

Chapter 7.1. Selected Topics: Arborescences formalism and last passage

In this chapter we want to give an explicit characterization of invariant measures and hitting probabilities for finite irreducible Markov chains.

Let us consider a Markov chain on a state space S . As usual, we associate to the chain a directed graph, where an edge (x, y) is present if $p_{x,y} > 0$, and we consider the probabilities as weights on the directed edges. Now imagining that, at time $t = 0$, at each point of S there is a friendly particle. Each particle performs a Markov chain with transition probabilities $p_{x,y}$. Moreover they cannot see beyond the vertex they occupy at any given time, so each particle just proceeds independently until they meet another particle on the same vertex of S . As they are very friendly, once two or more particles meet, they continue their random journey together; one usually says that the particles **coalesce** in this case. Notice that each particle still performs a Markov chain with the same transition probability, except that they are not independent.

Furthermore, these unusual particles do the following: each time they (alone or in a group) leave a vertex to jump to a new one, they paint a colored arrow pointing to the vertex where they are going. And as soon as they arrive at a new vertex, they erase any outgoing colored arrow that some particle may have painted before. If we take a snapshot of the system after some steps, observing the vertices and painted arrows, we see the following:”

- There are two types of vertices in S : those occupied by particles, that we will call the *roots* and have no outgoing painted arrows; and the empty ones, from which there is exactly one outgoing painted arrow (the arrow outgoing from x will point to the vertex where the last group of particles leaving x jumped to).
- There is at least one root.
- From each vertex x , there is exactly one path of painted arrows from x to a root (this path has length 0 if x is itself a root). Thus, the procedure induces a random partition of S into subsets, where all points that can reach the same root are in the same component of the partition. Each component is then called an *in-arborescence*.
- If we think that the painted arrows are a subset of the set of original edges of S , we see that we are considering a subgraph of the original graph, which contains all the vertices but only some of the edges of the original graph. Moreover, each in-arborescence (each connected component) has exactly one edge less than the number of vertices, and as such is acyclic (there is no closed “painted” path).
- A subgraph with these properties, precisely described below in mathematical terms, is called a *spanning in-forest* (as a union of in-arborescences).

If the original chain is finite and aperiodic, if we wait sufficiently long, all the particles will meet, and we will see a unique in-arborescence evolving randomly over time. Remarkably, the dynamics of in-forests (or of a single in-arborescence), despite preserving memories of the past history of the particles (the edge across which they jumped the last time they left a vertex), is itself a Markov chain. For this “derived” Markov chain many quantities (e.g. the invariant measure) can be written with explicit formulas. Although these formulas are combinatorially complex, they yield tractable expressions in certain regimes, see Exercise 0.1.

Arborescences

Definition 0.1 (in-arborescence). Let S be a finite directed graph. A **spanning in-forest** \mathfrak{F} is a subgraph of S such that

- \mathfrak{F} contains all the vertices in S (*spanning* property). In particular we identify \mathfrak{F} with a subset of the edges on the original graph.
- Each vertex $x \in S$, has at most one outgoing arrow in \mathfrak{F} .
- \mathfrak{F} is acyclic.

A vertex $z \in S$ with no incoming edges is called a **leaf** of \mathfrak{F} . A vertex $r \in S$ with no outgoing edges in \mathfrak{F} , is called a **root**. The set of vertices x such that $x \rightarrow r$ in \mathfrak{F} is called a rooted **in-arborescence**¹, and r is the root of the in-arborescence.

A spanning in-forest with just one root r is called a **spanning in-arborescence rooted at r** . Equivalently a spanning in-arborescence rooted at r is a subgraph θ of S such that

- θ contains all the vertices in S .
- For each vertex $x \in S \setminus \{r\}$, there is exactly one outgoing edge from x , while there is no outgoing arrow from r .
- For each vertex $x \in S$, there is a directed path $x \rightarrow r$.

To get an idea, if we represent a given Markov chain with a weighted directed graph

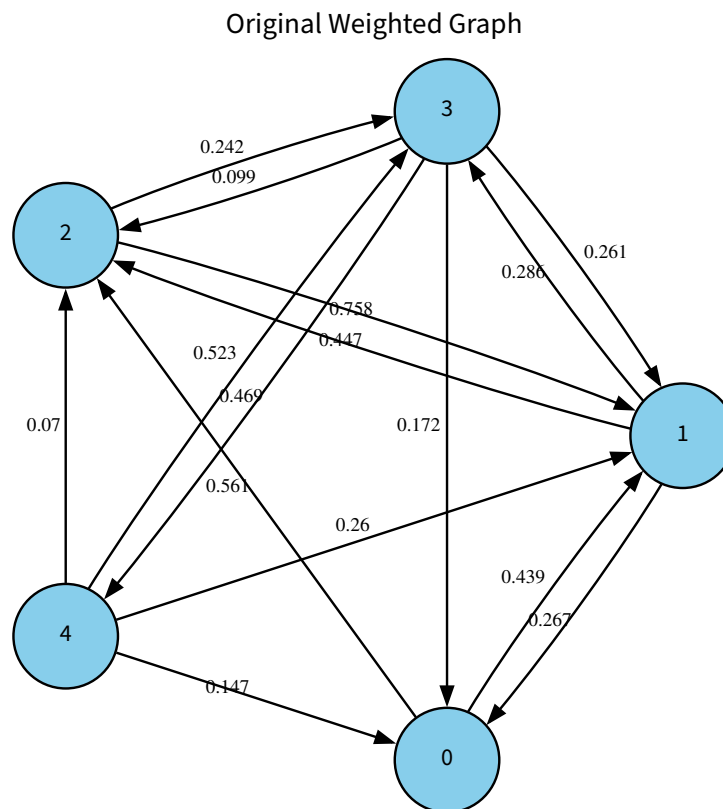


Figure 1

¹*Arborescence* etymology comes from the Latin *arbor* (tree), *arborescere* (to grow into a tree), words with proto indoeuropean roots. In Mathematics, the term *tree* usually refers to undirected acyclic graphs, while *arborescence* is the directed equivalent. Usually an *arborescence* or *out-arborescence* has edges directed outward from the root, while an *in-arborescence* (our case) has edges “toward” the root. In-forests and in-arborescences are sometimes called anti-forests and anti-arborescences in Graph Theory, emphasizing the direction of the edges.

Last passage Markov chain

We want to define transition probabilities (and thus a Markov chain) on Θ following this simple idea: when the chain is in a state (in-arborescence) θ , we look at all the edges outgoing from the root r of θ . Then, with probability $p_{r,y}$ we jump to a new in-arborescence (rooted at y) obtained as follows: we add to θ the edge (r, y) and remove the unique edge outgoing from y . Let us define it more precisely.

For $\theta \in \Theta_r$ and $r' \in S$ such that $p_{r,r'} > 0$, define $\theta^{r,r'}$ as the graph obtained from θ adding the directed edge (r, r') and removing the unique edge in θ outgoing from r' . This new graph $\theta^{r,r'}$ has the following properties:

- It contains all the vertices in S .
- From each vertex $x \in S \setminus \{r'\}$ there is exactly one outgoing edge at x .
- From each vertex $x \in S$, there is a directed path to r' . Indeed, either there was a path in θ from x to r passing through r' (and in this case x still is connected to r' since the edge we removed was not used in the path), or there was a path in θ from x to r not passing through r' (and in this case, the same path plus the edge (r, r') is a path from x to r').

Thus, $\theta^{r,r'} \in \Theta_{r'}$. We then define the transition probabilities on Θ as

$$q_{\theta,\theta'} = \begin{cases} p_{x,y} & \text{if } \theta \in \Theta_x \text{ and } \theta' = \theta^{x,y} \text{ for some } x, y \in S \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

and we let Q be the associated Markov operator $Q = (q_{\theta,\theta'})$. To fix ideas, let us denote $\mathbf{X} = (X_t)_{t \in \mathbb{N}}$ a generic Markov chain on S and $\xi = (\xi_t)_{t \in \mathbb{N}}$ a generic Markov chain on Θ .

One step of the Markov Chain on Θ



Figure 4: The Markov chain on Θ works as follows. When the state is the in-arborescence θ on the left, we can select each arrow outgoing from the root 0 with probability given by the original Markov chain (dashed edges). Say we select the arrow $(0, 2)$ (red dashed edge). Then we add this arrow and remove the one outgoing from 2. The result is the in-arborescence on the right.

We can define the root map $\iota: \Theta \rightarrow S$ that associates to an in-arborescence θ its root $\iota(\theta)$. It is important to note that the root of the Markov chain on Θ performs a Markov chain on S with transition probabilities given by the original

$p_{x,y}$. More formally, for all functions $f: S \rightarrow \mathbb{R}$

$$Q(f \circ \iota) = (Pf) \circ \iota \quad (2)$$

In other words, we can couple the Markov chains as $X_t = \iota(\xi_t)$. In particular, if μ is an invariant measure for Q (that is for the Markov chain ξ on Θ), then $m = \mu \circ \iota^{-1}$ is invariant for P (that is, for the original Markov chain \mathbf{X} on S).

We next want to check that if the original Markov chain is irreducible, so is the Markov chain on Θ .

Proposition 0.1. *Assume that the original Markov chain has a unique closed class (Θ is empty otherwise). The Markov chain on Θ is irreducible iff each transient state in the original Markov chain has a unique outgoing edge.*

Before proceeding with the proof, let us introduce some notation and preliminary result. Given $\theta \in \Theta$ and $x \neq \iota(\theta)$, denote by $e^{\theta,+}(x)$ the unique edge outgoing from x in θ , and $e^{\theta,-}(x)$ the unique edge incoming to $\iota(\theta)$ on the path $x \rightarrow \iota(\theta)$ in θ (the last edge of that path).

Lemma 0.1. *Assume that the original chain \mathbf{X} is irreducible. Then for any $\theta \in \Theta$ and initial point $z \in S$, there exists a finite path in S starting at z and ending at r , such that for every $x \neq r$, the last exit from x in the path is along the edge $e^{\theta,+}(x)$ outgoing from x in θ .*

Proof. We can order the points in S as $r = x_0, x_1, x_2, \dots, x_{|S|-1}$, so that the minimal number of steps needed to reach x_i from r is non-decreasing in i . To define the desired path, first go from z to r , and then build the path recursively by chaining $|S| - 1$ loops. First, go from r to $x_{|S|-1}$ along a path of minimal length, and then back to r following the unique path $x_{|S|-1} \rightarrow r$ in θ . Then proceed recursively for $|S| - 1$ loops, the i -th loop starting at r , going to $x_{|S|-i}$ along a path of minimal length, then back to r along the unique path $x_{|S|-i} \rightarrow r$ in θ . Since we follow path of minimal length, in the i -th loop, the path never passes through x_k for $k > |S| - i$, while on the way back $x_{|S|-i} \rightarrow r$ we only use edges in θ . Therefore, for each i , the last time the path passed through x_i it left x_i along an edge in θ . \square

Proposition 0.1. First, we prove that if the chain is irreducible, each transient state has a unique outgoing edge. Since there is a unique closed class C , the transient states are those in $S \setminus C$, and in the original graph they necessarily have at least one outgoing edge (being transient). All the elements of Θ are rooted in C (since Θ_x is empty for transient x). Since C is closed, outgoing edges in C point to elements of C itself. Thus, the Markov chain on Θ never changes the edges outgoing from transient states. Namely, if $\theta \rightarrow \theta'$, all the outgoing edges at transient states are the same for θ and θ' . Therefore, if the walk in Θ is irreducible, each transient state has a unique outgoing edge.

Next, we prove that, if each transient state has a unique outgoing edge (and there is a unique closed class), then these edges are part of each in-arborescence in Θ , and thus it is enough to restrict to the dynamics on the closed class. In other words, it is enough to prove that if the original chain is irreducible, the chain on Θ is irreducible. To this end, consider a path $(\theta_0, \theta_1, \dots)$ in Θ , and the path of its root $r_t := \iota(\theta_t)$. For a given $t \geq 0$, an edge (x, y) is in θ_t , iff:

- (x, y) was in θ_0 and the root never passed in x before t : $r_s \neq x$ for $s \leq t$.
- at the last (largest) time $s < t$ where $r_s = x$, it left x via y . In other words, $r_{s_x+1} = y$ where $s_x = \max\{s < t : r_s = x\}$.

Therefore, for any fixed initial in-arborescence θ_0 and target θ , we can use the finite path (z_0, z_1, \dots, z_N) built in Lemma 0.1, and follow the transitions $\theta_1 = \theta_0^{z_0, z_1}, \dots, \theta_t = \theta_{t-1}^{z_{t-1}, z_t}, \dots$. The root will have visited all the elements of S , leaving the last time along the edge in θ , and thus $\theta_N = \theta$. \square

Explicit invariant measure

Proposition 0.2. *The measure*

$$\mu_\theta := \prod_{e \in \theta} p_e \quad (3)$$

is invariant for the Markov chain on Θ . In particular, the measure

$$m_x := \sum_{\theta \in \Theta_x} \prod_{e \in \theta} p_e$$

is invariant for the Markov chain on S .

Remark. For each $\theta \in \Theta$, rooted say at $r = \iota(\theta)$, there is a one-to-one correspondence between edges (r, y) outgoing from r in the original graph (with $p_{r,y} > 0$), and in-arborescences θ' such that $q_{\theta',\theta} > 0$. Indeed, we can associate to the edge (r, y) , the in-arborescence θ' obtained by adding the edge (r, y) to θ and removing the edge $e^{\theta,-}(y)$. The map is one-to-one and, with the notation Equation 1, $\theta = (\theta')^{r,y}$. In particular

$$\sum_{\theta': q_{\theta',\theta} > 0} p_{e^{\theta',+}(r)} = \sum_{y \in S} p_{r,y} = 1 \quad (4)$$

Proposition 0.2. Fix $\theta \in \Theta$ and say that its root is $r \in S$. We need to prove that

$$\sum_{\theta' \in \Theta} \mu_{\theta'} q_{\theta',\theta} = \mu_\theta$$

On the left hand side, the terms in the sum with $q_{\theta',\theta} = p_{\iota(\theta'),r} > 0$ correspond to all the arborescences θ' that can reach θ in one step (with positive probability). That is, those arborescences θ' such that θ is obtained from θ' by adding the edge $(\iota(\theta'), r)$ and removing the unique edge outgoing from r in θ' , denoted $e^{\theta'}(r)$. Then we have, by Equation 3

$$\mu_{\theta'} q_{\theta',\theta} = \mu_\theta p_{e^{\theta'}(r)}$$

Therefore, in view of Equation 4

$$\sum_{\theta' \in \Theta} \mu_{\theta'} q_{\theta',\theta} = \mu_\theta \sum_{\theta' \in \Theta} p_{e^{\theta'}(r)} = \mu_\theta$$

Thus, μ_θ is invariant for the Markov chain on Θ .

Since ι projects the chain on Θ to the original chain on S , see Equation 2, $m = \mu \circ \iota^{-1}$ is invariant for the chain on S . \square

Exercises

Exercise 0.1. Consider an irreducible Markov Chain on a finite state space S with transition probabilities $p_{x,y}^0$. Let $c(x, y) \in \mathbb{R}$ be defined for all edges (x, y) such that $p_{x,y} > 0$. Fix $\beta > 0$ and define the tilted transition probabilities

$$p_{x,y}^\beta := \frac{e^{-\beta c(x,y)}}{Z_x^\beta} p_{x,y}^0$$

- Determine the value of Z_x^β and check that the Markov Chain with transition probabilities $p_{x,y}^\beta$ is still irreducible.

b. Explain why one can assume that $\min_{y: p_{x,y}^0 > 0} c(x, y) = 0$ and verify that in this case

$$\inf_y p_{x,y}^0 \leq Z_x^\beta \leq 1$$

- c. Let m^β be the invariant probability measure associated to the transition probabilities p^β . Characterize the limit $\lim_{\beta \rightarrow \infty} m^\beta$. *Hint: For the sake of simplicity, assume that the values $\sum_{e \in \theta} c(e)$ are distinct for distinct $\theta \in \Theta$, where Θ is the space of [rooted spanning in-arborescences](#).*
- d. Give an explicit answer in the case $c(x, y) = (V(y) - V(x))^+$, where $V: S \rightarrow \mathbb{R}$ is a given injective function, and for $a \in \mathbb{R}$, $a^+ = \max(a, 0)$. *Hint: For the sake of simplicity, assume that $p_{x,y}^0 > 0$ whenever $p_{y,x}^0 > 0$.*

Solution

a. Since $p_{x,\cdot}^\beta$ is a probability, we have

$$Z_x^\beta := \sum_y e^{-\beta c(x,y)} p_{x,y}^0$$

The Markov chain is still irreducible since $p_{x,y}^\beta > 0$ iff $p_{x,y}^0 > 0$.

b. If we change $c(x, y)$ to $c(x, y) - a(x)$, then $p_{x,y}^\beta$ does not change, regardless of $a(x)$. Therefore, to make the computation simpler, we can take $a(x) = \min_y c(x, y)$. In other words, we just assume $\min_y c(x, y) = 0$. In particular

$$\inf_y p_{x,y}^0 \leq \max_y e^{-\beta c(x,y)} p_{x,y}^0 \leq Z_x^\beta \leq \sum_y e^{-0} p_{x,y}^0 = 1$$

c. [We know that](#)

$$m_x^\beta = c^\beta \sum_{\theta \in \Theta_x} w^\beta(\theta), \quad w^\beta(\theta) := \prod_{e \in \theta} p_e^\beta$$

where $c^\beta > 0$ is the suitable normalization constant (which is just the same sum over $\theta \in \Theta$). Using the definition of $p_{x,y}^\beta$

$$m_x^\beta = c^\beta \sum_{\theta \in \Theta_x} \frac{e^{-\beta \sum_{e \in \theta} c(e)}}{\prod_{y \neq x} Z_y^\beta} w^0(\theta)$$

where the product in the denominator runs over all the $y \in S$ from which there is an outgoing edge, namely all the $y \in S$ except the root x . For $Z^\beta := \prod_x Z_x^\beta$ and $c(\theta) := \sum_{e \in \theta} c(e)$:

$$m_x^\beta = \frac{c^\beta}{Z^\beta} \sum_{\theta \in \Theta_x} Z_x^\beta e^{-\beta c(\theta)} w^0(\theta)$$

We know that for an irreducible chain Θ_x is non-empty for any $x \in S$. And simply estimating the sum with the largest summand (as above)

$$Z_x^\beta \max_{\theta \in \Theta_x} e^{-\beta c(\theta)} w^0(\theta) \leq m_x^\beta \frac{Z^\beta}{c^\beta} \leq |\Theta_x| Z_x^\beta \max_{\theta \in \Theta_x} e^{-\beta c(\theta)} w^0(\theta)$$

and thus using the bound in point b.

$$\min_y p_{x,y}^0 \min_{\theta \in \Theta_x} w^0(\theta) e^{-\beta \min_{\theta \in \Theta_x} c(\theta)} \leq m_x^\beta \frac{Z^\beta}{c^\beta} \leq |\Theta_x| \max_{\theta \in \Theta_x} w^0(\theta) e^{-\beta \min_{\theta \in \Theta_x} c(\theta)}$$

This implies that m_x^β concentrates (exponentially fast as $\beta \rightarrow \infty$) on the point x that carries the **minimal spanning arborescence** for the cost $c(x, y)$, that is on the root of the tree θ that minimizes $c(\theta)$. Indeed, for such an x and $y \neq x$, $m_y^\beta/m_x^\beta \rightarrow 0$ as $\beta \rightarrow \infty$.

- d. We want to show that in the case $c(x, y) = (V(y) - V(x))^+$ the root of the minimal spanning arborescence is the minimizer of $V(\cdot)$ (which is unique since we assumed V injective). One may give a combinatorial proof in this case using $c(x, y) - c(y, x) = V(y) - V(x)$. Instead, let us give a *Markov Chain* proof, making good use of the fact that the minimal spanning arborescence does not depend on the values of $p_{x,y}^0$, but just on the edges on which this probability is non-vanishing.

First assume that the invariant measure m^0 of $(p_{x,y}^0)$ is reversible, then $m_x^\beta := c^\beta e^{-\beta V(x)} Z_x^\beta m_x^0$ is reversible for $(p_{x,y}^\beta)$ since

$$m_x^\beta p_{x,y}^\beta = Z_x^\beta e^{-\beta V(x)} \frac{e^{-\beta(V(y)-V(x))^+}}{Z_x^\beta} m_x^0 p_{x,y}^0 = e^{-\beta \max(V(x), V(y))} m_x^0 p_{x,y}^0 = e^{-\beta \max(V(x), V(y))} m_y^0 p_{y,x}^0 = m_y^\beta p_{y,x}^\beta$$

Thus, since $V(x) + (V(y) - V(x))^+ = \max(V(y), V(x))$, reasoning as above

$$m_x^\beta = c^\beta m_x^0 \sum_y e^{-\beta \max(V(y), V(x))} p_{x,y}^0$$

concentrates on the points x minimizing

$$x \mapsto \min_{y: p_{x,y}^0 > 0} \max(V(y), V(x))$$

In particular, if $p_{x,x}^0 > 0$, it concentrates on the minimizer of $V(x)$.

If m^0 is not reversible, we can change $p_{x,y}^0$ to $q_{x,y}^0 := p_{x,y}^0 + p_{y,x}^0 m_y^0/m_x^0$ (for which m^0 is reversible) without changing the minimal spanning arborescence (which only depends on the $c(x, y)$), so the answer does not change even when m^0 is not reversible.

Exercise 0.2. Consider an irreducible Markov Chain on a finite state space S with transition probabilities $(p_{x,y})$. Recall that we consider S as a graph where the edges are the $e = (x, y)$ such that $p_{x,y} > 0$. The weight of a **rooted spanning in-forest** F is the product

$$w(F) := \prod_{e \in F} p_e$$

For $A \subset S$ nonempty, let \mathfrak{F}_A be the set of rooted spanning in-forests, whose set of roots is exactly A . For $x \in S$ and $r \in A$, define $\mathfrak{F}_A(x \rightarrow r)$ as the set of forests $F \in \mathfrak{F}_A$ such that x is in an in-arborescence rooted at r (in other words, following the unique path in F outgoing from x , one reaches $r \in A$).

Let now $r \in A \subset S$. Since the chain is irreducible and finite, $\mathbb{P}(\tau_A < \infty) = 1$, and let $h(x) \equiv h_{r,A}(x) := \mathbb{P}_x(X_{\tau_A} = r)$. Prove that

$$h(x) = \frac{\sum_{F \in \mathfrak{F}_A(x \rightarrow r)} w(F)}{\sum_{F \in \mathfrak{F}_A} w(F)} \quad (5)$$

Solution

Since the Markov Chain is irreducible and has an invariant measure (being the state finite), h is the **unique**

solution to the problem for the unknown u

$$(I - \mathbf{1}_{A^c}P)u = \mathbf{1}_r$$

Let $g(x)$ be the right hand side of Equation 5, we show that g also shows this equation, so $g = h$.

Clearly $g(x) = \mathbf{1}_r(x)$ for $x \in A$, so we only need to check that $g(x) = (Pg)(x)$ for $x \notin A$. The denominator $\sum_{F \in \mathfrak{F}_A} w(F)$ simplifies on both sides of this equation so it remains to verify

$$\sum_{G \in \mathfrak{F}_A(x \rightarrow r)} w(G) \stackrel{?}{=} \sum_y p_{x,y} \sum_{F \in \mathfrak{F}_A(y \rightarrow r)} w(F), \quad x \notin A$$

Since $\sum_y p_{x,y} = 1$ can we write this as

$$\sum_{z \neq x} \sum_{G \in \mathfrak{F}_A(x \rightarrow r)} p_{x,z} w(G) \stackrel{?}{=} \sum_{y \neq x} \sum_{F \in \mathfrak{F}_A(y \rightarrow r)} p_{x,y} w(F), \quad x \notin A$$

We can match terms on both sides, building a bijection between indices (z, G) (with $x \rightarrow r$ in G) on the left and (y, F) (with $y \rightarrow r$ in F) on the right as follows.

- If $z \rightarrow r$ in G , then we take $y = z$ and $G = F$, in other words these terms are on both sides.
- If $z \not\rightarrow r$, then (x, z) is not in G . We match such a (z, G) with the element (y, F) in the r.h.s., where y is the unique vertex such that $(x, y) \in G$; and F is the forest obtained from G removing (x, y) and adding (x, z) .

This is a bijection between summands in the l.h.s. and r.h.s. and clearly $p_{x,z}w(G) = p_{x,y}w(F)$.

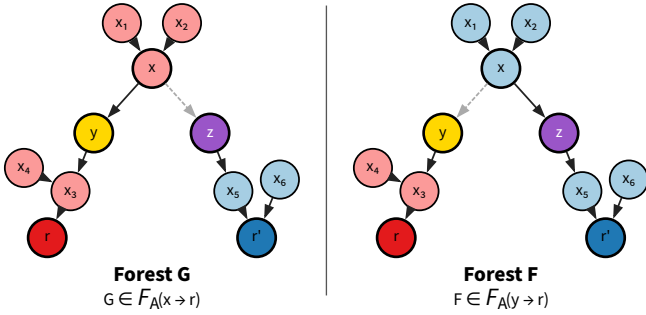


Figure 5: A pair (z, G) with $G \in \mathfrak{F}_A(x \rightarrow r)$ and $z \not\rightarrow r$, is mapped bijectively to a (y, F) with $F \in \mathfrak{F}_A(y \rightarrow r)$, just switching the successor of x .