Random subtree generation of a given graph

Luis Fredes
(Work with J.F. Marckert)

ERC GeoBrown
CIRM, january 2021

Motivation: Odded Schramm question

Conformally invariant scaling limits: an overview and a collection of problems

Oded Schramm

2.5. Lattice trees. We now present an example of a discrete model where we suspect that perhaps conformal invariance might hold. However, we do not presently have a candidate for the scaling limit.

Fix $n \in \mathbb{N}_+$, and consider the collection of all trees contained in the grid G that contain the origin and have n vertices. Select a tree T from this measure, uniformly at random.

Problem 2.8. What is the growth rate of the expected diameter of such a tree? If we rescale the tree so that the expected (or median) diameter is 1, is there a limit for the law of the tree as $n \to \infty$? What are its geometric and topological properties? Can the limit be determined?

It would be good to be able to produce some pictures. However, we presently do not know how to sample from this measure.

Problem 2.9. Produce an efficient algorithm which samples lattice trees approximately uniformly, or prove that such an algorithm does not exist.

Figure: Schramm ICM 2006.

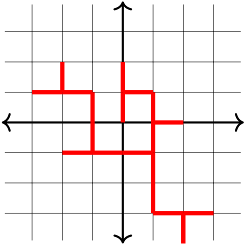
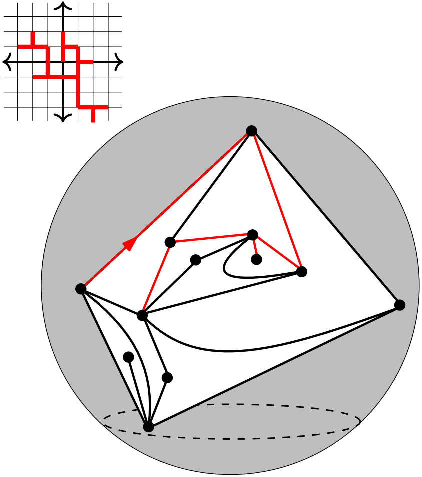
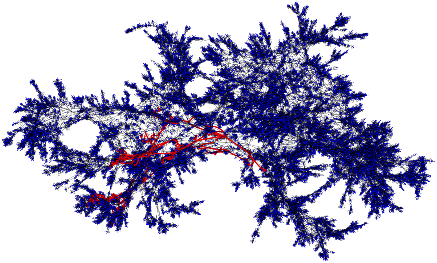


Figure: Subtree of size 20 containing the origin on $\mathbb{Z}^2$.



(a) tree-decorated quad. 10 faces, tree of size 6.



(b) Unif. tree-decorated quad. 90k faces and tree of size 500.

We try to contribute to Schramm's question in different ways:

- Trying to generalize known algorithms to a target size.
- Sampling (approx.) from the uniform measure in the set of subtrees of given size.
- Estimate scaling exponents.
- A new combinatorial proof of the Aldous-Broder theorem.

Chart of algorithms

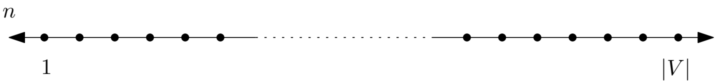
$\text{SubTree}(G, r, n) =$ set of subtrees of G containing r of size n.

$$\text{SubTree}(G, r) = \bigcup_{n=1}^{|V|} \text{SubTree}(G, r, n)$$

Chart of algorithms

$$\mathrm{SubTree}(G, r, n) = \text{set of subtrees of } G \text{ containing } r \text{ of size } n.$$

$$\mathrm{SubTree}(G, r) = \bigcup_{n=1}^{|V|} \mathrm{SubTree}(G, r, n)$$



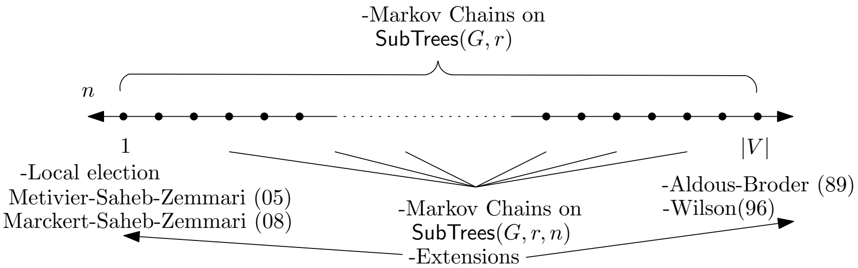
-Local election
Metivier-Saheb-Zemmari (05)
Marckert-Saheb-Zemmari (08)

-Aldous-Broder (89)
-Wilson(96)

Chart of algorithms

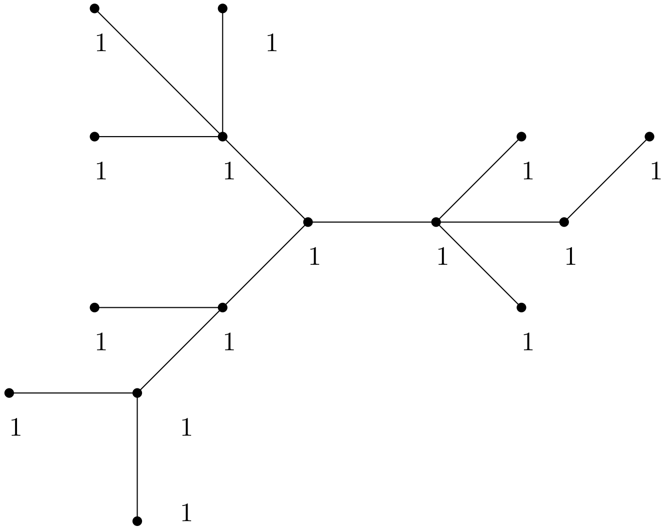
$\mathsf{SubTree}(G, r, n) =$ set of subtrees of G containing r of size n.

$$\mathsf{SubTree}(G, r) = \bigcup_{n=1}^{|V|} \mathsf{SubTree}(G, r, n)$$



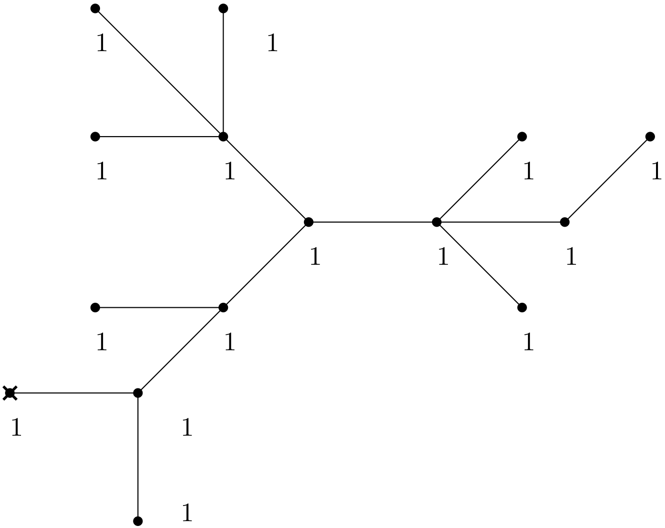
I. Local election to sample one vertex

If we cannot uniformly sample in $\mathsf{SubTree}(G, r, n)$ for G when it is a tree, we are hopeless!

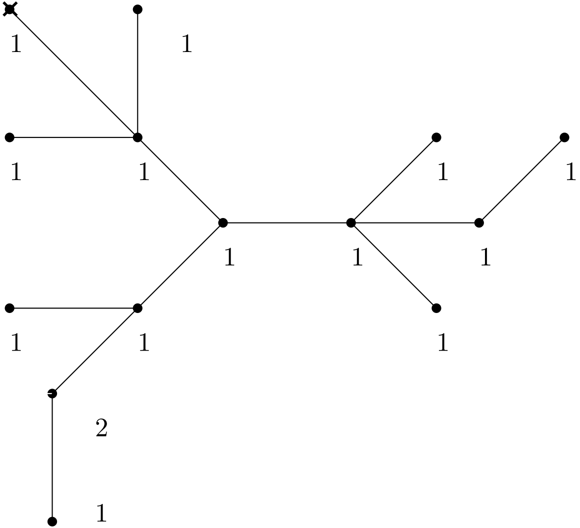


I. Local election to sample one vertex

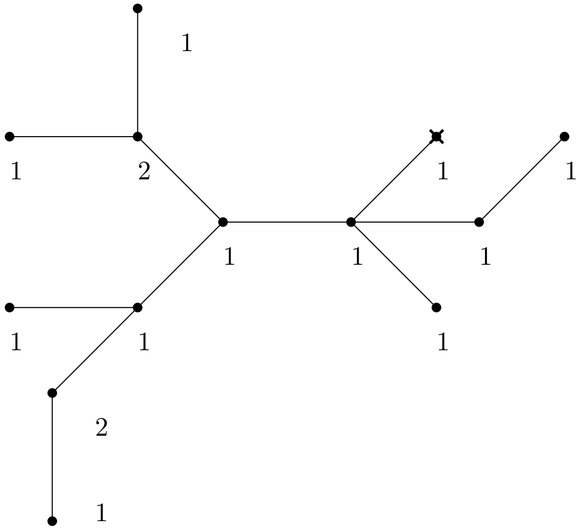
If we cannot uniformly sample in $\mathsf{SubTree}(G, r, n)$ for G when it is a tree, we are hopeless!



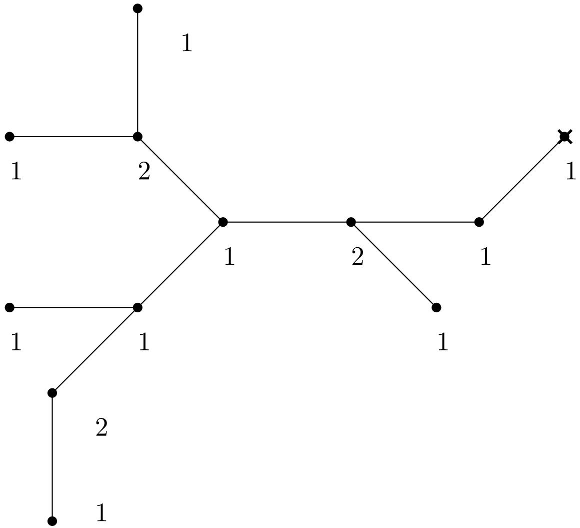
If we cannot uniformly sample in $\mathsf{SubTree}(G, r, n)$ for G when it is a tree, we are hopeless!



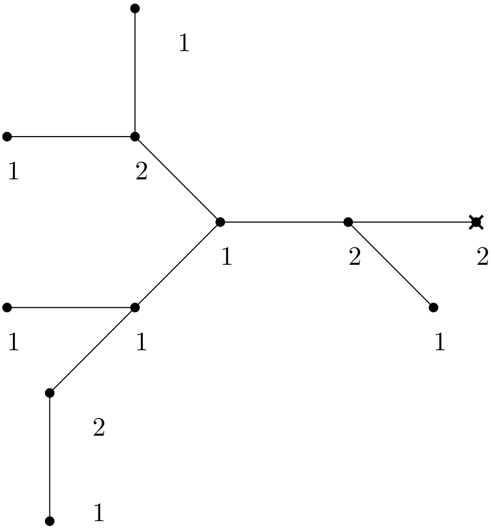
If we cannot uniformly sample in $\mathsf{SubTree}(G, r, n)$ for G when it is a tree, we are hopeless!



If we cannot uniformly sample in $\mathsf{SubTree}(G, r, n)$ for G when it is a tree, we are hopeless!

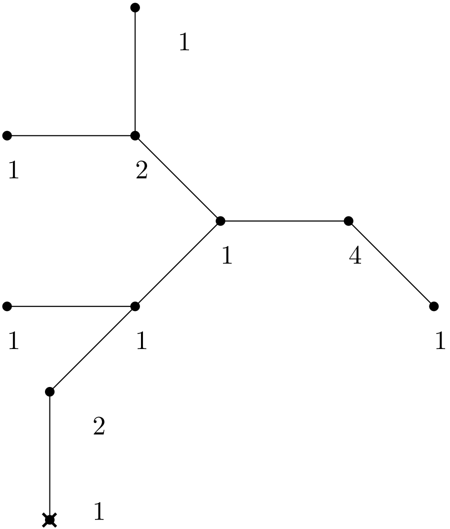


If we cannot uniformly sample in $\mathsf{SubTree}(G, r, n)$ for G when it is a tree, we are hopeless!

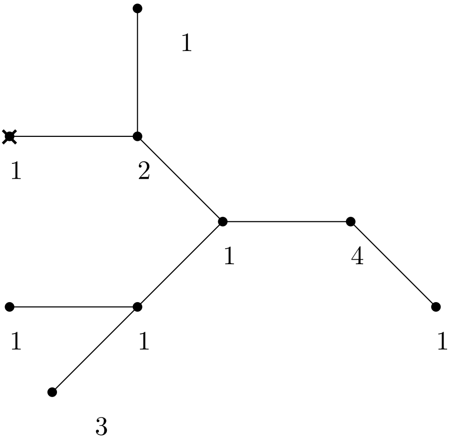


I. Local election to sample one vertex

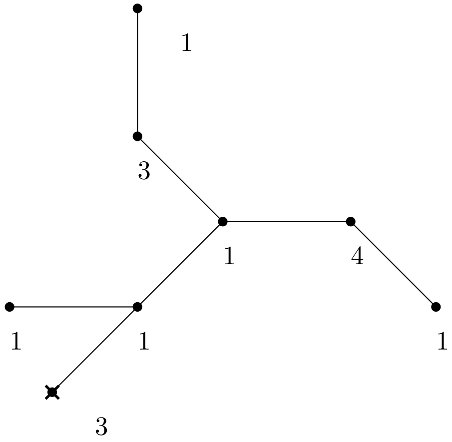
If we cannot uniformly sample in $\mathsf{SubTree}(G, r, n)$ for G when it is a tree, we are hopeless!



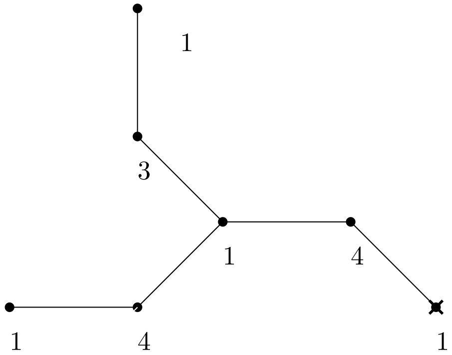
If we cannot uniformly sample in $\mathsf{SubTree}(G, r, n)$ for G when it is a tree, we are hopeless!



If we cannot uniformly sample in $\mathsf{SubTree}(G, r, n)$ for G when it is a tree, we are hopeless!

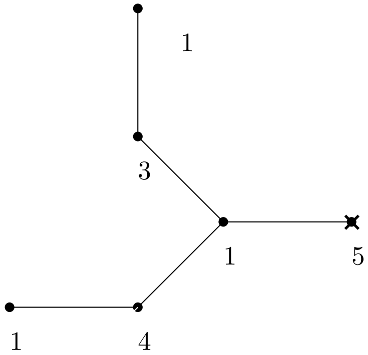


If we cannot uniformly sample in $\mathsf{SubTree}(G, r, n)$ for G when it is a tree, we are hopeless!



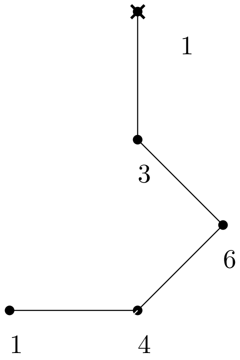
I. Local election to sample one vertex

If we cannot uniformly sample in SubTree(G, r, n) for G when it is a tree, we are hopeless!

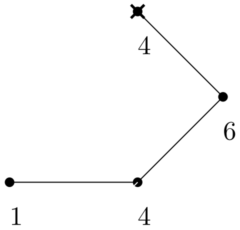


I. Local election to sample one vertex

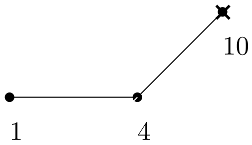
If we cannot uniformly sample in $\mathsf{SubTree}(G, r, n)$ for G when it is a tree, we are hopeless!



If we cannot uniformly sample in $\mathsf{SubTree}(G, r, n)$ for G when it is a tree, we are hopeless!

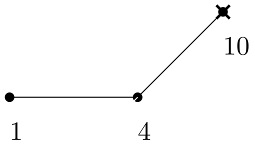


If we cannot uniformly sample in $\mathsf{SubTree}(G, r, n)$ for G when it is a tree, we are hopeless!



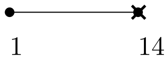
I. Local election to sample one vertex

If we cannot uniformly sample in SubTree(G, r, n) for G when it is a tree, we are hopeless!



I. Local election to sample one vertex

If we cannot uniformly sample in $\mathsf{SubTree}(G, r, n)$ for G when it is a tree, we are hopeless!



If we cannot uniformly sample in $\mathsf{SubTree}(G, r, n)$ for G when it is a tree, we are hopeless!

Theorem (Metivier-Saheb-Zemmari ('05) and Marckert-Saheb-Zemmari ('08))

The last vertex is uniform on V.

Theorem (Metivier-Saheb-Zemmari ('05) and Marckert-Saheb-Zemmari ('08))

The last vertex is uniform on V.

What is the distribution of the tree obtained by this method when n nodes remain (A.K.A. Evaporation(T, n))?

Theorem (Metivier-Saheb-Zemmari ('05) and Marckert-Saheb-Zemmari ('08))

The last vertex is uniform on V.

What is the distribution of the tree obtained by this method when n nodes remain (A.K.A. Evaporation(T, n))?

Proposition (F.- Marckert ('21+))

Let T be a tree on N vertices. Then

$$\mathbb{P}(\text{Evaporation}(T, n) = t) = \frac{(|L(t)| - 1)!(N - n)!}{(|L(t)| + N - n)!} \sum_{v \in L(t)} |\Delta_v|$$

where $L(t)$ is the set of leaves of t and Δ_v is the c.c. in $T - t$ attached to v.

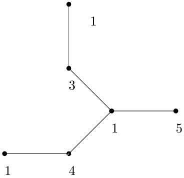


Figure: Tree of size 6 in the algorithm. Only the green one is considered in the probability.

Proof idea

Consider a collection $(X_s^j)_{j\in\mathbb{N}}$ of exponential r.v. of parameter s, then

$$
\begin{aligned}
m_n &:= \min\{X_1^j : j \in \{1,2,\dots,n\}\} \\
M_n &:= \max\{X_1^j : j \in \{1,2,\dots,n\}\} \\
&= m_n + M_n - m_n \\
&=^d m_n + M_{n-1} \\
&=^d X_n^1 + M_{n-1}
\end{aligned}
$$

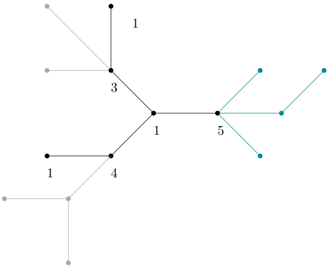


Figure: Tree of size 6 in the algorithm. Only the green one is considered in the probability.

Proof idea

Consider a collection $(X_s^j)_{j\in\mathbb{N}}$ of exponential r.v. of parameter s, then

$$
\begin{aligned}
m_n &:= \min\{X_1^j : j \in \{1,2,\dots,n\}\} \\
M_n &:= \max\{X_1^j : j \in \{1,2,\dots,n\}\} \\
&= m_n + M_n - m_n \\
&=^d m_n + M_{n-1} \\
&=^d X_n^1 + M_{n-1}
\end{aligned}
$$

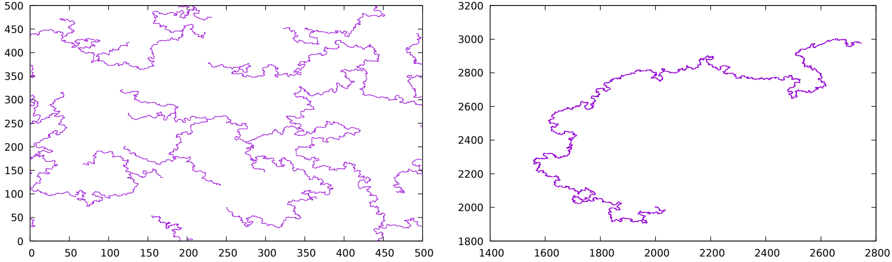


Figure: A tree Evaporation(T, 1000) on T a UST of resp. $(\mathbb{Z}/500\mathbb{Z})^2$ and $(\mathbb{Z}/4000\mathbb{Z})^2$

Fact: Reversible + symmetric Markov kernel $\implies$ Uniform measure is the unique invariant measure.

The fastest we obtained in practice: Starting from the tree $X_i = t \in \mathrm{SubTree}(G, r, n)$, the tree X_{i+1} is defined as follows:

1. Pick the oriented edge $\vec{e} = (u, u')$, where u is a uniform vertex, and conditional on u, u' is a uniform neighbor of u.
2. Add e to t:
    1. **The addition of $\vec{e}$ creates a new leaf:** Pick $\vec{e}' = (v, v')$ indep. of $\vec{e}$, following the same procedure to sample $\vec{e}$.
    If $t \cup \{e\} \setminus \{e'\}$ is a tree without the suppression of r, then $X_{i+1} = t \cup \{e\} \setminus \{e'\}$, else $X_{i+1} = t$.
    2. **The addition of $\vec{e}$ creates a cycle:** sample an edge e' according to $\mathrm{BreakCycle}(t \cup \{e\}, e)(\cdot)$ and define $X_{i+1} = t \cup \{e\} \setminus \{e'\}$.
    3. **Otherwise:** $X_{n+1} = t$

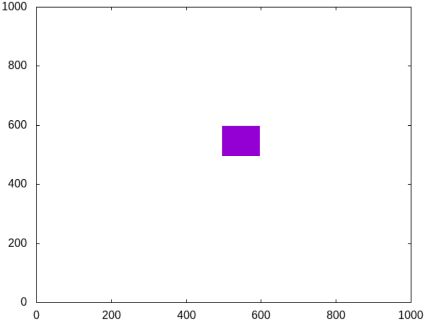
Fact: Reversible + symmetric Markov kernel $\implies$ Uniform measure is the unique invariant measure.

The fastest we obtained in practice: Starting from the tree $X_i = t \in \text{SubTree}(G, r, n)$, the tree X_{i+1} is defined as follows:

1. Pick the oriented edge $\vec{e} = (u, u')$, where u is a uniform vertex, and conditional on u, u' is a uniform neighbor of u.
2. Add e to t:
    1. **The addition of $\vec{e}$ creates a new leaf:** Pick $\vec{e}' = (v, v')$ indep. of $\vec{e}$, following the same procedure to sample $\vec{e}$.
    If $t \cup \{e\} \setminus \{e'\}$ is a tree without the suppression of r, then $X_{i+1} = t \cup \{e\} \setminus \{e'\}$, else $X_{i+1} = t$.
    2. **The addition of $\vec{e}$ creates a cycle:** sample an edge e' according to $\text{BreakCycle}(t \cup \{e\}, e)(\cdot)$ and define $X_{i+1} = t \cup \{e\} \setminus \{e'\}$.
    3. **Otherwise:** $X_{n+1} = t$

We impose (for reversibility purposes) for all graph g with excess 1 and for each pair of edges in the unique cycle that

$$\text{BreakCycle}(g, e)(e') = \text{BreakCycle}(g, e')(e)$$



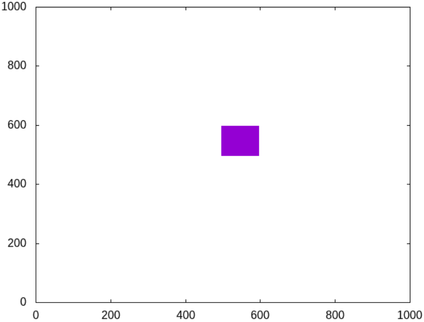


Figure: 1M and 100M iteration by frame.

$T_n =$ Uniform element in $\mathrm{SubTree}((\mathbb{Z}/n\mathbb{Z})^2, n)$
$\mathrm{W}(t)(\mathrm{H}(t)) =$ cols (lines) of $(\mathbb{Z}/n\mathbb{Z})^2$ containing at least one vertex of t.
$\mathrm{q}_i(T_n) =$ proportion of vertices of degree i in T_n.

Conjecture

1. *There exists* $\alpha \in [0.63, 0.67]$ *s.t.*

$$n^{-\alpha}(\mathrm{W}(T_n), \mathrm{H}(T_n)) \xrightarrow[n\to\infty]{(d)} (\boldsymbol{W}, \boldsymbol{H}) \quad \textit{non trivial r.v.}$$

2. *There exists* $\beta \in [3/4 - 0.01, 3/4 + 0.01]$ *s.t.*

$$n^{-\beta}\mathrm{d}_{\mathrm{T}_\mathbf{n}}(\mathrm{u_n}, \mathrm{v_n}) \xrightarrow[\mathrm{n}\to\infty]{(\mathrm{d})} \boldsymbol{D} \quad \textit{real r.v. a.s. non zero,}$$

where u_n *and* v_n *are independent uniformly chosen vertices of* T_n.

3. *There exists a constant vector satisfying* $\boldsymbol{q}_1 \in [0.2585 \pm 0.001]$, $\boldsymbol{q}_2 \in [0.506 \pm 0.001]$, $\boldsymbol{q}_3 \in [0.214 \pm 0.001]$, $\boldsymbol{q}_4 \in [0.02185 \pm 0.001]$

$$(\mathrm{q}_1(T_n), \mathrm{q}_2(T_n), \mathrm{q}_3(T_n), \mathrm{q}_4(T_n)) \xrightarrow[n\to\infty]{proba} (\mathrm{q}_1, \mathrm{q}_2, \mathrm{q}_3, \mathrm{q}_4),$$

T_n = Uniform element in SubTree$((\mathbb{Z}/n\mathbb{Z})^2, n)$
$\mathrm{W}(t)(\mathrm{H}(t))$ = cols (lines) of $(\mathbb{Z}/n\mathbb{Z})^2$ containing at least one vertex of t.
$\mathrm{q}_i(T_n)$ = proportion of vertices of degree i in T_n.

tree size	1000	2500	5000	8100
number of simulations	5039	5486	6111	5232
Initial rectangle tree shape	40×25	50×50	50×100	90×90
Nb Steps of the chain	$150M$	$1G$	$25G$	$200G$

To estimate the exponents we use

$$\alpha \sim \log(\mathrm{Mean}(W(T_n))/\mathrm{Mean}(W(T_m)))/\log(n/m)$$

(n,m)	$(1000, 2500)$	$(2500, 5000)$	$(5000, 8100)$
Estimation of α (median)	0.641	0.654	0.644
Estimation of α (mean)	0.642	0.657	0.638
Estimation of β (median)	0.756	0.735	0.751
Estimation of β (mean)	0.746	0.746	0.754

T_n = Uniform element in $\mathrm{SubTree}((\mathbb{Z}/n\mathbb{Z})^2, n)$
$\mathrm{W}(t)(\mathrm{H}(t))$ = cols (lines) of $(\mathbb{Z}/n\mathbb{Z})^2$ containing at least one vertex of t.
$\mathrm{q}_i(T_n)$ = proportion of vertices of degree i in T_n.

Degree proportion	T_n	Spanning Tree
q_1	≈ 0.2585	$\frac{8}{\pi^2}\left(1-\frac{2}{\pi}\right) \approx 0.294$
q_2	≈ 0.506	$\frac{4}{\pi}\left(2-\frac{9}{\pi}+\frac{12}{\pi^2}\right) \approx 0.447$
q_3	≈ 0.214	$2\left(1-\frac{2}{\pi}\right)\left(2-\frac{6}{\pi}+\frac{12}{\pi^2}\right) \approx 0.222$
q_4	≈ 0.02185	$\left(\frac{4}{\pi}-1\right)\left(1-\frac{2}{\pi}\right) \approx 0.036$

Both **Wilson's algorithm** and **Aldous-Broder algorithm** sample from the uniform distribution when we consider simple random walks, but they are more general.

Both **Wilson's algorithm** and **Aldous-Broder algorithm** sample from the uniform distribution when we consider simple random walks, but they are more general.

Consider a Markov kernel M with unique invariant distribution ρ.

Both **Wilson's algorithm** and **Aldous-Broder algorithm** sample from the uniform distribution when we consider simple random walks, but they are more general.

Consider a Markov kernel M with unique invariant distribution ρ.

Sometimes we consider the edges of (t, r) oriented towards the root, we write $\vec{e}$.

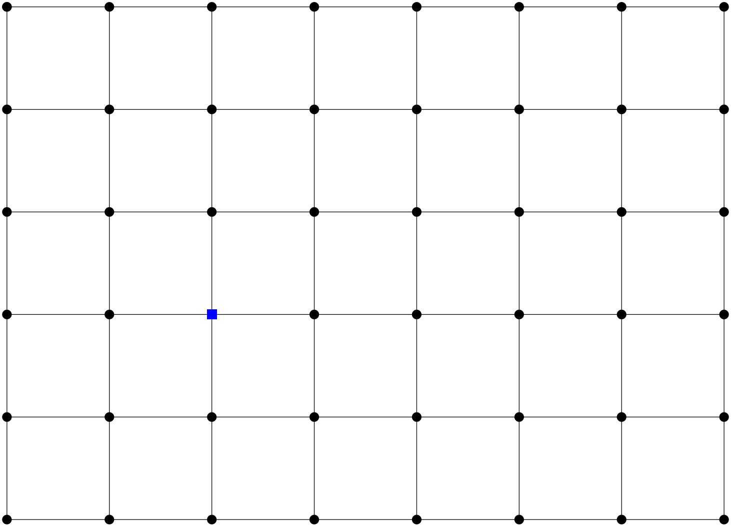


Figure: Pick any vertex as root (square vertex)

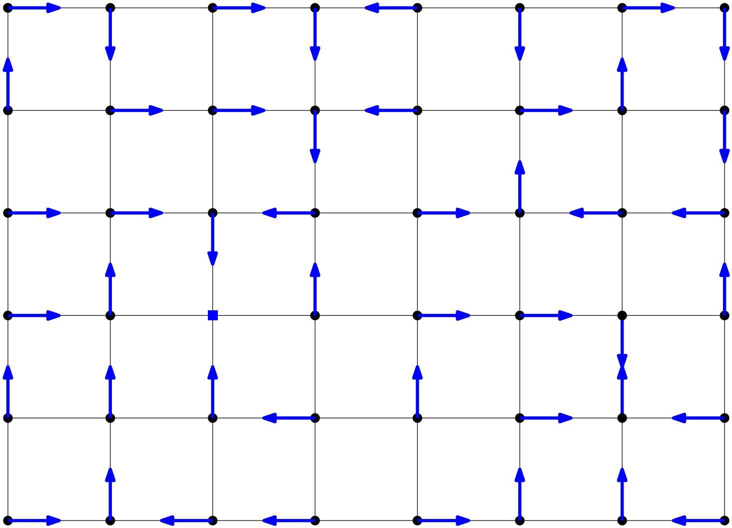


Figure: Pick one outgoing edge for each $v \in V \setminus \{r\}$ following the markov kernel M.

III.1. Wilson (Cycle popping version)

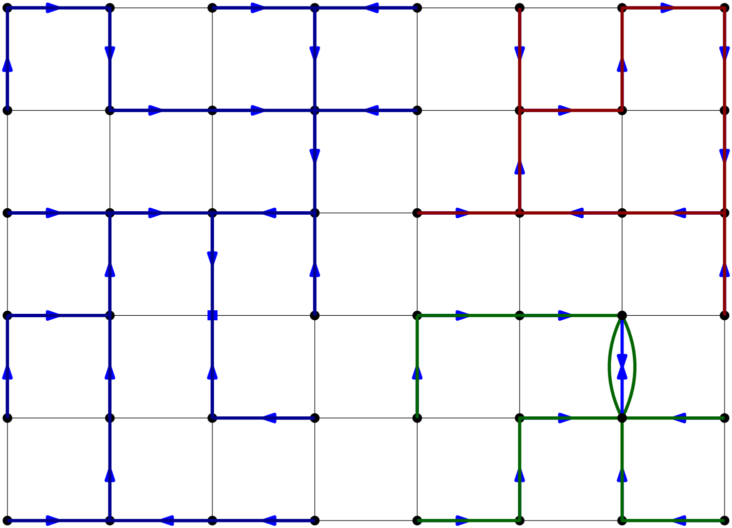


Figure: The oriented edges induce a graph.

III.1. Wilson (Cycle popping version)

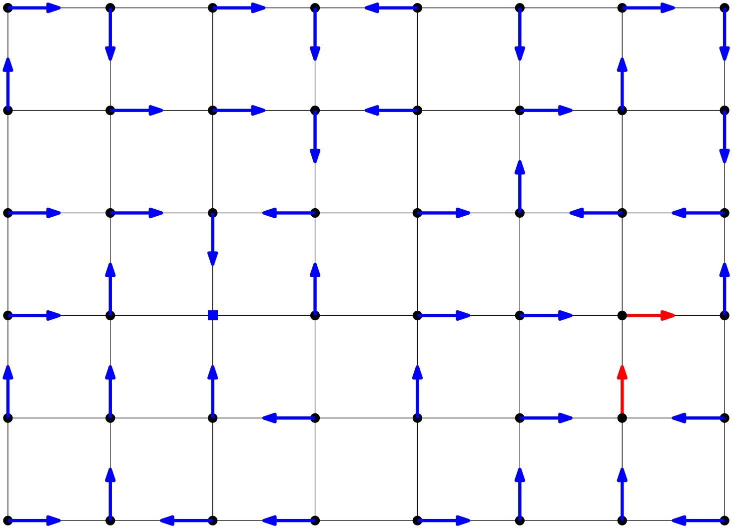


Figure: If there is a cycle pick one and re-sample the outgoing edges of the vertices on it.

III.1. Wilson (Cycle popping version)

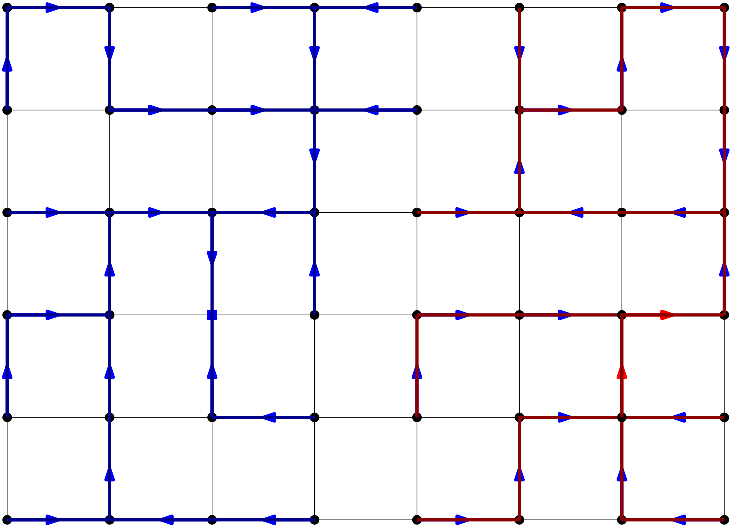


Figure: Induced graph.

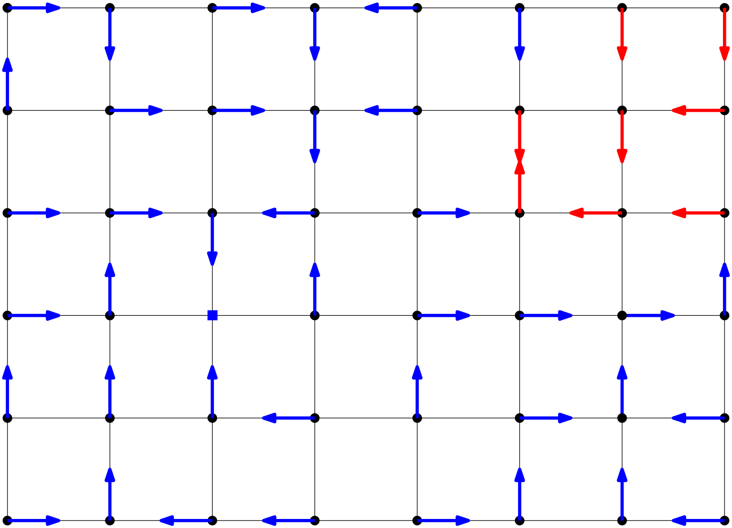
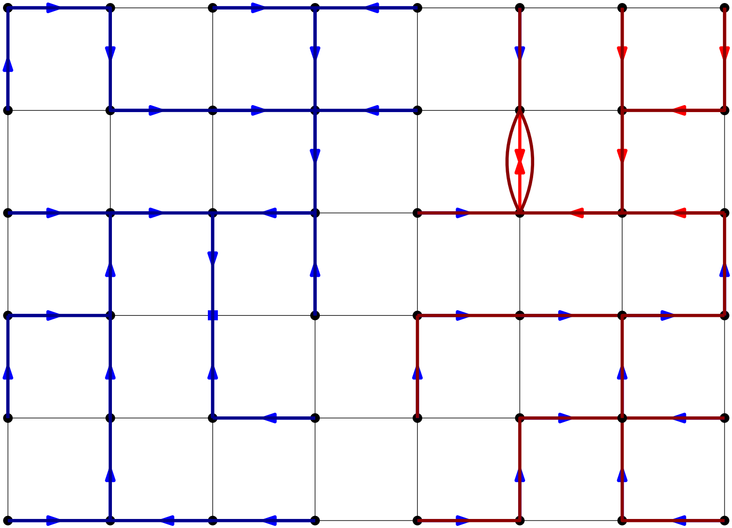
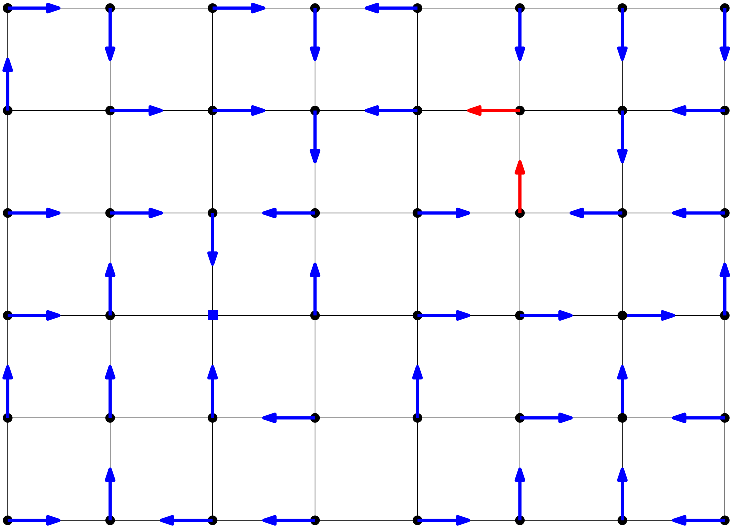


Figure: Pick a cycle and resample again.

III.1. Wilson (Cycle popping version)





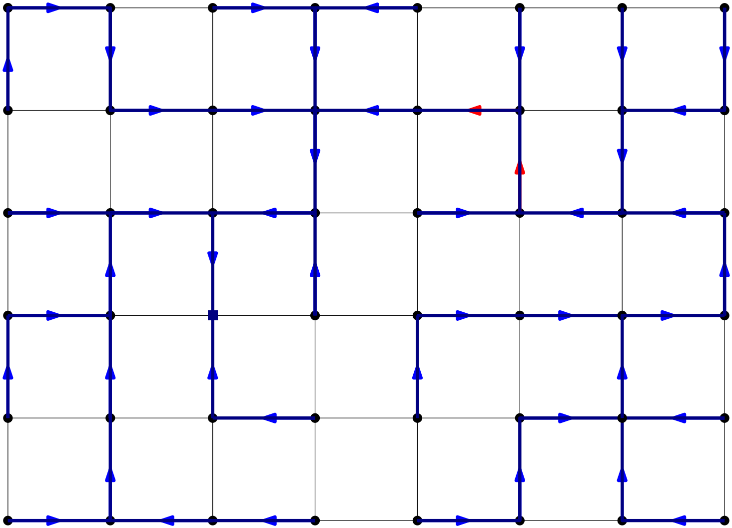


Figure: Stop when there is no more cycle, i.e. a tree.

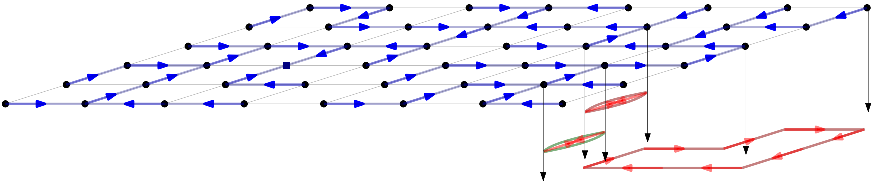


Figure: Heap of cycles $\times$ Tree

Call $(\mathcal{H}, \mathcal{T})$ the r.v. associated to the heap of cycles and rooted tree of the cycle popping.

Theorem (Wilson ('96))

For any finite graph the cycle popping ends with probability 1. Moreover, for any heap of cycles H and any tree $T \in \mathrm{SubTree}(G, r, |V|)$ one has

$$\mathbb{P}((\mathcal{H}, \mathcal{T}) = (H, T)) = \mathbb{P}(\mathcal{H} = H)\mathbb{P}(\mathcal{T} = T) = P(H)P(T),$$

where for any multiset of oriented edges $P(S) = \prod_{\vec{e} \in S} M_{\vec{e}}$.

Fix a root $r \in V$ and associate to each vertex in $V \setminus \{r\}$ a random uniform outgoing edge. Call τ the connected component of the root.

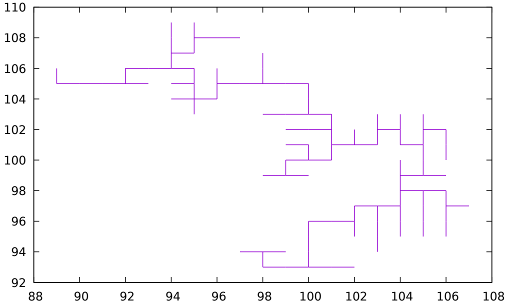


Figure: Simulation of τ on $(\mathbb{Z}/100\mathbb{Z})^2$, 3536949 simulations were needed to get a tree of size at least 100.

Fix a root $r \in V$ and associate to each vertex in $V \setminus \{r\}$ a random uniform outgoing edge. Call τ the connected component of the root.

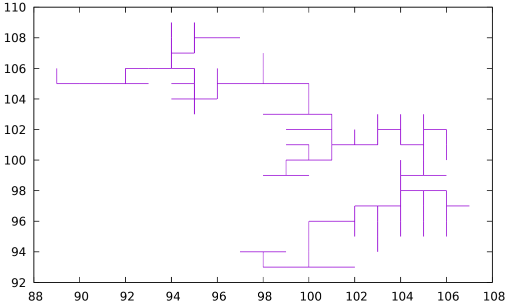


Figure: Simulation of τ on $(\mathbb{Z}/100\mathbb{Z})^2$, 3536949 simulations were needed to get a tree of size at least 100.

Problem!

The distribution of $(\tau \big| |\tau| = n)$ does not have full support in general. The connected components different from τ have one cycle, then they cannot have size 1.

Combinatorial prelude

- **Inversion lemma** Wilson's algorithm construct all possible heap of cycles which do not contain r, summing over this set

$$\sum_H P(H) = \frac{1}{\sum_{H \text{ trivial}} P(H)} = \frac{1}{\det(I - M^{(r)})}$$

- **Matrix tree theorem**

$$\det(I - M^{(r)}) = \sum_{T \in \mathrm{SubTree}(G, r, |V|)} P(T)$$

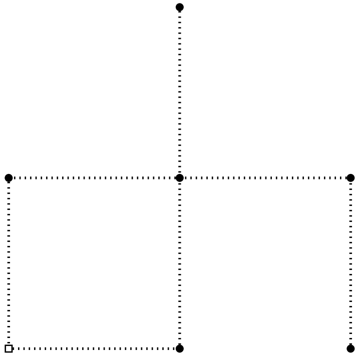
To keep in mind

The output tree $\mathcal{T}$ satisfies

$$\mathbb{P}(\mathcal{T} = T) = \frac{\prod_{\vec{e} \in t} M_{\vec{e}}}{\det(I - M^{(r)})}$$

where each edge $\vec{e}$ in T is oriented towards the root r

Consider an M-walk W in the invariant regime started at $r \in V$ up to the cover time.
Denote by **FirstEntrance**$(W) = (t, r)$, where r is the starting point of W and t is the spanning tree formed by the first edge used to visit each vertex.



III.2. Extension to the non-reversible case

Theorem (Aldous-Broder ('89))

*For M positive and **reversible** Markov kernel with invariant distribution ρ. For any $T \in \mathrm{SubTree}(G, r, |V|)$ one has*

$$\mathbb{P}\left(\boldsymbol{FirstEntrance}(W) = (T, r)\right) = \frac{\prod_{\vec{e} \in t} M_{\vec{e}}}{\sum_{w \in V} \det(I - M^{(w)})},$$

III.2. Extension to the non-reversible case

Theorem (Aldous-Broder ('89))

*For M positive and **reversible** Markov kernel with invariant distribution ρ. For any $T \in \mathrm{SubTree}(G, r, |V|)$ one has*

$$\mathbb{P}\left(\boldsymbol{FirstEntrance}(W) = (T, r)\right) = \frac{\prod_{\vec{e} \in t} M_{\vec{e}}}{\sum_{w \in V} \det(I - M^{(w)})},$$

Define for a Markov kernel M with unique invariant measure ρ, the Markov kernel $\overleftarrow{M}$ as

$$\overleftarrow{M}_{x,y} = \rho_y / \rho_x M_{y,x}$$

III.2. Extension to the non-reversible case

Theorem (Aldous-Broder ('89))

*For M positive and **reversible** Markov kernel with invariant distribution ρ. For any $T \in \mathrm{SubTree}(G, r, |V|)$ one has*

$$\mathbb{P}\left(\boldsymbol{FirstEntrance}(W) = (T, r)\right) = \frac{\prod_{\vec{e} \in t} M_{\vec{e}}}{\sum_{w \in V} \det(I - M^{(w)})},$$

Define for a Markov kernel M with unique invariant measure ρ, the Markov kernel $\overleftarrow{M}$ as

$$\overleftarrow{M}_{x,y} = \rho_y / \rho_x M_{y,x}$$

Theorem (F.- Marckert ('21+))

For M positive with invariant distribution ρ. For any $T \in \mathrm{SubTree}(G, r, |V|)$ one has

$$\mathbb{P}\left(\boldsymbol{FirstEntrance}(W) = (T, r)\right) = \frac{\prod_{\vec{e} \in t} \overleftarrow{M}_{\vec{e}}}{\sum_{w \in V} \det(I - \overleftarrow{M}^{(w)})},$$

Both normalization constants are the same. In particular

$$\det(I - M^{(v)}) = \det(I - \overleftarrow{M}^{(v)}),$$

since for $\vec{C}$ oriented cycle $\prod_{\vec{e}\in\vec{C}} M_{\vec{e}} = \prod_{\vec{e}\in\overleftarrow{C}} \overleftarrow{M}_{\vec{e}}$.

Numerators are different when ρ is not reversible with respect to M.

The edges are directed from each node u toward its direct ancestor $a(u)$. For a tree $t \in \mathrm{SubTree}(G, r)$,

$$\prod_{\vec{e}\in t} M_{\vec{e}} = \prod_{u\in t\neq\{r\}} M_{u,a(u)} = \textbf{Const.}\ \rho_r \prod_{u\in t\neq\{r\}} \rho_u M_{u,a(u)}$$

$$\prod_{\vec{e}\in t} \overleftarrow{M}_{\vec{e}} = \prod_{u\in t\neq\{r\}} \left[M_{a(u),u}\, \rho_{a(u)}/\rho_u\right] = \textbf{Const.}\ \rho_r \prod_{u\in t\neq\{r\}} \rho_{a(u)} M_{a(u),u}.$$

The Aldous-Broder proof is purely probabilistic!

Consider the following chain $X_i = (t, r)$. To define X_{i+1} do as follows

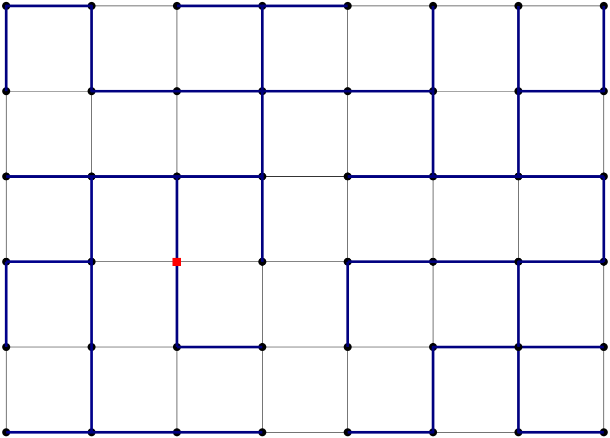


Figure: $X_i = (t, r)$

The Aldous-Broder proof is purely probabilistic!

Consider the following chain $X_i = (t, r)$. To define X_{i+1} do as follows

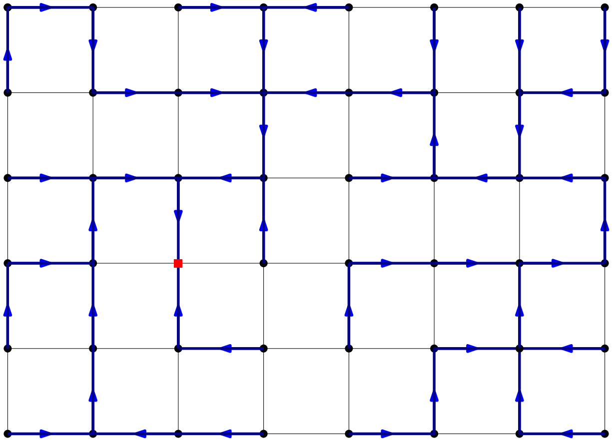


Figure: Orient the edges towards r

The Aldous-Broder proof is purely probabilistic!

Consider the following chain $X_i = (t, r)$. To define X_{i+1} do as follows

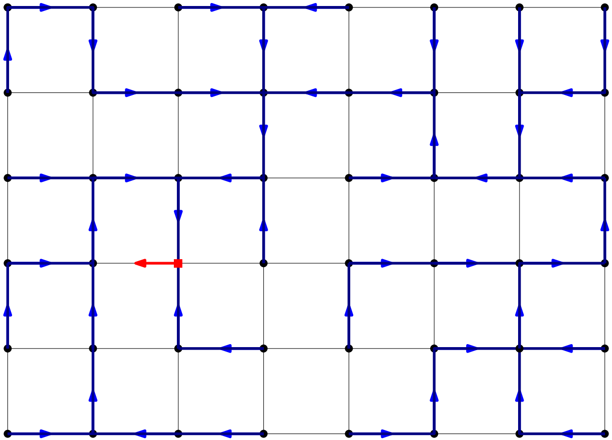


Figure: Make a step from the root following the kernel M.

The Aldous-Broder proof is purely probabilistic!

Consider the following chain $X_i = (t, r)$. To define X_{i+1} do as follows

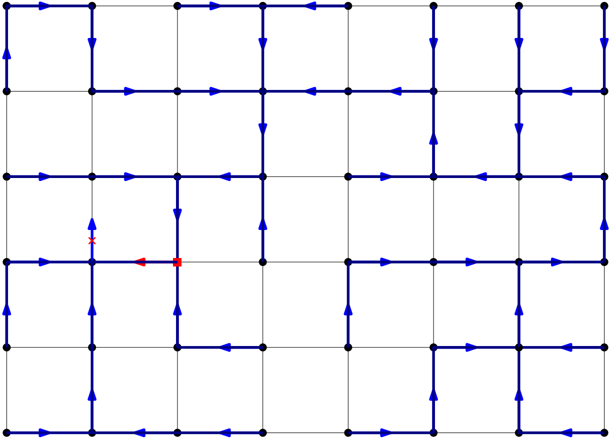


Figure: Suppress the outgoing edge in the destination point

Consider the following chain $X_i = (t, r)$. To define X_{i+1} do as follows

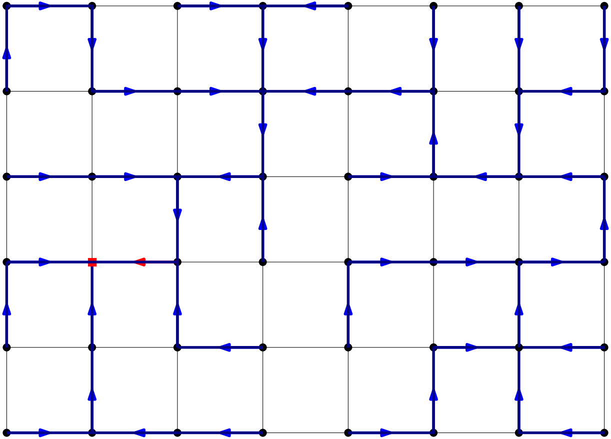


Figure: Change the root to the destination point

Consider the following chain $X_i = (t, r)$. To define X_{i+1} do as follows

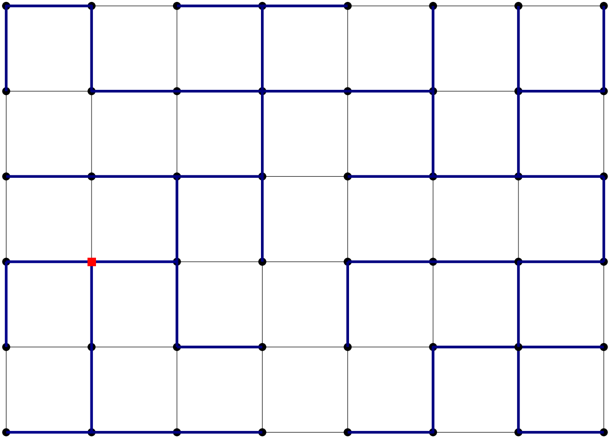


Figure: Define this resulting rooted tree as X_{i+1}

The Aldous-Broder proof is purely probabilistic!

Two facts:

- For w a deterministic walk up to the cover time one has

$$\textbf{FirstEntrance}(w) = \textbf{LastExit}(\overleftarrow{w})$$

- Markov chain tree theorem

$$\rho_v = \frac{\sum_{t \in \mathsf{SubTree}(G,v,|V|)} \prod_{\vec{e} \in t} \overleftarrow{M}_{\vec{e}}}{Z} = \frac{\det(I - M^{(v)})}{Z}$$

The Aldous-Broder proof is purely probabilistic!

Two facts:

- For w a deterministic walk up to the cover time one has

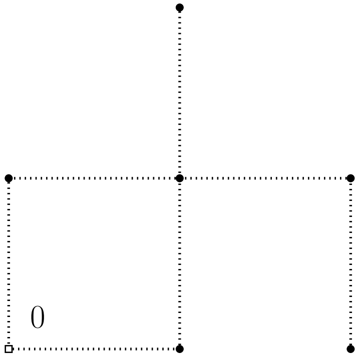
$$\textbf{FirstEntrance}(w) = \textbf{LastExit}(\overleftarrow{w})$$

- Markov chain tree theorem

$$\rho_v = \frac{\sum_{t \in \mathsf{SubTree}(G,v,|V|)} \prod_{\vec{e} \in t} \overleftarrow{M}_{\vec{e}}}{Z} = \frac{\det(I - M^{(v)})}{Z}$$

The proof uses a coupling from the past argument + both precedent facts.

III.2. Labeled extension



Denote by **Pionner**$(W) = ($**FirstEntrance**$(W), L)$ where L is the labeling.
$H_D(a, b) =$ probability starting from a that a walk following M escapes D at b.
$\overleftarrow{H}_D(a, b) =$ probability starting from a that a walk following $\overleftarrow{M}$ escapes D at b.

$$\begin{aligned}
&\mathbb{P}(\textbf{Pionner}(W) = ((t, r), \ell)) \\
&= \mathbb{1}_{\ell_{\mathbf{0}}=r}\rho_{\ell_{\mathbf{0}}} \prod_{i=0}^{n-2} \left[H_{\{\ell_{\leq i}\}}(\ell_i, a(\ell_{i+1}))M_{a(\ell_{i+\mathbf{1}}),\ell_{i+\mathbf{1}}}\right] \\
&= \left(\mathbb{1}_{\ell_{\mathbf{0}}=r}\rho_{\ell_{n-\mathbf{1}}} \prod_{i=0}^{n-2} \left[\overleftarrow{H}_{\{\ell_{\leq i}\}}(a(\ell_{i+1}), \ell_i)\right] Z\right) \frac{\prod_{\vec{e}\in t} \overleftarrow{M}_{\vec{e}}}{Z}
\end{aligned}$$

Can we prove using combinatorics that

$$\sum_{\ell} \mathbb{1}_{\ell_{\mathbf{0}}=r}\, \rho_{\ell_{n-\mathbf{1}}} \prod_{i=0}^{n-2} \left[\overleftarrow{H}_{\{\ell_{\leq i}\}}(a(\ell_{i+1}), \ell_i) \right] Z = 1?$$

(the sum ranges over all decreasing labelings of the tree)

III.2. Combinatorial proof

Can we prove using combinatorics that

$$\sum_{\ell} \mathbb{1}_{\ell_0=r}\rho_{\ell_{n-1}} \prod_{i=0}^{n-2} \left[\overleftarrow{H}_{\{\ell_{\leq i}\}}(a(\ell_{i+1}),\ell_i)\right] Z = 1?$$

(the sum ranges over all decreasing labelings of the tree)
The Markov chain tree Theorem gives that $\rho_v = \det(I - M^{(v)})/Z$, so equivalently

$$\sum_{\ell} \mathbb{1}_{\ell_0=r} \prod_{i=0}^{n-2} \left[\overleftarrow{H}_{\{\ell_{\leq i}\}}(a(\ell_{i+1}),\ell_i)\right] \det(I - M^{(\ell_{n-1})}) = 1?$$

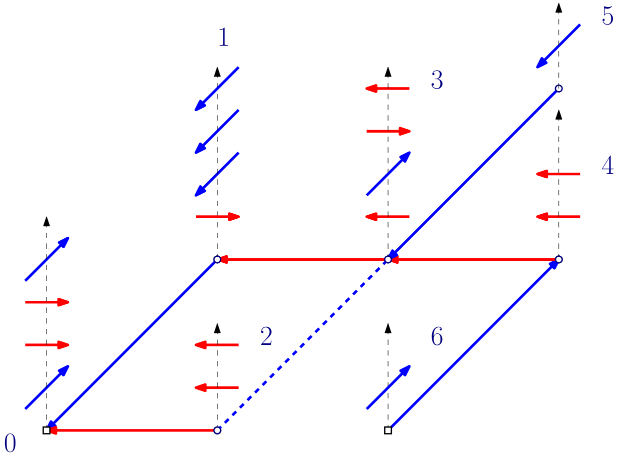


Figure: Path seen backward as a heap of outgoing edges

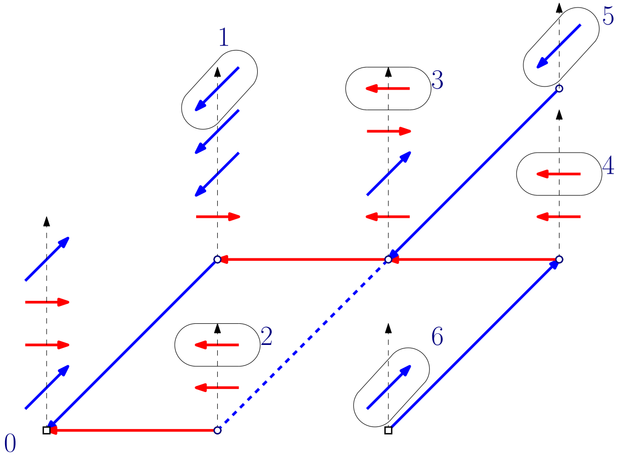


Figure: The tree edges are always on top of the piles.

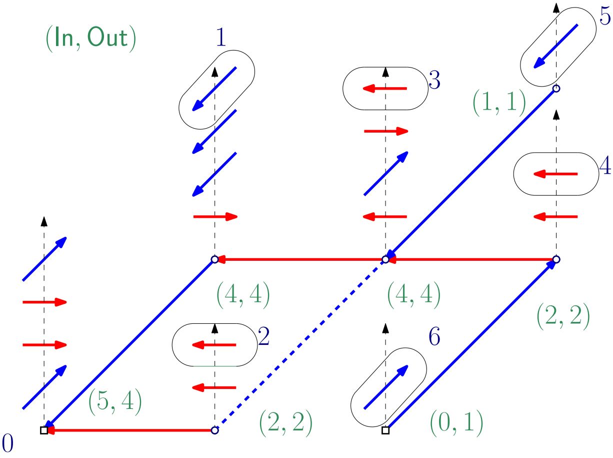


Figure: Count the incoming and outgoing edges

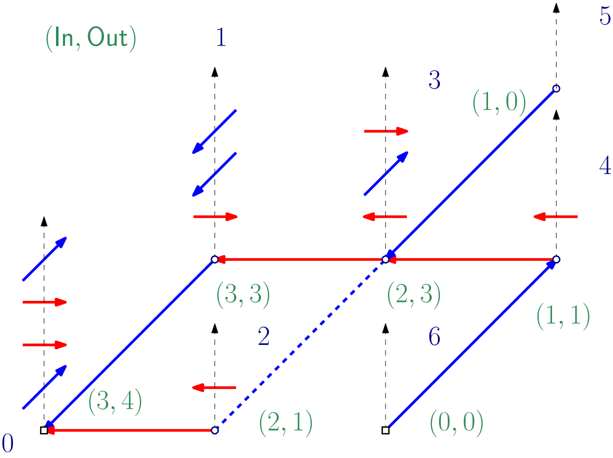


Figure: Pop-out the tree edges to construct H^{-t} (update (In,Out))

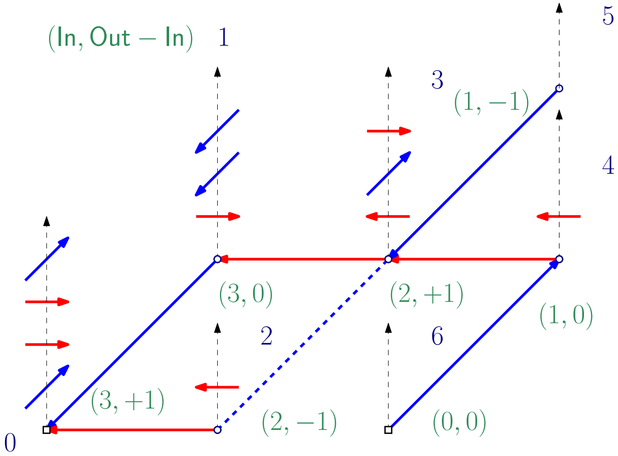


Figure: Convenient to keep an eye on (In,Out-In)

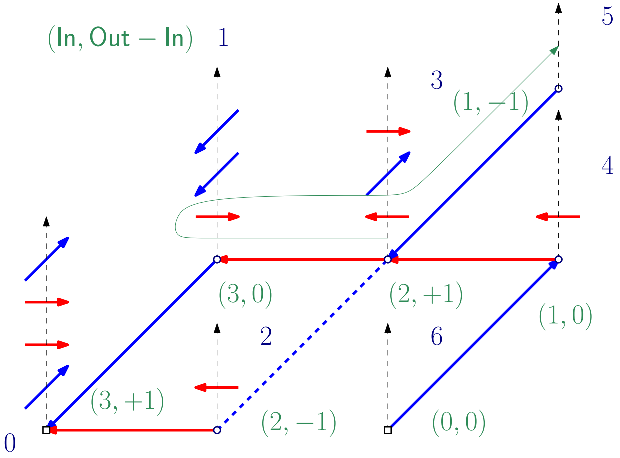


Figure: Play golf!

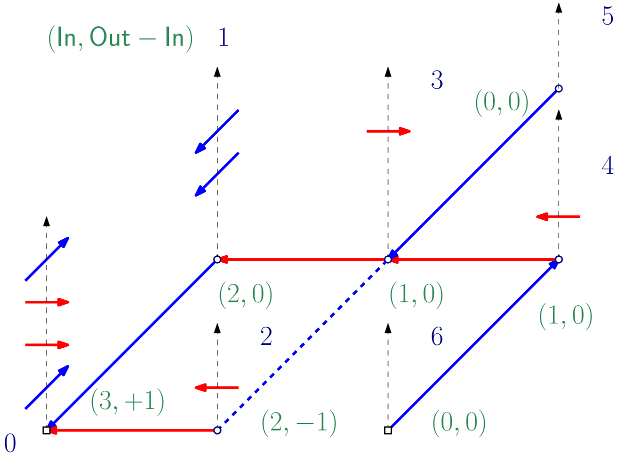


Figure: Supress the path and update (In,Out-In)

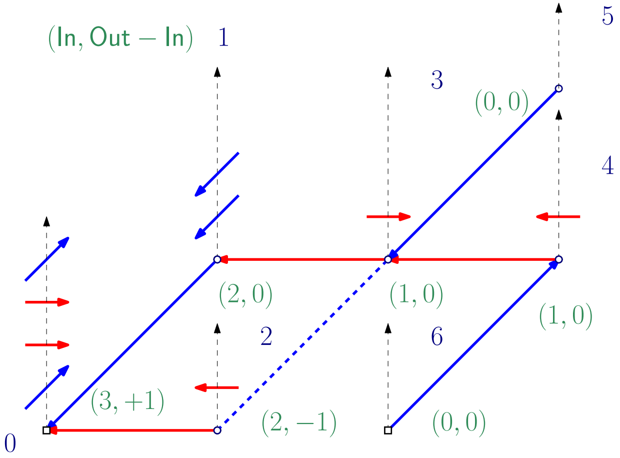


Figure: Let the pieces fall

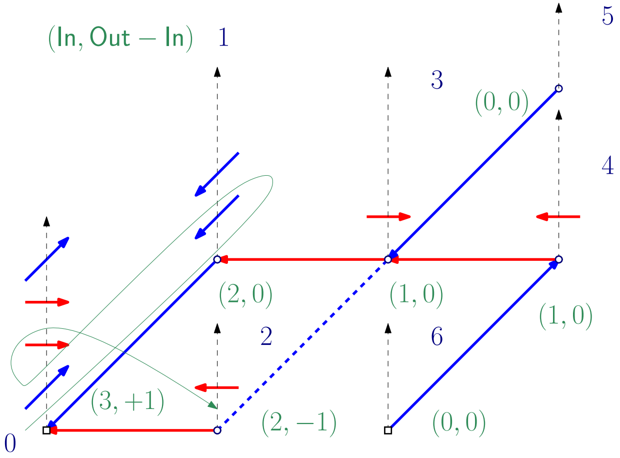


Figure: Continue playing golf with next emitting vertex.

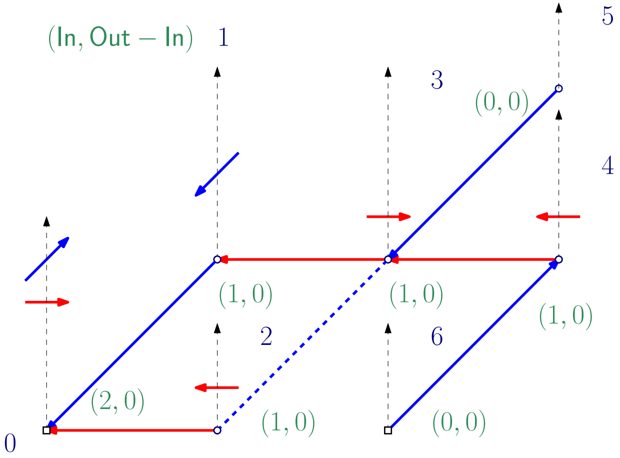


Figure: Supress the path and update (In,Out-In)

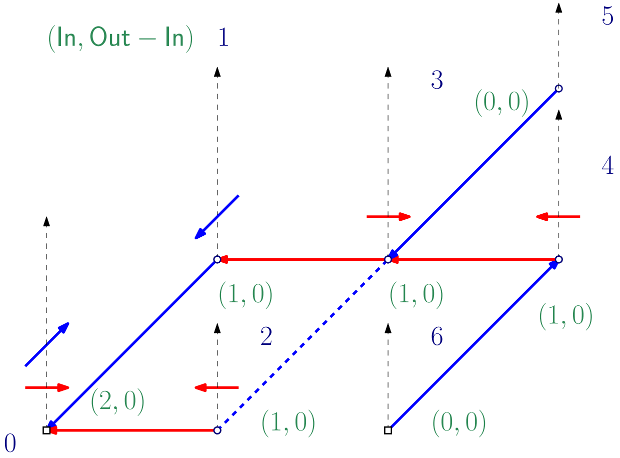


Figure: Let the pieces fall

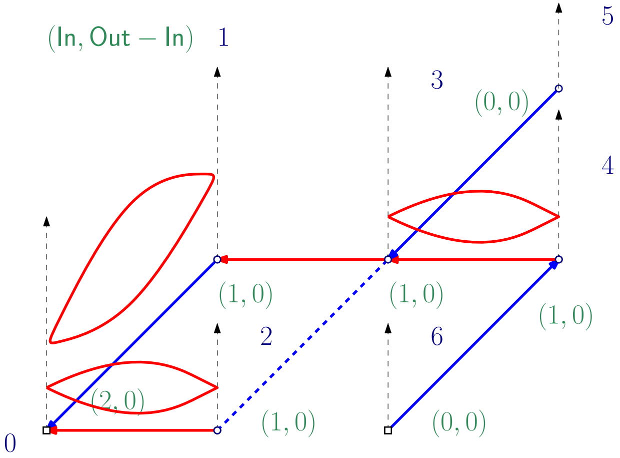


Figure: heap of cycles

The heap of outgoing edges H^{-t} is a heap only on $V \setminus \ell_{n+1}$ and $H^{-t} = \text{Golf} \times HC$. Recall we fix a treatment order to fix an ordering of the starting points in the Golf game.

$$\sum_{\ell} \mathbb{1}_{\ell_0 = r} \prod_{i=0}^{n-2} \left[\overleftarrow{H}_{\ell_{\leq i}}(a(\ell_{i+1}), \ell_i) \right] \det(I - M^{(\ell_{n-1})})$$

The heap of outgoing edges H^{-t} is a heap only on $V \setminus \ell_{n+1}$ and $H^{-t} = \text{Golf} \times HC$. Recall we fix a treatment order to fix an ordering of the starting points in the Golf game.

$$
\begin{aligned}
&\sum_{\ell} \mathbb{1}_{\ell_0 = r} \prod_{i=0}^{n-2} \left[\overleftarrow{H}_{\ell_{\leq i}}(a(\ell_{i+1}), \ell_i) \right] \det(I - M^{(\ell_{n-1})}) \\
&= \sum_{H^{-t} \text{ valid}} W(H^{-t}) \det(I - M^{(\ell_{n-1})})
\end{aligned}
$$

The heap of outgoing edges H^{-t} is a heap only on $V \setminus \ell_{n+1}$ and $H^{-t} = \text{Golf} \times HC$. Recall we fix a treatment order to fix an ordering of the starting points in the Golf game.

$$\begin{aligned}
&\sum_{\ell} \mathbb{1}_{\ell_0=r} \prod_{i=0}^{n-2} \left[\overleftarrow{H}_{\ell_{\leq i}}(a(\ell_{i+1}), \ell_i) \right] \det(I - M^{(\ell_{n-1})}) \\
&= \sum_{H^{-t} \text{ valid}} W(H^{-t}) \det(I - M^{(\ell_{n-1})}) \\
&= \sum_{(Golf, HC) \text{ valid}} W(Golf) \times W(HC) \det(I - M^{(\ell_{n-1})})
\end{aligned}$$

The heap of outgoing edges H^{-t} is a heap only on $V \setminus \ell_{n+1}$ and $H^{-t} = \text{Golf} \times HC$. Recall we fix a treatment order to fix an ordering of the starting points in the Golf game.

$$
\begin{aligned}
&\sum_{\ell} \mathbb{1}_{\ell_0 = r} \prod_{i=0}^{n-2} \left[\overleftarrow{H}_{\ell_{\leq i}}(a(\ell_{i+1}), \ell_i) \right] \det(I - M^{(\ell_{n-1})}) \\
&= \sum_{H^{-t} \text{ valid}} W(H^{-t}) \det(I - M^{(\ell_{n-1})}) \\
&= \sum_{(Golf, HC) \text{ valid}} W(Golf) \times W(HC) \det(I - M^{(\ell_{n-1})}) \\
&= \underbrace{\sum_{Golf \text{ valid}} W(Golf)}_{=1} \times \underbrace{\left(\sum_{\substack{HC \textbf{ heap of cycles} \\ \text{not containing } \ell_{n-1}}} W(HC) \right) \det(I - M^{(\ell_{n-1})})}_{=1}
\end{aligned}
$$

The first by a probabilistic algorithm.

Corollary (F.-Marckert ('21+))

*If W is a SRW stopped when $m < |V|$ vertices has been discovered, then the tree **FirstEntrance**(W) is not uniform in* $\mathrm{SubTree}(G, r, m)$.

Consider τ_A as the hitting time of the set A and recall that for a rooted tree (t, r) we let $a(v)$ denote the ancestor of v towards the root.

Proposition (F.-Marckert ('21+))

For any spanning tree t of G we have

$$\sum_{\ell} \prod_{i=0}^{n-2} \mathbb{P}_{a(\ell_{i+\mathbf{1}})} \left(\overleftarrow{\tau}_{\{\ell_i\}} < \overleftarrow{\tau}_{\{\ell_{i+\mathbf{1}}, \ldots, \ell_{n-\mathbf{1}}\}} \right) = 1,$$

where the sum ranges over the set of decreasing labeling of (t, r).
Moreover, this is not true if t is not a spanning tree.

Do I have more time?

No → Thanks!

Yes → Wait for it!

IV. Markov Chain in SubTree(G, r)

Assume $X_i = t$ is an element of SubTrees(G, r). To define X_{i+1}, proceed as follows. Pick independently, a random edge $\vec{e} \sim$ Uniform$(\vec{E}(G))$ and "a random choice $\mathbf{c}$" satisfying

$$\mathbb{P}(\mathbf{c} = +1) = p_{|t|}, \quad \mathbb{P}(\mathbf{c} = 0) = q_{|t|}, \quad \mathbb{P}(\mathbf{c} = -1) = r_{|t|},$$

- if $\mathbf{c} = +1$ then "try to add e": if $t \cup \{e\}$ is a tree, set $X_{i+1} = t \cup \{e\}$. If it has a cycle, then pick X_{i+1} according to BreakCycle$(t \cup \{e\}, e)$, else $X_{i+1} = t$.
- if $\mathbf{c} = 0$, do nothing, and set $X_{i+1} = t$,
- if $\mathbf{c} = -1$, then "try to remove $\vec{e}$": set $X_{i+1} = t \setminus \vec{e}$ if it is a tree and does not remove the root r, else $X_{i+1} = t$.

II.2. Markov Chain in SubTree(G, r)

Proposition (F.-Marckert ('21))

The MC previously defined is reversible and its unique invariant measure ρ_r on SubTree(G, r) *gives the same weight ν_n to each element in* SubTree(G, r, n)*, for all $1 \le n \le |V|$, that is $\rho_t = \nu_{|t|}$. The sequence $\nu_k : k \in \{1, 2, \ldots, |V|\}$ satisfies:*

$$\nu_m = \nu_1 \prod_{i=2}^{m} \left(\frac{p_{i-1}}{r_i} \right), \quad \forall m \in \{2, 3, \ldots, |V|\}$$

$$\sum_{n=1}^{|V|} \nu_n |\mathrm{SubTree}(G, r, n)| = 1$$

Remark

- *Tunning p, r, q one can target a size w.h.p. even concentrate in an interval.*
- *Conditioning on the size of the tree, we obtain the uniform distribution + simple conditions on p, q, r.*

II.2. Subcase: the graph G is a tree.

We obtain a coupling from the past and we give explicit bounds on the coupling time.

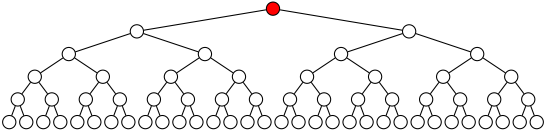
$$\text{Hypothesis M}: p_1 \leq p_2 \leq \cdots \leq p_{|V|-1}$$
$$r_2 \geq \ldots \geq r_{|V|}$$

II.2. Subcase: the graph G is a tree.

We obtain a coupling from the past and we give explicit bounds on the coupling time.

$$\text{Hypothesis M}: p_1 \leq p_2 \leq \cdots \leq p_{|V|-1}$$

$$r_2 \geq \ldots \geq r_{|V|}$$



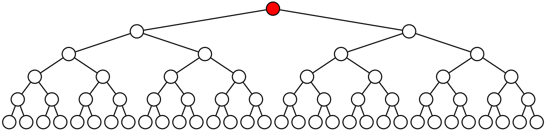
(a) Initialization

II.2. Subcase: the graph G is a tree.

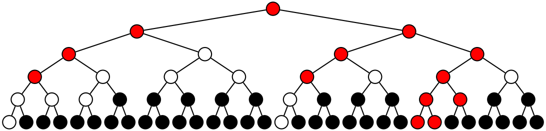
We obtain a coupling from the past and we give explicit bounds on the coupling time.

$$\text{Hypothesis M}: p_1 \leq p_2 \leq \cdots \leq p_{|V|-1}$$

$$r_2 \geq \ldots \geq r_{|V|}$$



(a) Initialization

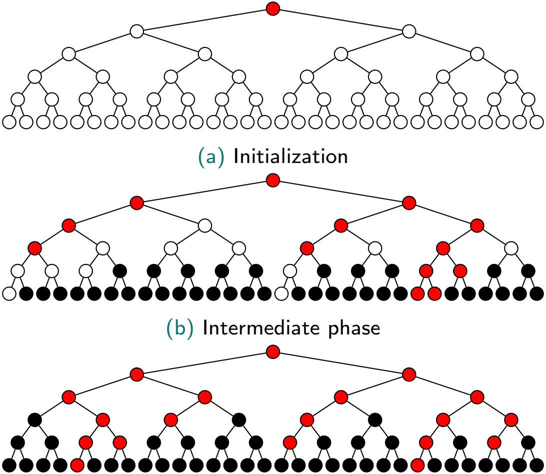


(b) Intermediate phase

We obtain a coupling from the past and we give explicit bounds on the coupling time.

$$\text{Hypothesis M}: p_1 \leq p_2 \leq \cdots \leq p_{|V|-1}$$

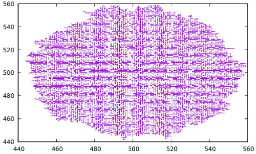
$$r_2 \geq \ldots \geq r_{|V|}$$



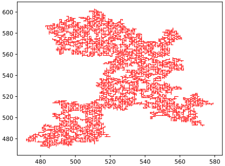
(a) Initialization

(b) Intermediate phase

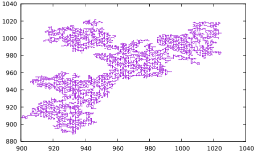
(c) Merged state



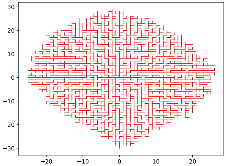
(a) FPP on the $(\mathbb{Z}/1000\mathbb{Z})^2$ with i.i.d. uniform labels on $[0, 1]$. Tree size 10k.



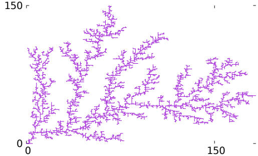
(b) Kruskal's tree of size 5k containing on $(\mathbb{Z}/1000\mathbb{Z})^2$.



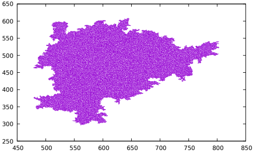
(c) Prim's tree of size 5k on $(\mathbb{Z}/2000\mathbb{Z})^2$.



(d) Tree Internal DLA with 2000 vertices.



(e) DLA tree with 5k $(\mathbb{Z}/1000\mathbb{Z})^2$.



(f) Size biased forest, tree component on $(\mathbb{Z}/2000\mathbb{Z})^2$.

THANKS!