

Temps de Mélange sur des graphes à communautés

Thomas POIRÉE

Encadré par Anna Ben-Hamou et Arnaud Guyader

Introduction : L'étude du temps de mélange est un problème classique dans l'étude des chaînes de Markov : au bout de combien de temps une chaîne de Markov ergodique est-elle "proche" de sa distribution stationnaire. Cette information peut être utile par exemple pour mettre en oeuvre des méthodes MCMC dans lesquelles on aimerait savoir combien de temps faire tourner la chaîne pour obtenir un résultat avec un niveau de précision donné.

Pour certaines chaînes le mélange a presque entièrement lieu à un temps précis et dans une fenêtre très courte. Ce phénomène, le *cutoff*, a été établie pour la marche aléatoire sans rebroussement sur des graphes aléatoires réguliers ou "creux". On s'intéresse ici à des graphes organisés en communautés et on voudrait savoir si ce phénomène est toujours présent. La géométrie d'un graphe joue un rôle déterminant dans son temps de mélange, la présence de "goulots d'étranglements" créés par des communautés trop isolées va ralentir significativement le mélange de la marche. On verra que dans ce cas, le mélange est gouverné par un temps d'attente géométrique et il n'y a plus cutoff. En revanche dans un régime avec beaucoup d'arêtes "inter-communautaires", la marche mélange vite et présente toujours le cutoff.

Table des matières

1	Temps de mélange des chaînes de Markov	3
1.1	Variation totale	3
1.2	Couplages	4
1.3	Marche aléatoire sur des graphes	4
1.4	Le phénomène de Cutoff	6
2	Marche aléatoire sur des graphes aléatoires	8
2.1	Modèle de configuration	8
2.2	Marche aléatoire sans rebroussement	9
3	Graphes à communautés	10
3.1	Modèle à deux communautés	10
3.2	Influence des communautés sur le temps de mélange	11
3.3	Résultats principaux	12

1 Temps de mélange des chaînes de Markov

On commence par présenter des notions de base sur le temps de mélange. On énonce aussi quelques propriétés utiles et on présente le phénomène de cutoff. Les preuves des résultats peuvent être trouvées dans [3] (chapitres 4-5-7). Dans toute la suite Ω est un ensemble fini.

1.1 Variation totale

Définition 1.1. *Etant données deux distributions μ et ν on définit la **distance en variation totale** par*

$$\|\mu - \nu\|_{TV} := \max_{A \in \Omega} |\mu(A) - \nu(A)| \quad (1)$$

Proposition 1.1. *Soient μ et ν deux distributions de probabilité sur Ω . Alors :*

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

$$\|\mu - \nu\|_{TV} = \sum_{x \in \Omega} (\mu(x) - \nu(x))_+ \quad (3)$$

On considère maintenant une chaîne de Markov $(X_t)_{t \geq 0}$ sur Ω de matrice de transition P irréductible et apériodique. On note π son unique distribution stationnaire, on s'intéresse aux quantités suivantes :

$$\mathcal{D}_x(t) = \|P^t(x, \cdot) - \pi\|_{TV} \text{ et } \mathcal{D}(t) = \max_{x \in \Omega} \mathcal{D}_x(t).$$

La quantité $\mathcal{D}(t)$ correspond à la distance à l'équilibre au temps t .

Le théorème de convergence pour les chaînes de Markov dit que $\mathcal{D}(t) \xrightarrow[t \rightarrow +\infty]{} 0$ on aimerait étudier à partir de quel t la distance est plus petite qu'un ε donné, ce qui mène à la définition suivante :

Définition 1.2. *Temps de mélange*

$$t_{mix}(\varepsilon) := \min\{t : \mathcal{D}(t) \leq \varepsilon\} \quad (4)$$

Il est parfois plus facile de contrôler l'écart entre les distributions partant de deux points différents, on définit ainsi

$$\bar{d}(t) := \max_{x, y \in \Omega} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \quad (5)$$

On a la relation utile suivante :

$$\mathcal{D}(t) \leq \bar{d}(t) \leq 2\mathcal{D}(t). \quad (6)$$

1.2 Couplages

Un couplage de deux mesures de probabilité est un couple de variables aléatoires (X, Y) défini sur un même espace de probabilité et tel que les marginales X et Y sont respectivement de loi μ et ν . Le couplage est une technique très utile pour l'étude des temps de mélange, la proposition suivante met en évidence le lien entre couplage et distance entre deux lois.

Proposition 1.2. *Soient μ et ν deux mesures de probabilité sur Ω . Alors*

$$\|\mu - \nu\|_{TV} = \inf\{\mathbb{P}(X \neq Y) : (X, Y) \text{ est un couplage de } \mu \text{ et } \nu\} \quad (7)$$

Un **couplage markovien** est un processus $(X_t, Y_t)_{t \geq 0}$ tel que (X_t) et (Y_t) sont des chaînes de Markov de transition P mais de point de départ potentiellement différent. Supposons que ce couplage reste ensemble après sa première rencontre, c'est à dire que si $X_s = Y_s$ alors $X_t = Y_t$ pour tout $t \geq s$.

Comme $P^t(x, z) = \mathbb{P}_{(x,y)}(X_t = z)$ et $P^t(y, z) = \mathbb{P}_{(x,y)}(Y_t = z)$, (X_t, Y_t) est un couplage de $P^t(x, \cdot)$ avec $P^t(y, \cdot)$ la proposition 1.2 implique donc

$$\|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \leq \mathbb{P}_{(x,y)}(X_t \neq Y_t). \quad (8)$$

Soit $\tau := \inf\{n \geq 0, X_n = Y_n\}$, en remarquant que $\mathbb{P}_{(x,y)}(X_t \neq Y_t) = \mathbb{P}_{(x,y)}(\tau > t)$, et en appliquant 6 on obtient :

$$\mathcal{D}(t) \leq \max_{(x,y)} \mathbb{P}_{(x,y)}(\tau > t). \quad (9)$$

Ainsi en construisant des couplages on pourra obtenir des bornes supérieures sur \mathcal{D}

1.3 Marche aléatoire sur des graphes

Un graphe $G = (V, E)$ est composé d'un ensemble de sommets V et d'un ensemble d'arêtes $E \subset \{\{x, y\} : x, y \in V, x \neq y\}$. On dit que x est voisin de y si $\{x, y\} \in E$ (on note $x \sim y$).

La marche aléatoire simple sur G passe d'un sommet à l'autre en choisissant uniformément parmi ses voisins. Autrement dit, en notant $\deg(x)$ le cardinal de $\{y \in V, x \sim y\}$

$$P(x, y) = \begin{cases} \frac{1}{\deg(x)} & \text{si } x \sim y \\ 0 & \text{sinon.} \end{cases} \quad (10)$$

Si tous les sommets ont le même degré d , on dit que le graphe est d -régulier. Comme on l'a dit précédemment, le problème dans les graphes à communauté est la présence de goulots d'étranglement, on va donc essayer de formaliser cette notion et de la relier au temps de mélange.

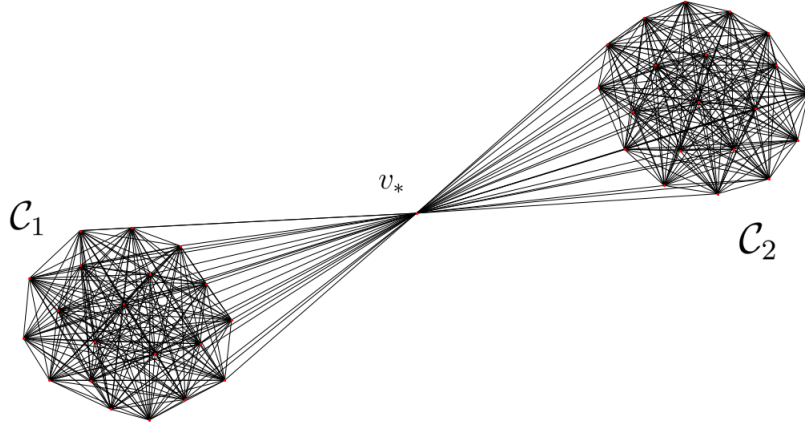


FIGURE 1 – deux graphes complets à 20 sommets collés en v_*

Définition 1.3. Soit S un sous-ensemble de Ω , on définit le "**bottleneck ratio**" de S par

$$\Phi(S) := \frac{Q(S, S^c)}{\pi(S)} \quad (11)$$

où

$$Q(S, S^c) = \sum_{\substack{x \in S \\ y \in S^c}} \pi(x)P(x, y)$$

Et le "**bottleneck ratio**" de la chaîne entière par

$$\Phi_* := \min_{S: \pi(S) \leq 1/2} \Phi(S) \quad (12)$$

Pour une marche aléatoire sur un graphe on a une forme plus explicite du "bottleneck ratio" pour un ensemble S . En effet on voit facilement que $Q(S, S^c) = \frac{|\partial S|}{2|E|}$ ou ∂S est la frontière entre S et S^c donc $\Phi(S) = \frac{|\partial S|}{\sum_{x \in S} \deg(x)}$. Le théorème suivant n'est en général pas suffisant si on cherche des informations précises sur le temps de mélange mais sera souvent utile car c'est un outil facile d'utilisation en particulier pour l'étude de graphes.

Théorème 1.3.1.

$$t_{mix}(\varepsilon) \geq \frac{1 - 2\varepsilon}{2\Phi_*} \quad (13)$$

Exemple 1.1. Deux cliques collées

Considérons le cas extrême de deux graphes complets à n sommets qu'on a collés avec un nouveau sommet v_* (voir figure 1). Autrement dit $G = (\mathcal{C}_1 \cup \mathcal{C}_2, E)$

avec $\mathcal{C}_1 \cap \mathcal{C}_2 = \{v_*\}$ et $\forall x \neq y, (x, y) \in E \Leftrightarrow \{(x \in \mathcal{C}_1 \text{ et } y \in \mathcal{C}_1) \text{ ou } (x \in \mathcal{C}_2 \text{ et } y \in \mathcal{C}_2)\}$. On a en particulier

$$\deg(v) = \begin{cases} 2n & \text{si } v = v_* \\ n+1 & \text{sinon.} \end{cases} \quad (14)$$

Ainsi en prenant $S = \mathcal{C}_1 \setminus \{v_*\}$ on trouve $\Phi(S) = \frac{n}{n^2}$ ce qui donne

$$t_{mix}(\varepsilon) \geq 1/2(1 - 2\varepsilon)n \quad (15)$$

On verra dans la section suivante que l'ordre en n est le bon mais que la dépendance en ε est mauvaise.

1.4 Le phénomène de Cutoff

Pour observer des comportements intéressants de $t_{mix}(\varepsilon)$ il est plus pertinent d'opter pour une approche asymptotique sur la taille de l'espace d'états. On considère donc une suite de chaînes de Markov de matrices de transitions $(P^{(n)})_{n \geq 1}$ sur une suite d'espaces d'états $(\Omega^{(n)})_{n \geq 1}$ et on s'intéresse au comportement de $t_{mix}^{(n)}(\varepsilon)$ quand $n \rightarrow +\infty$.

C'est en adoptant ce point de vue qu'a émergé le phénomène de cutoff, il a été observé pour la première fois par Diacons et Shahshahani (1981) dans le contexte du mélange de cartes.

Définition 1.4. *Une suite de chaîne a un **cutoff** si*

$$\lim_{n \rightarrow +\infty} \frac{t_{mix}^{(n)}(\varepsilon)}{t_{mix}^{(n)}(1 - \varepsilon)} = 1 \quad (16)$$

Le cutoff a lieu avec une fenêtre w_n si :

$$\lim_{\alpha \rightarrow +\infty} \liminf_{n \rightarrow +\infty} \mathcal{D}^{(n)}(t_{mix}^{(n)} - \alpha w_n) = 1 \quad (17)$$

Et

$$\lim_{\alpha \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathcal{D}^{(n)}(t_{mix}^{(n)} + \alpha w_n) = 0 \quad (18)$$

Avec $w_n = o(t_{mix}^{(n)})$

En d'autres termes, le premier terme dans le développement asymptotique de $t_{mix}(\varepsilon)$ ne dépend pas de ε , c'est la raison pour laquelle on parle souvent de "temps de mélange" pour la valeur arbitraire $1/4$ de ε . On soupçonne que de nombreuses chaînes présentent ce phénomène mais le prouver est souvent une tâche très difficile car elle demande une étude très fine du temps de mélange.

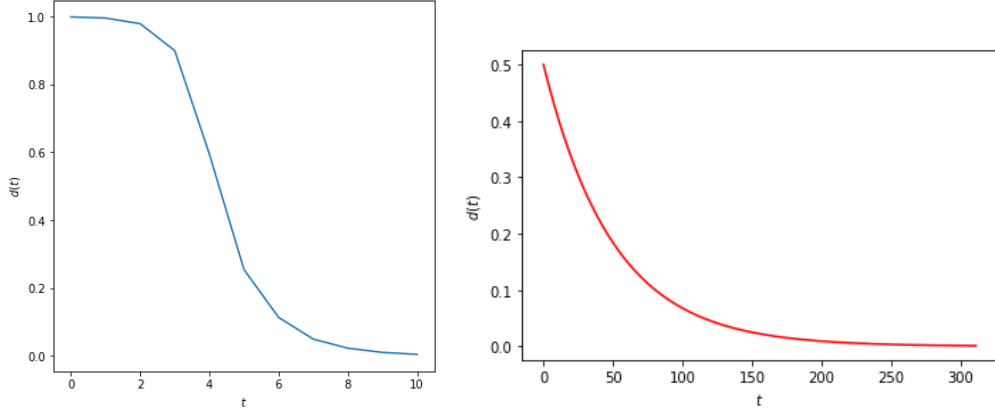


FIGURE 2 – Cutoff pour la marche aléatoire sur un graphe 6-régulier à 1000 sommets (gauche), marche aléatoire sur les deux cliques sans cutoff (droite)

Un exemple sans cutoff : Reprenons l'exemple 1.1, on va montrer que cette chaîne ne présente pas de cutoff.

Borne sup :

On va coupler une chaîne $(X_n)_{n \geq 0}$ partant de x et une chaîne $(Y_n)_{n \geq 0}$ telle que $Y_0 \sim \pi$ avec π la distribution stationnaire donnée par :

$$\pi(x) = \begin{cases} \frac{1}{n+1} & \text{si } v = v_* \\ \frac{1}{2(n+1)} & \text{sinon.} \end{cases}$$

On construit une bijection ϕ entre les éléments de \mathcal{C}_1 et ceux de \mathcal{C}_2 vérifiant $\phi(v_*) = v_*$. Les chaînes évoluent de la façon suivante :

Si X_0 et Y_0 ne sont pas dans la même clique (par exemple $X_0 \in \mathcal{C}_1$ et $Y_0 \in \mathcal{C}_2$). Soit $(U_t)_{t \geq 1}$ une suite I.I.D d'éléments tirés uniformément dans \mathcal{C}_1 (la suite des innovations). Tant que X_t et Y_t ne se sont pas rencontrés on pose $X_t = U_t$ et $Y_t = \phi(U_t)$. Une fois que les chaînes se rencontrent elles ne se quittent plus ($X_s = Y_s = U_s$) .

Si X_0 et Y_0 sont dans la même clique, alors pour tout $n \geq 1$, $X_n = Y_n = U_n$. On vérifie facilement que (X_t, Y_t) est un couplage de $P^t(x, \cdot)$ avec π , par la proposition 1.2 on a :

$$\mathcal{D}_x(t) \leq \mathbb{P}_{x, \pi}(X_t \neq Y_t)$$

Soit $\tau := \inf\{n \geq 0, X_n = Y_n\}$, Si X_0 et Y_0 sont dans la même clique, $\tau = 1$, sinon $\tau = \inf\{n \geq 1, U_n = v_*\}$. On obtient donc

$$\mathcal{D}(t) \leq 1/2(1 - 1/n)^t$$

et

$$t_{mix}(\varepsilon) \geq \frac{\log(\frac{1}{2\varepsilon})}{\log(1 - 1/n)} \simeq n \log(\frac{1}{2\varepsilon})$$

Borne Inf : On pose $\tau = \inf\{n \geq 1, X_n = v_*\}$, supposons sans perte de généralité que $x \in \mathcal{C}_1$

$$\begin{aligned}\mathcal{D}_x(t) &\geq \sum_{y \neq v_*} (P^t(x, y) - \pi(y))_+ \\ &= \sum_{y \neq v_*} (\mathbb{P}_x(X_t = y, \tau < t) + \mathbb{P}_x(X_t = y, \tau \geq t) - \pi(y))_+ \\ &= \sum_{y \neq v_*} \left(\frac{1}{2(n+1)} \mathbb{P}(\tau < t) + \frac{\mathbb{P}_x(\tau \geq t)}{n} \mathbb{1}_{y \in \mathcal{C}_1} - \pi(y) \right)_+\end{aligned}$$

Or $\frac{1}{2(n+1)} \mathbb{P}_x(\tau < t) - \pi(y) < 0$ donc

$$\begin{aligned}&= \sum_{\substack{y \neq v_* \\ y \in \mathcal{C}_1}} \left(\frac{1}{2(n+1)} \mathbb{P}_x(\tau < t) + \frac{1}{n} \mathbb{P}_x(\tau \geq t) - \frac{1}{2(n+1)} \right)_+ \\ &= \sum_{\substack{y \neq v_* \\ y \in \mathcal{C}_1}} \left(\frac{1}{2(n+1)} (1 - \mathbb{P}_x(\tau \geq t)) + \frac{1}{n} \mathbb{P}_x(\tau \geq t) - \frac{1}{2(n+1)} \right)_+ \\ &= \sum_{\substack{y \neq v_* \\ y \in \mathcal{C}_1}} \frac{n+2}{2n(n+1)} \mathbb{P}_x(\tau \geq t) \\ &= \frac{n+2}{2(n+1)} \mathbb{P}_x(\tau \geq t) \simeq \frac{1}{2} (1 - 1/n)^{t-1}\end{aligned}$$

On trouve finalement :

$$n \log(2\varepsilon) \leq t_{mix}(\varepsilon) \leq n \log\left(\frac{1}{2\varepsilon}\right) + 1 \quad (19)$$

Et il n'y a pas cutoff (voir figure 2).

2 Marche aléatoire sur des graphes aléatoires

Dans cette section on définit le cadre qui nous permettra d'étudier le temps de mélange de graphes, on présente le "modèle de configuration" utilisé pour générer les graphes aléatoires et on définit la marche aléatoire sans rebroussement.

2.1 Modèle de configuration

Pour montrer le cutoff sur une famille de graphe, on a besoin d'introduire une nouvelle source d'aléa. En effet même sur une classe raisonnable de graphes tels

que les graphes d-réguliers, on peut construire des cas pathologiques pour lesquels le temps de mélange sera sensiblement différent de celui d'un graphe "typique" de cette famille ce qui nous gêne pour construire un résultat général. Cette nouvelle source d'aléa offre plus de souplesse dans le maniement des graphes.

On définit ainsi le modèle de configuration, ce modèle à l'avantage de fournir un algorithme de génération uniforme de graphes aléatoires peu coûteux que l'on utilisera pour les simulations.

Soit V un ensemble de sommets fini et $\deg : V \mapsto \{2, 3, \dots\}$ une fonction degrés telle que

$$N := \sum_{v \in V} \deg(v) \quad (20)$$

est pair.

On va attacher à chaque sommet v $\deg(v)$ "demi-arêtes" formant ainsi l'ensemble

$$\mathcal{X} := \{(v, i), v \in V, 1 \leq i \leq \deg(v)\} \quad (21)$$

On choisit maintenant uniformément un appariement σ sur \mathcal{X} , c'est à dire qu'on tire une fonction σ sur \mathcal{X} vérifiant :

$$\forall x \in \mathcal{X}, \sigma(\sigma(x)) = x \quad (22)$$

$$\forall x \in \mathcal{X}, \sigma(x) \neq x \quad (23)$$

On obtient un graphe en identifiant chaque couple $\{(v, i), \pi((v, i)) = (w, j)\}$ de demi-arêtes comme une arête entre v et w . On note $G \sim \mathcal{G}(n, d)$ un graphe généré par le modèle de configuration avec $|V| = n$ et $\forall v \in V, \deg(v) = d$

Ce modèle a l'inconvénient de pouvoir produire des boucles et des multi-graphes c'est à dire qu'on peut avoir plusieurs arêtes entre deux mêmes sommets, cependant la probabilité pour qu'un tel événement se produise est bornée asymptotiquement et conditionnellement au fait que le graphe soit simple, le tirage est uniforme sur l'ensemble des graphes simples avec la même suite de degrés. Voir chapitre 7 de [5] pour la preuve de ces résultats et plus d'informations sur le modèle de configuration.

Un avantage de ce modèle est que le tirage de l'appariement (phase de "matching") peut être réalisé séquentiellement (voir code (?)) , ce qui sera exploité au cours des démonstrations.

2.2 Marche aléatoire sans rebroussement

La marche aléatoire sans rebroussement est une chaîne de Markov définie sur l'espace d'état \mathcal{X} des demi-arêtes. Deux demi-arêtes $x = (u, i)$ et $y = (v, j)$ sont voisines si $u = v$ et $i \neq j$. La marche sans rebroussement passe d'un état x à un état

y choisi uniformément parmi les voisins de $\sigma(x)$. On note $\deg((v, i)) = \deg(v) - 1$, la matrice de transition de la chaîne est :

$$P(x, y) = \begin{cases} \frac{1}{\deg(\sigma(x))} & \text{si } y \text{ voisin de } \sigma(x) \\ 0 & \text{sinon.} \end{cases} \quad (24)$$

Cela signifie que en projetant la marche sur V , si la marche viens de passer de v à w , à la prochaine étape elle passera sur un élément de $\{z, z \sim wz \neq v\}$ choisi uniformément.

Un calcul direct montre que la probabilité uniforme sur \mathcal{X} est invariante.

On dispose du résultat suivant extrêmement précis de Lubetsky et Sly [4] sur la marche aléatoire sans rebroussement :

Théorème 2.2.1. *Soit $G \sim \mathcal{G}(n, d)$, $d \geq 3$ et $\varepsilon \in]0, 1[$. Soit $t_{mix}^{(n)}(\varepsilon)$ le temps de mélange de la marche aléatoire sans rebroussement sur G . Alors :*

$$t_{mix}^{(n)}(\varepsilon) = \log_{d-1}(dn) + O_{\mathbb{P}}(1).$$

Autrement dit, avec probabilité qui tend vers 1, la marche sans rebroussement présente le cutoff en $\log_{d-1}(n)$ avec fenêtre d'ordre constant.

Notre objectif principal est d'étudier l'influence des communautés sur ce résultat, on commence donc par définir un modèle de configuration "à communautés".

3 Graphes à communautés

3.1 Modèle à deux communautés

On s'intéresse au cas d -régulier mais la procédure peut se généraliser à une suite de degrés quelconques.

Soient V_0 et V_1 deux ensembles de sommets de cardinale n et $\mathcal{X}_0, \mathcal{X}_1$ les ensembles de demi-arêtes associées : $\mathcal{X}_k = \{(v, i), v \in V_k, 1 \leq i \leq d\}$. On note $N = dn$ le cardinal de \mathcal{X}_1 .

Soit $1 \leq p < N/2$ (qui peut dépendre de n) et $\alpha = \frac{p}{N}$.

On choisit p demi-arêtes distinctes uniformément et de façon indépendante dans \mathcal{X}_0 et \mathcal{X}_1 on note les ensembles de ces arêtes respectivement I_0 et I_1 . On note O_0 les arêtes non-tirées dans \mathcal{X}_0 et O_1 celles non-tirées dans \mathcal{X}_1 .

Une fois ces ensembles construits on choisit uniformément un appariement sur I_0 , un appariement sur I_1 et un appariement entre O_0 et O_1 . On obtient ainsi un appariement σ sur $\mathcal{X}_0 \cup \mathcal{X}_1$ et on note $G \sim \mathcal{G}_\alpha(n, d)$ un graphe ainsi généré.

On s'intéresse à la marche sans rebroussement associée à G (matrice de transition donnée par 24).

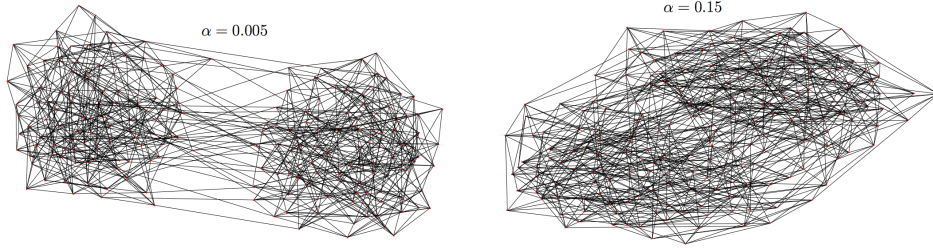


FIGURE 3 – Graphes générés avec le modèle à communautés ($n = 100$, $d = 8$)

3.2 Influence des communautés sur le temps de mélange

Le paramètre α va jouer un rôle déterminant sur le temps de mélange de la marche. Il représente la proportion d'arêtes qui permettent de passer d'une communauté à l'autre. À l'équilibre, ce paramètre correspond également à la probabilité de passer d'une communauté à une autre. En effet :

$$\mathbb{P}_\pi(X_{k+1} \in \mathcal{X}_1 | X_k \in \mathcal{X}_0) = \sum_{\substack{x \in \mathcal{X}_0^O \\ y \in \mathcal{X}_1^O}} \mathbb{P}_\pi(X_{k+1} = y | X_k = x) = \sum_{\substack{x \in \mathcal{X}_0^O \\ y \in \mathcal{X}_1^O}} \frac{1}{2N} = \alpha \quad (25)$$

Comme on peut le voir sur la figure 3 plus α est petit, plus les communautés sont séparées et plus on voit apparaître un goulot d'étranglement entre celles-ci. On peut le vérifier par le calcul, par définition la frontière entre \mathcal{X}_0 et \mathcal{X}_1 est donnée par \mathcal{X}_0^O on obtient donc :

$$\Phi(\mathcal{X}_0) = \frac{|I_0|}{N} = \alpha \quad (26)$$

En appliquant le théorème 13 on obtient :

$$t_{mix}^{(n)} := t_{mix}^{(n)}(1/4) \geq \frac{1}{4\alpha} \quad (27)$$

Pour $\alpha \ll \frac{1}{\log_{d-1}(dn)}$ le comportement de la chaîne sera donc sensiblement différent que celui donné par 2.2.1.

Les deux régimes qui nous intéressent sont donc

$$\alpha \ll \frac{1}{\log_{d-1}(dn)} \text{ et } \alpha \gg \frac{1}{\log_{d-1}(dn)}.$$

Dans le premier cas, le temps de mélange est gouverné par celui de la chaîne "projetée" sur les communautés. En effet intuitivement la marche mettra plus longtemps que $\log_{d-1}(dn)$ pour sortir d'une communauté et donc celle-ci sera mélangée avant que la marche ne puisse s'échapper. Les probabilités de transition pour la

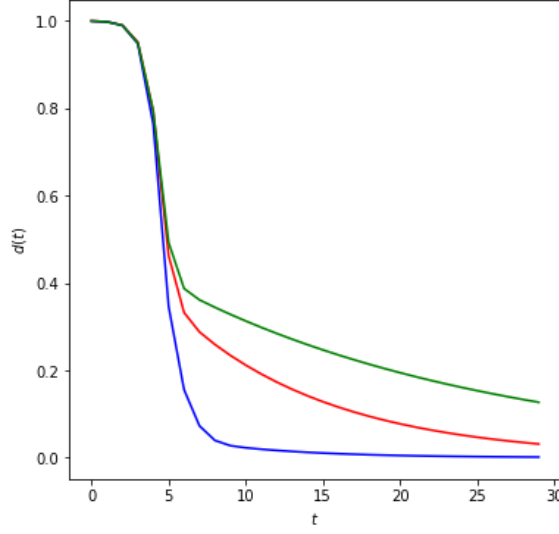


FIGURE 4 – Distance à l'équilibre pour la marche sur un graphe 10-régulier à 400 sommet avec $\alpha = 0.06$ (rouge), $\alpha = 0.01$ (vert) et $\alpha = 0.11$ (bleu)

chaîne projeté seront donc approximativement données par 25.

Posons :

$$Q = \begin{pmatrix} 1 - \alpha & \alpha \\ \alpha & 1 - \alpha \end{pmatrix}$$

On peut montrer que pour la chaîne de transition Q la distance à l'équilibre vérifie

$$\frac{(1 - 2\alpha)^t}{2} \leq \mathcal{D}(t) \leq (1 - 2\alpha)^t \quad (28)$$

On retrouve un résultat très similaire à celui des deux cliques collée 19, cas extrême de graphe communautaire, et il n'y a pas de cutoff.

Dans le second cas, la marche change suffisamment souvent de communauté pour qu'elles n'aient pas trop d'influence sur le temps de mélange et la chaîne présente toujours le cutoff.

3.3 Résultats principaux

Théorème 3.3.1. Soit $\alpha \gg \frac{1}{\log(dn)}$ et $G_n \sim \mathcal{G}_\alpha(n, d)$. Alors :
 $\forall \varepsilon > 0, \exists \delta > 0,$

$$\mathbb{P}(\mathcal{D}^{(n)}(\log_{d-1}(2dn) - \delta \log(\log(2dn))) < 1 - \varepsilon) \xrightarrow{n \rightarrow +\infty} 0 \quad (29)$$

et $\forall \varepsilon > 0, \exists \delta > 0,$

$$\mathbb{P}(\mathcal{D}^{(n)}(\log_{d-1}(2dn) + \delta \log(\log(2dn))) > \varepsilon) \xrightarrow{n \rightarrow +\infty} 0 \quad (30)$$

Autrement dit, avec probabilité qui tend vers 1, la marche aléatoire sans rebroussement présente le cutoff en $\log_{d-1}(2dn)$ avec une fenêtre d'ordre $\log(\log(2dn))$.

Pour la démonstration on s'appuie en grande partie sur la preuve du cutoff dans le cas sans communautés. Le début est essentiellement le même.

Démonstration. Preuve de la borne inférieure : On note $\mathcal{X} = \mathcal{X}_0 \cup \mathcal{X}_1$ l'ensemble des états. Soit $x_0 \in \mathcal{X}$.

$$\mathbb{P}(\mathcal{D}(t) < 1 - \varepsilon) \leq \mathbb{P}(\mathcal{D}_{x_0}(t) < 1 - \varepsilon)$$

Soit $\mathcal{B}_t(x)$ l'ensemble des demi-arêtes que la marche sans rebroussement peut atteindre en moins de t pas et soit A sa frontière.

$$\begin{aligned} \mathcal{D}_{x_0}(t) &\geq P^t(x_0, A) - \pi(A) \\ &\geq 1 - \frac{|\partial \mathcal{B}_t(x_0)|}{2N} \\ &\geq 1 - \frac{(d-1)^t}{2N} \end{aligned}$$

Soit $\delta > 0$, en prenant $t = \log_{d-1}(2N) - \delta \log(\log(2N))$ on trouve

$$\mathcal{D}_{x_0}(t) \geq 1 - (d-1)^{-\delta \log(\log(2N))} \quad (31)$$

On obtient 29 (de façon déterministe) en prenant N suffisamment grand dans 31.

Preuve de la borne supérieure : Soit $a > 0$ et $K = \frac{\log(N)}{6 \log(d-1)}$. On dit que x est une K -racine si $\mathcal{B}_K(x)$ est un arbre et on note \mathcal{R} l'ensemble des K -racines. Soit $s = a \log \log(N)$ pour $a > 0$. On a :

$$\begin{aligned} \mathcal{D}_{x_0}(s+t) &= \sum_y \left(\frac{1}{2N} - P^{s+t}(x_0, y) \right)_+ \\ &= \sum_y \left(\frac{1}{2N} - \sum_x P^s(x_0, x) P^t(x, y) \right)_+ \\ &\leq \sum_y \sum_x P^s(x_0, x) \left(\frac{1}{2N} - P^t(x, y) \right)_+ \\ &\leq \sum_x P^s(x_0, x) \mathcal{D}_x(t) \\ &\leq P^s(x_0, \mathcal{X} \setminus \mathcal{R}) + \max_{x \in \mathcal{R}} \mathcal{D}_x(t) \end{aligned}$$

Le lemme suivant nous permet de ne considérer la maximisation que sur les K-racines.

Lemme 3.3.2. $\max_{x_0 \in \mathcal{X}} P^s(x_0, \mathcal{X} \setminus \mathcal{R}) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0$

Si on admet ce lemme pour le moment il nous reste donc à montrer que $\mathbb{P}(\max_{x \in \mathcal{R}} \mathcal{D}_x(t) > \varepsilon) \xrightarrow[n \rightarrow +\infty]{} 0$.

$$\begin{aligned} \mathcal{D}_x(t) &= \sum_y \left(\frac{1}{2N} - P^t(x, y) \right)_+ = \sum_y \left(\frac{1}{N} - P^t(x, \sigma(y)) \right)_+ \\ &\leq \sum_{y \in \mathcal{R} \setminus \mathcal{B}_K(x)} \left(\frac{1}{2N} - P^t(x, \sigma(y)) \right)_+ + \frac{|\mathcal{B}_K(x) \cup \mathcal{X} \setminus \mathcal{R}|}{2N} \end{aligned}$$

Le deuxième terme tend vers 0 en probabilité, en effet $|\mathcal{B}_K(x)| \leq (d-1)^K = o(N)$. Et

$$|\mathcal{X} \setminus \mathcal{R}| = \sum_{x \in \mathcal{X}} P^s(x, \mathcal{X} \setminus \mathcal{R}) = o_{\mathbb{P}}(N)$$

Par 3.3.2

En utilisant la propriété de symétrie de la marche $P(x, \sigma(y)) = P(y, \sigma(x))$ (22) on obtient la représentation suivante :

$$P^t(x, \sigma(y)) = \sum_{\substack{u \in \partial \mathcal{B}_x(t/2) \\ v \in \partial \mathcal{B}_y(t/2)}} P^{t/2}(x, u) P^{t/2}(y, v) \mathbf{1}_{v=\sigma(u)} \quad (32)$$

On va à présent exploiter la propriété de génération séquentielle de l'appariement du modèle de configuration pour explorer les boules $\mathcal{B}_x(t/2)$ et $\mathcal{B}_y(t/2)$ en ne conservant que les chemins sans cycle. On obtiendra ainsi une approximation sous forme d'arbre des deux boules (Figure 5) ce qui permet une ré-écriture très simple de 32. On verra ensuite que les chemins exclus par ce procédé sont en fait négligeables.

Processus d'exploration :

1. On commence par révéler le voisinage de x jusqu'en $t/2$. Au moment de matcher chaque demi-arête u , si $\sigma(u)$ a déjà été révélée on n'effectue pas l'appariement et on les marque en rouge.
2. Pour y on fait la même chose mais en évitant de matcher les demi-arêtes avec celles déjà révélées en 1. . Si on obtient donc une arête $\sigma(u)$ de u dans celles déjà révélées on marque ces deux demi-arêtes en rouge également.
3. On s'arrête quand il n'y a plus de demi-arêtes compatibles et on complète l'appariement en réalisant un appariement uniforme sur toutes les demi-arêtes restantes.

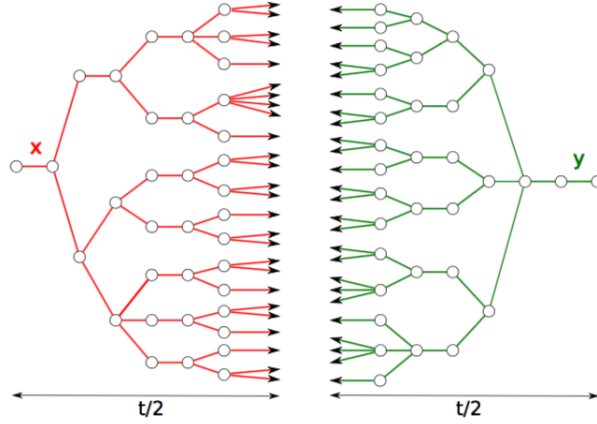


FIGURE 5 – Approximation sous forme d'arbre (référence : [2])

On note $\mathcal{F}_x = \{ \text{demi-arêtes révélées en 1. à distance } t/2 \text{ de } x \text{ (demi-arêtes non marquées en rouge dans } \partial\mathcal{B}_x(t/2)) \}$ et $\mathcal{F}_y = \{ \text{demi-arêtes révélées en 2. à distance } t/2 \text{ de } y \text{ (demi-arêtes non marquées en rouge dans } \partial\mathcal{B}_y(t/2)) \}$ les ensembles ainsi obtenus.

On a :

$$\begin{aligned} P^t(x, \sigma(y)) &\geq \sum_{\substack{u \in \mathcal{F}_x \\ v \in \mathcal{F}_y}} \frac{1}{(d-1)^t} \mathbb{1}_{v=\sigma(u)} = \frac{1}{2N} \sum_{\substack{u \in \mathcal{F}_x \\ v \in \mathcal{F}_y}} \frac{1}{\log(N)^b} \mathbb{1}_{v=\sigma(u)} \\ &= \frac{1}{2N} \times \frac{1}{\log(N)^b} \left(\sum_{\substack{u \in \mathcal{F}_x \cap \mathcal{X}_0^I \\ v \in \mathcal{F}_y \cap \mathcal{X}_0^I}} \mathbb{1}_{v=\sigma(u)} + \sum_{\substack{u \in \mathcal{F}_x \cap \mathcal{X}_1^I \\ v \in \mathcal{F}_y \cap \mathcal{X}_1^I}} \mathbb{1}_{v=\sigma(u)} + \sum_{\substack{u \in \mathcal{F}_x \cap \mathcal{X}_0^O \\ v \in \mathcal{F}_y \cap \mathcal{X}_1^O}} \mathbb{1}_{v=\sigma(u)} + \sum_{\substack{u \in \mathcal{F}_x \cap \mathcal{X}_1^O \\ v \in \mathcal{F}_y \cap \mathcal{X}_0^O}} \mathbb{1}_{v=\sigma(u)} \right) \end{aligned}$$

On note : $I_x^0 = \mathcal{F}_x \cap I_0$, $I_y^0 = \mathcal{F}_y \cap I_0$, $O_x^0 = \mathcal{F}_x \cap O_0$, $O_y^0 = \mathcal{F}_y \cap O_0 \dots$

On pose $s = \frac{b \log(\log(N))}{\log(d-1)}$ avec $b > 0$, et $t = \frac{\log(2N)}{\log(d-1)} + s$. Ainsi , $\frac{1}{(d-1)^t} = \frac{1}{2N} \times \frac{1}{(d-1)^s}$

On va à présent utiliser le lemme 6.1 de [2] :

Lemme: Soit \mathcal{I} un ensemble pair, $\{w_{i,j}\}_{(i,j) \in \mathcal{I} \times \mathcal{I}}$ un tableau de poids positifs et σ un appariement uniforme sur \mathcal{I} . Alors pour tout $a > 0$:

$$\mathbb{P}\left(\sum_{i \in \mathcal{I}} w_{i, \sigma(i)} \leq m - a\right) \leq \exp\left\{-\frac{a^2}{4\theta m}\right\},$$

avec $m = \frac{1}{\mathcal{I}-1} \sum_{i \in \mathcal{I}} \sum_{j \neq i} w_{i,j}$ et $\theta = \max_{i \neq j} (w_{i,j} + w_{j,i})$

En prenant $\mathcal{I} = \mathcal{X}_0 \cap I_0$ et $w_{i,j} = \mathbb{1}_{i \in I_x^0} \mathbb{1}_{j \in I_y^0} \frac{1}{\log(N)^b}$ on a $m = \frac{|I_x^0| \times |I_y^0|}{\log(N)^b (1-\alpha)N}$ et

$\theta = \frac{1}{2\log(N)^b}$. On obtient ainsi :

$$\begin{aligned}
\mathbb{P}\left(\frac{1}{\log(N)^b} \sum \mathbb{1}_{v=\sigma(u)} - \frac{|I_x^0| \times |I_y^0|}{\log(N)^b(1-\alpha)N} \leq -\varepsilon/4 | \mathcal{F}_x, \mathcal{F}_y\right) &\leq \exp\left(-\frac{\varepsilon^2 \log(N)^{2b} N(1-\alpha)}{32|I_x^0| \times |I_y^0|}\right) \\
&\leq \exp\left(-\frac{\varepsilon^2 \log(N)^{2b} N}{32N \log(N)^b}\right) \\
&= \exp\left(-\frac{1}{32} \varepsilon^2 \log(N)^b\right) \\
&= o\left(\frac{1}{N^2}\right)
\end{aligned}$$

Pour b suffisamment grand.

Le même argument avec $\mathcal{I} = O_0 \cup O_1$ donne :

$$\begin{aligned}
\mathbb{P}\left(\frac{1}{\log(N)^b} \sum \mathbb{1}_{v=\sigma(u)} - \frac{|I_x^0| \times |I_y^0|}{\log(N)^b \alpha N} \leq -\varepsilon/4 | \mathcal{F}_x, \mathcal{F}_y\right) &\leq \exp\left(-\frac{\varepsilon^2 \log(N)^{2b} N \alpha}{32|I_x^0| \times |I_y^0|}\right) \\
&\leq \exp\left(-\frac{1}{32} \varepsilon^2 \log(N)^b \alpha\right)
\end{aligned}$$

Comme $\alpha \gg \frac{1}{\log(N)}$ on peut prendre b suffisamment grand pour que

$$\exp(-(\varepsilon^2 \log(N)^b \alpha)/32) = o\left(\frac{1}{N^2}\right)$$

En prenant une borne union sur les couples (x, y) on obtient qu'avec probabilité qui tend vers 1, pour tout $x \in \mathcal{R}$ et tout $y \in \mathcal{R} \setminus \mathcal{B}_K(x)$

$$\begin{aligned}
P^t(x, \sigma(y)) &\geq \frac{1}{2N} \left(\frac{|I_x^0| \times |I_y^0|}{\log(N)^b N(1-\alpha)} + \frac{|I_x^1| \times |I_y^1|}{\log(N)^b N(1-\alpha)} + \frac{|O_x^0| \times |O_y^1|}{\log(N)^b N \alpha} + \frac{|O_x^1| \times |O_y^0|}{\log(N)^b N \alpha} - \varepsilon \right) \\
&\geq \frac{1}{2N} \left(\frac{|I_x^0| \times |I_y^0| 2}{(d-1)^t(1-\alpha)} + \frac{|I_x^1| \times |I_y^1| 2}{(d-1)^t(1-\alpha)} + \frac{|O_x^0| \times |O_y^1| 2}{(d-1)^t \alpha} + \frac{|O_x^1| \times |O_y^0| 2}{(d-1)^t \alpha} - \varepsilon \right) \\
&:= \frac{1}{2N} (\mathcal{A}_1 + \mathcal{A}_2 + \mathcal{A}_3 + \mathcal{A}_4 - \varepsilon)
\end{aligned}$$

Et donc :

$$\sum_{y \in \mathcal{R} \setminus \mathcal{B}_K(x)} \left(\frac{1}{2N} - P^t(x, \sigma(y)) \right)_+ \leq \varepsilon + \sum_{y \in \mathcal{R} \setminus \mathcal{B}_K(x)} \frac{1}{2N} (1 - (\mathcal{A}_1 + \mathcal{A}_2 + \mathcal{A}_3 + \mathcal{A}_4))_+ \quad (33)$$

On va maintenant commencer par montrer que $|I_x^k| \simeq (1-\alpha)|\mathcal{F}_x \cap \mathcal{X}_k|$ et $|O_x^k| \simeq \alpha|\mathcal{F}_x \cap \mathcal{X}_k|$.

Conditionnellement à \mathcal{F}_x :

$$|I_x^k| \succeq \mathcal{Bin}(|\mathcal{F}_x \cap \mathcal{X}_k|, 1 - \alpha_t), \text{ avec } 1 - \alpha_t = \frac{N - p - 2(d-1)^{t/2}}{N}$$

L'inégalité de Hoeffding donne donc :

$$\begin{aligned} \mathbb{P}\left(\frac{I_x^k}{(d-1)^{t/2}} \leq \frac{(1-\alpha)|\mathcal{F}_x \cap \mathcal{X}_k|}{(d-1)^{t/2}} - \varepsilon | \mathcal{F}_x\right) &\leq \\ \mathbb{P}\left(\frac{\mathcal{Bin}(|\mathcal{F}_x \cap \mathcal{X}_k|, 1 - \alpha_t)}{(d-1)^{t/2}} \leq \frac{(1-\alpha_t)|\mathcal{F}_x \cap \mathcal{X}_k|}{(d-1)^{t/2}} - \varepsilon | \mathcal{F}_x\right) &= \\ \mathbb{P}\left(\frac{\mathcal{Bin}(|\mathcal{F}_x \cap \mathcal{X}_k|, 1 - \alpha_t)}{(d-1)^{t/2}} \leq \frac{(1-\alpha)|\mathcal{F}_x \cap \mathcal{X}_k|}{(d-1)^{t/2}} - \varepsilon + \frac{(\alpha - \alpha_t)|\mathcal{F}_x \cap \mathcal{X}_k|}{(d-1)^{t/2}} | \mathcal{F}_x\right) &\leq \exp\left(-\frac{2\varepsilon^2(d-1)^t}{|\mathcal{F}_x \cap \mathcal{X}_k|}\right) \\ &\leq \exp(-2\varepsilon^2(d-1)^{t/2}) \\ &= o(1/N^2) \end{aligned}$$

On obtient donc avec forte probabilité que pour tout $x \in \mathcal{R}$,

$$\frac{I_x^k}{(d-1)^{t/2}} > \frac{(1-\alpha)|\mathcal{F}_x \cap \mathcal{X}_k|}{(d-1)^{t/2}} - \varepsilon + \frac{2|\mathcal{F}_x \cap \mathcal{X}_k|}{N} > \frac{(1-\alpha)|\mathcal{F}_x \cap \mathcal{X}_k|}{(d-1)^{t/2}} - \varepsilon + o(1)$$

De même $|O_x^k| \succeq \mathcal{Bin}(|\mathcal{F}_x \cap \mathcal{X}_k|, \frac{p-2(d-1)^{t/2}}{N})$ et les mêmes arguments donnent qu'avec probabilité qui tend vers 1 pour tout $x \in \mathcal{R}$:

$$\frac{O_x^k}{(d-1)^{t/2}} > \frac{\alpha|\mathcal{F}_x \cap \mathcal{X}_k|}{(d-1)^{t/2}} - \varepsilon + o(1)$$

En injectant dans 33 on trouve

$$\begin{aligned} \sum_{y \in \mathcal{R} \setminus \mathcal{B}_K(x)} \left(\frac{1}{2N} - P^t(x, \sigma(y))\right)_+ &\leq 16\varepsilon - 8\varepsilon^2 + \sum_{y \in \mathcal{R} \setminus \mathcal{B}_K(x)} \frac{1}{2N} \left(1 - 2(1-\alpha) \frac{|\mathcal{F}_x \cap \mathcal{X}_0| \times |\mathcal{F}_y \cap \mathcal{X}_0|}{(d-1)^t} \right. \\ &\quad - 2(1-\alpha) \frac{|\mathcal{F}_x \cap \mathcal{X}_1| \times |\mathcal{F}_y \cap \mathcal{X}_1|}{(d-1)^t} \\ &\quad - 2\alpha \frac{|\mathcal{F}_x \cap \mathcal{X}_0| \times |\mathcal{F}_y \cap \mathcal{X}_1|}{(d-1)^t} \\ &\quad \left. - 2\alpha \frac{|\mathcal{F}_x \cap \mathcal{X}_1| \times |\mathcal{F}_y \cap \mathcal{X}_0|}{(d-1)^t} \right)_+ + o(1) \end{aligned}$$

En effet si on développe le premier produit on obtient

$$-\frac{2}{1-\alpha} \left(\frac{(1-\alpha)|\mathcal{F}_x \cap \mathcal{X}_0|}{(d-1)^{t/2}} - \varepsilon + o(1) \right) \left(\frac{(1-\alpha)|\mathcal{F}_y \cap \mathcal{X}_0|}{(d-1)^{t/2}} - \varepsilon + o(1) \right) =$$

$$\begin{aligned} & \frac{2\varepsilon|\mathcal{X}_0 \cap \mathcal{F}_x|}{(d-1)^{t/2}} + \frac{2\varepsilon|\mathcal{X}_0 \cap \mathcal{F}_y|}{(d-1)^{t/2}} - \frac{2}{1-\alpha}\varepsilon^2 + \frac{(1-\alpha)|\mathcal{X}_0 \cap \mathcal{F}_x| \times |\mathcal{X}_0 \cap \mathcal{F}_y|}{(d-1)^t} + o(1) \\ & \leq 4\varepsilon - 2\varepsilon^2 + \frac{(1-\alpha)|\mathcal{X}_0 \cap \mathcal{F}_x| \times |\mathcal{X}_0 \cap \mathcal{F}_y|}{(d-1)^t} + o(1) \end{aligned}$$

Les autres se traitent de la même façon.

Notons $Q_x = \mathbb{P}_x^G(X_{t/2} \in \mathcal{X}_0)$ et $Q_y = \mathbb{P}_y^G(X_{t/2} \in \mathcal{X}_0)$

Lemme 3.3.3.

$$\mathbb{P}\left(\frac{|\mathcal{F}_x \cap \mathcal{X}_0|}{(d-1)^{t/2}} \leq Q_x - \varepsilon\right) = o(1/N^2)$$

et de même

$$\mathbb{P}\left(\frac{|\mathcal{F}_x \cap \mathcal{X}_1|}{(d-1)^{t/2}} \leq 1 - Q_x - \varepsilon\right) = o(1/N^2)$$

Démonstration.

$$\begin{aligned} \frac{|\mathcal{F}_x \cap \mathcal{X}_1|}{(d-1)^{t/2}} & \geq Q_x - \sum_{\substack{z \in \mathcal{B}_x(t/2) \\ z \text{ rouge}}} \mathbb{P}_x^G(X \text{ touche } z \text{ avant } t/2) \\ & \geq Q_x - \frac{|\text{rouge}|}{(d-1)^K} \end{aligned}$$

Car x est une K -racine donc un cycle ne peut se créer qu'à distance au moins K de x .

Le nombre de demi-arêtes rouges dans l'arbre généré autour de x est stochastiquement dominé par une binomiale $\mathcal{Bin}(2(d-1)^{t/2}, \frac{(d-1)^{t/2}}{N-2(d-1)^{t/2}})$.

Par l'inégalité de Bennett :

$$\begin{aligned} \mathbb{P}(\mathcal{Bin}(n, p) - np \geq np) & \leq \exp\left(-\frac{(np)^2}{2(np + \frac{np}{3})}\right) \\ & \leq \exp(-np/4) \end{aligned}$$

Ici, $np = \frac{2(d-1)^t}{N-2(d-1)^{t/2}} \sim 2(\log(N)^b)$ donc

$$\mathbb{P}\left(\frac{|\mathcal{F}_x \cap \mathcal{X}_1|}{(d-1)^{t/2}} \leq Q_x - \frac{4\log(N)^b}{N^{1/6}}\right) = o(1/N^2)$$

□

On obtient donc avec probabilité qui tends vers 1 , pour tout $x \in \mathcal{R}$

$$\begin{aligned} \sum_{y \in \mathcal{R} \setminus \mathcal{B}_k(x)} \left(\frac{1}{2N} - P^t(x, \sigma(y)) \right)_+ &\leq 16\varepsilon - 16\varepsilon^2 + \sum_{y \in \mathcal{R} \setminus \mathcal{B}_k(x)} \frac{1}{2N} (1 - 2(1 - \alpha)Q_x Q_y \\ &\quad - 2(1 - \alpha)(1 - Q_x)(1 - Q_y) - 2\alpha Q_x(1 - Q_y) - 2\alpha Q_y(1 - Q_x))_+ \\ &\leq 16\varepsilon - 16\varepsilon^2 + \sum_{y \in \mathcal{R} \setminus \mathcal{B}_k(x)} \frac{1}{2N} (1 - 2\alpha)((2Q_x - 1)(2(1 - Q_y) - 1))_+ \end{aligned}$$

À présent, on va montrer que Q_x est très proche de $1/2$.
Soit $d > 0$ et $m = \lfloor d \log(N) \rfloor$, par l'inégalité de Markov on a

$$\mathbb{P}(Q_x > 1/2 + \varepsilon) \leq \frac{\mathbb{E}[Q_x^m]}{(1/2 + \varepsilon)^m}. \quad (34)$$

Soit $X^{(1)}, X^{(2)}, \dots, X^{(m)}$, m marches aléatoire sans rebroussement indépendantes sur G . Alors :

$$\mathbb{E}[Q_x^m] = \mathbb{P}_x(X_{t/2}^{(1)} \in \mathcal{X}_0, X_{t/2}^{(2)} \in \mathcal{X}_0, \dots, X_{t/2}^{(m)} \in \mathcal{X}_0) = \prod_{i=1}^m \mathbb{P}_x(X_{t/2}^{(i)} \in \mathcal{X}_0 | X_{t/2}^{(1)}, X_{t/2}^{(2)}, \dots, X_{t/2}^{(i-1)} \in \mathcal{X}_0)$$

Pour évaluer cette probabilité on va générer les marches les une après les autres avec le graphe. Pour cela, à l'étape k si la i -ème marche $X_{k-1}^{(i)} \in \mathcal{X}_i$:

1. Soit $X_{k-1}^{(i)}$ n'est pas appariée :

On tire alors une Bernoulli $\mathcal{B}_k^{(i)}$ de paramètre $\alpha_k = \frac{p - p_k^i}{N - N_k^i}$ (p_k^i étant le nombre de demi-arêtes "out" déjà révélées et N_k^i le nombre de demi-arêtes de \mathcal{X}_i déjà révélées). Si $\mathcal{B}_k^{(i)} = 0$ on apparie $X_{k-1}^{(i)}$ avec une demie-arête uniformément choisie parmi les demie-arêtes "in" non-appariées de \mathcal{X}_i . Sinon, on l'apparie avec une demi-arête "out" de \mathcal{X}_{1-i} .

2. Soit $X_{k-1}^{(i)}$ est déjà appariée :

On choisie $X_k^{(i)}$ uniformément sur les voisins de $\sigma(X_{k-1}^{(i)})$.

Supposons que les $l - 1$ premières marches vérifient la propriété voulue. La l -ème marche peut :

- Soit suivre la trajectoire d'une des $l - 1$ précédentes sur plus de $c \log \log(N)$ étapes ce qui se produit avec une probabilité inférieure à $\frac{m}{(d-1)^{c \log \log(N)}}$.
- Soit se retrouver en un temps $s < c \log \log(N)$ sur une demie-arête z qui n'avait pas encore été révélée. À partir de la on créer un couplage entre $(X_k^l)_{k \geq s}$ et une suite $(U_k^l)_{k \geq s}$ de demie-arêtes générées en tirant à chaque étape une demi-arête uniformément dans \mathcal{X}_i si la Bernoulli utilisée pour

g n rer la marche al atoire vaut 0 et uniform ment dans \mathcal{X}_{1-i} sinon. Ce couplage est valide jusqu'au premier temps T o  on choisi une-demi ar te d j  appari e (cas 2. dans la g n ration de la marche).

  l' tape k il y'a au plus $2k$ demi-ar tes appari es donc par une borne union on   :

$$\mathbb{P}(T \leq t) \leq \frac{t^2}{N}$$

Et ainsi,

$$\left| \mathbb{P}(X_{t/2}^{(l)} \in \mathcal{X}_0 | X_{t/2}^{(1)}, X_{t/2}^{(2)}, \dots, X_{t/2}^{(l-1)} \in \mathcal{X}_0) - \mathbb{P}(U_{t/2}^{(l)} \in \mathcal{X}_0 | U_{t/2}^{(1)}, U_{t/2}^{(2)}, \dots, U_{t/2}^{(l-1)} \in \mathcal{X}_0) \right| \leq \frac{(t/2 - s)^2}{N}$$

On a de plus :

$$\mathbb{P}(U_{t/2}^{(l)} \in \mathcal{X}_0 | U_{t/2}^{(1)}, U_{t/2}^{(2)}, \dots, U_{t/2}^{(l-1)} \in \mathcal{X}_0) = \begin{cases} \mathbb{P}(\sum_{i=s}^{t/2} \mathcal{B}_k^{(l)} \text{ est paire}) & \text{si } z \in \mathcal{X}_0 \\ \mathbb{P}(\sum_{i=s}^{t/2} \mathcal{B}_k^{(l)} \text{ est impaire}) & \text{sinon.} \end{cases}$$

Consid rions une suite IID $(\mathcal{B}_{k \geq 1}^*)$ de variables al atoire de Bernouilli de param tre α . Alors :

$$\left\| \sum_{i=s}^{t/2} \mathcal{B}_k^{(l)} - \sum_{i=1}^{t/2-s} \mathcal{B}_k^* \right\|_{TV} \leq \sum_{i=1}^{t/2-s} \|\mathcal{B}_{k+s}^{(l)} - \mathcal{B}_k^*\|_{TV} \leq (t/2-s) \max\left\{ \frac{p - (t/2 - s)}{N}, \frac{p}{N - (t/2 - s)} \right\}$$

Or d'apr s l'hypoth se sur α le terme de droite tend vers 0. On se ram ne donc au calcul de la probabilit  pour qu'une binomiale soit paire.

$$\begin{aligned} \mathbb{P}(\mathcal{Bin}(t/2 - c \log \log(N), \alpha) = \text{paire}) &= 1/2(1 + (1 - 2\alpha)^{t/2 - c \log \log(N)}) \\ &= 1/2(1 + (1 - \frac{2}{o(\log(N))})^{t/2 - c \log \log(N)}) \\ &= 1/2 + o(1) \end{aligned}$$

Et

$$\mathbb{P}(\mathcal{Bin}(t/2 - c \log \log(N), \alpha) = \text{impaire}) = 1 - \mathbb{P}(\mathcal{Bin}(t/2 - c \log \log(N), \alpha) = \text{paire}) = 1/2 - o(1)$$

Finalement on trouve

$$\mathbb{E}[Q_x^m] \leq \prod_{i=1}^m \left(\frac{m}{(d-1)^{c \log \log(N)}} + o(1) + 1/2 + o(1) \right) \leq (1/2 + o(1))^m$$

Puis en injectant dans 34 on obtient pour d suffisamment grand :

$$\mathbb{P}(Q_x > 1/2 + \varepsilon) \leq \left(\frac{1/2 + o(1)}{1/2 + \varepsilon} \right)^m = o\left(\frac{1}{N^2}\right)$$

On conclue la démonstration en remarquant que $1 - Q_y = \mathbb{P}_y^G(X_{t/2} \in \mathcal{X}_1)$ et en appliquant le même raisonnement. \square

Lemme 3.3.2 :

Démonstration. Soit $K = \frac{\log(N)}{6 \log(d-1)}$, $s = a \log(\log(N))$ et $l = \frac{\log(N)}{5 \log(d-1)}$. On génère à nouveau la boule sans rebroussement de rayon l autour de x_0 séquentiellement. On fait au total au plus : $1 + (d-1) + (d-1)^2 + \dots + (d-1)^{l-1} \leq (d-1)^l$ appariements.

À chaque appariements il y a au plus $(d-1)^l$ demi-arêtes non appariées mais déjà révélées (celles qu'on n'apparie pas pour ne pas créer de cycle). Le nombre de mauvais évènements durant la génération de $\mathcal{B}_l(x_0)$ est stochastiquement dominée par une binomiale $\text{Bin}((d-1)^l, \frac{(d-1)^l}{N-2(d-1)^l})$. On utilise que $\mathbb{P}(\text{Bin}(n, p) \geq 2) \leq (np)^2$ on a donc :

$$\begin{aligned} \mathbb{P}(\text{nombre de mauvais matching} \geq 2) &\leq \left(\frac{(d-1)^{2l}}{N-2(d-1)^l} \right)^2 = O(N^{-2}(d-1)^{\frac{4 \log(N)}{5 \log(d-1)}}) \\ &= O(N^{-6/5}) \\ &= o(1/N) \end{aligned}$$

Par une borne union,

$$\begin{aligned} &\mathbb{P}(\exists x_0 \in \mathcal{X}, \mathcal{B}_l(x_0) \text{ contient strictement plus d'un cycle}) \\ &\leq N \mathbb{P}(\mathcal{B}_l(x_0) \text{ contient strict. plus d'un cycle}) \leq N \times N^{-6/5} \end{aligned} \quad (35)$$

Donc avec probabilité qui tend vers 1 chaque sommet à au plus un cycle dans sa boule de rayon l . Fixons maintenant un graphe G vérifiant cette propriété.

$$\mathbb{P}_x(X_t \text{ n'est pas une } K\text{-racine} \mid G) \leq 2 \times 2^{-t}, \forall t \leq l - K = \frac{\log(N)}{30 \log(d-1)}.$$

En effet si $\mathcal{B}_l(x)$ est déjà un arbre c'est trivial. Sinon, $\mathcal{B}_l(x)$ contient un unique cycle \mathcal{C} . Soit $d(z, \mathcal{C})$ la distance minimale d'un chemin sans rebroussement entre z et \mathcal{C} . Si $\exists t < l - k, d(X_t, \mathcal{C}) < d(X_{t+1}, \mathcal{C})$, alors $X_{t+1}, X_{t+2}, \dots, X_{l-K}$ sont des K -racines. En effet dès que la marche s'éloigne de \mathcal{C} sur un des arbres disjoints partant de \mathcal{C} , elle ne peut que continuer à s'éloigner de \mathcal{C} . Or la probabilité que $d(X_{t+1}, \mathcal{C}) = d(X_t, \mathcal{C}) + 1$ est au moins $1/2$ car $d \geq 3$, sauf si $d(X_t, \mathcal{C}) = 1$ ce qui ne peut arriver qu'une fois.

En prenant $t = s$ on trouve $\mathbb{P}_x(X_s \text{ n'est pas une } K\text{-racine} \mid G) \leq 2 \times 2^{-s}$ \square

Théorème 3.3.4. Soit $\alpha \ll \frac{1}{\log(dn)}$ et $G_n \sim \mathcal{G}_\alpha(n, d)$. Alors :

Démonstration. \square

Ouverture :

Pour 3.3.1 on s'est limité à une fenêtre d'ordre $\log(\log(N))$ pour faciliter les démonstrations, en utilisant des argument plus fins comme dans [4] on peut imaginer adapter la preuve pour retrouver une fenêtre constante.

On peut également s'intéresser au cas non régulier, au cas à plus de deux communautés voir à un nombre de communautés aléatoire.

Enfin nous avons traité les régimes $\alpha \ll \frac{1}{\log_{d-1}(N)}$ et $\alpha \gg \frac{1}{\log_{d-1}(N)}$, il reste à étudier le régime "critique" $\alpha \sim \frac{1}{\log_{d-1}(N)}$

Références

- [1] Anna Ben-Hamou. Temps de mélange de marches aléatoires sur des graphes aléatoires.
- [2] Anna Ben-Hamou and Justin Salez. Cutoff for non-backtracking random walks on sparse random graphs. *The Annals Probability*, 2017.
- [3] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, 2017.
- [4] E. Lubetsky and A. Sly. Cutoff phenomena for random walks on random regular graphs. *Duke Mathematical Journal*, 2010.
- [5] Remco Van Der Hofstad. *Random Graphs and Complex Networks*. 2013.