

# Spectral estimation through Kirchoff's random forest.

Work in progress

S. Barthélémy<sup>1</sup>, F. Castell<sup>2</sup>, A. Gaudillière<sup>2</sup>, C. Mélot<sup>2</sup>, M. Quattropiani<sup>3</sup>, N. Tremblay<sup>1</sup>

<sup>1</sup>GIPSA-lab, Grenoble

<sup>2</sup>I2M, Marseille

<sup>3</sup>Sapienza Università, Roma

GraphLearn, CIRM Marseille, November 2022

# Table of Contents

- 1 Kirchoff's random forest. Estimating  $S(q)$ ,  $q$  fixed.
- 2 Replicated forests. Estimating  $S^{(l)}(q)$ ,  $q$  fixed.
- 3 Coupled forests. Estimating  $q \mapsto S^{(k)}(q)$ ,  $k \leq l$ .
- 4 Estimating the spectral cdf.

## Setup, Notations.

- **Weighted non oriented graph:**  $G = (\mathcal{X}, w)$ ,  
 $\mathcal{X}$  set of vertices,  $n = |\mathcal{X}|$  ;  
 $w : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}^+$  weight function;  
For  $(x, y) \in \mathcal{X} \times \mathcal{X}$ ,  $(x, y)$  is an edge iff  $w(x, y) > 0$ .

Assumption:

$$w(x, y) = w(y, x).$$

## Setup, Notations.

- **Weighted non oriented graph:**  $G = (\mathcal{X}, w)$ ,  
 $\mathcal{X}$  set of vertices,  $n = |\mathcal{X}|$  ;  
 $w : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}^+$  weight function;  
For  $(x, y) \in \mathcal{X} \times \mathcal{X}$ ,  $(x, y)$  is an edge iff  $w(x, y) > 0$ .

Assumption:

$$w(x, y) = w(y, x).$$

- **Graph Laplacian  $L$ :**  $w(x) := \sum_{y \neq x} w(x, y)$  .

$$L(x, y) = w(x, y) \text{ for } x \neq y, \quad L(x, x) = -w(x).$$

$(-L)$  is a positive symmetric matrix with eigenvalues:

$$\lambda_0 = 0 \leq \lambda_1 \leq \dots \leq \lambda_{n-1}.$$

# Aim of the talk

Let  $\mu_L$  be the spectral measure of  $L$ :  $\mu_L = \frac{1}{n} \sum_{i=0}^{n-1} \delta_{-\lambda_i}$ .

- 1 Describe **good Monte-Carlo estimates** of the Stieltjes transform of  $\mu_L$ :

$$\forall q > 0, S(q) = \int \frac{1}{q - z} d\mu_L(z) = \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{q + \lambda_i};$$

# Aim of the talk

Let  $\mu_L$  be the spectral measure of  $L$ :  $\mu_L = \frac{1}{n} \sum_{i=0}^{n-1} \delta_{-\lambda_i}$ .

- 1 Describe **good Monte-Carlo estimates** of the Stieltjes transform of  $\mu_L$ :

$$\forall q > 0, S(q) = \int \frac{1}{q - z} d\mu_L(z) = \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{q + \lambda_i};$$

- 2 and its derivatives:

$$\forall q > 0, \forall k \in \mathbb{N}, S^{(k)}(q) = \frac{(-1)^k k!}{n} \sum_{i=0}^{n-1} \frac{1}{(q + \lambda_i)^k};$$

# Aim of the talk

Let  $\mu_L$  be the spectral measure of  $L$ :  $\mu_L = \frac{1}{n} \sum_{i=0}^{n-1} \delta_{-\lambda_j}$ .

- 1 Describe **good Monte-Carlo estimates** of the Stieltjes transform of  $\mu_L$ :

$$\forall q > 0, S(q) = \int \frac{1}{q - z} d\mu_L(z) = \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{q + \lambda_i};$$

- 2 and its derivatives:

$$\forall q > 0, \forall k \in \mathbb{N}, S^{(k)}(q) = \frac{(-1)^k k!}{n} \sum_{i=0}^{n-1} \frac{1}{(q + \lambda_i)^k};$$

- 3 Derive from these estimates, estimates of the cumulative distribution function:  $\forall t \geq 0, F(t) = \frac{1}{n} \sum_{j=0}^{n-1} \mathbf{1}_{\lambda_j \leq t}$ .

# Aim of the talk

Let  $\mu_L$  be the spectral measure of  $L$ :  $\mu_L = \frac{1}{n} \sum_{i=0}^{n-1} \delta_{-\lambda_j}$ .

- 1 Describe **good Monte-Carlo estimates** of the Stieltjes transform of  $\mu_L$ :

$$\forall q > 0, S(q) = \int \frac{1}{q - z} d\mu_L(z) = \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{q + \lambda_i};$$

- 2 and its derivatives:

$$\forall q > 0, \forall k \in \mathbb{N}, S^{(k)}(q) = \frac{(-1)^k k!}{n} \sum_{i=0}^{n-1} \frac{1}{(q + \lambda_i)^k};$$

- 3 Derive from these estimates, estimates of the cumulative distribution function:  $\forall t \geq 0, F(t) = \frac{1}{n} \sum_{j=0}^{n-1} \mathbf{1}_{\lambda_j \leq t}$ .

This can be done with a **computational cost of order  $n \log(n)$** .

# Aim of the talk

Let  $\mu_L$  be the spectral measure of  $L$ :  $\mu_L = \frac{1}{n} \sum_{i=0}^{n-1} \delta_{-\lambda_j}$ .

- 1 Describe **good Monte-Carlo estimates** of the Stieltjes transform of  $\mu_L$ :

$$\forall q > 0, S(q) = \int \frac{1}{q - z} d\mu_L(z) = \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{q + \lambda_i};$$

- 2 and its derivatives:

$$\forall q > 0, \forall k \in \mathbb{N}, S^{(k)}(q) = \frac{(-1)^k k!}{n} \sum_{i=0}^{n-1} \frac{1}{(q + \lambda_i)^k};$$

- 3 Derive from these estimates, estimates of the cumulative distribution function:  $\forall t \geq 0, F(t) = \frac{1}{n} \sum_{j=0}^{n-1} \mathbf{1}_{\lambda_j \leq t}$ .

This can be done with a **computational cost of order  $n \log(n)$** .

Main tool: **Kirchoff's random forest**.

# Table of Contents

- 1 Kirchoff's random forest. Estimating  $S(q)$ ,  $q$  fixed.
- 2 Replicated forests. Estimating  $S^{(l)}(q)$ ,  $q$  fixed.
- 3 Coupled forests. Estimating  $q \mapsto S^{(k)}(q)$ ,  $k \leq l$ .
- 4 Estimating the spectral cdf.

# Kirchoff's random forest

## Definition

A **rooted oriented spanning forest on  $G$**  is a collection of oriented trees (i.e. connected subgraphs without loops), oriented from the leaves to the root, exhausting the set of vertices.

$$\mathcal{F} = \{ \text{rooted oriented spanning forest on } G \}$$

For  $\phi \in \mathcal{F}$ ,  $\rho(\phi) = \{ \text{roots of } \phi \}$ ,  $w(\phi) = \prod_{e \text{ edge of } \phi} w(e)$ .

# Kirchoff's random forest

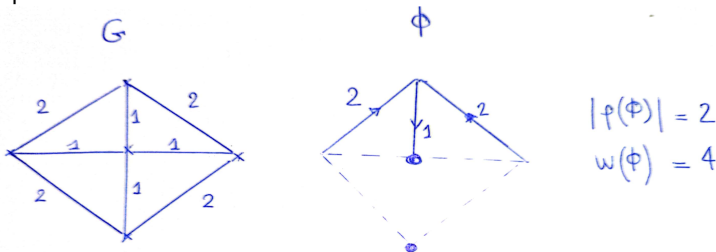
## Definition

A **rooted oriented spanning forest on  $G$**  is a collection of oriented trees (i.e. connected subgraphs without loops), oriented from the leaves to the root, exhausting the set of vertices.

$$\mathcal{F} = \{ \text{rooted oriented spanning forest on } G \}$$

For  $\phi \in \mathcal{F}$ ,  $\rho(\phi) = \{ \text{roots of } \phi \}$ ,  $w(\phi) = \prod_{e \text{ edge of } \phi} w(e)$ .

Example:



# Kirchoff's random forest

Kirchoff's random forest. Parameter  $q > 0$ .

$\Phi_q$  is a random variable with values in  $\mathcal{F}$  such that

$$\forall \phi \in \mathcal{F}, \mathbb{P}(\Phi_q = \phi) = \frac{q^{|\rho(\phi)|} w(\phi)}{Z(q)}, \text{ with } Z(q) = \sum_{\phi \in \mathcal{F}} q^{|\rho(\phi)|} w(\phi).$$

# Kirchoff's random forest

Kirchoff's random forest. Parameter  $q > 0$ .

$\Phi_q$  is a random variable with values in  $\mathcal{F}$  such that

$$\forall \phi \in \mathcal{F}, \mathbb{P}(\Phi_q = \phi) = \frac{q^{|\rho(\phi)|} w(\phi)}{Z(q)}, \text{ with } Z(q) = \sum_{\phi \in \mathcal{F}} q^{|\rho(\phi)|} w(\phi).$$

Link with  $L$  (Kirchoff 1847 [2])

$$Z(q) = \det(q\text{Id} - L).$$

# Kirchoff's random forest

Kirchoff's random forest. Parameter  $q > 0$ .

$\Phi_q$  is a random variable with values in  $\mathcal{F}$  such that

$$\forall \phi \in \mathcal{F}, \mathbb{P}(\Phi_q = \phi) = \frac{q^{|\rho(\phi)|} w(\phi)}{Z(q)}, \text{ with } Z(q) = \sum_{\phi \in \mathcal{F}} q^{|\rho(\phi)|} w(\phi).$$

Link with  $L$  (Kirchoff 1847 [2])

$$Z(q) = \det(q\text{Id} - L).$$

Corollary.

$$|\rho(\Phi_q)| \stackrel{(d)}{=} \sum_{j=0}^{n-1} X_j, \text{ where the } X_j \text{ are independent, } X_j \sim \mathcal{B}\left(\frac{q}{q+\lambda_j}\right). \\ \mathbb{E}(|\rho(\Phi_q)|) = \sum_{j=0}^{n-1} \frac{q}{q+\lambda_j} = qnS(q), \text{ var}(|\rho(\Phi_q)|) \leq \mathbb{E}(|\rho(\Phi_q)|).$$

# Sampling the Kirchoff's random forest.

## Wilson's algorithm:

Let  $X$  be a discrete time Markov chain on  $\mathcal{X} \cup \{\dagger\}$  with transition probabilities

$$P(x, \dagger) = \frac{q}{q + w(x)}, \quad P(x, y) = \frac{w(x, y)}{q + w(x)}, \quad P(\dagger, \dagger) = 1.$$

where  $\dagger$  is a cemetery point.



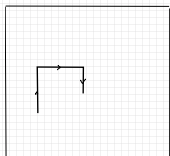
# Sampling the Kirchoff's random forest.

## Wilson's algorithm:

Let  $X$  be a discrete time Markov chain on  $\mathcal{X} \cup \{\dagger\}$  with transition probabilities

$$P(x, \dagger) = \frac{q}{q + w(x)}, \quad P(x, y) = \frac{w(x, y)}{q + w(x)}, \quad P(\dagger, \dagger) = 1.$$

where  $\dagger$  is a cemetery point.



Choose a point  $x \in \mathcal{X}$ . From  $x$ , run  $X$  until death (hitting of  $\dagger$ ).

Erase the loops when they appear.



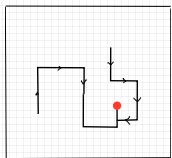
# Sampling the Kirchoff's random forest.

## Wilson's algorithm:

Let  $X$  be a discrete time Markov chain on  $\mathcal{X} \cup \{\dagger\}$  with transition probabilities

$$P(x, \dagger) = \frac{q}{q + w(x)}, \quad P(x, y) = \frac{w(x, y)}{q + w(x)}, \quad P(\dagger, \dagger) = 1.$$

where  $\dagger$  is a cemetery point.



Choose a point outside the drawn trajectory.

Run  $X$  until death, or when it reaches the drawn trajectory.

Erase the loops.

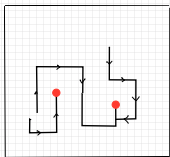
# Sampling the Kirchoff's random forest.

## Wilson's algorithm:

Let  $X$  be a discrete time Markov chain on  $\mathcal{X} \cup \{\dagger\}$  with transition probabilities

$$P(x, \dagger) = \frac{q}{q + w(x)}, \quad P(x, y) = \frac{w(x, y)}{q + w(x)}, \quad P(\dagger, \dagger) = 1.$$

where  $\dagger$  is a cemetery point.



Choose a point outside the drawn trajectory.

Run  $X$  until death, or when it reaches the drawn trajectory.

Erase the loops.

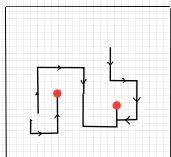
## Sampling the Kirchoff's random forest.

### Wilson's algorithm:

Let  $X$  be a discrete time Markov chain on  $\mathcal{X} \cup \{\dagger\}$  with transition probabilities

$$P(x, \dagger) = \frac{q}{q + w(x)}, \quad P(x, y) = \frac{w(x, y)}{q + w(x)}, \quad P(\dagger, \dagger) = 1.$$

where  $\dagger$  is a cemetery point.



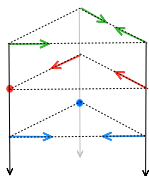
Go on until exhaustion of  $\mathcal{X}$ .

**Theorem: Wilson 1996 [4], Marchal 2000 [3]**

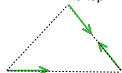
The output of the algorithm is one sample of  $\Phi_q$ , whatever the choice of the starting points (**abelianity**).

## Sampling the Kirchoff's random forest.

**Stack representation, Diaconis & Fulton 1991 [1]:** To each vertex  $x$ , associate an infinite stack of arrays  $x \rightarrow y$ , or stops  $x \rightarrow \dagger$ , sampled from the transition probabilities of  $X$ . Erase the cycles viewed from the top.



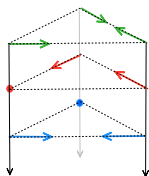
View from the top



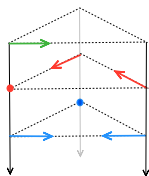


## Sampling the Kirchoff's random forest.

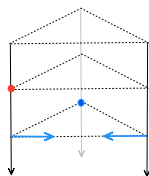
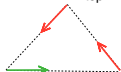
**Stack representation, Diaconis & Fulton 1991 [1]:** To each vertex  $x$ , associate an infinite stack of arrays  $x \rightarrow y$ , or stops  $x \rightarrow \dagger$ , sampled from the transition probabilities of  $X$ . Erase the cycles viewed from the top.



View from the top



View from the top

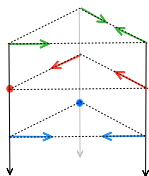


View from the top

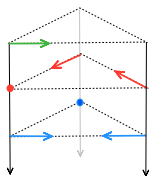


# Sampling the Kirchoff's random forest.

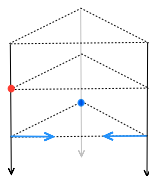
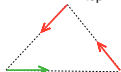
**Stack representation, Diaconis & Fulton 1991 [1]:** To each vertex  $x$ , associate an infinite stack of arrays  $x \rightarrow y$ , or stops  $x \rightarrow \dagger$ , sampled from the transition probabilities of  $X$ . Erase the cycles viewed from the top.



View from the top



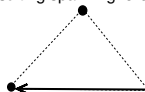
View from the top



View from the top



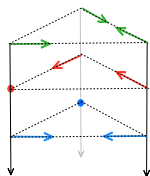
Resulting spanning forest:



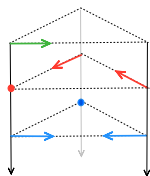
$S = 8$

# Sampling the Kirchoff's random forest.

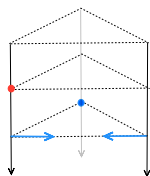
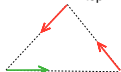
**Stack representation, Diaconis & Fulton 1991 [1]:** To each vertex  $x$ , associate an infinite stack of arrays  $x \rightarrow y$ , or stops  $x \rightarrow \dagger$ , sampled from the transition probabilities of  $X$ . Erase the cycles viewed from the top.



View from the top



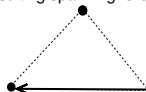
View from the top



View from the top



Resulting spanning forest:

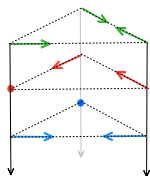


$S = 8$

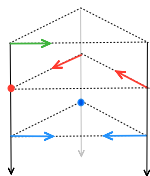
Useful to

# Sampling the Kirchoff's random forest.

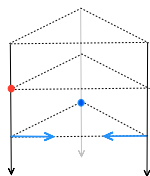
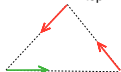
**Stack representation, Diaconis & Fulton 1991 [1]:** To each vertex  $x$ , associate an infinite stack of arrays  $x \rightarrow y$ , or stops  $x \rightarrow \dagger$ , sampled from the transition probabilities of  $X$ . Erase the cycles viewed from the top.



View from the top



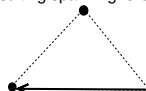
View from the top



View from the top



Resulting spanning forest:



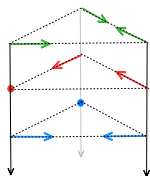
$S = 8$

Useful to

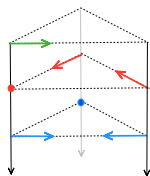
- Prove the abelianity.

# Sampling the Kirchoff's random forest.

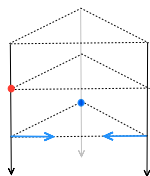
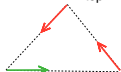
**Stack representation, Diaconis & Fulton 1991 [1]:** To each vertex  $x$ , associate an infinite stack of arrays  $x \rightarrow y$ , or stops  $x \rightarrow \dagger$ , sampled from the transition probabilities of  $X$ . Erase the cycles viewed from the top.



View from the top



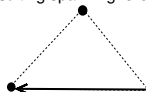
View from the top



View from the top



Resulting spanning forest:



$S = 8$

Useful to

- Prove the abelianity.
- Study the computational cost  $C$  of the algorithm.

## Sampling the Kirchoff's random forest.

Computational cost for sampling the forest  $\Phi_q$ :

$$\begin{aligned} C &= (\text{cost for sampling an arrow}) \times (\text{number of arrows used} \\ &= \log(n) \times S_q \end{aligned}$$

in the stacks)

## Sampling the Kirchoff's random forest.

Computational cost for sampling the forest  $\Phi_q$ :

$$\begin{aligned} C &= (\text{cost for sampling an arrow}) \times (\text{number of arrows used} \\ &= \log(n) \times S_q \end{aligned}$$

### Proposition

Let  $W = \text{diag}(w(x), x \in \mathcal{X})$ , and  $\bar{\lambda} = \frac{1}{n} \text{Trace}(W) = \frac{1}{n} \sum_{k=0}^{n-1} \lambda_k$ . Then ,

$$\mathbb{E}(S_q) = \text{Trace}((q\text{Id} - L)^{-1}(q\text{Id} + W)) \leq n \left( 1 + \frac{\bar{\lambda}}{q} \right).$$

# Sampling the Kirchoff's random forest.

Computational cost for sampling the forest  $\Phi_q$ :

$$\begin{aligned} C &= (\text{cost for sampling an arrow}) \times (\text{number of arrows used} \\ & \quad \text{in the stacks}) \\ &= \log(n) \times S_q \end{aligned}$$

## Proposition

Let  $W = \text{diag}(w(x), x \in \mathcal{X})$ , and  $\bar{\lambda} = \frac{1}{n} \text{Trace}(W) = \frac{1}{n} \sum_{k=0}^{n-1} \lambda_k$ . Then ,

$$\mathbb{E}(S_q) = \text{Trace}((q\text{Id} - L)^{-1}(q\text{Id} + W)) \leq n \left( 1 + \frac{\bar{\lambda}}{q} \right).$$

**Proof:**  $\mathbb{E}(S_q) = \sum_{x \in \mathcal{X}} \mathbb{E}(S_q(x))$ , where  $S_q(x)$  is the number of arrays used in the stack of  $x$ . Using abelianity, and starting the algorithm from  $x$ ,  $S_q(x)$  has the same law as  $\ell_{T_{\dagger}}(x) =$  the number of times the random walk  $X$  visits  $x$  before death.

Now, if  $M(x, y) := \mathbb{E}_x(\ell_{T_{\dagger}}(y))$ ,  $M = (q\text{Id} - L)^{-1}(q\text{Id} + W)$ , and

$$\begin{aligned} \text{Trace}(M) &= \sum_{k=0}^{n-1} \langle u_k, (q\text{Id} - L)^{-1}(q\text{Id} + W)u_k \rangle \\ &= \sum_{k=0}^{n-1} \frac{1}{q + \lambda_k} \langle u_k, (q\text{Id} + W)u_k \rangle \leq \frac{1}{q} \text{Trace}(q\text{Id} + W) \end{aligned}$$

# Table of Contents

- 1 Kirchoff's random forest. Estimating  $S(q)$ ,  $q$  fixed.
- 2 Replicated forests. Estimating  $S^{(l)}(q)$ ,  $q$  fixed.
- 3 Coupled forests. Estimating  $q \mapsto S^{(k)}(q)$ ,  $k \leq l$ .
- 4 Estimating the spectral cdf.

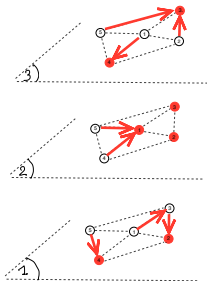
Kirchoff's random forest. Estimating  $S(q)$ ,  $q$  fixed.

Replicated forests. Estimating  $S^{(l)}(q)$ ,  $q$  fixed.

Coupled forests. Estimating  $q \mapsto S^{(k)}(q)$ ,  $k \leq l$ .

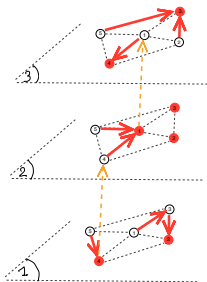
Estimating the spectral cdf.

# Replicated forest



Sample  $l$  independent Kirchoff's forests  $\Phi_q^{(1)}, \dots, \Phi_q^{(l)}$ .

# Replicated forest



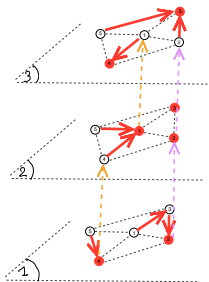
Sample  $l$  independent Kirchoff's forests  $\Phi_q^{(1)}, \dots, \Phi_q^{(l)}$ .  
For  $x \in \mathcal{X}$ ,  $\phi \in \mathcal{F}$ ,

$\rho_x(\phi) =$  root of the tree of  $\phi$  containing  $x$ .

$R^{(1)}(x) := \rho_x(\Phi_q^{(1)})$ ,  $R^{(k)}(x) := \rho_{R^{(k-1)}(x)}(\Phi_q^{(k)})$ .

$R^{(1)}(5) = 4$ ,  $R^{(2)}(5) = 1$ ,  $R^{(3)}(5) = 4$

# Replicated forest



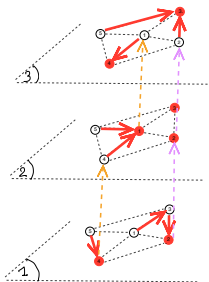
Sample  $l$  independent Kirchoff's forests  $\Phi_q^{(1)}, \dots, \Phi_q^{(l)}$ .  
For  $x \in \mathcal{X}$ ,  $\phi \in \mathcal{F}$ ,

$\rho_x(\phi) =$  root of the tree of  $\phi$  containing  $x$ .

$R^{(1)}(x) := \rho_x(\Phi_q^{(1)})$ ,  $R^{(k)}(x) := \rho_{R^{(k-1)}(x)}(\Phi_q^{(k)})$ .

Let  $\xi_{k,q} := \sum_{x \in \mathcal{X}} \mathbf{1}_{x=R^{(k)}(x)}$ .  $\xi_{3,q} = 2$ .

# Replicated forest



Sample  $l$  independent Kirchoff's forests  $\Phi_q^{(1)}, \dots, \Phi_q^{(l)}$ .  
For  $x \in \mathcal{X}$ ,  $\phi \in \mathcal{F}$ ,

$\rho_x(\phi)$  = root of the tree of  $\phi$  containing  $x$ .

$$R^{(1)}(x) := \rho_x(\Phi_q^{(1)}), \quad R^{(k)}(x) := \rho_{R^{(k-1)}(x)}(\Phi_q^{(k)}).$$

$$\text{Let } \xi_{k,q} := \sum_{x \in \mathcal{X}} \mathbf{1}_{x=R^{(k)}(x)}.$$

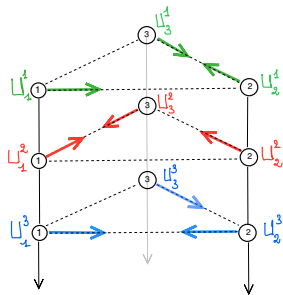
## Estimating the derivatives of $S(q)$

$$\mathbb{E}(\xi_{k,q}) = \sum_{j=0}^{n-1} \binom{q}{q+\lambda_j}^k = \frac{(-1)^k n S^{(k)}(q)}{k! q^k}, \quad \text{var}(\xi_{k,q}) \leq \mathbb{E}(\xi_{k,q}).$$

# Table of Contents

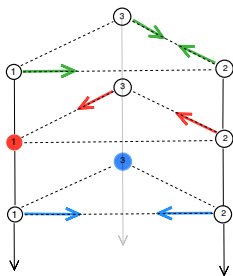
- 1 Kirchoff's random forest. Estimating  $S(q)$ ,  $q$  fixed.
- 2 Replicated forests. Estimating  $S^{(l)}(q)$ ,  $q$  fixed.
- 3 Coupled forests. Estimating  $q \mapsto S^{(k)}(q)$ ,  $k \leq l$ .
- 4 Estimating the spectral cdf.

# Coupled forest



Instead of associating to  $x \in \mathcal{X}$  a stack of stops  
 (sampled with probability  $\frac{q}{q+w(x)}$ ), and arrays  
 ( $x \rightarrow y$  sampled with probability  $\frac{w(x,y)}{q+w(x)}$ ), one can  
 associate a stack of couples  $(U, \text{array})$ ,  
 $U \sim \mathcal{U}([0, 1])$ ,  $\mathbb{P}(x \rightarrow y) = \frac{w(x,y)}{w(x)}$ .

## Coupled forest

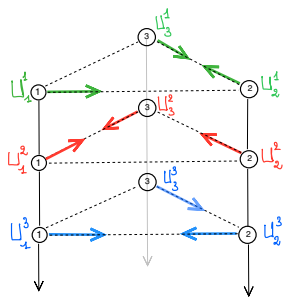


Instead of associating to  $x \in \mathcal{X}$  a stack of stops (sampled with probability  $\frac{q}{q+w(x)}$ ), and arrays ( $x \rightarrow y$  sampled with probability  $\frac{w(x,y)}{q+w(x)}$ ), one can associate a stack of couples  $(U, \text{array})$ ,

$$U \sim \mathcal{U}([0, 1]) , \mathbb{P}(x \rightarrow y) = \frac{w(x,y)}{w(x)} .$$

To sample  $\Phi_q$ , just replace each  $U \leq \frac{q}{q+w(x)}$  by a "stop"; otherwise, just keep the arrow.

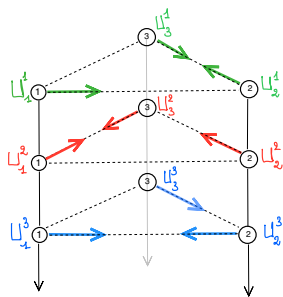
## Coupled forest



Instead of associating to  $x \in \mathcal{X}$  a stack of stops (sampled with probability  $\frac{q}{q+w(x)}$ ), and arrays ( $x \rightarrow y$  sampled with probability  $\frac{w(x,y)}{q+w(x)}$ ), one can associate a stack of couples  $(U, \text{array})$ ,  $U \sim \mathcal{U}([0, 1])$ ,  $\mathbb{P}(x \rightarrow y) = \frac{w(x,y)}{q+w(x)}$ .  
 To sample  $\Phi_q$ , just replace each  $U \leq \frac{q}{q+w(x)}$  by a "stop"; otherwise, just keep the arrow.

- One can sample the whole trajectory  $q \in [q_{min}; q_{max}] \mapsto \Phi_q$  from the same stacks (coupling).

## Coupled forest



Instead of associating to  $x \in \mathcal{X}$  a stack of stops (sampled with probability  $\frac{q}{q+w(x)}$ ), and arrays ( $x \rightarrow y$  sampled with probability  $\frac{w(x,y)}{q+w(x)}$ ), one can associate a stack of couples  $(U, \text{array})$ ,  $U \sim \mathcal{U}([0, 1])$ ,  $\mathbb{P}(x \rightarrow y) = \frac{w(x,y)}{w(x)}$ .

To sample  $\Phi_q$ , just replace each  $U \leq \frac{q}{q+w(x)}$  by a "stop"; otherwise, just keep the arrow.

- One can sample the whole trajectory  $q \in [q_{min}; q_{max}] \mapsto \Phi_q$  from the same stacks (coupling).
- If  $q' < q$ ,  $\frac{q'}{q'+w(x)} \leq \frac{q}{q+w(x)} \Rightarrow S_{q'} \geq S_q$ .  
 $\Rightarrow$  The number of samplings to build  $(q \in [q_{min}; q_{max}] \mapsto \Phi_q)$ , is  $S_{q_{min}}$ .

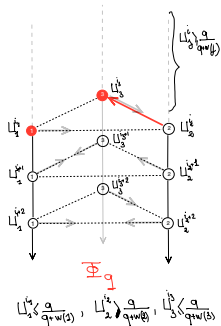
## Coupled forest

Computational cost of sampling  $q \in [q_{min}; q_{max}] \mapsto \Phi_q$ : a little more than  $S_{q_{min}}$ , since some sampled arrays have to be reread one time.

# Coupled forest

Computational cost of sampling  $q \in [q_{min}; q_{max}] \mapsto \Phi_q$ : a little more than  $S_{q_{min}}$ , since some sampled arrays have to be reread one time.

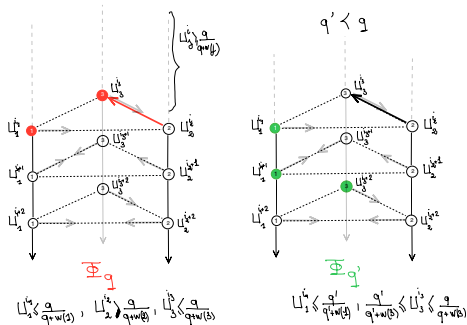
Sampling  $\Phi_{q'}$  from  $\Phi_q$ ,  $q' < q$ :



# Coupled forest

Computational cost of sampling  $q \in [q_{min}; q_{max}] \mapsto \Phi_q$ : a little more than  $S_{q_{min}}$ , since some sampled arrays have to be reread one time.

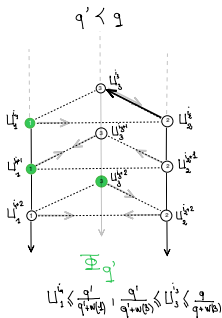
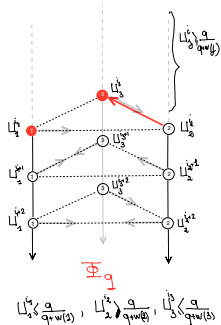
Sampling  $\Phi_{q'}$  from  $\Phi_q$ ,  $q' < q$ :



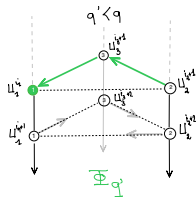
Reactivation of root 3.  
 Array 2  $\rightarrow$  3 is read again.

# Coupled forest

Computational cost of sampling  $q \in [q_{min}; q_{max}] \mapsto \Phi_q$ : a little more than  $S_{q_{min}}$ , since some sampled arrays have to be reread one time.  
**Sampling  $\Phi_{q'}$  from  $\Phi_q$ ,  $q' < q$ :**



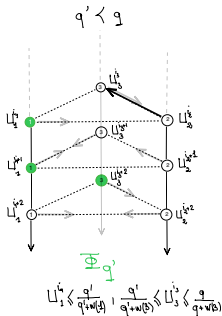
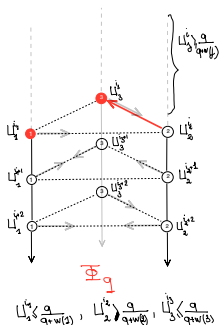
Reactivation of root 3.  
 Array 2  $\rightarrow$  3 is read again.



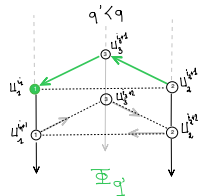
# Coupled forest

Computational cost of sampling  $q \in [q_{min}; q_{max}] \mapsto \Phi_q$ : a little more than  $S_{q_{min}}$ , since some sampled arrays have to be reread one time.

Sampling  $\Phi_{q'}$  from  $\Phi_q$ ,  $q' < q$ :



Reactivation of root 3.  
 Array 2  $\rightarrow$  3 is read again.



$R_{q',q}$  := number of reread arrows  $\leq$  size of the reactivated tree in  $\Phi_q$ .

$$\mathbb{E}(R_{q',q}) \leq n \left( 1 + \log \left( \frac{\alpha}{q_{min}} \right) \right), \text{ where } \alpha = \max_{x \in \mathcal{X}} w(x).$$

## Monte Carlo estimates of $q \mapsto S^{(k)}(q)$ , $k \leq l$ .

$$\text{For } k \leq l, m_k(q) := \sum_{j=0}^{n-1} \left( \frac{q}{q + \lambda_j} \right)^k = \mathbb{E}(\xi_{k,q})$$

- Estimator:  $q \in [q_{min}; q_{max}] \mapsto m_k(q)$  is estimated by  $q \in [q_{min}; q_{max}] \mapsto \hat{m}_k(q) := \frac{1}{S} \sum_{s=1}^S \xi_{k,q}^{(s)}$ , where  $(q \mapsto \xi_{k,q}^{(s)}; s \leq S)$  are i.i.d. copies of  $q \mapsto \xi_{k,q}^{(1)}$  sampled from  $S \times k$  coupled forests.
- Relative error  $\frac{|\hat{m}_k(q) - m_k(q)|}{m_k(q)} \leq \frac{1}{\sqrt{S m_k(q)}}$ .
- Computational cost:

$$S \times l \times \left[ n \log(n) \frac{\bar{\lambda}}{q_{min}} + n \log \left( \frac{\alpha}{q_{min}} \right) \right]$$

with  $\bar{\lambda} = \frac{1}{n} \sum_{x \in \mathcal{X}} w(x) \leq \alpha \leq n \bar{\lambda}$ . Taking  $q_{min}$  of the order of  $\bar{\lambda}$ , we end up with a computational cost of order  $S \times l \times n \log(n)$ .

# Table of Contents

- 1 Kirchoff's random forest. Estimating  $S(q)$ ,  $q$  fixed.
- 2 Replicated forests. Estimating  $S^{(l)}(q)$ ,  $q$  fixed.
- 3 Coupled forests. Estimating  $q \mapsto S^{(k)}(q)$ ,  $k \leq l$ .
- 4 Estimating the spectral cdf.

## Bounds on the spectral cdf.

- **Question:** Assume that we know  $q \in [q_{min}; q_{max}] \mapsto (m_k(q), k \leq l)$ , what can be said about  $F_L : t \in [0; 2\alpha] \mapsto \frac{1}{n} \sum_{j=0}^{n-1} \mathbf{1}_{\lambda_j \leq t}$ ? (note that  $\forall j, \lambda_j \leq 2\alpha$ ).

## Bounds on the spectral cdf.

- **Question:** Assume that we know  $q \in [q_{min}; q_{max}] \mapsto (m_k(q), k \leq l)$ , what can be said about  $F_L : t \in [0; 2\alpha] \mapsto \frac{1}{n} \sum_{j=0}^{n-1} \mathbf{1}_{\lambda_j \leq t}$ ? (note that  $\forall j, \lambda_j \leq 2\alpha$ ).
- **A moment problem.** Let  $Y_q = \frac{q}{q + \lambda_j}$ ,  $J \sim \mathcal{U}(\{0, \dots, n-1\})$ . Then,

$$Y_q \in \left[ \frac{q}{q + 2\alpha}; 1 \right], \quad \mathbb{E}(Y_q^k) = \frac{1}{n} m_k(q), \quad \text{and} \quad \mathbb{P} \left( Y_q \geq \frac{1}{2} \right) = F_L(q).$$

## Bounds on the spectral cdf.

- **Question:** Assume that we know  $q \in [q_{min}; q_{max}] \mapsto (m_k(q), k \leq l)$ , what can be said about  $F_L : t \in [0; 2\alpha] \mapsto \frac{1}{n} \sum_{j=0}^{n-1} \mathbf{1}_{\lambda_j \leq t}$ ? (note that  $\forall j, \lambda_j \leq 2\alpha$ ).
- **A moment problem.** Let  $Y_q = \frac{q}{q+\lambda_J}$ ,  $J \sim \mathcal{U}(\{0, \dots, n-1\})$ . Then,

$$Y_q \in \left[ \frac{q}{q+2\alpha}; 1 \right], \quad \mathbb{E}(Y_q^k) = \frac{1}{n} m_k(q), \quad \text{and} \quad \mathbb{P} \left( Y_q \geq \frac{1}{2} \right) = F_L(q).$$

- **Markov's bounds, Markov 1887.**

$$\mathcal{R}(m) := \left\{ \nu \text{ meas. on } \left[ \frac{q}{q+2\alpha}; 1 \right]; \int x^k d\nu(x) = \frac{m_k(q)}{n}, k = 0, \dots, l \right\}$$

$$\min_{\nu \in \mathcal{R}(m)} \nu\left(\left[\frac{1}{2}, 1\right]\right) \leq \mathbb{P}\left(Y_q \geq \frac{1}{2}\right) \leq \max_{\nu \in \mathcal{R}(m)} \nu\left(\left[\frac{1}{2}, 1\right]\right)$$

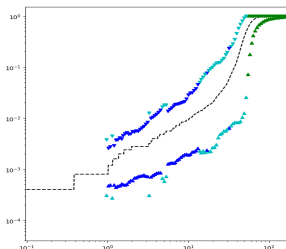
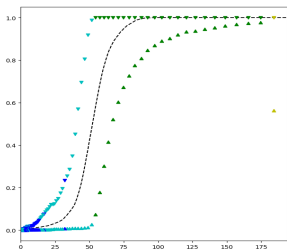
$$\nu_{opt}\left(\left[\frac{1}{2}, 1\right]\right) \rightarrow \nu_{opt} \text{ explicit} \leftarrow \nu_{opt}\left(\left[\frac{1}{2}, 1\right]\right)$$

## Bounds on the spectral cdf.

**Bunny graph:**  $l = 4$ ,  $S = 400$ .

yellow = 1 moment ; green = 2 moments; cyan = 3 moments; blue = 4 moments.

400 replicated forests,  $\alpha / q_0 = 1e+02$ , 2503 nodes



Markov's bounds are not tight as soon as the spectrum is not concentrated on a few points. Actually,

$$\text{card}(\text{Support}(\nu_{opt})) \leq 2 + l/2, \text{ and } 1/2 \in \text{Support}(\nu_{opt}).$$

## Estimating the spectral measure.

Bayesian framework:

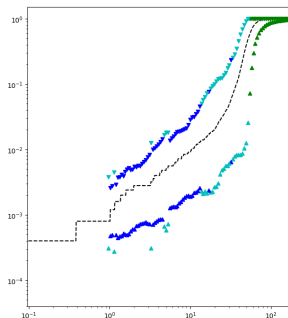
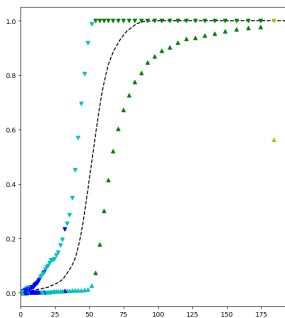
- To estimate:  $\nu_q = \text{law of } Y_q = \frac{1}{n} \sum_{j=0}^n \delta_{q/q+\lambda_j}$ .
- Observations:  $\hat{m}(q)$  noisy version of  $(\int x^k d\nu_q, k = 0, \dots, l)$ .
- Prior on the law of  $\nu_q$ :  $N_q = \frac{1}{n} \sum_{j=0}^n \delta_{Y_j}$ ,  $Y_j$  i.i.d  $\mathcal{U}([q/(q+2\alpha); 1])$ .
- Then, as  $n \rightarrow +\infty$ , the posterior concentrates on  $\hat{\nu}_q$  which maximizes the entropy  $S(\nu) = - \int_{q/(q+2\alpha)}^1 \nu(x) \log(\nu(x)) dx$  under the moment constraints  $\int x^k \nu(x) = \hat{m}_k(q)$ .
- We then estimate  $F_L(q) = P(Y_q \geq 1/2)$  by  $\hat{\nu}_q[1/2; 1]$ .

# Numerical Examples: $l = 4$ , $S = 400$ .

## Bunny graph:

yellow = 1 moment ; green = 2 moments; cyan = 3 moments; blue = 4 moments.

400 replicated forests,  $\alpha / q_0 = 1e+02$ , 2503 nodes

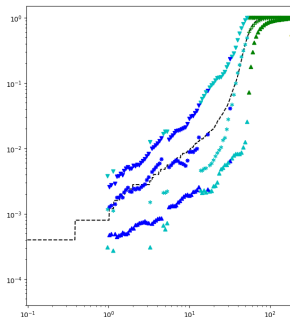
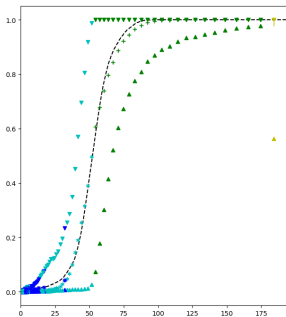


## Numerical Examples: $l = 4$ , $S = 400$ .

### Bunny graph:

yellow = 1 moment ; green = 2 moments; cyan = 3 moments; blue = 4 moments.

400 replicated forests,  $\alpha / q_0 = 1e+02$ , 2503 nodes

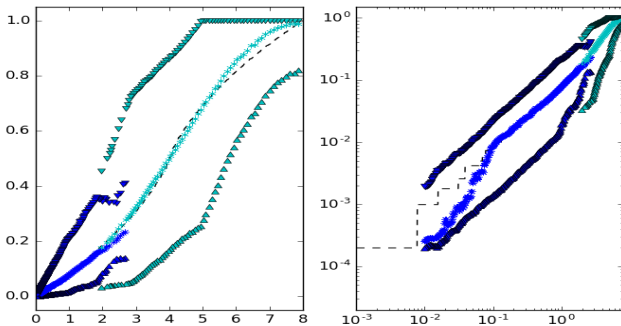


## Numerical Examples: $l = 4$ , $S = 400$ .

### Torus.

yellow = 1 moment ; green = 2 moments; cyan = 3 moments; blue = 4 moments.

400 replicated forests,  $\alpha / q_0 = 4e+02$ , 5041 nodes

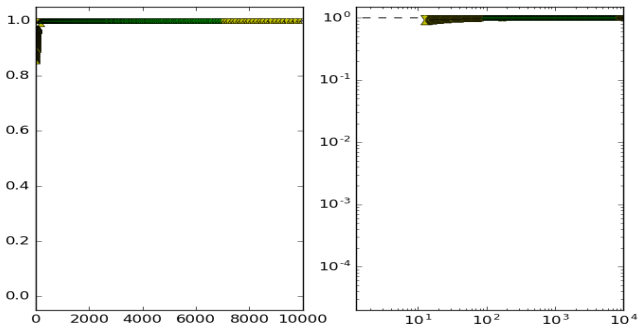


# Numerical Examples: $l = 4$ , $S = 400$ .

Star.

yellow = 1 moment ; green = 2 moments; cyan = 3 moments; blue = 4 moments.

400 replicated forests,  $\alpha / q_0 = 4e+02$ , 5000 nodes

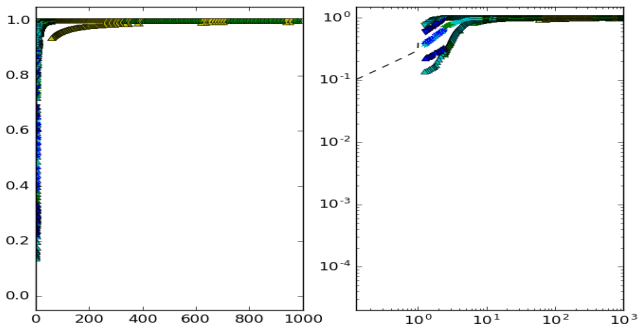


## Numerical Examples: $l = 4$ , $S = 400$ .

### Comet.

yellow = 1 moment ; green = 2 moments; cyan = 3 moments; blue = 4 moments.

400 replicated forests,  $\alpha / q_0 = 4e+02$ , 5000 nodes

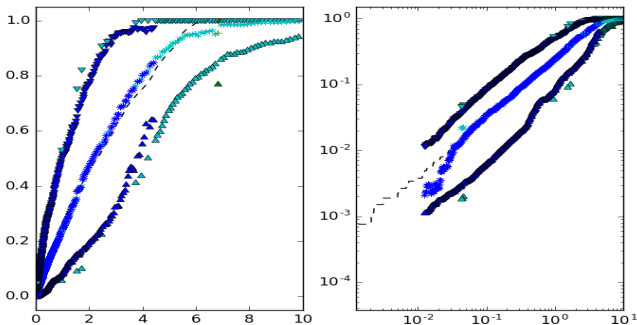


## Numerical Examples: $l = 4$ , $S = 400$ .

### Minnesota

yellow = 1 moment ; green = 2 moments; cyan = 3 moments; blue = 4 moments.

400 replicated forests,  $\alpha / q_0 = 4e+02$ , 2642 nodes

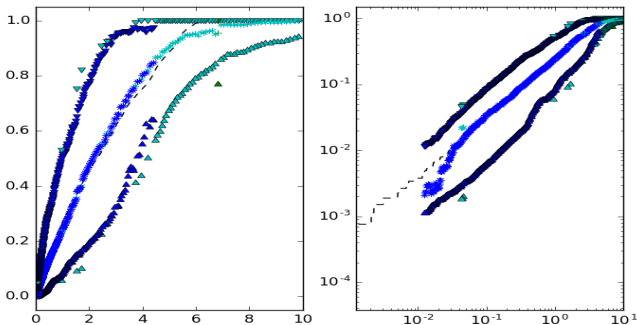


## Numerical Examples: $l = 4$ , $S = 400$ .

### Minnesota

yellow = 1 moment ; green = 2 moments; cyan = 3 moments; blue = 4 moments.

400 replicated forests,  $\alpha / q_0 = 4e+02$ , 2642 nodes

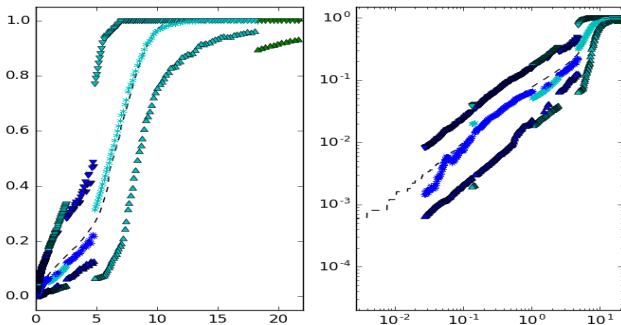


## Numerical Examples: $l = 4$ , $S = 400$ .

### Sensor

yellow = 1 moment ; green = 2 moments; cyan = 3 moments; blue = 4 moments.

400 replicated forests,  $\alpha / q_0 = 4e+02$ , 5000 nodes

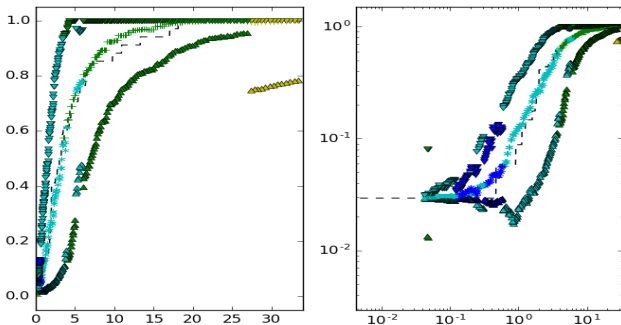


## Numerical Examples: $l = 4$ , $S = 400$ .

### Karate .

yellow = 1 moment ; green = 2 moments; cyan = 3 moments; blue = 4 moments.

400 replicated forests,  $\alpha / q_0 = 4e+02$ , 34 nodes

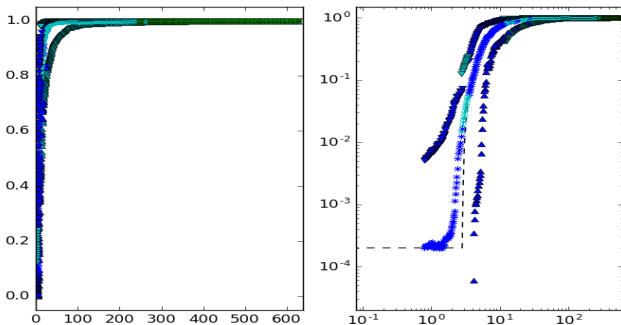


# Numerical Examples: $l = 4$ , $S = 400$ .

## Barabasi-Albert' graph .

yellow = 1 moment ; green = 2 moments; cyan = 3 moments; blue = 4 moments.

400 replicated forests,  $\alpha / q_0 = 4e+02$ , 5000 nodes

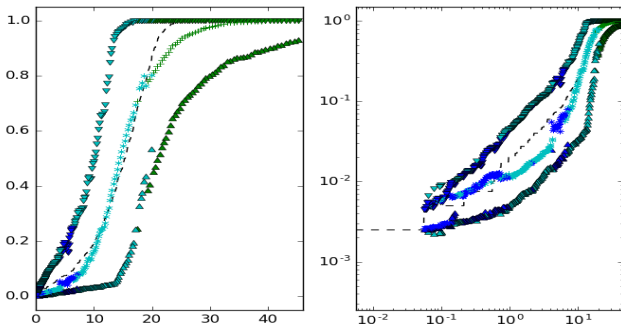


## Numerical Examples: $l = 4$ , $S = 400$ .

### Swiss Roll

yellow = 1 moment ; green = 2 moments; cyan = 3 moments; blue = 4 moments.

400 replicated forests,  $\alpha / q_0 = 4e+02$ , 400 nodes

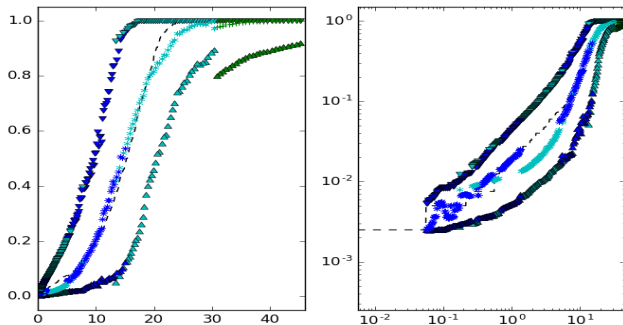


## Numerical Examples: $l = 4$ , $S = 1600$ .

### Swiss Roll

yellow = 1 moment ; green = 2 moments; cyan = 3 moments; blue = 4 moments.

1600 replicated forests,  $\alpha / q_0 = 4e+02$ , 400 nodes



# References

## About the random forest.



Diaconis, P., and Fulton, W., *A growth model, a game, an algebra, Lagrange inversion, and characteristic classes*, Rendiconti del Seminario Matematico Università e Politecnico di Torino, **Vol 49**, no 1, 1991, pp 95–119.



Kirchhoff, Gustav, *Ueber die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Vertheilung galvanischer Ströme geführt wird*, Annalen der Physik, **Vol 148**, no 12, 1847, pp 497–508.



Marchal, P., *Loop-Erased Random Walks, Spanning Trees and Hamiltonian Cycles*, Electronic Communications in Probability, **Vol 5**, 2000, pp 39–50



Wilson, D. B., *Generating random spanning trees more quickly than the cover time*, in STOC '96, 1996.

## About others applications of the random forests



Avena, L., Gaudillière, A., *Two Applications of Random Spanning Forests*, Journal of Theoretical Probability, VOLUME = **Vol 31**, no 4, 2018, pp 1975–2004.



Avena, L., Castell, F. , Gaudillière, A. , Mélot., C. *Random forests and Network analysis*, J. Stat. Phys. 173 (2018), no 3-4, 984–1027.



Avena, L., Castell, F. , Gaudillière, A. , Mélot., C. *Intertwining wavelets or Multiresolution analysis on graphs through random forests* . Appl. Comput. Harmon. Anal. 48 (2020), no. 3, 949–992.