

Graphes aléatoires

Mesbahi Maroua Raphael Poix

23 mars 2017

Sommaire

1	Introduction :	2
2	Graphes aléatoires avec une loi de degrés arbitraire	3
2.1	Fonctions génératrices	3
2.1.1	Loi du degré	3
2.1.2	Degré moyen d'un sommet	3
2.1.3	Puissance	4
2.1.4	Distance moyenne	4
2.2	Exemples	5
2.2.1	Loi de Poisson	5
2.2.2	Loi exponentielle	5
2.2.3	Loi puissance	5
2.3	Graphe à degrés spécifiques	6
2.4	Composantes connexes	7
2.4.1	Théorie	7
2.4.2	Applications	9
2.5	Nombre de voisins, distance moyenne entre deux sommets	9
2.5.1	Théorie	9
2.5.2	Applications	12
3	Simulations et résultats	13
3.1	Algorithmes utilisés	13
3.1.1	Algorithme de Molloy-Reed	13
3.1.2	Algorithme de Havel-Hakimi	13
3.1.3	Parcours du graphe	14
3.2	Résultats	14
4	Conclusion	17

1 Introduction :

Un graphe aléatoire est un graphe généré par un processus aléatoire. Le premier modèle a été popularisé par Paul Erdos et Alfred Renyi dans une série d'articles publiés entre 1953 et 1968. La théorie qu'ils ont développée était essentiellement basée sur le fait que la présence ou l'absence d'une arête entre deux sommets était indépendante de la présence ou l'absence d'une autre arête. De ce fait :

- On considère que la probabilité qu'une arête existe est de valeur p .
- On peut donc déduire, qu'un graphe $G = (V, A)$ où $Card(V) = N$ et où chaque sommet v est connecté en moyenne à z arêtes, que $p = z/(N - 1)$. Pour N très grand, p est égale à z/N .
- La probabilité pour qu'un sommet soit de degré k est donnée par : $p_k = \binom{N}{k} p^k (1 - p)^{N-k} \approx \frac{z^k e^{-z}}{k!}$
- Pour N assez grand p_k est donc égal au degré d'un sommet d'un graphe aléatoire qui suit une loi de Poisson.

Démonstration. En effet, en prenant $p_k = \binom{N}{k} p^k (1 - p)^{N-k}$ et en posant $p = \lambda/N$ on obtient :

$$\frac{\lambda^k}{k!} * \frac{N(N-1)\dots(N-k+1)}{N^k} * \left(1 - \frac{\lambda}{N}\right)^N * \left(1 - \frac{\lambda}{N}\right)^{-k},$$

avec :

1. $\frac{\lambda^k}{k!}$ est constant,
2. $\lim_{N \rightarrow +\infty} \frac{N(N-1)\dots(N-k+1)}{N^k} = 1$,
3. $\lim_{N \rightarrow +\infty} \left(1 - \frac{\lambda}{N}\right)^N = e^{-\lambda}$,
4. $\lim_{N \rightarrow +\infty} \left(1 - \frac{\lambda}{N}\right)^{-k} = 1$.

Le tout tend bien pour N très grand, vers une loi de poisson

$$\frac{\lambda^k}{k!} e^{-\lambda}$$

□

Les graphes aléatoires sont utilisés dans plusieurs domaines. Ils servent à modéliser certains problèmes tels que l'épidémiologie, les relations d'amitié dans certaines communautés ou les appels téléphoniques. Néanmoins, dans plusieurs de ces domaines, la distribution des degrés ne suit pas une loi de Poisson (elle peut être radicalement opposée).

Contre-exemple :

- Internet : Les recherches empiriques, basées sur l'observation et l'expérimentation ont prouvé que la loi du degré du sommet est asymétrique négative (c'est-à-dire que la moyenne se déplace plus rapidement que la médiane). Mais aussi, internet est un graphe orienté (un lien entre deux pages n'existe que dans un seul sens) ce détail est essentiel pour étudier l'accès d'une page à partir d'une autre.
- Réseaux de collaborations : se modélisent avec un graphe biparti qui ne suit pas une loi de Poisson.

Le but est donc de généraliser l'étude des graphes aléatoires aux degrés qui ne suivent pas une loi de Poisson.

2 Graphes aléatoires avec une loi de degrés arbitraire

Étude de graphes uni-partis, non orientés, avec une loi sur les degrés arbitraire. Comment ces graphes sont-ils générés ? On distingue deux cas :

1. Il s'agit de se donner une loi de probabilité pour les degrés et de considérer que les graphes sont par ailleurs totalement aléatoires. Ainsi on se donne une suite de degrés dont les degrés sont choisis pour approximer au mieux la loi voulue. Enfin, on tire aléatoirement un graphe possédant une telle suite de degrés.
2. En ce qui concerne les graphes de très grande taille, il s'agit d'étudier une unique suite de degrés, le but étant de se rapprocher le plus possible de la loi de probabilité souhaitée et ce en faisant la moyenne uniforme sur tous les graphes possédant une telle suite.

2.1 Fonctions génératrices

On définit la fonction génératrice du nombre de "premiers voisins" :

$$G_0(x) = \sum_{k=0}^{\infty} p_k x^k,$$

avec p_k la probabilité qu'un sommet choisi au hasard soit de degré k . G_0 est normalisée telle que $G_0(1) = 1$. Cette propriété ainsi que celles qui suivent seront vraies pour la majorité des fonctions génératrices que nous allons traiter. Cette modélisation permet de calculer de façon simple certaines quantités :

2.1.1 Loi du degré

p_k est obtenue à la k^{ieme} dérivée de G_0 car :

$$G_0^{(k)}(t) = \sum_{i \geq k} \frac{i!}{(i-k)!} p_i t^{i-k},$$

donc :

$$G_0^{(k)}(0) = k! p_k,$$

alors :

$$p_k = G_0^{(k)}(0) \frac{1}{k!}.$$

2.1.2 Degré moyen d'un sommet

$$z = \langle k \rangle = \sum_k k p_k$$

Démonstration. On sait que $E[X] = \sum_k k p_k$, et on a que

$$G_0'(1) = \sum_k k p_k 1^{k-1},$$

donc :

$$G_0'(1) = \sum_k k p_k = E[X].$$

□

2.1.3 Puissance

Si la loi de distribution d'une propriété X sur un objet est donnée par une certaine fonction génératrice, alors la loi de distribution sur la somme de X sur m éléments indépendants est donnée par la m^{eme} puissance de cette fonction génératrice. C'est-à-dire que la fonction génératrice d'une somme de variables aléatoires indépendantes est le produit des fonctions génératrices (résultat très classique). Comme ici tous les sommets ont la même fonction génératrice pour leur degré, on peut élever cette fonction à des puissances supérieures pour observer les probabilités sur plusieurs sommets. On a :

$$[G_0(x)]^2 = \left[\sum_k p_k x^k \right]^2 = \sum_{jk} p_j p_k x^{j+k},$$

Il est clair que le coefficient devant x^n est bien la somme de tous les $p_j p_k$ tels que $j + k = n$. Nous allons le démontrer de façon plus générale.

Démonstration. En faisant la somme de n variables aléatoires (dans ce cas identiques) indépendante, on a que :

$$[G_{X_1+\dots+X_n}(s)] = E[s^{X_1+\dots+X_n}] = E \left[\prod_{k=1}^n s^{X_k} \right] = \prod_{k=1}^n E[s^{X_k}] = [G_X(s)]^n.$$

□

2.1.4 Distance moyenne

Si on part d'un sommet aléatoire donné n , on s'intéresse à la distribution des degrés pour les premiers voisins atteints.

- Une arête quelconque prise au hasard atteint un nœud n de degré k avec une probabilité proportionnelle à $k p_k$ car il y a k façon d'atteindre un tel sommet. Si on normalise la distribution pour que sa somme vaille 1, on obtient la série génératrice souhaitée :

$$\frac{\sum_k k p_k x^k}{\sum_k k p_k} = x \frac{G'_0(x)}{G'_0(1)}.$$

- Si on considère un sommet choisi au hasard, si l'on suit ses arcs sortant pour explorer le voisinage immédiat, la distribution des degrés de chaque voisin direct est définie par l'équation précédente. On divise par x pour ne pas compter le sommet par lequel on arrive, on obtient la fonction génératrice des premiers voisins d'un sommet :

$$G_1(x) = \frac{G'_0(x)}{G'_0(1)} = \frac{1}{z} G'_0(x),$$

car

$$G'_0(1) = z.$$

- Puisque $G_1(x)^k$ est la fonction génératrice de la somme du nombre de nœuds sortants de k voisin (par la proposition de la puissance). On obtient la série génératrice des voisins à distance 2 :

$$\sum_k p_k G_1^k = G_0(G_1(x)),$$

par définition de la fonction génératrice.

- En itérant le même principe, on obtient le nombre moyen de voisins à distance 3 donné par $G_0(G_1(G_1(x)))$ et ainsi de suite.
- On peut donc calculer l'espérance du nombre de voisins à distance 2 :

$$z_2 = \frac{d}{dx} G_0(G_1(1)) = G'_0(1)G'_1(1) = G''_0(1).$$

2.2 Exemples

2.2.1 Loi de Poisson

C'est l'exemple le plus simple pour ces types de graphes. La distribution des degrés suit une loi binomiale, qui tend vers une loi de Poisson si le nombre de sommets est très grand. Dans ces graphes, la probabilité d'existence ou non d'une arête entre deux sommets choisis aléatoirement est $p = z/N$. $G_0(x)$ est donné par la formule suivante :

$$\sum_{k=0}^N \binom{N}{k} p^k (1-p)^{N-k} x^k = (1-p+px)^N \xrightarrow{N \rightarrow +\infty} e^{z(x-1)}.$$

On remarque que $G'_0(x) = ze^{z(x-1)}$, d'où $G'_0(1) = z$. Et donc :

$$G_1(x) = \frac{G'_0(x)}{z} = \frac{ze^{z(x-1)}}{z} = e^{z(x-1)} = G_0(x).$$

Cette propriété spécifique aux graphes suivant une loi de Poisson pour la distribution des degrés, rend l'étude de ceux-ci simple.

2.2.2 Loi exponentielle

La distribution des degrés suit ici une loi exponentielle, c'est-à-dire $p_k = (1 - e^{-1/c})e^{-k/c}$, avec c constante. La fonction génératrice est alors :

$$G_0(x) = \sum_{k=0}^{+\infty} (1 - e^{-1/c})e^{-k/c} x^k = (1 - e^{-1/c}) \sum_{k=0}^{\infty} e^{-k/c} x^k = (1 - e^{-1/c}) \frac{1}{1 - xe^{-1/c}} = \frac{(1 - e^{-1/c})}{1 - xe^{-1/c}},$$

et ainsi :

$$G_1(x) = \frac{G'_0(x)}{G'_0(1)} = (1 - e^{-1/c}) \frac{e^{-1/c}}{(1 - xe^{-1/c})^2} \frac{(1 - e^{-1/c})^2}{(1 - e^{-1/c})(e^{-1/c})} = \left(\frac{1 - e^{-1/c}}{1 - xe^{-1/c}} \right)^2.$$

2.2.3 Loi puissance

Certains réseaux, comme Internet peuvent être modélisés par ce que l'on appelle réseau invariant d'échelle, où les degrés du graphe suivent une loi de puissance. La loi des degrés est la suivante :

$$p_k = Ck^{-\tau} e^{-k/\kappa},$$

pour $k \geq 0$ où C , τ et κ sont des constantes. Pour des raisons de normalisation, on pose $C = [Li_{\tau}(e^{(-1/\kappa)})]^{-1}$. Ainsi :

$$p_k = \frac{k^{-\tau} e^{-k/\kappa}}{Li_{\tau}(e^{(-1/\kappa)})},$$

pour $k \geq 0$ où $Li_n(x)$ est la fonction polylogarithme. Désormais, on peut calculer certaines propriétés. La fonction génératrice est donnée par :

$$G_0(x) = \sum_k x^k \frac{k^{-\tau} e^{-k/\kappa}}{Li_\tau(e^{(-1/\kappa)})}.$$

Ainsi :

$$G_0(x) = \frac{1}{Li_\tau(e^{(-1/\kappa)})} \sum_k \frac{(xe^{-1/\kappa})^k}{k^\tau}.$$

Par définition de $Li_\tau(x)$ qui est égal à $\sum_k \frac{x^k}{k^\tau}$ on obtient :

$$G_0(x) = \frac{Li_\tau(xe^{-1/\kappa})}{Li_\tau(e^{-1/\kappa})}.$$

Lorsque κ tend vers l'infini, la fonction génératrice se simplifie en :

$$G_0(x) = \frac{Li_\tau(x)}{\zeta(\tau)},$$

où $\zeta(\tau)$ est la fonction zêta de Riemann définie par $\sum_k \frac{1}{n^\tau}$. D'autre part, le nombre moyen de voisins d'un sommet aléatoirement choisi est égal à :

$$z = G'_0(1) = \frac{Li_{\tau-1}(e^{-1/\kappa})}{Li_\tau(e^{-1/\kappa})},$$

car $G'_0(x) = \frac{1}{Li_\tau(e^{-1/\kappa})} (Li_\tau(xe^{-1/\kappa}))'$ et $Li'_\tau(x) = \frac{Li_{\tau-1}(x)}{x}$. Ainsi on peut déduire que le nombre moyen de voisins de distance 1 d'un sommet aléatoirement choisi est donnée par :

$$z_2 = G''_0(1) = \frac{Li_{\tau-2}(e^{-1/\kappa}) - Li_{\tau-1}(e^{-1/\kappa})}{Li_\tau(e^{-1/\kappa})}.$$

Enfin, la fonction génératrice du degré de voisins de distance 1 :

$$G_1(x) = \frac{G'_0(x)}{G'_0(1)} = \frac{Li_{\tau-1}(xe^{-1/\kappa})}{x Li_{\tau-1}(e^{-1/\kappa})}.$$

2.3 Graphe à degrés spécifiques

Dans quelques cas, on veut juste reproduire un modèle d'un graphe que l'on connaît déjà, que l'on a pu observer dans la réalité (par exemple, ceux décrits en introduction). Dans ces cas là, on connaît le nombre de sommets n_k étant de degré k . Ainsi on peut écrire la fonction génératrice (qui est un polynôme fini) comme ceci :

$$G_0(x) = \frac{\sum_k n_k x^k}{\sum_k n_k}.$$

Le dénominateur sert à normaliser, c'est-à-dire qu'on considère $p_k = \frac{n_k}{\sum_k n_k}$.

2.4 Composantes connexes

2.4.1 Théorie

On pose $H_1(x)$, la fonction génératrice de la taille de la composante connexe qu'on obtient après avoir suivi un arc choisi au hasard. On exclut le cas où $H_1(x)$ est une composante géante. La probabilité d'avoir un cycle étant en $1/N$, on peut la négliger pour N grand. Grâce à la propriété de la puissance on remarque que la fonction génératrice de la taille de la composante formée de k sous composantes distinctes s'exprime comme la puissance k_{ieme} de $H_1(x)$. Donc $H_1(x)$ doit vérifier l'équation suivante (avec q_i la probabilité que le sommet initial ait i arêtes sortantes) :

$$H_1(x) = xq_0 + xq_1H_1(x) + xq_2[H_1(x)]^2 + \dots$$

d'où on obtient :

$$H_1(x) = xG_1(H_1(x)).$$

En partant d'un sommet plutôt que d'une arête, il vient :

$$H_0(x) = xG_0(H_1(x)).$$

On veut résoudre l'équation précédente afin d'obtenir $H_0(x)$, à partir de $G_0(x)$ et $G_1(x)$ par substitution. Pour ensuite avoir la probabilité qu'un sommet soit dans une composante de taille s en faisant la s_{ieme} dérivée. Cependant cet équation est difficile à résoudre en théorie. Néanmoins, on peut déduire que :

$$\langle s \rangle = H'_0(1) = 1 + G'_0(1)H'_1(1),$$

et

$$H'_1(1) = 1 + G'_1(1)H'_1(1),$$

ainsi on obtient :

$$\langle s \rangle = 1 + \frac{G'_0(1)}{1 - G'_1(1)}.$$

Démonstration.

$$\langle s \rangle = H'_0(1),$$

et :

$$H'_1(1) = \frac{-1}{G'_1(1) - 1}.$$

Sachant que $G'_0(1) = \langle z \rangle$ et en notant $z_1 = z$ le nombre moyen de voisins ainsi que z_2 le nombre moyen de voisins de distance 2 on obtient :

$$\langle s \rangle = 1 + \frac{z_1^2}{z_1 - z_2}.$$

On remarque que $\langle s \rangle$ diverge pour $G'_1(1) = 1$. C'est le moment de la transition i.e. où la première composante géante apparaît. Cette condition peut aussi se traduire par :

$$\sum_k k(k-2)p_k = 0,$$

car on a :

$$G'_1 - 1 = 0,$$

et on sait que :

$$G_1(x) = \frac{G'_0(x)}{G'_0(1)} = \frac{1}{z} G'_0(x),$$

donc :

$$G'_1 = \frac{1}{z} G''_0(x).$$

On sait aussi que :

$$G_0(x) = \sum_k p_k x^k,$$

donc :

$$G''_0(x) = \sum_k p_k k(k-1) x^{k-2}.$$

On obtient donc :

$$\begin{aligned} G'_1 - 1 &= \frac{1}{z} G''_0(1) - 1 \\ \iff G'_1 - 1 &= \frac{1}{z} \sum_k p_k k(k-1) - 1 \\ \iff G'_1 - 1 &= \frac{1}{z} (\sum_k p_k k(k-1) - p_k k) \\ \iff G'_1 - 1 &= \frac{1}{z} \sum_k p_k k(k-2) = 0 \end{aligned}$$

D'où le résultat. □

Désormais, étant donné que cette somme est monotone et croissante lorsqu'on rajoute des arêtes au graphe, on en déduit que la composante géante existe si et seulement si la somme est positive. Toutes les propriétés sur la fonction génératrice s'appliquent quand il y a une composante géante. Mais, par définition, $H_0(x)$ génère la probabilité de distribution de la taille des composantes en excluant la composante géante. Donc, $H_0(1) \neq 1$. On obtient plutôt : $H_0(1) = 1 - S$ où S est la partie du graphe occupée par la composante géante. On peut utiliser cette propriété afin de calculer la taille de cette composante :

$$\begin{aligned} \iff S &= 1 - H_0(1) \\ \iff S &= 1 - G_0(H_1(1)) \\ \iff S &= 1 - G_0(u) \end{aligned}$$

où $u = H_1(1)$ qui est la plus petite solution non négative de $u = G_1(u)$. Étant donné que $H_0(1) \neq 1$ on normalise l'espérance en divisant par $H_0(1)$. On peut donc retrouver la taille moyenne d'une composante (finie) :

$$\langle s \rangle = \frac{H'_0(1)}{H_0(1)}.$$

On sait que :

$$H'_0(1) = G_0(H_1(1)) + H'_1(1)G'_0(H_1(1)).$$

Et :

$$H'_1(1) = G_1(H_1(1)) + H'_1(1)G'_1(H_1(1)).$$

Donc :

$$H'_1(1) = \frac{G_1(H_1(1))}{1 - G'_1(H_1(1))}.$$

Ainsi :

$$H'_0(1) = G_0(H_1(1)) + \frac{G'_0(H_1(1))G_1(H_1(1))}{1 - G'_1(H_1(1))}.$$

Donc :

$$\begin{aligned} \langle s \rangle &= \frac{1}{H_0(1)} [G_0(H_1(1)) + \frac{G'_0(H_1(1))G_1(H_1(1))}{1 - G'_1(H_1(1))}], \\ &\iff \langle s \rangle = 1 + \frac{zu^2}{[1 - S][1 - G'_1(u)]}, \end{aligned}$$

qui est équivalente à l'équation de $\langle s \rangle$ retrouvée auparavant en posant $S=0$, $u=1$.

2.4.2 Applications

1. Dans un graphe aléatoire avec une loi de Poisson pour la distribution des degrés, on a déjà vu que $G_0(x) = G_1(x) = e^{z(x-1)}$, et on trouve donc que $1 - S = u$ est solution de $u = G_0(u)$ et de façon équivalente, $S = 1 - e^{-zS}$. La taille moyenne de la composante géante est donnée par :

$$\langle s \rangle = \frac{1}{1 - z + zS}.$$

Ces deux résultats sont bien connus.

2. Pour les graphes dont les degrés suivent une loi puissance pure (κ qui tend vers l'infini) et dont la fonction génératrice est donnée par $G_0(x) = \frac{Li_\tau(x)}{\zeta(\tau)}$ on a :

$$u = G_1(u) = \frac{Li_{\tau-1}(u)}{u\zeta(\tau-1)}.$$

Ainsi :

- Lorsque $\tau \leq 2$ on obtient $u = 0$ et ainsi $S = 1$ (car dans ce cas $\zeta(\tau-1)$ n'est pas définie). Ceci implique que la composante géante contient n'importe quel sommet choisi aléatoirement avec une probabilité qui tend vers 1 lorsque κ tend vers l'infini.
- D'autre part, lorsque $\tau > 2$, la probabilité qu'un sommet appartienne à la composante géante est strictement inférieure à 1 car $u > 0$.

2.5 Nombre de voisins, distance moyenne entre deux sommets

2.5.1 Théorie

On s'intéresse maintenant au nombre de voisins de degré m , c'est-à-dire ceux qui sont à distance m d'un sommet choisi aléatoirement. Nous avons déjà parlé de la probabilité de distribution des

premiers et seconds voisins, générée par $G_0(x)$ et $G_0(G_1(x))$. En étendant le même principe, on trouve que la distribution des m^{ieme} voisins est générée par :

$$G_0(\underbrace{G_1(\dots G_1(x))}_{m-1 \text{ fois}}).$$

On l'écrit de la façon suivante :

$$G^{(m)}(x) = \begin{cases} G_0(x) & \text{si } m=1 \\ G^{(m-1)}(G_1(x)) & \text{si } m \geq 2 \end{cases}$$

On trouve le nombre moyen de m^{ieme} voisins avec la dérivée de la fonction ci-dessus en $x = 1$:

$$z_m = \frac{dG^{(m)}}{dx}(1) = G'_1(1)G'^{(m-1)}(1) = G'_1(1)z_{m-1}.$$

En sachant que $z_1 = z = G'_0(1)$, on peut trouver une expression plus pratique pour z_m par induction :

$$z_m = G'_1(1)^{m-1} z_1 = \left[\frac{z_2}{z_1}\right]^{m-1} z_1.$$

A partir de ce résultat, on va pouvoir estimer la longueur l du plus court chemin entre deux sommets choisis aléatoirement sur le graphe. Ce résultat est atteint approximativement quand le nombre total de voisins d'un sommet atteignable avec cette distance est égal au nombre de sommets du graphe. C'est-à-dire quand :

$$1 + \sum_{m=1}^l z_m = N$$

On peut donc estimer l :

$$\begin{aligned} \sum_{m=1}^l z_m &= N - 1 \\ \iff z_1 \frac{1 - (\frac{z_2}{z_1})^l}{1 - \frac{z_2}{z_1}} &= N - 1 \\ \iff z_1 (1 - (\frac{z_2}{z_1})^l) &= (N - 1)(1 - \frac{z_2}{z_1}) \\ \iff -(\frac{z_2}{z_1})^l &= \frac{(N - 1)(1 - \frac{z_2}{z_1}) - z_1}{z_1} \\ \iff (\frac{z_2}{z_1})^l &= \frac{(N - 1)(z_2 - z_1) + z_1^2}{z_1^2} \\ \iff l &= \frac{\ln((N - 1)(z_2 - z_1) + z_1^2) - \ln(z_1^2)}{\ln(\frac{z_2}{z_1})}. \end{aligned}$$

Et ainsi, dans les cas courants où $z_1 \ll N$ et $z_1 \ll z_2$, on peut réduire ce résultat à :

$$l = \frac{\ln(N/z_1)}{\ln(z_2/z_1)} + 1.$$

Démonstration.

$$l = \frac{\ln((N-1)(z_2 - z_1) + z_1^2) - \ln(z_1^2)}{\ln(\frac{z_2}{z_1})}.$$

Comme N est très grand, on peut approximer par :

$$\begin{aligned} & \frac{\ln((N)(z_2 - z_1) + z_1^2) - 2\ln(z_1)}{\ln(\frac{z_2}{z_1})}, \\ &= \frac{\ln(Nz_2 - Nz_1 + z_1^2) - 2\ln(z_1)}{\ln(\frac{z_2}{z_1})}, \\ &= \frac{\ln(N\frac{z_2}{z_1} - N + z_1) - \ln(z_1)}{\ln(\frac{z_2}{z_1})}, \\ &= \frac{\ln(N(\frac{z_2}{z_1} - 1) + z_1) - \ln(z_1)}{\ln(\frac{z_2}{z_1})}. \end{aligned}$$

Et comme $z_1 \ll N$ et $z_1 \ll z_2$, on a donc :

$$\begin{aligned} & \frac{\ln(N(\frac{z_2}{z_1})) - \ln(z_1)}{\ln(\frac{z_2}{z_1})}, \\ &= \frac{\ln(N) + \ln(\frac{z_2}{z_1}) - \ln(z_1)}{\ln(\frac{z_2}{z_1})}, \\ &= \frac{\ln(N/z_1)}{\ln(z_2/z_1)} + 1. \end{aligned}$$

□

Ce n'est qu'une approximation pour deux raisons

1. Elle est obtenue à partir d'estimations. En général, la réponse exacte dépend surtout de la structure du graphe observé.
2. Il faut que tous les sommets soient atteignables à partir d'un sommet de départ choisi aléatoirement.

On peut donc dire que ce résultat est erroné lorsque l'on regarde un graphe sans composante connexe géante. Cependant, cette équation permet de trouver des résultats importants :

1. On remarque que la distance moyenne entre deux sommets pour un graphe aléatoire (en fonction de la distribution des degrés) devrait augmenter de façon logarithmique en fonction de N . Ce résultat est très connu pour un bon nombre de cas particulier de graphes.
2. La distance moyenne peut être calculée juste en connaissant le nombre de premiers et deuxièmes voisins (qui sont des propriétés locales).

Cela voudrait donc dire que si deux graphes possèdent les mêmes valeurs pour z_1 et z_2 , alors la distance moyenne sera la même pour chacun d'eux.

2.5.2 Applications

Nous allons maintenant voir ce que ces résultats impliquent pour des graphes aléatoires que nous connaissons déjà bien.

— Distribution des degrés selon une loi de Poisson :

Dans ce cas là, nous avons déjà trouvé que $z_2 = z_1^2$, ce qui facilite fortement les calculs. On obtient donc :

$$l = \frac{\ln(N/z_1)}{\ln(z_2/z_1)} + 1 = \frac{\ln(N/z_1)}{\ln(z_1)} + 1 = \frac{\ln(N) - \ln(z_1)}{\ln(z_1)} + 1 = \frac{\ln(N)}{\ln(z_1)} - 1 + 1 = \frac{\ln(N)}{\ln(z_1)}.$$

— Distribution des degrés selon une loi Puissance :

Ici, nous avons vu précédemment que

$$z_1 = G'_0(1) = \frac{Li_{\tau-1}(e^{-1/k})}{Li_{\tau}(e^{-1/k})},$$

et

$$z_2 = G''_0(1) = \frac{Li_{\tau-2}(e^{-1/k}) - Li_{\tau-1}(e^{-1/k})}{Li_{\tau}(e^{-1/k})}.$$

On a donc :

$$l = \frac{\ln(N Li_{\tau}(e^{-1/k}) / Li_{\tau-1}(e^{-1/k}))}{\ln([Li_{\tau-2}(e^{-1/k}) - Li_{\tau-1}(e^{-1/k})] / Li_{\tau-1}(e^{-1/k}))} + 1 = \frac{\ln(N) + \ln(Li_{\tau}(e^{-1/k}) / Li_{\tau-1}(e^{-1/k}))}{\ln([Li_{\tau-2}(e^{-1/k}) / Li_{\tau-1}(e^{-1/k})] - 1)} + 1.$$

En faisant tendre k vers $+\infty$, cela devient :

$$l = \frac{\ln N + \ln(\zeta(\tau) / \zeta(\tau - 1))}{\ln([\zeta(\tau - 2) / \zeta(\tau - 1)] - 1)} + 1.$$

On remarque que ces expressions ne prennent pas de valeurs réelles positives pour $\tau < 3$.

3 Simulations et résultats

Nous allons maintenant vous présenter les résultats découlant de nos simulations.

3.1 Algorithmes utilisés

L'idée ici était tout d'abord de générer un graphe aléatoire selon une suite de degrés, puis de le parcourir afin d'obtenir la taille de la plus grosse composante connexe. Les degrés ont été générés à partir d'une loi puissance. Pour $\kappa > 0$ fixé, on obtient le degré d'un sommet avec la formule suivante :

$$\lceil k = -\kappa \ln(1 - r) \rceil,$$

avec $0 \leq r < 1$. Ensuite, nous acceptons ce résultat avec une probabilité $k^{-\tau}$ (c'est-à-dire que si on refuse, on recommence l'opération jusqu'à ce que le degré soit validé).

A partir de ça, nous avons testé deux façons différentes de générer nos graphes.

3.1.1 Algorithme de Molloy-Reed

Cet algorithme consiste à tirer deux sommets au hasard, en coefficientant par les degrés de chacun. C'est-à-dire qu'un sommet de degré 9 a trois fois plus de chance d'être choisi qu'un sommet de degré 3 par exemple. Le problème est qu'on a très peu de chance d'obtenir un graphe qui respecte la suite de degrés, mais en moyenne, on devrait s'en approcher.

3.1.2 Algorithme de Havel-Hakimi

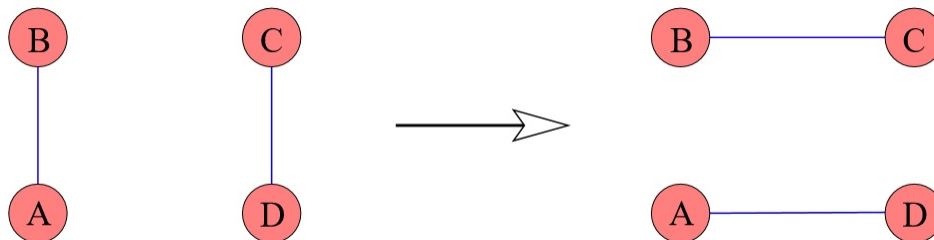
Dans celui-là, la façon de procéder est bien différente : plutôt que de construire un graphe aléatoire respectant la suite de degrés fixée (c'est-à-dire en choisir un uniformément parmi tous ceux qui possèdent une telle suite de degrés), nous allons en créer un non aléatoire, puis tenter de le rendre aléatoire.

— Construction du graphe

Le procédé est très simple, pour poser une arête, nous prenons le sommet qui a le plus de places disponibles (*degré – nombreVoisins* maximal), et nous le relions avec n'importe quel sommet libre. Nous répétons cette opération tant que la suite de degrés n'est pas parfaitement respectée.

— Pour rendre notre graphe aléatoire, nous allons le modifier. Nous sélectionnons deux arêtes aléatoirement, puis nous les intervertissons (comme vous le voyez sur la figure 1).

FIGURE 1 – Inversion d'arêtes



Cette étape ne fait pas parti de l'algorithme à la base, mais est nécessaire afin d'éviter d'obtenir un graphe trop prévisible.

Cela n'est bien sur possible que si D n'est pas déjà relié à A , de même pour B et C . Nous répétons cette opération autant de fois que la somme des degrés de notre graphe.

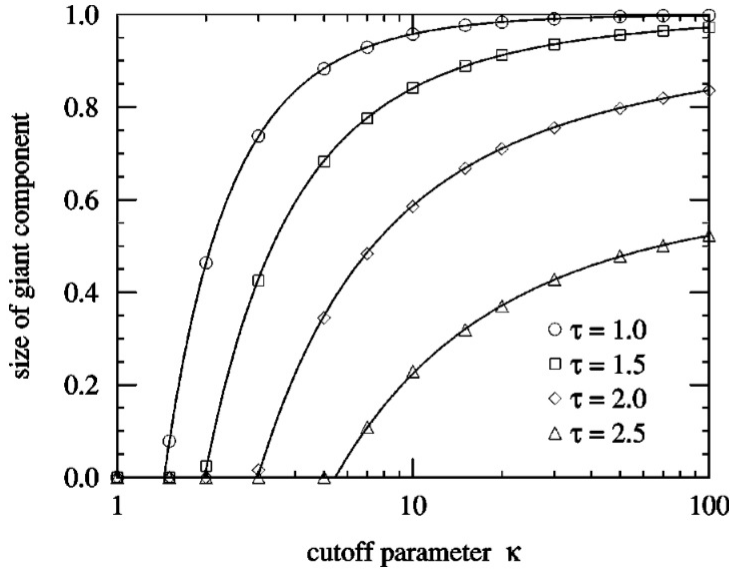
3.1.3 Parcours du graphe

Une fois le graphe généré, nous utilisons un parcours en profondeur afin de calculer la taille de la composante connexe géante. Le principe est assez simple, nous partons d'un sommet quelconque, et nous comptons 1. Puis nous allons voir tous les voisins de ce sommet, et nous comptons encore 1 pour chacun, et ainsi de suite. Pour être sûr de ne pas passer deux fois sur le même sommet, nous le marquons. Nous obtenons ainsi un résultat entier pour chaque composante connexe, il ne reste plus qu'à extraire le maximum de ceux-ci.

3.2 Résultats

Nous avons à notre disposition des résultats de simulation déjà concluants (vérifiés par des calculs). Notre but était tout d'abord d'essayer d'obtenir des résultats similaires à l'aide de nos algorithmes.

FIGURE 2 – Résultats attendus



Les résultats ci-dessus sont obtenus pour des graphes à un million de sommets. Mais réaliser d'aussi grands objets n'était pas possible avec notre programme (pour vous donner un ordre de grandeur, nous générons 500 graphes pour être le plus précis possible, et un graphe à un million de sommets prend une dizaine de minutes avant d'être créé et parcouru), nous avons donc dû nous limiter à dix mille sommets. Voici ce que nous avons obtenu avec les deux algorithmes détaillés plus haut.

FIGURE 3 – Résultats : Algorithme de Molloy-Reed

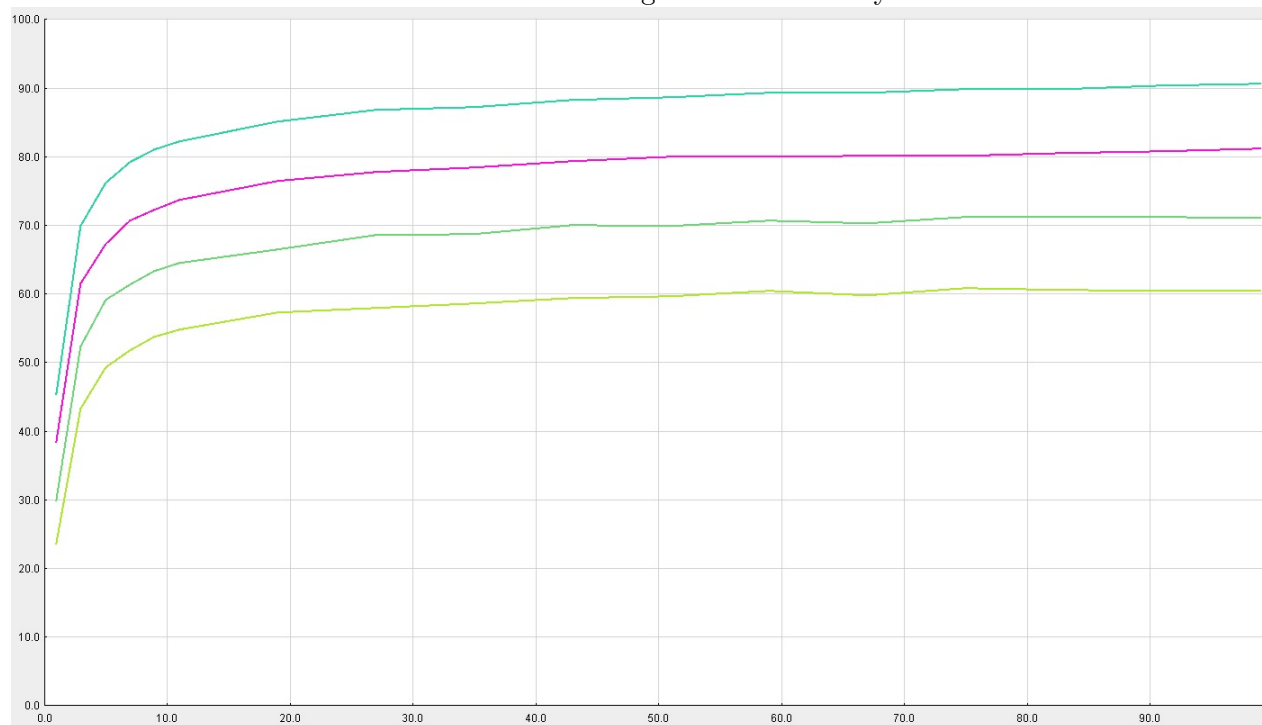
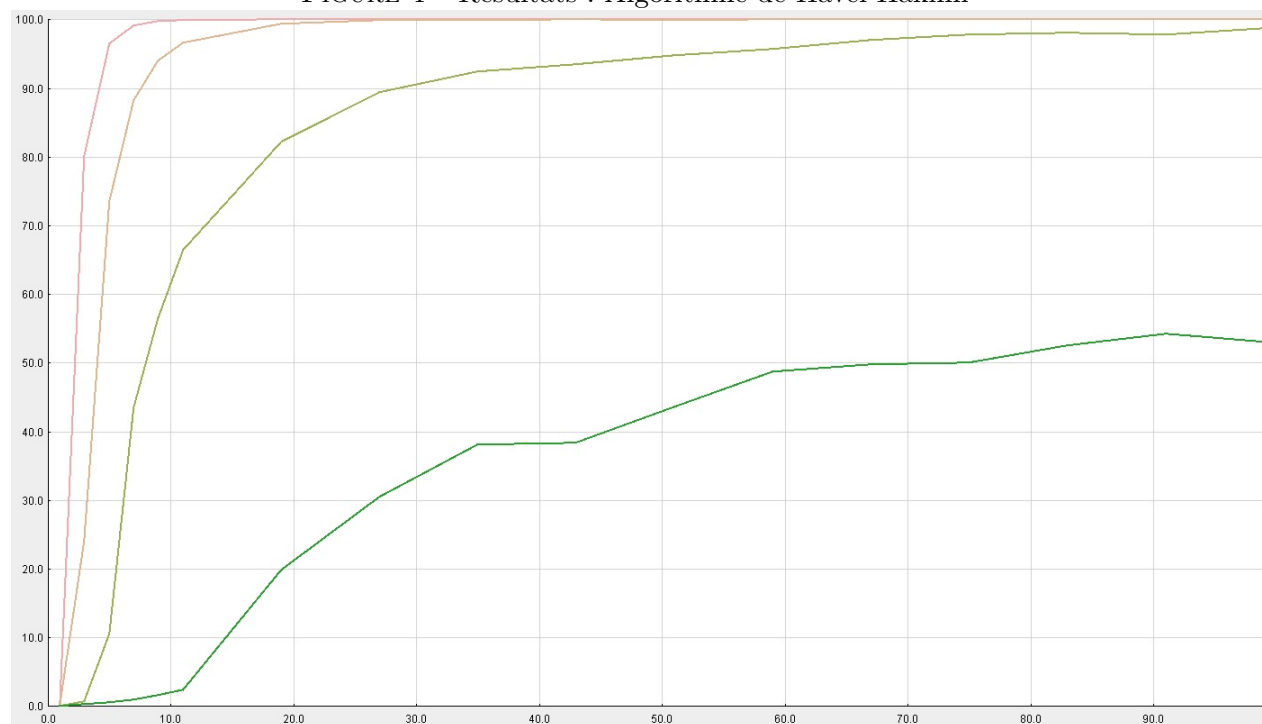


FIGURE 4 – Résultats : Algorithme de Havel-Hakimi



On remarque clairement une similitude entre la figure 2 et la figure 4. Bien que cela ne prouve rien, on sait donc que nos résultats sont proches de ceux des précédentes recherches. Par contre, dans le cas de la figure 3, on voit clairement que les résultats sont très éloignés, comparés aux précédentes simulations. On peut penser que cela viendrait du fait que nous créons avec l'algorithme de Molloy-Reed, des graphes certes aléatoires, mais qui peuvent ne pas respecter la suite de degrés imposée. Il serait donc plausible d'avoir des résultats différents.

4 Conclusion

Pour conclure, l'étude des graphes aléatoires nous a permis de nous rendre compte à quel point le sujet est riche et intéressant. Effectivement, ils nous permettent de modéliser des réseaux de la vie réelles et ainsi arriver à des conclusions qui permettent des avancées dans plusieurs domaines notamment le web et l'épidémiologie. De plus, on a vu que le choix du modèle est important. De ce fait, étant donné que le modèle de Erdős-Rényi fournit nécessairement une distribution des degrés "Poissonienne", qui est éloigné de nombreux réseaux connus, le but a été d'étudier un modèle aléatoire se rapprochant de la vie réelle. Grâce à la fonction génératrice et aux nombreux outils qu'elle fournit, nous sommes arrivés à prouver des résultats tels que la distribution des degrés pour les n_{iemes} voisins atteints, ou l'existence et la taille de la composante connexe géante. Enfin, à la suite d'une simulation numérique, nous avons pu prouver "l'exactitude" des résultats mathématiques.