Graphes aléatoires : taille de la plus grande composante connexe

Emmanuel Dupuy Lucien Pech
Sujet proposé par Philippe Marchal

28 Juin 2007

Résumé

Nous nous intéressons à l'étude des graphes aléatoires simples $\mathcal{G}(n, p)$, où n est le nombre de sommets, et p la probabilité qu'une paire de sommets donnée soit une arête. Plus précisément, nous cherchons à évaluer la taille de la plus grande composante connexe de $\mathcal{G}(n, c/n)$ quand n tend vers l'infini. Nous démontrerons que cette taille a pour ordre de grandeur :

- $\ln(n)$ si $c < 1$,
- $\alpha_c n$ si $c > 1$, avec $0 < \alpha_c < 1$,
- $n^{2/3}$ si $c = 1$.

Table des matières

1 Introduction et premiers résutats

1.1 Définitions et notations

Nous commençons par définir les notions utiles par la suite :

Définition 1.1. *$\mathcal{G}(n,p)$ représente un graphe aléatoire simple non orienté à n sommets, dont la loi de probabilité est la suivante : si, pour toute paire de sommets distincts i et j, on note $A_{i,j}$ l'événement : "$\{i,j\}$ est une arête" (c'est à dire que les deux sommets soient reliés), les $A_{i,j}$ sont tous de probabilité p, et sont indépendants.*

Définition 1.2. *Pour un graphe $G \in \mathcal{G}(n,p)$ on appelle composante connexe tout sous-ensemble C de sommets de G, tel que toute paire de points de C peut être reliée par une suite d'arêtes adjacentes, et qui soit maximal pour cette propriété. Les différentes composantes connexes d'un graphe en forment donc une partition.*

Pour un point i du graphe, on note $\mathcal{C}(i)$ l'unique composante connexe contenant i, et on appele taille d'une composante connexe le nombre de sommets qui la composent.

On note enfin $|\mathcal{C}_{max}|$ la taille de la (ou des) composante(s) connexe(s) comportant le plus de sommets.

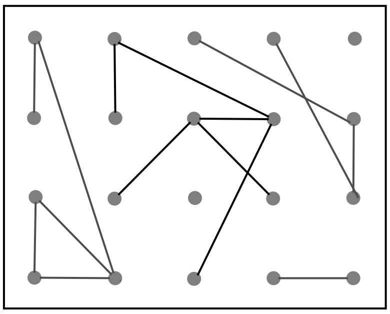


FIG. 1 – Exemple de graphe : 3 composantes connexes de tailles 12, 7 et 1

Définition 1.3. *On note $\mathcal{B}(n,p)$ la loi binomiale de paramètres n et p. On rappelle que X suit une telle loi si $\mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$, et qu'on a alors $\mathbb{E}[X] = np$ et $Var(X) = np(1-p)$.*

Définition 1.4. *On note $\mathcal{P}(c)$ la loi de Poisson de paramètre c. On rappelle que X suit une telle loi si $\mathbb{P}(X=k) = \exp(-c)\frac{c^k}{k!}$, et qu'on a alors $\mathbb{E}[X] = c$.*

Définition 1.5. *Soient A et B deux variables aléatoires réelles. On dit que A domine stochastiquement B si*

$$\forall x \in \mathbb{R}, \quad \mathbb{P}(A \geqslant x) \geqslant \mathbb{P}(B \geqslant x).$$

1.2 Résultats admis

Sont regroupés ici des résultats techniques généraux, dont les preuves peuvent facilement se trouver dans les ouvrages de référence traitant de la théorie des probabilités.

Théorème 1.1 (Théorème d'arrêt). *Si* $(X_n)_{n\in\mathbb{N}}$ *est une martingale adaptée à la filtration* $(\mathcal{F}_n)$*, et* T *est un temps d'arrêt, alors* $(X_{n\wedge t})_{n\in\mathbb{N}}$ *est une martingale.*

Théorème 1.2 (Inégalité des grandes déviations, loi binomiale)**.** *Soit* $\varepsilon > 0$.
$\exists\, \alpha_\varepsilon, \beta_\varepsilon > 0$ *tels que :*

- $\mathbb{P}\big(\mathcal{B}(n, \frac{t}{n}) \geqslant (1+\varepsilon)\, t\big) \leqslant \exp(-\alpha_\varepsilon t)$,
- $\mathbb{P}\big(\mathcal{B}(n, \frac{t}{n}) \leqslant (1-\varepsilon)\, t\big) \leqslant \exp(-\beta_\varepsilon t)$.

On pourra se référer à [1] (pages 237 à 239) pour une preuve de cette inégalité.

Théorème 1.3. *Soient* A *et* B *deux variables aléatoires réelles.* A *domine stochastiquement* B *si et seulement s'il existe deux variables aléatoires* A' *et* B' *définies sur un même espace de probabilité, de lois respectives les lois de* A *et* B*, et telles que* $A \geqslant B$ *presque sûrement.*

Théorème 1.4. *Soient* A *et* B *deux variables aléatoires dans* $\mathbb{R}^+$. A *domine stochastiquement* B *si et seulement si, pour toute fonction croissante* $f\colon \mathbb{R}^+ \to \mathbb{R}^+$*, on a* $\mathbb{E}\big[f(A)\big] \geqslant \mathbb{E}\big[f(B)\big]$.

1.3 Exploration du graphe

Pour calculer la taille d'une composante connexe $\mathcal{C}(i)$, la méthode que nous utilisons est connue sous le nom de *parcours en profondeur*. Il s'agit du processus d'exploration suivant : à tout instant $t \in [\![0, n]\!]$, on associe à chaque sommet (ou point) un unique état : actif, mort ou neutre. On suppose que les sommets sont numérotés, et que i est le premier.

On note Y_t le nombre de sommets actifs à l'instant t. Au temps $t = 0$, le seul point i est actif, c'est le point de départ du processus ; les autres sont neutres. Ensuite à chaque instant $t \in [\![1, n]\!]$, on procède comme suit :

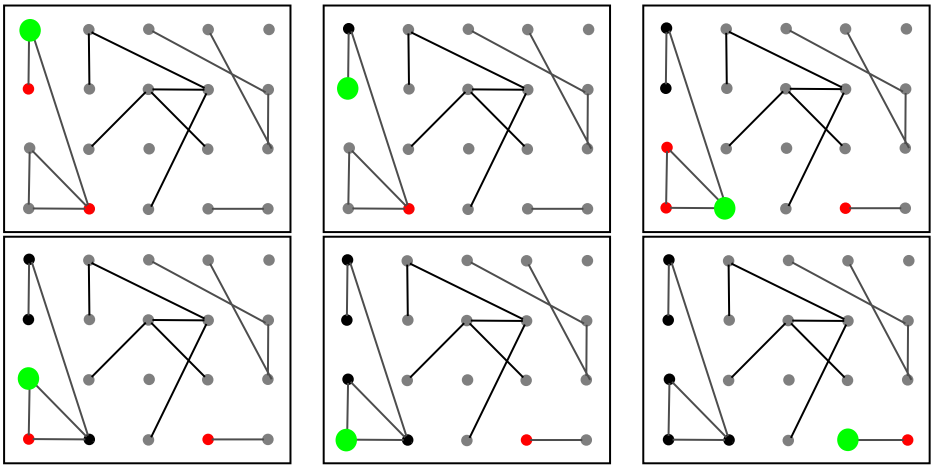
- Si $Y_{t-1} \neq 0$, j_t est le premier sommet actif. Sinon, c'est le premier sommet neutre.
- Les sommets neutres qui sont reliés à j_t deviennent actifs. On note η_t leur nombre.
- Le point j_t prend l'état "mort".

Autrement dit, on a

$$Y_t = \begin{cases} Y_{t-1} + \eta_t - 1 & \text{si} \quad Y_{t-1} \geqslant 0, \\ \eta_t & \text{sinon.} \end{cases} \tag{1}$$

Le nombre de sommets morts au temps t^+ vaut toujours t, et le processus s'arrête lorsque tous les points ont été explorés (*i.e.* sont "morts"). On voit que Y_t s'annule lorsqu'une composante connexe a été explorée en entier.

Exemple : On donne les premières étapes de l'exploration du graphe proposée plus haut. À chaque étape, on représente le sommet j_t par un grand disque vert, les sommets actifs en rouge, les neutres en gris et les morts en noir.



Pour déterminer la taille de $\mathcal{C}(i)$, nous cherchons donc à déterminer la loi de Y_t, donc celle de η_t. Au temps t^-, il y a $t-1$ sommets morts et Y_{t-1} sommets actifs, auxquels il faut enlever j_t si $Y_{t-1} = 0$.

Il reste alors $n-(t-1)-Y_{t-1}-1_{Y_{t-1}=0}$ sommets neutres. La probabilité que j_t soit lié à chacun d'entre deux vaut p, et ces évènements sont indépendants, conformément à la loi de $\mathcal{G}(n,p)$. On en déduit donc que $\mathbb{E}[\eta_t \mid Y_{t-1}]$ suit une loi binomiale $\mathcal{B}(n_{t-1}, p)$, avec $n_t = n - t - Y_t - 1_{Y_t=0}$.

On voit que, au moins au début de l'exploration, Y_t se comporte comme une marche aléatoire sur $\mathbb{Z}$ de loi de saut $\mathcal{B}(n,p) - 1$. On sait que le comportement de ce processus dépend du signe de l'espérance de la loi de saut, qui vaut ici $np-1$, ce qui nous conduit à poser $p = \frac{c}{n}$.

2 Les cas $c < 1$ et $c > 1$

2.1 Notations, résultats communs

2.1.1 Notations

On modifie le processus Y_t en Y'_t de la façon suivante : $Y'_0 := 1$; si $Y'_t = y_i$, alors $Y'_{t+1} := Y'_t + Z'_t - 1$, où Z'_t est une variable aléatoire de loi $\mathcal{B}(n - t - y_t, p)$. Par analogie avec Y_t, on note $n'_t := n - t - Y'_t$. On note T le temps d'entrée du processus Y'_t en 0, *i.e.* $T := \inf(t \in \mathbb{N}^* : (Y'_t = 0))$. Y'_t a même loi que Y_t pour $t < T$, et T a même loi que la taille de la composante connexe $\mathcal{C}(i)$.

On note Y^*_t la marche aléatoire issue de 1 de loi de transition $\mathcal{P}(c) - 1$ (avec $\mathcal{P}(c)$ la loi de Poisson de paramètre c). On a donc $Y^*_t = 1 + (\sum_{i=1}^t Z^*_i) - t$, où les Z^*_i sont des variables aléatoires indépendantes de loi $\mathcal{P}(c)$. On note T^* le temps d'entrée en 0, *i.e.* $T^* := \inf(t \in \mathbb{N}^* : Y^*_t = 0))$.

2.1.2 Idées de preuves

Pour évaluer la taille de la composante connexe $\mathcal{C}(i)$, nous allons évaluer la loi de T en comparant les processus Y'_t et Y^*_t. Pour cela, nous allons montrer le :

Théorème 2.1. *Soit $t \in \mathbb{N}$ fixé. Alors $\lim_{n\to\infty} \mathbb{P}(T \leqslant t) = \mathbb{P}(T^* \leqslant t)$.*

On a $Y'_t = 1 + \sum_{k=1}^t Z'_t - t$, et $Y^*_t = 1 + \sum_{k=1}^t Z^*_t - t$. Soit $t \in \mathbb{N}^*$ fixé. On a

$$\mathbb{P}(Y'_t = 0) = \sum_{z_1 + \ldots + z_t = t-1} \prod_{i=1}^t \mathbb{P}(Z'_i = z_i).$$

On fixe $(z_1, \ldots, z_t) \in \mathbb{N}^t$ tel que $z_1 + \ldots + z_t = t - 1$.

$\mathbb{P}((Z'_1, \ldots, Z'_t) = (z_1, \ldots, z_t)) = \prod_{k=1}^t \mathbb{P}(Z'_k = z_k)$ où les Z'_k sont des variables aléatoires de loi $\mathcal{B}(n - k - z_1 - \ldots - z_{k-1}, \frac{c}{n})$.

De même : $\mathbb{P}((Z^*_1, \ldots, Z^*_t) = (z_1, \ldots, z_t)) = \prod_{k=1}^t \mathbb{P}(Z^*_k = z_k)$.

Pour évaluer $\mathbb{P}(Z'_k = z_k)$, nous aurons besoin du lemme suivant :

Lemme 2.1. *Soient m_n suite à valeurs dans $\mathbb{N}$ telle que $m_n \underset{n\to\infty}{\sim} n$, $k \in \mathbb{N}$ fixé. Alors $\mathbb{P}(\mathcal{B}(m_n, \frac{n}{c}) = k) \underset{n\to\infty}{\to} \mathbb{P}(\mathcal{P}(c) = k)$.*

Démonstration. On a $\mathbb{P}(\mathcal{B}(m_n, \frac{n}{c}) = k) = \binom{m_n}{k}(\frac{c}{n})^k(1 - \frac{c}{n})^{m_n - k}$.

Or $\binom{m_n}{k} \underset{n\to\infty}{\sim} \frac{m_n^k}{k!} \underset{n\to\infty}{\sim} \frac{n^k}{k!}$ et $(1 - \frac{c}{n})^{m_n - k} \underset{n\to\infty}{\to} \exp(-c)$, donc

$$\mathbb{P}(\mathcal{B}(m_n, \frac{n}{c}) = k) \underset{n\to\infty}{\to} \exp(-c)\frac{c^k}{k!} = \mathbb{P}(\mathcal{P}(c) = k).$$

$\square$

Comme $n-k-z_1-\ldots-z_{k-1} \underset{n\to\infty}{\sim} n$, d'après le lemme 2.1 on a :

$$\mathbb{P}(Z'_k = z_k) \underset{n\to\infty}{\to} \exp(-c)\frac{c^{k-1}}{(k-1)!}.$$

Donc :

$$\lim_{n\to\infty} \mathbb{P}\big((Z'_1,\ldots,Z'_t) = (z_1,\ldots,z_t)\big) = \mathbb{P}\big((Z^*_1,\ldots,Z^*_t) = (z_1,\ldots,z_t)\big).$$

En sommant sur toutes les valeurs possibles de $(z_1,\ldots,z_t)$, la somme étant finie avec la contrainte $z_1+\cdots+z_t = t-1$, on obtient :

$$\lim_{n\to\infty} \mathbb{P}(T \leqslant t) = \mathbb{P}(T^* \leqslant t).$$

2.1.3 Étude de Y^*_t

On a vu qu'on pouvait, dans une certaine mesure, ramener l'étude de Y'_t à celle de Y^*_t.

$Y^*(t)$ est une marche aléatoire de loi de saut $\mathcal{P}(c)-1$, d'espérance $c-1$.

Calcul de $P(T^* < \infty)$

Pour étudier le comportement de $\mathbb{P}(\mathcal{P}(c) = i)$, on introduit les séries génératrices :

$$p(x) := \sum_{i=0}^{\infty} \mathbb{P}(\mathcal{P}(c) = i)x^i.$$

$$q(x) := \sum_{i=0}^{\infty} \mathbb{P}(T^* = i)x^i.$$

On a $p(x) = \exp(c(x-1))$.

On a $\mathbb{P}(T = k \mid Y^*_1 = s) = \sum_{i_1+\ldots+i_s=k-1} \prod_{j=1}^{s} \mathbb{P}(T = i_j)$.

Donc :

$$\begin{aligned}
\mathbb{P}(T = k) &= \sum_{s=0}^{\infty} \mathbb{P}(Y^*_1 = s) \sum_{i_1+\ldots+i_s=k-1} \prod_{j=1}^{s} \mathbb{P}(T = i_j) \\
&= \sum_{s=0}^{\infty} \mathbb{P}(\mathcal{P}(c) = s) \sum_{i_1+\ldots+i_s=k-1} \prod_{j=1}^{s} \mathbb{P}(T = i_j).
\end{aligned}$$

Et donc, en utilisant l'égalité précédente dans la définition de $q(x)$, on obtient :

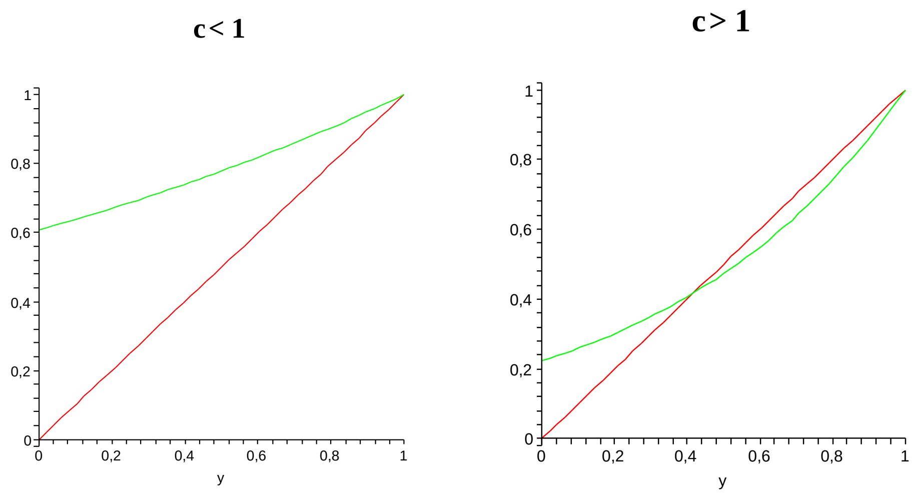


FIG. 2 – $y \mapsto \exp(c(y-1))$ et $y \mapsto y$, pour $c < 1$ et $c > 1$

$$\begin{aligned}
q(x) &= \sum_{k=0}^{\infty} \mathbb{P}(T = k) x^k \\
&= \sum_{k=0}^{\infty} \Big(\sum_{s=0}^{\infty} \mathbb{P}(\mathcal{P}(c) = s) \sum_{i_1+\ldots+i_s=k-1} \prod_{j=1}^{s} \mathbb{P}(T = i_j) \Big) x^k \\
&= x \sum_{s=0}^{\infty} \mathbb{P}(\mathcal{P}(c) = s) \sum_{k=0}^{\infty} (\sum_{i_1+\ldots+i_s=k-1} \prod_{j=1}^{s} \mathbb{P}(T = i_j)) x^{k-1} \\
&= x \sum_{s=0}^{\infty} \mathbb{P}(\mathcal{P}(c) = s) q(x)^s \\
&= x p(q(x)).
\end{aligned}$$

Donc si $y(x) := \frac{q(x)}{x}$, y satisfait l'équation fonctionnelle $y(x) = p(xy(x)) = \exp(c(xy(x) - 1))$. Donc :

$$\mathbb{P}(T^* < \infty) = \sum_{i=0}^{\infty} \mathbb{P}(T^* = i) = \frac{q(1)}{1} = y(1)$$

satisfait l'équation $y = \exp(c(y-1))$.

Par une étude de fonction, on montre facilement que l'équation $y = \exp(c(y-1))$ admet 1 pour unique solution si $c < 1$ et deux solutions y_c et 1 si $c > 1$, avec $0 < y_c < 1$ (cf. figure 2.1.3).

2.1.4 Loi de Y'_t

Calculons la loi de Y'_t, pour s'affranchir des variables aléatoires Z_i, dont la loi dépend de Y_i.

On fixe le sommet initial i, et un autre sommet quelconque v. À chacune des t étapes, le sommet tué a une probabilité $1-p$ de ne pas être rattaché au sommet v. Donc le sommet v à une probabilité $(1-p)^t$ d'être toujours neutre après t étapes.

Comme il y a initialement $n-1$ sommets neutres, n'_t est de loi $\mathcal{B}(n-1,(1-p)^t)$.

Donc Y'_t est de loi $\mathcal{B}(n-1, 1-(1-p)^t) + 1 - t$.

2.2 Le cas $c < 1$

Le but de cette partie est de montrer le résultat suivant :

Théorème 2.2. *Soit $c < 1$. Il existe $\beta_c > 0$ tel que $\mathbb{P}(|\mathcal{C}_{max}| > \beta_c \ln(n)) \underset{n\to\infty}{\longrightarrow} 0$.*

2.2.1 Majoration de T

On commence par majorer la loi de Y'_t par une loi plus simple, qui nous permettra d'appliquer le théorème des grandes déviations.

Soit $t \in \mathbb{N}^*$.

$$\begin{aligned}\mathbb{P}(T > t) &\leqslant \mathbb{P}(Y'_t > 0)\\ &= \mathbb{P}(\mathcal{B}(n-1, 1-(1-p)^t) + 1 - t > 0)\\ &= \mathbb{P}(\mathcal{B}(n-1, 1-(1-p)^t) \geqslant t).\end{aligned}$$

On a $1-(1-p)^t \leqslant tp$, car $(1-p)^t = 1 - tp + \int_0^p (p-u)t(t-1)(1-u)^{t-2}du$, et $(p-u)t(t-1)(1-u)^{t-2} \geqslant 0$ sur $[0,\, p]$.

$n-1 < n$ et $1-(1-p)^t \leqslant tp$, donc $\mathbb{P}(\mathcal{B}(n-1, 1-(1-p)^t) \geqslant t) \leqslant \mathbb{P}(\mathcal{B}(n, tp) \geqslant t)$.

D'après le théorème 1.2 (inégalités des grandes déviations), il existe $\alpha_c > 0$ tel que $\mathbb{P}(\mathcal{B}(n, tp) \geqslant t) < \exp(-\alpha t)$. Intuitivement, cela signifie que la probabilité d'obtenir des grandes valeurs décroît de manière exponentielle.

On a donc $\mathbb{P}(T > t) < \exp(-\alpha t)$.

2.2.2 Majoration de $|\mathcal{C}_{max}|$

On rappelle que si i est le sommet initial dans le parcours en profondeur, T correspond à la taille de $\mathcal{C}(i)$.

Soit $\beta_c > \frac{1}{\alpha_c}$. On pose $t := \beta_c \ln(n)$ dans l'inégalité qui précéde, et on obtient :

$$\mathbb{P}(T > \beta_c \ln(n)) < n^{-\alpha_c \beta_c}.$$

Donc $\mathbb{P}(|\mathcal{C}(i)| > \beta_c \ln(n)) < n^{-\alpha_c\beta_c}$.

Comme il y a n choix possibles pour le premier sommet et $1-\alpha_c\beta_c < 0$, on obtient :

$$\mathbb{P}(|\mathcal{C}_{max}| > \beta_c \ln(n)) < n^{1-\alpha_c\beta_c} = o_{n\to\infty}(1).$$

2.3 Le cas $c > 1$

2.3.1 Grandes déviations

Comme dans le cas $c < 1$, on commence par majorer Y'_t par une loi plus simple, qui nous permettra d'appliquer l'inégalité des grandes déviations.

On rappelle que $\mathbb{P}(Y'_t \leqslant 0) = \mathbb{P}(\mathcal{B}(n-1, 1-(1-p)^t) \leqslant t)$.

Soient $0 < \varepsilon, \delta < 1$, tels que $(1+\delta)(1-y) < 1$.

Commençons par montrer le :

Lemme 2.2. *Il existe N_1 tel que $\forall n \geqslant N_1$, $\mathbb{P}\left(T > (1+\delta)\, n\, (1-y_c)\right) < \frac{\varepsilon}{2}$.*

Démonstration. Soit $\alpha := (1+\delta)(1-y_c)$. On montre facilement que :

$$1 - \left(1 - \frac{c}{n}\right)^{\lfloor \alpha n \rfloor} \underset{n\to\infty}{\to} 1 - \exp(-c\alpha).$$

Donc :

$$\begin{aligned}\mathbb{P}\left(Y_{\lfloor \alpha n\rfloor} \leqslant 0\right) &= \mathbb{P}\left(\mathcal{B}\left(n-1, 1-\left(1-\frac{c}{n}\right)^{\lfloor \alpha n\rfloor}\right) \leqslant \lfloor \alpha n \rfloor - 1\right) \\ &\underset{n\to\infty}{\to} \mathbb{P}(\mathcal{B}(n, 1-\exp(-c\alpha) \leqslant \alpha n).\end{aligned}$$

Donc $\exists N \in \mathbb{N}$ tel que $\forall n \geqslant N$,

$$|\mathbb{P}\big(Y_{\lfloor \alpha n\rfloor} \leqslant 0\big) - \mathbb{P}\big(\mathcal{B}\left(n, 1-\exp(-c\alpha)\right) \leqslant \alpha n\big)| \leqslant \frac{\varepsilon}{4}.$$

On a $1-\exp(-c\alpha) < \alpha$, donc d'après le théorème 1.2, $\exists q > 0$ tel que :

$$\mathbb{P}\big(\mathcal{B}(n, 1-\exp(-c\alpha)) \geqslant \alpha n\big) \leqslant \exp(-qn).$$

Donc $1-\exp(-qn) \leqslant \mathbb{P}\big(\mathcal{B}(n, 1-\exp(-c\alpha)) < \alpha n\big)$. Donc il existe $N_1 \geqslant N$ tel que $\forall n \geqslant N_1$, $1-\frac{\varepsilon}{4} \leqslant 1-\exp(-qn) \leqslant \mathbb{P}(\mathcal{B}(n, 1-\exp(-c\alpha)) < \alpha n)$.

Donc $\forall n \geqslant N_1$,

$$\begin{aligned}1 - \frac{\varepsilon}{2} &< \mathbb{P}(Y_{\lfloor \alpha n\rfloor} \leqslant 0) \\ &\leqslant \mathbb{P}(T \leqslant \lfloor \alpha n \rfloor) \\ &\leqslant \mathbb{P}(T \leqslant \alpha n).\end{aligned}$$

Donc $\forall n \geqslant N_1$, $\mathbb{P}\left(T > (1+\delta)\, n\, (1-y_c)\right) < \frac{\varepsilon}{2}$. $\square$

L'autre inégalité est légèrement plus délicate (et il y a sans doute un moyen bien plus simple de procéder !) :

Lemme 2.3. *Il existe* $N_2, t_0 \in \mathbb{N}$ *tels que* $\forall n \geqslant N_2, \mathbb{P}\big(t_0 < T < (1-\delta)\, n\, (1-y_c)\,\big) < \frac{\varepsilon}{2}$.

Démonstration. Soit $\alpha := (1-\delta)(1-y_c)$.

Pour n assez grand, on minore $\ln(1-\frac{c}{n})$ par $-\frac{c}{n} - \frac{c^2}{n^2}$.

Comme précédemment, on peut montrer qu'il existe $q > 0$ tel que $\mathbb{P}(Y'_t \leqslant 0) \leqslant \exp(-qn)$.

En utilisant le fait que $\mathbb{P}(T = t) \leqslant \mathbb{P}(Y'_t \leqslant 0)$, en sommant sur toutes les valeurs de t, on montre qu'il existe t_0, N_2 tels que $\forall n \geqslant N_2$, $\mathbb{P}(t_0 < T < (1-\varepsilon)\alpha) < \frac{\varepsilon}{2}$. □

En résumé, on a montré qu'il existe $N, t_0 \in \mathbb{N}$ tels que $\forall n \geqslant N$,

$$\mathbb{P}\big(T \in [t_0, (1-\delta)n(1-y)] \cup [(1+\delta)n(1-y), \infty[\big) < \varepsilon.$$

D'après le paragraphe 2.1, on a $y_c = \mathbb{P}(T^* < \infty) = \lim_{t_0 \to \infty} \mathbb{P}(T^* < t_0)$.

Donc $\exists\, t_0$ suffisamment grand tel que $y_c - \varepsilon \leqslant \mathbb{P}(T^* \leqslant t_0) \leqslant y_c$.

Comme $\lim_{n\to\infty} \mathbb{P}(T \leqslant t_0) = \mathbb{P}(T^* \leqslant t_0)$, $\exists\, N_\varepsilon$ tel que $\forall\, n \geqslant N_\varepsilon$:

$$y_c - 2\varepsilon \leqslant \mathbb{P}(T \leqslant t_0) \leqslant y_c + \varepsilon.$$

On a donc $\forall\, n \geqslant N_\varepsilon$:

$$1 - y_c - 2\varepsilon \leqslant \mathbb{P}\big((1-\delta)n(1-y_c) < T < (1+\delta)n(1-y_c)\big) < 1 - y_c - 3\varepsilon.$$

2.3.2 Modification de la procédure

Revenons au processus d'exploration décrit à la partie 1.3.

On dira qu'une composante connexe est "petite" si $|\mathcal{C}(v)| \leqslant t_0$, et qu'elle est "géante" si $(1+\delta)n(1-y_c) \leqslant |\mathcal{C}(v)| \leqslant (1+\delta)n(1-y_c)$ (le cas restant étant d'après ce qui précède un cas pathologique).

$y_c + \varepsilon < 1$, donc $\exists\, s_\varepsilon > 0$ tel que $(y_c+\varepsilon)^{s_\varepsilon} < \varepsilon$.

On modifie légèrement le processus d'exploration décrit à la partie 1.3 ; on arrête la procédure dès que :

- ou bien un composante "géante" a été trouvée,
- ou bien un composante "pathologique" a été trouvée,
- ou bien s "petites" composantes ont été trouvées.

Comme à chaque étape, seules des petites composantes ont été enlevées, le nombre de sommets restants est $m \geqslant n - st_0 = n - O(1)$. Donc les probabilités, après suppression des sommets, d'obtenir une composante "géante", "pathologique" ou "petite" restent asymptotiquement les mêmes.

La probabilité de trouver une composante "pathologique" est inférieure à εs, et la probabilité de n'avoir trouvé que des "petites" composantes est inférieure ou égale à $(y_c + \varepsilon)^s < \varepsilon$, donc la probabilité de trouver une composante "géante" est supérieure ou égale à $1 - \varepsilon'$, où $\varepsilon' = (s+1)\varepsilon$ peut être rendu aussi petit qu'on le souhaite.

En résumé, on a montré que si $c > 1$, $0 < \varepsilon' < 1$, $0 < \delta < 1$, et $0 < y_c < 1$ est tel que $1 - y_c$ soit solution de $1 - \exp(-c\alpha) = \alpha$, alors il existe N tel que $\forall\, n \geqslant N$, la probabilité pour que $\mathcal{C}_{max} \in \big[(1-\delta)n(1-y_c), (1+\delta)n(1-y_c)\big]$ dans un graphe de $\mathcal{G}(n, \frac{c}{n})$ est supérieure à $1 - \varepsilon'$.

3 Le cas critique : $c = 1$

Nous allons montrer que la taille de la plus grande composante connexe de $\mathcal{G}(n, 1/n)$ est de l'ordre de $n^{2/3}$. Plus précisément, nous allons prouver ces deux théorèmes :

Théorème 3.1. *Pour $A < 20$ et $n > 10$ on a :*

$$\mathbb{P}\left(|\mathcal{C}_{max}| \geqslant An^{2/3}\right) \leqslant \frac{3}{A^{3/2}}.$$

Théorème 3.2. *Pour $\delta < \frac{1}{10}$ et $n > 100$ on a :*

$$\mathbb{P}\left(|\mathcal{C}_{max}| \leqslant \delta n^{2/3}\right) \leqslant 12\delta^{3/5}.$$

3.1 Borne supérieure

Pour montrer le théorème 3.1, nous allons prouver qu'on peut majorer ainsi la taille de toutes les composantes connexes de $\mathcal{G}(n, 1/n)$:

Théorème 3.3. *Pour tout $H \in [\![3, n-3]\!]$ on a :*

$$\mathbb{P}\left(|\mathcal{C}(i)| \geqslant H^2\right) \leqslant \frac{3}{H}.$$

3.1.1 Comparaison à une martingale

Pour arriver au théorème 3.3, l'idée est de dominer stochastiquement Y_t par une martingale, ce qui permet d'utiliser le théorème d'arrêt pour évaluer le temps de passage en 0.

On pose donc :

$$\begin{cases} S_0 & = & 1, \\ S_{t+1} & = & S_t + \xi_{i+1} - 1 \quad \forall t \geqslant 0. \end{cases}$$

où $(\xi_t)_{1 \leqslant t \leqslant n}$ sont des variables aléatoires indépendantes, toutes de loi $\mathcal{B}(n, 1/n)$.

Cette définition est à comparer à (1).

Théorème 3.4. *(S_t) est une martingale adaptée à la filtration $(\mathcal{F}_t)_{t \geqslant 0} = \big(\sigma(\xi_1, \ldots, \xi_t)\big)_{t \geqslant 0}$.*

Démonstration. On montre facilement que :

- $\forall\, t \in \mathbb{N}$, S_t est $\mathcal{F}_t$-mesurable,
- $\forall\, t \in \mathbb{N}$, $\mathbb{E}[S_t] < \infty$,
- $\mathbb{E}[S_{t+1} - S_t \,|\, \mathcal{F}_t] = \mathbb{E}[\xi_{t+1}] - 1 = 0.$

□

Théorème 3.5. *Pour tout $t \in [\![0, n]\!]$, S_t domine stochastiquement Y_t.*

Démonstration. La démonstration est facile, si on utilise le théorème 1.3. En effet, η_t est stochastiquement dominée par toute variable aléatoire de loi $\mathcal{B}(n, 1/n)$. On peut donc trouver η'_t de même loi que η_t et ξ'_t de même loi que $\mathcal{B}(n, 1/n)$, vérifiant $\eta'_t < \xi'_t$ presque sûrement.

En posant de nouveau :

$$Y'_t := \begin{cases} Y_{t-1} + \eta_t - 1 & \text{si} \quad Y_{t-1} \geqslant 0, \\ \eta_t & \text{sinon.} \end{cases} \qquad \text{et} \quad S'_t := 1 + \sum_{i=1}^{t} (\xi_i - 1),$$

on a $S'_t \sim S_t$ et $Y'_t \sim Y_t$. Alors, par une récurrence immédiate sur t, on a $Y'_t < S'_t$ p.s., ce qui conclut la preuve. $\square$

On introduit maintenant, pour $H > 0$, le temps d'arrêt suivant :

$$\gamma_H := \inf\{t \geqslant 1 : S_t = 0 \text{ ou } S_t \geqslant H\}.$$

Théorème 3.6. *γ_H est un temps d'arrêt fini presque sûrement.*

Corollaire 3.1. $\mathbb{E}[S_{\gamma_H}] = 1$.

Démonstration. Puisque $S_{\gamma_H \wedge t}$ est une martingale (d'après le théorème 1.1) positive, elle converge presque sûrement vers S_{γ_H}. Le théorème de convergence dominée achève la preuve. $\square$

3.1.2 Évaluation du temps d'arrêt

On va montrer le théorème suivant :

Théorème 3.7. *Pour $H \in [\![3, n-3]\!]$, on a :*

$$\mathbb{P}(\gamma_H \geqslant H^2) \leqslant \frac{2}{H}.$$

On part de l'inégalité de Markov :

$$\mathbb{P}\left(\gamma_H \geqslant H^2\right) \leqslant \frac{\mathbb{E}[\gamma_H]}{H^2}. \tag{2}$$

Pour prouver le théorème 3.7, on cherche donc à majorer $\mathbb{E}[\gamma_H]$.

Le lemme suivant ramène cette majoration à celle de $\mathbb{E}[S^2_{\gamma_H}]$:

Lemme 3.1. $\mathbb{E}[\gamma_H] = \frac{n}{n-1}\big(\mathbb{E}[S^2_{\gamma_H}] - 1\big)$.

Démonstration. $(S_t^2 - (1-\frac{1}{n})t)$ est une martingale, en effet :

- $\forall\, t \in \mathbb{N}$, $S_t^2 - (1-\frac{1}{n})t$ est $\mathcal{F}_t$-mesurable,
- $\forall\, t \in \mathbb{N}$, $\mathbb{E}[|S_t^2 - (1-\frac{1}{n})t|] \leqslant \mathbb{E}[S_t^2] + t < \infty$,
- $\mathbb{E}[S_{t+1}^2 - (1-\frac{1}{n})(t+1) - (S_t^2 - (1-\frac{1}{n})t) \,|\, \mathcal{F}_t]$
  $= \mathbb{E}\big[S_{t+1}^2 - S_t^2 \,|\, \mathcal{F}_t\big] - (1-\frac{1}{n})$
  $= \mathbb{E}\big[(\xi_{t+1}-1)^2 + 2S_t(\xi_{t+1}-1) \,|\, \mathcal{F}_t\big] - (1-\frac{1}{n})$
  $= 0.$

D'après le théorème 1.1, pour tout $t \in \mathbb{N}$, $S_{t\wedge\gamma_H}^2 - (1-\frac{1}{n})(t\wedge\gamma_H)$ est encore une martingale, donc $\mathbb{E}[t\wedge\gamma_H] = \frac{n}{n-1}\big(\mathbb{E}[S_{t\wedge\gamma_H}^2] - 1\big)$. Or, comme $|S_{t\wedge\gamma_H}| < H$, on a :

$$\mathbb{E}[S_{t\wedge\gamma_H}^2] \underset{t\to\infty}{\longrightarrow} \mathbb{E}[S_{\gamma_H}^2].$$

On déduit donc d'abord, par le lemme de Fatou, que γ_H est intégrable, puis le lemme 3.1 par convergence dominée. □

Majoration de $\mathbb{E}[S_{\gamma_H}^2]$

Lemme 3.2. *Conditionnellement à $S_{\gamma_H} \geqslant H$, $S_{\gamma_H} - H$ est stochastiquement dominée par toute variable aléatoire de loi $\mathcal{B}(n, 1/n)$.*

Démonstration. On remarque que, conditionnellement à l'événement $\{S_{\gamma_H} \geqslant H\} \cap \{\gamma_H = l\} \cap \{S_{l-H} = H - r\}$, on a $S_{\gamma_H} - H = \xi_l - r$, où ξ_l est de loi $\mathcal{B}(n, 1/n)$.

Il suffit donc de montrer que si ξ est de loi $\mathcal{B}(n, 1/n)$, $\xi - r$ sachant que $\xi \geqslant r$ est stochastiquement dominée par la distribution $\mathcal{B}(n, 1/n)$, ce qui est évident : si on voit ξ comme une somme de n v.a. i.i.d de loi de Bernoulli de paramètre p, et que la somme des J premières vaut r, $\xi - r$ sachant J est de loi $\mathcal{B}(n - J, 1/n)$. □

Lemme 3.3. $\forall\, H \geqslant 0$, $\mathbb{E}\left[S_{\gamma_H}^2 \,|\, S_{\gamma_H} \geqslant H\right] \leqslant H^2 + 2H + 2.$

Démonstration. La fonction $f\colon x \mapsto 2Hx + x^2$ est croissante sur $\mathbb{R}^+$. On en déduit, d'après le lemme 3.2 et le théorème 1.4, que $\mathbb{E}[f(S_{\gamma_H} - H)] \leqslant \mathbb{E}[f(X)]$. On a donc $\mathbb{E}[S_{\gamma_H}^2 - H^2 \,|\, S_{\gamma_H} \geqslant H] \leqslant 2H + 2 - \frac{1}{n}$, d'où le résultat. □

3.1.3 Fin des preuves

Preuve du théorème 3.7 En utilisant les lemmes 3.3 et 3.1 ainsi que l'inégalité de Markov, on a :

$$\begin{aligned} 1 + \left(1 - \frac{1}{n}\right)\mathbb{E}[\gamma_H] &= \mathbb{E}[S_{\gamma_H}^2] \\ &= \mathbb{P}(S_{\gamma_H} \geqslant H)\mathbb{E}\left[S_{\gamma_H}^2 \,|\, S_{\gamma_H} \geqslant H\right] \\ &\leqslant \frac{\mathbb{E}[S_{\gamma_H}]}{H}(H^2 + 2H + 2). \end{aligned}$$

On en déduit, d'après le corollaire 3.1 :

$$\mathbb{E}[\gamma_H] \leqslant \left(\frac{H^2+H+2}{H}\right)\left(1+\frac{1}{n-1}\right).$$

Pour $H \in [\![3, n-3]\!]$, on a donc $\mathbb{E}\left[\gamma_H\right] \leqslant H+3$, ce qui permet, à partir de (2), de prouver le théorème 3.7.

Fin de la preuve du théorème 3.1 On peut maintenant prouver le théorème 3.3. En effet, on a :

$$\begin{aligned} \mathbb{P}\left(|\mathcal{C}(i)| \geqslant H^2\right) & \leqslant \mathbb{P}\left(Y_{\gamma_H \wedge H^2} > 0\right) \\ & \leqslant \mathbb{P}\left(S_{\gamma_H \wedge H^2} > 0\right) \\ & \leqslant \mathbb{P}\left(S_{\gamma_H} \geqslant H\right) + \mathbb{P}\left(\gamma_H \geqslant H^2\right) \end{aligned}$$

En utilisant l'inégalité de Markov et le théorème 3.7, on obtient bien :

$$\mathbb{P}\left(|\mathcal{C}(i)| \geqslant H^2\right) \leqslant \frac{3}{H}.$$

On peut enfin prouver le théorème 3.1 en remarquant que :

$$\mathbb{P}(|\mathcal{C}_{max}| \geqslant H^2) \leqslant \mathbb{P}(U_H \geqslant H^2), \tag{3}$$

où U_H est le nombre de sommets appartenant à une composante connexe de taille supérieure à H^2.

Or, $\mathbb{E}[U_H] = n\mathbb{P}\left(|\mathcal{C}(i)| \geqslant H^2\right)$, donc en utilisant l'inégalité de Markov, l'inégalité (3) devient :

$$\mathbb{P}(|\mathcal{C}_{max}| \geqslant H^2) \leqslant \frac{n\mathbb{P}\left(|\mathcal{C}(i)| \geqslant H^2\right)}{H^2}.$$

Le théorème 3.1 s'en déduit en posant $H^2 := An^{2/3}$.

3.2 Borne inférieure

Pour prouver le théorème 3.2, on cherche à minorer l'espérance du temps de premier passage de Y_t par 0. Pour cela, on procède en deux étapes : on montre d'abord que la probabilité pour que Y_t dépasse un certain seuil avant une date fixée est relativement grande, puis que Y_t peut rester strictement positive assez longtemps.

3.2.1 Étape 1 : montée

On se donne un seuil $h > 0$ et une date T_1. On pose :

$$\tau_h := \inf\{t \leqslant T_1 \ : \ Y_t \geqslant h\},$$

si cet ensemble est non vide, et $\tau_h := T_1$ sinon.

Nous allons montrer qu'en choisissant bien T_1, on a le résultat suivant :

Théorème 3.8. *Pour $h \in [3, \frac{\sqrt{n}}{4}]$ et $n \geqslant 16$:*

$$\mathbb{P}(\tau_h = T_1) \leqslant \frac{32h^3}{n}.$$

Ici encore, on part de l'inégalité de Markov, et on cherche à majorer $\mathbb{E}[\tau_h]$.

Si on suppose $0 < Y_{t-1} \leqslant h$, on a :

$$\begin{aligned}
\mathbb{E}[Y_t^2 - Y_{t-1}^2 \mid Y_{t-1}] &= \mathbb{E}[(\eta_t - 1)^2 + 2(\eta_t - 1)Y_{t-1} \mid Y_{t-1}] \\
&= \mathbb{E}[(\eta_t - 1)^2 \mid Y_{t-1}] + 2\mathbb{E}[(\eta_t - 1) \mid Y_{t-1}]Y_{t-1} \\
&= \frac{n_{t-1}}{n}(1 - \frac{1}{n}) + 2Y_{t-1}(\frac{n_{t-1}}{n} - 1) \\
&= \frac{n - Y_{t-1} - (t-1)}{n}(1 - \frac{1}{n}) - 2Y_{t-1}(\frac{Y_{t-1} + t - 1}{n}) \\
&\geqslant \frac{n - h - t}{n}(1 - \frac{1}{n}) - 2h\frac{h+t}{n} \\
&\geqslant 1 - \frac{h + t + 1}{n} - 2h\frac{h+t}{n}.
\end{aligned}$$

Si on suppose de plus que $2 \leqslant h < \frac{\sqrt{n}}{4}$ et $t \leqslant T_1 = \frac{n}{8h}$, on a :

$$\begin{aligned}
1 - \frac{h+t+1}{n} - 2h\frac{h+t}{n} &\geqslant 1 - \frac{1}{4\sqrt{n}} - \frac{1}{8h} - \frac{1}{n} - 2(\frac{1}{16} + \frac{1}{8}) \\
&\geqslant \frac{5}{8} - \frac{1}{4\sqrt{n}} - \frac{1}{16} - \frac{1}{n} \\
&\geqslant \frac{1}{2} \quad \text{si } n \geqslant 64.
\end{aligned}$$

Sous ces conditions, on a donc :

$$\mathbb{E}[Y_t^2 - Y_{t-1}^2 \mid Y_{t-1}] \geqslant \frac{1}{2}. \tag{4}$$

De plus (4) est encore vraie pour $Y_{t-1} = 0$. En effet :

$$\mathbb{E}[Y_t^2] = \mathbb{E}[\eta_t^2] = \frac{n_{t-1}}{n}(1 - \frac{1}{n}) + 1 \geqslant 1.$$

Ceci permet de montrer que $Y^2_{t\wedge\tau_h} - \frac{t\wedge\tau_h}{2}$ est une sous-martingale, donc que $\mathbb{E}(Y^2_{t\wedge\tau_h} - \frac{t\wedge\tau_h}{2}) \geqslant 0$. Finalement, en prenant $t := T_1$, on obtient :

$$\mathbb{E}[\tau_h] \leqslant 2\mathbb{E}[Y^2_{\tau_h}]. \tag{5}$$

Lemme 3.4. *Conditionnellement à $Y_{\tau_h} \geqslant h$, $Y_{\tau_h} - h$ est stochastiquement dominée par toute variable aléatoire de loi $\mathcal{B}(n, 1/n)$.*

Démonstration. La preuve est identique à celle du lemme 3.2. □

En utilisant encore le théorème 1.4 appliqué à la fonction $f \colon x \mapsto 2hx + x^2$, on obtient alors :

$$\mathbb{E}[(Y_{\tau_h} - h)^2 + 2h(Y_{\tau_h} - h)] \leqslant 2h + 2 - \frac{1}{n},$$

d'où :

$$\begin{aligned} \mathbb{E}[Y^2_{\tau_h}] &\leqslant h^2 + 2h + 2 - \frac{1}{n} \\ &\leqslant 2h^2, \quad \text{si } h \geqslant 3. \end{aligned} \tag{6}$$

En combinant (6) et (5), on obtient que $\mathbb{P}(\tau_h = T_1) \leqslant \frac{\mathbb{E}[\tau_h]}{T_1} \leqslant \frac{4h^2}{T_1}$, ce qui achève la preuve du théorème 3.8.

3.2.2 Étape 2 : descente

On se donne maintenant une autre date $T_2 > T_1$, et on pose

$$\tau_0 := \inf\{s \ : \ Y_{T_h+s} = 0\},$$

si cet ensemble est non vide, et $\tau_0 := T_2$ sinon.

On suppose dans cette partie que $Y_{\tau_h} \geqslant h$ (évènement noté A_h), et on cherche à majorer $\mathbb{P}(\tau_0 < T_2)$. Plus précisément, on va montrer :

Théorème 3.9. $\mathbb{P}(\tau_0 < T_2 \mid A_h) \leqslant \frac{2T_2}{h^2}$.

Démonstration. On pose $M_t := h - \min(h, Y_{\tau_t+s})$.

Si $M_t > 0$ et $0 < M_{t-1} < h$, alors on a :

$$\begin{aligned} M_t^2 - M_{t-1}^2 &= (h - Y_{\tau_h-t-1} - \eta_{\tau_h-t} - 1)^2 - (h - Y_{\tau_h-t-1})^2 \\ &= (\eta_{\tau_h-t} - 1)^2 - 2(\eta_{\tau_h-t} - 1)M_{t-1}. \end{aligned}$$

En utilisant la preuve de (4), on obtient, pour $2 \leqslant h \leqslant \frac{\sqrt{n}}{4}$ et $t \leqslant T_2 \leqslant \frac{n}{8h}$:

$$
\begin{aligned}
\mathbb{E}[M_t^2 - M_{t-1}^2 \mid Y_{\tau_h+s-1}, \tau_h] &= \frac{\tau h - t - 1}{n}(1 - \frac{1}{2}) + 2(1 - \frac{\tau h - t - 1}{n}) M_{t-1} \\
&\leqslant 1 + \frac{2h(2t+h)}{n} \\
&\leqslant 1 + \frac{1}{2} + \frac{1}{8} \\
&\leqslant 2.
\end{aligned}
$$

Cette inégalité est encore vraie si $M_{t-1} > h$ ou $M_t = 0$.

On en déduit que $\mathbb{E}[M^2_{\tau_0 \wedge t} - 2(\tau_0 \wedge t) \mid A_h]$ est une surmartingale, ce qui prouve :

$$\forall t \quad \mathbb{E}[M^2_{\tau_0 \wedge T_2} \mid A_h] \leqslant 2\mathbb{E}[\tau_0 \wedge T_2 \mid A_h].$$

En posant $t := T_2$, on obtient :

$$
\begin{aligned}
\mathbb{E}[M^2_{\tau_0 \wedge T_2} \mid A_h] &\leqslant 2\mathbb{E}[\tau_0 \wedge T_2 \mid A_h] \\
&\leqslant 2T_2.
\end{aligned}
$$

On en déduit que :

$$
\begin{aligned}
\mathbb{P}(\tau_0 < T_2 \mid A_h) &\leqslant \mathbb{P}(M_{\tau_0 \wedge T_2} \geqslant h \mid A_h) \\
&\leqslant \frac{\mathbb{E}[M_{\tau_0 \wedge T_2} \mid A_h]}{h^2} \\
&\leqslant \frac{2T_2}{h^2}.
\end{aligned}
$$

□

3.2.3 Conclusion

Les théorèmes 3.8 et 3.9 permettent de montrer que :

$$\mathbb{P}(\tau_0 < T_2) \leqslant \mathbb{P}(\tau_h = T_1) + \mathbb{P}(\tau_0 < T_2 \mid A_h) \leqslant \frac{32h^3}{n} + \frac{2T_2}{h^2}. \tag{7}$$

On pose $T_2 := \delta n^{2/3}$ et on choisit h pour optimiser l'inégalité (7) :

$$
\begin{aligned}
f(h) &= \frac{32h^3}{n} + \frac{2\delta n^{2/3}}{h^2} \\
f'(h) &= \frac{96h^2}{n} - \frac{4\delta n^{2/3}}{h^3} \\
&= \frac{4}{nh^3}(24h^5 - \delta n^{5/3}).
\end{aligned}
$$

$f(h)$ est minimale pour $h^* = \frac{\delta^{1/5} n^{1/3}}{24^{1/5}}$.

On a alors :

$$f(h^*) = \delta^{3/5}(\frac{32}{24^{3/5}} + 2 \times 24^{2/5}) < 12\delta^{3/5}.$$

Pour respecter les conditions $h^* \leqslant \frac{\sqrt{n}}{4}$ et $T_2 \leqslant \frac{n}{8h}$, il suffit de supposer $\delta < \frac{1}{10}$ et $n > 100$. (7) devient ainsi :

$$\mathbb{P}(\tau_0 < T_2) \leqslant 12\delta^{3/5}.$$

Or, si $|\mathcal{C}_{max}| < T_2$, alors Y_t ne saurait rester strictement positive entre τ_h et $\tau_h + T_2$, donc nécessairement $\tau_0 < T_2$.

On en déduit que :

$$\mathbb{P}(|\mathcal{C}_{max}| < T_2) \leqslant \mathbb{P}(\tau_0 < T_2) \leq 12\delta^{3/5}.$$

Ce qui achève la preuve du théorème 3.2.

Références

[1] Noga Alon et Joel H. Spencer, *The probabilistic method*, Wiley, 1992.

[2] Bela Bollobás, *Random graphs*, Academic Press, 1985.

[3] Asaf Nachmias et Yuval Peres, *The critical random graph, with martingales*, `arXiv :math.PR/0512201`, 2005.