transition function $P(t)$ that λ is an invariant distribution if and only if

$$\lambda Q = 0 . \tag{3.6}$$

Then we derive from (3.5) that λ is an invariant distribution for X if and only if μ is an invariant distribution for Y, where

$$\mu_i = q_i \lambda_i , \quad i \in E . \tag{3.7}$$

In particular if Q is irreducible then there exists a unique invariant distribution for X. Moreover, one may obtain an expression of the invariant distribution directly from the Q-matrix through the limit:

$$\lambda_j = \lim_{n \to \infty} \left(\frac{1}{n} \sum_{k=0}^{n} e^{kQ} \right)_{ij} , \quad \text{for all } i \in E.$$

As well as in discrete time these expressions of the invariant distribution are not really satisfactory since they are obtained as limits of sequences involving all the powers of Q. The matrix tree theorem in Section 3.5 provides a much more exploitable formula.

3.3 The Markov chain tree theorem

Any Q-matrix, naturally induces a digraph denoted $G(Q) = (E, A(Q))$ in the following way: for all $i, j \in E$, $(i, j) \in A(Q)$ if and only if $q_{ij} > 0$. When it is associated to a graph, a Q-matrix is also called the *weighted Laplacian matrix* of the digraph (the Laplacian matrix of the graph is actually obtained from Q by replacing q_{ij}, for $i \neq j$, by 1, see Section 3.5). It is plain that $G(Q)$ is connected if and only if Q is irreducible. The value q_{ij} is then called the *weight* of (i,j). A digraph which is obtained from a Q-matrix in this way is called a *weighted digraph.* The weighted digraph above is generated from the Q-matrix,

$$Q = \begin{pmatrix} -4 & 2 & 0 & 2 & 0 & 0 \\ 0 & -7 & 7 & 0 & 0 & 0 \\ 0 & 0 & -5 & 3 & 2 & 0 \\ 0 & 0 & 0 & -4 & 1 & 3 \\ 0 & 0 & 3 & 0 & -3 & 0 \\ 1 & 0 & 0 & 5 & 0 & -6 \end{pmatrix}.$$

A subgraph of a digraph $G = (E, A)$ is a digraph $G' = (E', A')$ such that $E' \subset E$ and $A' \subset A$. Then note that a subgraph of a weighted digraph $G(Q)$ is obtained as $G(Q')$, where Q' is a sub matrix of Q in which some entries have been replaced by 0 and the diagonal has been properly diminished so that is it a Q-matrix. For instance, the following subgraph of the weighted digraph presented in Figure 1,

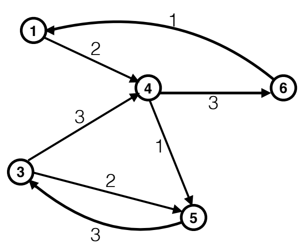

Figure 2: A subgraph of the weighted digraph in Figure 1.

is obtained from the Q-matrix,

$$Q' = \begin{pmatrix} -4 & 0 & 2 & 0 & 0 \\ 0 & -5 & 3 & 2 & 0 \\ 0 & 0 & -4 & 1 & 3 \\ 0 & 3 & 0 & -3 & 0 \\ 1 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

A *rooted tree* is a connected digraph with no cycle such that all the arcs are directed toward a particular vertex called the *root.* A *rooted spanning tree* of a digraph is a rooted tree with the same set of vertices. In the example of Figure 1 there are 9 spanning trees rooted at vertex 5, see the figure below.

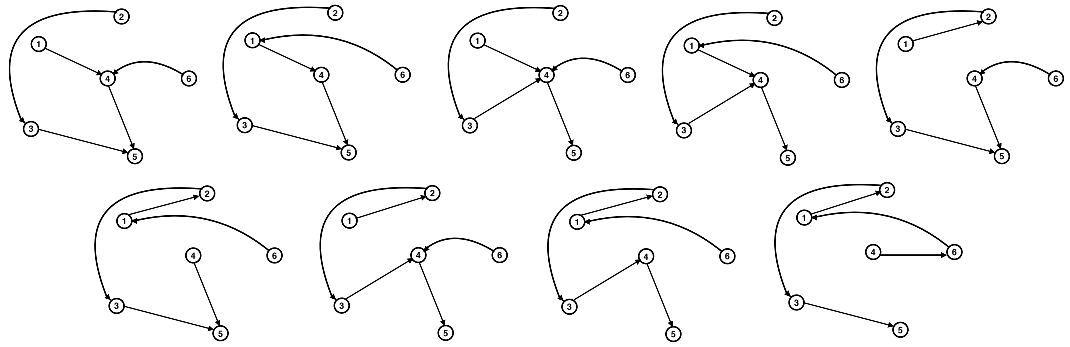

Figure 3: The 9 spanning trees rooted at vertex 5.

The weight of the subgraph $G' = G(Q')$ is obtained as the product of all the weights of its edges, that is the product of all the positive entries of Q'. It is denoted by $w(G')$ or $w(Q')$. More formally, with $Q' = (q'_{ij})$,

$$w(G') = \prod_{i,j:q'_{ij}>0} q'_{ij} \,.$$

The following result, known as the Markov chain tree theorem, provides an explicit formula for the invariant distribution of any irreducible continuous time Markov chain with finite state space, in terms of the weights of the spanning trees associated to its graph. Let $X = (X_t)_{t\geq 0}$ be such a Markov chain with state space E and Q-matrix Q. For $i \in E$, let $\mathcal{T}_i$ be the set of spanning trees of $G(Q)$ which are rooted at i and define,

$$\Sigma_i = \sum_{T\in\mathcal{T}_i} w(T) \quad \text{and} \quad \Sigma = \sum_{i\in E} \Sigma_i \,.$$

Theorem 3.2. *The Markov chain X is irreducible if and only if $\Sigma_i > 0$, for all $i \in E$. When this is the case, the unique invariant distribution λ of X is given by,*

$$\lambda_i = \Sigma_i/\Sigma \,, \quad i \in E \,. \tag{3.8}$$

Proof. The Markov chain X is irreducible if and only if its associated graph is strongly connected. Then it is easy to check that a graph is strongly connected if and only if for each $i \in E$ there is a spanning tree rooted at i. In order to check that $(\lambda_i)_{i\in E}$ is an invariant distribution, it is enough to show that $(\Sigma_i)_{i\in E}$ is an invariant measure for X. Then according to (3.6), we have to prove that for all $j \in E$,

$$\sum_{i\neq j} \Sigma_i q_{ij} = -\Sigma_j q_{jj} \,. \tag{3.9}$$

Since X is irreducible, $q_{jj} < 0$, for all j. Then multiply the right hand side of (3.9) by $-\sum_{k\neq j} \frac{q_{jk}}{q_{jj}} = 1$ and expand de sums Σ_i and Σ_j according to their definition. We obtain

$$\sum_{i\neq j} \sum_{T\in\mathcal{T}_i} q_{ij} w(T) = \sum_{k\neq j} \sum_{T\in\mathcal{T}_j} q_{jk} w(T) \,. \tag{3.10}$$

It remains to convince ourself that this identity is true.

Let us first define a cycle-rooted tree as tree rooted at some vertex i to which we attach an edge directed from i to any other vertex of this tree. Then note that by removing from this cycle-rooted tree an edge (j, k) along the cycle containing i (which actually is the only cycle), we obtain a new tree rooted at j.

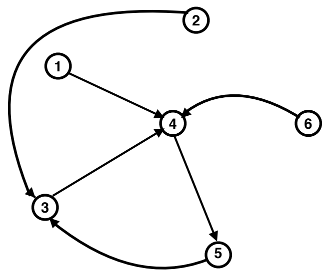

Figure 4: A tree cycle-rooted

Now in formula (3.10), the terms $q_{ij}w(T)$ and $q_{jk}w(T)$ are the weights of cycle rooted trees whose cycle contains j. Then by performing the double sum $\sum_{i\neq j}\sum_{T\in\mathcal{T}_i}$ or the double sum $\sum_{k\neq j}\sum_{T\in\mathcal{T}_j}$ one obtains the total weight of all cycle rooted trees whose cycle contains vertex j. On the right hand side, we enumerate them through the arc which is directed toward j, whereas on the left hand side we enumerate them through the arc which is issued from j. $\square$

Then note that the same result can be obtained for discrete time Markov chains. Let Y be such a chain on the finite state space E and let P be its transition matrix. The notion of weighted directed graph associated to Y is defined in the same way as in continuous time. The only difference in this case is that the digraph can have loops. The weight of an arc (i, j) of a subgraph is then p_{ij} and the weight of the subgraph is the product of all the weights of its arcs. Then let $G(P)$ be the directed graph of Y, let $\mathcal{T}_i^P$ be the set of spanning trees of $G(P)$ rooted at vertex i and let $w^P(T)$ be the weight of $T \in \mathcal{T}_i^P$. Define

$$\Sigma_i^P = \sum_{T\in\mathcal{T}_i^P} w^P(T) \quad \text{and} \quad \Sigma^P = \sum_{i\in E}\Sigma_i^P\,,$$

then the following result can be proved exactly in the same way as Theorem 3.2. Its proof is left as an exercise.

Theorem 3.3. *The Markov chain Y is irreducible if and only if $\Sigma_i^P > 0$, for all $i \in E$. When this is the case, the unique invariant distribution μ of Y is given by,*

$$\mu_i = \Sigma_i^P/\Sigma^P\,, \quad i \in E\,. \tag{3.11}$$

Note that formula (3.11) can be recovered from formula (3.8) by using the relation (3.7). Indeed, from (3.5), the weight of a spanning tree t of $G(P)$ rooted at i is

$w^P(t) = w(t)/\prod_{j\neq i} q_j$. Hence $\Sigma_i^P/q_i = \Sigma_i/q$, where $q = \prod_j q_j$.

The Markov chain tree theorem was first discovered in the 80's by Leighton and Rivest [14, 15], althought some particular form already existed from the 60's. The proof which is given here is excerpt from Ventcel and Freidlin [11].

Despite Theorems 3.2 and 3.3 provide explicit forms of the invariant measure, the latter are not fully satisfactory since they bear on the sets of spanning trees of the associated weighted graph. Determining these sets require some work, see for instance Figure 2. However this result has a lot of applications. It was used by [1], [7] and [21] to produce algorithms for building uniform spanning trees in a graph, which leads to the construction of loop erased random walks in an infinite network. The Kirchhoff's formula which will be established in Subsection 3.5 complete Theorems 3.2 and 3.3 and allows us to compute the invariant measures in a direct way.

3.4 Lifting the Markov chain to its spanning trees

We will now see how to associate to any irreducible Markov chain $(X_t)_{t\geq 0}$ on a finite set E with Q-matrix Q, a Markov chain on the set of spanning trees of the graph $G(Q) = (E, A(Q))$.

Let us first define the projection of a Markov chain. Let E' be a finite set and $p : E' \to E$ be a surjective map. Let $(X'_t)_{t\geq 0}$ be a Markov chain on E' with Q-matrix $Q' = (q'_{ij})_{i,j\in E'}$. Assume that for all $j \in E$,

$$\sum_{m\in p^{-1}(j)} q'_{km} = \sum_{m\in p^{-1}(j)} q'_{lm}\,, \quad \text{whenever } p(k) = p(l).$$

Define for $i, j \in E$,

$$q_{ij} = \sum_{m\in p^{-1}(j)} q'_{km}\,,$$

where k is any element of E' such that $p(k) = i$. Then $X = p(X')$ is a Markov chain on E with Q-matrix $Q = (q_{ij})_{i,j\in E}$. Furthermore if ν is an invariant measure for X', then μ defined by

$$\mu(i) = \sum_{k\in p^{-1}(i)} \nu(k)$$

is an invariant measure for X. This property is quite simple to check and is left as an exercise. We will refer to it as the *projection property of Markov chains.*

We now define the *lifting operation* of the Markov chain $(X_t)_{t\geq 0}$. Let $\mathcal{T} = \cup_{i\in E}\mathcal{T}_i$ be the set of all spanning trees of the graph $G(Q)$. Assume that $(X_t)_{t\geq 0}$ is irreducible and define the map $p : \mathcal{T} \to E$ which assigns to each tree $t \in \mathcal{T}$ its root, i.e. $p(\mathcal{T}_i) = i$. From the irreducibility of $(X_t)_{t\geq 0}$, this mapping is surjective. Then let $(X_t^{\mathcal{T}})_{t\geq 0}$ be a Markov chain on $\mathcal{T}$ whose transition probabilities $(q_{st}^{\mathcal{T}})_{s,t\in\mathcal{T}}$ are defined as follows:

Let $i \in E$, $s \in \mathcal{T}_i$ and $j \in E$ be such that $q_{ij} > 0$. Take out from s the unique edge coming out from j and add to s the edge (i, j). One obtains a new tree t which is rooted at j, i.e. $t \in \mathcal{T}_j$. Then set $q^{\mathcal{T}}_{st} = q_{ij}$ and for all pair $s \neq t$ which are not obtained in this way, set $q^{\mathcal{T}}_{st} = 0$. The transition rates $q^{\mathcal{T}}_{st}$ define a Q-matrix on the set $\mathcal{T}$.

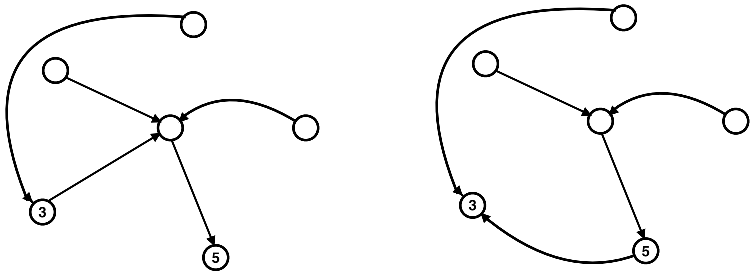

Figure 5: Lifting a transition between $i = 5$ and $j = 3$

It is easy to check from the projection property recalled above that X is the projection of $X^{\mathcal{T}}$ by the map p. Moreover, the above definition of $Q^{\mathcal{T}}$ induces a natural pathwise construction of the chain $X^{\mathcal{T}}$ from the paths of the original chain X. More specifically, let $i \in E$ and $t \in \mathcal{T}_i$, then to each path of X with $X_0 = i$ corresponds a unique path of $X^{\mathcal{T}}$ such that $X^{\mathcal{T}}_0 = t$.

Theorem 3.4. *Let X be an irreducible Markov chain on the set E with Q-matrix Q. The lifted Markov chain $(X^{\mathcal{T}}_t)_{t\geq 0}$ is irreducible and*

$$\lambda^{\mathcal{T}}(t) = w(t)/\Sigma\,, \quad t \in \mathcal{T} \tag{3.12}$$

is the invariant distribution of $X^{\mathcal{T}}$.

Proof. Let us first prove that the Markov chain $(X^{\mathcal{T}}_t)_{t\geq 0}$ is irreducible. As noticed before the statement of Theorem 3.4, for fixed $i \in E$ and $t \in \mathcal{T}_i$ the path $X_n(\omega), n \geq 0$ with $X_0(\omega) = i$ naturally induces a path $X^{\mathcal{T}}_n(\omega), n \geq 0$ with $X^{\mathcal{T}}_0(\omega) = t$. It is readily seen that for $i \neq j$, the sets $\mathcal{T}_i$ and $\mathcal{T}_j$ are connected through $X^{\mathcal{T}}$. So it is enough to prove that for any $s, t \in \mathcal{T}_i$, if $X^{\mathcal{T}}_0(\omega) = t$, then there is a path of $X^{\mathcal{T}}$ which leads to s.Let k be a leaf of s and call $s_{k\to i}$ the branch leading from k to i in s. Then let $i \to k_1 \to \cdots \to k_l \to k$ be a path of X from i to k. Run X from i to k through this path and then let X go back to i through the branch $s_{k\to i}$. Then $X^{\mathcal{T}}$ equals a new tree rooted at i and containing the path $s_{k\to i}$. Now run X from i to k_l and go back to i through the branch $s_{k_l\to i}$ and so on until the branch $s_{k_1\to i}$ is constructed.

Repeating this procedure for each leaf of s we finally obtain a tree which is composed of all branches going from the leaves of s to i, that is the tree s itself.

Then we now check that $\lambda^{\mathcal{T}}(t)$ is an invariant distribution for $Q^{\mathcal{T}}$. We only need to prove that for all $t \in \mathcal{T}$, $\sum_{s\in\mathcal{T}} w(s)q^{\mathcal{T}}_{st} = 0$, which is equivalent to

$$\sum_{s\in\mathcal{T}, s\neq t} w(s)q^{\mathcal{T}}_{st} = -w(t)q^{\mathcal{T}}_{tt}.$$

Since $(X_t^{\mathcal{T}})$ is irreducible, $q_{tt}^{\mathcal{T}} > 0$ and $-\sum_{l\neq t} \frac{q_{tl}^{\mathcal{T}}}{q_{tt}^{\mathcal{T}}} = 1$. Multiplying the right hand side of the above equality by this last term gives,

$$\sum_{s\in\mathcal{T}, s\neq t} w(s)q_{st}^{\mathcal{T}} = \sum_{l\in\mathcal{T}, l\neq t} w(t)q_{tl}^{\mathcal{T}}\,. \tag{3.13}$$

But for each neighbor s of t, there is a unique neighbor u of t such that $w(s)q_{st}^{\mathcal{T}} = w(t)q_{tu}^{\mathcal{T}}$ and this establishes a bijection in the set of neighbors of t. Summing overall neighbors of t, we obtain, $\sum_{s\in\mathcal{T},s\neq t} w(s)q_{st}^{\mathcal{T}} = \sum_{u\in\mathcal{T},u\neq t} w(t)q_{tu}^{\mathcal{T}}$, which is equation (3.13). □

This result was actually used by Anantharam and Tsoucas in [2] to provide a new proof of the Markov chain tree theorem. Recently Biane and Chapuy, in [5] and [6] have been interested in describing relations between the invariant distribution of a Markov chain (X_t) and its lifted version $(X_t^{\mathcal{T}})$.

3.5 The matrix tree theorem.

In this section we shall prove a quite general version of Kirchhoff formula which is known as the matrix tree theorem. The original result is dated from 1847, [12] and the general form which is presented here is due to Tutte [20]. Recall that a graph is a rooted forest if it consists in a disjoint union of rooted trees. When each vertex of a rooted forest has a type $i \in [d]$, we will call this forest a multitype rooted forest or more specifically a d-type rooted forest. This matrix tree theorem regards special d-type rooted forests, called elementary forests.

Definition 1. *An elementary forest on $[d]$, is a d-type rooted forest which contains exactly one vertex of each type.*

As all forests considered in this section are elementary forests, we simply call them forests. The roots of all the trees composing a forest $\mathbf{f}$ will be called the roots of the forest and this set will be denoted by roots($\mathbf{f}$). The digraph of Figure 6 is a forest on the set $E = [7]$ with roots($\mathbf{f}$) $= \{1, 5, 7\}$.

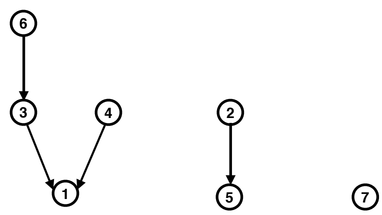

Figure 6: A forest $\mathbf{f}$ on [7] with roots($\mathbf{f}$) $= \{1, 5, 7\}$

For each couple of vertices i and j of E we define a variable x_{ij} and to each forest $\mathbf{f}$ on E we attach the monomial $x_{\mathbf{f}}$ which is obtained as the product of the variables

x_{ij} for all arcs (i,j) of $\mathbf{f}$. For example, the monomial of the forest $\mathbf{f}$ in Figure 6 is

$$x_{\mathbf{f}} = x_{63}x_{31}x_{41}x_{25}\,.$$

Note that a vertex of $\mathbf{f}$ is both a root and a leaf if and only if its index does not appear in $x_{\mathbf{f}}$. In particular, if all vertices of $\mathbf{f}$ are leaves, then $x_{\mathbf{f}} = 0$. If $\mathbf{f}$ is a weighted digraph with Laplacian matrix Q then by giving to each variable x_{ij} the value q_{ij}, we see that the monomial $x_{\mathbf{f}}$ corresponds to the weight of $\mathbf{f}$. Note also that the data of the monomial $x_{\mathbf{f}}$ allows us to recover the forest $\mathbf{f}$.

Now set as usual $E = [d]$, let us use the notation $\mathrm{x} = (x_{ij})_{i,j\in E}$ and fix a subset $I \subseteq E$. Then we define $G_d^{(I)}(\mathrm{x})$, the generating function of all the forests whose set of roots is I, by

$$G_d^{(I)}(\mathrm{x}) = \sum_{\mathbf{f}:\mathrm{roots}(\mathbf{f})=I} x_{\mathbf{f}}\,.$$

By convention we set $G_d^{(\emptyset)}(\mathrm{x}) \equiv 0$ and $G_d^{(E)}(\mathrm{x}) \equiv 1$. On the other hand, let us define for $d \geq 2$ the matrix

$$H_d(\mathrm{x}) = \begin{pmatrix} (x_{12}+\cdots+x_{1d}) & -x_{12} & -x_{13} & \dots & -x_{1d} \\ -x_{21} & (x_{21}+\cdots+x_{2d}) & -x_{23} & \dots & -x_{2d} \\ \vdots & & \ddots & & \vdots \\ -x_{d1} & -x_{d2} & -x_{d,d-1} & \dots & (x_{d1}+\cdots+x_{d,d-1}) \end{pmatrix},$$

and set $H_1(\mathrm{x}) \equiv 0$. Note that by giving nonnegative values q_{ij} to the variables x_{ij} the matrix $H_d(\mathrm{x})$ corresponds to $-Q$ where Q is the Q-matrix $Q = (q_{ij})_{i,j\in E}$.

The following result is known as the matrix tree theorem.

Theorem 3.5. *The generating function $G_d^{(I)}(\mathrm{x})$ of forests rooted at I satisfies the equality*

$$G_d^{(I)}(\mathrm{x}) = \det\big(H_d^{(I)}(\mathrm{x})\big)\,,$$

where $H_d^{(I)}(\mathrm{x})$ is the squared matrix obtained by delating rows and columns which indices belong to I in $H_d(\mathrm{x})$ and where by convention we set $\det\big(H_d^{(E)}(\mathrm{x})\big) \equiv 1$.

Proof. Note that for $d \geq 2$, $\det H_d^{(\emptyset)}(\mathrm{x}) \equiv 0$ since the sum of the rows in $H_d(\mathrm{x})$ is 0 and this is also true for $d = 1$ by convention. The proof will be done by induction on d. For $d = 1$, the result follows by convention, since in this case we necessarily have $I = \emptyset$ or $I = E$. Let us now assume that the result is true for $d-1$, where $d \geq 2$, and for all $I \subset E$.

We will assume that $I \neq \emptyset$ and $I \neq E$, since the result is always true by convention in these cases. Then first observe that both $G_d^{(I)}(\mathrm{x})$ and $\det\big(H_d^{(I)}(\mathrm{x})\big)$ are polynomial of degree $d-|I|$. Indeed, for $\det\big(H_d^{(I)}(\mathrm{x})\big)$ we can show by induction, expending this determinant along a row or a column that all monomials contain $d-|I|$ terms. Moreover, this determinant cannot be 0 for all x, since $I \neq \emptyset$. On the other hand, a forest with d vertices and $|I|$ roots contains $d-|I|$ arcs, so that all monomials of $G_d^{(I)}(\mathrm{x})$ have $d-|I|$ terms. Now note the following property of both polynomials: for every monomial, there is j such that none of the variables x_{ij}, $i=1,\dots,d$ appear in this monomial. (For $G_d^{(I)}(\mathrm{x})$ this is true for instance if j is a leaf of a forest F rooted at I. Indeed, since there are no directed edges toward j, no variable x_{ij} occurs in the monomial x_F.) But this is clearly true anyway for both polynomials since they are of degree $d-|I|\le d-1$. Indeed, if there was a monomial that would contain a variable x_{ij} for all j, then the polynomial would have degree d.

Thanks to this property, in order to verify that both polynomials are identical, it is enough to verify this identity for each j by assuming that $x_{ij}=0$, for all i. Moreover, we do not lose any generality by assuming that $j=d$.

So, let us set $x_{id}=0$ for all i. Then $G_d^{(I)}(\mathrm{x})$ is left with monomials $x_{\mathbf{f}}$ such that d is a leaf of $\mathbf{f}$. Let us denote by $\overline{G}_d^{(I)}(\mathrm{x})$ this polynomial. If $d\in I$, then the terms of $G_d^{(I)}(\mathrm{x})$ remaining in $\overline{G}_d^{(I)}(\mathrm{x})$ correspond to forests on E in which d is both a root and leaf, which is the same as $G_{d-1}^{(J)}(\mathrm{x})$, where $J=I\setminus\{d\}$. If $d\notin I$, then each monomial in $\overline{G}_d^{(I)}(\mathrm{x})$ corresponds to a forest on E in which d is a leaf, but not a root. Any such forest is actually constructed from a forest on $[d-1]$, with roots set J to which d is attached by an arc directed to any of its vertices. This means that in this case, $\overline{G}_d^{(I)}(\mathrm{x})$ is obtained by multiplying $G_{d-1}^{(J)}(\mathrm{x})$ by the factor $x_{d1}+\cdots+x_{d,d-1}$. Each term x_{dj} of this sum corresponds to the edge (d,j) which attached to the forests in $G_{d-1}^{(J)}(\mathrm{x})$. Then we obtained the following expression for $\overline{G}_d^{(I)}(\mathrm{x})$, with $J=I\setminus\{d\}$,

$$\overline{G}_d^{(I)}(\mathrm{x})=\begin{cases} G_{d-1}^{(J)}(\mathrm{x}), & \text{if } d\in I,\\ (x_{d1}+\cdots+x_{d,d-1})G_{d-1}^{(J)}(\mathbf{x}), & \text{if } d\notin I.\end{cases} \tag{3.14}$$

On the other hand, setting $x_{id}=0$ for all i in $H_d^{(I)}(\mathrm{x})$ gives a matrix which can be written in a block form as

$$\overline{H}_d(\mathrm{x})=\begin{pmatrix} H_{d-1}(\mathrm{x}) & 0\\ -x_{d1}-\cdots-x_{d,d-1} & x_{d1}+\cdots+x_{d,d-1}\end{pmatrix}.$$

If $d\in I$, then the last row and column will be deleted so that $\det\overline{H}_d^{(I)}(\mathrm{x})=\det H_{d-1}^{(J)}(\mathrm{x})$. If $d\notin I$, then by expanding the determinant of $\overline{H}_d(\mathrm{x})$ along the last column, we obtain $x_{d1}+\cdots+x_{d,d-1}$ multiplied by $\det\overline{H}_{d-1}^{(J)}(\mathrm{x})$, so that

$$\det\overline{H}_d^{(I)}(\mathrm{x})=\begin{cases} \det\overline{H}_{d-1}^{(J)}(\mathrm{x}), & \text{if } d\in I,\\ (x_{d1}+\cdots+x_{d,d-1})\det\overline{H}_{d-1}^{(J)}(\mathrm{x}), & \text{if } d\notin I.\end{cases} \tag{3.15}$$

From the induction assumption we have $\overline{G}_{d-1}^{(J)}(\mathrm{x})=\det\overline{H}_{d-1}^{(J)}(\mathrm{x})$ and the result follows by comparing (3.14) and (3.15). $\square$

Examples: 1. For $d = 2$ and $I = \{1\}$ then there is only one forest on E which is given by the edge $(2,1)$. Hence $G_2^{(\{1\})}(\mathbf{x}) = x_{21}$. On the other hand,

$$Q_2(\mathbf{x}) = \begin{pmatrix} x_{12} & -x_{12} \\ -x_{21} & x_{21} \end{pmatrix},$$

so that $\det(Q_2^{(\{1\})}(\mathbf{x})) = x_{21}$.

2. Let $d = 3$ and $I = \{1\}$, then the three forests on $[3]$ are given below:

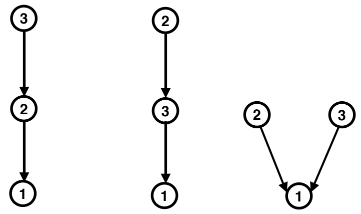

Figure 7: The three forests on [3] with roots set $\{1\}$

The corresponding generating function is

$$G_3^{(\{1\})}(\mathbf{x}) = x_{32}x_{21} + x_{23}x_{31} + x_{21}x_{31}\,.$$

On the other hand, delating the first row and column in $Q_3(\mathbf{x})$, we obtain the determinant,

$$\det \begin{pmatrix} (x_{21} + x_{23}) & -x_{23} \\ -x_{32} & (x_{31} + x_{32}) \end{pmatrix} = (x_{21} + x_{23})(x_{31} + x_{32}) - x_{23}x_{32}\,,$$

which is equal to $G_3^{(\{1\})}(\mathbf{x})$.

Let $X = (X_t)_{t\geq 0}$ be an E-valued continuous time Markov chain with Q-matrix Q. We will denote by $Q^{(i)}$ the $(d-1)\times(d-1)$ matrix which is obtained by deleting row and column i in Q. Recall from Subsection 3.3 that $\mathcal{T}_i$ is the set of spanning trees of G which are rooted at i and that $w(T)$ denotes the weight of $T \in \mathcal{T}_i$. In particular, we can express the invariant distribution of a continuous time Markov chain with sate space E and Q-matrix Q in a more explicit way than in Theorem 3.2.

Corollary 1. *Let X be an irreducible continuous time Markov chain with state space E and Q-matrix Q. Then*

$$\det(-Q^{(i)}) = \sum_{T\in\mathcal{T}_i} w(T)\,, \quad i \in E\,,$$

and the unique invariant distribution λ of X is given by,

$$\lambda_i = \det(-Q^{(i)})/\Sigma\,,$$

where $\Sigma = \sum_{i=1}^{d} \det(-Q^{(i)})$.

Proof. For any x, $G_d^{(\{i\})}(\mathrm{x})$ represents the total weight of spanning trees rooted at i in the complete graph endowed with the adjacency matrix (x_{ij}) (where $x_{ii} = 0$, for all i). Then taking $x_{ij} = q_{ij}$, we obtain the graph of X and we conclude from Theorems 3.2 and 3.5. □

4 Enumeration of multitype forests.

We will now consider multitype trees and forests whose set of types is $[d]$. From now on, when talking about trees or forests, we will always mean rooted multitype plane trees or rooted multitype plane forests on the set $[d]$. Recall that a plane forest (or an ordered forest) is a forest which is embedded in the plan. It is equivalent to say that the vertices of the forest are ordered. Recall that in a multitype forest, children of each vertex are placed so that children of type 1 are on the left, then children of type 2 and so on, see Figure 8. For such a forest $\mathbf{f}$, we call the Laplacian matrix of $\mathbf{f}$, the matrix $K(\mathbf{f}) = (k_{ij}(\mathbf{f}))_{i,j\in[d]}$ ($k_{ij} = k_{ij}(\mathbf{f})$ when no confusion is possible) such that for $i, j \in [d]$, $i \neq j$, k_{ij} is the number of vertices of type i whose parent has type j, and for all $i \in [d]$,

$$-k_{ii} = r_i + \sum_{j\neq i}^{d} k_{ij}, \tag{4.16}$$

where r_i is the number of roots of type i. If $\mathrm{n} = (n_1, \dots, n_d)$ denotes the vector such that n_i is the total number of vertices of type i in $\mathbf{f}$, then the coupe (K, n) is called the *total progeny of the forest.*

Definition 2. *A simple forest is a forest such that for each i, at most one vertex of type i has children. If for some type i, no vertex has children, then a special vertex is marked. Moreover a simple forest contains at least one vertex of each type.*

Note that an elementary forest is a simple forest whose Laplacian matrix has at most one entry equal to 1 on each line, the other values being less or equal than 0.

Lemma 1. *Let r_i and k_{ij}, $i, j \in [d]$, $i \neq j$ be nonnegative integers satisfying* (4.16) *and assume that $r_1 + \dots + r_d > 0$. Then the number of simple forests with total progeny (K, n) is* $\det(-k_{ij})_{i,j\in[d]}$.

Proof. To each simple forest, we can associate an elementary forest in an obvious way: we start from the roots, then their children are those who themselves have children or those who are marked, and so on. An example of a simple forest and its associated elementary forest is presented in Figure 7.

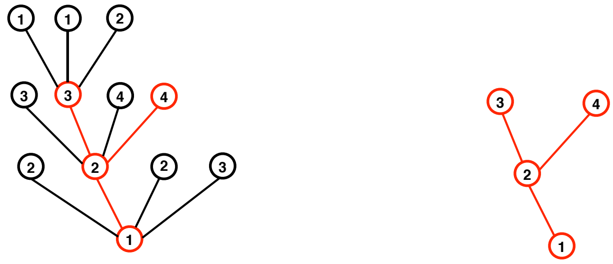

Figure 8: A simple forest and its associated elementary forest.

Let us code elementary forests by vectors $(i, j_i)_{i\in[d]}$, where j_i is the parent of i and $j_i = d+1$ if i is a root. Then, inverting the above procedure, we see that to each elementary forest $(i, j_i)_{i\in[d]}$, we can associate exactly $\prod_{i\in[d]} k_{ij_i}$ simple forests, where $k_{i(d+1)} = r_i$. Indeed, for each i, there are k_{ij_i} possibilities to choose the vertex of type i which has got children or who is marked. Hence, $\prod_{i\in[d]} k_{ij_i}$ is the number of simple forests to which we can associate the elementary forest $(i, j_i)_{i\in[d]}$. Then in order to obtain the total number of simple forests with Laplacian matrix $(k_{ij})_{i,j\in[d]}$, it remains to perform the sum of these monomials over all the elementary forests on $[d]$, that is

$$\sum_{(i,j_i)_{i\in[d]}} \prod_{i\in[d]} k_{ij_i}.$$

But we see that this sum is the same as the generating function of elementary trees on $[d+1]$ and rooted at $d+1$,

$$G_{d+1}^{(\{d+1\})}(\mathrm{x}) = \sum_{\mathbf{t}:\mathrm{roots}(\mathbf{t})=d+1} x_{\mathbf{t}}\,.$$

taken at $x_{ij} = k_{ij}$, this is exactly $\det(-k_{ij})$ from Theorem 3.5. $\square$

We emphasize that from this lemma, the number simple forests with total progeny (K, n) only depends on the Laplacian matrix K. Indeed, it is easy to see that vertices of type i whose parent has type i, that is $n_i + k_{ii}$, are not involved in the counting presented above.

Let us now extend the notion of spanning tree of a graph to this of a spanning forest of a multigraph. Let $G = ([d], A)$ be a multigraph and let $I \subseteq [d]$, then a spanning forest of a multigraph $G = ([d], A)$ is a subgraph of G which is an elementary forest rooted at I. To a multigraph $G = (E, A)$, we associate the matrix $L(G) = (l(G)_{ij})_{i,j\in[d]}$, where for $i \neq j$, $l_{ij} = l_{ij}(G)$ is the number of edges directed from i to j and $l_{ii} = -\sum_{j\neq i} l_{ij}$. The matrix L will be called the Laplacian matrix of G, although definitions may differ from a text to another.

Corollary 2. *Let $G = ([d], A)$ be a multigraph with Laplacian matrix L and fix $I \subseteq E$, a non empty subset of E. Let $L^{(I)}$ be the matrix obtained by delating rows* *and colomns which indices belong to I in L. Then the number of spanning forests rooted at I of G obtained by labeling its edges is* $\det(-L^{(I)})$.

Proof. Let us first prove the result for $I = \{1\}$. Let us construct the following set of simple trees from the multigraph G. Fix a spanning tree t of G. Then unfold the multigraph G in order to obtain a simple multitype tree mt on $[d]$ as follows: we start from the elementary tree on $[d]$ defined by our spanning tree. Then on each vertex of t we graft vertices corresponding to its incident edges in G by increasing order of the types of the children and by respecting the order of the edges in G. If some type has no children in mt, then we mark the vertex which belongs to t. This construction is illustrated in Figure 8.

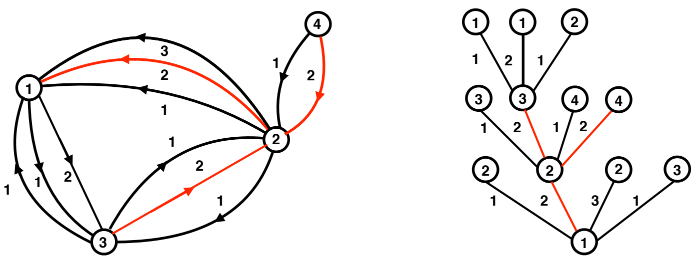

Figure 9: A multigraph and one of its unfolded simple trees.

It is clear that two different spanning trees rooted at 1 in G provide two different simple trees rooted at 1. Conversely, each simple tree, once folded, produces G with some distinguished spanning tree. Moreover, the simple trees issued from this construction have Laplacian matrix

$$k_{ij} = l_{ij} \quad \text{if } (i,j) \neq (1,1) \quad \text{and} \quad k_{11} = l_{11} - 1.$$

Then we have defined a bijection between spanning trees in G rooted at 1 and simple multitype trees rooted at 1 with Laplacian matrix (k_{ij}). From Lemma 1, the number of these simple trees is

$$\det(-k_{ij})\,,$$

which is equal to $\det(-L^{(\{1\})})$. Indeed, the first column of K can be decomposed as $k_{\cdot 1} = l_{\cdot 1} +^t (-1, 0, \ldots, 0)$ and the determinant of L is 0.

The general case of a subset $I \subset E$ is done from the same arguments. □

The Laplacian matrix of the example of the proof is

$$L = \begin{pmatrix} -1 & 0 & 1 & 0 \\ 3 & -4 & 1 & 0 \\ 2 & 1 & -3 & 0 \\ 0 & 2 & 0 & -2 \end{pmatrix},$$

and the number of spanning trees of $G = ([d], A)$ rooted at $\{1\}$ is $\det(-L^{(\{1\})}) = 22$.

Now we shall enumerate labeled forests according to the degree of their vertices. For the remainder of this section, r_i, n_i and k_{ij}, $i, j \in [d]$ will be integers satisfying the following conditions:

$$r_i \geq 0,\ r_1 + \cdots + r_d \geq 1,\ k_{ij} \geq 0, \text{ for } i \neq j,\ -k_{ii} = r_i + \sum_{j \neq i} k_{ij} \text{ and } n_i \geq -k_{ii} \geq 1.$$

Definition 3. *To each forest* $\mathbf{f}$ *with* n_i *vertices of type* i *and to each vertex of type* i *in* $\mathbf{f}$*, we associate an integer in* $[n_i]$*, which is called its* label*. Then* $\mathbf{f}$ *is called a* labeled plane forest*. Let* $\mathscr{L}$ *be the set of* labeled plane forests *with* n_i *vertices of type* i*,* r_i *roots of type* i*, in which* k_{ij} *vertices of type* i *have a parent of type* j*.*

Then we define the set of forests with given indegree.

Definition 4. *Let* $\mathbf{c} = (c_{i,j,k})_{i,j \in [d], k \in [n_j]}$ *be a tuple of non-negative integers such that* $k'_{ij} = \sum_{k=1}^{n_j} c_{i,j,k}$*, where* $k'_{ij} = k_{ij}$ *if* $i \neq j$ *and* $k'_{ii} = n_i + k_{ii}$*. We will denote by* $\mathscr{L}(\mathbf{c})$ *the subset of* $\mathscr{L}$ *of* labeled plane forests with indegree $\mathbf{c}$*, that is the set of labeled plane forests in which the vertex of type* j *with label* k *has* $c_{i,j,k}$ *children of type* i*. Then* $\mathbf{c}$ *is called the* indegree tuple *of the forest* $\mathbf{f} \in \mathscr{L}(\mathbf{c})$*.*

We first show that the cardinality of the set $\mathscr{L}(\mathbf{c})$ does not depend on the indegree $\mathbf{c}$. Recall that in this definition, k'_{ii} represents the number of vertices of type i whose parent has type i, so that for all $i, j \in [d]$, the integer k'_{ij} is equal to the number of vertices of type i whose parent has type j.

Lemma 2. *For any* $\mathbf{c}$ *and* $\mathbf{c}'$ *such that* $\sum_{k=1}^{n_j} c_{i,j,k} = \sum_{k=1}^{n_j} c'_{i,j,k}$*, for all* $i, j \in [d]$*,*

$$|\mathscr{L}(\mathbf{c})| = |\mathscr{L}(\mathbf{c}')|.$$

In other words, the number of labeled plane multitype forests with given indegrees tuples only depends on the adjacency matrix.

Proof. Assume first that for fixed i_0, j_0 and $1 \leq k_1, k_2 \leq n_{j_0}$, $0 \leq c'_{i_0,j_0,k_1} = c_{i_0,j_0,k_1} - 1$ and $c'_{i_0,j_0,k_2} = c_{i_0,j_0,k_2} + 1$ and $c'_{i,j,k} = c_{i,j,k}$, whenever $(i,j,k) \neq (i_0, j_0, k_1)$ and $(i,j,k) \neq (i_0, j_0, k_2)$. Then we can build a labeled plane forest $\mathbf{f}'$ with indegree $\mathbf{c}'$ from a labeled plane forest $\mathbf{f}$ with indegree $\mathbf{c}$ by removing the subtree rooted at the first child of type i_0 of k_1-th vertex of type j_0 in $\mathbf{f}$ and by graphing it on the k_2-th vertex of type j_0 in $\mathbf{f}$ so that the root of type i_0 of this new subtree is the first child of type i_0 of the k_2-th vertex of type j_0. This transformation is clearly a bijection between $\mathscr{L}(\mathbf{c})$ and $\mathscr{L}(\mathbf{c}')$.

Then for any indegree tuples $\mathbf{c}$ and $\mathbf{c}'$ such that $\sum_{k=1}^{n_j} c_{i,j,k} = \sum_{k=1}^{n_j} c'_{i,j,k}$ and any labeled plane forests $\mathbf{f}$ and $\mathbf{f}'$ with respective indegrees $\mathbf{c}$ and $\mathbf{c}'$, we can transform $\mathbf{f}$ into $\mathbf{f}'$ through elementary operations as described above. This induces a bijection between $\mathscr{L}(\mathbf{c})$ and $\mathscr{L}(\mathbf{c}')$ and the result is proved. $\square$

Theorem 4.1. *The number of labeled plane multitype forests with given indegrees tuples is*

$$|\mathscr{L}(\mathbf{c})| = det(-k_{ij}) \prod_{j=1}^{d} (n_j - 1)!$$

Proof. Assume first that $\mathbf{c} = (c_{i,j,k})_{i,j \in [d], k \in [n_j]}$ is such that for each $j \in [d]$ there is at most one $k \in [n_j]$, such that $c_{i,j,k} > 0$ for some $i \in [d]$. Then the corresponding unlabeled forest is a simple forest. We have proved in Lemma 1 that the number of simple forests with Laplacian matrix $K = (k_{ij})$ is $\det(-k_{ij})_{i,j \in [d]}$. Given a simple forest there are exactly $(n_i - 1)!$ ways to give labels to the $n_i - 1$ vertices of type i who have no children or who are not marked. Hence there are $\det(-k_{ij}) \prod_{j=1}^{d} (n_j - 1)!$ labeled simple forests, that is $|\mathscr{L}(\mathbf{c})| = \det(-k_{ij}) \prod_{j=1}^{d} (n_j - 1)!$ for this particular indegree tuple. Then the result follows from Lemma 2. $\square$

Let us finally mention the following consequence of our results which can be found in Proposition 11 of [4]. A (multitype) Cayley trees is just a connected graph with no cycles. (It is not embedded in the plan.) A (multitype) Cayley forest is a finite set of Cayley trees.

Corollary 3. *The number of labeled multitype Cayley forests with given indegrees tuples is*

$$det(-k_{ij}) \frac{\prod_{j=1}^{d} (n_j - 1)!}{\prod_{i \in [d]} r_i! \prod_{i,j \in [d], k \in [n_j]} c_{i,j,k}!}.$$

Proof. When enumerating labeled plane forests with indegree $\mathbf{c}$, we count $\mathbf{f}, \mathbf{f}' \in \mathcal{L}(\mathbf{c})$ such that $\mathbf{f}'$ can be obtained by permuting in $\mathbf{f}$ the $c_{i,j,k}$ subtrees whose roots are the $c_{i,j,k}$ children of type i of the kth vertex of type j, for some $i, j \in [d]$ and $k \in [n_j]$ or by exchanging the trees whose roots have the same type in the whole forest. But in this case, $\mathbf{f}$ and $\mathbf{f}'$ are the same Cayley forest. Therefore, we still have to divide the number $\prod_{j=1}^{d} (n_j - 1)! \det(-k_{ij})$ by $\prod_{i \in [d]} r_i! \prod_{i,j \in [d], k \in [n_j]} c_{i,j,k}!$, that is

$$|\mathcal{L}(\mathbf{c})| = \frac{\prod_{j=1}^{d} (n_j - 1)!}{\prod_{i \in [d]} r_i! \prod_{i,j \in [d], k \in [n_j]} c_{i,j,k}!} \det(-k_{ij}).$$

$\square$

5 Branching trees and forests.

5.1 Definitions

All the random variables we will consider here will be defined on a reference probability space $(\Omega, \mathcal{F}, P)$. Let $\nu := (\nu_1, \dots, \nu_d)$, where ν_i is some distribution on $\mathbb{Z}_+^d$. We consider a population of individuals which reproduce independently of each other at each generation. Individuals of type i give birth to n_j children of type $j \in [d]$ with probability $\nu_i(n_1, \dots, n_d)$. Then ν is called the progeny distribution of the population.

For $i, j \in [d]$, we denote by m_{ij} the mean number of children of type j, given by an individual of type i, i.e.

$$m_{ij} = \sum_{(n_1,\dots,n_d)\in\mathbb{Z}_+^d} n_j \nu_i(n_1, \dots, n_d)\,.$$

We say that ν is non singular if there is $i \in [d]$ such that $\nu_i(\mathrm{n} : n_1 + \dots + n_d = 1) < 1$. The matrix $M = (m_{ij})$ is said to be irreducible if for all i, j, $m_{ij} < \infty$ and there exists $n \geq 1$ such that $m_{ij}^{(n)} > 0$, where $m_{ij}^{(n)}$ is the ij entry of the matrix M^n. If moreover the power n does not depend on (i, j), then M is said to be primitive. In the latter case, according to Perron-Frobenius theory, the spectral radius ρ of M is the unique eigenvalue which is positive, simple and with maximal modulus. If $\rho \leq 1$, then the population will become extinct almost surely, whereas if $\rho > 1$, then with positive probability, the population will never become extinct. We say that ν is subcritical if $\rho < 1$, critical if $\rho = 1$ and supercritical if $\rho > 1$. We sometimes say that μ is irreducible, primitive, (sub)critical or supercritical, when this is the case for M.

In what follows, we will always assume that the progeny distribution ν is non degenerate, primitive and critical or subcritical. Then under this assumption we can define almost surely finite multitype branching trees, as follows: we start from some vertex of type $i \in [d]$ at generation 0. Then the tree grows from one generation to the other as follows. For $n \geq 1$, conditionally on the joint progeny at generation n, each vertex of this generation gives birth to children at generation $n+1$ according to its own progeny distribution, independently of the other vertices and then die. This stochastic evolution of the population is called the *branching property*. It produces a multitype tree which is almost surely finite according to what is recalled above. Then we embed this tree in the plan so that it is a multitype plane tree, as defined in the previous section. A multitype branching forest with progeny distribution ν is a sequence of independent multitype branching trees with progeny distribution ν, who are ordered between themself. The forest is said to be finite or an infinite according to whether it contains a finite or infinite number of trees. We will first pay a special attention to finite forests. A finite multitype branching forest is a random variable which will usually be denoted by F_{r}, where $\mathrm{r} = (r_1, \dots, r_d) \in \mathbb{Z}_+^d$ is such that $r_1 + \dots, r_d > 0$ and r_i is the number of trees in F_{r} whose root is of type i. We will say that the forest in rooted at r. (Recall that in the ordering of the plane forest F_{r}, we placed trees whose root is of type 1 first, then trees whose root is of type 2, and so on.) Let us denote by $\mathscr{F}_{\mathrm{r}}$ the set of multitype forests which are rooted at r. Then the law of F_{r} is given by

$$P(F_{\mathrm{r}} = \mathbf{f}) = \prod_{u\in\mathbf{f}} \nu_{c(u)}(p(u))\,, \quad \mathbf{f} \in \mathscr{F}_{\mathrm{r}}, \tag{5.17}$$

where $c(u)$ is the type of the vertex u and $p(u) = (p_1(u), \dots, p_d(u))$ is its progeny, that is $p_i(u)$ is the number of children of type i of the vertex u.

Let us note that it is more common to call $\mathscr{F}_{\mathrm{r}}$ a multitype Bienaymé-Galton-Watson forest rather than a multitype branching forest which is used in a more general context. We choose this terminology as it is simpler.

5.2 Total progeny of multitype branching forests

Recall that the total progeny $p(\mathbf{f})$ of a multitype forest $\mathbf{f}$, is the couple (K, n), where $K = (k_{ij})_{i,j\in[d]}$ is the Laplacian matrix of $\mathbf{f}$ and $\mathrm{n} = (n_1, \dots, n_d)$ is the vector of $\mathbb{Z}_+^d$ such that n_i is the number of vertices of type i in $\mathbf{f}$. We denote by $p(F_{\mathrm{r}})$ the total progeny of the multitype branching forest F_{r}.

The next theorem gives and expression of the total progeny of any multitype branching forest in terms of its progeny distribution.

Theorem 5.1. *Assume that the progeny distribution ν is non-degenarate, primitive and critical or subcritical. Then for all $\mathrm{r} \in \mathbb{Z}_+^d$, such that $r_1 + \dots + r_d > 0$, for all Laplacian matrix K and for all $\mathrm{n} \in \mathbb{Z}_+^d$, such that*

$$r_i \geq 0,\ r_1 + \dots + r_d \geq 1,\ k_{ij} \geq 0, \text{ for } i \neq j,\ -k_{ii} = r_i + \textstyle\sum_{j\neq i} k_{ij} \text{ and } n_i \geq -k_{ii},$$

$$P(p(F_{\mathrm{r}}) = (K, \mathrm{n})) = \frac{\det(-\overline{K})}{\bar{n}_1\bar{n}_2 \dots \bar{n}_d} \prod_{j=1}^{d} \nu_j^{*n_j}(k'_{1j}, \dots, k'_{dj}), \tag{5.18}$$

where $\bar{n}_i = n_i \vee 1$, $k'_{ij} = k_{ij}$ if $i \neq j$ and $k'_{ii} = n_i + k_{ii}$ and $\overline{K}$ is the matrix K to which we removed the line i and the column i, for all i such that $n_i = 0$. Moreover, we set $\nu_j^{*0} = \delta_0$.

Proof. Let us first assume that $n_i \geq 1$, for all $i \in [d]$. Recall Definitions 3 and 4 of $\mathscr{L}$ and $\mathscr{L}(\mathrm{c})$. To each $\mathbf{f} \in \mathscr{F}_{\mathrm{r}}$ such that $p(\mathbf{f}) = (K, \mathrm{n})$ correspond $\prod_{i\in[d]} n_i!$ forests in $\mathscr{L}$, so that we can write

$$\begin{aligned} P(p(F_{\mathrm{r}}) = (K, \mathrm{n})) &= \sum_{\mathbf{f}\in\mathscr{F}_{\mathrm{r}},\, p(\mathbf{f})=(K,\mathrm{n})} P(F_{\mathrm{r}} = \mathbf{f}) \\ &= \frac{1}{\prod_{i\in[d]} n_i!} \sum_{\mathbf{f}\in\mathscr{L},\, p(\mathbf{f})=(K,\mathrm{n})} P(F_{\mathrm{r}} = \mathbf{f}) . \end{aligned}$$

Then let us decompose $\mathscr{L}$ as the union of $\mathscr{L}(\mathrm{c})$.

$$P(p(F_{\mathrm{r}}) = (K, \mathrm{n})) = \frac{1}{\prod_{i\in[d]} n_i!} \sum_{\mathrm{c}:\sum_{k=1}^{n_j} c_{i,j,k} = k'_{ij}} \sum_{\mathbf{f}\in\mathscr{L}(\mathrm{c})} P(F_{\mathrm{r}} = \mathbf{f}).$$

But from (5.17), for fixed c such that $\sum_{k=1}^{n_j} c_{i,j,k} = k'_{ij}$, the value of $P(F_{\mathrm{r}} = \mathbf{f})$ is the same for all $\mathbf{f} \in \mathscr{L}(\mathrm{c})$ and is $\prod_{j\in[d]} \prod_{k=1}^{n_j} \nu_j(c_{1,j,k}, \dots, c_{d,j,k})$, we obtain from Theorem 4.1,

$$\begin{aligned} & P(p(F_{\mathrm{r}}) = (K, \mathrm{n})) \\ = \; & \frac{\det(-k_{ij}) \prod_{j=1}^{d} (n_j - 1)!}{\prod_{i\in[d]} n_i!} \sum_{\mathrm{c}:\sum_{k=1}^{n_j} c_{i,j,k} = k'_{ij}} \prod_{j\in[d]} \prod_{k=1}^{n_j} \nu_j(c_{1,j,k}, \dots, c_{d,j,k}) \\ = \; & \frac{\det(-K)}{n_1 n_2 \dots n_d} \prod_{j=1}^{d} \nu_j^{*n_j}(k'_{1j}, \dots, k'_{dj}). \end{aligned}$$

Finally assume that $n_d = 0$ and $n_i \geq 1$, for $i \neq d$. Then we obtain (5.18) by using the same arguments, but replacing $\mathscr{F}_{\mathrm{r}}$, $\mathscr{L}$ and $\mathscr{L}(\mathrm{c})$ by the same sets defined on $[d-1]$. □

5.3 Coding trees and forests

A way to order a given tree is to use a search algorithm. The most commonly used search algorithms come from computer science. We will only use the depth first search algorithm (dfsa) and the breadth first search algorithm (bfsa). The dfsa consists in ordering the vertices of a tree by starting from the root and by visiting the vertices in the lexicographical order. In the bfsa, we order the vertices by visiting each generation from the left to the right, the first generation being this of the root. These definitions should be clear from Figure 10.

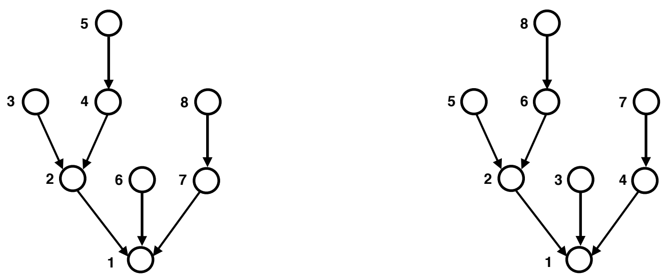

Figure 10: The same tree ordered in the depth first search order (left) and in the breadth first search order (right).

A plane multitype forest $\mathbf{f} = \{\mathbf{t}_1, \mathbf{t}_2, \dots\}$ will be ordered by ordering each tree, one after the other, see Figure 11. Note that this ordering of $\mathbf{f}$ does not depend on the types of the vertices. Such a forest will be called a forest ordered according to the bfso (breadth first search order).

For $i \in [d]$, we will denote by e_i the i-th unit vector of $\mathbb{Z}^d$. Then for $\mathrm{k} = (k_1, \dots, k_d)$ and $\mathrm{m} = (m_1, \dots, m_d)$ elements of $\mathbb{Z}_+^d$, we write $\mathrm{k} \leq \mathrm{m}$ if $k_i \leq m_i$, for all

$i \in [d]$ and we write $\mathrm{k} < \mathrm{m}$ if $k_i \le m_i$, for all $i \in [d]$ and $k_j < m_j$, for some $j \in [d]$.

Let us now define coding paths for multitype forests. Let u be some vertex of a forest $\mathbf{f}$. Then recall that we denoted by $p(u) = (p_1(u), \dots, p_d(u))$ the progeny of the vertex u, where $p_i(u)$ is the number of children of type i of u. Let $u_1, u_2, \dots$ be the ordered sequence of vertices of $\mathbf{f}$ and let $c(k)$ be the type of u_k. Then define the matrix $M_k = (m_{ij}^{(k)})_{i,j\in[d]}$, whose all the column are equal to 0 except the $c(k)$-th column which is given by

$$m_{i,c(k)}^{(k)} = p_i(u_k), \quad \text{for } i \neq c(k) \text{ and } \quad m_{c(k),c(k)}^{(k)} = p_{c(k)}(u_k) - 1\,.$$

We now construct a multi-indexed sequence of matrices (K_{h}), $\mathrm{h} = (h_1, \dots, h_d) \in \mathbb{Z}_+^d$ of $M_d(\mathbb{Z})$ associated with the forest $\mathbf{f}$. First we set $K_0 = 0$ and we define the matrices K_{h}, $0 \le \mathrm{h} \le \mathrm{n}$ by visiting vertices of $\mathbf{f}$ in their bfso. Suppose the matrix K_{h} is constructed up to the visit of vertex u_{k-1}. Then the next visited vertex is u_k and we set,

$$K_{\mathrm{h}+\mathrm{e}_{c(k)}} = K_{\mathrm{h}} + M_k\,,$$

and so on. It is not difficult to check from this construction that if n_i is the total number of vertices in $\mathbf{f}$, then the matrix K_{n} which is obtained once all vertices have been explored corresponds to the Laplacian matrix of $\mathbf{f}$. Moreover, the forest $\mathbf{f}$ can be reconstructed from the sequence $(K_{\mathrm{h}}, 0 \le \mathrm{h} \le \mathrm{n})$. The latter sequence will be denoted $K(\mathbf{f})$. Let us illustrate this construction on an example for $d = 2$:

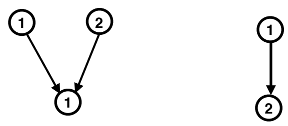

Figure 11: A 2-type forest with 2 trees.

The coding multi-indexed sequence of matrices $(K_{\mathrm{h}}, 0 \le \mathrm{h} \le \mathrm{n})$ corresponding to the forest of Figure 11 is $K_{(0,0)} = 0$ and then $K_{(1,0)}$, $K_{(2,0)}$, $K_{(2,1)}$, $K_{(2,2)}$ and $K_{(3,2)}$ are given respectively by

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} ; \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix} ; \begin{pmatrix} -1 & 0 \\ 1 & -1 \end{pmatrix} ; \begin{pmatrix} -1 & 1 \\ 1 & -2 \end{pmatrix} ; \begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix} .$$

We can check that the terminal value of this sequence, that is corresponds to the Laplacian matrix of the forest. More specifically, the progeny of this forest is

$$p(\mathbf{f}) = \left(\begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix}, (3,2) \right).$$

For any set E of real numbers, we denote by $\mathrm{M}_d(E)$ the set of squared matrices of dimension d with E-valued entries. We will consider the following subsets of $\mathrm{M}_d(\mathbb{Z})$,

$$\begin{aligned}\mathrm{M}_+ &= \{(x_{ij})_{i,j\in[d]} \in \mathrm{M}_d(\mathbb{Z}) : m_{ij} \geq 0, \;\; i,j \in [d],\, i \neq j\},\\ \mathrm{M}_{+,1} &= \{(x_{ij})_{i,j\in[d]} \in \mathrm{M}_+ : m_{ii} \geq -1, \;\; i \in [d]\}.\end{aligned}$$

The subset M_+ is call the set of essentially nonnegative matrices. It is very involved in Markov chains and branching processes. Indeed, q-matrices are essentially nonnegative matrices as well as Laplacian matrices of multitype branching forests. For a matrix $X \in \mathrm{M}_d(\mathbb{Z})$ we will use the notation $\overline{X} := X^t(1,1,\dots,1)$, so that $\overline{X}$ corresponds to the row vector:

$$\overline{X} = \left(\sum_{j=1}^{d} x_{ij}\,, i \in [d]\right).$$

We will consider sequences with values in $\mathrm{M}_d(\mathbb{Z})$ and indexed by $\mathbb{Z}_+^d$ as follows:

$$X_{\mathrm{k}} = (x_{k_j}^{i,j})_{i,j\in[d]}, \quad \mathrm{k} = (k_1,\dots,k_d) \in \mathbb{Z}_+^d,$$

that is all entries of column j of X_{k} are indexed by k_j.

Definition 5. *A sequence $(X_{\mathrm{k}},\, \mathrm{k} \in \mathbb{Z}_+^d)$ of matrices in M_+ is said to be downward skip free if*

(*i*) $X_0 = 0$,

(*ii*) *for all $i \in [d]$ and $\mathrm{k} \in \mathbb{Z}_+^d$, $X_{\mathrm{k}+\mathrm{e}_i} - X_{\mathrm{k}} \in \mathrm{M}_{+,1}$.*

Note that the sequence $K(\mathbf{f})$ defined above relatively to a forest $\mathbf{f}$ is a downward skip free sequence of matrices.

The following result allows us to define the first passage time of a downward skip free sequence of matrices.

Lemma 3. *Let $(X_{\mathrm{k}},\, \mathrm{k} \in \mathbb{Z}_+^d)$ be a downward skip free sequence of matrices and $\mathrm{r} = (r_1,\dots,r_d) \in \mathbb{Z}_+^d$. If there exists $\mathrm{k} \in \mathbb{Z}_+^d$ such that $\overline{X}_{\mathrm{k}} = -\mathrm{r}$, then there is some particular index $\mathrm{T}_{\mathrm{r}}(X) \in \mathbb{Z}_+^d$ (denoted by T_{r} when no confusion is possible) such that*

$$\overline{X}_{\mathrm{T}_{\mathrm{r}}} = -\mathrm{r} \quad \text{and} \quad \mathrm{T}_{\mathrm{r}} \leq \mathrm{q} \quad \text{whenever} \quad \overline{X}_{\mathrm{q}} = -\mathrm{r}. \tag{5.19}$$

The index T_{r} will be called the first passage time of the multi-indexed sequence $(X_{\mathrm{k}},\, \mathrm{k} \in \mathbb{Z}_+^d)$ at level $-\mathrm{r}$ and will be denoted by

$$\mathrm{T}_{\mathrm{r}} = \inf\{\mathrm{k} : \overline{X}_{\mathrm{k}} = -\mathrm{r}\}.$$

Let $\mathrm{n}, \mathrm{r} \in \mathbb{Z}_+^d$ and the matrix K such that

$$r_i \geq 0,\ r_1 + \cdots + r_d \geq 1,\ k_{ij} \geq 0, \text{ for } i \neq j,\ -k_{ii} = r_i + \textstyle\sum_{j \neq i} k_{ij} \text{ and } n_i \geq -k_{ii} \geq 1$$

and define the following set of finite sequences of downward skip free sequences of matrices,

$$\Sigma_{\mathrm{n},K} = \{(X_{\mathrm{k}},\, 0 \leq \mathrm{k} \leq \mathrm{n}) \text{ downward skip free} : T_{\mathrm{r}} = \mathrm{n} \text{ and } X_{T_{\mathrm{r}}} = K\}.$$

Denote by $\mathscr{F}_{K,\mathrm{n}}$ the subset of $\mathscr{F}_{\mathrm{r}}$ of forests with progeny (K, n).

Proposition 1. *The mapping*

$$\begin{array}{rcl} \psi : \mathscr{F}_{K,\mathrm{n}} & \to & \Sigma_{\mathrm{n},K} \\ \mathbf{f} & \mapsto & K(\mathbf{f}) \end{array}$$

is a bijection.

This result is intuitively clear and we will not prove it here. It actually extends a well known result for single type forest, which is called the Lukasiewicz-Harris coding of plane forests, see Proposition 1.1 in [16].

5.4 Application to multitype branching forests

Recall from the beginning of this section the definition of a branching forest F_{r}, with progeny distribution $\nu = (\nu_1, \ldots, \nu_d)$. We assume again that ν is non degenerate, primitive and critical or subcritical. Recall also the definition of the downward skip free sequence of matrices $K(F_{\mathrm{r}})$ which, from our definition, is a multi-indexed downward skip free sequence of random matrices.

In this section, $(\mathrm{X}_{\mathrm{k}} = (X_{k_j}^{i,j})_{i,j \in [d]},\ \mathrm{k} \in \mathbb{Z}_+^d)$ will be a downward skip free M_+-valued random walk, that is $(\mathrm{X}_{\mathrm{k}},\ \mathrm{k} \in \mathbb{Z}_+^d)$ is almost surely a downward skip free sequence of matrices, as defined in the previous subsection and for all $\mathrm{k}, \mathrm{m} \in \mathbb{Z}_+^d$ such that $\mathrm{k} \leq \mathrm{m}$, $\mathrm{X}_{\mathrm{m}} - \mathrm{X}_{\mathrm{k}}$ is independent of the family of random matrices $(\mathrm{X}_{\mathrm{s}}, \mathrm{s} \leq \mathrm{k})$. Note that this special construction implies that the processes defined by the column vectors $(X_k^{i,j},\ i \in [d])_{k \geq 0}$, for $j \in [d]$ are independent $\mathbb{Z}^d$-valued random walks. Let us denote by

$$T_{\mathrm{r}} = (T_{\mathrm{r}}^{(1)}, \ldots, T_{\mathrm{r}}^{(d)}) = \inf\{\mathrm{k} : \overline{\mathrm{X}}_{\mathrm{k}} = -\mathrm{r}\},$$

the first passage time of $(\mathrm{X}_{\mathrm{k}},\ \mathrm{k} \in \mathbb{Z}_+^d)$ at level $-\mathrm{r}$, as defined in Lemma 3 if it exists and set $T_{\mathrm{r}}^{(i)} = \infty$, for all $i \in [d]$, if it does not exist.

Proposition 2. *Let* $\mathrm{r} \in \mathbb{Z}_+^d$*, such that* $r_1 + \cdots + r_d > 1$*, and* F_{r} *be a branching forest with progeny distribution* ν*. Then there exists a downward skip free* M_+*-valued random walk* $(\mathrm{X}_{\mathrm{k}},\ \mathrm{k} \in \mathbb{Z}_+^d)$ *with step distribution*

$$P(X_1^{1,j} = a_1,\, \ldots, X_1^{d,j} = a_d) = \nu_j(a_1, \ldots, a_{j-1}, a_j + 1, a_{j+1}, \ldots, a_d), \quad i \in [d], \tag{5.20}$$

for $a_i \in \mathbb{Z}_+$, if $i \neq j$ and $a_j \geq -1$, such that

$$K(F_r) = (\mathrm{X}_\mathrm{k}, 0 \leq \mathrm{k} \leq T_\mathrm{r}).$$

Proof. Although this result is intuitively clear from the branching property and more particularly, from the definition of the law of F_c:

$$P(F_\mathrm{c} = \mathbf{f}) = \prod_{u \in \mathbf{f}} \nu_{c(u)}(p(u)), \quad \mathbf{f} \in \mathscr{F}_\mathrm{c},$$

and the coding presented in Proposition 1, its proof requires more developed arguments. We will not do it here. □

Proposition 1 means that any multitype branching forest is encoded by a downward skip free M_+-valued random walk whose step distribution is given by the progeny distribution. Actually, we presented this result only in the (sub)critical case in the sole concern of simplify the presentation but a version of this result can be obtained for general multitype branching forests. This proposition shows that in the (sub)critical case, $P(T_\mathrm{r}^{(i)} < \infty) = 1$, for all $i \in [d]$, since (sub)critical branching forests issued from r are almost surely finite.

Now we are interested in the law of the first passage time T_r of a downward skip free M_+-valued random walk $(\mathrm{X}_\mathrm{k},\ \mathrm{k} \in \mathbb{Z}_+^d)$ with step distribution given by (5.20). When $d = 1$, it is known that the process $(T_r,\ r \geq 0)$ is a renewal process, that is an increasing random walk. It laws is characterized by its generating function which is given from the generation function of the downward skip free random walk X. Let us extend this result for $d \geq 1$.

Let us define the generating function $\Phi_i : [0,1]^d \to [0,1]$ of T_{e_i} by

$$E(s^{T_{\mathrm{e}_i}}) = \Phi_i(s), \quad s \in [0,1]^d.$$

Recall that $P(T_\mathrm{r}^{(i)} < \infty) = 1$, for all $i \in [d]$ and $\mathrm{r} \in \mathbb{Z}_+^d$ and observe that from the independence and stationarity of the increments of $(\mathrm{X}_\mathrm{k},\ \mathrm{k} \in \mathbb{Z}_+^d)$, for all $\mathrm{r}, \mathrm{r}' \in \mathbb{Z}_+^d$,

$$T_{\mathrm{r}+\mathrm{r}'} \stackrel{(\mathrm{d})}{=} T_\mathrm{r} + \tilde{T}_{\mathrm{r}'}\,.$$

where $\tilde{T}_{\mathrm{r}'}$ is an independent copy of $T_{\mathrm{r}'}$. We derive from this identity that the generating function of T_r satisfies

$$E(s^{T_\mathrm{r}}) = \Phi(s)^\mathrm{r}, \quad s \in [0,1]^d,$$

where we set $\Phi(s) = (\Phi_1(s), \ldots, \Phi_d(s))$ and $\Phi(s)^\mathrm{r} = \Phi_1(s)^{r_1} \ldots \Phi_d(s)^{r_d}$. Let us define the generating function of $\mathrm{r} + \overline{X}_\mathrm{r}$ by

$$\varphi^\mathrm{r}(s) = E(s^{\mathrm{r}+\overline{X}_\mathrm{r}}).$$

Proposition 3. *The generating function of T_{r} is related to this of $\mathrm{r}+\overline{X}_{\mathrm{r}}$ through the following relationship.*

$$\Phi(s)^{\mathrm{r}} = s^{\mathrm{r}}\varphi^{\mathrm{r}}(\Phi(s)). \tag{5.21}$$

Proof. Since $(\mathrm{X}_{\mathrm{k}},\, \mathrm{k}\in\mathbb{Z}_+^d)$ is downward skip free, $T_{\mathrm{r}}\ge \mathrm{r}$ and $\overline{X}_{\mathrm{r}}\ge -\mathrm{r}$, so that we can write

$$T_{\mathrm{r}} = \mathrm{r} + T^{(\mathrm{r})}_{\mathrm{r}+\overline{X}_{\mathrm{r}}},$$

where

$$\mathrm{X}^{(\mathrm{r})}_{\mathrm{k}} = \mathrm{X}_{\mathrm{r}+\mathrm{k}} - \mathrm{X}_{\mathrm{r}},\ \mathrm{k}\in\mathbb{Z}_+^d \quad\text{and}\quad T^{(\mathrm{r})}_{\mathrm{n}} = \inf\{\mathrm{k} : \overline{\mathrm{X}}^{(\mathrm{r})}_{\mathrm{k}} = -\mathrm{n}\}\,.$$

Then from the independence between $(\mathrm{X}^{(\mathrm{r})}_{\mathrm{k}},\, \mathrm{k}\in\mathbb{Z}_+^d)$ and $(\mathrm{X}_{\mathrm{k}},\, \mathrm{k}\le \mathrm{r})$, we derive that

$$\begin{aligned}
\Phi(s)^{\mathrm{r}} &= E\big(s^{\mathrm{r}+T^{(\mathrm{r})}_{\mathrm{r}+\overline{X}_{\mathrm{r}}}}\big)\\
&= s^{\mathrm{r}}\sum_{\mathrm{k}\ge 0} E(s^{T_{\mathrm{k}}})P(\mathrm{r}+\overline{X}_{\mathrm{r}}=\mathrm{k})\\
&= s^{\mathrm{r}}\varphi^{\mathrm{r}}(\Phi(s)),
\end{aligned}$$

which proves our result. □

The solution of equation (5.21) can be made explicit in terms of ν from Lagrange-Good inversion formula. We derive here a more complete result from the coding of multitype forests. When $d=1$ a an explicit way to characterize the law of T_r is given by the so-called ballot theorem (or Kemperman's identity). Let us extend this identity for $d\ge 1$. The next theorem is called the multivariate ballot theorem. It provides the joint law of the first passage time T_{r} and the value of the matrix $X_{T_{\mathrm{r}}}$ in terms of the law of the step distribution of the random walk.

Theorem 5.2. *Let $\mathrm{n},\mathrm{r}\in\mathbb{Z}_+^d$ and the matrix K such that*

$r_i\ge 0$, $r_1+\dots+r_d\ge 1$, $k_{ij}\ge 0$, for $i\ne j$, $-k_{ii}=r_i+\sum_{j\ne i}k_{ij}$ and $n_i\ge -k_{ii}\ge 1$

then the joint law of $(T_{\mathrm{r}},X_{T_{\mathrm{r}}})$ is given by

$$P(T_{\mathrm{r}}=\mathrm{n},\, \mathrm{X}_{T_{\mathrm{r}}}=K) = \frac{\det(-K)}{n_1n_2\dots n_d}P(\mathrm{X}_{\mathrm{n}}=K).$$

Proof. It is a direct application of Theorem 5.1 and Proposition 2. □

Let us emphasize the fact that Theorem 5.2 is true for more general sequences of random matrices. Indeed, we can easily extend this result to cyclically interchangeable downward skip free sequences of random matrices. Let $(\mathrm{X}_{\mathrm{k}},\, \mathrm{k}\in\mathbb{Z}_+^d)$ be such a sequence. Fix $\mathrm{m},\mathrm{n}\in\mathbb{Z}_+^d$ such that $\mathrm{m}\le\mathrm{n}$ and define the sequence $(\mathrm{X}^{(\mathrm{m})}_{\mathrm{k}}, 0\le\mathrm{k}\le\mathrm{n})$ by

$$\mathrm{X}^{(\mathrm{m}),i,j}_{k_j} = \begin{cases} \mathrm{X}^{i,j}_{m_j+k_j} - \mathrm{X}^{i,j}_{m_j} & \text{if } k_j\le n_j-m_j\\ \mathrm{X}^{i,j}_{k_j-(n_j-m_j)} + \mathrm{X}^{i,j}_{n_j} - \mathrm{X}^{i,j}_{m_j} & \text{if } n_j-m_j\le k_j\le n_j. \end{cases} \tag{5.22}$$

A downward skip free sequence of random matrices $(\mathrm{X}_{\mathrm{k}},\,\mathrm{k}\in\mathbb{Z}_+^d)$ is said to be cyclically interchangeable if for all $\mathrm{m},\mathrm{n}\in\mathbb{Z}_+^d$ such that $\mathrm{m}\le\mathrm{n}$,

$$(\mathrm{X}_{\mathrm{k}}^{(\mathrm{m})},0\le\mathrm{k}\le\mathrm{n})\stackrel{(\mathrm{d})}{=}(\mathrm{X}_{\mathrm{k}},0\le\mathrm{k}\le\mathrm{n}).$$

We will now apply this coding to the expression of the law of the total number of vertices with a given degree. In order to simplify the presentation we will restrict ourself to the number of leaves in the multitype forest, that is the number of vertices with no children. Let F_{r} is be a multitype branching forest. Recall that $p(F_{\mathrm{r}})$ denotes the total progeny of F_{r}. Then let $l(F_{\mathrm{r}})=(l_1(F_{\mathrm{r}}),\dots,l_d(F_{\mathrm{r}}))$ be its number of leaves, that is $l_i(F_{\mathrm{r}})$ is the total number of vertices of type i with no children in F_{r}.

Theorem 5.3. *Assume that the progeny distribution ν is non-degenarate, primitive and critical or subcritical. Then for all $\mathrm{r}\in\mathbb{Z}_+^d$, such that $r_1+\dots+r_d>0$, for all Laplacian matrix K and for all $\mathrm{n}\in\mathbb{Z}_+^d$, such that $r_i\ge0$, $r_1+\dots+r_d\ge1$ and $k_{ij}\ge0$, for $i\ne j$, $-k_{ii}=r_i+\sum_{j\ne i}k_{ij}$, $n_i\ge-k_{ii}$, $n_i\ge l_i$,*

$$P(p(F_{\mathrm{r}})=(K,\mathrm{n}),l(F_{\mathrm{r}})=\mathrm{l})=$$
$$\frac{\det(-\overline{K})}{\bar n_1\bar n_2\dots\bar n_d}\prod_{j=1}^{d}\begin{pmatrix}n_j\\ l_j\end{pmatrix}\nu_j(0)^{l_j}(1-\nu_j(0))^{n_j-l_j}\bar\nu_j^{*(n_j-l_j)}(h'_{1j},\dots,h'_{dj})\,,\quad(5.23)$$

where $\bar\nu_j(0)=0$ and $\bar\nu_j(\mathrm{k})=\nu_j(\mathrm{k})/(1-\nu_j(0))$, for $k\in\mathbb{Z}_+^d\setminus\{0\}$ and $h'_{ij}=k_{ij}+l_je_j$ if $i\ne j$ and $h'_{ii}=n_i+k_{ii}$. Moreover, $\bar n_i=n_i\vee1$, and $\overline{K}$ is the matrix K to which we removed the line i and the column i, for all i such that $n_i=0$ and we set $\nu_j^{*0}=\delta_0$.

Proof. From Proposition 2, the number of leaves $l(F_{\mathrm{r}})$ of type j corresponds to the number of jumps of size $-e_j$ in the sequence $((\mathrm{X}_{k_j}^{i,j},i\in[d]),0\le k_j\le T_{\mathrm{r}}^{(j)})$. Denote by Δ_j this number and set $\Delta=(\Delta_1,\dots,\Delta_d)$. Then according to Proposition 2,

$$P(p(F_{\mathrm{r}})=(K,\mathrm{n}),l(F_{\mathrm{r}})=\mathrm{l})=P(T_{\mathrm{r}}=\mathrm{n},\,\mathrm{X}_{T_{\mathrm{r}}}=K,\Delta=\mathrm{l}).$$

Note that under the conditional probability $P(\,\cdot\,|\,\Delta=\mathrm{l})$, the coding random walk $(\mathrm{X}_{\mathrm{k}},\,\mathrm{k}\in\mathbb{Z}_+^d)$ is a downward skip free sequence of random matrices with cyclically interchangeable increments. Indeed this number is invariant under the cyclic permutations defined in (5.22). Therefore, from Theorem 5.2 and the remark after this theorem, we can write

$$P(T_{\mathrm{r}}=\mathrm{n},\,\mathrm{X}_{T_{\mathrm{r}}}=K\,|\,\Delta=\mathrm{l})=\frac{\det(-K)}{n_1n_2\dots n_d}P(\mathrm{X}_{\mathrm{n}}=K\,|\,\Delta=\mathrm{l})\,,$$

so that

$$P(T_{\mathrm{r}}=\mathrm{n},\,\mathrm{X}_{T_{\mathrm{r}}}=K,\,\Delta=\mathrm{l})=\frac{\det(-K)}{n_1n_2\dots n_d}P(\mathrm{X}_{\mathrm{n}}=K,\,\Delta=\mathrm{l})\,.$$

Then let us develop this last term.

$$
\begin{aligned}
& P(\mathrm{X_n} = K,\, \Delta = \mathrm{l}) \\
= \;& \prod_{j\in[d]} P(\mathrm{X}_{n_j}^{\cdot,j} = k_{ij},\, \Delta_j = l_j) \\
= \;& \prod_{j\in[d]} \sum_{1\le m_1<\dots<m_{l_j}\le n_j} P(X_h^{\cdot,j} - X_{h-1}^{\cdot,j} = -e_j, h \in \{m_1,\dots,m_{l_j}\}) \\
& \times P(\sum_{h\notin\{m_1,\dots,m_{l_j}\}} X_h^{\cdot,j} - X_{h-1}^{\cdot,j} = k_{\cdot j} + l_j e_j,\, X_h^{\cdot,j} - X_{h-1}^{\cdot,j} \neq -e_j,\, h \notin \{m_1,\dots,m_{l_j}\}).
\end{aligned}
$$

Then this last probability can be written as

$$
\begin{aligned}
& P(\sum_{h\notin\{m_1,\dots,m_{l_j}\}} X_h^{\cdot,j} - X_{h-1}^{\cdot,j} = k_{\cdot j} + l_j e_j,\, X_h^{\cdot,j} - X_{h-1}^{\cdot,j} \neq -e_j,\, h \notin \{m_1,\dots,m_{l_j}\}) \\
= \;& P(X_{n_j-l_j}^{\prime\cdot,j} = k_{\cdot j} + e_j l_j) P(X_h^{\cdot,j} - X_{h-1}^{\cdot,j} \neq -e_j,\, h \notin \{m_1,\dots,m_{l_j}\}),
\end{aligned}
$$

where $(X_k^{\prime\cdot,j})_{k\ge 0}$ has law $\bar{\nu}_j$ defined in the statement. Then it remains to observe that

$$
\left\{ \begin{array}{l}
P(X_h^{\cdot,j} - X_{h-1}^{\cdot,j} = -e_j, h \in \{m_1,\dots,m_{l_j}\}) = \nu_j(0)^{l_j} \\
P(X_h^{\cdot,j} - X_{h-1}^{\cdot,j} \neq -e_j,\, h \notin \{m_1,\dots,m_{l_j}\}) = (1-\nu_j(0))^{n_j-l_j}.
\end{array} \right.
$$

□

References

[1] D. Aldous: The random walk construction of uniform spanning trees and uniform labelled trees. SIAM J. Discrete Math. 3 (1990), no. 4, 450–465.

[2] V. Anantharam and P. Tsoucas: A proof of the Markov chain tree theorem. *Statist. Probab. Lett.* 8 (1989), no. 2, 189–192.

[3] K.B. Athreya and P.E. Ney: *Branching processes.* Dover Publications, Inc., Mineola, NY, 2004.

[4] O. Bernardi and A.H. Morales: Counting trees using symmetries. *J. Combin. Theory* Ser. A 123 (2014), 104–122.

[5] P. Biane: Polynomials associated with finite Markov chains. In memoriam Marc Yor – Séminaire de Probabilités XLVII, 249–262, Lecture Notes in Math., 2137, Springer, Cham, 2015.

[6] P. Biane and G. Chapuy: Laplacian matrices and spanning trees of tree graphs. *Preprint* arXiv:1505.04806v4.

[7] A. Broder: Generating random spanning trees. In *Proc. 30'th IEEE Symp. Found. Comp. Sci.*, pages 442–447, 1989.

[8] L. Chaumont and R. Liu: Coding multitype forests: application to the law of the total population of branching forests. *Trans. Amer. Math. Soc.* 368, no. 4, 2723-2747, (2016).

[9] S. Chewi and V. Anantharam: A combinatorial proof of a formula of Biane and Chapuy. *Preprint,* arXiv:1705.10925.

[10] P. Flajolet and R. Sedgewick: *Analytic combinatorics.* Cambridge University Press, Cambridge, 2009.

[11] A.D. Ventcel and M.I. Freidlin: Small random perturbations of dynamical systems. *Uspehi Mat. Nauk*, 25(1 (151)):3–55, 1970.

[12] G. Kirchhoff: Über die Auflösung der Gleichungen, auf welche man bei der Untersuchungen der linearen Vertheilung galvanisher Ströme geführt wird. *Ann. Phys. Chem.* 72, 497–508, (1847).

[13] J. Norris: *Markov chains.* Cambridge University Press.

[14] F.T. Leighton and R.L. Rivest Estimating a probability using finite memory. In Foun- dations of computation theory (Borgholm, 1983), volume 158 of *Lecture Notes in Comput. Sci.*, 255–269. Springer, Berlin, 1983.

[15] F.T. Leighton and R.L. Rivest: The Markov chain tree theorem. M.I.T Laboratory for Computer Science, *Technical Report*, MIT/LCS/TM-249, 1983.

[16] J.-F. Le Gall: Random trees and applications. *Probab. Surv.* 2 (2005), 245–311.

[17] R. Lyons and Y. Peres: *Probability on trees and networks.* Cambridge Series in Statistical and Probabilistic Mathematics, 42. Cambridge University Press, New York, 2016.

[18] J. Pitman and W. Tang: Tree formulas, mean first passage times and Kemeny's constant of a Markov chain. *Preprint* arXiv:1603.09017

[19] R.P. Stanley: *Enumerative combinatorics.* Vol. 2. Cambridge Studies in Advanced Mathematics, 62. Cambridge University Press, Cambridge, 1999.

[20] W.T. Tutte: The dissection of equilateral triangles into equilateral triangles. *Proc. Cambridge Philos. Soc.* 44, 463–482, (1948).

[21] D.B. Wilson: Generating random spanning trees more quickly than the cover time. *Proceedings of the Twenty-eighth Annual ACM Symposium on the Theory of Computing* (Philadelphia, PA, 1996), 296–303, ACM, New York, 1996.