

Marches aléatoires et arbres couvrants uniformes sur des graphes connexes

Sylvain Chabredier et Gaspard Choné-Ducasse
sous la direction de Rémy Mahfouf et Théo Lenoir

Avril-Juin 2022

Résumé

L'objet de ce mémoire est l'étude d'arbres couvrants sur le quadrillage \mathbb{Z}^d et d'établir un résultat de connexité dû à Pemantle ([Pem91], 1991). On présentera aussi des résultats sur des graphes quelconques, pour lesquels, en associant à chaque arête un poids, on viendra à parler de marches aléatoires, tension et conductance effective. Ce mémoire se base en grande partie sur les notes de cours de Tom Hutchcroft à l'université de Cambridge ([Hut21], 2021).

Dans la section **2**, on étudie les fonctions harmoniques sur des graphes et on introduit la notion de tension entre ensembles. Ces fonctions particulières répondent au problème de Dirichlet : pour toute fonction quelconque sur un sous graphe, on peut la prolonger en unique fonction harmonique sur le reste du graphe. On définira aussi des éléments de calculs différentiels sur les graphes comme le gradient ou la divergence ce qui nous permettra d'établir le principe de Dirichlet qui donne une expression analytique de la conductance effective entre ensembles. A partir de ce théorème, on montre le principe de monotonie de Rayleigh.

Dans la section **3**, on introduit les arbres couvrants et pour les graphes finis, on définit une loi dite uniforme de probabilité de piocher un tel arbre. A l'aide des marches aléatoires, on propose deux algorithmes introduits par Aldous ([Ald90], 1990) et Broder ([Bro89], 1989) d'une part et par Wilson ([Wil96], 1996) d'autre part permettant d'échantillonner la loi uniforme.

Finalement, dans la section **4**, on traite le cas des graphes infinis en définissant la loi uniforme sur les graphes infinis comme la limite de la loi uniforme sur des exhaustions. On étend ensuite l'algorithme de Wilson pour démontrer le théorème de Pemantle.

Nous tenons à remercier chaleureusement Rémy Mahfouf et Théo Lenoir pour nous avoir proposé ce sujet et pour nous avoir encadré et conseillé tout au long de ce mémoire.

Table des matières

| | | |
|----------|---|-----------|
| 1 | Introduction | 3 |
| 2 | Principes de Dirichlet et de monotonie de Rayleigh | 3 |
| 2.1 | Fonctions harmoniques | 3 |
| 2.2 | Principe de Dirichlet | 5 |
| 2.3 | Principe de monotonie de Rayleigh | 7 |
| 3 | Échantillonnage d'arbres couvrants | 7 |
| 3.1 | L'algorithme d'Aldous Broder | 8 |
| 3.2 | Propriétés spatiales de Markov | 12 |
| 3.3 | L'algorithme de Wilson | 14 |
| 3.3.1 | Procédé d'effacement de cycles | 14 |
| 3.3.2 | Algorithme de Wilson | 17 |
| 4 | Cas des graphes infinis | 18 |
| 4.1 | Forêts couvrantes uniformes sur des graphes infinis | 19 |
| 4.2 | Algorithme de Wilson | 21 |
| 4.3 | Théorème de Pemantle | 23 |
| 5 | Conclusion | 26 |

1 Introduction

On étudie dans ce mémoire les arbres d'un graphe connexe en définissant une mesure sur ces arbres présentant une certaine uniformité. La section 2 donne des outils généraux pour étudier les graphes. La section 3 étudie les arbres sur des graphes finis, sur lesquels définir la mesure précédemment citée ne pose pas de problème. Cette section étudie alors des propriétés remarquables de cette mesure et son lien avec les marches aléatoires sur le graphe en question. Dans la section 4, le problème des graphes infinis est étudié, où il n'est plus aussi facile de définir la mesure voulue. On se ramène alors à des sous-graphes finis de plus en plus grands et on essaye de passer à la limite pour définir la mesure voulue. Cependant, en créant la mesure de cette façon, toutes les propriétés ne sont pas conservées : la mesure créée n'est plus concentrée sur les sous-graphes connexes par exemple. Elle conserve cependant des propriétés remarquables et possède toujours un lien avec les marches aléatoires. Finalement, la fin du mémoire étudie le graphe particulier \mathbb{Z}^d et sa connexité : La mesure créée est concentrée sur les sous-graphes connexes de \mathbb{Z}^d pour $d \leq 4$ et est concentrée sur les sous-graphes non connexes de \mathbb{Z}^d pour $d \geq 5$. On traitera seulement les cas $d \geq 5$ et $d \leq 2$.

2 Principes de Dirichlet et de monotonie de Rayleigh

Dans ce chapitre on cherche à démontrer les principes de Dirichlet et de monotonie de Rayleigh qui seront nécessaires à la bonne définition de la loi uniforme sur les arbres couvrants dans des graphes infinis. Dans cet objectif, on introduit la notion de fonction harmonique et de calcul différentiel discret sur un graphe et on verra apparaître une fonction particulière : la tension entre ensembles.

On adoptera pour l'ensemble du mémoire les quelques notations usuelles suivantes. Un graphe G est une paire (V, E) où V est l'ensemble des sommets et E l'ensemble des arêtes, on se restreint à l'étude des graphes simples. On se donne une fonction de pondération $c : E \mapsto]0, +\infty[$ telle que $c(v) := \sum_{e \in E_v} c(e) < \infty$. Pour v une arête de G , on appelle $c(v)$ la conductance de l'arête v . Par exemple, pour le quadrillage \mathbb{Z}^d , on donnera à toutes les arêtes la même conductance. On travaillera sur des graphes non orientés, mais il sera parfois utile de voir les arêtes non orientées comme deux arêtes orientées ayant chacune la conductance de l'arête non orientée de laquelle elle émane, on notera ainsi l'ensemble E^\rightarrow des arêtes orientées ainsi formées à partir des arêtes de G . Pour e un élément de E^\rightarrow , on note e^- sa queue et e^+ sa tête et pour tout sommet v , on définit $E_v^\rightarrow := \{e \in E^\rightarrow, e^- = v\}$, l'ensemble des arêtes émanant de v .

L'outil fondamental de notre étude est les marches aléatoires sur les graphes. Ce sont les chaînes de Markov de fonction de transition P définie par $P(e^-, e^+) = \frac{c(e)}{c(e^-)}$ pour toute arête e de E^\rightarrow . On notera $X = (X_n)_{n \geq 0}$ une telle marche. Le conditionnement par " $X_0 = v$ " d'un évènement A sera noté $\mathbb{P}_v(A)$. On dit qu'une mesure π est invariante pour la chaîne si elle vérifie l'égalité $\pi P = P$ dans une écriture matricielle. On rappelle qu'une chaîne de Markov est dite irréductible si tous les états communiquent, dans cette situation il existe au plus une mesure invariante.

2.1 Fonctions harmoniques

Dans cette section, nous introduisons les fonctions harmoniques sur des graphes et résolvons le problème de Dirichlet qui est de prolonger une fonction sur un graphe par une fonction harmonique.

Définition 2.1. On dit qu'une fonction $f : V \rightarrow \mathbb{R}$ est harmonique sur $W \subset V$ si pour tout $u \in W$:

$$f(u) = \sum_{v \in V} P(u, v) f(v).$$

Ces fonctions vérifient le principe du maximum.

Lemme 2.2 (Principe du maximum). Soit f une fonction harmonique sur W , on a pour tout $u \in W$:

$$f(u) \leq \sup_{v \sim u} f(v). \quad (1)$$

De plus, si W est connexe et il existe $u \in W$ tel que $f(u) = \sup_{v \in V} f(v)$, alors f est constante sur W et sa frontière.

Preuve. L'inégalité découle directement de la définition :

$$f(u) \stackrel{\text{def}}{=} \sum_{v \in V} P(u, v) f(v) \leq \sum_{v \sim u} P(u, v) \cdot (\sup_{w \sim u} f(w)) = \sup_{w \sim u} f(w).$$

Soit u réalisant le maximum de f , il y a alors égalité dans l'inégalité (1). Il vient donc $f(v) = \sup_{w \sim u} f(w) = f(u)$ pour v voisin de u . On en déduit la deuxième affirmation par connexité.

La définition suivante donne un exemple d'une fonction harmonique particulière qu'on nomme tension. On peut la rapprocher de la tension dans un réseau électrique.

Définition 2.3 (Tension entre deux ensembles). Soit A et B deux ensembles disjoints de sommets d'un graphe fini G .

Soit τ_A le temps de première visite de A par une marche aléatoire défini par : $\tau_A = \inf\{n \geq 0, X_n \in A\}$. Par convention si $A = \emptyset$, on pose $\tau_A = +\infty$.

On définit enfin :

$$F(v) = \mathbb{P}_v(\tau_B < \tau_A).$$

C'est une fonction harmonique sur $V \setminus (A \cup B)$, identiquement nulle sur A et identiquement égale à 1 sur B . On la nomme tension de A vers B .

Preuve. Soit $u \notin A \cup B$, si l'on conditionne la marche selon X_1 , on obtient :

$$F(u) = \sum_{v \in V} \mathbb{P}_u(X_1 = v) \mathbb{P}_v(\tau_B < \tau_A) = \sum_{v \in V} P(u, v) F(v).$$

On a donc montré que F est harmonique sur $V \setminus (A \cup B)$.

Le résultat suivant donne l'existence et l'unicité d'une solution dans le problème de Dirichlet.

Théorème 2.4. Soit G un graphe connexe fini et A un sous-ensemble non-vide de V . On se donne une fonction $\phi : A \mapsto \mathbb{R}$ arbitraire. Alors il existe une unique fonction $h : V \mapsto \mathbb{R}$ telle que $h \upharpoonright_A = \phi$ et h est harmonique sur $V \setminus A$.

Preuve. Reformulons le problème. Soit $H_{V \setminus A} \subset \mathbb{R}^V$ l'espace des fonctions harmoniques sur $V \setminus A$, c'est un espace vectoriel. Le théorème est équivalent au fait que l'application linéaire de $H_{V \setminus A}$ vers \mathbb{R}^A associant à chaque fonction harmonique ses valeurs en sa frontière est un isomorphisme.

Injectivité. Soit $f \in H_{V \setminus A}$ identiquement nulle sur A . Comme $(-f)$ est dans $H_{V \setminus A}$ et est identiquement nulle sur A , il suffit de montrer que : $\sup_{v \in V} f(v) \leq 0$ pour avoir $f = 0$. Si le maximum de f est atteint sur A , on a le résultat. Supposons alors qu'il est atteint en $v \in V \setminus A$. Soit W la composante connexe

de v dans $V \setminus A$. Par le lemme 2.2, le maximum de f est atteint sur W et sa frontière qui intersecte A . Finalement f est identiquement nulle sur V et l'application est injective.

Surjectivité. Soit $\phi : A \mapsto \mathbb{R}$ quelconque. On définit $h : V \rightarrow \mathbb{R}$ par :

$$h(v) = \mathbb{E}_v[\phi(X_{\tau_A})].$$

La fonction est bien définie car G est fini et connexe donc τ_A est presque sûrement fini. De plus, $h \upharpoonright_A = \phi$, il reste donc juste à montrer qu'elle est harmonique. Comme dans la preuve précédente, on conditionne sur X_1 . Pour $u \notin A$, il vient :

$$h(u) = \sum_{v \in V} \mathbb{P}_v(X_1 = v) \mathbb{E}_v[\phi(X_{\tau_A}) | X_1 = v] = \sum_{v \in V} P(u, v) h(v).$$

Ce qui achève la preuve.

2.2 Principe de Dirichlet

Dans cette section on introduit la conductance entre ensembles puis on cherche à démontrer le principe de Dirichlet qui permet d'exprimer cette conductance à partir des fonctions sur le graphe. On suppose dans cette partie que G est fini.

Définition 2.5. On suppose G fini. Pour $A, B \subset V$ disjoints, on définit la conductance effective entre A et B comme

$$C_{\text{eff}}(A \leftrightarrow B) := \sum_{v \in A} c(v) \mathbb{P}_v(\tau_B < \tau_A^+)$$

où $\tau_A^+ = \inf\{n \geq 1, X_n \in A\}$.

On introduit à présent les notions de calcul différentiel sur G .

Définition 2.6. Pour une fonction $f : V \rightarrow \mathbb{R}$, on définit le gradient de f comme la fonction $\nabla f : E^{\rightarrow} \rightarrow \mathbb{R}$ par

$$\nabla f(e) := c(e)(f(e^+) - f(e^-)).$$

Pour une fonction $f : E^{\rightarrow} \rightarrow \mathbb{R}$ telle que $f(e^{\leftarrow}) = -f(e)$ pour tous $e \in E^{\rightarrow}$, on définit la divergence de f comme la fonction $\nabla \cdot f : V \rightarrow \mathbb{R}$ par

$$\nabla \cdot f(v) := \sum_{e \in E_v} f(e).$$

Pour une fonction $f : V \rightarrow \mathbb{R}$, on définit le laplacien de f comme la fonction $\Delta f : V \rightarrow \mathbb{R}$ par $\Delta f := -\nabla \cdot \nabla f$.

Remarque 2.7. Une fonction réelle f est harmonique sur $W \subset V$ si et seulement si $\Delta f = 0$ sur W .

Remarque 2.8. Pour $f : E^{\rightarrow} \rightarrow \mathbb{R}$ telle que $f(e^{\leftarrow}) = -f(e)$ pour tous $e \in E^{\rightarrow}$,

$$\sum_{v \in G} \nabla \cdot f(v) = \sum_{v \in G} \sum_{e \in E_v} f(e) = 0$$

car on retrouve chaque orientation de chaque arête 1 fois. On dit que la divergence totale de G est nulle.

Définition 2.9. L'espace des fonctions réelles sur V est naturellement muni d'un produit scalaire

$$\langle f, g \rangle := \frac{1}{2} \sum_{e \in E^{\rightarrow}} r(e) f(e) g(e)$$

où $r(e) := \frac{1}{c(e)}$ est la résistance de l'arête e .

Lemme 2.10. *Pour $A, B \subset V$, si on note F la tension de G entre A et B , alors*

$$\mathcal{C}_{\text{eff}}(A \leftrightarrow B) = \|\nabla F\|^2.$$

Preuve. On a

$$\mathcal{C}_{\text{eff}}(A \leftrightarrow B) = \sum_{v \in A} c(v) \mathbb{P}_v(\tau_B < \tau_A^+),$$

donc en conditionnant sur la première étape de la marche, on déduit

$$\begin{aligned} \mathcal{C}_{\text{eff}}(A \leftrightarrow B) &= \sum_{v \in A} c(v) \sum_{e \in E_v} \frac{c(e)}{c(v)} \mathbb{P}_{e^+}(\tau_B < \tau_A) = \sum_{v \in A} \sum_{e \in E_v} c(e) (\mathbb{P}_{e^+}(\tau_B < \tau_A) - \mathbb{P}_v(\tau_B < \tau_A)) \\ &= \sum_{v \in A} \sum_{e \in E_v} \nabla F(e) = \sum_{v \in A} (-\Delta F(v)). \end{aligned}$$

Mais F est harmonique sur $G - (A \cup B)$ et $\Delta = -\nabla \cdot \nabla$, donc d'après la remarque 2.8 ,

$$\sum_{v \in A} (-\Delta F(v)) = \sum_{v \in B} \Delta F(v) = \sum_{v \in G} F(v) \Delta F(v) = - \sum_{v \in G} F(v) \sum_{e \in E_v} \nabla F(e).$$

On conclut en remarquant que dans la double somme on retrouve en fait exactement une fois chaque arête orientée, donc en regroupant chaque arête et son opposé, on déduit

$$- \sum_{v \in G} F(v) \sum_{e \in E_v} \nabla F(e) = \sum_{e \in E \rightarrow} (F(e^+) - F(e^-)) \nabla F(e) = \sum_{e \in E} r(e) \nabla F(e)^2 = \|\nabla F\|^2.$$

On énonce finalement le principe de Dirichlet :

Théorème 2.11 (Principe de Dirichlet). *Si $A, B \subset V$ sont non vides et disjoints, alors*

$$\mathcal{C}_{\text{eff}}(A \leftrightarrow B) = \min\{\|\nabla F\|^2, F|_A = 0, F|_B = 1\} = \min\{\|\nabla F\|^2, F|_A \leq 0, F|_B \geq 1\}.$$

Preuve. $(F(v))_{v \in V} \in \mathbb{R}^V \mapsto \|F\|^2$ est coercive et G est fini donc \mathbb{R}^V est de dimension finie donc les minimums considérés sont bien atteints. On commence par montrer la première égalité. Pour $v \in V$, la dérivée partielle de $\|\nabla F\|^2 = \sum_{e \in E} r(e) (F(e^+) - F(e^-))^2$ par rapport à $F(v)$ est

$$\sum_{e \in E_v^+} 2r(e)(F(v) - F(e^+)) = -2 \sum_{e \in E_v^+} \nabla F(e) = -2\nabla \cdot \nabla F(v) = 2\Delta F.$$

Donc si F réalise le min voulu, elle est harmonique sur $V - (A \cup B)$ et les conditions sur A et B impliquent par le principe du maximum (lemme 2.2) que F est la tension entre A et B , on déduit donc la première égalité grâce au lemme 2.10.

Pour $F : V \rightarrow \mathbb{R}$ vérifiant $F|_A \leq 0, F|_B \geq 1$, on définit

$$F_0(v) := \begin{cases} 0 & \text{si } F(v) \leq 0 \\ F(v) & \text{si } 0 \leq F(v) \leq 1 \\ 1 & \text{si } F(v) \geq 1 \end{cases}$$

Remarquons que $|\nabla F_0(e)| \leq |\nabla F(e)|$ pour tout $e \in E$ donc $\|F_0\| \leq \|F\|$ et $F_0|_A = 0, F_0|_B = 1$, on déduit donc la deuxième égalité.

2.3 Principe de monotonie de Rayleigh

A partir du principe de Dirichlet, on démontre le principe de monotonie de Rayleigh.

Théorème 2.12 (Principe de monotonie de Rayleigh). *Si G est fini, et $A, B \subset V$ disjoints et non vides. Alors la conductance effective $\mathcal{C}_{\text{eff}}(A \leftrightarrow B, (G, c))$ est une fonction croissante de $c \in]0, \infty[^E$.*

Preuve. Si $c \leq c'$, et si on note F la tension entre A et B de (G, c') , alors $\mathcal{C}_{\text{eff}}(A \leftrightarrow B, (G, c)) \leq \|F\|^2 = \mathcal{C}_{\text{eff}}(A \leftrightarrow B, (G, c'))$.

Le raisonnement marche aussi lorsqu'on remplace une conductance non nulle par une conductance nulle : l'arête est de fait effacée. Donc lorsqu'on enlève des arêtes, la conductance diminue, donc lorsqu'on ajoute des arêtes (avec éventuellement des nouveaux sommets), la conductance augmente.

3 Échantillonnage d'arbres couvrants

On peut donc à présent commencer le coeur de l'étude : les arbres. On essaie en particulier de définir une certaine mesure, uniforme dans un sens, et de regarder les propriétés de cette mesure : sa relation avec les marches aléatoires, la génération de cette loi avec des algorithmes, ... Lorsque le graphe est infini, il y a a priori une infinité non dénombrable d'arbres sur ce graphe, et il est donc plus difficile de définir une notion canonique de loi uniforme, ce sera l'objet de la prochaine section. Dans cette section, on s'intéresse aux arbres sur des graphes finis. Il y a donc un nombre fini d'arbres sur un tel graphe, et parler de mesure sur un tel ensemble d'arbre revient à regarder la probabilité que l'on associe à chaque arbre. On pourrait donc définir une telle mesure comme la mesure uniforme sur les arbres (qui associe donc une probabilité identique à chaque arbre), mais cela est peu respectueux de l'importance de chaque arête, c'est à dire de la conductance de chaque arête. On définit plutôt une mesure qui dépend directement de la conductance de chaque arête, et on étudie certaines des propriétés remarquables de la mesure ainsi définie.

Définition 3.1 (Arbres couvrants). *Soient G un graphe connexe, fini et A un sous-graphe de G . On dit que A est un arbre couvrant si c'est un arbre (i.e. un graphe connexe et acyclique) et s'il contient tous les sommets de G . On notera \mathcal{A} l'ensemble des arbres couvrants.*

On définit aussi la loi uniforme sur l'ensemble des arbres couvrants. Pour tout arbre couvrant t de G on pose :

$$\mathbf{UST}(t) := \frac{\prod_{e \in t} c(e)}{\sum_{s \in \mathcal{A}} \prod_{e \in s} c(e)}.$$

Une variable aléatoire T suivant la loi \mathbf{UST} est appelé un arbre couvrant uniforme ou UST («uniform spanning tree»).

On peut remarquer que lorsque les conductances sont toutes égales à 1, la loi uniforme alors définie coïncide avec la «vraie» loi uniforme sur l'ensemble des arbres, donc finalement on peut voir la loi uniforme ainsi définie comme une généralisation de la loi uniforme standard.

On définit aussi un arbre couvrant orienté en partant d'un arbre couvrant de G et en choisissant une racine $\rho(t)$ et en orientant toutes les arêtes vers la racine. Le choix de la racine donne alors une bijection naturelle entre \mathcal{A}^\rightarrow l'ensemble des arbres couvrants orientés et $\mathcal{A} \times V$.



FIGURE 1 – Un même arbre que l'on a orienté de deux manières, on a représenté la racine par un point noir.

On va maintenant présenter deux algorithmes afin d'échantillonner les arbres couvrants : l'algorithme d'Aldous Broder et l'algorithme de Wilson.

3.1 L'algorithme d'Aldous Broder

Afin d'échantillonner un arbre couvrant, définissons l'actualisation $U(t, e)$ d'un arbre $t \in \mathcal{A}$.

- On prend $e \in E^{\rightarrow}$, en bleu dans la figure, tel que $e^- = \rho(t)$.
- Dans le sous-graphe $t \cup \{e\}$ chaque sommet a exactement une arête partant de lui. Un unique cycle passe donc par la racine. On note d , en rouge ci-dessous, l'unique arête de ce cycle telle que $d^- = e^+$.
- On pose finalement $U(t, e) = t \cup \{e\} \setminus \{d\}$ qui est donc un arbre couvrant de racine e^+ .

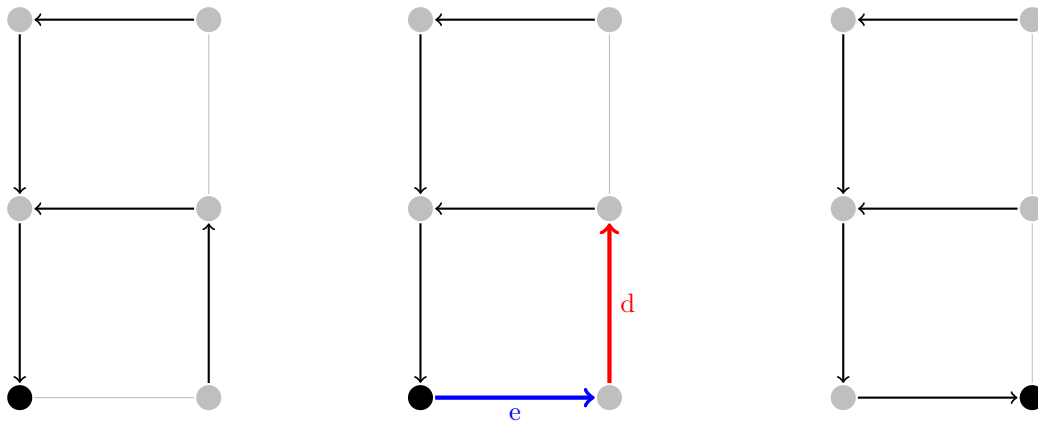


FIGURE 2 – Actualisation d'un arbre couvrant. A gauche, un arbre couvrant sur une grille 2x3. Au milieu, on a surligné les arrêtes d et e. A droite, l'arbre couvrant obtenu après actualisation.

Dans ce qui suit, on notera $X_{n,n+1}$ l'arête traversée par une marche aléatoire X entre le temps n et $n+1$.

Soient $X = (X_n)_{n \geq 0}$ une marche aléatoire sur G et T_0 un arbre couvrant orienté de G de racine X_0 . On pose pour $n \geq 1$, $T_n = U(T_{n-1}, X_{n-1,n})$. On a défini une chaîne de Markov, appelée chaîne UST, sur $\mathcal{A}^{\rightarrow}$ tel que $\rho(T_n) = X_n$ pour $n \geq 0$.

Le théorème suivant relie la chaîne UST aux marches aléatoires sur le graphe et à la loi uniforme sur les arbres couvrants

Théorème 3.2. Soit G un graphe connexe fini et π l'unique mesure invariante de la marche aléatoire sur G . Alors le produit $\mathbf{UST} \otimes \pi$ est une mesure invariante pour la chaîne UST.

La preuve du théorème découle du lemme qui suit.

Lemme 3.3 (Lemme de Kelly). Soit Q une matrice stochastique indexée par un ensemble dénombrable S . Si π est une mesure de probabilité sur S et il existe Q^* , une matrice stochastique indexée par S telle que pour tout u, v dans S , $\pi(u)Q(u, v) = \pi(v)Q^*(v, u)$ alors π est une mesure invariante pour Q .

Preuve. Pour $v \in S$, il suffit d'écrire :

$$\pi Q(v) = \sum_{u \in S} \pi(u)Q(u, v) \stackrel{\text{hyp}}{=} \sum_{u \in S} \pi(v)Q^*(v, u) = \pi(v).$$

Preuve. Démontrons le théorème. Soit Q la matrice de transition associée à la chaîne UST. On cherche à construire l'opération inverse de l'actualisation et une matrice Q^* pour appliquer le lemme de Kelly. Soit $t \in \mathcal{A}^\rightarrow$ un arbre orienté.

- On se donne $e \in E^\rightarrow$, en bleu dans la figure, tel que $e^- = \rho(t)$.
- On ajoute e à t . Le sous-graphe obtenu contient un cycle auquel appartient $\rho(t)$.
- On définit $R(t, e)$ l'arbre obtenu en retirant l'unique arête orientée f , en rouge ci-dessous, pointant vers $\rho(t)$ appartenant à ce cycle.

Notons que $\rho(R(t, e)) = f^+$ et que si $e \in E^\rightarrow$ est tel que $e^- = \rho(t)$ et que d est définie comme dans ce qui précède on a : $R(U(t, e), d) = t$.

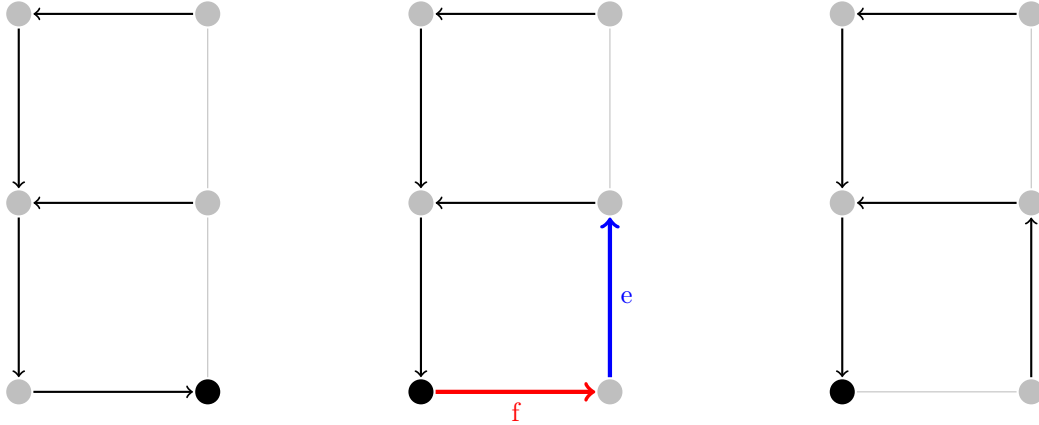


FIGURE 3 – Algorithme inversant l'actualisation de l'arbre de la figure 2

Soit X une marche aléatoire sur G . On définit une matrice stochastique Q^* sur \mathcal{A}^\rightarrow par :

$$Q^*(s, t) := \sum_{e \in E^\rightarrow} \frac{c(e)}{c(\rho(s))} \mathbb{1}(e^- = \rho(s), R(s, e) = t).$$

On remarque que l'on a $Q(s, t) > 0$ si et seulement si il existe une unique arête $e(s, t)$ telle que $U(s, e) = t$. On peut alors trouver d tel que $s = R(t, d)$, ce qui implique $Q(t, s) > 0$. De même, on a $Q^*(t, s) > 0$ si

et seulement si il existe $f(s, t)$, unique, telle que $R(t, f) = s$ et il vient alors $Q(s, t) > 0$. On peut donc écrire pour $s, t \in \mathcal{A}^\rightarrow$:

$$Q(s, t) = \frac{c(e(s, t))}{c(\rho(s))} \mathbb{1}(Q(s, t) > 0) \quad \text{and} \quad Q^*(t, s) = \frac{c(f(s, t))}{c(\rho(t))} \mathbb{1}(Q^*(t, s) > 0).$$

De plus, pour $s, t \in \mathcal{A}^\rightarrow$, si $Q(s, t) > 0$ alors $s = t \cup \{f(s, t)\} \setminus \{e(s, t)\}$. On a donc :

$$\begin{aligned} \left[c(\rho(s)) \prod_{e \in s} c(e) \right] Q(s, t) &= \left[c(\rho(s)) \prod_{e \in s} c(e) \right] \frac{c(e(s, t))}{c(\rho(s))} \mathbb{1}(Q(s, t) > 0) \\ &= \left[c(\rho(t)) \prod_{e \in t} c(e) \right] \frac{c(f(s, t))}{c(\rho(t))} \mathbb{1}(Q^*(t, s) > 0) \\ &= \left[c(\rho(t)) \prod_{e \in t} c(e) \right] Q^*(t, s). \end{aligned}$$

Ainsi Q , Q^* et $\mathbf{UST} \otimes \pi$ vérifient les hypothèses du lemme 3.3. Donc $\mathbf{UST} \otimes \pi$ est une mesure invariante pour Q .

Propriété 3.4. *Supposons G fini et connexe. Alors la chaîne UST est irréductible.*

Cherchons maintenant à échantillonner les arbres couvrants.

Définition 3.5. *Soient G un graphe fini, connexe et $(X_n)_{n \in \mathbb{N}}$ une marche aléatoire sur G . Pour tout sommet $v \in V$, on note $e(v, X)$ l'arête pointant vers v qui est traversée par X la première fois que la marche atteint v . On définit alors :*

$$AB(X) = \{e(v, X)^{\leftarrow}, v \neq X_0\}$$

où e^{\leftarrow} est l'arête d'orientation inverse de e .

On vérifie aisément que $AB(X)$ est presque sûrement un arbre couvrant de racine X_0 .

Le théorème suivant assure que l'algorithme alors défini échantillonne selon la loi uniforme.

Théorème 3.6 (Broder 1989, Aldous 1990). *Soient G un graphe fini, connexe et $(X_n)_{n \in \mathbb{N}}$ une marche aléatoire sur G de point de départ v . Alors $AB(X)$ est distribuée selon la loi uniforme sur les arbres couvrants enracinés en v .*

Preuve. Soit π la mesure de probabilité invariante sur G .

On se donne Y_0 distribué selon π ainsi que $(Y_i)_{i \geq 0}$ et $(Y_{-i})_{i \geq 0}$ des marches aléatoires sur G , de point de départ Y_0 et conditionnellement indépendante selon Y_0 . On appelle $Y = (Y_n)_{n \in \mathbb{Z}}$ marche aléatoire stationnaire bi-infinie sur G . Il découle de la définition de Y et de la propriété de Markov que :

1. Y est stationnaire : $(Y_{n+k})_{n \in \mathbb{Z}}$ est de même loi que $(Y_n)_{n \in \mathbb{Z}}$ pour $k \in \mathbb{Z}$.
2. Y est réversible : $(Y_{-n})_{n \in \mathbb{Z}}$ est de même loi que $(Y_n)_{n \in \mathbb{Z}}$.

Pour $n \in \mathbb{Z}$ et $v \in V$, on définit de plus $l_n(Y, v)$ comme l'arête orientée qui est traversée par Y en partant de v pour la dernière fois avant le temps n . On pose alors :

$$\mathcal{T}_n := \mathcal{T}_n(Y) = \{l_n(Y, v), v \neq Y_n\}.$$

On remarque que \mathcal{T}_n est presque sûrement un arbre couvrant de G enraciné en Y_n . En effet, toutes les arêtes pointent vers Y_n et aucune arête n'est émise de Y_n . De plus, pour $n \in \mathbb{Z}$, on a :

$$\mathcal{T}_n = U(\mathcal{T}_{n-1}, Y_{n-1, n}).$$

En effet :

$$l_{n+1}(Y, v) = \begin{cases} l_n(Y, v) & \text{si } v \neq Y_n \\ Y_{n, n+1} & \text{sinon.} \end{cases}$$

D'où :

$$\mathcal{T}_{n+1} = \mathcal{T}_n \cup \{Y_{n, n+1}\} \setminus \{l_n(Y, Y_{n, n+1}^+)\} = U(\mathcal{T}_n, Y_{n, n+1}).$$

Ainsi, conditionnellement à Y_0 , la distribution de $(\mathcal{T}_n)_{n \geq 0}$ est celle d'une chaîne UST sur G . De plus comme Y est une mesure stationnaire, il vient que $\mathcal{T}_n(Y) = \mathcal{T}_0((Y_{k+n})_{k \in \mathbb{Z}})$ est de même loi que $\mathcal{T}_0(Y)$ pour tout $n \in \mathbb{Z}$.

On en déduit donc que $\mathcal{T}_0(Y)$ est une mesure invariante pour les chaînes UST de G ce qui implique par le théorème 3.2 et la propriété 3.4 que $\mathcal{T}_0(Y)$ est de loi $\mathbf{UST} \otimes \pi$ qui est l'unique mesure invariante pour les chaînes UST.

Remarquons que $\mathcal{T}_0(Y) = AB((Y_{-i})_{i \geq 0})$, donc $AB((Y_{-i})_{i \geq 0})$ est distribué selon $\mathbf{UST} \otimes \pi$. De plus, l'arbre non orienté correspondant à $AB((Y_{-i})_{i \geq 0})$ est indépendant de Y_0 .

Ainsi, conditionnellement à Y_0 , $AB((Y_{-i})_{i \geq 0})$ est distribué selon la loi uniforme sur les arbres de G enracinés en Y_0 , ce qui achève la démonstration.

On démontre à partir du théorème 3.6 la remarquable formule de Kirchhoff qui donne la probabilité qu'une arête soit dans un arbre couvrant uniforme. Elle sera notamment utilisée lorsqu'on étudiera les graphes infinis.

Théorème 3.7 (Formule des résistances effectives de Kirchhoff). *Soient G un graphe fini, connexe, \mathcal{T} un arbre couvrant uniforme de G et $e \in E$ une arête entre u et v . Alors :*

$$\mathbb{P}(e \in \mathcal{T}) = c(e) \mathcal{R}_{\text{eff}}(u \leftrightarrow v)$$

où $\mathcal{R}_{\text{eff}}(u \leftrightarrow v) := \mathcal{C}_{\text{eff}}(u \leftrightarrow v)^{-1}$ (définition 2.5).

Preuve. Par le théorème précédent, on peut considérer \mathcal{T} comme étant généré par l'algorithme d'Aldous-Broder appliqué à une marche aléatoire X de point de départ u . Alors l'unique manière pour que e soit dans $AB(X)$ est que e soit l'arête traversée par X en entrant pour la première fois en v . Ainsi, par la propriété de Markov :

$$\mathbb{P}(e \in \mathcal{T}) = \frac{c(e)}{c(u)} + \mathbb{P}_u(\tau_u^+ < \tau_v) \mathbb{P}(e \in \mathcal{T}).$$

On peut finalement écrire :

$$\mathbb{P}(e \in \mathcal{T}) = \frac{c(e)}{c(u) \mathbb{P}_u(\tau_u^+ > \tau_v)} = c(e) \mathcal{R}_{\text{eff}}(u \leftrightarrow v).$$

3.2 Propriétés spatiales de Markov

On suppose dans cette section que le graphe G est fini et connexe.

On étudie dans cette partie une autre propriété remarquable des arbres couvrants uniformes. On va voir dans cette partie des ressemblance entre des arbres couvrants uniformes et des chaînes de Markov. On étudie en fait un analogue des propriétés de Markov des chaînes de Markov, mais ici pour les arbres couvrants uniformes. L'idée est de dire que demander qu'un certain nombre d'arêtes soient dans un arbre couvrant uniforme ou pas résulte en distribution à un arbre couvrant sur le reste du graphe.

Définition 3.8. *Pour $A, B \subset E$ disjoints, on pose $(G - A)/B$ la différence ensembliste de G par A quotientée par la relation d'équivalence induite par B : c'est le graphe G pour lequel on a contracté toutes les arêtes de B et auquel on a supprimé les arêtes de A .*

Théorème 3.9 (Propriété spatiale de Markov). *Soit \mathcal{T} un arbre couvrant uniforme de G , et soient $A, B \subset E$ des ensembles disjoints d'arêtes telles que l'évènement $\{A \cap \mathcal{T} = \emptyset, B \subset \mathcal{T}\}$ a une probabilité strictement positive. Alors la distribution conditionnelle de \mathcal{T} sur cet évènement coïncide avec celle de l'union de B avec un arbre couvrant uniforme de $(G - A)/B$.*

Preuve. On prouve le théorème dans le cas où $A \cup B$ est un singleton, le cas général s'en déduit par récurrence immédiate. On note \mathcal{A} l'ensemble des arbres sur G .

— Supposons donc pour commencer $A = \{f\}$ et $B = \emptyset$. Alors pour T arbre couvrant de G ,

$$\mathbb{P}(\mathcal{T} = T | f \notin \mathcal{T}) = \mathbb{1}_{f \notin T} \frac{\prod_{e \in T} c(e)}{\mathbb{P}(f \notin \mathcal{T}) \sum_{s \in \mathcal{A}} \prod_{e \in s} c(e)}.$$

Il suffit alors de remarquer qu'un ensemble T_0 d'arêtes de G est un arbre de $G - \{f\}$ si et seulement si c'est un arbre de G qui ne contient pas f , et on déduit alors, en notant \mathcal{A}' l'ensemble des arbres de $G - \{f\}$,

$$\mathbb{P}(f \notin \mathcal{T}) \sum_{s \in \mathcal{A}} \prod_{e \in s} c(e) = \frac{\sum_{s \in \mathcal{A}'} \prod_{e \in s} c(e)}{\sum_{s \in \mathcal{A}} \prod_{e \in s} c(e)} \sum_{s \in \mathcal{A}} \prod_{e \in s} c(e) = \sum_{s \in \mathcal{A}'} \prod_{e \in s} c(e)$$

d'où le résultat annoncé.

— Supposons à présent $A = \emptyset$ et $B = \{f\}$. Alors de même,

$$\mathbb{P}(\mathcal{T} = T | f \in \mathcal{T}) = \mathbb{1}_{f \in T} \frac{\prod_{e \in T} c(e)}{\mathbb{P}(f \in \mathcal{T}) \sum_{s \in \mathcal{A}} \prod_{e \in s} c(e)}$$

et il suffit de remarquer qu'un ensemble d'arêtes de G est l'union de $\{f\}$ avec un arbre de $G/\{f\}$ si et seulement c'est un arbre de G contenant f , et on déduit alors, en notant \mathcal{A}' l'ensemble des arbres de $G/\{f\}$,

$$\mathbb{P}(f \in \mathcal{T}) \sum_{s \in \mathcal{A}} \prod_{e \in s} c(e) = \frac{c(f) \sum_{s \in \mathcal{A}'} \prod_{e \in s} c(e)}{\sum_{s \in \mathcal{A}} \prod_{e \in s} c(e)} \sum_{s \in \mathcal{A}} \prod_{e \in s} c(e) = c(f) \sum_{s \in \mathcal{A}'} \prod_{e \in s} c(e)$$

donc en éliminant les $c(f)$ au numérateur et au dénominateur, on déduit le résultat.

On introduit à présent la notion d'ensemble local pour énoncer une seconde forme de propriété spatiale de Markov, utile dans la suite.

Définition 3.10. *On dit qu'un ensemble aléatoire $\mathcal{K} \subset E$ défini sur le même espace de probabilité que \mathcal{T} est un ensemble local pour \mathcal{T} si pour tout $W \subset E$, l'évènement $\{\mathcal{K} \subset W\}$ est mesurable par rapport à la tribu engendrée par la σ -algèbre générée par la restriction de \mathcal{T} à W .*

Cela signifie que si \mathcal{K} est un ensemble local pour \mathcal{T} , pour tout $W \subset E$, le fait que $\mathcal{K} \subset W$ « ne dépend » que de la restriction de \mathcal{T} à W : si on connaît la restriction de \mathcal{T} à W , alors on sait dire si $\mathcal{K} \subset W$, ou sachant la restriction de \mathcal{T} à W , la probabilité que $\mathcal{K} \subset W$ est soit 0 ou 1. Les ensembles locaux peuvent aussi être vus comme des équivalents des temps d'arrêt : on sait choisir selon la restriction de \mathcal{T} à W si on arrête \mathcal{K} à W .

Remarque 3.11. *Si \mathcal{K} est un ensemble local pour \mathcal{T} , sachant la valeur de \mathcal{T} , la probabilité que \mathcal{K} soit égal à un certain K est soit égale à 0 ou 1. En effet, la probabilité que \mathcal{K} soit incluse dans $W \subset E$ est soit 0 ou 1, et ce pour tout ensemble $W \subset E$.*

Il faut voir la prochaine propriété comme un équivalent pour les arbres couvrants uniformes avec des ensembles locaux de la propriété de Markov forte des chaînes de Markov avec un temps d'arrêt.

Proposition 3.12 (Propriété spatiale de Markov forte). *Soit \mathcal{T} un arbre couvrant uniforme de G , et \mathcal{K} un ensemble local pour \mathcal{T} . On note $\mathcal{K}_o := \mathcal{K} \cap \mathcal{T}$ et $\mathcal{K}_c := \mathcal{K} - \mathcal{K}_o$. Alors la distribution conditionnelle de \mathcal{T} sachant \mathcal{K} et la restriction de \mathcal{T} à \mathcal{K} coïncide avec celle de l'union de \mathcal{K}_o avec un arbre couvrant uniforme indépendant de $(G - \mathcal{K}_c)/\mathcal{K}_o$.*

Preuve. On calcule $\mathbb{P}(\mathcal{T} = T | \mathcal{K} = K, \mathcal{T} \cap \mathcal{K} = K_o)$ pour T, K, K_o tels que $T \cap K = K_o$. On a tout d'abord

$$\mathbb{P}(\mathcal{K} = K, \mathcal{T} \cap \mathcal{K} = K_o) = \mathbb{P}(\mathcal{K} = K, \mathcal{T} \cap K = K_o),$$

mais en faisant le même raisonnement que dans la remarque 3.11, $\mathbb{P}(\mathcal{K} = K | \mathcal{T} \cap K = K_o) \in \{0, 1\}$, donc pour que l'évènement sur lequel on conditionne ne soit pas de probabilité nulle, il faut qu'on ait $\mathbb{P}(\mathcal{K} = K | \mathcal{T} \cap K = K_o) = 1$, ce qu'on suppose donc vrai. On a alors

$$\begin{aligned} \mathbb{P}(\mathcal{T} = T | \mathcal{K} = K, \mathcal{T} \cap \mathcal{K} = K_o) &= \frac{\mathbb{P}(\mathcal{T} = T, \mathcal{K} = K, \mathcal{T} \cap \mathcal{K} = K_o)}{\mathbb{P}(\mathcal{K} = K, \mathcal{T} \cap \mathcal{K} = K_o)} \\ &= \frac{\mathbb{P}(\mathcal{T} = T, \mathcal{K} = K)}{\mathbb{P}(\mathcal{K} = K, \mathcal{T} \cap \mathcal{K} = K_o)} \\ &= \frac{\mathbb{P}(\mathcal{T} = T)\mathbb{P}(\mathcal{K} = K | \mathcal{T} = T)}{\mathbb{P}(\mathcal{K} = K, \mathcal{T} \cap K = K_o)} \\ &= \frac{\mathbb{P}(\mathcal{T} = T)}{\mathbb{P}(\mathcal{K} = K | \mathcal{T} \cap K = K_o)\mathbb{P}(\mathcal{T} \cap K = K_o)} \\ &= \frac{\mathbb{P}(\mathcal{T} = T)}{\mathbb{P}(\mathcal{T} \cap K = K_o)}. \end{aligned}$$

D'où

$$\mathbb{P}(\mathcal{T} = T | \mathcal{K} = K, \mathcal{T} \cap \mathcal{K} = K_o) = \mathbb{P}(\mathcal{T} = T | \mathcal{T} \cap K = K_o) = \mathbb{P}(\mathcal{T} = T | (\mathcal{K} - K_o) \cap \mathcal{T} = \emptyset, K_o \subset \mathcal{T}).$$

D'après la propriété spatiale de Markov (3.9), on déduit donc

$$\mathbb{P}(\mathcal{T} = T | \mathcal{K} = K, \mathcal{T} \cap \mathcal{K} = K_o) = \mathbf{UST}_{(G - K_o)/K_o}(T).$$

Remarque 3.13. Si v_1, \dots, v_k sont des sommets de G , et $S_{\mathcal{T}}(v_1, \dots, v_k)$ le plus petit sous-arbre contenant v_1, \dots, v_k d'un arbre couvrant uniforme \mathcal{T} , alors $S_{\mathcal{T}}(v_1, \dots, v_k)$ est un ensemble local pour \mathcal{T} .

Preuve.

$$\{S_{\mathcal{T}}(v_1, \dots, v_k) \subset W\} = \bigcup_{T_0 \subset W, v_1, \dots, v_k \text{ connectés dans } T_0} \{\mathcal{T} \cap W = T_0\}.$$

3.3 L'algorithme de Wilson

On suppose dans cette section que le graphe G est fini et connexe.

On donne ici un autre algorithme générant un arbre couvrant uniforme, qui a le mérite d'être généralisable lorsqu'on parlera de graphes infinis dans la section suivante.

3.3.1 Procédé d'effacement de cycles

Pour un chemin orienté Γ dans G , on note $\Gamma_{n,n+1}$ pour la n -ième arête orientée traversée par Γ , $\Gamma_n = \Gamma_{n,n+1}^-$, Γ^m la troncature de Γ après la m -ième arête, et finalement, si Γ est fini, Γ^{\leftarrow} le chemin inverse de Γ défini par $(\Gamma^{\leftarrow})_n = \Gamma_{|\Gamma|-n}$ de longueur $|\Gamma|$.

Supposons Γ fini. On définit la séquence de temps suivante : $l_0(\Gamma) := 0$ et $l_{n+1}(\Gamma) := 1 + \max\{t, \Gamma_t = \Gamma_n\}$ tant que $l_n(\Gamma) < |\Gamma|$.

L'effacement des cycles de Γ est le chemin $LE(\Gamma)$ défini par $LE(\Gamma)_i = \Gamma_{l_i(\Gamma)}$ pour i tel que $l_i(\Gamma) \leq |\Gamma|$.

La figure 4 illustre ce procédé.

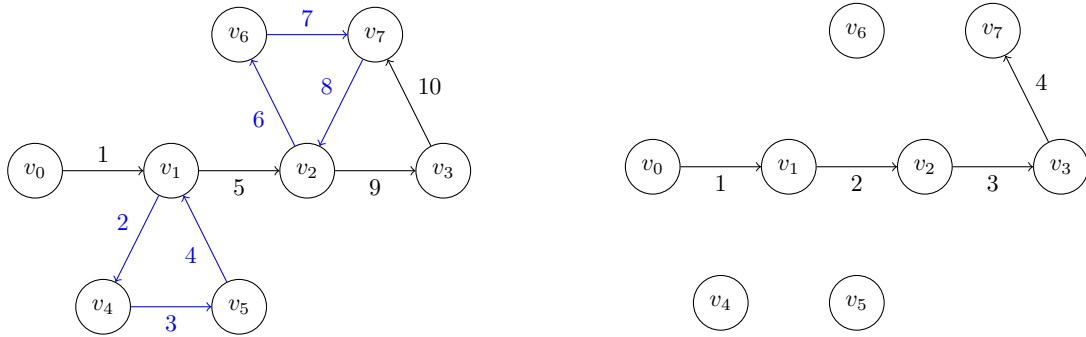


FIGURE 4 – L'effacement des cycles d'une marche. La figure de gauche représente une marche aléatoire dont les arêtes sont indexées par leur ordre respectif de passage, et la figure de droite représente l'effacement des cycles de cette marche. Les arêtes en bleu représentent les cycles qui sont effacés par le procédé LE .

Le théorème suivant peut être vu comme une sorte de lemme que l'on va utiliser pour montrer que le prochain algorithme de Wilson génère un arbre couvrant uniforme.

Théorème 3.14. Soit \mathcal{T} un arbre couvrant uniforme de G , u, v deux sommets de G et Y une marche aléatoire commençant en u et arrêtée lorsqu'elle atteint v . L'unique chemin entre u et v dans \mathcal{T} a la même distribution que $LE(Y)$.

Preuve. 1^{er} point : On commence par montrer que si Γ est un chemin de G tel que $\{\Gamma_n, n \geq 0\} = V$, et $\tau_v = \inf\{n, \Gamma_n = v\}$, alors le chemin entre Γ_0 et v dans $AB(\Gamma)$ a les mêmes arêtes que $LE((\Gamma^{\tau_v})^{\leftarrow})$.

En effet, dans les deux ensembles d'arêtes, l'unique arête e dont l'une de ses deux extrémités est v est clairement la même. Notons w l'autre extrémité. La marche aléatoire arrive dans w pour la première fois grâce à l'arête a , puis fait éventuellement des cycles avant de prendre e . Pour l'ensemble généré par AB , a sera l'unique autre arête dont l'une de ses deux extrémités est w , et pour celui généré par LE , comme le parcours est inversé, toutes les boucles seront éliminées et puisque la marche (non inversé) passait par a avant de traverser w pour la première fois, a sera aussi l'unique autre arête dont l'une de ses deux extrémités est w de l'ensemble généré par LE (cf figure 5). En itérant ce raisonnement, on déduit l'égalité des deux ensembles.

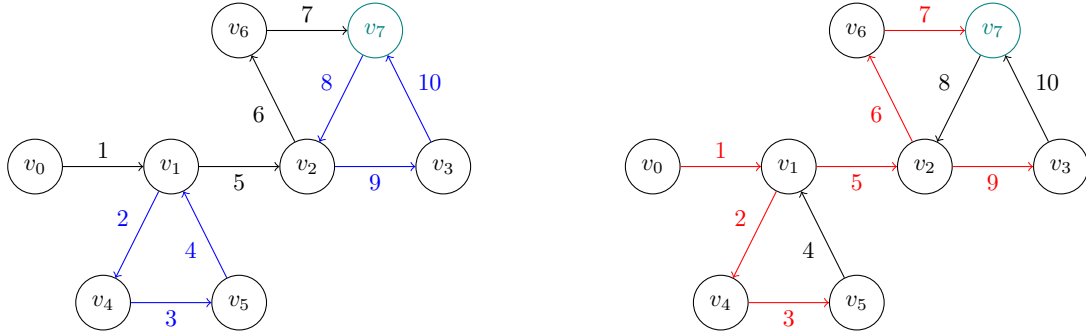


FIGURE 5 – La figure de gauche représente l'effacement des cycles du chemin inverse (les cycles effacés sont en bleu), et la figure de droite représente le procédé AB en rouge appliqué à la même marche. On voit alors ce qu'il se passe : pour v_7 , l'élimination des cycles permet de retrouver l'arête par laquelle la marche aléatoire a pour la première fois visité v_7 .

2^{ème} point : On pose G'_λ le graphe G auquel on ajoute une arête e_0 entre deux de ses sommets u et v de conductance λ . Par propriété de Markov spatiale, la distribution conditionnelle d'un UST de G'_λ sachant qu'il ne contient pas e_0 est celle d'un UST de G . Pour une marche aléatoire X sur G'_λ qui démarre à u , $AB(X)$ contient e_0 si et seulement si X traverse e_0 lorsqu'il entre pour la première fois dans v , et $AB(X)$ est distribué comme un UST de G'_λ donc si Z^λ est une marche aléatoire sur G'_λ qui commence à u et conditionnée à passer par v avant de traverser e_0 , $AB(Z^\lambda)$ est distribuée comme un UST de G . On prend aussi Z une marche aléatoire sur G démarrant en u et conditionnée pour ne pas retourner en u avant de rencontrer v .

On note c_∞ un chemin de G démarrant en u et terminant en v , et qui est sans cycle. Pour t un arbre, on note $\text{ch}(t)$ le chemin de t entre u et v , et on veut montrer que

$$\mathbb{P}(\text{ch}(AB(Y^\lambda)) = c_\infty) \xrightarrow{\lambda \rightarrow \infty} \mathbb{P}(\text{ch}(AB(Y)) = c_\infty)$$

et en déduire que le chemin entre u et v dans Y est distribué comme le chemin de u à v dans un UST de G . L'idée est que de revenir en u sans passer par e_0 après va « coûter » trop cher à la marche lorsque λ est grand.

On note $\mathcal{A}_n(T_0)$ l'ensemble des chemins c de longueur n sur G démarrant en u , tels que v n'apparait que

en étant le dernier sommet rencontré par c et le chemin de u à v dans $AB(c)$ est égal à c_∞ . On a alors

$$\mathbb{P}(\text{ch}(AB(Z^\lambda)) = c_\infty) = \sum_{n \in \mathbb{N}} \sum_{c \in \mathcal{A}_n} \mathbb{P}((Z^\lambda)^n = c).$$

Soit X^λ une marche aléatoire sur G'_λ commençant en u et X une marche aléatoire sur G commençant aussi en u , et $e_0(X^\lambda)$ l'évènement X passe par v avant de traverser e_0 , alors pour c un chemin de longueur $n + 2$ et u, c_1, \dots, c_n, v ses sommets, on a

$$\mathbb{P}((Z^\lambda)^{n+2} = c) = \frac{\mathbb{P}(X_1 = c_1, \dots, X_n = c_n, X_{n+1} = v, e_0(X^\lambda))}{\mathbb{P}(e_0(X^\lambda))}$$

qui est nul si un c_i vaut u et est suivi d'un v , ce qu'on suppose qu'il n'arrive donc pas. Donc

$$\mathbb{P}((Z^\lambda)^{n+2} = c) = \frac{c(u, c_1) \dots c(c_n, v)}{c(u|G'_\lambda) \dots c(c_n) \mathbb{P}(e_0(X^\lambda))}$$

où $c(a, b)$ est la conductance entre les sommets a et b , nulle s'il n'y a pas d'arête, et $c(u|G'_\lambda)$ la conductance de u dans G'_λ .

Donc

$$\mathbb{P}((Z^\lambda)^{n+2} = c) = \left(\frac{c(u|G)}{c(u|G'_\lambda)} \right)^l \frac{c(u, c_1) \dots c(c_n, v)}{c(u|G) \dots c(c_n) \mathbb{P}(e_0(X^\lambda))}$$

avec l le nombre de fois où c rencontre u ,

$$= \left(\frac{c(u|G)}{\lambda + \sum_{w \sim u, w \in G} c(u, w)} \right)^l \frac{\mathbb{P}(X^{n+2} = c)}{\mathbb{P}(e_0(X^\lambda))}.$$

Mais si on note de même $\mathcal{B}_n(T_0)$ l'ensemble des chemins b de longueur n sur G commençant en u tels que b ne rencontre v qu'en son dernier sommet visité, alors

$$\mathbb{P}(e_0(X^\lambda)) = \sum_{n \in \mathbb{N}} \sum_{b \in \mathcal{B}_n} \mathbb{P}((X^\lambda)^n = b)$$

et il y a alors deux cas : soit b visite u plus de 2 fois, alors $\mathbb{P}((X^\lambda)^n = b) = \frac{C}{(\lambda+c)^l}$ avec $l \geq 2$, et C, c constants en λ (mais dépendant de b), sinon $\mathbb{P}((X^\lambda)^n = b) = \frac{C}{\lambda+c}$, donc par théorème de convergence dominée (par la même expression avec $\lambda = 1$), on déduit

$$\mathbb{P}(e_0(X^\lambda)) \underset{\lambda \rightarrow \infty}{\sim} \frac{\mathbb{P}(\tau_u^+(X) > \tau_v(X)) c(u|G)}{\lambda}$$

et finalement

$$\begin{aligned} \mathbb{P}(\text{ch}(AB(Z^\lambda)) = c_\infty) &= \frac{\sum_{n \in \mathbb{N}} \sum_{c \in \mathcal{A}_n} \left(\frac{c(u|G)}{\lambda+c(u|G)} \right)^{l(c)} \mathbb{P}(X^n = c)}{\mathbb{P}(e_0(X^\lambda))} \\ &= \frac{\frac{c(u|G)}{\lambda+c(u|G)}}{\mathbb{P}(e_0(X^\lambda))} \sum_{n \in \mathbb{N}} \sum_{c \in \mathcal{A}_n} \left(\frac{c(u|G)}{\lambda+c(u|G)} \right)^{l(c)-1} \mathbb{P}(X^n = c) \end{aligned}$$

et un théorème de convergence dominé appliqué comme précédemment élimine les chemins retournant en u avant de retourner en v , donc

$$\sum_{n \in \mathbb{N}} \sum_{c \in \mathcal{A}_n} \left(\frac{c(u|G)}{\lambda + c(u|G)} \right)^{l(c)-1} \mathbb{P}(X^n = c) \xrightarrow{n \rightarrow \infty} \mathbb{P}(\tau_u^+(X) > \tau_v(X))$$

et l'équivalent de $\mathbb{P}(e_0(X^\lambda))$ permet de conclure

$$\mathbb{P}(\text{ch}(AB(Z^\lambda)) = c_\infty) \xrightarrow{\lambda \rightarrow \infty} \mathbb{P}(\text{ch}(AB(Z)) = c_\infty)$$

et puisque le terme de gauche est constant, on déduit que $\text{ch}(AB(Z))$ est distribué comme le chemin entre u et v d'un UST de G .

3^{ème} point : Pour finir, on montre que la marche aléatoire commençant en u , arrêtée lorsqu'elle atteint v pour la première fois et conditionnée pour atteindre v avant de retourner en u a la même distribution que la marche aléatoire commençant en v , arrêtée lorsqu'elle atteint u pour la première fois et conditionnée pour atteindre u avant de retourner en v , inversée en temps.

En effet, pour $c := (u, c_1, \dots, c_n, v)$ un chemin ne retournant pas en u , et pour X et X' des marches aléatoires sur G qui commencent respectivement en u et v , alors

$$\frac{\mathbb{P}(X^{n+2} = c)}{\mathbb{P}(\tau_u^+(X) > \tau_v(X))} = \frac{c(u, c_1) \dots c(c_n, v)}{\mathbb{P}(\tau_u^+(X) > \tau_v(X)) c(u) \dots c(c_n)} = \frac{c(u, c_1) \dots c(c_n, v)}{\mathcal{C}(u \leftrightarrow v) c(c_1) \dots c(c_n)}.$$

Mais la conductance effective ne dépend pas de l'ordre dans lequel on met les 2 ensembles d'après le principe de Dirichlet (Théorème 2.11), donc

$$\frac{\mathbb{P}(X^{n+2} = c)}{\mathbb{P}(\tau_u^+(X) > \tau_v(X))} = \frac{c(v, c_n) \dots c(c_1, u)}{\mathbb{P}(\tau_v^+(X') > \tau_u(X')) c(v) c(c_n) \dots c(c_1)} = \frac{\mathbb{P}((X'^{n+2})^\leftarrow = c)}{\mathbb{P}(\tau_v^+(X') > \tau_u(X'))}$$

ce qui conclut.

Résumons ce qu'on vient de faire : on a d'abord montré que le chemin entre u et v dans $AB(X)$ est $LE((X^{\tau_v})^\leftarrow)$. Puis que si Y est un marche aléatoire démarrant en u et conditionnée pour atteindre v avant de revenir en u , alors le chemin de $AB(Y)$ entre u et v est distribué comme le chemin entre u et v d'un UST, ce qui a permis de conclure avec le premier point que $LE((Y^{\tau_v})^\leftarrow)$ est distribué comme le chemin entre u et v d'un UST, et le dernier point montre que $(Y^{\tau_v})^\leftarrow$ est distribué comme une marche aléatoire Z démarrant en v arrêtée quand elle arrive en u et conditionnée pour toucher u avant de revenir en v , donc $LE(Z)$ est distribué comme le chemin entre u et v d'un UST, qui est le chemin entre v et u d'un UST, on donc montré le théorème.

3.3.2 Algorithme de Wilson

On définit finalement l'algorithme de Wilson.

On énumère les sommets de G $\{v_0, \dots, v_n\}$. On construit une suite croissante de sous-arbres de G de la manière suivante :

- Soit \mathcal{T}_0 l'arbre de sommet v_0 et sans arête.
- Pour $1 \leq i \leq n$, en supposant \mathcal{T}_{i-1} construit, on prend X^i une marche aléatoire sur G partant de v_i et arrêtée lorsqu'elle rencontre un sommet de \mathcal{T}_{i-1} , on pose alors \mathcal{T}_i comme l'union de \mathcal{T}_{i-1} avec $LE(X^i)$.

L'arbre final \mathcal{T}_n est un arbre de G , le prochain théorème de Wilson affirme même que celui-ci est distribué comme un arbre couvrant uniforme de G . La figure 6 illustre une étape de cette construction.

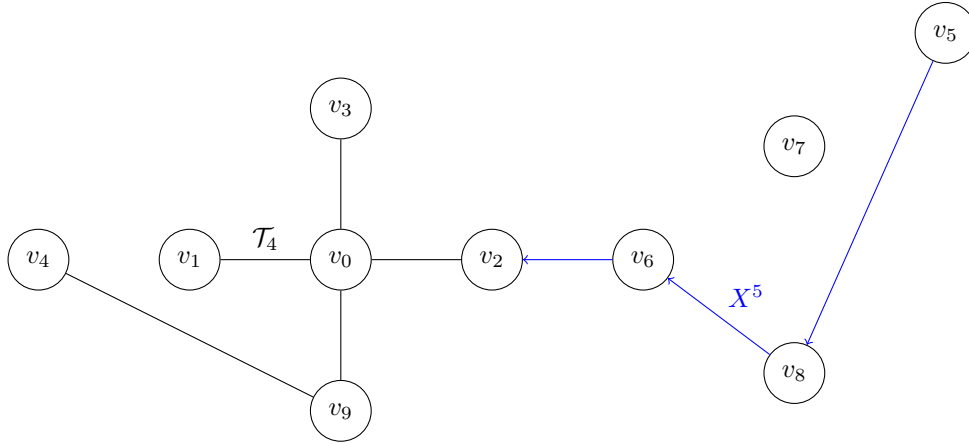


FIGURE 6 – Algorithme de Wilson

Théorème 3.15 (Algorithme de Wilson). *L'arbre aléatoire \mathcal{T}_n obtenu par le précédent algorithme de Wilson est distribué comme un arbre couvrant uniforme de G .*

Preuve. On montre par récurrence que \mathcal{T}_i a la même distribution que $S_{\mathcal{T}}(v_0, \dots, v_i)$ (cf remarque 3.13) où \mathcal{T} est un arbre couvrant uniforme de G .

L'initialisation est vraie.

Pour \mathcal{T} un UST de G , et pour $i \geq 0$, on pose $\mathcal{T}'_i = S_{\mathcal{T}}(v_0, \dots, v_i)$. Le sous-arbre \mathcal{T}'_i est distribué comme \mathcal{T}_i et est un ensemble local d'après la remarque 3.13 pour \mathcal{T} donc d'après la propriété de Markov spatiale forte (Proposition 3.12), la distribution conditionnelle de \mathcal{T} sachant \mathcal{T}'_i est égale à l'union de \mathcal{T}'_i avec un arbre couvrant uniforme de G/\mathcal{T}'_i . Le théorème 3.14 affirme alors que le chemin liant v_{i+1} aux sommets de \mathcal{T}'_i dans \mathcal{T} a la même distribution que l'effacement des cycles d'une marche aléatoire sur G/\mathcal{T}'_i stoppée lorsqu'elle atteint le sommet représentant \mathcal{T}'_i . On déduit que $S_{\mathcal{T}}(v_0, \dots, v_{i+1})$ a la même distribution que \mathcal{T}_{i+1} , ce qui conclut la récurrence.

Le théorème se déduit du résultat de la récurrence pour $i = n$.

4 Cas des graphes infinis

On travaille à présent sur des graphes infinis : on suppose G infini, dénombrable, localement fini et connexe dans cette partie.

On veut développer la même sorte d'objets que ceux développés dans la section précédente mais à présent sur des graphes infinis. Le problème est alors qu'on ne peut plus définir a priori de loi canonique uniforme sur les arbres couvrants puisqu'il y a a priori une infinité de tels arbres sous des hypothèses raisonnables et possiblement même un nombre indénombrable. On définira alors une mesure en prenant en quelque sorte la limite sur une exhaustion de G de **USF**, et on verra 2 manières relativement naturelles de le faire. Mais l'objet issu de la loi créée ne sera pas forcément un arbre, mais une forêt.

On peut déjà mentionner que des problèmes de mesurabilité semblent se poser, puisque si on définit une mesure sur l'ensembles des sous-ensembles de E , comme celui-ci est indénombrable, la mesure n'est par caractérisée par ses valeurs en les atomes.

4.1 Forêts couvrantes uniformes sur des graphes infinis

Définition 4.1. *On appelle forêt sur un graphe G un sous-ensemble de E ne contenant pas de cycle.*

On peut donc voir une forêt comme une union d'arbres, ce qui donne le sens du nom.

Définition 4.2. *On appelle ici une exhaustion de G un suite croissante (V_n) d'ensembles de sommets de G tels que $\bigcup_{n \in \mathbb{N}} V_n = V$ et tel que V_n est connexe.*

Pour une telle exhaustion (V_n) , on définit le graphe G_n de sommets V_n et contenant les arêtes de G reliant 2 points de V_n . Autrement dit, on peut écrire $E_n = E \cap (V_n \times V_n)$.

On définit aussi le graphe G_n^ comme le graphe obtenu en regroupant tous les sommets de $V - V_n$ en un seul sommet δ_n , contenant les arêtes de G reliant 2 points de V_n et pour tout point v de V_n une arête reliant v à δ_n de conductance la somme des conductances des arêtes entre v et un point de $V - V_n$.*

On cherche maintenant à définir une mesure sur l'ensemble des forêts. Une manière qui peut sembler assez naturelle est de se munir d'une exhaustion (V_n) de G , donc pour $A, B \subset E$ finis, alors $A, B \subset V_n$ à partir d'un certain rang, et on veut alors poser

$$\mathbf{FUSF}(\{\omega \subset E, A \cap \omega = \emptyset, B \subset \omega\}) := \lim_{n \rightarrow \infty} \mathbf{UST}_{G_n}(\{t, A \cap t = \emptyset, B \subset t\})$$

et

$$\mathbf{WUSF}(\{\omega \subset E, A \cap \omega = \emptyset, B \subset \omega\}) := \lim_{n \rightarrow \infty} \mathbf{UST}_{G_n^*}(\{t, A \cap t = \emptyset, B \subset t\})$$

ce qui définirait deux uniques mesures **FUSF** et **WUSF** sur la tribu engendrée par les évènements $\{\omega \subset E, A \cap \omega = \emptyset, B \subset \omega\}$ pour $A, B \subset E$ finis, d'après le théorème d'Hahn-Kolmogorov.

Le théorème 4.3 et la proposition 4.5 suivants assurent que cette construction marche.

Théorème 4.3 (Définition de forêts couvrantes uniformes). *Pour (V_n) une exhaustion de G , les limites faibles $\mathbf{FSUF} := \lim_{n \rightarrow \infty} \mathbf{UST}_{G_n}$ et $\mathbf{WSUF} := \lim_{n \rightarrow \infty} \mathbf{UST}_{G_n^*}$ sont bien définies et ne dépendent pas de l'exhaustion choisie.*

Preuve. Il ne reste donc plus qu'à montrer que les limites $\lim_{n \rightarrow \infty} \mathbf{UST}_{G_n^*}(\{t, A \cap t = \emptyset, B \subset t\})$ et $\lim_{n \rightarrow \infty} \mathbf{UST}_{G_n}(\{t, A \cap t = \emptyset, B \subset t\})$ existent et ne dépendent pas de l'exhaustion choisie.

- Montrons pour commencer par récurrence sur $k \in \mathbb{N}$ que si $|B| \leq k$, alors $\lim_{n \rightarrow \infty} \mathbf{UST}_{G_n}(\{t, B \subset t\})$ est bien défini et ne dépend pas de l'exhaustion.

Supposons donc le résultat vrai pour $k \in \mathbb{N}$, et soit $B \subset E$ de cardinal $k + 1$. On prend donc $e \in B$ d'extrémités u et v et on pose $B' := B - \{e\}$. Alors pour n assez grand, d'après la propriété spatiale de Markov (3.9),

$$\mathbf{UST}_{G_n}(\{t, B \subset t\}) = \mathbf{UST}_{G_n/B'}(\{t, \{e\} \subset t\}) \mathbf{UST}_{G_n}(\{t, B' \subset t\}),$$

donc d'après la formule des résistances effectives de Kirchhoff (3.7),

$$\mathbf{UST}_{G_n}(\{t, B \subset t\}) = c(e)\mathcal{R}_{\text{eff}}(u \leftrightarrow v, G_n/B')\mathbf{UST}_{G_n}(\{t, B' \subset t\}).$$

Mais d'après le principe de Rayleigh (2.12), $\mathcal{R}_{\text{eff}}(u \leftrightarrow v, G_n/B') = \frac{1}{\mathcal{C}_{\text{eff}}(u \leftrightarrow v, G_n/B')}$ décroît lorsque n grandit, donc converge vers un réel positif qu'on note $\mathcal{R}_{\text{eff}}^F(u \leftrightarrow v, G/B', (V_n))$, on déduit alors

$$\mathbf{UST}_{G_n}(\{t, B \subset t\}) \xrightarrow[n \rightarrow \infty]{} c(e)\mathcal{R}_{\text{eff}}^F(u \leftrightarrow v, G/B', (V_n)) \lim_{n \rightarrow \infty} \mathbf{UST}_{G_n}(\{t, B' \subset t\}).$$

Il suffit donc de montrer que $\mathcal{R}_{\text{eff}}^F(u \leftrightarrow v, G/B', (V_n))$ ne dépend pas de l'exhaustion choisie : on prend deux exhaustions (V_n) et (V'_n) et on crée une nouvelle exhaustion qui mélange les deux : on pose $W_0 := V_0$ et pour n impair $W_n := V'_m$ où $m := \inf\{l \geq n, W_{n-1} \subset V'_l\}$ et pour n pair $W_n := V_m$ où $m := \inf\{l \geq n, W_{n-1} \subset V_l\}$. (W_n) est une exhaustion de G donc $\mathcal{R}_{\text{eff}}^F(u \leftrightarrow v, G)$ converge, mais par construction elle converge donc à la fois vers $\mathcal{R}_{\text{eff}}^F(u \leftrightarrow v, G, (V_n))$ et $\mathcal{R}_{\text{eff}}^F(u \leftrightarrow v, G, (V'_n))$, donc les deux sont égales, ce qui conclut la récurrence.

- Pour G_n^* , le seul problème vient de l'application de principe de Rayleigh qui ne s'applique pas à première vue. En réalité, on peut faire la même démonstration que celle du principe de Rayleigh : si on note F la tension entre u et v dans G'_n , et qu'on note F_0 égale à F sur V_n , à $F(\delta_n)$ sur $V_{n+1} - V_n$ et sur δ_{n+1} , alors on a par principe de Dirichlet (2.11)

$$\mathcal{C}_{\text{eff}}(u \leftrightarrow v, G_{n+1}) \leq \|F_0\| = \|F\| = \mathcal{C}_{\text{eff}}(u \leftrightarrow v, G_n).$$

Donc on déduit $\mathcal{R}_{\text{eff}}(u \leftrightarrow v, G_n) \leq \mathcal{R}_{\text{eff}}(u \leftrightarrow v, G_{n+1})$: la résistance effective croît et la récurrence précédente marche donc aussi pour G_n^* .

- Finalement, on conclut par formule d'inclusion-exclusion : pour $A, B \subset G$ finis, et n suffisamment grand,

$$\mathbf{UST}_{G_n}(\{t, A \cap t = \emptyset, B \subset t\}) = \sum_{S \subset A} (-1)^{|S|} \mathbf{UST}_{G_n}(\{t, B \cup S \subset t\}).$$

Définition 4.4. Une variable aléatoire ayant pour loi \mathbf{WUSF}_G est alors appelé une forêt couvrante uniforme connectée, et une ayant pour loi \mathbf{FUSF}_G est appelé une forêt uniforme libre.

Un objet ayant une des deux telles lois n'est pas forcément connexe presque sûrement : dans les exhaustions, deux points sont connectés mais il est possible qu'en passant à la limite le chemin entre ces deux points soit de longueur tendant vers l'infini, dans ce cas à la limite les 2 points ne sont plus connectés. C'est donc pour cela qu'on parle de forêt et plus d'arbre.

On peut aussi trouver une forme d'explication aux qualifications « libre » et « connecté » : dans le premier cas, les « bords » de G_n sont simplement enlevés alors que dans le second cas ils sont de force connectés à un sommet.

Proposition 4.5. Les 2 mesures de probabilités \mathbf{FUSF} et \mathbf{WUSF} sont concentrées sur l'ensemble des forêts couvrantes de G pour lesquelles chaque composante connexe est infinie.

Preuve. Montrons d'abord qu'elles sont concentrées sur l'ensemble des forêts : si $A \subset E$ est un ensemble fini d'arêtes contenant un cycle, alors pour une exhaustion (V_n) et n suffisamment grand, on a $\mathbf{UST}_{G_n}(\{t, A \subset t\}) = 0$ donc $\mathbf{FUSF}_G(\{t, A \subset t\}) = 0$. Mais l'ensemble $\mathcal{A} := \{A \subset E, A \text{ contient un cycle}\}$ est dénombrable : $\mathcal{A} = \bigcup_{n \in \mathbb{N}} \mathcal{A}_n$ où $\mathcal{A}_n = \{A, A \text{ contient un cycle, } |A| = n\}$ qui est dénombrable car partie de E^n . Donc $\mathbf{FUSF}_G(\mathcal{A}) = 0$: \mathbf{FUSF}_G est concentré sur les forêts de G .

Pour $u \in V$, on dit que $A \subset E$ sépare u de l'infini si A est fini et tout chemin infini partant de u qui ne passe par aucune arête de A rencontre un nombre fini d'arêtes (possiblement plusieurs fois), et on note $\text{cut}(u)$ l'ensemble de ces ensembles B . Alors pour $A \in \text{cut}(u)$, pour une exhaustion (V_n) et n assez grand, un ensemble $t \subset E_n$ vérifiant $t \cap A = \emptyset$ n'est pas connexe donc $\mathbf{FUSF}_G(\{t, t \cap A\}) = 0$ et on montre comme précédemment que $\text{cut}(u)$ est dénombrable donc

$$\mathbf{FUSF}_G(t, \text{la composante connexe de } u \text{ est finie}) = \mathbf{FUSF}_G\left(\bigcup_{A \in \text{cut}(u)} \{t, t \cap A = \emptyset\}\right) = 0$$

et puisque l'ensemble des sommets est dénombrable on déduit

$$\mathbf{FUSF}_G(\{t, t \text{ contient une composante connexe finie}\}) = \mathbf{FUSF}_G\left(\bigcup_{v \in V} \{t, \text{la composante connexe de } v \text{ est finie}\}\right) = 0.$$

4.2 Algorithme de Wilson

On étudie dans cette partie l'algorithme de Wilson déjà étudié sur les graphes finis, mais maintenant sur les graphes infinis. On verra qu'il donne un objet ayant pour loi **WUSF**. On introduit d'abord quelques définitions et propriétés qui permettent d'énoncer l'algorithme de Wilson.

Définition 4.6. G est dit récurrent en $v \in V$ si une marche aléatoire démarrant en v revient presque sûrement en son point d'origine.

Propriétés 4.7. 1. La notion de récurrence ne dépend pas du point de départ. On dit alors qu'un graphe est récurrent s'il est récurrent en l'un de ses points, transient sinon.

2. Pour v un sommet d'un graphe transient, une marche aléatoire sur ce graphe rencontre seulement un nombre fini de fois v presque sûrement.
3. Une marche aléatoire sur un graphe récurrent rencontre presque sûrement tous les sommets du graphe
4. Pour $d \geq 1$, le graphe \mathbb{Z}^d est récurrent si et seulement si $d \leq 2$ et est transient sinon.

On pourra trouver une preuve de ces propriétés dans [Gal06], de la 1ère et de la 3ème dans le lemme 13.4.3, de la 2ème dans la proposition 13.4.1, et de la 4ème propriété dans l'exemple avant le lemme 13.4.3.

On peut alors adapter l'algorithme de Wilson à des graphes infinis.

Énumérons les sommets de G $\{v_1, \dots, v_n, \dots\}$. On construit une suite croissante (\mathcal{F}_n) de forêts de la manière suivante :

- Si G est récurrent, on pose \mathcal{F}_0 la forêt avec sommet $\{v_1\}$ et aucune arête. S'il est transient, on pose \mathcal{F}_0 la forêt sans sommet et sans arête.
- Supposons \mathcal{F}_{n-1} donné, on prend alors X^n une marche aléatoire commençant en v_n et arrêtée lorsqu'elle rencontre un sommet de \mathcal{F}_{n-1} , et continuant à l'infini si elle ne rencontre aucun sommet de \mathcal{F}_{n-1} . On pose alors \mathcal{F}_n la forêt union de \mathcal{F}_{n-1} et de l'effacement des cycles de X^n .

Finalement, on pose $\mathcal{F} := \bigcup_{n \in \mathbb{N}} \mathcal{F}_n$.

Le prochain théorème est remarquable et donne une meilleure perspective sur les forêts couvrantes uniformes connectées : la loi de \mathcal{F} est **WUSF**. On comprend alors beaucoup mieux ce que sont les forêts couvrantes uniformes connectées. A contrario on ne dispose pas de tel algorithme pour caractériser les

forêts couvrantes uniformes libres et ces dernières sont globalement moins bien comprises que les forêts couvrantes uniformes connectées.

Théorème 4.8 (Algorithme de Wilson). *La forêt aléatoire \mathcal{F} qu'on vient juste de définir a loi **WUSF**.*

Preuve. On commence par le cas où G est transient. On se munit d'une exhaustion (V_n) de G . Soit $B \subset E$ fini et soit m tel que v_1, \dots, v_m contienne toutes les extrémités des arêtes de B . Alors pour n assez grand, on applique l'algorithme de Wilson à G_n^* sur les sommets $\delta_n, v_1, \dots, v_m$ et on a alors d'après l'algorithme de Wilson dans les graphes finis (3.15)

$$\mathbf{UST}_{G_n^*}(\{t, B \subset t\}) = \mathbb{P}(B \subset \text{Wilson}(\delta_n, v_1, \dots, v_m))$$

où $\mathbb{P}(B \subset \text{Wilson}(\delta_n, v_1, \dots, v_m))$ est la probabilité que B soit inclus dans la forêt créée par l'algorithme de Wilson arrêté après l'étape v_m .

On prend donc X^1, \dots, X^m des marches aléatoires sur G vues comme marches aléatoires sur G_n^* jusqu'au moment où elles touchent un point de G représenté par δ_n dans G_n^* , commençant en v_1, \dots, v_m , \mathcal{T}_i^n l'arbre créée à l'étape i de l'algorithme de Wilson sur G_n^* , comme dans sa construction, et on note $\tau_i^{(n)} = \inf\{l, X_l^1 \in \mathcal{T}_{i-1}^n\}$, et finalement $\tau_i = \inf\{l, X_l^1 \in \mathcal{F}_{i-1}\}$ où \mathcal{F}_{i-1} est la forêt de l'algorithme de Wilson dans G . Donc

$$\mathbf{UST}_{G_n^*}(\{t, B \subset t\}) = \mathbb{P}\left(B \subset \bigcup_{i=1}^m LE((X^i)^{\tau_i^{(n)}})\right).$$

Mais puisque G est transient, il existe presque sûrement un certain $N = N(X^1, \dots, X^n)$, tel que pour $n \geq N$, $B \subset \bigcup_{i=1}^m LE((X^i)^{\tau_i^{(n)}}) \Leftrightarrow B \subset \bigcup_{i=1}^m LE((X^i)^{\tau_i})$, donc en posant M le plus petit de ces entiers N ,

on a $\mathbb{P}\left(\bigcap_{n=1}^{\infty} \{M \geq n\}\right) = 0$ donc par théorème de continuité décroissante, $\lim_{n \rightarrow \infty} \mathbb{P}(M \geq n) = 0$ et

$$\begin{aligned} \mathbf{UST}_{G_n^*}(\{t, B \subset t\}) &= \mathbb{P}\left(B \subset \bigcup_{i=1}^m LE((X^i)^{\tau_i^{(n)}}) \mid M \leq n\right) \cdot \mathbb{P}(M \leq n) + \\ &\quad \mathbb{P}\left(B \subset \bigcup_{i=1}^m LE((X^i)^{\tau_i^{(n)}}) \mid M > n\right) \cdot \mathbb{P}(M > n), \end{aligned}$$

$$\begin{aligned} \mathbf{UST}_{G_n^*}(\{t, B \subset t\}) &= \mathbb{P}\left(B \subset \bigcup_{i=1}^m LE((X^i)^{\tau_i})\right) \cdot \mathbb{P}(M \leq n) + \mathbb{P}\left(B \subset \bigcup_{i=1}^m LE((X^i)^{\tau_i^{(n)}}) \mid M > n\right) \cdot \mathbb{P}(M > n) \\ &\xrightarrow{n \rightarrow \infty} \mathbb{P}\left(B \subset \bigcup_{i=1}^m LE((X^i)^{\tau_i})\right). \end{aligned}$$

Donc

$$\mathbb{P}\left(B \subset \bigcup_{i=1}^m LE((X^i)^{\tau_i})\right) = \mathbf{WUSF}(\{t, B \subset t\})$$

ce qui conclut le cas transient.

Dans le cas récurrent, on raisonne similairement : on prend v_1, \dots, v_m similairement, on applique l'algorithme de Wilson sur G_n^* aux sommets v_1, \dots, v_m (en commençant en v_1 et pas en δ_n), et on adopte les mêmes autres notations jusque τ_i .

Puisque G est récurrent, d'après la 3eme propriété 4.7, il existe presque sûrement un certain $N(X^1, \dots, X^n)$, tel que pour $n \geq N$, $\bigcup_{i=1}^m LE((X^i)^{\tau_i^{(n)}}) = \bigcup_{i=1}^m LE((X^i)^{\tau_i})$, donc en posant M le plus petit de ces entiers N , on a $\mathbb{P}\left(\bigcap_{n=1}^{\infty} \{M \geq n\}\right) = 0$ donc par théorème de continuité décroissante, $\lim_{n \rightarrow \infty} \mathbb{P}(M \geq n) = 0$ et on conclut de la même manière.

Remarque 4.9. *La preuve précédente montre dans le cas où G est récurrent que $\mathbf{WUSF}_G = \mathbf{FUSF}_G$. En revanche, en général, ce n'est pas le cas : si G est un arbre transient, par construction, la mesure \mathbf{FUSF}_G est un dirac en E , mais d'après l'algorithme de Wilson, et puisque l'arbre est transient, un objet ayant pour loi \mathbf{WUSF}_G a une probabilité strictement positive de ne pas être connexe puisque la deuxième marche a une probabilité strictement positive de ne pas rencontrer la première. Donc dans un tel cas, les 2 mesures ne coïncident pas.*

4.3 Théorème de Pemantle

L'objectif de cette partie est d'étudier la connexité des forêts couvrantes uniformes de \mathbb{Z}^d et d'établir le théorème suivant.

Théorème 4.10 (Pemantle, 1991). *La forêt uniforme de \mathbb{Z}^d est presque sûrement connexe si et seulement si $d \leq 4$.*

Remarque 4.11. *Cet énoncé ne fait pas la différence entre les lois \mathbf{WUSF} et \mathbf{FUSF} car elles coïncident sur \mathbb{Z}^d . On ne démontrera pas ce résultat ici et on prouvera le théorème 4.10 pour la loi \mathbf{WUSF} .*

La preuve originale de Pemantle dans [Pem91] (Théorème (4.2)) n'utilise pas l'algorithme de Wilson qui n'était pas disponible à Pemantle à l'époque contrairement à la preuve qu'on propose ici qui est due à Lyons et Peres ([LP16], Proposition (10.20)).

Dans la suite de cette partie, on démontre le théorème pour les cas $d \leq 2$ et $d \geq 5$. Dans un premier temps on se penche sur le cas $d \leq 2$ puis on étudie le cas $d \geq 5$.

Pour $d \leq 2$, le théorème 4.10 découle de la propriété suivante.

Proposition 4.12. *Soit G un graphe infini, connexe, localement fini et récurrent. Alors la forêt couvrante uniforme de G est presque sûrement connexe.*

L'énoncé est bien posé car sous ces hypothèses et par la remarque 4.9 on a bien que les lois \mathbf{WUSF} et \mathbf{FUSF} coïncident.

Preuve. Comme G est récurrent, il vient que dans l'algorithme de Wilson, pour tout n de \mathbb{N} , X^n rencontre presque sûrement F_{n-1} . On en déduit donc que la forêt tirée par l'algorithme est presque sûrement connexe ce qui implique par le théorème 4.8 que les forêts couvrantes uniformes sont presque sûrement connexe.

Démontrons le théorème 4.10 pour $d \geq 5$. Pour ce faire, on établit d'abord le théorème suivant.

Théorème 4.13 (Erdős et Taylor, 1963). *Soient $d \geq 1$ et X et Y deux marches aléatoires indépendantes sur \mathbb{Z}^d partant de l'origine. Alors $|\{X_n, n \geq 0\} \cap \{Y_n, n \geq 0\}|$ est presque sûrement infini si $d \leq 4$ et presque sûrement fini si $d \geq 5$.*

Preuve. Pour $d \leq 2$, comme \mathbb{Z}^d est récurrent, il vient que pour tout $n \in \mathbb{N}$, il existe presque sûrement $m \in \mathbb{N}$ tel que $X_n = Y_m$, ce qui implique le théorème. On suppose maintenant $d \geq 3$.

On écrit \asymp , \succeq et \preceq pour les égalités et inégalités valables à multiplication par une constante positive près indépendante de n .

Pour $x, y \in G$ et $n \in \mathbb{N}$, on pose $p_n(x, y)$ la probabilité qu'une marche aléatoire sur G commençant en x soit en y au temps n . Un théorème de Pólya permet d'écrire pour $G = \mathbb{Z}^d$, $p_n(0, 0) = 0$ si n est impair et $p_n(0, 0) \asymp n^{-d/2}$ si n est pair.

Pour $N \in \mathbb{N} \cup \{\infty\}$, on définit $I_N := \sum_{n=0}^N \sum_{m=0}^N \mathbb{1}(X_n = Y_m)$. Comme on a fait l'hypothèse que $d \geq 3$, on travaille sur des graphes transients : une marche aléatoire ne passe presque sûrement qu'un nombre fini de fois par un sommet v donné. Ainsi $|\{X_n, n \geq 0\} \cap \{Y_n, n \geq 0\}|$ est fini p.s. si et seulement si I_∞ est fini presque sûrement. Il suffit donc de montrer que I_∞ est presque sûrement fini pour $d \geq 5$ et est presque sûrement infini pour $d \leq 4$.

Dans un premier temps, calculons $\mathbb{E}[I_N]$ pour $N \geq 0$:

$$\begin{aligned} \mathbb{E}[I_N] &\stackrel{\text{def}}{=} \sum_{n=0}^N \sum_{m=0}^N \sum_{x \in \mathbb{Z}^d} p_n(0, x) p_m(0, x) = \sum_{n=0}^N \sum_{m=0}^N p_{n+m}(0, 0) \\ &= \sum_{k=0}^{2N} \min\{k+1, 2N-k+1\} p_k(0, 0). \end{aligned}$$

On en déduit que :

$$\mathbb{E}[I_N] \asymp \sum_{k=0}^{2N} k p_k(0, 0) \asymp \sum_{k=0}^{2N} k^{1-d/2} \asymp \begin{cases} N^{2-d/2}, & \text{si } d \leq 3 \\ \log(N), & \text{si } d = 4 \\ 1, & \text{si } d \geq 5. \end{cases}$$

On en déduit donc que I_∞ est presque sûrement fini pour $d \geq 5$. Pour prouver le résultat pour $d \leq 4$, on estime le second moment de I_N .

On définit pour $N \in \mathbb{N} \cup \{\infty\}$ et $x, y \in \mathbb{Z}^d$, $\mathcal{G}_N(x, y) := \sum_{n=0}^N p_n(x, y)$. Comme \mathbb{Z}^d est stable par translation, on remarque que pour $x \in \mathbb{Z}^d$:

$$\begin{aligned} \sum_{y \in \mathbb{Z}^d} \mathcal{G}_N(x, y)^2 &= \sum_{y \in \mathbb{Z}^d} \sum_{n=0}^N \sum_{m=0}^N p_n(x, y) p_m(x, y) = \sum_{n=0}^N \sum_{m=0}^N p_{n+m}(x, x) \\ &= \sum_{n=0}^N \sum_{m=0}^N p_{n+m}(0, 0) = \mathbb{E}[I_N]. \end{aligned}$$

On cherche maintenant à calculer $\mathbb{E}[I_N^2]$ et à comparer cette quantité à $\mathbb{E}[I_N]^2$.

$$\begin{aligned}
\mathbb{E}[I_N^2] &= \mathbb{E}\left[\left(\sum_{x \in \mathbb{Z}^d} \sum_{n=0}^N \sum_{m=0}^N \mathbb{1}(X_n = x) \mathbb{1}(Y_m = x)\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{x \in \mathbb{Z}^d} \sum_{n=0}^N \mathbb{1}(X_n = x) \sum_{m=0}^N \mathbb{1}(Y_m = x)\right)^2\right] \\
&= \mathbb{E}\left[\sum_{x, y \in \mathbb{Z}^d} \left(\sum_{i=0}^N \sum_{j=0}^N \mathbb{1}(X_i = x) \mathbb{1}(Y_j = y)\right) \left(\sum_{k=0}^N \sum_{l=0}^N \mathbb{1}(X_k = y) \mathbb{1}(Y_l = x)\right)\right] \\
&= \sum_{x, y \in \mathbb{Z}^d} \left(\sum_{i=0}^N \sum_{k=0}^N \mathbb{E}[\mathbb{1}(X_i = x) \mathbb{1}(X_k = y)]\right) \left(\sum_{j=0}^N \sum_{l=0}^N \mathbb{E}[\mathbb{1}(Y_j = x) \mathbb{1}(Y_l = y)]\right) \\
&= \sum_{x, y \in \mathbb{Z}^d} \left(\sum_{i=0}^N \sum_{k=0}^N \mathbb{E}[\mathbb{1}(X_i = x) \mathbb{1}(X_k = y)]\right)^2
\end{aligned}$$

car X et Y sont indépendants et de même loi.

On cherche alors à rapprocher cette quantité à la fonction \mathcal{G} définie précédemment. Pour ce faire, on remarque de plus que :

$$\begin{aligned}
\sum_{i=0}^N \sum_{k=0}^N \mathbb{E}[\mathbb{1}(X_i = x) \mathbb{1}(X_k = y)] &\leq \sum_{0 \leq i \leq k \leq N} \mathbb{E}[\mathbb{1}(X_k = x | X_i = y)] + \mathbb{E}[\mathbb{1}(X_k = y | X_i = x)] \\
&\leq \sum_{0 \leq i \leq k \leq N} p_i(0, y) p_{k-i}(y, x) + p_i(0, x) p_{k-i}(x, y) \\
&\leq \sum_{0 \leq i, k \leq N} p_i(0, y) p_k(y, x) + p_i(0, x) p_k(x, y) \\
&= \mathcal{G}_N(0, y) \mathcal{G}_N(y, x) + \mathcal{G}_N(0, x) \mathcal{G}_N(x, y).
\end{aligned}$$

Finalement, à partir de la remarque faite plus haut, on peut écrire :

$$\begin{aligned}
\mathbb{E}[I_N^2] &\leq \sum_{x, y \in \mathbb{Z}^d} [\mathcal{G}_N(0, x) \mathcal{G}_N(x, y) + \mathcal{G}_N(0, y) \mathcal{G}_N(y, x)]^2 \\
&\leq \sum_{x, y \in \mathbb{Z}^d} 2 [(\mathcal{G}_N(0, x) \mathcal{G}_N(x, y))^2 + (\mathcal{G}_N(0, y) \mathcal{G}_N(y, x))^2] \\
&= 4 \sum_{x, y \in \mathbb{Z}^d} (\mathcal{G}_N(0, x) \mathcal{G}_N(x, y))^2 = 4 \sum_{x \in \mathbb{Z}^d} [\mathcal{G}_N(0, x)^2 \sum_{y \in \mathbb{Z}^d} \mathcal{G}_N(x, y)^2] \\
&= 4 \sum_{x \in \mathbb{Z}^d} \mathcal{G}_N(0, x)^2 \mathbb{E}[I_N] = 4 \mathbb{E}[I_N]^2.
\end{aligned}$$

Rappelons l'inégalité de Paley-Zigmund qui permet de comparer $\mathbb{E}[X]^2$ et $\mathbb{E}[X^2]$. Pour X une variable aléatoire positive de second moment fini et pour $\epsilon \geq 0$, on a :

$$\mathbb{P}(X \geq \epsilon \mathbb{E}[X]) \geq (1 - \epsilon)^2 \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}.$$

Ce qui donne dans notre cas :

$$\mathbb{P}(I_N \geq \frac{1}{2} \mathbb{E}[I_N]) \geq \frac{1}{16}.$$

Or pour $d \leq 4$, $\mathbb{E}[I_N] \rightarrow \infty$ quand $N \rightarrow \infty$, donc :

$$\mathbb{P}(I_N = \infty) \geq \mathbb{P}(\#\{N \geq 1 \text{ tel que } I_N \geq \frac{1}{2} \mathbb{E}[I_N]\} = \infty).$$

D'où par le lemme de Fatou :

$$\mathbb{P}(I_N = \infty) \geq \limsup_{N \rightarrow \infty} \mathbb{P}(I_N \geq \frac{1}{2} \mathbb{E}[I_N]) \geq \frac{1}{16}.$$

Or par la loi du zéro-un de Kolmogorov, $\mathbb{P}(I_N = \infty) \in \{0, 1\}$. De plus, par ce qui précède $\mathbb{P}(I_N = \infty) \neq 0$, donc $\mathbb{P}(I_N = \infty) = 1$. Le théorème est donc bien établi.

A partir de ce résultat, on peut finir la preuve du théorème 4.10 dans le cas $d \geq 5$.

Preuve. On se donne $(X(v))_{v \in V}$ une collection de marches aléatoires indépendantes partant chacune de $v \in V$.

Pour u et v dans V et pour $j \geq 0$, on pose $\mathcal{B}_j(u, v)$ l'évènement " $X(u)$ et $X(v)$ ne s'intersectent pas à partir du rang j ". On note $\alpha(u, v) = \mathbb{P}(\mathcal{B}_0(u, v))$ la probabilité que $X(u)$ et $X(v)$ ne s'intersectent pas.

Soit $u \neq v$ dans V . Pour $n \geq j$, en conditionnant sur les n premiers pas des marches $X(u)$ et $X(v)$:

$$\alpha(X_n(u), X_n(v)) \geq \mathbb{P}(\mathcal{B}_j(u, v) \mid \{X_m(u/v) \text{ tq } m \leq n\}).$$

Ce qui implique par la loi du zéro-un de Lévy :

$$\liminf_{n \rightarrow \infty} \alpha(X_n(u), X_n(v)) \geq \lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{B}_j(u, v) \mid \{X_m(u/v) \text{ tq } m \leq n\}) = \mathbb{1}(\mathcal{B}_j(u, v)) \text{ p. s.}$$

Or on sait par le théorème 4.13 que les marches sont presque sûrement d'intersections finies, donc $\lim_{j \rightarrow \infty} \mathbb{1}(\mathcal{B}_j(u, v)) = 1$ presque sûrement. On en déduit que :

$$\liminf_{n \rightarrow \infty} \alpha(X_n(u), X_n(v)) \geq 1.$$

Soit $\epsilon > 0$, on peut trouver n tel que $\alpha(X_n(u), X_n(v)) \geq 1 - \epsilon$. On prend dans l'algorithme de Wilson $v_1 = X_n(u)$ et $v_2 = X_n(v)$. Ainsi la forêt obtenue n'est pas connexe avec une probabilité supérieure à $1 - \epsilon$. En faisant tendre ϵ vers 0, on obtient le résultat voulu.

5 Conclusion

Il reste encore beaucoup de questions non évoquées par ce mémoire, on peut par exemple se poser la question d'à quelle condition \mathbf{FUSF}_G et \mathbf{WUSF}_G coïncident ou le nombre de composantes connexes de \mathbf{WUSF}_G . Ces questions ont été résolues (on peut consulter par exemple [LP16]) mais il y en a encore beaucoup qui ne sont pas résolues : par exemple, on peut montrer comme dans la dernière preuve de ce mémoire que la probabilité qu'un objet ayant loi \mathbf{WUSF}_G soit connexe est 0 ou 1 mais le cas de \mathbf{FUSF}_G est encore inconnu.

Références

- [Ald90] David J Aldous. The random walk construction of uniform spanning trees and uniform labelled trees. *SIAM Journal on Discrete Mathematics*, 3(4) :450–465, 1990.
- [BLPS01] Itai Benjamini, Russell Lyons, Yuval Peres, and Oded Schramm. *Annals of Probability* 29, pages 1–65, 2001.
- [Bro89] Andrei Z Broder. Generating random spanning trees. In *FOCS*, volume 89, pages 442–447, 1989.
- [Gal06] Jean-François Le Gall. *Intégration, probabilités et processus aléatoires*. Septembre 2006.
- [Hut21] Tom Hutchcroft. *Random walks and uniform spanning trees*. University of Cambridge, January 2021.
- [LP16] Russell Lyons and Yuval Peres. *Probability on Trees and Networks*, volume 42 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, New York, 2016. Available at <https://rdlyons.pages.iu.edu/>.
- [Pem91] Robin Pemantle. Choosing a spanning tree for the integer lattice uniformly. *Annals of Probability* 19, pages 1559–1574, 1991.
- [Wil96] David Bruce Wilson. Generating random spanning trees more quickly than the cover time. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 296–303, 1996.