

École Normale Supérieure
Département De Mathématiques et Applications

Procédure d'isolement d'un sommet appliqué au graphe d'Erdős–Rényi

Rapport de Stage de M1

Ons Rameh
ons.rameh@ens.fr

Sous la direction de
M. Cyril Marzouk

Résumé

Le graphe d'Erdős–Rényi est l'un des plus importants modèles de graphe aléatoire. Dans un premier temps, on se propose d'expliquer certaines de ses caractéristiques et de donner sa limite d'échelle au sens de Gromov-Hausdorff.

On introduit ensuite l'algorithme d'isolation d'un sommet et formule une conjecture sur le nombre d'étapes pour isoler un sommet dans le graphe d'Erdős–Rényi. Finalement, on explique une piste de résolution qui n'a pas abouti.

Stage réalisé du 8 Février 2021 au 8 Juin 2021

Table des matières

1	Introduction	2
2	Le graphe d’Erdős–Rényi et ses premières propriétés	4
2.1	Algorithme de parcours d’un graphe	4
2.1.1	Parcours en profondeur	4
2.1.2	Parcours en largeur	6
2.2	Transition de phase et fenêtre critique :	7
3	Comportement critique de la taille et du surplus	8
3.1	Énoncés des théorèmes	8
3.2	Preuve du théorème 3.2	9
3.3	Preuve du corollaire 3.1	12
4	Étude de la structure du graphe dans la fenêtre critique	13
4.1	Topologie de Gromov-Hausdorff	13
4.2	Arbres réels et leurs constructions	14
4.3	La limite continue du graphe d’Erdős–Rényi	16
5	Justification heuristique de la conjecture	21
5.1	Etude des moments :	22
5.2	Lien avec les arbres branchants Markoviens	23
5.3	Limite d’échelle des chaînes de Markov décroissantes	25
6	Essai de preuve de la conjecture	27
6.1	Convergence de la fonction génératrice	30
6.2	Étude de la compacité de $Y_{n,s}$	31
6.3	Unicité de la limite	32
6.4	De la convergence de $Z_{n,s}$ vers la convergence de $B_{n,s}$	34
7	Conclusion	35
	Annexes	37

1 Introduction

Un graphe non orienté est la donnée d'un ensemble de sommets pouvant être liés par des arêtes. Plus formellement, un graphe G est la donnée d'un ensemble $\mathcal{V} = \{v_1, \dots, v_n\}$ et d'un sous ensemble $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. On dit que v_i et v_j sont reliés si et seulement si $(v_i, v_j) \in \mathcal{E}$. On considère pour simplifier $\mathcal{V} = [n] = \{1 \dots n\}$ pour les graphes à n sommets. Le graphe est enraciné si on s'est fixé un sommet particulier qu'on appellera racine ρ . Par ailleurs, un arbre est un graphe connexe et acyclique. La notion de graphe aléatoire a été introduite par Paul Erdős et Alfred Rényi dans une série d'articles publiée entre 1959 et 1968. Ils ont été reçus avec une énorme attention et ont ouvert la voie à d'autres mathématiciens pour qu'ils s'impliquent dans l'étude des propriétés des graphes aléatoires.

Le graphe d'Erdős–Rényi $ER(n, p)$ est défini la première fois en 1960 par Erdős et Rényi dans leur article “On the evolution of random graphs” [12]. Dans ce modèle, on choisit un graphe aléatoire à n sommets en décidant de joindre deux sommets avec une probabilité p de manière indépendante des autres paires de sommets. De prime abord, ce modèle a l'air simple. Comme il est discret, on a tendance à vouloir développer des calculs. Par exemple, la probabilité d'avoir k arêtes dans le graphe est explicitée par :

$$\frac{\binom{\binom{n}{2}}{k}}{2^{\binom{n}{2}}}$$

La méthode purement combinatoire échoue dans la mise en évidence des propriétés subtiles de ce modèle. En effet, la première caractéristique du modèle binomial d'Erdős–Rényi est l'apparition d'une transition de phase. Elle a été mise en évidence dans l'article [12]. Si la valeur de la probabilité de connexion est de l'ordre de $\frac{a}{n}$ avec $a < 1$, la taille de la plus grande composante connexe est de l'ordre de $O(\log(n))$. Si $a > 1$, on observe l'apparition d'une composante géante d'une taille de l'ordre n . Ce phénomène est perceptible à travers la figure (1). Pour le régime critique c'est-à-dire $a = 1$, la taille de la plus grande composante connexe est de l'ordre $n^{2/3}$ ce qui justifie le terme de “double saut” quand il s'agit de caractériser cette transition de phase.

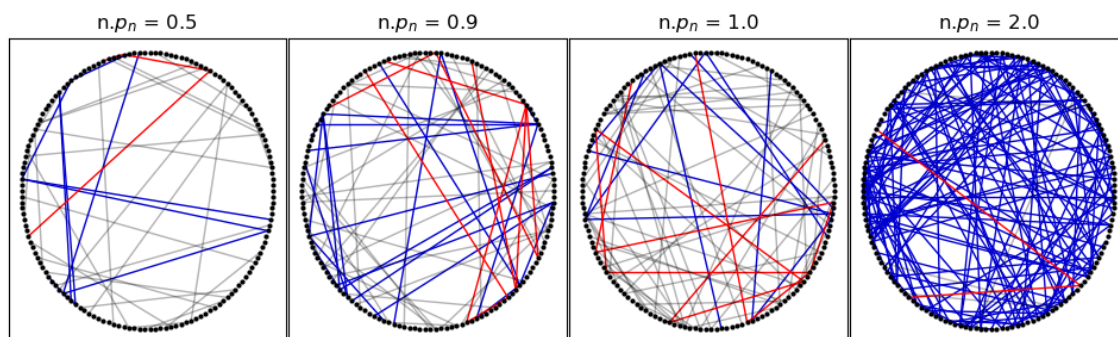


FIGURE 1 – Simulation de graphe d'Erdős–Rényi pour $n = 150$ avec des probabilités différentes. On affiche en bleu la plus grande composante connexe et en rouge, la deuxième plus grande composante connexe. Le reste des arêtes sont tracés en gris.

Aldous étudie plus précisément le régime critique en explicitant le comportement des tailles des composantes connexes ainsi que leurs surplus¹. Son article [7] est sûrement l'un des résultats les plus fondamentaux dans l'étude des régimes critiques. En réalité, l'aspect le plus remarquable est d'exprimer les graphes à travers des processus stochastiques qu'on maîtrise. On peut alors réinterpréter les propriétés du monde des processus à travers le prisme des graphes aléatoires. Addario-Berry, Broutin et Goldschmidt étendent l'étude du régime critique en s'intéressant à l'aspect géométrique. Dans l'article [2], ils prouvent que les composantes connexes du graphe d'Erdős–Rényi vues comme des espaces métriques munis de la distance naturelle renormalisée par $n^{-1/3}$ admettent une limite au sens de la convergence de Gromov-Hausdorff. L'objet limite est essentiellement un arbre réel continu aléatoire où on identifie des paires de points. Dans [1], ils donnent aussi une méthode de construction qui permet de comprendre la structure des cycles à l'infini.

Par ailleurs, le sujet des algorithmes de destruction des arbres a suscité l'intérêt de plusieurs mathématiciens. Il a été introduit pour la première fois par Meir et Moon dans [21] et [22] en 1969. Initialement, la procédure s'effectue sur un arbre. Soit T un arbre enraciné quelconque, à chaque itération, on sélectionne aléatoirement une arête qu'on coupe. On obtient ainsi deux arbres T_1 et T_2 . A l'issue de cette étape, on ne garde que la composante connexe qui contient la racine. On s'arrête quand la composante connexe contient uniquement la racine. L'objet d'étude de Meir et Moon est le nombre d'étapes nécessaire pour isoler la racine. Ils appliquent cet algorithme à l'arbre de Cayley² T_n et démontrent le théorème suivant :

Théorème 1.1 (Meir et Moon [21]). *On note $\mu(n)$ l'espérance du nombre d'étape et $\sigma^2(n)$ sa variance. On obtient :*

$$\mu(n) \sim \sqrt{\frac{1}{2}\pi n} \quad \text{et} \quad \sigma^2(n) \sim (2 - \frac{\pi}{2})n$$

Ultérieurement, Panholzer dans [23] prouve que la loi limite est la loi de Rayleigh en se basant sur l'étude des limites des moments. Janson [17] donne une autre preuve pour une classe plus large, celle des arbres de Galton–Watson à variance finie conditionné dans le régime critique. Cette classe englobe les arbres de Cayley en considérant la loi de Poisson de paramètre 1.

Théorème 1.2 (Panholzer [23]). *On note κ_n le nombre d'étapes pour isoler un sommet dans un arbre de Cayley T_n . On a :*

$$\frac{\kappa_n}{\sqrt{n}} \xrightarrow{d} R$$

où R est une variable aléatoire qui suit la loi de Rayleigh de densité $xe^{-\frac{x^2}{2}}$

L'objectif de ce stage est d'étudier le comportement asymptotique du nombre d'étape pour isoler un sommet dans le graphe $ER(n, p_n)$ en se plaçant dans la fenêtre critique étudiée par Aldous.

1. le nombre minimal d'arête qu'il faut enlever pour obtenir un arbre
 2. un arbre aléatoire uniforme à n sommets

Conjecture 1. Soit $t \in \mathbb{R}$, on considère $p(n) = \frac{1}{n} + \frac{t}{n^{4/3}}$ et on note κ_n^{ER} le nombre d'étapes pour isoler un sommet dans la plus grande composante connexe. Alors, il existe $\alpha_t \in \mathbb{R}$

$$\frac{\kappa_n^{ER}}{n^{1/3}} \xrightarrow{d} \alpha_t R$$

où R est une loi de Rayleigh.

Dans la section (2), on définira les méthodes d'exploration et les marches associées. Puis, on donnera les premières caractéristiques du modèle d'Erdős–Rényi. Ensuite, dans la section (3), on s'intéressera au comportement asymptotique de la suite des tailles et des surplus des composantes connexes. Par ailleurs, on définira la notion de convergence de Gromov-Hausdorff ainsi que la notion d'arbre continu réel dans la section (4) pour pouvoir étudier plus en détail l'aspect géométrique. Puis, on expliquera dans la section (5) l'intuition derrière cette conjecture. Enfin, on donnera dans la section (6) un schéma potentiel de preuve qui n'a pas encore abouti.

2 Le graphe d'Erdős–Rényi et ses premières propriétés

Le modèle fait apparaître une transition de phase lorsque la probabilité p varie ainsi qu'une limite lorsqu'on n tend vers l'infini dans la fenêtre critique. Afin d'étudier les propriétés de ce graphe, on commence par définir l'arbre de parcours en profondeur et en largeur.

2.1 Algorithme de parcours d'un graphe

On définit deux manières déterministes pour parcourir un graphe G . On considère que les sommets de G sont initialement à valeurs dans $[n]$.

2.1.1 Parcours en profondeur

On note \mathcal{O}_i la pile de sommets actifs à l'étape i et \mathcal{A}_i l'ensemble de sommets explorés à l'étape i .

A l'étape 0, on initialise la pile \mathcal{O}_0 avec le sommet (1). L'ensemble \mathcal{A}_0 est vide et $c(0) = 1$. A chaque étape, deux cas se présentent :

- La pile \mathcal{O}_i n'est pas vide. On se donne v_i , l'élément en haut de la pile et \mathcal{N}_i l'ensemble de ses voisins dans $[n] \setminus \mathcal{O}_i \cup \mathcal{A}_i$. A l'issue de cette étape, le sommet v_i est "exploré" et on pose $\mathcal{A}_{i+1} = \mathcal{A}_i \cup \{v_i\}$. On rajoute les éléments de \mathcal{N}_i à \mathcal{O}_i au sommet de la pile dans l'ordre croissant.
- La pile \mathcal{O}_i est vide. Alors, on fait les mêmes opérations en se donnant v_i le plus petit élément de $[n] \setminus \mathcal{A}_i$. On pose $c_{i+1} = c_i + 1$

A la fin de cette procédure, c_n correspond au nombre de composantes connexes du graphe G . On a aussi défini un nouvel ordre $\{v_0, \dots, v_{n-1}\}$ sur les sommets de G . On construit la forêt d'exploration en profondeur **oDFS**(G) où l'arête xy existe si

seulement si x est exploré avant y et $y \in \mathcal{N}_i$ ou réciproquement.

On définit aussi la marche en profondeur par

$$X_i = |\mathcal{O}_i \setminus v_i| - (c_i - 1)$$

ainsi que la fonction hauteur par $H(i) = \text{distance du sommet } v_i \text{ à la racine}$.

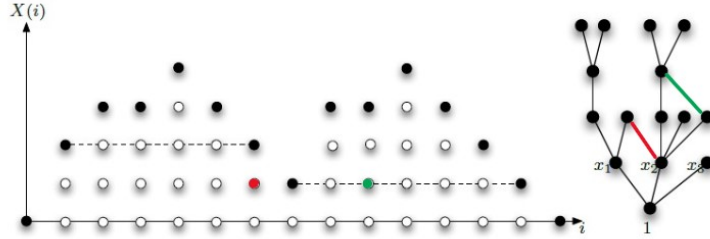


FIGURE 2 – En noir, la marche aléatoire en profondeur associée au graphe à droite. Les points coloriés en rouge et en vert correspondent aux arêtes en surplus. La figure est extraite de l'article [2].

Intuitivement, lorsqu'on parcourt le graphe, l'arête uv est en surplus si lors de l'exploration, les deux sommets ont été actifs simultanément. En effet, si v est exploré avant le sommet u , deux cas se présentent :

- u est actif : uv est en surplus et à l'étape précédente, u et v étaient actifs.
- u n'est actif et il sera un enfant de u et uv appartiendra à $\mathbf{oDFS}(G)$.

Cette remarque justifie la définition suivante :

Définition 2.1. *L'ensemble des arêtes admissibles d'un arbre T est l'ensemble :*

$$\mathcal{P}(T) = \{uv \notin T; \exists i \text{ tel que } u, v \in \mathcal{O}_i\}.$$

Si on se donne un arbre T , l'ensemble des arêtes qu'on peut ajouter pour obtenir un graphe connexe G tel que $\mathbf{oDFS}(G) = T$ est l'ensemble des arêtes admissibles. Le cardinal de cet ensemble peut être calculé par le lemme suivant démontré dans [2] :

Lemme 2.1 (Lemme 6 [2]). *Le cardinal de $\mathcal{P}(T)$ est égal à $a(T) = \sum_{i=1}^{n-1} X(i)$ où X est la marche en profondeur associée à T .*

On remarque que les fonctions X et H ne tiennent pas compte des arêtes en surplus. Pour ne pas perdre la structure du graphe, on code les arêtes en surplus avec un ensemble de marques $\mathcal{Q}(G)$, voir un exemple à la figure 2. On met une marque à la position (i, j) si $j \leq X(i)$ et s'il y a une arête entre le sommet v_i et $v_{k(i,j)}$ où $k(i, j) = \inf\{k > i, X(k) = j\}$.

Définition 2.2. *Soit $\mathcal{Q} \subset \mathbb{Z} \times \mathbb{Z}$ et T un arbre discret, le graphe $G(T, \mathcal{Q})$ est construit à partir de l'arbre T en ajoutant les arêtes $v_i v_{k(i,j)}$ pour $(i, j) \in \{(i', j') \in \mathcal{Q} : 0 \leq i' < i; 0 \leq j' < X(i')\}$.*

On remarque que $G(\mathbf{oDFS}(G), \mathcal{Q}(G)) = G$.

2.1.2 Parcours en largeur

La seule différence entre les deux procédures consiste à remplacer le mot “pile” par “file” c’est-à-dire qu’on traite les sommets les moins récents en priorité au lieu de manipuler le dernier sommet introduit dans la pile.

On note \mathcal{O}_i la file de sommets actifs à l’étape i et \mathcal{A}_i l’ensemble de sommets explorés à l’étape i .

À l’étape 0, on initialise la file \mathcal{O}_0 avec le sommet (1). L’ensemble \mathcal{A}_0 est vide et $c(0) = 1$. À chaque étape, deux cas se présentent :

- La file \mathcal{O}_i n’est pas vide. On pose v_{i+1} , l’élément au début de la file et \mathcal{N}_{i+1} l’ensemble de ses voisins dans $[n] \setminus \mathcal{O}_i \cup \mathcal{A}_i$. Le sommet v_{i+1} est “exploré” et on pose $\mathcal{A}_{i+1} = \mathcal{A}_i \cup \{v_{i+1}\}$. On rajoute les éléments de \mathcal{N}_{i+1} à \mathcal{O}_i à la fin de la file dans l’ordre croissant.
- La file \mathcal{O}_i est vide. Alors, on fait les mêmes opérations en se donnant v_{i+1} le plus petit élément de $[n] \setminus \mathcal{A}_i$. On pose $c_{i+1} = c_i + 1$

On a défini un nouvel ordre $\{v_1, \dots, v_n\}$ sur les sommets de G ainsi qu’un ordre sur les composantes connexes $\{\mathcal{C}_i\}$.

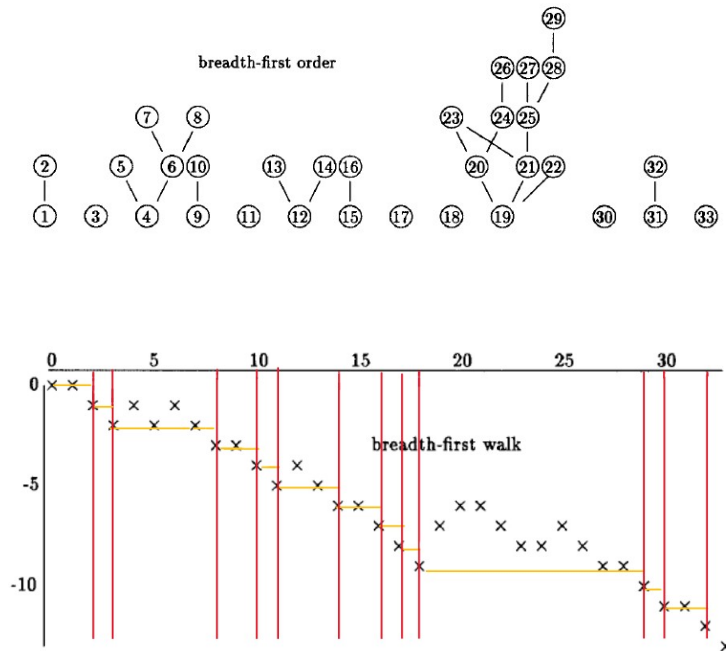


FIGURE 3 – La marche aléatoire associée au graphe ci-dessus. Les sommets sont numérotés selon le parcours en largeur. Les traits rouges délimitent les composantes connexes. La figure est extraite de l’article d’Aldous [7].

Par ailleurs, on définit la marche aléatoire en largeur $Y(i) = |\mathcal{O}_i| - c_i$ ainsi que la forêt d’exploration en largeur $\mathbf{oBFS}(G)$ où l’arête xy existe si et seulement si x est exploré avant y et $y \in \mathcal{N}_i$ ou réciproquement.

On note :

$$\zeta(j) = |\mathcal{C}_1| + \cdots + |\mathcal{C}_j|$$

et on remarque que :

$$Y(\zeta(j)) = -j \text{ et } Y(i) \geq -j \text{ pour tout } \zeta(j) < i < \zeta(j+1).$$

La taille de la composante connexe correspond au temps que passe la marche aléatoire au dessus de son ancien minimum. Intuitivement, l'étude du comportement de la marche aléatoire permettra de retrouver le comportement asymptotique des tailles des composantes connexes.

2.2 Transition de phase et fenêtre critique :

Le comportement des clusters dépend radicalement de la valeur de $p(n)n$. Alors que la plus grande composante connexe \mathcal{C}_{max} est de l'ordre de $\log(n)$ lorsque $p(n)n$ est inférieure à 1, une composante connexe géante de taille $\Theta(n)$ apparaît lorsque $p(n)n > 1$. Ce changement abrupte de comportement justifie l'appellation transition de phase.

Théorème 2.1 ([16]). *On note $a = np(n)$*

- *Pour $a > 1$, pour tout $\nu \in (\frac{1}{2}, 1)$, il existe δ tel que*

$$\mathbb{P}(|\mathcal{C}_{max}| - \alpha_a n| \geq n^\nu) = O(n^{-\delta}). \quad (1)$$

- *Pour $a < 1$, pour tout $\alpha > \frac{1}{a-1-\log(a)}$, il existe $\delta > 0$ tel que*

$$\mathbb{P}(|\mathcal{C}_{max}| \geq \alpha \log(n)) = O(n^{-\delta}). \quad (2)$$

- *Pour $a < 1$, pour tout $\alpha < \frac{1}{a-1-\log(a)}$, il existe $\delta > 0$ tel que*

$$\mathbb{P}(|\mathcal{C}_{max}| \leq \alpha \log(n)) = O(n^{-\delta}). \quad (3)$$

La preuve de ce théorème se base sur la comparaison entre le graphe d'Erdős–Rényi et le processus de branchement de Galton–Watson. A chaque étape i de l'exploration, chaque sommet qui n'a pas été vu a une probabilité $\frac{a}{n}$ d'être un voisin de v_i . On en déduit que $\mathcal{L}(Y_i - Y_{i-1} | Y_1, \dots, Y_{i-1}) = \mathcal{B}(n - (i-1) - Y_{i-1}, \frac{a}{n})$. Intuitivement, au début de l'exploration et pour n assez grand, le graphe se comporte comme un processus de branchement poissonnien de moyenne a . Le lemme suivant précise cette relation :

Lemme 2.2. *Pour tout k, n, p , on considère $\mathcal{C}(1)$ la composante connexe de $ER(n, p)$ contenant le sommet 1 alors*

$$\begin{aligned} \mathbb{P}(|\mathcal{C}(1)| \geq k) &\leq \mathbb{P}_{n,p}(T^\geq \geq k) \\ \mathbb{P}(|\mathcal{C}(1)| \geq k) &\geq \mathbb{P}_{n-k,p}(T^\leq \geq k) \end{aligned}$$

où T^\geq, T^\leq sont des processus de branchement binomiaux de paramètre (n, p) et $(n-k, p)$ respectivement.

3 Comportement critique de la taille et du surplus

Dans le régime critique, on évite à la fois un réseau très peu dense avec beaucoup de composantes connexes et un effet cluster où la majorité des points sont reliés. Il paraît donc naturel d'étudier le comportement critique du graphe d'Erdős–Rényi. On considère une probabilité de connexion égale à $\frac{1}{n} + \frac{t}{n^{4/3}}$. Ce choix est en partie justifié par le théorème suivant démontré dans [16] :

Théorème 3.1. *Pour tout t , on considère une probabilité de connexion égale à $p(n) = \frac{1}{n} + \frac{t}{n^{4/3}}$ alors il existe une constante b telle que pour tout $\omega > 1$:*

$$\mathbb{P} \left(\frac{n^{2/3}}{\omega} \leq |\mathcal{C}_{max}| \leq \omega n^{2/3} \right) \geq 1 - \frac{b}{\omega}$$

Ce théorème prouve que la taille de la composante connexe maximale est de l'ordre de $n^{2/3}$. Par ailleurs, les suites $n^{-2/3}|\mathcal{C}_{max}|$ et $n^{2/3}/|\mathcal{C}_{max}|$ sont tendues ce qui motive l'étude de leurs limites.

3.1 Énoncés des théorèmes

Nous avons déjà mentionné qu'on pouvait retrouver les tailles des composantes connexes à partir de la marche aléatoire en largeur. Aldous prouve dans [7] que la marche aléatoire renormalisée dans le cas $p(n) = \frac{1}{n} + \frac{t}{n^{4/3}}$ converge vers un mouvement brownien avec drift parabolique, défini pour tout $t \geq 0$ par :

$$W^t(s) = W(s) + ts - \frac{s^2}{2}$$

où W est le mouvement brownien standard. On définit aussi le mouvement brownien réfléchi :

$$B^t(s) = W^t(s) - \min_{0 \leq s' \leq s} W^t(s').$$

Par ailleurs, on souhaite comprendre le comportement des surplus du graphe d'Erdős–Rényi. On remarque qu'on peut compter le nombre de surplus lors de l'exploration. On associe simplement à la marche un processus de comptage $(N_n^t(s))_{0 \leq s \leq n}$ en ajoutant $+1$ pour chaque arête entre le sommet v_i en cours d'exploration et un sommet appartenant à $\mathcal{O}_i \cup \mathcal{A}_i$. Aldous prouve que $(N_n^t(s))$ converge vers un processus de comptage défini par la propriété :

$$N^t(s) - \int_0^s B^t(u) du \text{ est une martingale .}$$

Le processus N^t est bien défini grâce au théorème de décomposition de Doob-Meyer. De manière informelle, conditionnellement au passé, on a un saut entre $[s, s + ds]$ avec une probabilité $B^t(s)ds$:

$$\mathbb{P} \left(\text{saut entre } [s, s + ds] \mid B^t(u), u \leq s \right) = B^t(s)ds.$$

Les limites sont résumées par le théorème suivant qu'on redémontre dans la sous-section suivante :

Théorème 3.2 (Aldous [7]). Soit $t \in \mathbf{R}$, on note Y_n^t la marche aléatoire en largeur du graphe $ER\left(n, \frac{1}{n} + \frac{t}{n^{4/3}}\right)$. On obtient la convergence en loi

$$\left[(n^{-1/3}Y_n^t(\lfloor n^{2/3}s \rfloor), N_n^t(\lfloor n^{2/3}s \rfloor)), s \geq 0\right] \xrightarrow{d} ((W^t(s), N^t(s)), s \geq 0)$$

quand $n \rightarrow \infty$ pour la topologie de la convergence uniforme sur tout intervalle de temps compact.

On rappelle la définition d'une excursion.

Définition 3.1. Une excursion γ du processus brownien réfléchi est un intervalle $[l(\gamma), r(\gamma)]$ tel que

$$B^t(l(\gamma)) = B^t(r(\gamma)) = 0 \quad , \quad B^t(s) > 0 \quad \text{pour } s \in (l(\gamma), r(\gamma)).$$

On note $|\gamma| = r(\gamma) - l(\gamma)$ et $\mu(\gamma)$ le nombre de points du processus ponctuel associé à N^t dans l'intervalle $[l(\gamma), r(\gamma)]$.

Au vu des arguments heuristiques, il paraît naturel de s'attendre à ce que les tailles des composantes connexes convergent vers la longueur des excursions du brownien réfléchi. Formellement, Aldous démontre le théorème suivant :

Corollaire 3.1 (Aldous [7]). On note :

- $(C_n^t(j))_{j \in \mathbb{N}}$ la taille des composantes de $ER\left(n, \frac{1}{n} + \frac{t}{n^{4/3}}\right)$ dans l'ordre décroissant
- $(S_n^t(j))_{j \in \mathbb{N}}$ le nombre de surplus correspondant.
- $(|\gamma_j|)_{j \in \mathbb{N}}$ les tailles des excursions du brownien B^t rangées dans l'ordre décroissant.
- Pour tout j , μ_j le nombre de point dans l'intervalle $[l(\gamma_j), r(\gamma_j)]$ du processus ponctuel associé.

On obtient la convergence suivante :

$$(n^{-2/3}C_n^t(j), S_n^t(j)) \xrightarrow{d} (|\gamma_j|, \mu_j),$$

au sens ℓ^2 par rapport à la première composante et au sens de la topologie produit par rapport à la seconde.

3.2 Preuve du théorème 3.2

On se propose de démontrer le théorème 3.2. On commence par définir une nouvelle marche aléatoire. Pour éviter que les sauts soient concentrés aux temps entiers, on les éparpille en choisissant uniformément un temps sur les intervalles entiers. Plus formellement, on définit pour tout $u \in [0, 1]$:

$$Z_n^t(i-1+u) = Z_n^t(i-1) - u + \sum_{j=1}^{ch(i)} \mathbf{1}_{U_{ij} \leq u}$$

où $ch(i)$ est le nombre d'enfant du sommet i et $(U_{ij})_{1 \leq i \leq n, 1 \leq j \leq ch(i)}$ des variables uniformes sur $(0, 1)$ indépendantes. Étant donné que Z_n^t est un processus markovien

de saut, on peut le décomposer en une somme d'une martingale M_n^t et un processus A_n^t continu et à variation bornée. Intuitivement, le processus sera l'accumulation des accroissements infinitésimaux $\mathbb{E}(Z_n^t(s+ds) - Z_n^t(s) \mid \sigma(Z_n^t(u), u \leq s))$. On obtient alors :

$$A_n^t(u) = \int_0^u (a_n(s) - \underbrace{1}_{\text{du au drift}}) ds$$

où

$$a_n(s) ds = \mathbb{P}(\text{une nouvelle arête est découverte entre } s \text{ et } s+ds \mid \sigma(Z_n^t(u), u \leq s))$$

Pour $v \in [n]$, on définit $\zeta_n^{-1}(v)$ l'indice de la composante connexe à laquelle appartient le sommet v rangé dans l'ordre de leur découverte. Le sommet $v(\lfloor s \rfloor)$ ne peut pas être relié à lui-même. Une arête avec un sommet déjà vu ne compte pas dans la marche aléatoire. Le nombre d'arêtes dont $v(\lfloor s \rfloor)$ est l'un des côtés et qui compte dans la marche aléatoire est égale $n - \nu_n(s)$ avec

$$\nu_n(s) = \lfloor s \rfloor - 1 + |\mathcal{O}_{\lfloor s \rfloor - 1}| + \zeta_n^{-1}(\lfloor s \rfloor) - \zeta_n^{-1}(\lfloor s \rfloor - 1) + \sum_j \mathbf{1}_{U_{ij} \leq (s - \lfloor s \rfloor + 1)} = s + \zeta_n^{-1}(\lfloor s \rfloor) - Z_n^t(s).$$

Par construction du graphe d'Erdős-Rényi, une arête éligible appartient au graphe avec une probabilité $p(n)$. De plus, sachant qu'une arête existe, elle sera découverte en un temps uniforme entre 0 et 1. On obtient alors :

$$a_n = (n - \nu_n(s)) \frac{p(n)}{1 - (s - \lfloor s \rfloor)p(n)}.$$

Grâce au théorème de décomposition de Doob-Meyer, on peut aussi décomposer $(M_n^t)^2$ en une somme d'une martingale et d'un processus B_n^t croissant et continu. Grâce au même type de raisonnement, on obtient $B_n^t(u) = \int_0^u (a_n(s)) ds$. L'objectif est d'utiliser le théorème 7.1.4b de [13]. Si l'on montre que pour tout s_0 :

$$n^{-2/3} B_n^t(n^{2/3} s_0) \xrightarrow{p} s_0, \quad (4)$$

$$n^{-2/3} E \sup_{s \leq n^{2/3} s_0} |M_n^t(s) - M_n^t(s^-)|^2 \rightarrow 0. \quad (5)$$

alors on pourra conclure que :

$$n^{-1/3} M_n^t(n^{2/3} s) \xrightarrow{d} W^t$$

Par ailleurs, si on montre que :

$$n^{-1/3} \sup_{s \leq n^{2/3} s_0} |A_n^t(s) + n^{-1} s^2 / 2 - n^{-1/3} s t| \xrightarrow{p} 0, \quad (6)$$

alors on obtient :

$$n^{-1/3} Z_n^t(n^{2/3} s) \xrightarrow{d} W^t + t s - \frac{1}{2} s^2$$

Étant donné que les sauts de M_n^t sont limités à 1, la convergence (5) est trivialement vérifiée. Comme $A_n^t(u) = B_n^t(u) - u$, (4) se ramène à prouver :

$$n^{-2/3} A_n^t(n^{2/3} s_0) \rightarrow_p 0$$

ce qui est une conséquence de (6). Au vu de l'expression de A_n^t , on se ramène à démontrer pour tout s_0 ,

$$n^{-2/3} \sup_{s \leq n^{2/3} s_0} |Z_n^t(s)| \rightarrow_p 0. \quad (7)$$

On pose $T_n = \min\{s \leq s_0 n^{2/3}; |Z_n^t(s)| > K n^{1/3}\}$. L'inégalité de Markov donne :

$$\mathbb{P} \left(\sup_{s \leq n^{2/3} s_0} |Z_n^t(s)| > K n^{1/3} \right) = \mathbb{P} (|Z_n^t(T_n)| > K n^{1/3}) \leq \frac{\mathbb{E}(|Z_n^t(T_n)|)}{K n^{1/3}}.$$

Il reste à contrôler $\mathbb{E}(|Z_n^t(T_n)|)$:

$$\begin{aligned} \mathbb{E}(|Z_n^t(T_n)|) &\leq \mathbb{E}(|M_n^t(T_n)|) + \mathbb{E}(|A_n^t(T_n)|) \\ &\leq \mathbb{E}(|M_n^t(T_n)|^2)^{1/2} + \mathbb{E}(|A_n^t(T_n)|) \\ &\leq \mathbb{E}(|B_n^t(T_n)|)^{1/2} + \mathbb{E}(|A_n^t(T_n)|) \\ &\leq \int_0^{s_0 n^{2/3}} \frac{np(n)}{1 - (s - [s])p(n)} ds + \alpha n^{1/3} \\ &\quad + \int_0^{s_0 n^{2/3}} \left| \frac{s}{n} - \frac{t}{n^{1/3}} + \frac{st}{n^{4/3}} \right| ds + (s_0 n^{2/3}) 4 \frac{K n^{1/3}}{n} \\ &\leq \alpha n^{1/3} + 4s_0 K. \end{aligned}$$

On se donne $\epsilon > 0$ et on prend $K = \epsilon n^{1/3}$

$$\mathbb{P} \left(n^{-2/3} \sup_{s \leq n^{2/3} s_0} |Z_n^t(s)| > \epsilon \right) \leq \frac{\alpha}{\epsilon n^{1/3}} + \frac{4s_0}{n^{1/3}} \rightarrow 0$$

ce qui conclut la première partie de cette preuve.

On s'intéresse maintenant au processus N_n^t qui est un processus de comptage. D'après le théorème 12.6 de [9], il suffit de montrer la convergence des lois finidimensionnelles. Comme les accroissements sont indépendants, il suffit de montrer la convergence des intensités qui est une conséquence de la convergence de la marche aléatoire. En effet, conditionnellement à la donnée Z_n^t , N_n^t est le processus de comptage associé au processus ponctuel binomial de paramètre p entre $\min_{u \leq *}(Z_n^t(u))$ et $Z_n^t(*)$. Après renormalisation, le processus ponctuel binomial converge vers le processus ponctuel de Poisson d'intensité la mesure de Lebesgue au sens de la convergence faible des distributions de probabilité. On conclut que pour tout s , $N_n^t(s)$ converge vers $N^t(s)$. \square

3.3 Preuve du corollaire 3.1

On rappelle qu'on a montré la convergence des marches aléatoires et qu'on cherche à déduire la convergence des longueurs des excursions. On expliquera dans ses grandes lignes la preuve du corollaire 3.1 démontré par Aldous dans [7]. Comme on n'utilisera pas ces lemmes techniques par la suite, on va les énoncer sans les prouver.

Le premier permet de passer de la convergence de la marche X à la convergence vague des processus ponctuels. Les processus ponctuels sont interprétés comme des mesures ce qui justifie l'utilisation de la convergence vague :

$$\Xi^{(n)} = \{(n^{-2/3}\gamma_n(i), n^{-2/3}C_n^t(i)) : i \geq 1\} \rightarrow \Xi^{(\infty)} = \{(l(\gamma), |\gamma|), \gamma \text{ une excursion } B^t\}$$

où $v(\gamma_n(i))$ est le dernier sommet de la $(i - 1)$ composante découverte.

Lemme 3.1. *Soit $f : [0, \infty) \rightarrow \mathbb{R}$ est une fonction continue. On note \mathcal{E} est un ensemble d'intervalles non vide tel que :*

$$f(r) = f(l) = \min_{s \leq l} f(s) \\ f(s) > f(l)$$

Supposons que pour $e_1, e_2 \in \mathcal{E}$ tels que $l_1 < l_2$, on a $f(l_1) > f(l_2)$. On suppose aussi $\overline{\bigcup_{e \in \mathcal{E}} (l, r)}$ est de mesure nulle. On se donne f_n qui converge uniformément sur tout compacts ainsi que $t_{n,i}$ tel que :

- $0 = t_{n,1} < t_{n,2} < t_{n,3} \cdots$ et $\lim_{i \rightarrow \infty} t_{n,i} = \infty$;
- $f_n(t_{n,i}) = \min_{u \leq t_{n,i}} f_n(u)$,
- $\max_{i: t_{n,i} \leq s_0} (f_n(t_{n,i}) - f_n(t_{n,i+1})) \rightarrow 0$ quand $n \rightarrow \infty$, pour tout $s_0 < \infty$.

Alors, $\Xi^{(n)} = \{(t_{n,i}, t_{n,i+1} - t_{n,i}) ; i \geq 1\}$ converge vaguement quand n tend vers l'infini vers $\Xi = \{(l, r - l) : (l, r) \in \mathcal{E}\}$

Le deuxième lemme consiste à passer de la convergence vague à la convergence dans ℓ^2 de la suite des points rangés dans l'ordre décroissant. On note π la projection sur la seconde coordonnée et la fonction $\text{ord} : l_+^2(\Gamma) \rightarrow l_{\searrow}^2$ qui ordonne les termes de la suite dans l'ordre décroissant.

Lemme 3.2. *Soit $Y^{(n)} \in l_+^2(\Gamma^n)$ pour tout $1 < n \leq \infty$ et $\Xi^{(n)}$ le processus ponctuel biaisé par la taille associé. Supposons $\Xi^{(n)} \rightarrow_d \Xi^{(\infty)}$, tel que $\Xi^{(\infty)}$ est un processus ponctuel*

$$\text{Si } (s, y) \in \Xi \text{ alors } \sum_{\{y' : (s', y') \in \Xi, s' < s\}} y' = s ,$$

$$\max \{y : (s, y) \in \Xi \text{ avec } s > s_0\} \rightarrow_p 0 \text{ lorsque } s_0 \rightarrow \infty$$

$$\sup \{s : (s, y) \in \Xi^{(\infty)} \text{ pour un certain } y\} = \infty \text{ a.s.}$$

alors $Y^{(\infty)} = \text{ord } \pi(\Xi^{(\infty)})$ est dans l_{\searrow}^2 , et $\text{ord } Y^{(n)} \rightarrow_d \text{ord } Y^{(\infty)}$.

Les hypothèses du lemme 3.1 sont vérifiées par le mouvement brownien et le restent pour W^t par absolue continuité. On obtient grâce au théorème 3.2 et le théorème de représentation de Skorokhod, la convergence presque sûre de $\Xi^{(n)}$ vers $\Xi^{(\infty)}$ au sens de la convergence vague. Intuitivement, on s'attend à découvrir les grandes composantes très vite. En effet, $\Xi^{(n)}$ est le processus ponctuel biaisé par la taille associé à $Y^{(n)}$ qui sont les tailles des composantes dans l'ordre de leur apparition. La vérification des hypothèses est très technique donc on choisit de ne pas la détailler. Le lemme 3.2 s'applique et on conclut la convergence au sens ℓ_2 . \square

4 Étude de la structure du graphe dans la fenêtre critique

On a étudié le comportement asymptotique des tailles et des surplus des composantes connexes. Il est naturel d'étendre l'étude à la forme géométrique du graphe. Il reste à donner du sens à la limite d'un graphe ce qui fait l'objet du prochain paragraphe.

4.1 Topologie de Gromov-Hausdorff

Il est naturel de voir les graphes comme des espaces métriques compacts. Une notion intuitive de distance entre deux graphes G_1 et G_2 serait de les plonger dans un espace métrique (\mathcal{M}, d) et d'évaluer la distance au sens de Hausdorff :

$$d_H(K, K') = \inf \{ \epsilon > 0 : K \subseteq F_\epsilon(K') \text{ and } K' \subseteq F_\epsilon(K) \}.$$

où $F_\epsilon(K) = \{x \in \mathcal{M} : d(x, K) \leq \epsilon\}$ est le ϵ voisinage de K dans \mathcal{M} .

Plus formellement, on définit la distance de Gromov-Hausdorff entre deux espaces métriques compacts X et X' ayant des points distingués ρ et ρ' par :

$$d_{GH}(X, X') = \inf \{ d_H(\phi(X), \phi'(X')) \vee d(\phi(\rho), \phi(\rho')) \},$$

en prenant l'infimum sur tous les espaces métriques (\mathcal{M}, d) et tous les plongements isométriques ϕ (resp. ϕ') de X (resp. X') dans \mathcal{M} . De part son expression, cette définition est difficile à manipuler dans les démonstrations. On donne une formulation alternative qui se base sur la notion de correspondance.

Définition 4.1. *Une correspondance entre X et Y est la donnée d'un sous-ensemble $R \subset X \times Y$ tel que pour tout $x \in X, y \in Y$, il existe $x' \in X, y' \in Y$ tel que $(x, y') \in R$ et $(x', y) \in R$*

Autrement dit, on généralise la notion d'application surjective en autorisant un point à avoir plusieurs images. On a en quelque sorte tordu X pour le transformer en Y et on mesure la déformation par :

$$\text{dis}(\mathcal{R}) = \sup \{ |d_X(x, x') - d_Y(y, y')| : (x, y), (x', y') \in \mathcal{R} \}.$$

On mesure la proximité des deux graphes en choisissant la correspondance qui tord le moins. Formellement, on obtient l'équivalence des deux définitions (théorème 7.3.25 [11]) :

$$d_{\text{GH}}(X, Y) = \frac{1}{2} \inf_{\mathcal{R} \in \mathcal{C}(X, Y)} \text{dis}(\mathcal{R}).$$

4.2 Arbres réels et leurs constructions

Afin de pouvoir prouver un théorème de convergence, il faut trouver la bonne notion de limite potentielle. On commence d'abord à s'intéresser aux arbres et on définit ainsi les arbres réels.

Définition 4.2. *Un arbre réel enraciné (\mathcal{T}, ρ, d) est un espace métrique compact tel que pour tous sommets $x, y \in \mathcal{T}$:*

- *Il existe une unique géodésique de x vers y c'est à dire il existe une unique isométrie $f_{x,y} : [0, d(x, y)] \mapsto \mathcal{T}$ tel que $f_{x,y}(0) = x$ et $f_{x,y}(d(x, y)) = y$. L'image de cette application sera noté $[x, y]$*
- *Si $q : [0, 1] \mapsto \mathcal{T}$ est une application injective et continue tel que $q(0) = x$ et $q(1) = y$ alors $q([0, 1]) = [x, y]$.*

Comme on a pu voir, le parcours en profondeur encode un arbre discret à travers la marche en profondeur associée. L'idée fondamentale derrière les arbres réels est de les coder avec la fonction hauteur. Pour mesurer la distance entre deux sommets x, y dans un arbre discret, on cherche l'ancêtre commun $x \wedge y$ et on somme la distance relative à $x \wedge y$. La définition de d_h se base sur cette idée. Plus précisément, si on se donne $h : [0, \infty) \mapsto [0, \infty)$ une fonction continue à support compact telle que $h(0) = 0$, on a une pseudo-distance :

$$d_h(x, y) = h(x) + h(y) - 2 \inf_{x \wedge y \leq t \leq x \vee y} h(t), \quad x, y \in [0, \infty).$$

On définit l'arbre $\mathcal{T}_h = [0, \infty) / \sim_h$ où $x \sim_h y$ si et seulement si $d_h(x, y) = 0$. On notera τ_h l'application de projection associée. Par ailleurs, on obtient aussi une mesure qui est le tiré en arrière par la projection τ_h de la mesure de Lebesgue.

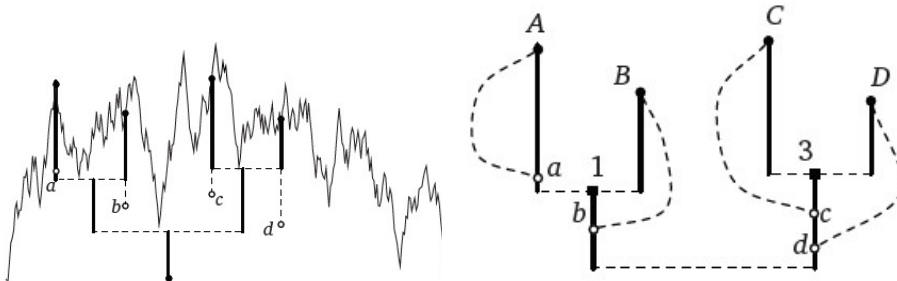


FIGURE 4 – Une réalisation de l'excursion brownienne et un sous-graphe engendré par la racine et les feuilles A,B,C,D. La figure est extraite de l'article [1].

Un arbre réel particulièrement étudié est l'arbre aléatoire continu brownien d'Aldous (Brownian continuous real tree) défini par $[0, \infty)/ \sim_h$ où $h(t) = 2e(t)$ l'excursion brownienne. Il a été défini et étudié par David Aldous dans ([4], [5], [6]) qui a prouvé la limite d'échelle suivante :

Théorème 4.1 (Aldous [6]). *Soit T_n l'arbre de Cayley et \mathcal{T} est un CRT alors on a la convergence en loi suivante au sens Gromov-Hausdorff :*

$$n^{-1/2}T_n \rightarrow \mathcal{T}.$$

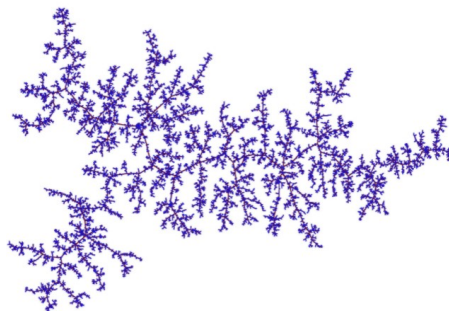


FIGURE 5 – Une réalisation de l'arbre de Cayley à 27179 sommets (approximant l'arbre brownien d'Aldous). La figure est extraite du site d'Igor Kotchemski.

Ce résultat reste naturel au vu du théorème de Marckert et Mokaddem dans [18] et de la construction du CRT. Ils démontrent la convergence au sens de Skorokhod de la marche aléatoire en profondeur ainsi que la hauteur associée à l'arbre planaire de Galton–Watson conditionné. Si on considère un arbre Galton–Watson poissonnien qu'on numérote ses sommets et on oublie l'ordre et la racine, on obtient un arbre de Cayley. Grâce à cette construction, la marche aléatoire associée à Galton–Watson et celle associée ont la même loi. Le théorème suivant est une conséquence du théorème de Marckert et Mokaddem dans [18] :

Théorème 4.2 (Marckert et Mokaddem [18]). *Soit X^n la marche en profondeur associée à T_n l'arbre de Cayley et H^n la fonction hauteur associée, on obtient la convergence en loi au sens de Skorokhod suivante :*

$$\left(\frac{1}{\sqrt{n}} (X^n(\lfloor nt \rfloor), 0 \leq t \leq 1), \frac{1}{\sqrt{n}} (H^n(\lfloor nt \rfloor), 0 \leq t \leq 1) \right) \xrightarrow{d} (e, 2e). \quad (8)$$

où e est l'excursion brownienne.

L'objet de notre étude est un graphe avec un nombre d'arêtes en surplus qui converge vers une constante. On étend l'idée d'encoder les arêtes en surplus à travers des marques en dessous de la fonction hauteur. La notion d'arête disparaît à l'infini et sa réminiscence s'exprime par l'identification de certains points.

Soit $\mathcal{Q} \subset \mathbb{R}^+ \times \mathbb{R}^+$, h une fonction continue à support compact telle que $h(0) = 0$, on définit :

$$\mathcal{Q} \cap h = \{(x, y) \in \mathcal{Q} : 0 \leq y < h(x)\}.$$

Pour tout $(\xi_x, \xi_y) \in \mathcal{Q} \cap (h/2)$, on connecte les deux points $\tau(\xi_x)$ et $\tau(\inf\{u \geq \xi_x; h(u) = 2\xi_y\})$. Le graphe ainsi construit sera noté $g(h, \mathcal{Q})$.

On munit ce graphe de la distance naturelle définie par l'expression suivante :

$$\min \left(d_h(x, x'), \inf_{\substack{i_1, \dots, i_r, \\ \epsilon_1, \dots, \epsilon_r}} \left(d_h(x, x_{i_1}^{\epsilon_1}) + \sum_{j=1}^{r-1} d_h(x_{i_j}^{\epsilon_j+1}, x_{i_{j+1}}^{\epsilon_{j+1}}) + d_h(x_{i_r}^{\epsilon_r+1}, x') \right) \right) \quad (9)$$

où on a noté (x_i^0, x_i^1) les points qu'on a identifiés. Il faut s'imaginer qu'on peut tester tous les raccourcis possibles et on prend le chemin optimal après le passage au quotient.

4.3 La limite continue du graphe d'Erdős–Rényi

Dans cette section, on se propose d'expliquer la convergence des graphes d'Erdős–Rényi étudiée par Addario-Berry, Broutin et Goldschmidt dans [2]. On commence par s'intéresser à chaque composante connexe. On note $G(m, p)$ la loi d'un graphe d'Erdős–Rényi sachant qu'il est connexe. On rappelle que $a(T)$ est le nombre d'arêtes admissibles de l'arbre T (voir définition 2.1). On construit $\tilde{G}(m, p)$ à partir de l'arbre \tilde{T}_m^p suivant la loi $\mathbb{P}(\tilde{T}_m^p = T) \propto (1-p)^{-a(T)}$ auquel on ajoute des arêtes indépendamment et avec une probabilité p parmi les arêtes admissibles.

Propriété 4.1. *Pour tout m, p , le graphe $G(m, p)$ a la même loi que $\tilde{G}(m, p)$.*

Preuve : On se donne un graphe G à m sommets. Par définition du graphe aléatoire $\tilde{G}(m, p)$, on a :

$$\begin{aligned} \mathbb{P}(\tilde{G}(m, p) = G) &\propto (1-p)^{-a(\mathbf{oDFS}(G))} p^{s(G)} (1-p)^{a(\mathbf{oDFS}(G)) - s(G)} \\ &\propto p^{s(G)} (1-p)^{-s(G)} \\ &\propto p^{m-1+s(G)} (1-p)^{\binom{m}{2} - m + 1 - s(G)} \propto \mathbb{P}(G(m, p) = G) \end{aligned}$$

ce qui suffit pour conclure. □

On rappelle qu'on note $(C_n^t(i))_i$ les tailles des composantes connexes du graphe d'Erdős–Rényi rangées dans l'ordre décroissant. Par ailleurs, on notera $\mathcal{C}_n^t(i)$ la composante connexe qui correspond à la taille $C_n^t(i)$. La propriété 4.1 se réinterprète comme suit : “sachant (C_n^t) , les composantes (\mathcal{C}_n^t) ont la même loi que $(\tilde{G}(C_n^t(i), p))_i$ ”. Comme le corollaire 3.1 nous donne le comportement limite de $(C_n^t(i))_i$, on peut se restreindre à l'étude de $\tilde{G}(m, p)$. L'essentiel de la preuve consiste à contrôler le comportement asymptotique de la marche aléatoire en profondeur et du processus ponctuel associés à $\tilde{G}(m, p)$.

La marche aléatoire tiltée : La première étape consiste à étudier la marche aléatoire associée au graphe $(\tilde{G}(m, p))$ lorsque $mp^{2/3} \rightarrow \sigma$ ce qui est donné par le théorème suivant :

Théorème 4.3 (Addario-Berry et al. [2]). *On suppose que $mp^{2/3} \rightarrow \sigma$ et on note \tilde{X}_m la marche aléatoire en profondeur associée à $\tilde{G}(m, p)$. On note $e^{(\sigma)}$ l'excursion brownienne de longueur σ . Alors, on a la convergence suivante dans $\mathbb{D}([0, \sigma], \mathbb{R}^+)^3$*

$$\left((m/\sigma)^{-1/2} \tilde{X}_m(\lfloor (m/\sigma)t \rfloor), 0 \leq t \leq \sigma \right) \xrightarrow{d} (\tilde{e}^{(\sigma)}(t), 0 \leq t \leq \sigma). \quad (10)$$

où \tilde{e} est définie par :

$$\mathbb{P}(\tilde{e}^{(\sigma)} \in \mathcal{B}) = \frac{\mathbb{E} \left[\mathbf{1}_{\{e^{(\sigma)} \in \mathcal{B}\}} \exp \left(\int_0^\sigma e^{(\sigma)}(t) dt \right) \right]}{\mathbb{E} \left[\exp \left(\int_0^\sigma e^{(\sigma)}(t) dt \right) \right]} \quad (11)$$

pour tout borélien \mathcal{B} .

Remarque. *Comme la limite est continue, on a l'équivalence entre la convergence au sens de Skorokhod et la convergence uniforme.*

Preuve : Pour simplifier les calculs, on note e l'excursion brownienne de taille 1. On rappelle que :

$$(e^{(\sigma)}(t), 0 \leq t \leq \sigma) \stackrel{d}{=} (\sqrt{\sigma} \cdot e(t/\sigma), 0 \leq t \leq \sigma).$$

Quitte à renormaliser, on se ramène au cas $\sigma = 1$.

On note X_m la marche aléatoire en profondeur associée à T_m l'arbre de Cayley à m sommets. L'étape fondamentale est de remarquer que pour tout $f : \mathbb{D}([0, 1], \mathbb{R}^+) \rightarrow \mathbb{R}^+$ continue borné :

$$\mathbb{E} \left[f \left(m^{-1/2} \tilde{X}_m(\lfloor mt \rfloor), 0 \leq t \leq 1 \right) \right] = \frac{\mathbb{E} \left[f(\bar{X}_m) (1-p)^{-m^{3/2}} \int_0^1 \bar{X}_m(t) dt \right]}{\mathbb{E} \left[(1-p)^{-m^{3/2}} \int_0^1 \bar{X}_m(t) dt \right]} \quad (12)$$

où $\bar{X}_m(t) = m^{-1/2} X_m(\lfloor mt \rfloor)$. D'après le théorème 4.2 et le théorème de Skorokhod, on se place dans un espace probabilisé tel que :

$$(\bar{X}_m(t), 0 \leq t \leq 1) \xrightarrow{p.s.} (e(t), 0 \leq t \leq 1).$$

Comme la convergence a lieu dans $\mathbb{D}([0, \sigma], \mathbb{R}^+)$ et que $h \in \mathbb{D}([0, \sigma], \mathbb{R}^+) \rightarrow \int_0^1 h(t) dt$ est continue, on a aussi :

$$(1-p)^{-m^{3/2}} \int_0^1 \bar{X}_m(t) dt \xrightarrow{p.s.} \exp \left(\int_0^1 e(t) dt \right).$$

3. L'espace des fonctions càdlàg à valeurs dans \mathbb{R}^+ muni de la topologie de Skorokhod.

Il paraît naturel de s'attendre à la convergence de $(m^{-1/2}\tilde{X}_m)$ vers \tilde{e} . Plus précisément, il nous manque un argument d'uniforme intégrabilité pour échanger la limite et l'espérance dans (12). Montrons que $(1-p)^{-a(T_m)}$ est borné dans L^2 . Pour m suffisamment grand, on a $p \leq 2m^{-2/3} \leq 1/2$ et

$$\begin{aligned} \mathbb{E}((1-p)^{-2a(T_m)}) &\leq \mathbb{E}(e^{4pa(T_m)}) \\ &\leq \int_0^{m^2} 4pe^{4px} \mathbb{P}(a(T_m) \geq x) dx \\ &\leq \int_0^{m^2} e^{8m^{-3/2}x} \mathbb{P}(\|X_m\| \geq \sqrt{m}m^{-3/2}x) dx. \end{aligned}$$

On remarque par ailleurs que le nombre de surplus est majoré par $m\|X_m\|$ donc

$$\mathbb{P}(a(T_m) \geq x) \leq \mathbb{P}(m\|X_m\| \geq x)$$

La preuve du théorème se ramène à majorer $\mathbb{P}(m\|X_m\| \geq x)$ ce qui est donné par le lemme ci-dessous qui nous permet de conclure. \square

Lemme 4.1. *Il existe $C \geq 0$ et $\alpha > 0$ tel que pour tout $m \in \mathbb{N}$ et $x \geq 0$:*

$$\mathbb{P}(\|X_m\| \geq x\sqrt{m}) \leq Ce^{-\alpha x^2}$$

La preuve du lemme est donnée dans l'annexe.

Les marques : On rappelle que pour tout arbre T et $\mathcal{Q} \subset \mathbb{Z} \times \mathbb{Z}$, le graphe $G(T, \mathcal{Q})$ construit par l'ajout des arêtes v_i et $v_{k(i,j)}$ pour $(i, j) \in \{(i', j') \in \mathcal{Q} : 0 \leq i'; 0 \leq j' < X(i')\}$ à l'arbre T .

Par construction de $\tilde{G}(m, p)$, on a $\tilde{G}(m, p) \stackrel{d}{=} G(\tilde{X}(m, p), \mathcal{Q}^p)$ où \mathcal{Q}^p est un processus ponctuel binomial d'intensité p indépendant de $\tilde{X}(m, p)$. Comme on suppose dans cette section $mp^{2/3} \rightarrow \sigma$, on s'attend à voir apparaître un processus ponctuel de Poisson.

Lemme 4.2 (Addario-Berry et al. [2]). *On se donne $p = p(m)$ tel que $mp^{2/3} \rightarrow \sigma$. On tire aléatoirement un arbre tilté \tilde{T}_m^p et on note \tilde{X}_m la marche en profondeur associée. On considère maintenant un processus ponctuel binomial d'intensité p qu'on normalise*

$$\mathcal{P}_m = \{((m/\sigma)^{-1}i, (m/\sigma)^{-1/2}j) : (i, j) \in \mathcal{Q}^p\}.$$

On obtient la convergence suivante au sens de Skorokhod pour la première coordonnée et au sens de Hausdorff pour la seconde coordonnée :

$$\left((m/\sigma)^{-1/2}\tilde{X}_m(\lfloor (m/\sigma)\cdot \rfloor), \mathcal{P}_m \cap ((m/\sigma)^{-1/2}\tilde{X}_m(\lfloor (m/\sigma)\cdot \rfloor)) \right) \xrightarrow{d} (\tilde{e}^{(\sigma)}, \mathcal{P} \cap \tilde{e}^{(\sigma)}) \quad (13)$$

lorsque $n \rightarrow \infty$, où \mathcal{P} est un processus ponctuel de Poisson d'intensité la mesure de Lebesgue sur $\mathbb{R}^+ \times \mathbb{R}^+$ et indépendant $\tilde{e}^{(\sigma)}$.

On rappelle qu'on a construit le graphe $g(h, \mathcal{Q})$ en connectant les deux points $\tau(\xi_x)$ et $\tau(\inf\{u \geq \xi_x; h(u) = 2\xi_y\})$ dans l'arbre \mathcal{T}_h . Au vu des résultats 4.2 et 4.3, la limite naturelle de $n^{-1/3}\tilde{G}(m, p)$ est $g(2\tilde{e}^{(\sigma)}, \mathcal{P})$.

Théorème 4.4 (Addario-Berry et al. [2]). *Supposons $m(n)^{-2/3}n \rightarrow \sigma$ et $p(n)n \rightarrow 1$, on a :*

$$n^{-1/3}\tilde{G}(m, p) \xrightarrow{d} \mathcal{C}^{(\sigma)} = g(2\tilde{e}^{(\sigma)}, \mathcal{P}).$$

Preuve : Quitte à renormaliser, on peut supposer $\sigma = 1$. D'après (13) et le théorème de Skorokhod, on peut trouver un espace probabilisé tel qu'on a la convergence presque sûre suivante :

$$(m^{-1/2}\tilde{X}_m(\lfloor m \cdot \rfloor), \mathcal{P}_m \cap (m^{-1/2}\tilde{X}_m(\lfloor m \cdot \rfloor))) \xrightarrow{p.s} (\tilde{e}, \mathcal{P} \cap \tilde{e}). \quad (14)$$

On note pour simplifier $\bar{X}^m(t) = m^{-1/2}\tilde{X}_m(\lfloor m \cdot \rfloor)$ et $\bar{H}^m(t) = m^{-1/2}\tilde{H}^m(\lfloor m \cdot \rfloor)$. Un lemme technique (lemme 16 de [2]) dont la preuve est donnée dans l'article montre qu'il existe K et $m_0 \geq 0$ tel que pour tout $m \geq m_0$:

$$\mathbb{P}(\|\tilde{X}_m - \tilde{H}^m/2\| \geq m^{3/8}) \leq Km^{-1/16}. \quad (15)$$

On en déduit que sur ce même espace de probabilité, la fonction hauteur converge vers $2\tilde{e}$.

A ω fixé, on a $|\mathcal{P} \cap \tilde{e}| = s(\omega)$ est fini et comme $|\mathcal{P}_m \cap \bar{X}^m|$ est entier, la suite est constante à partir d'un certain rang $M(\omega)$. Étant donné que la convergence est au sens de Hausdorff, on peut numéroter les points de chaque ensemble de manière à avoir :

$$\max_{1 \leq \ell \leq s} \{|i_\ell^m - \xi_\ell^x| + |j_\ell^m - \xi_\ell^y|\} \rightarrow 0. \quad (16)$$

On note $k_\ell^m = m^{-1}k(mi_\ell^m, m^{1/2}j_\ell^m)$ et $u(\xi) = u(\xi^x, \xi^y) = \inf\{u \geq \xi^x : \tilde{e}(u) = \xi^y\}$. On obtient aussi la convergence suivante :

$$\max_{1 \leq \ell \leq s} |k_\ell^m - u(\xi_\ell)| \rightarrow 0. \quad (17)$$

Autrement dit, on retrouve le fait que le nombre de surplus converge vers une constante. On n'a pas juste conservé le nombre de surplus mais les positions des arêtes en surplus convergent "sans discontinuité".

D'abord, on remarque que le graphe G' défini en identifiant $v_{mi_\ell^m}$ et $v_{mk_\ell^m}$ est à une distance d'au plus s du graphe $\tilde{G}(m, p)$. En effet, il suffit de considérer la correspondance identité. On se ramène donc à l'étude de la convergence de $m^{-1/2}G'$.

On peut interpréter les convergences (16) et (17) comme une correspondance des points carrefours. La meilleure vitesse qu'on puisse espérer pour la convergence du graphe est limitée par ε_m défini par :

$$\varepsilon_m = \max \left\{ \max_{1 \leq \ell \leq s} |i_\ell^m - \xi_\ell^x|, \max_{1 \leq \ell \leq s} |k_\ell^m - u(\xi_\ell)| \right\}.$$

On construit une correspondance entre \tilde{T}_m et \tilde{T} en associant les points dont l'abscisse correspondante ne dépasse pas ϵ_m . Formellement, on définit :

$$(v_i, t) \in \mathcal{R}_m \iff \left| \frac{i}{m} - t \right| \leq \epsilon_m.$$

On remarque que :

$$d_{\text{GH}}(n^{-1/3}G'_m, \mathcal{C}) \leq \frac{s+1}{2} \text{dis}(\mathcal{R}_m).$$

En effet, pour tout $x, x' \in G'$, $y, y' \in \mathcal{C}$, si $d(x, x') > d(y, y')$, on considère le chemin optimal entre y et y' et le chemin entre x et x' en passant par les carrefours correspondants. Si le chemin entre y et y' passe par plus de s fois par un carrefour, il existe au moins un point par lequel on passe deux fois. Autrement dit, il n'est pas optimal. Donc, on passe par au plus de s points carrefours. Comme on a choisi un chemin entre x et x' qui passe par les points carrefours correspondants, on s'est ramené à calculer la distorsion entre les arbres DFS. On obtient l'inégalité annoncée. Il reste maintenant à montrer que $\text{dis}(\mathcal{R}_m) \rightarrow 0$.

On se donne $x, x' \in \tilde{T}_m^p$, $y, y' \in \tilde{T}$ et on note $I = |n^{-1/3}d_{\tilde{T}_m^p}(v_i, v_j) - d_{\tilde{T}}(\tau(t), \tau(u))|$, alors

$$\begin{aligned} I &\leq |n^{-1/3}d_{\tilde{T}_m^p}(v_i, v_i \wedge v_j) - d_{\tilde{T}}(\tau(t), \tau(t) \wedge \tau(u)) \\ &\quad + n^{-1/3}d_{\tilde{T}_m^p}(v_i \wedge v_j, v_j) - d_{\tilde{T}}(\tau(t) \wedge \tau(u), \tau(u))| \\ &\leq |n^{-1/3}m^{1/2} \left(m^{-1/2}\tilde{H}^m(i) + m^{-1/2}\tilde{H}^m(j) - 2 \min_{r \in [i, j] \cap \mathbb{Z}} m^{-1/2}\tilde{H}^m(r) \right) \\ &\quad - \left(2\tilde{e}(t) + 2\tilde{e}(u) - 4 \inf_{v \in [t, u]} \tilde{e}(v) \right)| + 2n^{-1/3} \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Le terme en $n^{-1/3}$ vient de l'inégalité :

$$\left| d_T(v_0, v_i \wedge v_j) - \min_{i \leq k \leq j} H(k) \right| \leq 1.$$

On conclut donc la convergence du graphe au sens de Gromov-Hausdorff. \square

Grâce au théorème précédent et au théorème de représentation de Skorohod, on déduit la convergence en loi au sens ℓ^2 par rapport à la première coordonnée et de la topologie produit par rapport à la deuxième coordonnée :

$$(n^{-2/3}C_n^t, n^{-1/3}\mathcal{C}_n^t) \xrightarrow{d} (\mathbf{C}, \mathcal{C}).$$

où $\mathbf{C} \stackrel{d}{=} (|\gamma_j|)_j$ et conditionnellement à la loi \mathbf{C} , $(\mathcal{C}(i))$ sont indépendants et $\mathcal{C}(i) \stackrel{d}{=} \mathcal{C}(\mathbf{C}(i))$.

Dans [2], les auteurs prouvent que la convergence par rapport à la seconde coordonnée a lieu en l^4 :

Théorème 4.5 (Addario-Berry et al. [2]). *En gardant les notations précédentes, on a :*

$$(n^{-2/3}C_n^t, n^{-1/3}\mathcal{C}_n^t) \xrightarrow{d} (\mathbf{C}, \mathcal{C}).$$

La première est au sens ℓ^2 et la convergence de la seconde coordonnée a lieu par rapport à la métrique suivante :

$$d(\mathbf{A}, \mathbf{B}) = \left(\sum_{i=1}^{\infty} d_{\text{GH}}(A_i, B_i)^4 \right)^{1/4} .$$

5 Justification heuristique de la conjecture

On rappelle que l’objectif du stage était l’étude du nombre de coupes nécessaires pour isoler un sommet dans le graphe d’Erdős–Rényi dans la fenêtre critique. Comme on a pu le voir, l’arbre de Cayley converge au sens de Gromov-Hausdorff vers un CRT alors que les graphes d’Erdős–Rényi convergent vers un graphe construit de manière très proche. On va “juste” recoller un nombre fini de sommets d’un CRT. Quand bien même que l’excursion n’est pas la même, on peut s’attendre à avoir le même comportement.

On peut aussi voir la conjecture comme une conséquence de la généralisation du théorème 1.1. En effet, on déduit directement de la propriété 4.1 le corollaire suivant :

Corollaire 5.1. *Une composante de $ER(n, p)$ conditionnée à avoir m sommets et un surplus s est un graphe connexe uniformément choisi parmi ceux ayant m sommets et $m + s - 1$ arêtes.*

Essentiellement, à partir du moment où l’on connaît la taille et le surplus qui sont étudiés dans [7], la composante connexe a une loi uniforme. En particulier, le paramètre t n’intervient plus. Le théorème 1.1 démontre la convergence pour des arbres de Cayley soit pour un surplus égale à 0. C’est en ceci qu’il s’agit d’une conséquence de la généralisation du théorème 1.1.

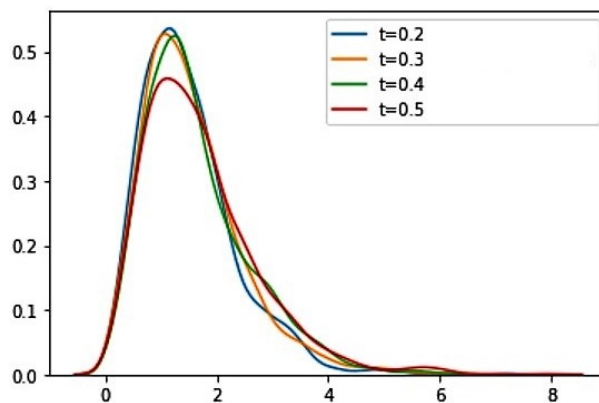


FIGURE 6 – Apparition de la densité de Rayleigh pour $n = 1000$

Simulation du processus : On a simulé des graphes d’Erdős–Rényi pour différentes valeurs de t en prenant des valeurs de n de plus en plus grandes. On voit effectivement apparaître la densité d’une loi de Rayleigh dont le paramètre varie avec t (figure 6).

5.1 Etude des moments :

Comme on connaît le comportement asymptotique de la taille et du surplus, on peut se restreindre à l'étude des graphes uniformes à n arêtes et s surplus. On commence par étudier les moments du temps de coupe en s'appuyant sur la méthode de Meir et Moon dans [21] qu'on commence par présenter. L'idée est d'utiliser les séries génératrices pour transformer une relation de récurrence en une EDO. On note :

- $\kappa_{m,s}$ le nombre d'étapes pour isoler la racine d'un graphe connexe à m sommets et s surplus tiré uniformément,
- $A_{m,s}$ le nombre de graphes étiquetés connexes à n sommets et s surplus,
- $W_k = \sum_l A_{l,k+1} \frac{x^l}{l!}$,
- $\mu(m, s) = \mathbb{E}(\kappa_{m,s})$ et $Y_s(x) = \sum \mu(m, s) \frac{x^m}{(m-1)!}$.

Dans un arbre, lorsqu'on coupe une arête, on le décompose en deux composantes connexes : l'une de taille k qui contient la racine et l'autre qu'on oublie. Pour un graphe à n sommets et s surplus, on peut le déconnecter en une composante connexe de taille k et i arêtes en surplus et une autre composante qu'on oublie. On peut aussi enlever une arête qui participe à un cycle. Dans ce cas, le nombre de surplus diminue de 1.

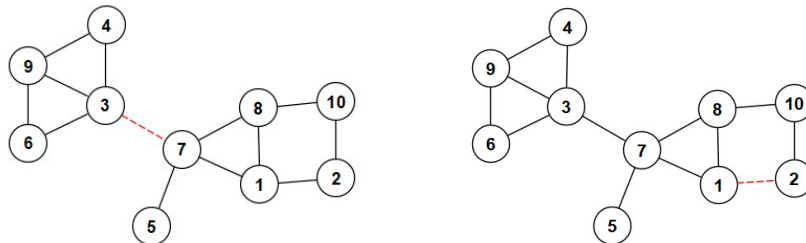


FIGURE 7 – Les deux cas de figure qui peuvent apparaître. Si on retire l'arête (3, 7), on déconnecte le graphe en deux composantes connexes. Si on retire l'arête (1, 2), le graphe reste connexe.

On obtient alors l'équation de récurrence suivante :

$$(m + s - 1)A_{m,s}\mathbb{P}(\kappa_{m,s} = x) = \sum_{k=1}^{m-1} \sum_{i=0}^s \binom{m-1}{k-1} k(m-k)A_{k,i}A_{m-k,s-i}\mathbb{P}(\kappa_{k,i} = x-1) + \left(\frac{m(m-1)}{2} - (m+s-2) \right) A_{m,s-1}\mathbb{P}(\kappa_{m,s-1} = x-1).$$

On en déduit :

$$(m + s - 1)A_{m,s}\mu(m, s) = \sum_{k=1}^{m-1} \sum_{i=0}^s \binom{m-1}{k-1} k(m-k)A_{k,i}A_{m-k,s-i}(\mu(k, i) + 1) + \left(\frac{m(m-1)}{2} - (m+s-2) \right) A_{m,s-1}(\mu(m, s-1) + 1).$$

On obtient finalement :

$$x\partial_x Y_s + (s-1)Y_s = x^2\partial_x^2 W_{s-1} + sW_{s-1} + 1/2x^2\partial_x^2 Y_{s-1} \\ - xY_{s-1} - (s-2)Y_{s-1} + \sum_{i=0}^s x^2(\partial_x Y_i)(\partial_x W_{s-i-1}).$$

On note $\theta = 1 - W_{-1}$. Si $s = 0$, on retrouve le résultat de Meir et Moon

$$Y_0 = -(1 - \theta) \log(\theta).$$

On identifie ainsi $\mu(m, 0)$ dont on peut étudier le comportement asymptotique. Si $s = 1$, on obtient :

$$2\frac{\theta-1}{\theta}\partial_\theta Y_1 = R(\theta) + \left(Y_0 + \theta + \frac{1}{\theta}\right) \frac{(1-\theta)^3}{\theta^2}$$

où R est une fraction rationnelle. Contrairement aux articles ([21],[26],[27]) qui emploient la même méthode, on obtient des termes en $\theta^2 \log(\theta)$ ce qui indique que le logarithme persistera et nous conduit à s'intéresser directement à la convergence de la variable aléatoire.

5.2 Lien avec les arbres branchants Markoviens

On reprend une preuve donnée par Bertoin dans [8]. On commence par expliquer la preuve dans le cadre de l'arbre de Cayley. L'idée est d'encoder le processus de coupe de tout l'arbre en un autre arbre qui contiendrait toutes les informations.

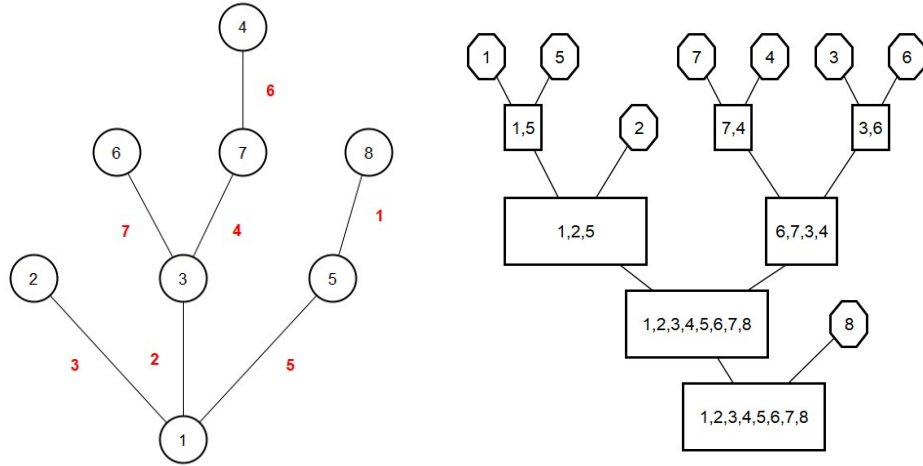


FIGURE 8 – Un arbre à 8 sommets et le processus de logging associé.

Logging process : On se donne un arbre T à n sommets et un ordre de coupe $\mathbf{E} = e_1, \dots, e_{n-1}$. L'idée est de voir chaque composante comme un bloc qui contient ces sommets. On commence par un bloc qui contient tous les sommets. A chaque

où \bar{q}_n est une mesure sur \mathcal{S}^\downarrow (l'ensemble des partitions de la masse 1 ordonné dans l'ordre décroissant) défini par

$$\bar{q}_n(f) = \sum_{\lambda \in \mathcal{P}_n} q_n(\lambda) f\left(\frac{\lambda}{n}\right).$$

Le théorème principal de l'article [15] est le suivant :

Théorème 5.1 (Haas et Miermont [15]). *Soit $q = (q_n(\lambda))_{\lambda \in \mathcal{P}_n}$ satisfaisant l'hypothèse 1, on se donne T_n ayant la distribution P_n^q qu'on voit comme un espace métrique compact. On a la convergence en loi au sens de Gromov-Hausdorff :*

$$\frac{1}{n^\gamma \ell(n)} T_n \xrightarrow[n \rightarrow \infty]{d} \mathcal{T}_{\gamma, \nu}. \quad (19)$$

Remarque. *En réalité, il démontre le théorème au sens de Gromov Hausdorff Prokhorov étant donné que l'arbre réel est naturellement muni d'une mesure quotient. Dans ce cadre, on munira les arbres discrets de la mesure uniforme sur les feuilles.*

Dans notre exemple, l'arbre construit après le logging est un arbre de branchement markovien. Plus exactement, Pavlov démontre dans [24] que la série de lois q_n correspondante est :

$$q_n(m) = \frac{\binom{n}{m} m^{m-2} (n-m)^{n-m-2} n(n-m)}{(n-1)n^{n-2}}. \quad (20)$$

On remarque que l'hypothèse 1 est vérifiée pour $\gamma = 1/2$, $\ell(n) = 1$ et

$$\nu(ds) = \frac{1}{\sqrt{2\pi s^3(1-s)^3}} ds.$$

En effet, on a

$$n^{3/2}(1-m/n)q_n(m) \sim \frac{1}{\sqrt{2\pi(m/n)^3(1-m/n)}}.$$

Grâce au théorème (5.1), on démontre la convergence de l'arbre de logging vers l'arbre brownien d'Aldous et on sait que la distance entre deux sommets suit la loi de Rayleigh. En réalité, ce qui nous intéresse lorsqu'on veut démontrer 1.2 est le lemme 28 dans [15] qui étudie la convergence de la distance d'une feuille à la racine. La démonstration de ce lemme se base essentiellement sur l'article [14] qui étudie les limites d'échelle des chaînes de Markov décroissantes.

5.3 Limite d'échelle des chaînes de Markov décroissantes

Dans l'article [14], Haas et Miermont considèrent une chaîne de Markov décroissante X à valeurs dans \mathbb{N} . L'hypothèse essentielle est de considérer qu'à l'état n , il est peu probable (probabilité de l'ordre $c_\epsilon n^{-\gamma}$) d'accomplir un saut de taille $(1-\epsilon)n$. Formellement, on se donne la matrice de transition $(p_{n,k} \ k \leq n)$ et on note :

$$p_n^*(dx) = \sum_{k=0}^n p_{n,k} \delta_{k/n}(dx). \quad (21)$$

On suppose :

Hypothèse 2. *Il existe une séquence a_n de la forme $a_n = n^\gamma \ell(n)$ où $\gamma > 0$ et ℓ une fonction $(0, \infty) \mapsto (0, \infty)$ à variation lente. Il existe aussi une mesure μ positive non nulle et finie sur $[0, 1]$ tel qu'on a la convergence faible suivante :*

$$a_n(1-x)p_n^*(dx) \xrightarrow[n \rightarrow \infty]{d} \mu(dx). \quad (22)$$

On définit :

$$B_n = \inf\{k \in \mathbb{N} : p_{X_n(k), X_n(k)} = 1\}.$$

Autrement dit, B_n est l'instant à partir duquel on devient stationnaire qui est presque sûrement fini. On commence par définir la loi limite. On note pour commencer :

$$\psi(\lambda) := \int_{[0,1]} \frac{1-x^\lambda}{1-x} \mu(dx)$$

On définit $(\xi_t, t \geq 0)$ un processus de Lévy croissant tel que :

$$\mathbb{E}[\exp(-\lambda \xi_t)] = \exp(-t\psi(\lambda)), \quad t, \lambda \geq 0.$$

On notera également :

$$\tau : t \in [0, \infty) \rightarrow \int_0^t \exp(-\gamma \xi_r) dr$$

et τ^{-1} son inverse en posant $\tau^{-1}(t) = \infty$ pour $t > I := \int_0^\infty \exp(-\gamma \xi_r) dr$.

Théorème 5.2 (Haas et Miermont). *Pour tout t et n , on définit $Y_n(t)$ par :*

$$Y_n(t) = \frac{X_n(\lfloor a_n t \rfloor)}{n}.$$

Alors, sous l'hypothèse 2, on a la convergence en loi au sens de Skorokhod :

$$Y_n(t) \rightarrow Y(t) := \exp(-\xi_{\tau(t)})$$

conjointement avec la convergence en loi suivante :

$$\frac{B_n}{a_n} \rightarrow \sigma := \int_0^\infty \exp(-\gamma \xi_r) dr.$$

Remarque. *Dans notre cas, X_n correspondra au nombre de sommets dans la composante connexe qui contient la racine. L'étude de la convergence de Y_n permet d'étudier la convergence de la distance à une feuille dans les arbres de logging. On ne donnera pas la preuve du théorème 5.2 puisqu'on expliquera comment la réadapter dans la section suivante.*

6 Essai de preuve de la conjecture

On souhaite modifier les résultats précédents pour les adapter au cas du graphe d'Erdős–Rényi. Comme on connaît le comportement asymptotique de la taille et du surplus, on commence par se restreindre à l'étude des graphes uniformes à n arêtes et s surplus (voir le corollaire 5.1).

Contrairement au cas de l'arbre de Cayley, si on conserve la procédure de logging telle qu'elle est, on n'obtiendra pas une structure markovienne. En effet, lorsqu'on supprime une arête, on risque de ne pas diviser le graphe en deux. Dans ce cas, on garde le même nombre de sommets dans ce bloc. A l'étape suivante, si on avait une structure markovienne, on a la même loi de subdivision pour ce bloc. Autrement dit, on peut recommencer un nombre arbitraire de fois ce qui est exclu. On doit donc conserver le nombre de surplus dans les blocs.

On alloue le symbole $\#$ pour souligner la présence d'un surplus lors du processus de logging. La figure 6 montre le changement à effectuer. Lorsqu'on oublie les noms

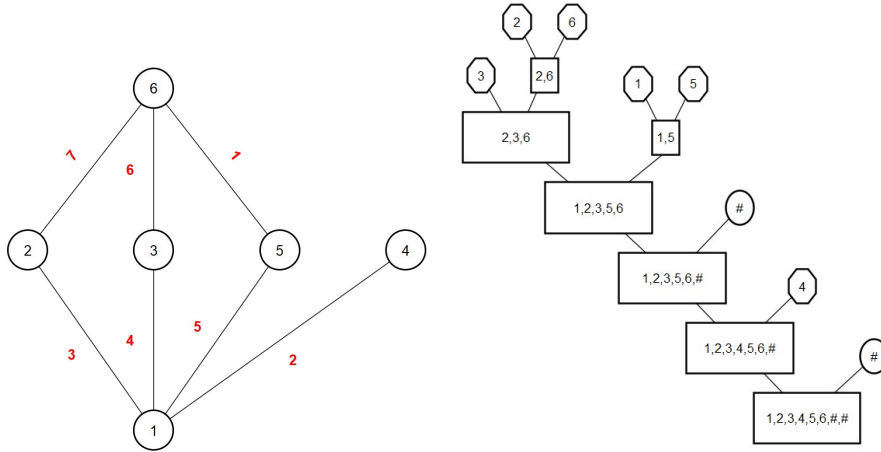


FIGURE 10 – Un graphe à 6 sommets et deux surplus et l'arbre de logging correspondant.

attribués à chaque sommet, on garde le nombre de surplus dans chaque bloc en attribuant un poids à chaque sommet. On modifie légèrement la définition [14] en :

Définition 6.1. Soit $(q_{n,s}, n, s \geq 1)$ une séquence de distributions de probabilités sur $\mathcal{P}_{n,s}$ l'ensemble des partitions du couple entier n, s tel que $q_{n,s}((n, s)) < 1$ pour tout $n, s \geq 1$. On définit la distribution markovienne branchante $P_{n,s}^q$ sur l'ensemble des arbres pondérés à n feuilles comme suit :

- $P_{1,0}^q$ est la loi d'un arbre à une feuille de longueur G une variable aléatoire géométrique de paramètre $q_{1,0}(\emptyset)$.
- pour tout $n \geq 2$, $P_{n,s}$ est la loi d'un arbre construit en reliant les arbres $(T^{(i)})_{1 \leq i \leq p(\Lambda)}$ à une racine où la distribution de Λ est q_n et conditionnellement à celle-ci, les arbres $(T^{(i)})_{1 \leq i \leq p(\Lambda)}$ sont indépendants et de loi P_{Λ_i} .

Il faut s'imaginer qu'on a mis un sac à $n + s$ balles dont s balles sont colorées qui représentent les arêtes en surplus. Le nombre de balles colorées représente le poids. L'étude du comportement de $\kappa_{n,s}$ se ramène à prouver la conjecture suivante :

Conjecture 2. *Pour tout s et n , on note $X_{n,s}(k)$ le nombre de sommets dans la composante connexe contenant 1 après k coupes et $S_{n,s}(k)$ le nombre de surplus correspondant. On définit $Y_n(t)$ par :*

$$Y_n(t) = \frac{X_{n,s}(\lfloor \sqrt{nt} \rfloor)}{n}.$$

Par ailleurs, on définit $B_{n,s}$ est l'instant à partir duquel on devient stationnaire qui est presque sûrement fini. *reverif* Alors, sous l'hypothèse 2, on a la convergence en loi au sens de Skorokhod :

$$Y_{n,s}(t) \rightarrow Y(t)$$

conjointement avec la convergence en loi suivante :

$$\frac{B_{n,s}}{\sqrt{n}} \rightarrow \sigma.$$

Plan : L'idée initiale était de s'inspirer de l'article de [14]. On résumera les étapes en précisant les points qui restent en suspens :

- La première étape est d'étudier la convergence de la fonction génératrice de $X_{n,s}(1)$. La probabilité que la première coupe produise des blocs de taille m , $n - m$ avec respectivement s , $s - u$ surplus est donnée pour $n/2 < m < n$ par :

$$q_{n,s}(m, u) = \frac{\binom{n}{m}(n-m)mA_{m,u}A_{n-m,s-u}}{A_{n,s}(n+s-1)}$$

On rappelle que $A_{n,s}$ est le nombre de graphes connexes à n sommets et s surplus. D'après l'article de Wright [27], on peut donner un équivalent à l'infini à s fixé si n tend vers l'infini :

$$A_{n,s} \sim f_s n^{n-2+\frac{3s}{2}} \quad (23)$$

ce qui permet de donner l'équivalent

$$n^{3/2}(1-m/n)\tilde{q}_{n,s}(m) \sim \frac{1}{\sqrt{2\pi(m/n)^3(1-m/n)}} \sum_{u=0}^s \frac{f_u f_{s-u}}{f_s} \left(\frac{m}{n}\right)^{3s/2} \left(1 - \frac{m}{n}\right)^{3(s-u)/2}$$

soit la convergence de $\sqrt{n}(1 - G_{n,s}(\lambda))$ vers une fonction ψ_s où

$$\forall \lambda, G_{n,s}(\lambda) = \mathbb{E} \left(\left(\frac{X_{n,s}(1)}{n} \right)^\lambda \right).$$

Le calcul de l'équivalent est donné dans la section 6.1.

- La deuxième étape consiste à s'assurer que $(Y_{n,s})_n$ est tendue. On le démontre en se basant sur le théorème suivant d'Aldous :

Lemme 6.1 (Aldous [9] Theorem 16.10). *Soit $(F_n, n \geq 0)$ une suite à valeurs dans \mathbb{D} , on note pour tout n , $\mathcal{J}(F_n)$ l'ensemble des temps d'arrêts pour la filtration associée à F_n . Supposons que pour tout $t > 0, \varepsilon > 0$:*

1. $\lim_{a \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\sup_{s \in [0, t]} F_n(s) > a \right) = 0$
2. $\lim_{\theta_0 \rightarrow 0} \limsup_{T \in \mathcal{J}} \sup_{\substack{(F_n), T \leq t_0 \leq \theta \leq \theta_0 \\ n \rightarrow \infty}} \sup \mathbb{P} (|F_n(T) - F_n(T + \theta)| > \varepsilon) = 0,$

alors $(F_n, n \geq 0)$ est tendue par rapport à la topologie de Skorokhod.

La vérification des hypothèses est expliquée dans la section 6.2.

- A partir de la compacité de Y_n , on déduit celle de :

$$Z_{n,s}(t) = Y_{n,s}(\tau_n^{-1}(t))$$

où

$$\tau_n^{-1}(t) = \inf \left\{ u \geq 0 : \int_0^u Y_{n,s}^{-1/2}(r) dr > t \right\}.$$

Si on extrait une sous-suite convergente de $Y_{n,s}$, la limite sera càdlàg et décroissante alors le théorème 1.5 du chapitre 6 [13] montre la convergence de $Z_{n,s}$ vers Z' défini par :

$$Z'(t) = Y'(\tau_{Y'}^{-1}(t))$$

où

$$\tau_{Y'}(u) = \begin{cases} \int_0^u Y'(r)^{-\frac{1}{2}} dr & \text{si } Y'(u) > 0 \\ \infty & \text{sinon} \end{cases}$$

et $\tau_{Y'}^{-1}(t) = \inf \{ u \geq 0 : \tau_{Y'}(u) > t \}$. A partir de cette étape, l'objectif est de démontrer que la limite Z' est unique en étudiant sa structure et de déduire ainsi l'unicité de la limite de Y .

- Une étape qu'on n'a pas traité est l'étude de la convergence de $S_{n,s}$. On admet qu'il existe $s_\infty \in \mathbb{D}$ tel que pour presque tout t :

$$S_{n,s}(\lfloor \sqrt{nt} \rfloor) \rightarrow s_\infty(t).$$

En plus, on admet que $(Y_{n,s}, Z_{n,s}, S_{n,s}(\lfloor \sqrt{n} * \rfloor))$ soit tendue. L'idée sera de mieux comprendre s_∞ à partir de la construction du graphe donné dans [1]. Le noyau du graphe à la limite d'échelle est un graphe 3 régulier tiré selon une loi explicite. Comprendre la loi l'écart des sauts de s_∞ revient à comprendre l'algorithme d'isolement des sommets appliqué à ce graphe 3-régulier. La loi des délais entre les sauts est reliée entre autre à la distance entre les sommets de degré 3.

- Dans l'article [14], ils démontrent dans le cas $s = 0$ que $Z' \stackrel{d}{=} e^{-t\xi_t}$ en prouvant que

$$(Z'(t)^\lambda \exp(\psi_0(\lambda)t), t \geq 0)$$

est une martingale càdlàg par rapport à sa filtration naturelle. Dans le cas général, on démontre que :

$$Z'(t)^\lambda \exp \left(\int_0^t \psi_{s_\infty(u)}(\lambda) du \right)$$

est une martingale càdlàg par rapport à sa propre filtration. En effet, elle héritera de la structure de martingale de

$$Z_{n,s}(t)^\lambda \left(\prod_{i=0}^{\lfloor a_n \tau_n^{-1}(t) \rfloor - 1} G_{X_{n,s}(i), S_{n,s}(i)}(\lambda) \right)^{-1}.$$

Les passages à la limite sont justifiés dans la section 6.3.

- Si la limite s_∞ est déterministe ou indépendante de Z' , on aurait l'unicité de la limite. Cependant, au vu de la structure limite de ces graphes, ceci semble peu probable. La preuve d'unicité n'est donc à ce stade pas complète.
- Enfin, le passage de la convergence de $Z_{n,s}$ vers celle de $B_{n,s}$ se base sur le même principe que [14]. Elle est expliquée dans la section 6.4.

On va maintenant donner les démonstrations des étapes prouvées.

6.1 Convergence de la fonction génératrice

On rappelle que $A_{n,s}$ est le nombre de graphes connexes à n sommets et s surplus. D'après l'article de Wright [27], on peut donner un équivalent à l'infini à s fixé si n tend vers l'infini :

$$A_{n,s} \sim f_s n^{n-2+\{(3s)/2\}} \quad (24)$$

On obtient d'une manière analogue à l'article [8] de Bertoin :

$$\begin{aligned} q_{n,s}(m, u) &= \frac{\binom{n}{m} (n-m) m A_{m,u} A_{n-m,s-u}}{A_{n,s}(n+s-1)} \\ &\sim \frac{\binom{n}{m} (n-m) m f_u^{m+\frac{3u-4}{2}} f_{s-u}^{m+\frac{3(s-u)-4}{2}}}{(n+s-1) f_s^{m+\frac{3s-4}{2}}} \\ &\sim q_{n,0}(m, 0) \frac{n-1}{n+s-1} \frac{f_u f_{s-u}}{f_s} \frac{m^{(3/2)s} (n-m)^{(3/2)(s-u)}}{n^{3s/2}} \\ &\sim q_{n,0}(m, 0) \frac{f_u f_{s-u}}{f_s} \left(\frac{m}{n}\right)^{(3/2)s} \left(1 - \frac{m}{n}\right)^{(3/2)(s-u)} \end{aligned}$$

Donc, on obtient :

$$n^{(3/2)} (1-m/n) \tilde{q}_{n,s}(m) \sim \frac{1}{\sqrt{2\pi(m/n)^3(1-m/n)}} \sum_{u=0}^s \frac{f_u f_{s-u}}{f_s} \left(\frac{m}{n}\right)^{3s/2} \left(1 - \frac{m}{n}\right)^{(3/2)(s-u)}$$

En tenant compte de l'échangeabilité, on a :

$$p_{n,k}(m, s) = \frac{m}{n} q_{n,k}(m, s)$$

où $(p_{n,k})$ est la matrice de transition de $(X_{n,s}, S_{n,s})$. On obtient que pour tout λ , et $s \in \mathbb{N}$,

$$\sqrt{n}(1 - G_{n,s}(\lambda)) \rightarrow \psi_s(\lambda) \quad (25)$$

où :

$$G_{n,s}(\lambda) = \mathbb{E} \left(\left(\frac{X_{n,s}(1)}{n} \right)^\lambda \right)$$

6.2 Étude de la compacité de $Y_{n,s}$

Les inégalités données dans la section 3 de l'article [14] pour $s = 0$ restent correctes. Il existe $c_1(\lambda), c_2(\lambda), n_0(\lambda)$ pour tout n et $s \leq s_0$,

$$\begin{aligned} 1 - G_{n,s}(\lambda) &\leq \frac{c_1(\lambda)}{\sqrt{n}} \\ -\ln(G_{n,s}(\lambda)) &\leq \frac{c_2(\lambda)}{\sqrt{n}} \leq c_3(\lambda) \\ \ln(G_{n,s}(\lambda)) &\leq -\frac{c_4(\lambda)}{\sqrt{n}} \end{aligned}$$

On définit les martingales suivantes :

$$\begin{aligned} \chi_{n,s}^{(\lambda)}(k) &= \left(\frac{X_{n,s}(k)}{n}\right)^\lambda + \sum_{i=0}^{k-1} \left(\frac{X_{n,s}(i)}{n}\right)^\lambda (1 - G_{X_{n,s}(i), S_{n,s}(i)}(\lambda)) \\ \Upsilon_n^{(\lambda)}(k) &= \left(\frac{X_{n,s}(k)}{n}\right)^\lambda \left(\prod_{i=0}^{k-1} G_{X_{n,s}(i), S_{n,s}(i)}(\lambda)\right)^{-1} \end{aligned}$$

On se propose d'utiliser le critère d'Aldous pour montrer que la suite $(Y_{n,s}(t))_n$ est tendue. Il suffit de montrer que :

$$\lim_{\theta_0 \rightarrow 0} \limsup_{n \rightarrow \infty} \sup_{T \in \mathcal{J}(Y_{n,s}), T \leq t} \sup_{\theta \leq \theta_0} \mathbb{P}(|Y_{n,s}(T) - Y_{n,s}(T + \theta)| > \varepsilon) = 0$$

On se donne un temps d'arrêt T borné par t . Comme $Y_{n,s}$ est décroissant,

$$\begin{aligned} \mathbb{E} \left[|Y_{n,s}(T) - Y_{n,s}(T + \theta)|^\lambda \right] &\leq \mathbb{E} \left[Y_{n,s}^\lambda(T) - Y_{n,s}^\lambda(T + \theta) \right] \\ &\leq \mathbb{E} \left(\chi_{n,s}^{(\lambda)}(\lfloor \sqrt{n}T \rfloor) - \chi_{n,s}^{(\lambda)}(\lfloor \sqrt{n}(T + \theta) \rfloor) \right) \\ &\quad + n^{-\lambda} \mathbb{E} \left[\sum_{i=\lfloor a_n T \rfloor}^{\lfloor a_n(T+\theta) \rfloor - 1} X_{n,s}^\lambda(i) (1 - G_{X_{n,s}(i), S_{n,s}(i)}(\lambda)) \right] \end{aligned}$$

Par soucis de lisibilité, on omettra les doubles indices si ceci ne conduit pas à une confusion. On applique le théorème d'arrêt de Doob :

$$\begin{aligned} \mathbb{E} \left[|Y_{n,s}(T) - Y_{n,s}(T + \theta)|^\lambda \right] &\leq n^{-\lambda} \mathbb{E} \left[\sum_{i=\lfloor a_n T \rfloor}^{\lfloor a_n(T+\theta) \rfloor - 1} X_{n,s}^\lambda(i) (1 - G_{X_{n,s}(i), S_{n,s}(i)}(\lambda)) \right] \\ &\leq c_1(\lambda) n^{-\lambda} \mathbb{E} \left[\sum_{i=\lfloor a_n T \rfloor}^{\lfloor a_n(T+\theta) \rfloor - 1} \frac{X_n^\lambda(i)}{\sqrt{X_n(i)}} \right] \\ &\leq \frac{c_1(\lambda) c_2'(\lambda - \gamma)}{a_n} \mathbb{E} [\lfloor a_n(T + \theta) \rfloor - \lfloor a_n T \rfloor] \\ &\leq c_1(\lambda) (\theta + a_n^{-1}) \end{aligned}$$

On conclut grâce à l'inégalité de Markov.

6.3 Unicité de la limite

On se donne $n > n_0(\lambda)/\varepsilon$, On considère une sous suite $Y_{n_k, s}$ qui converge vers Y' qu'on notera de la même manière par soucis de simplicité. On a la convergence de Z_n (extraite) vers Z' . On peut réécrire Z' comme solution d'un point fixe :

$$Z'(t) = Y' \left(\int_0^t Z'(r)^\gamma dr \right), \quad t \geq 0$$

Le premier objectif est de montrer que $Z'(t)^\lambda \exp \left(\int_0^t \psi_{s_{\infty}(u)}(\lambda) du \right)$ est une martingale cadlag. L'idée est de d'abord montrer que $Z'(t)^\lambda \exp \left(\int_0^t \psi_{s_{\infty}(u)}(\lambda) du \right)$ arrêtée est une martingale comme limite d'une martingale puis de conclure.

On opère un changement de temps sur $\Upsilon_n^{(\lambda)}(k)$ qui est une martingale par rapport à la filtration associée à (X_n, S_n) alors $\Upsilon_n^{(\lambda)}(\lfloor \sqrt{n}t \rfloor)$ est une martingale locale par rapport à $(Y_n, S_n(\lfloor \sqrt{n} * \rfloor))$. Comme $\tau_n^{-1}(t)$ est un temps d'arrêt borné par t , alors :

$$M_{n,s}^{(\lambda)}(t) = Z_{n,s}(t)^\lambda \left(\prod_{i=0}^{\lfloor a_n \tau_n^{-1}(t) \rfloor - 1} G_{X_{n,s}(i), S_{n,s}(i)}(\lambda) \right)^{-1}.$$

Comme on l'a précisé pour qu'on puisse passer à la limite, on considère d'abord les martingales arrêtées par :

$$T_{n,\varepsilon} := \inf \{ t \geq 0 : Z_{n,s}(t) \leq \varepsilon \}.$$

On note aussi :

$$T_\varepsilon := \inf \{ t \geq 0 : Z'(t) \leq \varepsilon \}.$$

On se place sur un espace probabilisé tel que Z_n, S_n converge presque sûrement par le théorème de Skorokhod. On montre de la même manière que dans [14] en utilisant le lemme de Fatou que C définie par l'ensemble des ε tels que

$$\mathbb{P}(\exists t_1, t_2 \geq 0 : t_1 \neq t_2, Z'(t_1) = Z'(t_2) = \varepsilon) > 0 \text{ ou } \mathbb{P}(Z'(T_\varepsilon -) = \varepsilon > Z'(T_\varepsilon)) > 0.$$

est au plus dénombrable. Autrement dit, on pourra prendre $\varepsilon \xrightarrow{\varepsilon \notin C} 0$. L'intérêt de cette étape consiste à pouvoir utiliser le lemme 5 de [14]. On en déduit qu'avec une probabilité 1 :

$$T_{n,\varepsilon} \rightarrow T_\varepsilon \quad \text{et} \quad (Z_{n,s}(t \wedge T_{n,\varepsilon}), t \geq 0) \rightarrow (Z'(t \wedge T_\varepsilon), t \geq 0).$$

On commence d'abord par montrer que :

$$M_n^{(\lambda)}(t \wedge T_{n,\varepsilon}) \rightarrow Z'(t \wedge T_\varepsilon)^\lambda \exp \left(\int_0^{t \wedge T_\varepsilon} \psi_{s_{\infty}(u)}(\lambda) du \right)$$

ce qui revient d'après le lemme 5 de [14] à montrer que :

$$\left(\prod_{i=0}^{\lfloor a_n \tau_n^{-1}(t \wedge T_{n,\varepsilon}) \rfloor - 1} G_{X_{n,s}(i), S_{n,s}(i)}(\lambda) \right)^{-1} \rightarrow \exp \left(\int_0^{t \wedge T_{n,\varepsilon}} \psi_{s_\infty(u)}(\lambda) du \right). \quad (26)$$

Comme on a $t < T_{n,\varepsilon}$, si $r < \lfloor a_n \tau_n^{-1}(t) \rfloor$, alors $X_n(\lfloor r \rfloor) > \varepsilon n > n_0(\lambda)$. Alors, $\ln(G_{X_{n,s}(\lfloor r \rfloor), S_{n,s}(\lfloor r \rfloor)})$ est bien défini. On en déduit

$$\begin{aligned} & \left(\prod_{i=0}^{\lfloor a_n \tau_n^{-1}(t \wedge T_{n,\varepsilon}) \rfloor - 1} G_{X_{n,s}(i), S_{n,s}(i)}(\lambda) \right)^{-1} = \exp \left(\sum_{i=0}^{\lfloor a_n \tau_n^{-1}(t \wedge T_{n,\varepsilon}) \rfloor - 1} -\ln(G_{X_{n,s}(i), S_{n,s}(i)})(\lambda) \right) \\ & = \exp \left(\int_0^{\lfloor a_n \tau_n^{-1}(t \wedge T_{n,\varepsilon}) \rfloor} -\ln(G_{X(\lfloor r \rfloor), S(\lfloor r \rfloor)})(\lambda) dr \right) \\ & = \exp \left(\int_0^{a_n \tau_n^{-1}(t \wedge T_{n,\varepsilon})} -\ln(G_{X(\lfloor r \rfloor), S(\lfloor r \rfloor)})(\lambda) dr - \int_{\lfloor a_n \tau_n^{-1}(t \wedge T_{n,\varepsilon}) \rfloor}^{a_n \tau_n^{-1}(t \wedge T_{n,\varepsilon})} -\ln(G_{X(\lfloor r \rfloor), S(\lfloor r \rfloor)})(\lambda) dr \right) \\ & = \exp \left(\int_0^{(t \wedge T_{n,\varepsilon})} -\ln(G_{nZ(\lfloor r \rfloor), S(\lfloor a_n \tau_n^{-1}(r) \rfloor)})(\lambda) \sqrt{nZ_n(r)} dr - \int_{\lfloor a_n \tau_n^{-1}(t \wedge T_{n,\varepsilon}) \rfloor}^{a_n \tau_n^{-1}(t \wedge T_{n,\varepsilon})} -\ln(G_{X(\lfloor r \rfloor), S(\lfloor r \rfloor)})(\lambda) dr \right). \end{aligned}$$

On s'intéresse d'abord à la première intégrale. Comme tout est croissant et qu'on cherche à prouver la convergence vers un processus continu, il suffit de prouver la convergence ponctuelle pour montrer la convergence sur tous compacts d'après le théorème de Dini. On veut utiliser le théorème de convergence dominée.

Comme $(Z_{n,s}(t \wedge T_{n,\varepsilon}), t \geq 0) \rightarrow (Z'(t \wedge T_\varepsilon), t \geq 0)$ au sens de Skorokhod, $Z_n(r) \rightarrow Z'(r)$ pour tout $r \in \mathcal{C}(Z')$ donc presque partout. Pour presque tout r , on a

$$(nZ_n(r), S(\lfloor a_n \tau_n^{-1}(r) \rfloor)) \rightarrow (\infty, s_\infty(r)).$$

D'après l'équation (25), on obtient la convergence :

$$-\ln(G_{nZ(\lfloor r \rfloor), S(\lfloor a_n \tau_n^{-1}(r) \rfloor)})(\lambda) \sqrt{nZ_n(r)} \rightarrow \psi_{s_\infty(r)}.$$

On a déjà remarqué que si $r < a_n \tau_n^{-1}(t \wedge T_{n,\varepsilon})$ alors $X_n(\lfloor r \rfloor) > n_0(\lambda)$ ce qui nous permet d'obtenir une domination :

$$-\ln G_{X_{n,s}(\lfloor r \rfloor), S_{n,s}(\lfloor r \rfloor)}(\lambda) \leq \frac{c_2(\lambda)}{\sqrt{X_{n,s}(\lfloor r \rfloor)}},$$

soit

$$-\ln(G_{nZ(\lfloor r \rfloor), S(\lfloor a_n \tau_n^{-1}(r) \rfloor)})(\lambda) \sqrt{nZ_n(r)} \leq c_2(\lambda) \text{ et } M_n^{(\lambda)}(t \wedge T_{n,\varepsilon}) \leq \exp(c_\lambda t).$$

D'après le théorème de convergence dominée, on obtient (26). L'autre terme converge vers 0 par le même raisonnement que dans [14]. Comme $M_n^{(\lambda)}(t \wedge T_{n,\varepsilon}) \leq \exp(c_\lambda t)$, pour tout t , $(M_n^{(\lambda)}(t \wedge T_{n,\varepsilon}))_{n \geq n_0(\lambda)}$ est uniformément intégrable. D'après l'équation

(26), on en déduit que $Z'(t \wedge T_{n,\varepsilon})^\lambda \exp\left(\int_0^{t \wedge T_{n,\varepsilon}} \psi_{s_\infty(u)}(\lambda) du\right)$ est une martingale.

De la même manière, on remarque que :

$$Z'(t \wedge T_{n,\varepsilon})^\lambda \exp\left(\int_0^{t \wedge T_{n,\varepsilon}} \psi_{s_\infty(u)}(\lambda) du\right) \rightarrow Z'(t)^\lambda \exp\left(\int_0^t \psi_{s_\infty(u)}(\lambda) du\right)$$

et

$$Z'(t \wedge T_{n,\varepsilon})^\lambda \exp\left(\int_0^{t \wedge T_{n,\varepsilon}} \psi_{s_\infty(u)}(\lambda) du\right) \leq e^{t \sup_{s' < s} (\psi_{s'}(\lambda))}$$

ce qui suffit pour conclure que $Z'(t)^\lambda \exp\left(\int_0^t \psi_{s_\infty(u)}(\lambda) du\right)$ est une martingale.

6.4 De la convergence de $Z_{n,s}$ vers la convergence de $B_{n,s}$

La suite $(Z_{n,s})$ vérifie :

$$Z_{n,s}(t)^\lambda \leq \frac{c_{\lambda,\beta} M_n^{(\lambda)}(t) + 1}{t^\beta}.$$

On explique maintenant les étapes qui mènent à la conclusion. On remarque $B_{n,s}/a_n$, la première fois que $Y_{n,s}$ touche 0, vérifie :

$$\int_0^\infty Z_{n,s}(r)^\gamma dr = \sigma_{n,s} = B_{n,s}/a_n.$$

On en conclut que $\sigma_{n,s}$ est tendue et on se ramène donc à un espace sur lequel $(Y_{n,s}, \sigma_{n,s}, Z_{n,s}, S_{n,s}(\sqrt{n^*}))$ converge presque sûrement. On extrait une sous suite. On note σ' la limite. De la même manière, on a :

$$\sigma' = \liminf_{n \rightarrow \infty} \inf \{t \geq 0 : Y_{n,s}(t) = 0\} \geq \inf \{t \geq 0 : Y_s(t) = 0\} = \sigma.$$

On cherche donc à montrer $\mathbb{E}(\sigma') \leq \mathbb{E}(\sigma)$. On commence par appliquer le lemme de Fatou pour intervertir la limite inférieure et espérance :

$$\mathbb{E}[\sigma'] \leq \liminf_n \mathbb{E}\left[\int_0^\infty Z_{n,s}(r)^{1/2} dr\right].$$

On remarque par ailleurs que :

$$\mathbb{E}\left[\int_0^\infty Z_s(r)^{1/2} dr\right] = \mathbb{E}[\sigma].$$

Il suffit donc de prouver que :

$$\mathbb{E}\left[\int_0^\infty Z_{n,s}(r)^{1/2} dr\right] \xrightarrow{n \rightarrow \infty} \mathbb{E}\left[\int_0^\infty Z_s(r)^{1/2} dr\right]. \quad (27)$$

Par le théorème de Fubini, comme $Z_{n,s} \rightarrow Z_s$ converge pour presque tout s presque sûrement, $Z_{n,s} \rightarrow Z_s$ converge presque sûrement pour presque tout s . On applique le théorème de convergence dominé deux fois et on obtient :

$$\int_0^\infty \mathbb{E}[Z_{n,s}(r)^\gamma] dr \rightarrow \int_0^\infty \mathbb{E}[Z(r)^\gamma] dr < \infty. \quad (28)$$

On conclut avec une autre utilisation de Fubini. \square

7 Conclusion

Ce rapport étudie les graphes d’Erdős–Rényi sous plusieurs angles. Dans un premier temps, on a expliqué le phénomène de transition de phase. Puis, on s’est intéressé au comportement critique de la taille et du surplus des composantes connexes. Finalement, on a étudié la limite d’échelle de ce graphe au sens de Gromov-Hausdorff.

Dans un second temps, on s’est concentré sur l’algorithme d’isolement des sommets. D’abord, on a rappelé le chemin de preuve pour le comportement du nombre de coupes dans l’arbre de Cayley. Ensuite, on a donné le plan de preuve pour le cas des graphes d’Erdős–Rényi. On s’est essentiellement ramené à l’étude du comportement du processus des surplus au cours de l’isolement qui semble indiquer la nécessité d’une meilleure compréhension de la structure des graphes 3-réguliers aléatoires qui apparaissent comme le noyau asymptotique des graphes.

Références

- [1] Louigi Addario-Berry, Nicolas Broutin, and Christina Goldschmidt. Critical random graphs : Limiting constructions and distributional properties. *ELECTRONIC JOURNAL OF PROBABILITY*, 15 :741–, 08 2010.
- [2] Louigi Addario-Berry, Nicolas Broutin, and Christina Goldschmidt. The continuum limit of critical random graphs. *Probability Theory and Related Fields*, 152 :367–406, 09 2012.
- [3] Louigi Addario-Berry, Luc Devroye, and Svante Janson. Sub-Gaussian tail bounds for the width and height of conditioned galton–watson trees. *Ann Probab*, 41, 11 2010.
- [4] David Aldous. The continuum random tree. I. *The Annals of Probability*, 19, 01 1991.
- [5] David Aldous. The continuum random tree II : an overview. page 23–70, 1991.
- [6] David Aldous. The continuum random tree III. *The Annals of Probability*, 21, 01 1993.
- [7] David Aldous. Brownian excursions, critical random graphs and the multiplicative coalescent. *The Annals of Probability*, 25, 08 1996.
- [8] Jean Bertoin. Fires on trees. *Annales de l’Institut Henri Poincaré Probabilités et Statistiques*, 48, 11 2010.
- [9] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics : Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.
- [10] Burgess Davis and David McDonald. An elementary proof of the local central limit theorem. *Journal of Theoretical Probability*, 8 :693–701, 07 1995.
- [11] Burago Dmitri. *A course in metric geometry / Dmitri Burago, Yuri Burago, Sergei Ivanov*. Graduate studies in mathematics. American Mathematical Society, Providence (R.I.), C 2001.

- [12] Paul Erdos and Alfred Renyi. On the evolution of random graphs. *Publ. Math. Inst. Hungary. Acad. Sci.*, 5 :17–61, 1960.
- [13] S. Ethier and Thomas Kurtz. *Markov Processes : Characterization and Convergence*, volume 43. 01 1986.
- [14] Benedicte Haas and Grégory Miermont. Self-similar scaling limits of non-increasing Markov chains. *Bernoulli*, 17, 09 2009.
- [15] Benedicte Haas and Grégory Miermont. Scaling limits of Markov branching trees, with applications to Galton-Watson and random unordered trees. *Ann Probab*, 40, 11 2012.
- [16] Remco Hofstad. *Random graphs and complex networks*. 01 2016.
- [17] Svante Janson. Random cutting and records in deterministic and random trees. *Random Struct. Algorithms*, 29 :139–179, 09 2006.
- [18] Jean-François Marckert and Abdelkader Mokraddem. The depth first processes of galton–watson trees converge to the same brownian excursion. *Ann Probab*, 31, 07 2003.
- [19] Cyril Marzouk. Scaling limits of random bipartite planar maps with a prescribed degree sequence. *Random Structures & Algorithms*, 53, 12 2016.
- [20] Colin McDiarmid. *Concentration*, pages 195–248. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998.
- [21] A. Meir and J. Moon. Cutting down random trees. *Journal of The Australian Mathematical Society - J AUST MATH SOC*, 11, 08 1970.
- [22] A. Meir and J. Moon. Cutting down recursive trees. *Bellman Prize in Mathematical Biosciences*, 21 :173–181, 1974.
- [23] Alois Panholzer. Cutting down very simple trees. *Quaestiones Mathematicae*, 29 :211–227, 06 2006.
- [24] Yuri Pavlov. The asymptotic distribution of maximum tree size in a random forest. *Theory of Probability & Its Applications*, 22, 01 1977.
- [25] Lajos Takács. A combinatorial theorem for stochastic processes. *Bulletin of the American Mathematical Society*, 71(4) :649 – 650, 1965.
- [26] E. Wright. The number of sparsely edged graphs. *Journal of Graph Theory*, 1 :317 – 330, 12 1977.
- [27] E. Wright. The number of connected sparsely edged graphs. III. Asymptotic results. *Journal of Graph Theory*, 4 :393 – 407, 10 2006.

Annexe

Preuve du lemme technique 4.1

Remarque. *On présente une démonstration qui est le recollement de la preuve du théorème 1.1 dans [3] et de la preuve du lemme 4 dans [19] parce que l'inégalité (5) issue de l'application du théorème du scrutin [25] dans la preuve originale de l'article [2] est dans le mauvais sens et elle ne permet pas de conclure.*

Preuve : Afin de comprendre la marche en profondeur X_m associée à T_m , on commence par se ramener à l'arbre de Galton–Watson conditionné.

Comme on l'a déjà précisé, un arbre de Cayley est essentiellement un arbre de Galton–Watson planaire poissonnien de paramètre 1 conditionnellement à $|\mathcal{T}| = m$ uniformément étiqueté tel qu'on a oublié l'ordre des enfants. En effet, les arbres de Galton–Watson sont par construction planaires c'est-à-dire pour chaque sommet, on se fixe un ordre sur ces enfants ([15] pour une définition précise). Il s'en suit qu'on a un ordre inhérent sur les sommets. La probabilité d'obtenir un arbre T est proportionnelle à $\frac{e^{-m}}{\prod_{i \in T} d_i!}$. Pour chaque arbre planaire, il lui correspond $\prod_{i \in T} d_i!$ ainsi que $m!$ étiquetages possibles. La probabilité d'obtenir un arbre ne dépend plus de celui-ci ce qui prouve l'assertion. On en déduit que le processus X_m a la même loi que la marche aléatoire en largeur associée à $\mathcal{T}(\mathcal{P}(1))$ conditionnée à $|\mathcal{T}(\mathcal{P}(1))| = m$.

Il nous reste à étudier la marche aléatoire en largeur associée à $\mathcal{T}(\mathcal{P}(1))$. On rappelle que \mathcal{O}_i est la file de sommets actifs à l'étape i et on note $\mathcal{Q}_i = |\mathcal{O}_i|$. A chaque étape du parcours en largeur d'un arbre $\mathcal{T}(\mathcal{P}(1))$, on a exploré un sommet et on a rajouté ses enfants à la file. On en déduit que $\mathcal{Q}_1 = 1$ et que pour tout $i > 1$, $\mathcal{Q}_{i+1} = \mathcal{Q}_i - 1 + \xi_i$ où $(\xi_i)_{i \in \mathbb{N}}$ des variables aléatoires indépendantes qui suivent la loi de Poisson de paramètre 1. L'exploration de l'arbre s'arrête dès qu'il n'y a plus de sommets dans la file c'est-à-dire $\mathcal{Q}_i = 0$. Le processus d'exploration s'écrit comme suit :

$$Y(j) = \sum_{i=1}^j (\xi_i - 1).$$

On conclut que

$$\mathbb{P}(\|X_m\| \geq x) = \mathbb{P}\left(\max_{j \leq n} Y(j) | Y_n = -1; Y_i \geq 0 \forall i < n\right).$$

Si on conditionne seulement par rapport à l'évènement $Y_n = -1$, la marche associée aux $(\xi_i)_{i \in \mathbb{N}}$ peut descendre en dessous de -1 , comme on peut le voir dans la figure 7. Afin de construire une excursion qui reste au dessus du niveau -1 , on effectue un argument de rotation.

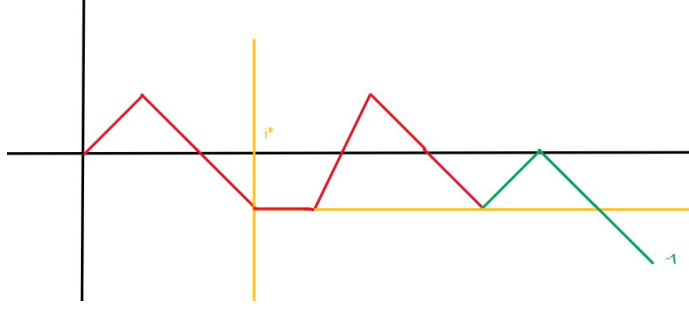


FIGURE 11 – L’argument de rotation appliqué à la suite $\xi = (1, -1, -1, 0, 2, -1, -1)$. La courbe en rouge correspond à la marche associée à ξ . La courbe verte est la marche après la rotation.

Pour illustrer l’argument de rotation, on se donne pour tout $1 \leq i \leq n$, $x_i \geq 0$ et tels que $\sum_i x_i = -1$. On note i^* le premier indice qui réalise le minimum :

$$i^*(x) = \min \operatorname{argmin}_{0 \leq l \leq n} \sum_{j=1}^l x_j.$$

On remarque que si on pose $\tilde{x}_i = x_{i+i^*(x)}$ ⁴, alors pour tout $0 \leq l < n$, $\sum_{j=1}^l x_j \geq 0$. On applique la procédure en prenant $x_i = \xi_i - 1$. On conserve la notation \tilde{x} pour les incréments après la rotation et on note également $\tilde{Y}_i = \sum_{j=1}^i \tilde{x}_j$. Finalement, on remarque que (\tilde{Y}_i) a la même loi que (Y_i) conditionnellement à l’évènement $Y_n = -1$; $Y_i \geq 0 \forall i < n$ et que l’écart entre le maximum de Y et le minimum ne varie que de 1 au plus :

$$\max_{j \leq n} \tilde{Y}_j \leq \max_{j \leq n} Y_j - \min_{j \leq n} Y_j.$$

On déduit de ce qui précède :

$$\begin{aligned} \mathbb{P}(\|X_m\| \geq x) &= \mathbb{P}\left(\max_{j \leq n} Y_j \geq x \mid Y_n = -1; Y_i \geq 0 \forall i < n\right) \\ &\leq \mathbb{P}\left(\max_{j \leq n} Y_j - \min_{j \leq n} Y_j \geq x \mid Y_n = -1\right). \end{aligned}$$

On va montrer que :

$$\mathbb{P}\left(\max_{j \leq n} Y_j - \min_{j \leq n} Y_j \geq 3x \mid Y_n = -1\right) \leq 4\mathbb{P}\left(\min_{j \leq \lfloor n/2 \rfloor} Y_j \leq -x \mid Y_n = -1\right). \quad (29)$$

On note :

$$M_1 = \max_{0 \leq k \leq r/2} Y_k, \quad m_1 = \min_{0 \leq k \leq r/2} Y_k, \quad M_2 = \max_{r/2 \leq k \leq r} Y_k, \quad m_2 = \min_{r/2 \leq k \leq r} Y_k.$$

On remarque que m_1 , $Y_{n/2} - M_1$, $-1 - M_2$ et $m_2 - Y_{n/2}$ ont la même loi conditionnellement à $Y_n = -1$ et que le terme dans la probabilité de gauche de l’équation (29)

4. Les indices sont pris modulo n

se réécrit $\max(M_1, M_2) - \min(m_1, m_2)$. Si tous ces termes ($m_1, Y_{n/2} - M_1, -1 - M_2$ et $m_2 - Y_{n/2}$) sont supérieures à $-x$ alors

$$\begin{aligned} M_1 - m_1 &< Y_{n/2} + x + x < m_2 + 3x < 3x \\ M_1 - m_2 &< Y_{n/2} + x - Y_{n/2} + x < 3x \\ M_2 - m_1 &< -1 + x + x < 2x \\ M_2 - m_2 &< -1 + x + x - Y_{n/2} < -1 + 2x + (x - M_1) < 3x. \end{aligned}$$

Autrement dit, on aurait $\max(M_1, M_2) - \min(m_1, m_2) < 3x$. On en déduit (29).

La preuve du lemme se ramène à majorer $\mathbb{P}(\min_{j \leq [n/2]} Y_j \leq -x | Y_n = -1)$. On utilise la structure markovienne en appliquant la propriété de Markov au temps $[n/2]$. On note pour simplifier $\theta_k(x) = \mathbb{P}(Y_k = -x)$

$$\begin{aligned} \mathbb{P}\left(\min_{0 \leq k \leq [n/2]} Y_k \leq -x \mid Y_n = -1\right) &= \frac{\mathbb{P}(\min_{0 \leq k \leq [n/2]} Y_k \leq -x \text{ et } Y_n = -1)}{\mathbb{P}(Y_n = -1)} \\ &= \mathbb{E}\left(\mathbf{1}_{\min_{0 \leq k \leq [n/2]} Y_k \leq -x} \frac{\theta_{n-[n/2]}(S_{[n/2]} + 1)}{\theta_n(-1)}\right). \end{aligned}$$

En appliquant le théorème central limite local (théorème 1.1 [10]), on montre que :

$$C := \sup_{r \geq 1, x \in \mathbf{Z}} \frac{\theta_{r-[r/2]}(x)}{\theta_r(0)} < \infty.$$

Par ailleurs, comme les incréments de $-S$ sont majoré par 1, le théorème de McDiarmid (théorème 2.7 [20] et la remarque à la fin de la section 2) montre que :

$$\mathbb{P}\left(\max_{0 \leq k \leq n} -S_k \geq x\right) \leq \exp\left(-\frac{x^2}{2n + 2x/3}\right).$$

Si $x > n$, alors $\mathbb{P}(\max_{0 \leq k \leq n} -S_k \geq x) = 0$. Il existe $C_1 > 0$ qui ne dépend pas de x et n telle que

$$\mathbb{P}\left(\max_{0 \leq k \leq n} -S_k \geq x\right) \leq \exp\left(-\frac{C_1 x^2}{n}\right).$$

On en déduit :

$$\mathbb{P}(\|X_m\| \geq x) \leq C_2 \exp\left(-\frac{C_3 x^2}{n}\right)$$

pour des constants C_2 et C_3 qui ne dépendent pas de n et x . □