

Cutoff pour un modèle de vote

Emile Avérous

Encadrante: Perla Sousi*

Laboratoire: Statistical Laboratory, University of Cambridge

2023

1 Introduction

Considérons une chaîne de Markov $(X_t)_t$ sur un ensemble fini Ω , avec matrice de transition P . Sous l'hypothèse que cette chaîne est irréductible et apériodique, la distribution de X_t converge, lorsque $t \rightarrow +\infty$, vers l'unique mesure de probabilité stationnaire π sur Ω , et ceci quelque soit l'état de départ X_0 . On définira alors, pour tout $\varepsilon > 0$, le temps de mélange $t_{mix}(\varepsilon)$ comme le temps nécessaire pour que X_t s'approche à distance ε de la distribution stationnaire, pour n'importe quel état de départ X_0 .

L'étude de temps de mélange de chaînes de Markov est un sujet très vaste, et de nombreuses méthodes ont été mises en place, suivant les modèles, pour obtenir des bornes sur des temps de mélange de certaines chaînes de Markov. Dans mon stage, je me suis intéressé notamment au phénomène de cutoff, observé initialement pour des modèles de mélange de cartes, puis pour de nombreux autres modèles. Ce phénomène se définit par le fait que le temps de mélange $t_{mix}(\varepsilon)$ ne dépende presque plus de ε , ou en d'autres termes que la convergence de la chaîne de Markov X_t vers la distribution stationnaire π se fait sur une période très courte devant le temps de mélange.

Le modèle principal auquel je me suis intéressé est celui du noisy voter model, défini comme suit: dans un ensemble fini S , chaque point possède un vote pouvant prendre n valeurs. Alors la chaîne évolue de la façon suivante:

- Chaque point $x \in S$, indépendamment des autres votes, change son vote en un vote uniforme dans $1, \dots, n$ à des temps suivant un processus de Poisson d'intensité 1.
- Pour tous points $x, y \in S$, le point x change son vote en celui de y à des temps suivant un processus de Poisson d'intensité $q(x, y) \geq 0$.

Un des objectifs du stage était de chercher à montrer que le noisy voter model sur un graphe possédait le phénomène de cutoff lorsque la taille du graphe tendait vers l'infini, sous certaines conditions de régularité. Dans le cas $n = 2$, le problème a été résolu par Cox, Peres et Steif dans [1] où ils démontrent l'existence de cutoff sous des conditions assez faibles. Cependant, leur preuve est spécifique au cas $n = 2$ et ne peut pas s'adapter de manière simple dans le cas d'un nombre quelconque de votes. Afin de généraliser à n votes, j'ai adapté la méthode de percolation d'information initialement introduite par Lubetzky et Sly pour le modèle d'Ising dans [2, 3, 4].

J'ai alors obtenu les résultats suivants:

Théorème 1. *Pour tout $n \geq 2$ entier, il existe un réel $q_0 = q_0(n) > 0$ tel que si les intensités $q(x, y)$ vérifient la condition*

$$\forall x \in S, \quad \sum_{y \in S} (q(x, y) + q(y, x)) \leq q_0,$$

*University of Cambridge, Cambridge, UK. Email: p.sousi@statslab.cam.ac.uk

alors lorsque $|S| \rightarrow +\infty$, le noisy voter model possède un cutoff au temps $\frac{1}{2} \log |S|$ et avec une fenêtre d'ordre 1.

Théorème 2. Pour tous $n \geq 2$ entier et réels $q_0, K > 0$, supposons que les intensités $q(x, y)$ vérifient la condition

$$\forall x \in S, \quad \sum_{y \in S} q(x, y) \leq q_0.$$

Supposons aussi que les $q(x, y)$ ne soient non nuls que sur les arêtes d'un certain graphe $G = (S, A)$ dont les boules vérifient les conditions suivantes:

$$\begin{cases} \forall x \in S, r > 0, & |B_G(x, r)| \leq Kr^K \\ \forall x, y \in S, r > 0, & \frac{|B_G(x, r)|}{|B_G(y, r)|} \leq K \end{cases} \quad (1)$$

alors lorsque $|S| \rightarrow +\infty$, le noisy voter model possède un cutoff au temps $\frac{1}{2} \log |S|$ et avec une fenêtre d'ordre 1.

Le stage s'est déroulé au Statistical Laboratory à l'université de Cambridge, et a été encadré par Perla Sousi. J'ai étudié plusieurs problèmes lors de mon stage, en lien avec les temps de mélange de chaînes de Markov. Initialement, j'ai étudié le phénomène de cutoff pour le noisy voter model, ce qui est l'objet principal de ce rapport. Ensuite, j'ai étudié les temps de mélange pour des dynamiques sur le random cluster model et le modèle de Potts, qui est une généralisation du modèle d'Ising. Enfin, j'ai cherché à montrer un résultat pour le temps de mélange d'un modèle de marche aléatoire sur une percolation dynamique, qui est présenté à la fin du rapport.

Pendant le stage, j'ai pu assister au séminaire de probabilité les semaines où il a eu lieu. Il y avait aussi un groupe de lecture où des doctorants ou post-doctorants présentaient des articles, j'ai pu y présenter les articles que j'avais étudié. J'ai ainsi pu interagir avec certaines personnes du département de probabilités, comme Jason Miller et Wendelin Werner.

2 Définitions

2.1 Temps de mélange, cutoff

On définit ici la notion de temps de mélange pour une chaîne de Markov X_t définie sur un ensemble fini. On définit une distance "en variation totale" entre deux mesures de probabilité μ, ν sur Ω par:

$$\|\mu - \nu\|_{TV} = \sup_{A \subset \Omega} |\mu(A) - \nu(A)|.$$

Une vérification simple donne une autre interprétation de cette distance

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|.$$

Lorsque la chaîne de Markov commence en $X_0 = x \in \Omega$, la convergence de X_t vers la mesure stationnaire π est décrite par la distance $\|\mathbb{P}_x(X_t \in \cdot) - \pi\|_{TV}$ qui tend vers 0 lorsque $t \rightarrow +\infty$. Afin de supprimer la dépendance en l'état initial x dans cette distance, on définit la distance

$$d(t) = \max_{x \in \Omega} \|\mathbb{P}_x(X_t \in \cdot) - \pi\|_{TV}.$$

Comme Ω est fini, cette distance $d(t)$ tend vers 0 lorsque $t \rightarrow +\infty$. On définit à présent le temps de mélange de la chaîne de Markov en posant, pour tout $\varepsilon > 0$,

$$t_{mix}(\varepsilon) = \inf\{t \geq 0 \mid d(t) \leq \varepsilon\}.$$

On pose aussi $t_{mix} = t_{mix}(\frac{1}{4})$. Un des grands enjeux de l'étude de nombreuses chaînes de Markov est la compréhension de ces temps de mélange.

Considérons à présent la matrice de transition P , dont les lignes et les colonnes sont indicées par Ω . Alors la distribution de X_t sous \mathbb{P}_x est donnée par la x -ième colonne de P^t , et donc en notant Π la matrice dont toutes les colonnes correspondent à la distribution stationnaire π , on a $d(t) = \|P^t - \Pi\|$ où la norme $\|\cdot\|$ est donnée par

$$\|A - B\| = \frac{1}{2} \max_{x \in \Omega} \sum_{y \in \Omega} |A_{yx} - B_{yx}|.$$

Matriciellement, la convergence de P^t vers Π s'interprète de la façon suivante: les valeurs propres de P sont toutes de module inférieur à 1, et cette inégalité est stricte sauf pour 1 qui est valeur propre de P avec espace propre la droite engendrée par π (vu comme un vecteur colonne). Alors en notant

$$\lambda_* = \max\{|\lambda|, \lambda \in \text{Spec}(P) \setminus \{1\}\},$$

la distance de P^t à Π vaut environ λ_*^t . Plus précisément, on a

$$d(t)^{\frac{1}{t}} \rightarrow \lambda_* \quad (t \rightarrow +\infty).$$

En général, λ_* est très proche de 1, donc on a l'estimation $\log(\lambda_*) \approx \lambda_* - 1$. Ainsi, lorsque $t \rightarrow +\infty$, un temps caractéristique de l'évolution de $d(t)$ est le temps de relaxation

$$t_{rel} = \frac{1}{1 - \lambda_*}.$$

D'après cette étude qualitative, on s'attend à ce que $d(t)$ se comporte comme une exponentielle décroissante avec un temps caractéristique t_{rel} , et donc que $t_{mix}(\varepsilon)$ possède une décroissance en $\log(\frac{1}{\varepsilon})$. Mais en fait dans de nombreux cas, le temps de mélange est beaucoup plus grand que le temps de relaxation. De plus, pour certaines chaînes de Markov, on observe plutôt le comportement curieux que $t_{mix}(\varepsilon)$ ne dépend presque pas de ε . Plus précisément,

Définition 1. Soient $(X_t^{(n)})_t$ des chaînes de Markov sur les ensembles finis $\Omega^{(n)}$ respectivement. Notons $t_{mix}^{(n)}$ les temps de mélange respectifs des chaînes de Markov. Alors cette suite de chaînes de Markov a un cutoff si

$$\forall \varepsilon \in]0, 1[, \quad \frac{t_{mix}^{(n)}(1 - \varepsilon)}{t_{mix}^{(n)}(\varepsilon)} \rightarrow 1 \quad (n \rightarrow +\infty).$$

De plus, on dit qu'elle a un cutoff avec une fenêtre d'ordre $w(n)$ si $w(n) = o(t_{mix}^{(n)})$ et

$$\forall \varepsilon \in]0, 1[, \quad t_{mix}^{(n)}(\varepsilon) = t_{mix}^{(n)} + O_\varepsilon(w(n)) \quad (n \rightarrow +\infty).$$

Le phénomène de cutoff a été montré pour de nombreuses chaînes de Markov, mais il reste très mal compris. Essentiellement, le seul résultat général connu est que le phénomène de cutoff ne peut arriver que si le temps de mélange est grand devant le temps relaxation, ce qui donne la condition nécessaire

$$\frac{t_{mix}^{(n)}}{t_{rel}^{(n)}} \rightarrow 0 \quad (n \rightarrow +\infty).$$

Malheureusement, on peut aussi montrer que cette condition (appelée "condition du produit") n'est pas suffisante pour avoir le phénomène de cutoff. Afin de montrer le cutoff pour un système, il faut pouvoir étudier très précisément la convergence de la chaîne de Markov vers la mesure stationnaire, pour obtenir des bornes supérieures et inférieures très précises.

2.2 Chaînes de Markov à temps continu

On définit ici la notion de chaîne de Markov à temps continu sur un ensemble discret. On commence par définir un processus de Poisson, dont on admet l'existence:

Définition 2. *Un processus de Poisson (homogène) d'intensité $\lambda > 0$ est un ensemble aléatoire de points sur \mathbb{R}_+ tel que pour tout intervalle borné $B \subset \mathbb{R}_+$, le nombre $N(B)$ de points appartenant à B suit une loi de Poisson de paramètre $\lambda\nu(B)$, où $\nu(B)$ est la mesure de Lebesgue de B :*

$$\mathbb{P}(N(B) = k) = e^{-\lambda\nu(B)} \frac{(\lambda\nu(B))^k}{k!}.$$

A partir d'une chaîne de Markov à temps discret $(\tilde{X}_t)_{t \in \mathbb{N}}$ avec matrice de transition P , on peut définir une chaîne de Markov à temps continu $(X_t)_{t \geq 0}$ en prenant un processus de Poisson, et à chaque point t de ce processus, en modifiant la valeur de la chaîne de Markov selon la matrice de transition P . Ainsi, la chaîne de Markov à temps continu est précisément celle à temps discret où l'on attend un temps exponentiel à chaque état.

L'avantage d'une chaîne de Markov à temps continu est que l'on peut considérer facilement plusieurs types de transition. Par exemple, on peut définir une chaîne de Markov en disant qu'à des temps suivant un processus de Poisson de paramètre λ_1 , la chaîne est modifiée selon la matrice de transition P_1 , et à des temps suivant un processus de Poisson de paramètre λ_2 , la chaîne est modifiée selon la matrice de transition P_2 . Lorsque l'on dit ceci, on sous-entend que les processus de Poisson sont indépendants, et alors on admettra que l'on obtient bien une chaîne de Markov à temps continu, qui possède des transitions à des temps suivant un processus de Poisson de paramètre $\lambda_1 + \lambda_2$. On admettra aussi que le résultat se généralise à plus de types de transition.

La théorie des chaînes de Markov à temps discret s'adapte en grande partie à l'étude des chaînes de Markov à temps continu. La plus grande différence est le fait que la notion d'apériodicité n'a plus de sens, et ainsi il suffit qu'une chaîne soit irréductible pour que sa distribution au temps t converge vers une unique mesure stationnaire lorsque $t \rightarrow +\infty$ (dans le cas d'un ensemble de valeurs fini). Les notions de temps de mélange et de cutoff sont alors identiques au cas du temps discret, et en général il est assez simple d'obtenir des estimations des temps de mélange en temps discret à partir de ceux en temps continu (et inversement).

2.3 Le noisy voter model

Considérons un ensemble fini S , ainsi que pour tous $x, y \in S$ distincts, un réel $q(x, y)$ positif ou nul. Soit $n \geq 2$ un entier, et notons $[n] = \{1, 2, \dots, n\}$ l'ensemble des votes. Une configuration de votes sera un élément de $\Omega = [n]^S$, et le vote d'un sommet $x \in S$ dans une configuration $\omega \in \Omega$ sera $\omega(x)$. Le *noisy voter model* sur (S, q, n) est la chaîne de Markov à temps continu X_t à valeurs dans Ω définie de la manière suivante:

- Pour tout $x \in S$, suivant un processus de Poisson d'intensité 1, x met à jour son vote de manière aléatoire en un élément uniforme de $[n]$.
- Pour tous $x, y \in S$, suivant un processus de Poisson d'intensité $q(x, y)$, x met à jour son vote en copiant le vote de y , on dit que le sommet x se fait influencer par y .

On notera π la distribution stationnaire. Un point important est que cette distribution n'est pas explicite, contrairement à de nombreux modèles de spins comme le modèle d'Ising.

Ce modèle admet la propriété suivante de monotonie: à partir de deux configurations de votes ω, ω' , supposons qu'à chaque sommet x , si $\omega(x) = k$, alors $\omega'(x) = k$. Alors si les configurations subissent une étape de la chaîne de Markov à un sommet y , le vote de y a une plus grande probabilité de devenir k dans le cas de ω' que dans le cas de ω .

On définit alors le couplage monotone de la chaîne. Si l'on considère le noisy voter model parti de deux configurations initiales ω, ω' , alors on peut coupler ces deux chaînes en faisant en sorte que les mises à jour de ces deux chaînes soient identiques: si un sommet est mis à jour de manière aléatoire dans une des chaînes, il en est de même dans l'autre chaîne et il choisissent le même vote dans les deux cas, de plus si un sommet se fait influencer par un autre, il est en de même dans l'autre chaîne. Ce couplage s'étend à toutes les chaînes de Markov démarrant de n'importe quelle condition initiale $\omega \in [n]^S$, pour la suite de la preuve on supposera

toujours que les chaînes de Markov sont couplées de telle manière. On remarque le point suivant: considérons par exemple les chaînes X_t^1 et X_t^n parties des configurations où tous les votes valent 1 ou n respectivement. Si X_t est une chaîne partie d'une configuration arbitraire, alors on a pour tout t , $X_t^1 \leq X_t \leq X_t^n$ (composante par composante) sous le couplage monotone, et donc si $X_t^1 = X_t^n$ à un instant t , la chaîne X_t vaut aussi cette même valeur.

3 La percolation d'information

3.1 Le noisy voter model dual

Lorsque l'on considère le noisy voter model X_t après un certain temps t , il peut paraître assez compliqué d'étudier la distribution des différents votes. Une manière de faire ceci est de partir du temps t et de revenir en arrière. On commence par se donner la variable aléatoire \mathcal{U} composée de toutes les données des mises à jour des votes, c'est-à-dire l'ensemble des couples (x, s) où le vote de x est mis à jour aléatoirement au temps s , et des triplets (x, y, s) où x prend le vote de y au temps s . A partir de ces informations, on retrouve le vote d'un sommet x au temps t en revenant dans le temps. Considérons un marcheur qui est en x au temps t et qui remonte le temps. A tout moment, si il rencontre une mise à jour aléatoire en un point, il meurt. Si il rencontre une mise à jour où le point où il est se fait influencer par un point z , il bouge en z . S'il arrive au temps 0 sans mourir, il s'arrête.

Alors le vote de x au temps t se détermine de la façon suivante: si le marcheur est mort avant d'atteindre le temps 0, le vote de x est uniforme sur $[n]$ (car cela signifie que le vote de x provient d'un sommet qui a modifié son vote de manière aléatoire). Si le marcheur atteint le sommet y au temps 0, alors le vote de x au temps t est celui de y au temps 0. L'ensemble des mises à jour étant aléatoire, le marcheur effectue en fait une marche aléatoire (en temps continu) sur S , avec des intensités de transition $q(x, y)$ entre deux sommets x et y , et il meurt suivant un processus de Poisson d'intensité 1. Maintenant, les corrélations entre les différents votes au temps t s'expliquent simplement par le fait que lorsque deux marcheurs partis de deux points différents se retrouvent à un certain temps, ils restent ensemble jusqu'à mourir ou atteindre le temps 0. Conditionnellement à \mathcal{U} , deux sommets ont le même vote au temps t si les marcheurs correspondants se sont retrouvés, sinon leurs votes sont indépendants. Par exemple, sur la figure 1, les marcheurs partis de x et de y sont commencent en haut et descendent suivant la ligne en magenta (les lignes vertes sur la figure, ainsi que la deuxième partie de la légende, serviront pour la suite de la preuve). Le marcheur parti de x meurt avant d'atteindre le temps 0, donc son vote est uniforme sur $[n]$. Celui parti de y ne rencontre jamais celui parti de x , donc le vote de y est indépendant de celui de x conditionnellement à \mathcal{U} , et il atteint le temps 0, donc le vote de y est précisément le vote du sommet atteint dans la condition initiale. Notons que sur la figure, \mathcal{U} est donné par l'ensemble des flèches et des croix.

Remarquons tout d'abord que le vote à un sommet x au temps t dépend de la configuration initiale si et seulement si le marcheur associé atteint le temps 0, sinon le vote est uniforme. Comme le marcheur meurt suivant un processus de Poisson d'intensité 1, la probabilité que le vote de x dépende de la configuration initiale est e^{-t} . Un des avantages de ce modèle par rapport aux dynamiques sur d'autres modèles est que cette probabilité est indépendante de x et de la géométrie du problème. A partir de ce résultat, on en déduit que si on attend un temps $t = (1 + o(1)) \log |S|$, avec grande probabilité la configuration de spins au temps t ne dépend plus de la configuration initiale. Comme la configuration initiale aurait pu être choisie selon la distribution stationnaire π , la distribution au temps t est donc (conditionnellement) égale à la distribution stationnaire. On obtient notamment que les temps de mélange $t_{mix}(\varepsilon)$ sont bornés par $(1 + o(1)) \log |S|$. Mais en fait, il n'est pas nécessaire d'attendre aussi longtemps pour se rapprocher de la distribution stationnaire, on verra qu'il suffit d'un temps $(\frac{1}{2} + o(1)) \log |S|$.

3.2 Les clusters d'information

On regroupe à présent les sommets de S en des *clusters*, qui sont précisément les ensembles (aléatoires) de sommets dont toutes les marches aléatoires parties des sommets se rejoignent avant de mourir ou d'atteindre le temps 0. Deux points appartiennent au même cluster si et seulement si les deux marcheurs correspondants se rejoignent, et on peut vérifier que ceci forme bien une relation d'équivalence entre les sommets. Ainsi,

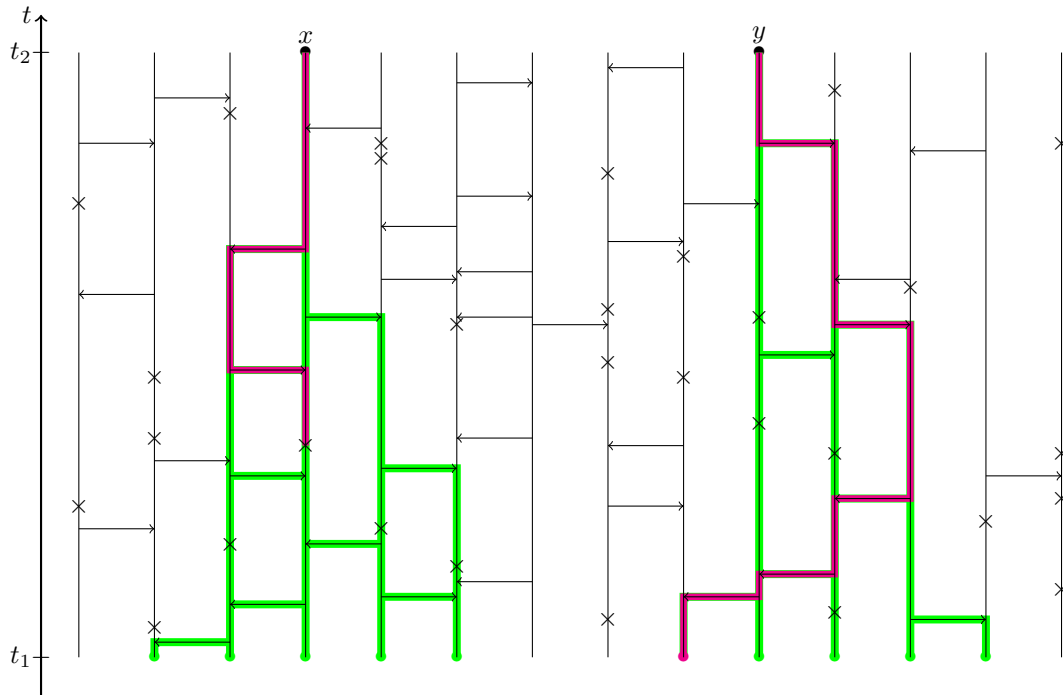


Figure 1: L'espace $S \times [t_1, t_2]$. Chaque sommet de S est représenté par une ligne verticale, les voisins correspondent aux lignes adjacentes. Les mises à jour aléatoires sont représentées par des croix, les influences sont représentées par des flèches allant du sommet influencé vers celui qui influence.

Pour les sommets x et y , on détermine les fonctions $\mathcal{F}_{\text{SUPP}}$ et \mathcal{F}_{UPD} entre les temps t_2 et t_1 . En magenta, on montre le calcul des fonctions support de mises à jour, obtenues en suivant un chemin descendant depuis le sommet de départ et suivant les flèches, en s'arrêtant sur une croix. Comme le chemin parti de x s'arrête avant d'atteindre le temps t_1 , $\mathcal{F}_{\text{SUPP}}(x, t_1, t_2)$ est vide, alors que $\mathcal{F}_{\text{SUPP}}(y, t_1, t_2)$ est un singleton composé de l'unique point magenta au temps t_1 . Les fonctions $\mathcal{F}_{\text{UPD}}(x, t_1, t_2)$ et $\mathcal{F}_{\text{UPD}}(y, t_1, t_2)$ sont données par les deux ensembles associés de points verts au temps t_1 (ainsi que le point magenta dans le cas de y).

conditionnellement sur l'ensemble des clusters, tous les sommets d'un même cluster ont les mêmes votes, et le vote d'un cluster est indépendant de tous les votes extérieurs au cluster.

Parmi les clusters, certains correspondent à des marches qui atteignent le temps 0, d'autres à des marches qui meurent avant. Dire que la distribution des votes ne dépend pas beaucoup de la configuration initiale au temps t revient à dire que les clusters dont les marches atteignent le temps 0 sont peu nombreux. On appelle ces clusters les clusters *rouges*.

Les clusters dont les marches meurent avant d'atteindre le temps 0 vont avoir une distribution qui va "correspondre" à la distribution stationnaire. Celle-ci peut être assez compliquée à comprendre, mais est particulièrement simple si on conditionne sur le fait que ces clusters sont tous des singletons, car il n'y a alors pas de corrélation entre les votes. Appelons alors un cluster *bleu* s'il est composé d'un unique point et que les marches associées meurent avant d'atteindre le temps 0. Les clusters restants seront les clusters *verts*. En d'autres termes, les clusters non rouges sont séparés selon s'ils sont un singleton ou pas.

L'idée de la preuve va être de montrer que dans le cas où les intensités $q(x, y)$ sont suffisamment faibles et que l'on attend un temps $\frac{1}{2} \log |S| + O(1)$, alors il y aura beaucoup plus de clusters bleus que de clusters rouges. Ainsi on se permet de ne plus regarder les clusters verts, et on va en fait conditionner sur ceux-ci.

Cette méthode, dite de percolation d'information, est très puissante pour de nombreux modèles sur un réseau. Elle ne nécessite pas de connaître beaucoup de choses sur la structure du problème, notamment la probabilité stationnaire π . Cependant, elle nécessite la monotonie de la chaîne, ainsi qu'une certaine propriété de localité des transitions de la chaîne, ce qui est un obstacle dans le cas où les intensités $q(x, y)$ sont plus larges.

Notons que la méthode telle qu'elle est présentée ci-dessus est adaptée au cas de faibles intensités $q(x, y)$, c'est-à-dire pour le théorème 1. Afin de démontrer le théorème 2, il est nécessaire de définir les clusters d'une autre manière plus compliquée, et c'est alors en utilisant la géométrie du graphe G que l'on pourra borner les tailles des clusters. Dans cette section, on se concentre sur le premier cas, qui va permettre d'illustrer les idées principales de la preuve.

3.3 Le déroulement de la preuve

La partie difficile de la preuve est de trouver une borne supérieure pour le temps de mélange t_{mix} . Pour faire cela, définissons la variable aléatoire $\mathcal{H}_{\text{VERT}}$ comme "l'histoire des clusters verts", c'est-à-dire l'ensemble des marches aléatoires partant d'un sommet appartenant à un cluster vert, ou encore le sous ensemble de \mathcal{U} constitué seulement des informations portant sur les clusters verts. On va travailler conditionnellement à cette variable aléatoire. Rappelons la définition de la distance à la stationarité

$$d(t) = \max_{\omega \in \Omega} \|\mathbb{P}_{\omega}(X_t \in \cdot) - \pi\|_{TV} = \max_{\omega \in \Omega} \|\mathbb{P}_{\omega}(X_t \in \cdot) - \mathbb{P}_{\pi}(X_t \in \cdot)\|_{TV}$$

où la deuxième inégalité est par stationarité de π . L'inégalité de Jensen donne

$$d(t) \leq \max_{\omega \in \Omega} \mathbb{E}(\|\mathbb{P}_{\omega}(X_t \in \cdot | \mathcal{H}_{\text{VERT}}) - \mathbb{P}_{\pi}(X_t \in \cdot | \mathcal{H}_{\text{VERT}})\|_{TV}).$$

On remplace alors l'espérance par un maximum sur toutes les valeurs possibles \mathcal{X} de $\mathcal{H}_{\text{VERT}}$:

$$d(t) \leq \max_{\omega} \sup_{\mathcal{X}} \|\mathbb{P}_{\omega}(X_t \in \cdot | \mathcal{H}_{\text{VERT}} = \mathcal{X}) - \mathbb{P}_{\pi}(X_t \in \cdot | \mathcal{H}_{\text{VERT}} = \mathcal{X})\|_{TV}.$$

où l'on admet que ce conditionnement peut être défini de manière correcte qui coïncide avec l'intuition. Maintenant, sous le couplage monotone des chaînes démarrant de ω et π , remarquons que les seuls endroits où les deux chaînes sont potentiellement en désaccord sont sur les sommets des clusters rouges, car les sommets des autres clusters ont des votes qui ne dépendent pas des conditions initiales. Ainsi, conditionnellement à $\mathcal{H}_{\text{VERT}} = \mathcal{X}$, les deux chaînes ne peuvent être qu'en désaccord sur S_{VERT}^c le complémentaire des sommets appartenant à des clusters verts. Comme X_t est donné par le couple $(X_t(S_{\text{VERT}}), X_t(S_{\text{VERT}}^c))$ (où pour $S' \subset S$, $X_t(S')$ est simplement X_t restreint à l'ensemble des sommets S'), et que l'on a l'égalité, simple à vérifier, $\|Loi(Y, Z) - Loi(Y, Z')\|_{TV} = \|Loi(Z) - Loi(Z')\|_{TV}$, on peut montrer que

$$d(t) \leq \max_{\omega} \sup_{\mathcal{X}} \|\mathbb{P}_{\omega}(X_t(S_{\text{VERT}}^c) \in \cdot | \mathcal{H}_{\text{VERT}} = \mathcal{X}) - \mathbb{P}_{\pi}(X_t(S_{\text{VERT}}^c) \in \cdot | \mathcal{H}_{\text{VERT}} = \mathcal{X})\|_{TV}.$$

Maintenant comme expliqué dans la partie précédente, on cherche à montrer que sur S_{VERT}^c , les sommets appartiennent presque tous à des clusters bleus, et donc que la distribution de $X_t(S_{\text{VERT}}^c)$ est proche de la distribution $\nu_{S_{\text{VERT}}^c}$ uniforme sur $[n]^{S_{\text{VERT}}^c}$. On obtient alors par inégalité triangulaire

$$d(t) \leq 2 \max_{\omega} \sup_{\mathcal{X}} \|\mathbb{P}_{\omega}(X_t(S_{\text{VERT}}^c)) \in \cdot | \mathcal{H}_{\text{VERT}} = \mathcal{X}\} - \nu_{S_{\text{VERT}}^c}\|_{TV}.$$

Maintenant que l'on s'est ramené à estimer la distance à une distribution uniforme, on peut utiliser un lemme originellement dû à Miller et Peres dans [5], que l'on adapte pour n valeurs à chaque sommet comme suit:

Lemme 3.1. *Soit S un ensemble fini et $\Omega = [n]^S$, et soit ν la mesure uniforme sur Ω . Pour tout $R \subset S$, soit φ_R une mesure sur $[n]^R$, et soit μ_0 une mesure sur les sous-ensembles de S . On définit alors une mesure μ sur Ω en choisissant d'abord un sous-ensemble $R \subset S$ avec la mesure μ_0 , puis en choisissant sa restriction à R selon φ_R et sa restriction à R^c de manière indépendante selon la mesure uniforme sur $[n]^{R^c}$. Alors la distance de μ à ν s'estime par*

$$\left\| \frac{\mu}{\nu} - 1 \right\|_{L^2(\nu)}^2 \leq \mathbb{E}(n^{|R \cap R'|}) - 1$$

où R, R' sont deux variables aléatoires indépendantes distribuées selon μ_0 .

D'après Cauchy-Schwarz, on a

$$d(t) \leq \max_{\omega} \sup_{\mathcal{X}} \left\| \frac{\mathbb{P}_{\omega}(X_t(S_{\text{VERT}}^c))}{\nu_{S_{\text{VERT}}^c}} - 1 \right\|_{L^2(\nu_{S_{\text{VERT}}^c})}.$$

On applique à présent le lemme précédent sur l'ensemble de sommets S_{VERT}^c . On sait que sur l'ensemble des clusters bleus, X_t est distribuée de manière uniforme. Soit donc μ_0 la distribution de l'union des clusters rouges (conditionnellement à $\mathcal{H}_{\text{VERT}} = \mathcal{X}$), alors en choisissant correctement les mesures φ_R , la mesure μ du lemme est égale à $\mathbb{P}_{\omega}(X_t(S_{\text{VERT}}^c))$, et ainsi d'après le lemme,

$$d(t)^2 \leq \sup_{\mathcal{X}} \mathbb{E}(n^{|S_{\text{ROUGE}} \cap S'_{\text{ROUGE}}|} | \mathcal{H}_{\text{VERT}} = \mathcal{X}) - 1$$

où $S_{\text{ROUGE}}, S'_{\text{ROUGE}}$ sont deux variables indépendantes distribuées (conditionnellement) comme l'union des clusters rouges. On remarque que le maximum sur ω peut être enlevé puisque le côté droit ne dépend plus de la condition initiale.

Afin d'estimer le nombre et la taille des clusters rouges, on introduit la probabilité suivante, pour tout $A \subset S$:

$$\Psi_A = \sup_{\mathcal{X}} \mathbb{P}(A \text{ est un cluster rouge} | \mathcal{H}_{A^c} = \mathcal{X}, A \text{ est soit un cluster rouge, soit une union de clusters bleus})$$

où \mathcal{H}_{A^c} est l'ensemble de toutes les marches partant d'un sommet de A^c , et ici le supremum sur \mathcal{X} est pris sur l'ensemble des valeurs de \mathcal{H}_{A^c} . On définit les Ψ_A ainsi car les événements conditionnels dans la définition de Ψ_A ont, en quelque sorte, des corrélations négatives. Plus précisément, on obtient:

Lemme 3.2. *Soient $(Y_{A,A'})_{A,A' \subset S}$ des variables aléatoires indépendantes de Bernoulli de paramètres respectifs $\Psi_A \Psi_{A'}$. Alors il est possible de coupler ces variables aléatoires avec la distribution de $(S_{\text{ROUGE}}, S'_{\text{ROUGE}})$ conditionnelle à $\mathcal{H}_{\text{VERT}}$ afin d'avoir*

$$|S_{\text{ROUGE}} \cap S'_{\text{ROUGE}}| \leq \sum_{\substack{A, A' \subset S \\ A \cap A' \neq \emptyset}} |A \cup A'| Y_{A,A'}.$$

Ensuite, les valeurs de Ψ_A peuvent être bornées. Afin de simplifier l'exposition de la méthode, on présente la borne suivante, qui ne suffit pas pour traiter le cas général des faibles intensités, mais seulement le cas où, en plus, les intensités sont à support sur un graphe G où les degrés sont uniformément bornés. Afin de simplifier l'exposition des idées de la percolation d'information, on ne va que donner les étapes de la preuve pour ce cas particulier.

Lemme 3.3. *Pour tout $\lambda > 0$, il existe des constantes $q_0(\lambda), C_0(\lambda) > 0$ telles que si pour tout $x \in S$, $\sum_{y \in S} q(x, y) \leq q_0$, alors*

$$\forall A \subset S, \Psi_A \leq C_0 e^{-t} e^{-\lambda \mathfrak{W}(A)}$$

où $\mathfrak{W}(A)$ est le plus petit ensemble connexe contenant A (où la relation d'adjacence est donnée par $q(x, y) + q(y, x) > 0$).

En combinant ces deux lemmes avec les inégalités d'avant, on obtient

$$d(t)^2 \leq \mathbb{E} \left(n^{\sum_{A \cap A' \neq \emptyset} |A \cup A'| Y_{A, A'}} \right) - 1.$$

où l'on a enlevé le conditionnement dans le côté droit car le côté droit ne dépend plus de $\mathcal{H}_{\text{VERT}}$. Maintenant, on peut utiliser le fait que les $Y_{A, A'}$ sont des variables de Bernoulli indépendantes pour obtenir

$$d(t)^2 \leq \prod_{A \cap A' \neq \emptyset} \left(1 + \Psi_A \Psi_{A'} (n^{|A \cup A'|} - 1) \right) - 1.$$

On peut remplacer $|A \cup A'|$ par $|A| + |A'|$, et utiliser l'inégalité $1 + x \leq e^x$ pour obtenir

$$d(t)^2 \leq \exp \left(\sum_{A \cap A' \neq \emptyset} \Psi_A \Psi_{A'} n^{|A| + |A'|} \right) - 1.$$

On remplace ensuite la somme par une somme sur $x \in S$ puis une double somme sur A, A' contenant x , ce qui ne peut que augmenter l'expression, et on obtient

$$d(t)^2 \leq \exp \left(\sum_{x \in S} \left(\sum_{x \ni A} \Psi_A n^{|A|} \right)^2 \right) - 1 \leq \exp \left(|S| \max_{x \in S} \left(\sum_{x \ni A} \Psi_A n^{|A|} \right)^2 \right) - 1.$$

On utilise l'inégalité $e^x - 1 \leq 2x$ pour $0 \leq x \leq 1$ et le fait que $d(t) \leq 1$ pour obtenir

$$d(t) \leq \sqrt{2|S|} \max_{x \in S} \sum_{x \ni A} \Psi_A n^{|A|}.$$

La borne sur Ψ_A donne pour d'assez faibles intensités en fonction de λ (en choisissant λ plus tard):

$$d(t) \leq C_0 \sqrt{2|S|} e^{-t} \max_{x \in S} \sum_{x \ni A} n^{|A|} e^{-\lambda \mathfrak{W}(A)}.$$

Mais maintenant si les intensités sont à support sur un graphe dont les degrés sont au plus d , on peut montrer que le nombre de $A \subset S$ contenant un point x donné et tels que $\mathfrak{W}(A) = k$ est borné par $(2d)^{k-1}$, et donc, comme $|A| \leq \mathfrak{W}(A)$:

$$\begin{aligned} d(t) &\leq C_0 \sqrt{2|S|} e^{-t} \sum_{k=1}^{+\infty} (2dne^{-\lambda})^k \\ &\leq C_0 \sqrt{2|S|} e^{-t} \end{aligned}$$

pour λ que l'on fixe enfin tel que $2dne^{-\lambda} = \frac{1}{2}$. Alors pour tout $\varepsilon > 0$, la distance $d(t)$ passe en-dessous de ε au plus au temps $\frac{1}{2} \log |S| + O(1)$.

3.4 La borne inférieure sur le temps de mélange

Dans cette section, la preuve présentée est identique que l'on souhaite montrer le théorème 1 ou 2, à part une preuve d'un lemme que l'on n'écrit pas ici. Afin de montrer la borne inférieure sur le temps de mélange dans le cas de faibles intensités, considérons la statistique

$$f(\omega) = \sum_{x \in S} \mathbb{1}(\omega(x) = n).$$

Sous la distribution stationnaire, f a pour moyenne $\mathbb{E}_\pi(f(X_t)) = \frac{|S|}{n}$. Mais sous la distribution de $X_t^{(n)}$ la chaîne de Markov débutée de tous les votes valant n , pour tout $x \in S$, le vote de x vaut n soit:

- si la marche aléatoire partant de x atteint le temps 0, ce qui arrive avec probabilité e^{-t}
- si la marche aléatoire partant de x meurt avant d'arriver au temps 0, et que le vote aléatoire de x est n , ce qui arrive avec probabilité $(1 - e^{-t})\frac{1}{n}$.

Alors la moyenne de f sous cette distribution est $\mathbb{E}(f(X_t^{(n)})) = e^{-t} + (1 - e^{-t})\frac{1}{n}$, qui est à distance de l'ordre de $|S|e^{-t}$ de la moyenne de f dans le cas de la distribution stationnaire. La clé est de dire que les fluctuations de f sont d'ordre $\sqrt{|S|}$ sous les distributions de X_t et π , et ainsi pour que les deux distributions soient proches, nécessairement e^{-t} doit être au plus de l'ordre de $\frac{1}{\sqrt{|S|}}$ ce qui est exactement la borne voulue. On a donc besoin du lemme suivant:

Lemme 3.4. *Il existe $\gamma > 0$ tel que pour toute distribution initiale et pour tout $t > 0$, on a*

$$\forall x \in S, \quad \sum_{y \in S} \text{cov}(\mathbb{1}(X_t(x) = n), \mathbb{1}(X_t(y) = n)) \leq \gamma.$$

On ne précise pas la dépendance de γ en les paramètres du problème, car on peut montrer ce lemme de plusieurs manières, suivant si on cherche à montrer le théorème 1 ou 2, γ dépend alors de l'ensemble des paramètres du problème. Ce lemme donne immédiatement

$$\text{Var}(f(X_t^{(n)})) \leq \gamma|S|.$$

Etant donné que $X_t^{(n)}$ converge vers la distribution stationnaire π , on en déduit que sous la distribution π , on a aussi

$$\text{Var}_\pi(f(X_t)) \leq \gamma|S|.$$

Alors l'inégalité de Bienaimé-Chebychev donne

$$\begin{aligned} \mathbb{P}(f(X_t^{(n)}) \geq e^{-t}/2) &\leq \alpha|S|e^{-2t} \\ \mathbb{P}_\pi(f(X_t) \leq e^{-t}/2) &\leq \alpha|S| \left(e^{-t} - \frac{2}{n} \right)^2 \end{aligned}$$

pour une certaine constante α . Pour tout $\varepsilon > 0$, on peut choisir $t = \frac{1}{2} \log |S| + O(1)$ tel que chacune des probabilités soit inférieure à $\frac{\varepsilon}{2}$, alors la distance en variation totale entre la distribution de $X_t^{(n)}$ et π est nécessairement supérieure ou égale à $1 - \varepsilon$.

4 Amélioration pour de grandes intensités

Dans cette section, j'explique comment il est possible d'adapter la méthode de la section précédente pour montrer le théorème 2. Dans ce cas, la preuve du lemme 3.3 n'est plus valide, car elle nécessite de pouvoir borner la distribution du cluster contenant un point x donné. Mais dans le cas de "grandes intensités" comme dans le théorème 2, on peut montrer que les tailles des clusters n'ont même pas nécessairement une espérance uniformément bornée. On ne présente que les idées de la preuve pour la borne supérieure sur le temps de mélange, qui est plus difficile et présente le plus de changements par rapport à la partie précédente.

4.1 Quelques résultats simples

Commençons par donner un lemme purement géométrique qui sera utile pour la suite. Ce lemme permet d'adapter la preuve de Lubetzky et Sly dans [4] à des graphes plus généraux que le tore $(\mathbb{Z}/N\mathbb{Z})^d$, notamment ici les graphes vérifiant les conditions de 1.

Lemme 4.1. *Soit $K > 0$. Alors il existe une constante $C = C(K) > 0$ telle que pour tout graphe G vérifiant les conditions de 1, et pour tout $R > 0$, il existe une partition $\mathcal{P}_1, \dots, \mathcal{P}_k$ de G telle que:*

- Pour tout indice i , $|\mathcal{P}_i| \leq KR^{2K}$.
- Pour tout indice i , le nombre d'indices j tels que \mathcal{P}_j intersecte $B(\mathcal{P}_i, R)$ est majoré par C .

On illustre simplement ce lemme avec un exemple. Pour le tore $(\mathbb{Z}/N\mathbb{Z})^d$, pour tout $N > R > 0$, on peut toujours partitionner le tore en pavés dont les côtés sont entre R et R^2 . Alors chaque pavé a une taille bornée par R^{2d} et pour chaque pavé, les pavés à distance au plus R de ce pavé sont ceux qui possèdent un sommet en commun avec lui, il y en a donc au plus 3^d (remarquer que cette constante ne dépend ni de N , ni de R).

On introduit à présent des notations en lien avec la représentation duale du noisy voter model. Considérons donc, à présent, des marcheurs aléatoires partant de chaque sommet, qui coalescent et meurent avec intensité 1. La distribution jointe des marches est simplement une fonction de la variable aléatoire, appelée \mathcal{U} dans la section précédente, dont on rappelle qu'elle contient toutes les informations sur les mises à jour des votes des sommets. Pour des raisons qui deviendront plus claires dans la suite, on renommera cette variable $\widehat{\mathcal{U}}$ dans la suite de la preuve. Pour deux temps $t_1 < t_2$, on notera aussi $\widehat{\mathcal{U}}[t_1, t_2]$ la variable aléatoire définie comme $\widehat{\mathcal{U}}$ mais ne contenant que les informations sur les mises à jour entre les temps t_1 et t_2 .

Définition 3. *Soient $A \subset S$ et $t_1 < t_2$ deux temps. Considérons les deux fonctions suivantes de $\widehat{\mathcal{U}}[t_1, t_2]$:*

- La fonction support de mises à jour $\mathcal{F}_{\text{SUPP}}(A, t_1, t_2)$ qui est l'ensemble des sommets de S au temps t_1 dont dépendent les votes sur A au temps t_2 . En d'autres termes, on obtient cet ensemble en considérant des marcheurs partant de chaque sommet de A au temps t_2 et en remontant le temps jusqu'au temps t_1 , $\mathcal{F}_{\text{SUPP}}(A, t_1, t_2)$ est alors l'ensemble des sommets occupés par les marcheurs au temps t_1 (qui peut être vide si tous les marcheurs meurent). On pourra aussi noter $\mathcal{F}_{\text{SUPP}, \widehat{\mathcal{U}}[t_1, t_2]}(A, t_1, t_2)$ pour expliciter la dépendance déterministe en $\widehat{\mathcal{U}}[t_1, t_2]$.
- La fonction des mises à jour $\mathcal{F}_{\text{UPD}}(A, t_1, t_2)$ qui est l'union sur tous les sous-ensembles de mises à jour $\widehat{\mathcal{U}}' \subset \widehat{\mathcal{U}}[t_1, t_2]$ de $\mathcal{F}_{\text{SUPP}, \widehat{\mathcal{U}}'}(A, t_1, t_2)$.

Ce deuxième objet peut paraître assez peu intuitif, mais on peut le comprendre comme ceci: un marcheur normal change de sommet lorsqu'il rencontre un sommet qui se fait influencer par un autre, et meurt lorsqu'il rencontre une mise à jour aléatoire. Lorsque l'on calcule \mathcal{F}_{UPD} , on peut le calculer comme pour $\mathcal{F}_{\text{SUPP}}$, sauf que le marcheur ne meurt pas lorsqu'il rencontre une mise à jour aléatoire, et lorsqu'il rencontre un sommet x qui se fait influencer par un sommet y , il se dédouble, un marcheur continue sur le sommet x alors que l'autre change pour aller sur le sommet y . La figure 1 montre un exemple de calcul de ces deux fonctions.

À partir de ces deux fonctions, on obtient l'estimation assez simple suivante:

Lemme 4.2. *Pour tout $x \in S$ et $h > 0$, on a:*

- Pour tout $t \geq h$,

$$\mathbb{P}(\mathcal{F}_{\text{SUPP}}(x, t-h, t) \neq \emptyset) = e^{-h}.$$

- Pour tout $t \geq h$ et $l \geq 10Kq_0h$ (où K, q_0 sont définis dans l'énoncé du théorème),

$$\mathbb{P}(\mathcal{F}_{\text{UPD}}(x, t-h, t) \not\subset B(x, l)) \leq e^{1-l}.$$

On peut à présent expliquer l'utilité du lemme 4.1. On va choisir une échelle de temps s_* , qui sera un grand entier, et on applique à présent le lemme 4.1 avec $R = s_*^2$ pour obtenir une partition de S que l'on appellera les "blocs", et que l'on notera par les lettres majuscules A, B . On dira que deux blocs sont adjacents s'ils sont à distance au plus s_*^2 l'un de l'autre. D'après le lemme chaque bloc a au plus C blocs adjacents (où C est défini dans le lemme et ne dépend que de K).

Considérons maintenant un bloc A , qui contient au plus un nombre de sommets polynomial en s_* (au plus Ks_*^{4K} plus précisément). Remarquons que si on souhaite appliquer le lemme précédent avec $h = s_*$, alors pour s_* assez grand, s_*^2 est une valeur acceptable de l pour appliquer le deuxième point. On sait alors que chaque point $x \in A$, avec une probabilité au moins $1 - e^{1-s_*^2}$, a une fonction de mise à jour qui reste dans la boule $B(x, s_*^2)$, et donc dans les blocs adjacents à A . Alors la probabilité que $\mathcal{F}_{\text{UPD}}(A, t_1, t_2)$ reste dans les blocs adjacents à A est au moins $1 - Ks_*^{4K}e^{1-s_*^2}$. Lorsque s_* est grand, cette probabilité est très proche de 1, et ainsi avec grande probabilité la fonction mise à jour d'un bloc reste dans les blocs adjacents pendant un temps s_* . Intuitivement, même si les intensités $q(x, y)$ sont “grandes”, à l'échelle des blocs, les marcheurs ne peuvent pas partir loin dans un temps s_* .

Maintenant, posons $t_* = \frac{1}{2} \log |S| + s_*$. On va chercher à montrer que pour des valeurs de s_* assez grandes (indépendantes de $|S|$), le système sera suffisamment mélangé au temps t_* . On divise alors l'intervalle de temps $[0, t_*]$ en les intervalles $[0, \frac{1}{2} \log |S|]$ et $]\frac{1}{2} \log |S|, t_*]$, que l'on appelle les intervalles *pré-mélange* et *post-mélange* respectivement.

4.2 L'intervalle post-mélange

On va définir les notions permettant d'étudier l'histoire d'une configuration dans l'intervalle post-mélange $]\frac{1}{2} \log |S|, t_*]$. Commençons par choisir un grand entier λ , on prendra ensuite s_* grand devant λ . On divise alors l'intervalle post-mélange en λ intervalles de taille égale $]\tau_{k-1}, \tau_k]$, $k \in \{0, 1, \dots, \lambda\}$, avec

$$\tau_k = \frac{1}{2} \log |S| + k \frac{s_*}{\lambda}.$$

On divise à nouveau chacun de ces intervalles en deux intervalles $\mathcal{I}_k =]\tau_{k-1}, \tau_k - 1]$ et $\mathcal{I}'_k =]\tau_k - 1, \tau_k]$. On nomme ces intervalles les phases *régulières* et *déférées* respectivement.

On définit à présent l'historique d'un sommet $x \in S$ dans l'intervalle post-mélange. Pour tout $t \in]\frac{1}{2} \log |S|, t_*]$, l'historique au temps t de x est un sous-ensemble $\mathcal{H}_x(t)$ de S défini par $\mathcal{H}_x(t_*) = \{x\}$ et on définit successivement:

- Pour $t \in \mathcal{I}'_k$, $\mathcal{H}_x(t)$ est donné par

$$\mathcal{H}_x(t) = \mathcal{F}_{\text{UPD}}(\mathcal{H}_x(\tau_k^+), t, \tau_k).$$

- Pour $t \in \mathcal{I}_k$, $\mathcal{H}_x(t)$ est donné par

$$\mathcal{H}_x(t) = \mathcal{F}_{\text{SUPP}}(\mathcal{H}_x((\tau_k - 1)^+), t, \tau_k - 1).$$

En d'autres termes, on développe l'historique de x en revenant dans le temps, en utilisant la fonction support de mises à jour dans les phases régulières, et la fonction des mises à jour dans les phases différées. Naturellement, on aurait voulu utiliser simplement la fonction $\mathcal{F}_{\text{SUPP}}$ pour définir l'historique. Mais l'intérêt des phases *déférées* apparaîtra plus tard dans la preuve. On peut aussi voir l'historique du point x comme un ensemble de chemins $P(t)$ qui démarrent à $P(t_*) = x$ et qui remontent le temps, comme les marcheurs définis précédemment, mais dans les phases différées, ils ne meurent pas et branchent lors d'une influence d'un sommet par un autre.

Considérons $\widehat{\mathcal{U}}]_{\frac{1}{2} \log |S|, t_*]$ qui contient toutes les informations sur les mises à jour dans l'intervalle post-mélange. On définit \mathcal{U}' le sous-ensemble constitué des mises à jour aléatoires dans les phases différées. Remarquons que l'historique est indépendant de \mathcal{U}' (car la fonction \mathcal{F}_{UPD} n'est pas influencée par la présence de mises à jour aléatoires), et est ainsi une fonction de \mathcal{U} le complémentaire de \mathcal{U}' dans $\widehat{\mathcal{U}}]_{\frac{1}{2} \log |S|, t_*]$. On appelle \mathcal{U}' *l'aléatoire déferé* et \mathcal{U} *l'aléatoire non déferé*. Plus tard dans la preuve, on conditionnera ensuite sur \mathcal{U} , le fait que l'on n'ait pas conditionné sur \mathcal{U}' sera très important. Notons que \mathcal{U} et \mathcal{U}' sont des variables aléatoires indépendantes.

Considérons maintenant Υ l'ensemble des sommets dans S dont l'historique survit jusqu'au temps $\frac{1}{2} \log |S|$ (c'est-à-dire qu'il est non vide à ce temps, ou encore qu'il y a un chemin P partant de x au temps t_* et survivant jusqu'au temps $\frac{1}{2} \log |S|$). Alors les sommets en dehors de Υ ont des votes qui ne peuvent pas dépendre de la condition initiale. On regroupe les sommets de Υ comme suit:

Définition 4. On définit la relation \sim sur Υ par $x \sim y$ s'il existe un chemin P dans l'historique de x et un chemin Q dans l'historique de y qui atteignent tous les deux le temps $\frac{1}{2} \log |S|$, ainsi que deux temps $t_x, t_y \in]\frac{1}{2} \log |S|, t_*]$, tels que $P(t_x)$ et $Q(t_y)$ soient dans deux blocs adjacents (ou dans le même bloc).

On étend alors \sim à sa clôture transitive, les classes d'équivalence pour cette relation sont les $(\Upsilon_i)_i$.

Notons \mathcal{H}_{Υ_i} l'historique de Υ_i , c'est-à-dire la collection des historiques des points de Υ_i . Alors pour tout i on note A_i l'ensemble des blocs couvrant $\mathcal{H}_{\Upsilon_i}(\frac{1}{2} \log |S|)$ et B_i l'ensemble des blocs couvrant $\mathcal{H}_{\Upsilon_i}(t_*) = \Upsilon_i$, les A_i et les B_i s'appellent les *composantes de blocs*, notons que ce ne sont pas des blocs mais des ensembles de blocs. On utilisera toujours A, B pour désigner des blocs, les notations A_i, B_i seront utilisées pour désigner des composantes de blocs. Notons aussi \mathcal{A} l'ensemble des composantes de blocs A_i .

On définit les composantes de blocs de cette manière pour deux raisons:

- Deux composantes de blocs différentes ont des historiques "distants", donc qui sont approximativement indépendantes.
- Différents blocs d'une même composante de blocs doivent avoir des historiques qui doivent tous se "regrouper" spatialement, ce qui arrive avec une faible probabilité si les blocs sont nombreux ou loins les uns des autres.

En rendant ces arguments rigoureux, on obtient le résultat suivant: pour deux ensembles de blocs R, S , on note $R \leftrightarrow S$ l'évènement qu'il existe i avec $R = A_i$ et $S = B_i$. Alors,

Lemme 4.3. Pour tous n, K, q_0 , il existe une constante $c > 0$, telle que pour λ assez grand, pour s_* assez grand comparé à λ , et pour toute collection de paires distinctes d'ensemble de blocs $(R_i, S_i)_i$, on a

$$\mathbb{P} \left(\bigcap_i \{R_i \leftrightarrow S_i\} \right) \leq \exp \left(-c \frac{s_*}{\lambda} \sum_i \mathfrak{W}(R_i \cup S_i) \right)$$

où pour un ensemble de blocs R , $\mathfrak{W}(R)$ est la plus petite taille d'un ensemble de blocs connexe (pour la relation d'adjacence) contenant R .

L'idée de la preuve du théorème est alors la suivante. On verra dans la suite que la distance en variation totale $d(t_*)$ est majorée par $\mathbb{E}(\bar{d}(t_*, \mathcal{U}))$ où l'espérance est sur \mathcal{U} et

$$\bar{d}(t_*, \mathcal{U}) = \sup_{\omega_0, \omega'_0} \|\mathbb{P}_{\omega_0}(X_{t_*} \in \cdot | \mathcal{U}) - \mathbb{P}_{\omega'_0}(X_{t_*} \in \cdot | \mathcal{U})\|_{TV}.$$

Mais conditionnellement à \mathcal{U} , X_{t_*} est une fonction de $X_{\frac{1}{2} \log |S|}(\mathcal{A})$ les votes sur l'union des composantes de blocs (qui est une fonction de \mathcal{U}), et de \mathcal{U}' . Selon le couplage monotone, la valeur de \mathcal{U}' est indépendante de la configuration initiale de votes, et donc la distance en variation totale entre les distributions conditionnelles de X_{t_*} est majorée par la distance en variation totale entre les distributions conditionnelles de $X_{\frac{1}{2} \log |S|}(\mathcal{A})$, et donc

$$\bar{d}(t_*, \mathcal{U}) \leq \sup_{\omega_0, \omega'_0} \|\mathbb{P}_{\omega_0}(X_{\frac{1}{2} \log |S|}(\mathcal{A}) \in \cdot | \mathcal{U}) - \mathbb{P}_{\omega'_0}(X_{\frac{1}{2} \log |S|}(\mathcal{A}) \in \cdot | \mathcal{U})\|_{TV}.$$

Cependant, cette inégalité n'est pas assez forte pour démontrer le résultat voulu. Pour améliorer ceci, on se place plutôt à un temps bien choisi où l'historique de Υ est "petit" (en un sens à définir). Pour chaque composante A_i , on cherche un temps τ_k où l'historique correspondant à A_i est petit. On pose alors, pour $1 \leq k \leq \lambda$:

$$\begin{cases} \chi_{i,k} &= \mathcal{H}_{\Upsilon_i}(\tau_k) \\ \Xi_{i,k} &= \prod_{x \in \chi_{i,k}} \frac{2n}{T_{x,k}} \end{cases}$$

où $T_{x,k}$ est le temps entre la dernière influence au sommet x dans la phase déferée \mathcal{I}'_k et τ_k (ou 1 si x ne se fait pas influencer dans l'intervalle), ou encore

$$T_{x,k} = \min(\tau_k - \sup\{t < \tau_k | x \text{ se fait influencer au temps } t\}, 1)$$

où par convention un supremum vide vaut $-\infty$. On verra dans le coeur de la preuve que l'on cherche à trouver des temps τ_k où les $\Xi_{i,k}$ sont petits. Posons alors k_i la valeur de k qui minimise $\Xi_{i,k}$. Posons aussi $\Xi_i = \Xi_{i,k_i}$ et $\chi_i = \chi_{i,k_i}$, le cut-set correspondant à A_i est alors le couple (k_i, χ_i) .

Remarquons que pour tout k , $\Xi_{i,k}$ est au moins égal à $\frac{2n}{T}$ avec T une variable exponentielle. Mais alors $\Xi_{i,k}$ ne peut pas être d'espérance finie. Cependant, cette variable a toujours des moments d'ordre inférieurs à 1, et en prenant le minimum de plusieurs telles variables, on arrive à avoir des variables Ξ_i de moments arbitrairement grands. Plus précisément:

Lemme 4.4. *Soient $n \geq 2, K, q_0 > 0$. Pour tout ensemble de blocs V , on pose*

$$\begin{aligned}\chi_k(V) &= \mathcal{H}_V(\tau_k) \\ \Xi_k(V) &= \prod_{x \in \chi_k(V)} \frac{2n}{T_{x,k}}\end{aligned}$$

où $T_{x,k}$ est, comme précédemment, le temps entre τ_k et la dernière influence à x dans la phase déferée \mathcal{I}'_k (ou 1 s'il y en a pas).

Si λ est assez large en fonction de n, K, q_0 et s_* est assez large en fonction de λ , alors on a

$$\mathbb{E} \left(\left(\min_k \Xi_k(V) \right)^4 \right) \leq 2^{\lambda+3} e^{|V|}.$$

On insiste sur le fait que les composantes de blocs A_i, B_i , ainsi que les $\chi_{i,k}, \Xi_{i,k}$, sont des fonctions déterministes de l'aléatoire non déferé \mathcal{U} .

4.3 L'intervalle pré-mélange

On s'intéresse maintenant à l'historique d'une configuration dans l'intervalle pré-mélange $[0, \frac{1}{2} \log |S|]$ conditionnellement à l'aléatoire non déferé \mathcal{U} (dont on rappelle qu'il ne contient que des informations sur les mises à jour dans l'intervalle post-mélange). Dans cet intervalle de temps, on définit l'historique d'un point $x \in S$ de manière naturelle, à partir de sa définition dans l'intervalle post-mélange:

$$\forall t \leq \frac{1}{2} \log |S|, \quad \mathcal{H}_x(t) = \mathcal{F}_{\text{SUPP}} \left(\mathcal{H}_x \left(\left(\frac{1}{2} \log |S| \right)^+ \right), t, \frac{1}{2} \log |S| \right).$$

Tout comme on a regroupé les blocs au temps t_* pour obtenir les composantes de blocs B_i , on regroupe à présent les composantes de blocs A_i au temps $\frac{1}{2} \log |S|$ afin d'obtenir les clusters comme dans la section 3.

Définition 5. *Pour deux composantes de blocs A_i, A_j , on écrit $A_i \sim A_j$ si l'une des deux conditions suivantes est satisfaite:*

- *Il existe deux points $x \in A_i, y \in A_j$ dont les historiques se rencontrent, c'est-à-dire qu'il y a un temps $t \in [0, \frac{1}{2} \log |S|]$ avec $\mathcal{H}_x(t) \cap \mathcal{H}_y(t) \neq \emptyset$.*
- *Il existe soit un point $x \in A_i$ et un temps $t \in [\frac{1}{2} \log |S| - s_*, \frac{1}{2} \log |S|]$ tel que $\mathcal{H}_x(t) \cap B(A_j, \frac{s_*^2}{3}) \neq \emptyset$, ou un point $y \in A_j$ et un temps $t \in [\frac{1}{2} \log |S| - s_*, \frac{1}{2} \log |S|]$ tel que $\mathcal{H}_y(t) \cap B(A_i, \frac{s_*^2}{3}) \neq \emptyset$.*

Encore une fois, on étend cette relation à sa clôture transitive, on définit les clusters $\mathcal{C}^{(j)}$ comme les classes d'équivalence de la relation obtenue.

Notons que les clusters sont des ensembles de composantes de blocs, qui sont eux mêmes des ensembles de blocs. On dit alors qu'un cluster \mathcal{C} est

- *bleu* s'il consiste en une unique composante A_i , et que son historique meurt avant d'atteindre le temps $\frac{1}{2} \log |S| - s_*$, sans quitter la boule $B(A_i, \frac{s_*^2}{3})$. En d'autres termes:

$$\mathcal{H}_{A_i} \left(\frac{1}{2} \log |S| - s_* \right) = \emptyset, \quad \bigcup_{\frac{1}{2} \log |S| - s_* < t < \frac{1}{2} \log |S|} \mathcal{F}_{\text{SUPP}} \left(A_i, t, \frac{1}{2} \log |S| \right) \subset B \left(A_i, \frac{s_*^2}{3} \right).$$

- *rouge* si son historique atteint le temps 0, c'est-à-dire

$$\mathcal{F}_{\text{SUPP}} \left(\mathcal{C}, 0, \frac{1}{2} \log |S| \right) \neq \emptyset$$

(où, par abus de notation, on a appliqué $\mathcal{F}_{\text{SUPP}}$ à \mathcal{C} en identifiant \mathcal{C} à l'ensemble des sommets contenus dans un bloc dans une composante de \mathcal{C}).

- *vert* sinon (son historique n'atteint pas le temps 0, mais survit "longtemps" ou voyage loin).

Comme dans la section 3, on va chercher à montrer qu'il y a peu de clusters rouges par rapport aux clusters bleus, conditionnellement à l'historique des clusters verts. Posons alors $\mathcal{H}_{\text{VERT}}$ l'historique de tous les clusters verts.

Pour un ensemble déterministe (toujours conditionnellement à \mathcal{U}) de composantes de blocs \mathcal{C} (qui est un cluster avec une certaine probabilité), on définit l'historique de l'extérieur

$$\mathcal{H}_{\mathcal{C}}^{-}(t) = \bigcup_{x \in A \in A_i \notin \mathcal{C}} \mathcal{H}_x(t).$$

On dit qu'une certaine valeur \mathcal{X} de $\mathcal{H}_{\mathcal{C}}^{-}$ est \mathcal{C} -compatible si conditionnellement à $\mathcal{H}_{\mathcal{C}}^{-} = \mathcal{X}$, \mathcal{C} est un cluster avec probabilité positive (c'est équivalent à dire que $\mathcal{H}_{\mathcal{C}}^{-}$ ne pénètre pas dans la boule $B(\mathcal{C}, \frac{s_*}{3})$ dans l'intervalle de temps $[\frac{1}{2} \log |S| - s_*, \frac{1}{2} \log |S|]$). De manière analogue à la section 3, on définit, pour \mathcal{C} un ensemble de composantes de blocs

$$\Psi_{\mathcal{C}} = \sup_{\mathcal{X}} \mathbb{P}(\mathcal{C} \text{ est un cluster rouge} | \mathcal{H}_{\mathcal{C}}^{-} = \mathcal{X}, \mathcal{C} \text{ est soit un cluster rouge, soit une union de clusters bleus}, \mathcal{U})$$

où le supremum est pris sur tous les historiques \mathcal{C} -compatibles. Remarquons que $\Psi_{\mathcal{C}}$ est une fonction de \mathcal{U} .

Dénotons par $[\cdot]^+$ la partie positive d'un réel. On rappelle que pour un ensemble de blocs V , $\mathfrak{W}(V)$ est la plus petite taille d'un ensemble de blocs connexe contenant V , par abus de notation on notera $\mathfrak{W}(\mathcal{C}) = \mathfrak{W}(\cup \mathcal{C})$ lorsque \mathcal{C} est un ensemble de composantes de blocs. Alors on a la borne suivante:

Lemme 4.5. *Soient $n \geq 2, K, q_0 > 0$. Alors il existe une constante $c = c(K, q_0) > 0$ (indépendante de n) telle que pour tout s_* suffisamment large en fonction de K, q_0 , on a (en rappelant que \mathcal{A} est l'ensemble des composantes de blocs)*

$$\forall \mathcal{C} \subset \mathcal{A}, \quad \Psi_{\mathcal{C}} \leq \frac{g(2s_*^4)}{\sqrt{|S|}} \exp \left(4 \sum_i |A_i| - cs_* \left[\mathfrak{W}(\mathcal{C}) - \sum_{A_i \in \mathcal{C}} \mathfrak{W}(A_i) \right]^+ \right).$$

Remarquons que les deux côtés de cette inégalité sont des fonctions de \mathcal{U} .

4.4 Preuve de la borne supérieure sur le temps de mélange

Dans cette partie, on utilise les résultats précédents pour démontrer, comme dans la section 3, une borne supérieure sur le temps de mélange de la chaîne de Markov. Tout d'abord, une vérification simple valide pour toute chaîne de Markov donne l'inégalité $d(t_*) \leq \bar{d}(t_*)$ où

$$\bar{d}(t_*) = \sup_{\omega_0, \omega'_0} \|\mathbb{P}_{\omega_0}(X_{t_*} \in \cdot) - \mathbb{P}_{\omega'_0}(X_{t_*} \in \cdot)\|_{TV}.$$

On utilise l'inégalité de Jensen pour conditionner par rapport à l'aléatoire non déferé \mathcal{U} afin d'obtenir $\bar{d}(t_*) \leq \mathbb{E}(\bar{d}(t_*, \mathcal{U}))$ avec

$$\bar{d}(t_*, \mathcal{U}) = \sup_{\omega_0, \omega'_0} \|\mathbb{P}_{\omega_0}(X_{t_*} \in \cdot | \mathcal{U}) - \mathbb{P}_{\omega'_0}(X_{t_*} \in \cdot | \mathcal{U})\|_{TV}.$$

On remarque à présent que conditionnellement à \mathcal{U} , la distribution de X_{t_*} se retrouve à partir de la distribution de X sur les cut-sets χ_i et de \mathcal{U}' . En effet, pour un sommet x , soit son historique n'atteint pas le temps

$\frac{1}{2} \log |S|$ et dans ce cas-là le vote de x est une fonction de \mathcal{U}' , soit x fait partie d'un Υ_i et son historique peut être retrouvé à partir des votes $X_{\tau_{k_i}}(\chi_i)$ et de \mathcal{U}' . Notons Z_χ la variable aléatoire $(X_{\tau_{k_i}}(\chi_i))_i$, on peut alors écrire $X_{t_*} = F_{\mathcal{U}}(Z_\chi, \mathcal{U}')$ pour une certaine fonction $F_{\mathcal{U}}$ dépendant de \mathcal{U} . Sous le couplage monotone, la variable \mathcal{U}' est indépendante de la condition initiale, donc on en déduit

$$\bar{d}(t_*, \mathcal{U}) \leq \sup_{\omega_0, \omega'_0} \|\mathbb{P}_{\omega_0}(Z_\chi \in \cdot | \mathcal{U}) - \mathbb{P}_{\omega'_0}(Z_\chi \in \cdot | \mathcal{U})\|_{TV}.$$

On conditionne maintenant, comme dans la section 3, par l'historique des clusters verts $\mathcal{H}_{\text{VERT}}$. En appliquant l'inégalité de Jensen puis en remplaçant l'espérance sur $\mathcal{H}_{\text{VERT}}$ par un supremum, on a

$$\bar{d}(t_*, \mathcal{U}) \leq \mathbb{E} \left(\sup_{\mathcal{X}} \sup_{\omega_0, \omega'_0} \|\mathbb{P}_{\omega_0}(Z_\chi \in \cdot | \mathcal{H}_{\text{VERT}} = \mathcal{X}, \mathcal{U}) - \mathbb{P}_{\omega'_0}(Z_\chi \in \cdot | \mathcal{H}_{\text{VERT}} = \mathcal{X}, \mathcal{U})\|_{TV} \middle| \mathcal{U} \right).$$

Comme dans la section 3, on se ramène à la distribution des votes sur les clusters rouges et bleus. Pour faire ceci, remarquons que conditionnellement à l'ensemble des clusters verts, la distribution de Z_χ est donnée par le couple (Z'_χ, Z''_χ) , où Z'_χ est la distribution de $(X_{\tau_{k_i}}(\chi_i))_i$ sur les clusters rouges et bleus, et Z''_χ est la distribution de $(X_{\tau_{k_i}}(\chi_i))_i$ sur les clusters verts. Mais comme l'historique des clusters verts n'atteint pas le temps 0, sous le couplage monotone, Z''_χ est indépendant de la condition initiale. Ainsi, on a

$$\bar{d}(t_*, \mathcal{U}) \leq \mathbb{E} \left(\sup_{\mathcal{X}} \sup_{\omega_0, \omega'_0} \|\mathbb{P}_{\omega_0}(Z'_\chi \in \cdot | \mathcal{H}_{\text{VERT}} = \mathcal{X}, \mathcal{U}) - \mathbb{P}_{\omega'_0}(Z'_\chi \in \cdot | \mathcal{H}_{\text{VERT}} = \mathcal{X}, \mathcal{U})\|_{TV} \middle| \mathcal{U} \right).$$

Une difficulté supplémentaire par rapport à l'étude de la section 3 est que la distribution des votes sur les clusters bleus n'est pas aussi simple qu'une distribution uniforme. On définit donc, pour tout i tel que A_i n'appartienne pas à un cluster vert, la distribution sur le cut-set correspondant dans le cas où $\{A_i\}$ est un cluster bleu:

$$\nu_i = (X_{k_i}(\chi_i) \in \cdot | \mathcal{H}_{\text{VERT}}, \mathcal{U}, \{A_i\} \text{ est un cluster bleu}).$$

Si tous les clusters étaient bleus, la distribution des votes serait donc la distribution produit

$$\nu = \prod_{A_i \notin \mathcal{A}_{\text{VERT}}} \nu_i$$

avec $\mathcal{A}_{\text{VERT}}$ l'ensemble des composantes de blocs appartenant à des clusters verts. Par inégalité triangulaire et inégalité de Cauchy-Schwarz, on obtient alors

$$\bar{d}(t_*, \mathcal{U}) \leq \mathbb{E} \left(\sup_{\mathcal{X}} \sup_{\omega_0} [\min(1, \|\mathbb{P}_{\omega_0}(Z'_\chi \in \cdot | \mathcal{H}_{\text{VERT}} = \mathcal{X}, \mathcal{U})/\nu - 1\|_{L^2(\nu)})] \middle| \mathcal{U} \right).$$

La norme L^2 est alors estimée par une généralisation du lemme 3.1 à des mesures produits non uniformes:

Lemme 4.6. *Soit S un ensemble fini, et $(V_i)_{i \in \mathcal{I}}$ une partition fixée de S . Pour tout i , soit ν_i une mesure sur $[n]^{V_i}$ et soit $\nu = \prod_{i \in \mathcal{I}} \nu_i$ la mesure produit correspondante sur $[n]^S$. Pour tout $\mathcal{J} \subset \mathcal{I}$, soit $\varphi_{\mathcal{J}}$ une mesure sur $[n]^{\cup_{j \in \mathcal{J}} V_j}$. Soit μ une mesure sur $[n]^S$ obtenue en choisissant un $\mathcal{J} \subset \mathcal{I}$ suivant une mesure μ_0 sur $\mathcal{P}(\mathcal{I})$, puis en choisissant les valeurs sur $\cup_{j \in \mathcal{J}} V_j$ via $\varphi_{\mathcal{J}}$, et les valeurs sur les autres V_i via des échantillons indépendants de lois ν_i . Alors*

$$\|\mu/\nu - 1\|_{L^2(\nu)} \leq \left[\sum_{\mathcal{J}, \mathcal{J}'} \left(\prod_{i \in \mathcal{J} \cap \mathcal{J}'} \min_{\omega_i \in [n]^{V_i}} \nu_i(\omega_i) \right)^{-1} \mu_0(\mathcal{J}) \mu_0(\mathcal{J}') - 1 \right]^{\frac{1}{2}}.$$

Dans notre cas, l'application du lemme fait apparaître les valeurs

$$\nu_i^* = \min_{\omega_i \in [n]^{V_i}} \nu_i(\omega_i)$$

et le lemme donne

$$\|\mathbb{P}_{\omega_0}(Z'_\chi \in \cdot | \mathcal{H}_{\text{VERT}} = \mathcal{X}, \mathcal{U}) / \nu - 1\|_{L^2(\nu)} \leq \left[\mathbb{E} \left(\prod_{A_i \in \mathcal{A}_{\text{ROUGE}} \cap \mathcal{A}'_{\text{ROUGE}}} \frac{1}{\nu_i^*} \middle| \mathcal{H}_{\text{VERT}} = \mathcal{X}, \mathcal{U} \right) - 1 \right]^{\frac{1}{2}}$$

où $\mathcal{A}_{\text{ROUGE}}, \mathcal{A}'_{\text{ROUGE}}$ sont deux copies i.i.d. de l'union des clusters rouges. Maintenant, tout comme pour le lemme 3.2, la définition des Ψ_A permet des couplages judicieux qui donnent l'inégalité suivante:

$$\mathbb{E} \left(\prod_{A_i \in \mathcal{A}_{\text{ROUGE}} \cap \mathcal{A}'_{\text{ROUGE}}} \frac{1}{\nu_i^*} \middle| \mathcal{H}_{\text{VERT}} = \mathcal{X}, \mathcal{U} \right) - 1 \leq \exp \left(\sum_{\mathcal{C} \cap \mathcal{C}' \neq \emptyset} \Psi_{\mathcal{C}} \Psi_{\mathcal{C}'} \prod_{A_i \in \mathcal{C}} \frac{1}{\nu_i^*} \prod_{A_i \in \mathcal{C}'} \frac{1}{\nu_i^*} \right).$$

Ainsi, on enlève le conditionnement sur $\mathcal{H}_{\text{VERT}}$, mais remarquons que le côté droit dépend encore de \mathcal{U} .

On sait comment borner les $\Psi_{\mathcal{C}}$, il reste à minorer les ν_i^* . Pour ceci, on va utiliser le fait que l'on n'ait pas conditionné sur \mathcal{U}' l'aléatoire déferé. En effet, pour chaque i , n'importe quelle configuration $\omega_i \in [n]^{\chi_i}$ apparaît avec probabilité positive sous ν_i , car tout point de χ_i peut recevoir une mise à jour aléatoire juste avant le temps τ_{k_i} , et cette information est contenue dans \mathcal{U}' . Plus précisément, il faut que chaque $x \in \chi_i$ reçoive une mise à jour aléatoire dans l'intervalle de temps $]\tau_{k_i} - T_{x, k_i}, \tau_{k_i}]$, ce qui arrive avec probabilité $(1 - \exp(-T_{x, k_i})) \geq \frac{1}{2} T_{x, k_i}$, et alors avec probabilité $\frac{1}{n}$, x obtient le vote voulu. On en déduit donc que

$$\nu_i^* \geq \prod_{x \in \xi_i} \frac{1}{2n} T_{x, k_i} = \frac{1}{\Xi_i}$$

et ainsi, en combinant toutes les étapes du raisonnement,

$$\bar{d}(t_*) \leq 2 \mathbb{E} \left(\sum_{\substack{\mathcal{C}, \mathcal{C}' \subset \mathcal{A} \setminus \mathcal{A}_{\text{GREEN}} \\ \mathcal{C} \cap \mathcal{C}' \neq \emptyset}} \Psi_{\mathcal{C}} \Psi_{\mathcal{C}'} \prod_{A_i \in \mathcal{C}} \Xi_i \prod_{A_i \in \mathcal{C}'} \Xi_i \right).$$

Pour l'instant, toutes les bornes obtenues étaient conditionnelles à \mathcal{U} , et on prend à présent l'espérance. On peut enfin utiliser les bornes sur les $\Psi_{\mathcal{C}}$ et les Ξ_i , et alors en utilisant des arguments géométriques sur le graphe, on peut obtenir la borne suivante

$$d(t_*) \leq 2 \exp \left(-c \frac{s_*}{\lambda} \right)$$

pour une constante c appropriée. Ainsi, en prenant s_* arbitrairement grand, $d(t_*)$ passe en dessous de n'importe quel $\varepsilon > 0$ fixé à l'avance.

5 Autres problèmes étudiés

Lors de mon stage, j'ai étudié d'autres problèmes, mais ces projets n'ont pas pu être menés à bout.

5.1 Temps de mélange pour une dynamique du modèle de Potts

On considère ici le modèle de Potts qui est une généralisation du modèle d'Ising. Considérons un graphe non orienté fini $G = (S, A)$, et $q \geq 2$ un entier. Alors le modèle de Potts est un élément aléatoire de $\Omega = \{1, \dots, q\}^S$, c'est-à-dire que chaque sommet de G a un spin entre 1 et q . La loi de probabilité sur Ω est donnée par, pour toute configuration $\omega \in \Omega$,

$$\mu_{q, \beta}(\omega) = \frac{1}{Z_{q, \beta}} \exp \left(\beta \sum_{\{x, y\} \in A} \delta_{\omega(x), \omega(y)} \right)$$

avec $Z_{q, \beta}$ une constante de normalisation. Pour $q = 2$, on obtient le modèle d'Ising.

On considère maintenant la dynamique de Glauber pour ce modèle, qui est une chaîne de Markov à temps continu sur Ω . Chaque sommet, suivant un processus de Poisson d'intensité 1, change de vote suivant

la loi $\mu_{q,\beta}$ conditionnellement aux sommets extérieurs à lui. La question est alors de comprendre le temps de mélange $t_{mix}(G, q, \beta)$ de cette chaîne de Markov. Le problème était de savoir s’il existait une constante $c > 0$ universelle telle que pour tous G, q, β , on ait

$$t_{mix}(G, q, \beta) \geq c \log |S|.$$

Pour le modèle d’Ising, c’est-à-dire lorsque $q = 2$, le résultat a été démontré par Ding et Peres dans [6], avec $c = \frac{1}{2} + o(1)$ lorsque $|S| \rightarrow +\infty$. Leur preuve est très courte et n’utilise presque aucune propriété du modèle, à part son caractère monotone: si on considère deux chaînes de Markov X, X' telles que $X_0 \leq X'_0$, on peut les coupler telles que pour tout t , $X_t \leq X'_t$ (c’est le couplage “monotone”, comme celui étudié pour le noisy voter model). Plus généralement, dans [7], il a été montré que la borne avec $c = \frac{1}{2} + o(1)$ est valide pour n’importe quelle chaîne de Markov qui possède une propriété de monotonie. Cependant, dans le cas du modèle de Potts, on n’a pas de couplage monotone comme on pourrait le faire dans le cas du noisy voter model, donc on ne peut pas conclure.

Enfin, pour des dynamiques *single-site*, c’est-à-dire où la modification d’un spin se fait en regardant seulement ses voisins, un résultat de Hayes et Sinclair dans [8] montre qu’il existe une constante $c(\Delta)$ dépendant du degré maximal Δ de G telle que $t_{mix} \geq c(\Delta) \log |S|$. On obtient ainsi le résultat voulu sous la condition que le degré de G soit borné.

Cependant, on s’attend à ce que la dynamique de Glauber pour le modèle de Potts ait tout de même certaines propriétés de monotonie: intuitivement, “le plus il y a de spins q , le plus chaque spin voudrait devenir un spin q ”.

J’ai montré que si on considérait la loi de probabilité $\bar{\mu}$ sur $\bar{\Omega} = \{0, 1\}^S$ donnée par le poussé en avant de $\mu_{q,\beta}$ par l’application qui à chaque sommet associe 1 si le spin à ce sommet est q et 0 sinon, alors la dynamique de Glauber sur ce nouvel espace probabilisé est bien monotone. Ainsi, cette dynamique de Glauber a bien un temps de mélange au moins égal à $(\frac{1}{2} + o(1)) \log |S|$. Cependant, il n’est pas clair comment passer de la dynamique “projetée” à la dynamique originelle, et ainsi je n’ai pas pu conclure à partir de ce point.

5.2 Marche aléatoire sur une percolation dynamique

Le dernier problème que j’ai étudié dans mon stage concerne les marches aléatoires sur une percolation dynamique. Considérons le tore $S = (\mathbb{Z}/N\mathbb{Z})^d$ et le graphe $G = (S, A)$ donné par le graphe canonique sur \mathbb{Z}^d . On notera $n = |S| = N^d$ la taille du tore. Choisissons $0 < p < 1$ et $\mu > 0$. Un *environnement* sera un élément $\eta \in \{0, 1\}^A$, qui détermine si chaque arête est ouverte ($\eta(e) = 1$) ou fermée ($\eta(e) = 0$). On considère la percolation dynamique η_t de paramètre p qui est une chaîne de Markov sur l’ensemble des environnements obtenue en considérant un processus de Poisson d’intensité μ sur chaque arête de A (tous indépendants les uns des autres). A chaque point de ce processus, l’arête s’ouvre avec probabilité p et se ferme avec probabilité $1 - p$, indépendamment de toute autre mise à jour. Ainsi, à partir du moment où toutes les arêtes ont été mises à jour au moins une fois, la distribution de l’environnement est celle d’une percolation de paramètre p sur le tore.

Considérons maintenant un marcheur aléatoire $X_t \in S$ qui se déplace sur le tore de la façon suivante: suivant un processus de Poisson de paramètre 1, il évolue en choisissant un de ses voisins dans le tore, puis en se déplaçant vers ce voisin si l’arête est ouverte dans η_t , et en restant sur place si l’arête est fermée dans η_t . De cette façon, si l’on fixait l’environnement, une distribution stationnaire de X_t serait la distribution uniforme sur S . Il est facile de voir que (X_t, η_t) est une chaîne de Markov (alors que X_t seul ne l’est pas !), de distribution stationnaire $U(S) \otimes \pi_p$ le produit tensoriel de la distribution uniforme sur S et de la percolation de paramètre p sur G .

La question est alors de trouver le temps de mélange de cette chaîne de Markov, initialement introduite dans [9]. Pour ceci, on s’autorise à modifier μ avec n . Le cas intéressant est celui où $\mu \rightarrow 0$ lorsque $n \rightarrow +\infty$, on se placera donc toujours dans ce cas. On distingue deux cas, selon la position relative de p par rapport à la valeur de la probabilité critique $p_c(d)$ dans \mathbb{Z}^d .

Dans [10], il est montré que le temps de mélange $t_{mix}(\varepsilon)$ est toujours supérieur à $c(n^2 + \frac{1}{\mu})$ pour ε assez petit par rapport à p et c une certaine constante positive. De plus, dans le cas où $p < p_c(d)$, il est montré plus précisément qu’il existe $C = C(d, \varepsilon, p) > 0$ tel que dès que le côté N du tore est assez grand, on a

$$\frac{n^2}{C\mu} \leq t_{mix}(\varepsilon) \leq \frac{Cn^2}{\mu}.$$

Le cas de $p > p_c(d)$ n'est pas encore totalement résolu. On s'attend à ce que ce temps de mélange soit plutôt de l'ordre de $n^2 + \frac{1}{\mu}$, ce qui est la borne inférieure trouvée dans le cas général. Dans le papier [11], il a été montré que lorsque $\theta(p) > \frac{1}{2}$ (où l'on rappelle que $\theta(p)$ est la probabilité qu'un point donné appartienne à la composante infinie de la percolation dans \mathbb{Z}^d), alors on a le résultat à un facteur polylogarithmique près, c'est-à-dire que $t_{mix}(\varepsilon) \leq (n^2 + \frac{1}{\mu})(\log n)^\alpha$ avec α dépendant de d, p .

Afin d'obtenir le résultat sans ce facteur polylogarithmique, et surtout sans la condition $\theta(p) > \frac{1}{2}$, il serait nécessaire d'obtenir le résultat suivant:

Conjecture 1. *Pour tous $d, p > p_c(d)$, il existe une constante $C > 0$ et un $\varepsilon > 0$ tels que pour tous N assez grand et $\mu > 0$, on ait*

$$\forall X, \eta, \quad \mathbb{P}_{X, \eta} \left(X_{\frac{C}{\mu}} \in \mathcal{G}_{\frac{C}{\mu}} \right) > \varepsilon$$

où $\forall X, \eta$ est un conditionnement sur toutes les conditions initiales, et $\mathcal{G}_{\frac{C}{\mu}}$ est la composante géante dans $\eta_{\frac{C}{\mu}}$.

On peut montrer qu'avec grande probabilité, si C est assez grand, on peut bien définir la composante géante dans $\eta_{\frac{C}{\mu}}$ pour toute condition initiale η , et ainsi cet énoncé fait sens. Le point clé dans cette conjecture est le fait que le temps $\frac{C}{\mu}$ auquel on regarde la chaîne est un temps fixé, et non pas un temps d'attente. La conjecture dit alors que sur un temps d'ordre $\frac{1}{\mu}$, avec probabilité positive, le marcheur passe une proportion positive de son temps dans la composante géante. Dans le papier [11], un résultat similaire est montré avec un temps d'attente, mais alors celui-ci dit seulement qu'une proportion de l'ordre de μ (modulo des facteurs polylogarithmiques en n) du temps est passé dans la composante géante, ce qui est beaucoup plus faible car μ tend vers 0 avec n .

J'ai démontré la conjecture dans le cas où X n'a aucun voisin dans μ , c'est-à-dire que le conditionnement $\forall X, \eta$ devient un conditionnement "à X a toutes ses arêtes adjacentes fermées dans η ".

Afin de pouvoir conclure avec ce résultat, il faudrait montrer qu'avec une probabilité positive, pour toute condition initiale, le marcheur se retrouve sans voisins dans l'environnement en un temps de l'ordre de $\frac{1}{\mu}$. Plus précisément, on aurait besoin de

Conjecture 2. *Pour tous $d, p > p_c(d)$, il existe une constante $C > 0$ et un $\varepsilon > 0$ tels que pour tous N assez grand et $\mu > 0$, on ait*

$$\forall X, \eta, \quad \mathbb{P}_{X, \eta} \left(X_{\frac{C}{\mu}} \text{ a aucun voisin dans } \eta_{\frac{C}{\mu}} \right) > \varepsilon.$$

Pour ceci, il faudrait qu'il se retrouve sur un point avec une seule arête adjacente ouverte, puis que cette arête se ferme pendant ce temps. La conjecture précédente est alors impliquée par la conjecture suivante très similaire:

Conjecture 3. *Pour tous $d, p > p_c(d)$, il existe une constante $C > 0$ et un $\varepsilon > 0$ tels que pour tous N assez grand et $\mu > 0$, on ait*

$$\forall X, \eta, \quad \mathbb{P}_{X, \eta} \left(X_{\frac{C}{\mu}} \text{ n'a qu'un seul voisin dans } \eta_{\frac{C}{\mu}} \right) > \varepsilon.$$

A priori, cette conjecture paraît plus simple que la précédente à réaliser puisqu'elle donne une condition locale sur le marcheur, qui ne nécessite pas de savoir si le marcheur se trouve dans la composante géante, objet qui peut être assez compliqué à manipuler sur le tore. La structure locale de la percolation est beaucoup plus facilement comprise que la structure globale (composante géante, taille des autres composantes, etc.). Cependant, la façon dont la chaîne de Markov évolue est très mal comprise, et il n'est pas clair comment démontrer cette conjecture.

References

- [1] J. T. Cox, Y. Peres, and J. E. Steif. *Cutoff for the noisy voter model*. 2014. arXiv: 1408.5122v2.
- [2] E. Lubetzky and A. Sly. “An exposition to information percolation for the Ising model”. In: *Ann. Fac. Sci. Toulouse Serie 6, Volume 24* (2015), pp. 745–761.
- [3] E. Lubetzky and A. Sly. “Universality of cutoff for the Ising model”. In: *Ann. Probab.* 45(6A) (2017), pp. 3664–3696.
- [4] E. Lubetzky and A. Sly. “Information percolation and cutoff for the stochastic Ising model”. In: *J. Amer. Math. Soc.* 29 (2016), pp. 729–774.
- [5] J. Miller and Y. Peres. “Uniformity of the uncovered set of random walk and cutoff for lamplighter chains”. In: *Ann. Probab.* 40(2) (March 2012), pp. 535–577.
- [6] J. Ding and Y. Peres. “Mixing time for the Ising model : a uniform lower bound for all graphs”. In: *Annales de l’I.H.P. Probabilités et Statistiques* 47.4 (2011), pp. 1020–1028.
- [7] E. Nestoridi. “Optimal strong stationary times for random walks on the chambers of a hyperplane arrangement”. In: *Probab. Theory Relat. Fields* 174 (2019), pp. 929–943.
- [8] T.P. Hayes and A. Sinclair. “A general lower bound for mixing of single-site dynamics on graphs”. In: *Ann. Appl. Probab.* 17.3 (2007), pp. 931–952.
- [9] Y. Peres, A. Stauffer, and J.E. Steif. “Random walks on dynamical percolation: mixing times, mean squared displacement and hitting times”. In: *Probab. Theory Relat. Fields* 162 (2015), pp. 487–530.
- [10] Y. Peres, P. Sousi, and J.E. Steif. *Quenched exit times for random walk on dynamical percolation*. 2017. arXiv: 1707.07619.
- [11] Y. Peres, P. Sousi, and J.E. Steif. “Mixing time for random walk on supercritical dynamical percolation”. In: *Probab. Theory Relat. Fields* 176 (2020), pp. 809–849.