

Graphe du web.  
Modélisation et loi de puissance.

Elias Perillier

Mai 2024

# Table des matières

<b>1</b>	<b>Motivations</b>	<b>3</b>
<b>2</b>	<b>Définition</b>	<b>3</b>
2.1	Graphes : Définitions et propriétés de base des graphes . . . . .	3
2.2	Graphes aléatoires . . . . .	4
2.3	Loi de puissance et convergence . . . . .	4
<b>3</b>	<b>Modèle pour le graphe du web</b>	<b>5</b>
3.1	Graphe du web . . . . .	5
3.2	Modèle jouet . . . . .	5
3.2.1	Conclusion . . . . .	7
3.3	Modèle LCD (Linearized Chord Diagram) . . . . .	7
3.3.1	Conclusion . . . . .	8

# 1 Motivations

Dans le cadre de ce projet d'étude, nous examinerons les propriétés des graphes aléatoires, en nous concentrant principalement sur ceux dont la distribution des degrés suit une loi de puissance, afin de les adapter à l'étude du graphe du web. En effet, le graphe du web est modélisé comme un graphe où les sommets représentent des pages internet et les arêtes des liens entre ces pages. Pour que ce modèle soit valide, la distribution des degrés des sommets doit suivre une loi de puissance. Nous analyserons donc deux modèles distincts : un premier modèle qui ne respecte pas cette propriété, et un second modèle qui la respecte. Cette analyse nous permettra de mieux comprendre leur conformité avec les caractéristiques observées dans le graphe du web.

## 2 Définition

### 2.1 Graphes : Définitions et propriétés de base des graphes

**Définition 2.1.** Un graphe est un couple  $G = (V, E)$  comprenant deux ensembles :

- $V$  : sommets ;
- $E$  : arêtes, chacune étant associée à un couple  $(u, v)$  ou une paire  $\{u, v\}$

**Remarque 2.2.** On appelle plus souvent l'ensemble des sommets et des arêtes  $V(G)$  et  $E(G)$ .

**Définition 2.3.** Étant donné un sommet  $u$ , définissons son ensemble de voisins  $N(u)$  comme l'ensemble des sommets connectés à  $u$  (aussi appelés voisins de  $u$ ), tandis que son ensemble de non-voisins  $N^c(u)$  est l'ensemble des sommets distincts de  $u$  et non connectés à  $u$ .

**Définition 2.4.** Le degré d'un sommet est le cardinal  $|N(u)|$ , et s'écrit  $\deg_G(u)$  ou simplement  $\deg(u)$ .

**Théorème 2.5.** Premier théorème de la théorie des graphes.

Soit  $G$  un graphe, alors

$$2|E(G)| = \sum_{u \in V(G)} \deg(u).$$

*Démonstration.* Preuve tirée de [1]. Lors de la sommation des degrés de chaque sommet, chaque arête est comptée exactement deux fois.  $\square$

**Définition 2.6.** Soit  $G$  et  $G'$  deux graphes. On dit que  $f$  est un isomorphisme de graphes si et seulement si  $f$  est une bijection entre les sommets de  $G$  et  $G'$  telle que les arêtes soient préservées.

De manière équivalente,  $f$  est un isomorphisme si et seulement si la propriété suivante est vérifiée :

$$\forall u, v \in V(G), \{u, v\} \in E(G) \iff \{f(u), f(v)\} \in E(G').$$

**Exemple 2.7.** Les deux graphes ci-dessous sont isomorphes.



FIGURE 1 – Exemples de graphes isomorphes

**Définition 2.8.** Un graphe orienté  $G = (V, A)$  est un couple formé de  $V$ , un ensemble appelé ensemble de nœuds, et  $A$ , un ensemble appelé ensemble d'arêtes. Les arêtes sont alors nommées arcs.

- Dans un graphe orienté, les arêtes sont appelées arcs et ont une orientation de  $u$  vers  $v$ . On les identifie donc un couple  $(u, v)$ .
- Un graphe mixte comporte à la fois des arêtes et des arcs ; on le définit alors plutôt comme le triplet  $G = (V, E, A)$ .

## 2.2 Graphes aléatoires

**Définition 2.9.** Nous pouvons définir un espace de probabilité sur les graphes d'un ordre donné  $n \geq 1$  comme suit. Fixons un ensemble de sommets  $V(G)$  composé de  $n$  éléments distincts, généralement pris comme  $[n] = \{1, 2, \dots, n\}$ , et fixons  $p \in [0, 1]$ . Définissons l'espace des graphes aléatoires d'ordre  $n$  avec une probabilité d'arête  $p$ , noté  $G(n, p)$ . On a la propriété suivante :

**Remarque 2.10.** Pour  $n \in [n], p \in [0, 1], G \in G(n, p)$ , on a

$$\mathbb{P}(G) = p^{|E(G)|} (1-p)^{\binom{n}{2} - |E(G)|}.$$

*Démonstration.* D'une part, chaque arête est présente avec une probabilité  $p$  et il y a  $|E(G)|$  arêtes. D'autre part, chaque paire de sommets connectés apparaît avec une probabilité de  $1-p$  et il y a en tout  $\binom{n}{2} - |E(G)|$  paires de sommets non connectés. D'où le résultat.  $\square$

**Remarque 2.11.** Pour  $G \in G(n, p)$  et  $x \in V(G)$ ,  $X = \deg(x)$  suit une loi binomiale de paramètres  $(n-1, p)$ .

## 2.3 Loi de puissance et convergence

**Définition 2.12.** Étant donné un graphe non orienté  $G$  et un entier  $k > 0$ , nous définissons  $N_{k,G}$  par

$$N_{k,G} = |\{x \in V(G) : \deg_G(x) = k\}|.$$

Le paramètre  $N_{k,G}$  est le nombre de sommets de degré  $k$  dans  $G$ .

Nous disons que la distribution des degrés de  $G$  suit une loi de puissance si pour chaque degré  $k$ ,

$$\frac{N_{k,G}}{|V(G)|} \sim k^{-\beta},$$

pour une constante réelle fixée  $\beta > 1$ .

**Exemple 2.13.** Soit  $G \in G(n, p)$ ,  $A \in V(G)$

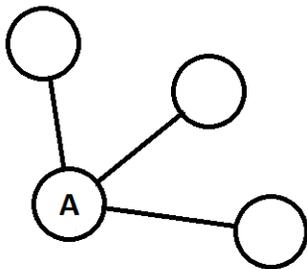


FIGURE 2 – Exemple d'un graphe aléatoire dans  $G(4, p)$ .

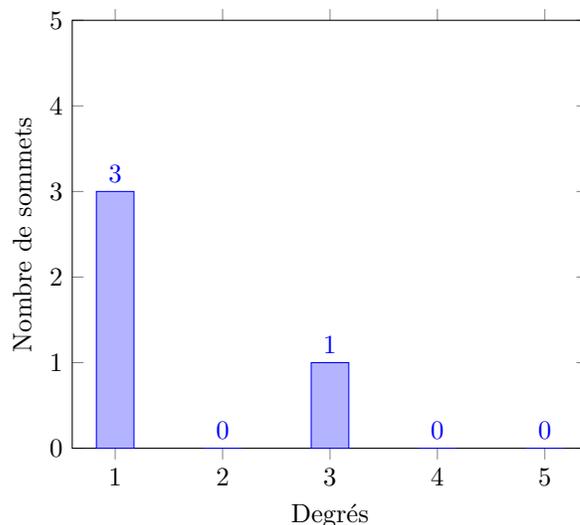


FIGURE 3 – Histogramme du nombre de sommets en fonction des degrés

**Définition 2.14.** soit  $\mathcal{P}$  une propriété de graphe (c'est-à-dire, une propriété préservée par isomorphisme). Nous disons que  $G \in G(n, p)$  satisfait  $\mathcal{P}$  asymptotiquement presque sûrement si

$$\lim_{n \rightarrow \infty} \mathbb{P}(G \in G(n, p) \text{ satisfait } \mathcal{P}) = 1.$$

**Définition 2.15.** Un événement se produit avec une probabilité extrême s'il se produit avec une probabilité d'au moins  $1 - \exp(-\Theta(\log^2 n))$  pour  $\Theta > 0$  et lorsque  $n \rightarrow \infty$ .

**Remarque 2.16.** Dans notre cas d'études, afin d'étudier la distribution des degrés, les graphes aléatoires que nous étudierons seront non orientés. De plus,  $n$  sera strictement supérieur à 0 et  $p$  différents de 0 et 1 afin d'éviter les cas triviaux.

### 3 Modèle pour le graphe du web

#### 3.1 Graphe du web

**Définition 3.1.** Le graphe du web, souvent noté  $W$ , est un graphe où l'ensemble des sommets de  $W$  représente l'ensemble des pages web, et les arêtes représentent les liens entre ces pages. Pour garantir un modèle solide, il doit posséder au moins les trois propriétés suivantes :

1. **"On-line property"** : Les graphes générés par le modèle évoluent avec le temps, de sorte que le nombre de sommets et d'arêtes change au fil du temps.
2. **Distribution du degré selon une loi de puissance** : Asymptotiquement presque sûrement, la distribution des degrés des graphes générés par le modèle suit une loi de puissance avec un exposant  $\beta > 2$ .
3. **"Small world property"** : Le modèle génère presque sûrement des graphes peu denses, c'est à dire des graphes où le nombre d'arêtes est très éloigné du nombre maximal d'arêtes possible

#### 3.2 Modèle jouet

**Lemme 3.2.** Soit  $X$  une variable aléatoire binomiale  $X \sim B(n, p)$  avec  $\mathbb{E}(X) = np$ ,  $t \geq 0$  et  $p \in ]0, 1[$ . On a les résultats suivants :

$$\begin{cases} \mathbb{P}(X \geq \mathbb{E}(X) + t) \leq \exp\left(-\mathbb{E}(X)\phi\left(\frac{t}{\mathbb{E}(X)}\right)\right), \\ \mathbb{P}(X \leq \mathbb{E}(X) - t) \leq \exp\left(-\mathbb{E}(X)\phi\left(\frac{-t}{\mathbb{E}(X)}\right)\right), \end{cases} \quad (1)$$

où  $\phi(x) = (1+x)\log(1+x) - x$  pour  $x \geq -1$  et  $\phi(x) = \infty$  pour  $x < -1$ .  
Si  $\epsilon \leq 1/2$ , alors :

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \epsilon\mathbb{E}(X)) \leq 2 \exp\left(-\frac{1}{3}\epsilon^2\mathbb{E}(X)\right). \quad (2)$$

*Démonstration.* (1) Les inégalités de Chernoff sont admise.

(2) Soit  $X$  une variable aléatoire binomiale  $X \sim B(n, p)$  avec  $\mathbb{E}(X) = np$ ,  $t \geq 0$  et  $p \in ]0, 1[$ ,  $\phi(x) = (1+x)\ln(1+x) - x$  pour  $x \geq -1$  et  $\phi(x) = \infty$  pour  $x < -1$  et  $\epsilon \leq 1/2$ .

$$\begin{aligned} \mathbb{P}(|X - \mathbb{E}(X)| \geq \epsilon\mathbb{E}(X)) &= \mathbb{P}(X - \mathbb{E}(X) \leq -t) + \mathbb{P}(X - \mathbb{E}(X) \geq t), \quad t = \epsilon\mathbb{E}(X) \\ &= \mathbb{P}(X \leq \mathbb{E}(X) - t) + \mathbb{P}(X \geq \mathbb{E}(X) + t) \\ &\leq \exp(-\mathbb{E}(X)\phi(-t/\mathbb{E}(X))) + \exp(-\mathbb{E}(X)\phi(t/\mathbb{E}(X))) \\ &\leq \exp(-\mathbb{E}(X)\phi(-\epsilon)) + \exp(-\mathbb{E}(X)\phi(\epsilon)) \end{aligned}$$

La fonction  $\phi$  est dérivable deux fois pour  $x > -1$  et on a :

$$\phi'(x) = \ln(1+x), \quad \phi''(x) = \frac{1}{1+x}.$$

De plus, on a :

$$\phi(0) = 0, \quad \phi'(0) = 0, \quad \phi''(0) = 1.$$

Comme  $\phi''$  est bornée, on a en utilisant les inégalités de Taylor-Lagrange pour approximer  $\phi(\epsilon)$  autour de  $\epsilon = 0$  :

$$|\phi(\epsilon) - \phi(0) - \epsilon\phi'(0)| = |\phi(\epsilon)| \leq \frac{\epsilon^2}{2} \|\phi''\|_\infty.$$

De même, on a :

$$|\phi(-\epsilon) - \phi(0) + \epsilon\phi'(0)| = |\phi(-\epsilon)| \leq \frac{\epsilon^2}{2} \|\phi''\|_\infty.$$

Soit  $u_\epsilon \in ]-\epsilon, 0[$ ,  $v_\epsilon \in ]0, \epsilon[$  tel que  $\phi(-\epsilon) = \frac{\epsilon^2}{2}\phi''(u_\epsilon)$  et  $\phi(\epsilon) = \frac{\epsilon^2}{2}\phi''(v_\epsilon)$ .

On a :

$$\phi''(u_\epsilon) = \frac{1}{1+u_\epsilon} \geq 1 \geq \frac{2}{3} \quad \text{et} \quad \phi''(v_\epsilon) = \frac{1}{1+v_\epsilon} \geq \frac{2}{3}.$$

Donc :

$$e^{-\mathbb{E}(X)\phi(-\epsilon)} + e^{-\mathbb{E}(X)\phi(\epsilon)} \leq 2e^{-\mathbb{E}(X)\frac{\epsilon^2}{3}}.$$

Finalement, on a bien :

$$\mathbb{P}(|X - \mathbb{E}(X)| > \epsilon\mathbb{E}(X)) \leq 2\exp(-\mathbb{E}(X)\phi(\epsilon)).$$

□

**Théorème 3.3.** Soit  $p \in ]0, 1[$  et  $G \in G(n, p)$ . Alors, pour tout  $v \in V(G)$ , il existe une constante  $c > 0$  et un entier  $N \in \mathbb{N}$  tels que :

$$\mathbb{P}\left(|\deg(v) - p(n-1)| \leq \sqrt{p(n-1)\ln(n)}\right) \geq 1 - e^{-c\ln(n)^2}.$$

*Démonstration.* Soit  $n \geq 1$ ,  $p \in ]0, 1[$ ,  $G \in G(n, p)$  et  $v \in V(G)$ . Soit

$$Y = |\{v \in V(G), |\deg(v) - p(n-1)| > \sqrt{p(n-1)\ln(n)}\}|$$

On a  $\mathbb{P}(Y = 0) = 1 - \mathbb{P}(Y = 1) \geq 1 - \mathbb{E}(Y)$  par Markov.

Il suffit donc de montrer qu'il existe une constante  $c > 0$  tel que  $\mathbb{E}(Y) \leq e^{-c\ln(n)^2}$ .

Soit  $X$  la variable aléatoire égale au degré de  $v$ . Comme elle suit une loi binomiale de paramètre  $(n-1, p)$ , on peut lui appliquer le lemme précédent avec  $\epsilon = \ln n / \sqrt{p(n-1)}$ .

En effet, comme cette suite est décroissante et tend vers 0, il existe un entier  $N \in \mathbb{N}$  tel que pour tout  $n \geq N$ , on a  $\epsilon < \frac{1}{2}$ .

On a donc :

$$\mathbb{P}\left(|\deg(v) - p(n-1)| \leq \sqrt{p(n-1)\ln(n)}\right) \leq 2\exp\left(-\frac{\ln^2 n}{3}\right).$$

Finalement, comme  $Y = \sum_{i=1}^n \mathbf{1}_{\{|\deg(v_i) - p(n-1)| \geq \sqrt{p(n-1)\ln(n)}\}}$ , nous avons :

$$\begin{aligned} \mathbb{E}[Y] &= \sum_{i=1}^n \mathbb{E}[\mathbf{1}_{\{|\deg(v_i) - p(n-1)| \geq \sqrt{p(n-1)\ln(n)}\}}] \\ &= \sum_{i=1}^n \mathbb{P}(|\deg(v_i) - p(n-1)| \geq \sqrt{p(n-1)\ln(n)}) \\ &\leq n\mathbb{P}\left(|\deg(v_1) - p(n-1)| \geq \sqrt{p(n-1)\ln(n)}\right) \\ &\leq 2n\exp\left(-\frac{\ln(n)^2}{3}\right) \\ &\leq 2n\exp\left(-\frac{\ln(n)^2}{6}\right)\exp\left(-\frac{\ln(n)^2}{6}\right) \end{aligned}$$

Comme  $2n\exp\left(-\frac{\ln(n)^2}{6}\right)$  tend vers 0, il existe un rang à partir duquel  $2n\exp\left(-\frac{\ln(n)^2}{6}\right) \leq 1$  et donc à partir duquel :

$$\mathbb{E}[Y] \leq \exp\left(-\frac{\ln(n)^2}{6}\right)$$

□

### 3.2.1 Conclusion

Nous avons donc montré que la probabilité que le degrés diffère de  $p(n-1) \pm \sqrt{p(n-1)\ln(n)}$  est extrêmement faible. Ainsi, pour tout  $v$  dans  $V(G)$ ,

$$\deg(v) = p(n-1) + O(\sqrt{p(n-1)\ln(n)}) = (1 + o(1))p(n-1).$$

Ainsi la distribution des degrés d'un graphe aléatoire  $G \in G(n, p)$  ne peut pas suivre une loi de puissance. Ce modèle n'est donc pas intéressant pour étudier le graph du Web.

### 3.3 Modèle LCD (Linearized Chord Diagram)

**Définition 3.4.** Soit  $G_t$  un graphe à l'étape  $t$  dans le modèle LCD, défini comme suit :

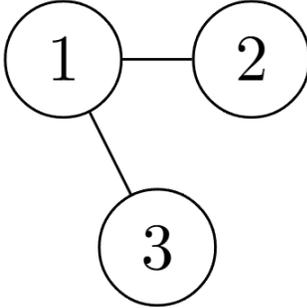
- Formation de  $G_t$  à partir de  $G_{t-1}$  :
  - On ajoute une arête entre le nouveau sommet  $v_t$  et un sommet  $v_i$  choisi aléatoirement parmi les sommets existants de  $G_{t-1}$ .
- Probabilité de choisir  $v_i$  :

$$\mathbb{P}(i = s) = \begin{cases} \frac{\deg_{G_{t-1}}(v_s)}{2^{t-1}} & \text{si } 1 \leq s \leq t-1, \\ \frac{1}{2^{t-1}} & \text{si } s = t. \end{cases}$$

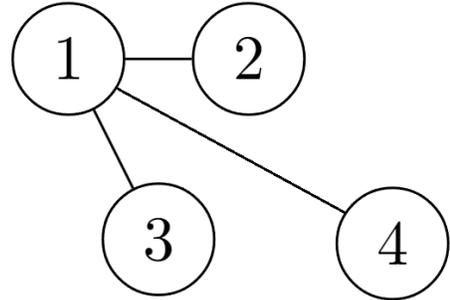
Ainsi, la probabilité que le nouveau sommet soit relié à un sommet existant est plus élevée si ce dernier a un degré élevé.

**Exemple 3.5.** Simulation de deux étapes du modèle LCD. Après calcul, on a :

- La probabilité que le sommet 4 soit relié au sommet 1 est de  $\frac{2}{5}$ .
- La probabilité que le sommet 4 soit relié au sommet 2 est de  $\frac{1}{5}$ .
- La probabilité que le sommet 4 soit relié au sommet 3 est de  $\frac{1}{5}$ .
- La probabilité que le sommet 4 soit relié à lui-même est de  $\frac{1}{5}$ .



(a) Étape 1 : Graphe initial à 3 sommets.



(b) Étape 2 : Ajout du sommet 4.

FIGURE 4 – Simulation des étapes du modèle LCD

**Théorème 3.6.** Dans le modèle LCD, fixons  $m$  un entier positif, et fixons  $\varepsilon > 0$ . Pour  $k$  un entier strictement positif, définissons

$$\alpha_{m,k} = \frac{2m(m+1)}{(k+m)(k+m+1)(k+m+2)}.$$

Alors, asymptotiquement presque sûrement, pour tous les  $k$  satisfaisant  $0 \leq k \leq t^{1/15}$ ,

$$(1 - \varepsilon)\alpha_{m,k} \leq \frac{N_{k,t}}{t} \leq (1 + \varepsilon)\alpha_{m,k}.$$

**Remarque 3.7.** Ce théorème nous dit que pour des  $t$  grands, avec une grande probabilité, la distribution des degrés des graphes dans le modèle LCD suit une loi de puissance avec un exposant  $\beta = 3$ .

### 3.3.1 Conclusion

Contrairement aux graphes aléatoires quelconques, ce modèle semble respecter les trois propriétés nécessaires à la construction d'un modèle pouvant servir à étudier le graphe du web.

## Références

- [1] Graph Theory Book, *Graph Theory : An Introductory Course*. Page 123, 2023.