Random subtree generation of a given graph

Luis Fredes
(Work with J.F. Marckert)

IMB
Journées de Probabilités et Statistique en Nouvelle Aquitaine
Avril 2023.

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.

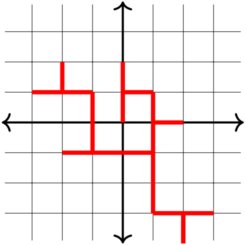
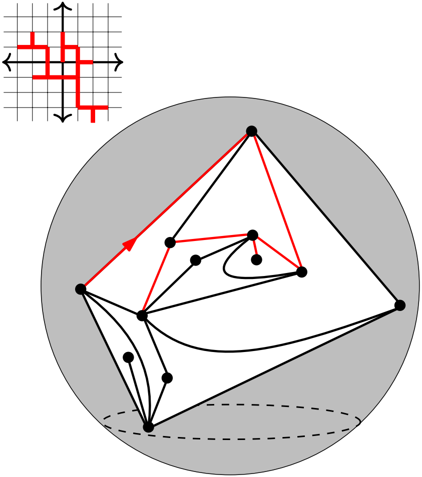
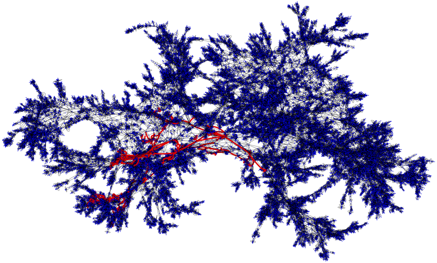


Figure: Subtree of size 20 containing the origin on $\mathbb{Z}^2$.



(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

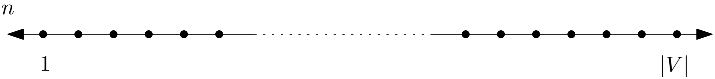
$\mathrm{SubTree}(G, r, n) =$ set of subtrees of G containing r of size n.

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

Chart of algorithms

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

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



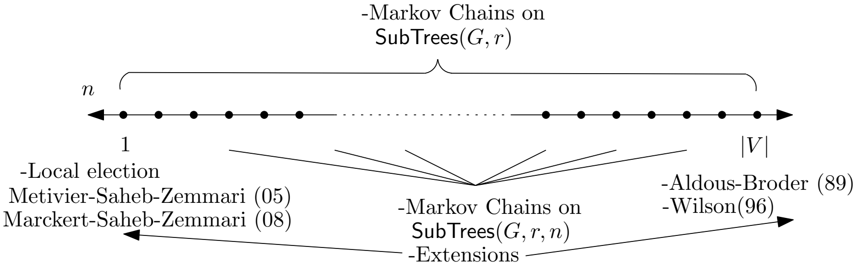
-Local election
Metivier-Saheb-Zemmari (05)
Marckert-Saheb-Zemmari (08)

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

Chart of algorithms

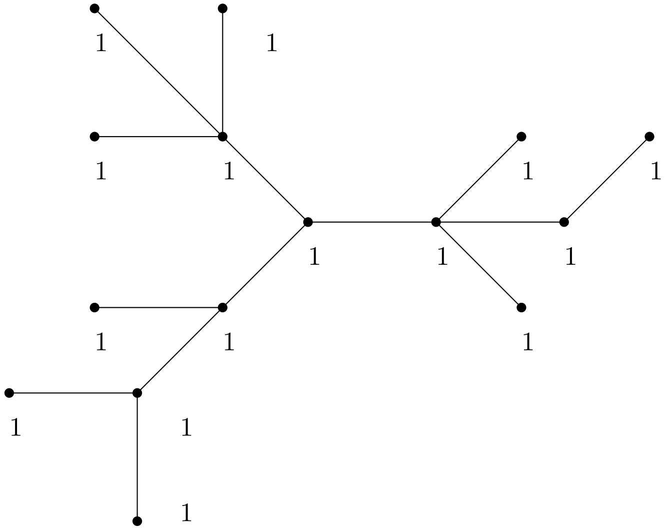
$\mathsf{SubTree}(G, r, n) =$ set of subtrees of G containing r of size n.

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

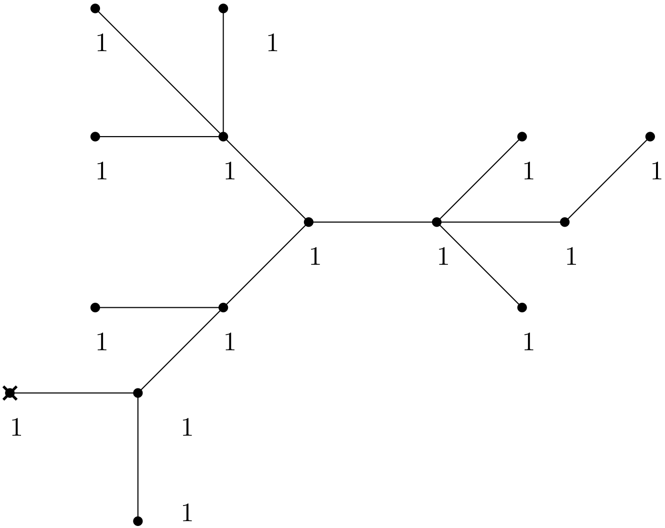


I. Local election to sample one vertex

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

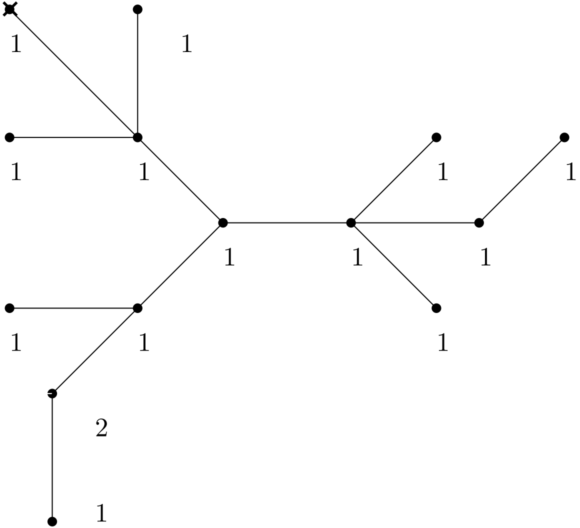


If we cannot uniformly sample in SubTree(G, r, n) for G when it is a tree, we are hopeless!

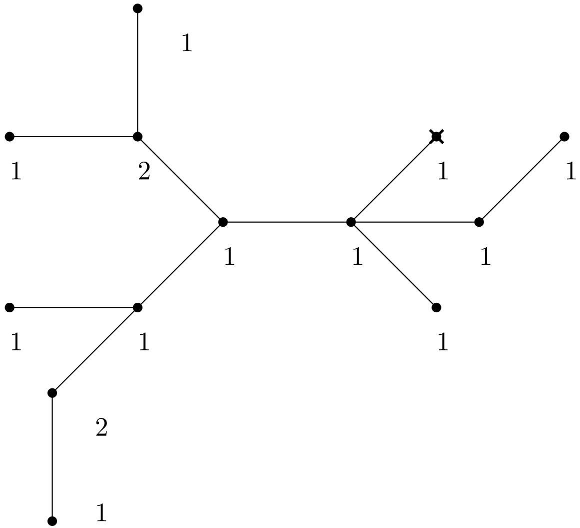


I. Local election to sample one vertex

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

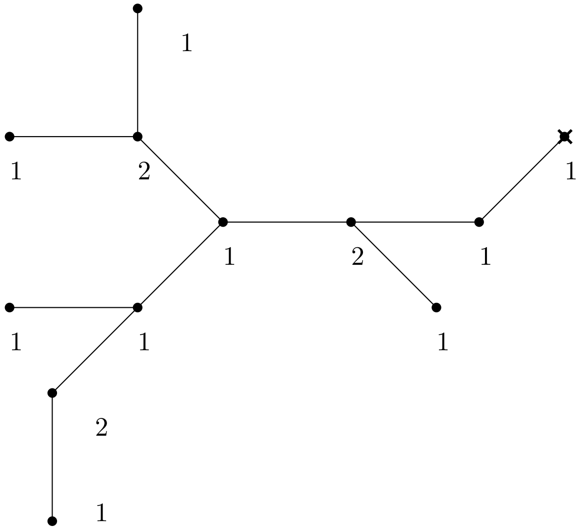


If we cannot uniformly sample in $\mathsf{SubTree}(G, r, n)$ for G when it is a tree, we are hopeless!



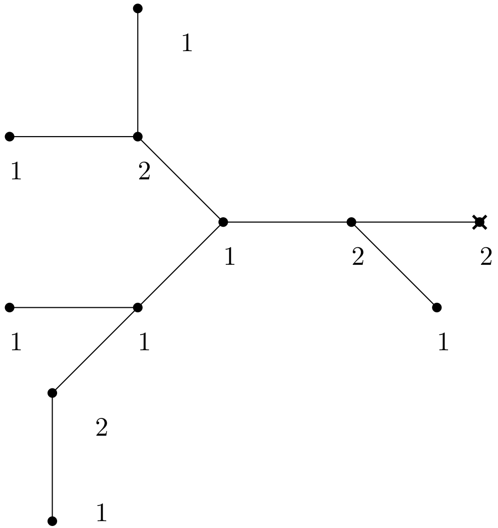
I. Local election to sample one vertex

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



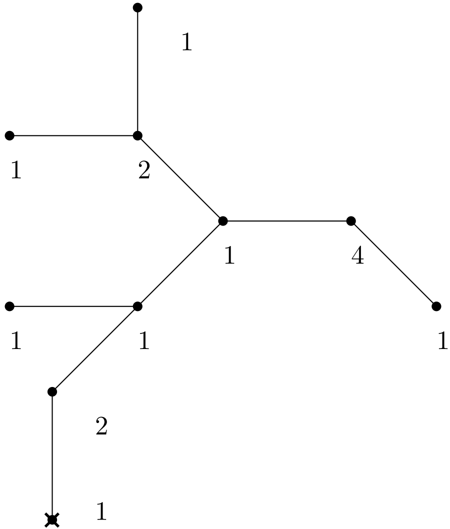
I. Local election to sample one vertex

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



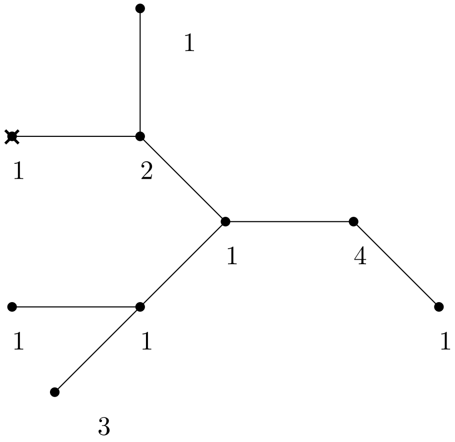
I. Local election to sample one vertex

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

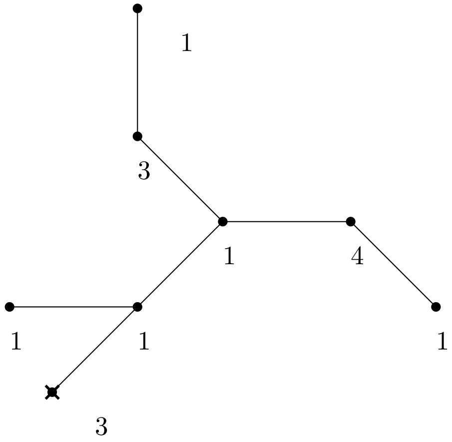


I. Local election to sample one vertex

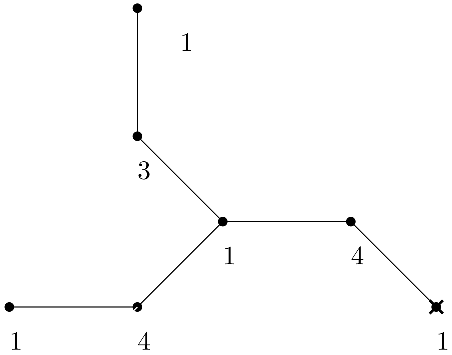
If we cannot uniformly sample in $\mathsf{SubTree}(G, r, n)$ for G when it is a tree, we are hopeless!



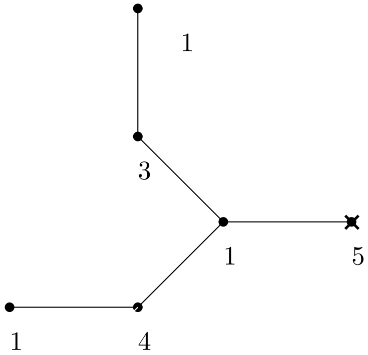
If we cannot uniformly sample in $\mathsf{SubTree}(G, r, n)$ for G when it is a tree, we are hopeless!



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

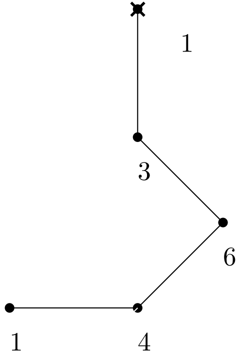


If we cannot uniformly sample in $\mathrm{SubTree}(G, r, n)$ for G when it is a tree, we are hopeless!



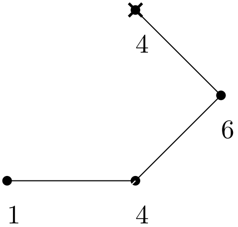
I. Local election to sample one vertex

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

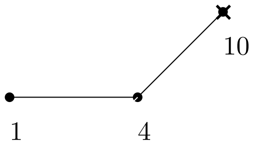


I. Local election to sample one vertex

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

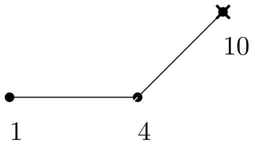


If we cannot uniformly sample in $\mathsf{SubTree}(G, r, n)$ for G when it is a tree, we are hopeless!



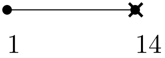
I. Local election to sample one vertex

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 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 ('22))

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.

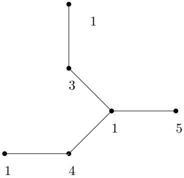


Figure: Tree of size 6 in the algorithm. Only the green one is considered in the probability.

Proof idea

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

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

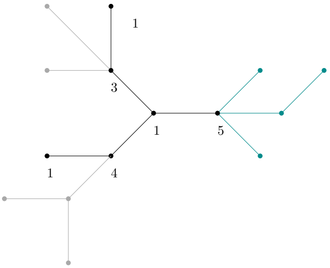


Figure: Tree of size 6 in the algorithm. Only the green one is considered in the probability.

Proof idea

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

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

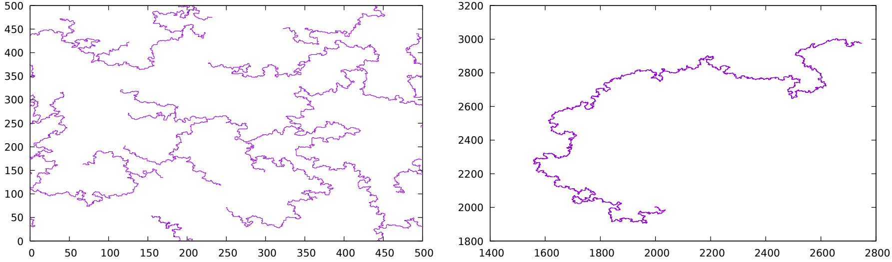


Figure: A tree Evaporation$(T, 1000)$ on T a UST of resp. $(\mathbb{Z}/500\mathbb{Z})^2$ and $(\mathbb{Z}/4000\mathbb{Z})^2$

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 of t, 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. **Otherwise, the addition of $\vec{e}$ creates a cycle:** sample an edge e' according to $\mathrm{BreakCycle}(t \cup \{e\}, e)(\cdot)$ and define $X_{i+1} = t \cup \{e\} \setminus \{e'\}$

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

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 of t, 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. **Otherwise, the addition of $\vec{e}$ creates a cycle:** sample an edge e' according to BreakCycle$(t \cup \{e\}, e)(\cdot)$ and define $X_{i+1} = t \cup \{e\} \setminus \{e'\}$

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

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

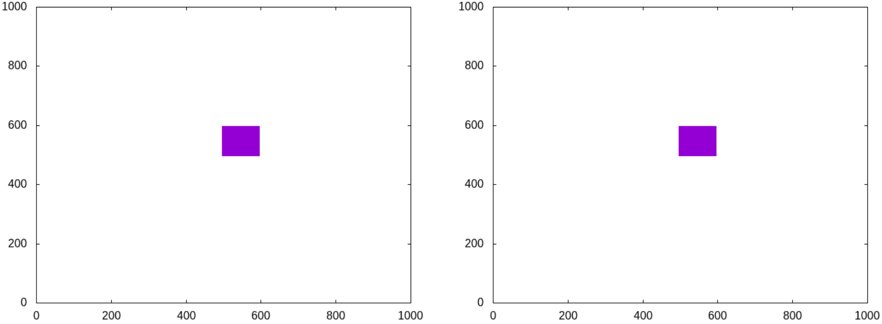


Figure: 1M and 100M iteration by frame.

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 (L. - Marckert ('22))

1. *There exists* $\alpha \in [0.65 \pm 0.02]$ *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* $\beta \in [3/4 \pm 0.01]$ *s.t.*

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

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

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

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

II.z Simulation results

$T_n =$ Uniform element in SubTree$((\mathbb{Z}/n\mathbb{Z})^2, n)$
$q_i(T_n) =$ 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.

III. Spanning trees.

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.

III. Spanning trees.

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 $\vec{e}$.

III.1. Wilson (Cycle popping version)

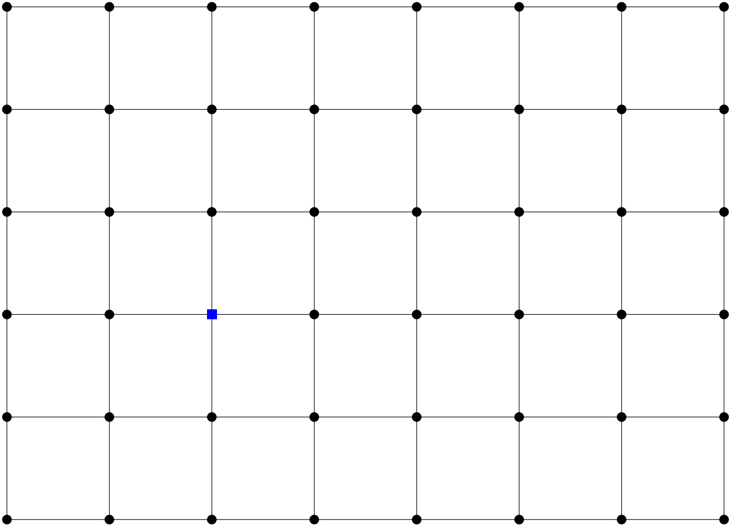


Figure: Pick any vertex as root (square vertex)

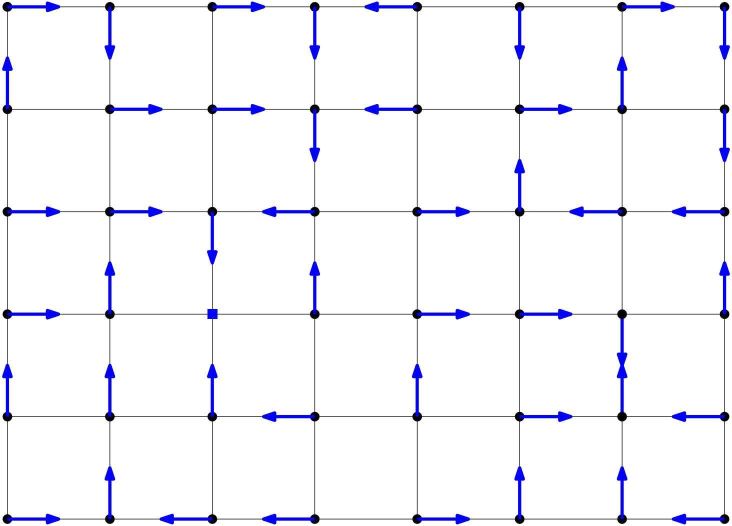


Figure: Pick one outgoing edge for each $v \in V \setminus \{r\}$ following the markov kernel M.

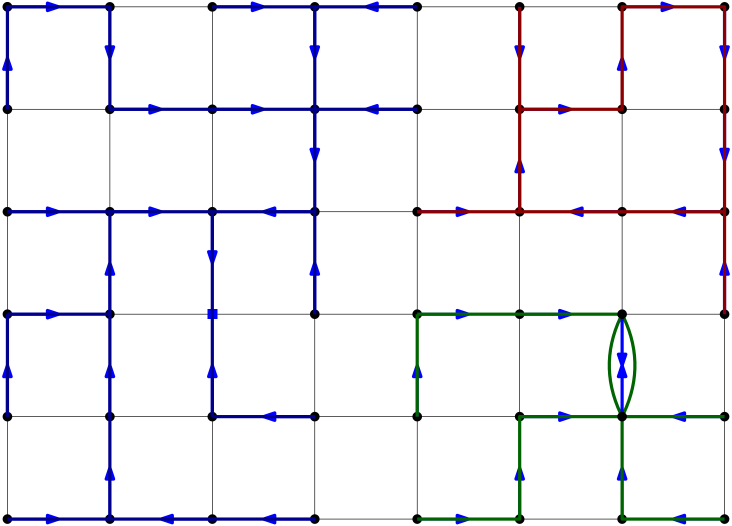


Figure: The oriented edges induce a graph.

III.1. Wilson (Cycle popping version)

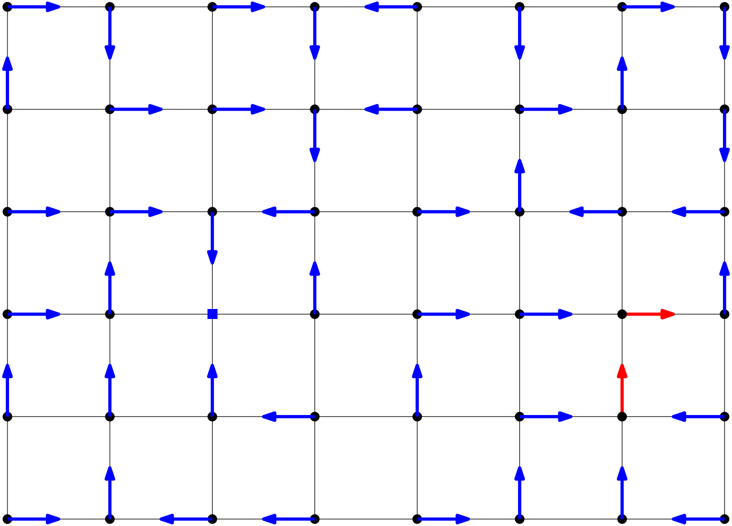


Figure: If there is a cycle pick one and re-sample the outgoing edges of the vertices on it.

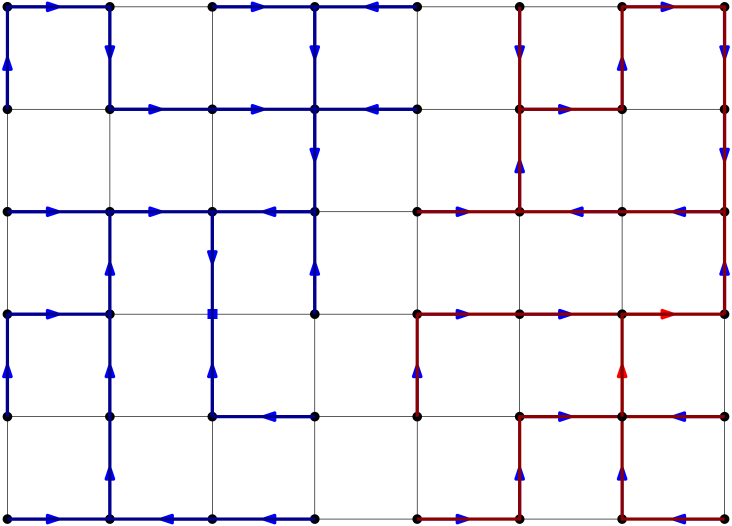


Figure: Induced graph.

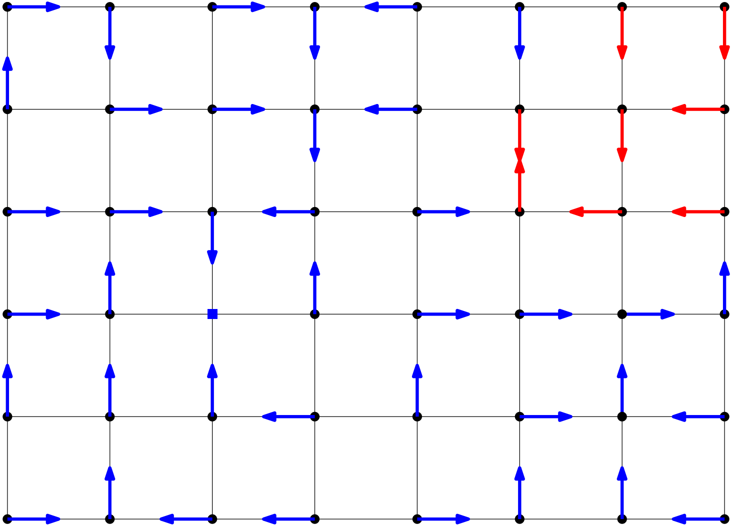
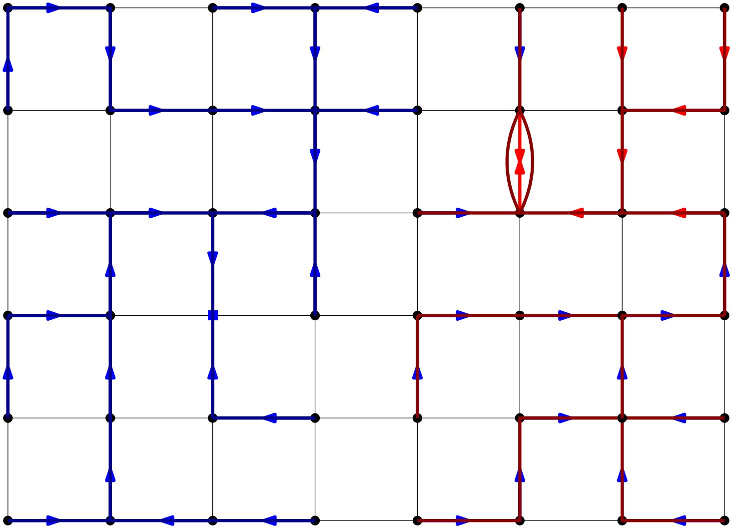
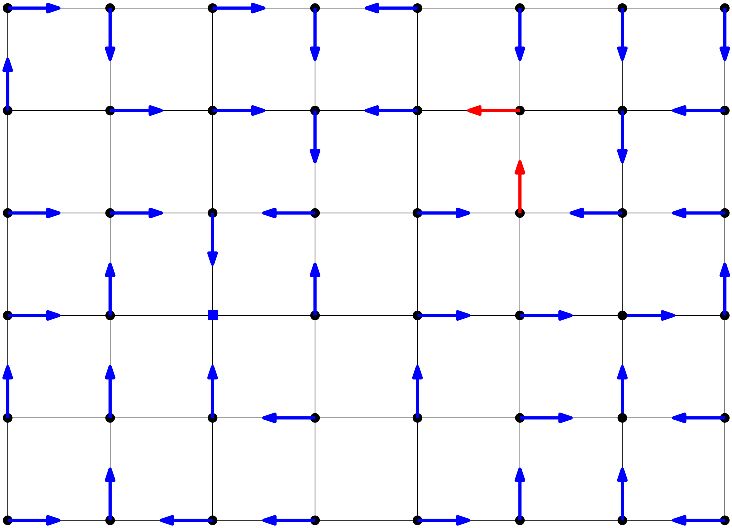


Figure: Pick a cycle and resample again.

III.1. Wilson (Cycle popping version)



III.1. Wilson (Cycle popping version)



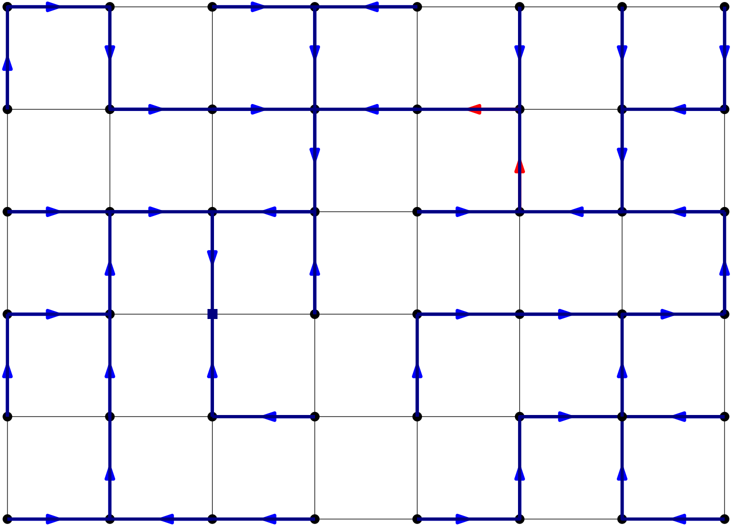


Figure: Stop when there is no more cycle, i.e. a tree.

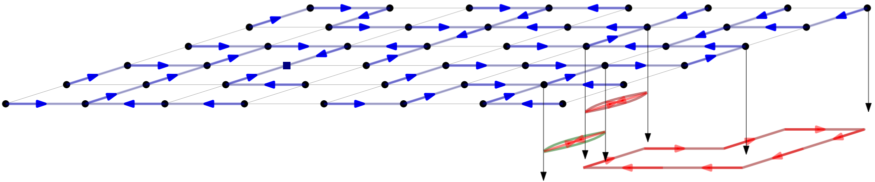


Figure: Heap of cycles × Tree

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

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

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.

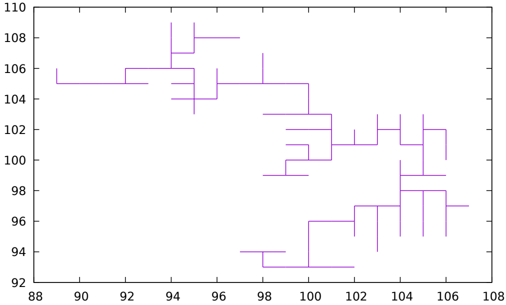


Figure: Simulation of τ on $(\mathbb{Z}/100\mathbb{Z})^2$, 3536949 simulations were needed to get a tree of size at least 100.

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.

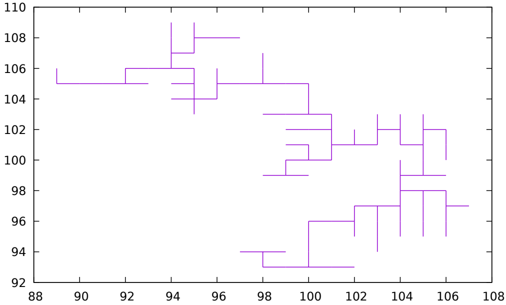
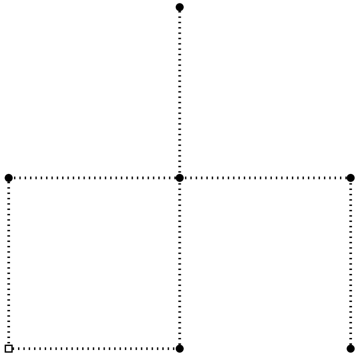


Figure: Simulation of τ on $(\mathbb{Z}/100\mathbb{Z})^2$, 3536949 simulations were needed to get a tree of size at least 100.

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.

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

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

Theorem (Aldous-Broder ('89))

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

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

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

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

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 ('22); Hua- Lyons- Tang ('21))

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

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

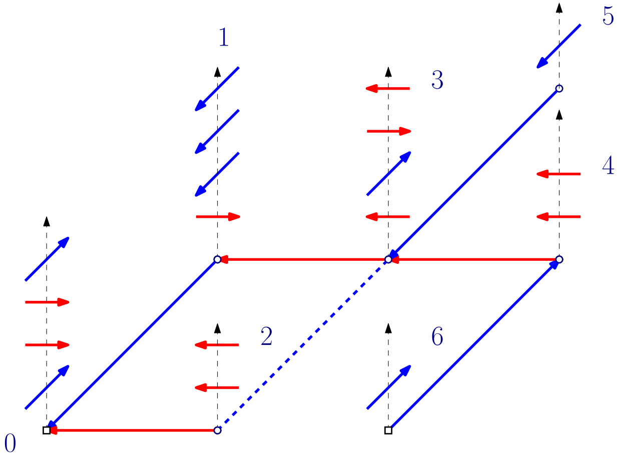


Figure: Path seen backward as a heap of outgoing edges

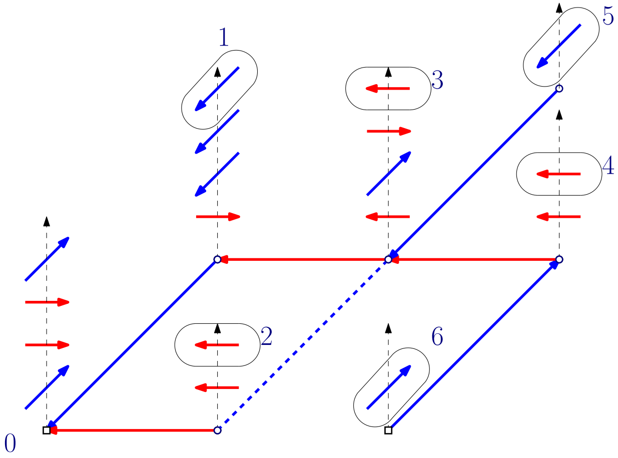


Figure: The tree edges are always on top of the piles.

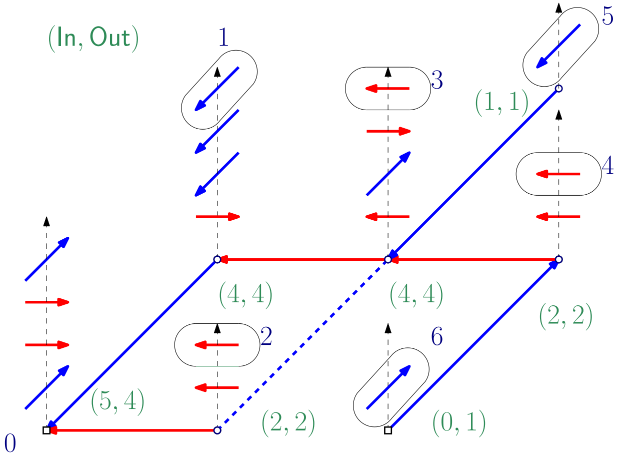


Figure: Count the incoming and outgoing edges

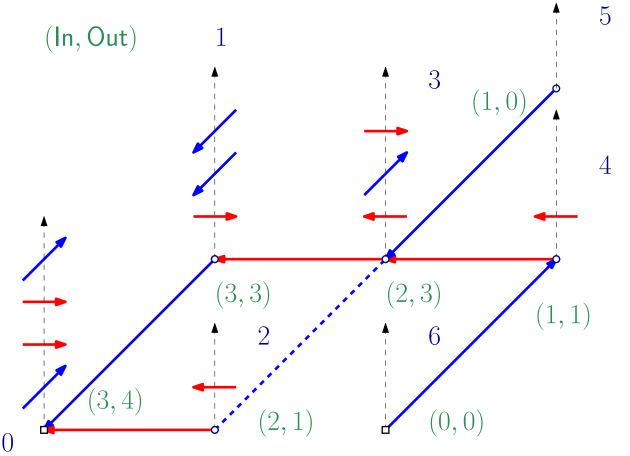


Figure: Pop-out the tree edges to construct H^{-t} (update (In,Out))

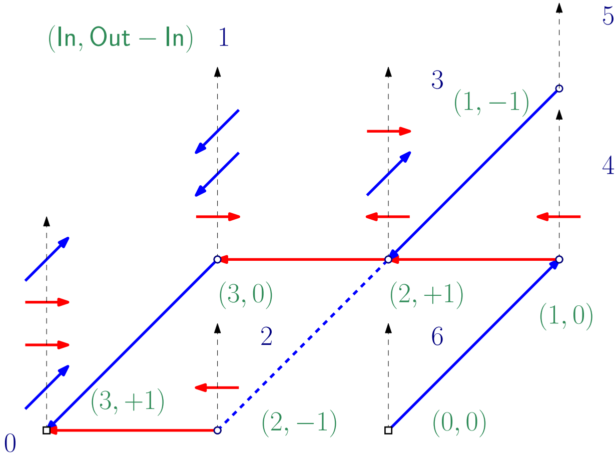


Figure: Convenient to keep an eye on (In,Out-In)

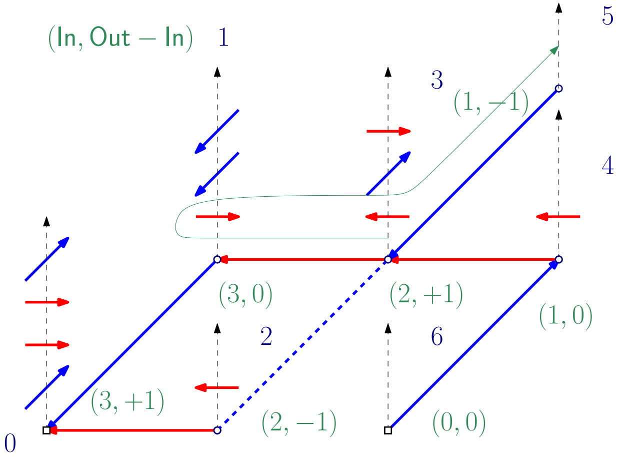


Figure: Play golf!

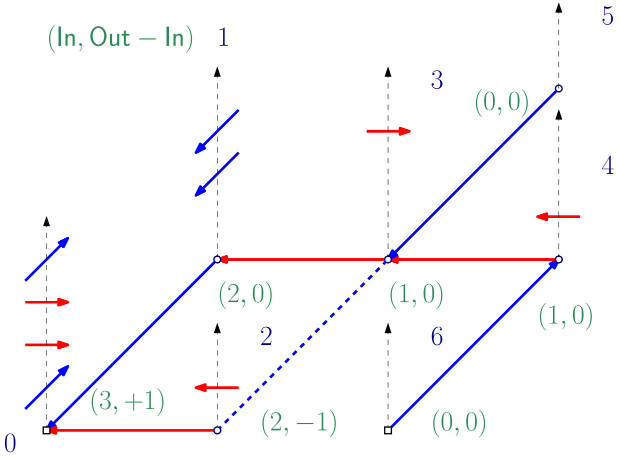


Figure: Supress the path and update (In,Out-In)

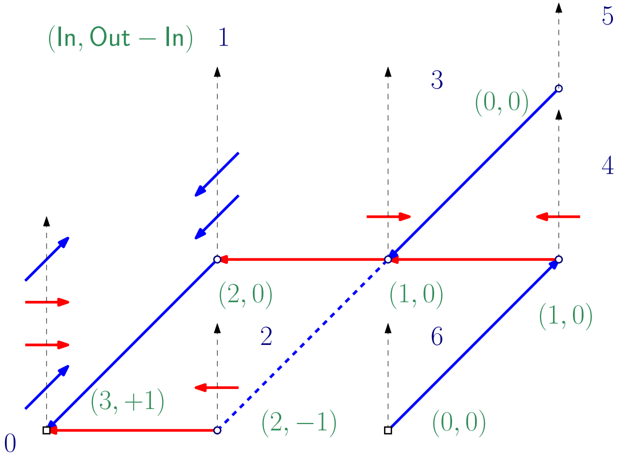


Figure: Let the pieces fall

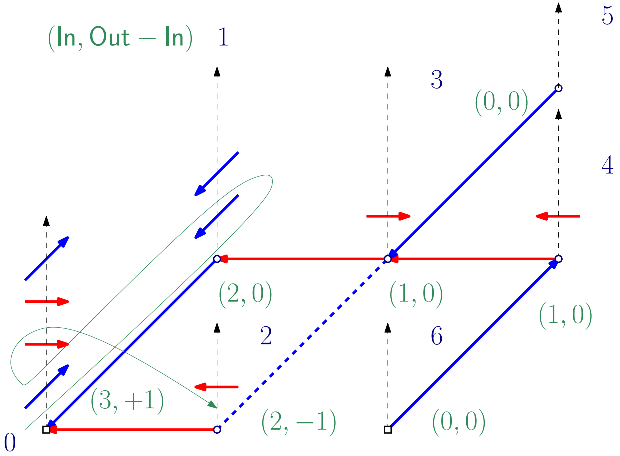


Figure: Continue playing golf with next emitting vertex.

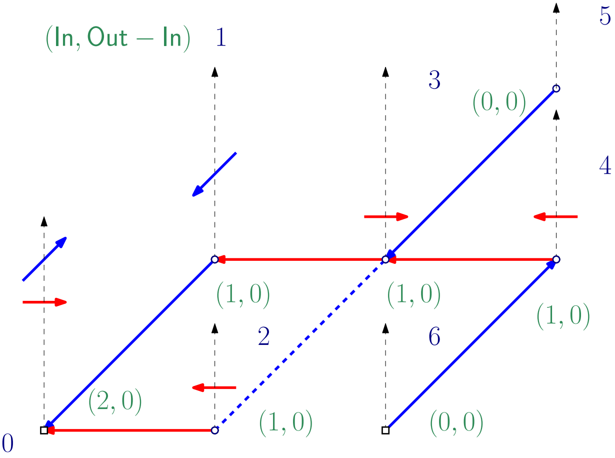


Figure: Supress the path and update (In,Out-In)

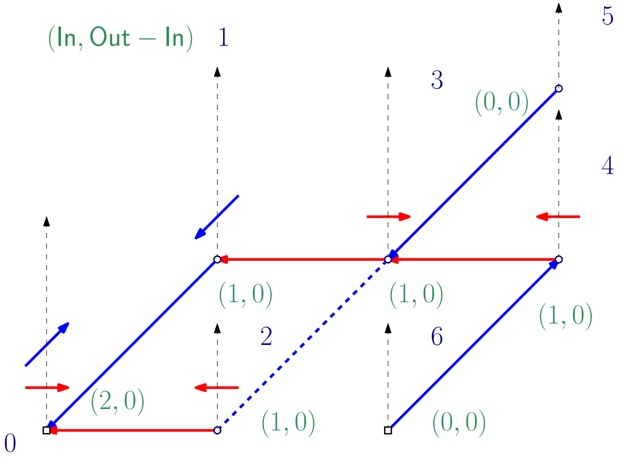


Figure: Let the pieces fall

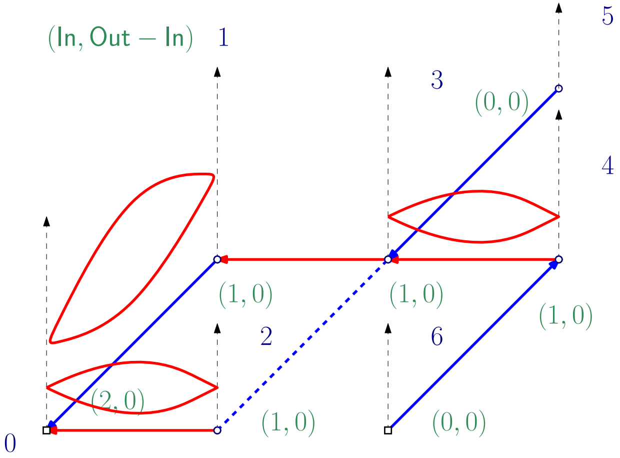


Figure: heap of cycles

Idea: Each path W started at r and up to its cover time can be decomposed as

$$W = T \times Golf \times HC$$

We proved that this is in fact a bijection.

III.2. Consequences

Idea: Each path W started at r and up to its cover time can be decomposed as

$$W = T \times Golf \times HC$$

We proved that this is in fact a bijection.

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* $\mathrm{SubTree}(G, r, m)$.

THANKS!