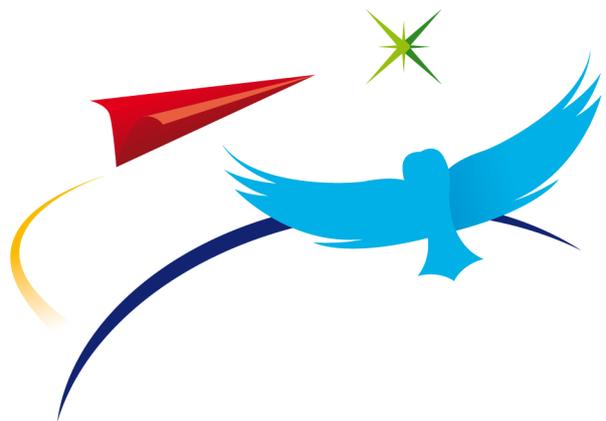


EISC-102 – Introduction à la théorie des graphes aléatoires

F. Simatos
12 avril 2019

ISAE



Institut Supérieur de l'Aéronautique et de l'Espace

S U P A E R O

EISC-102 – Introduction à la théorie des graphes aléatoires

F. Simatos
12 avril 2019

Cette œuvre est mise à disposition selon les termes de la **Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International**. Une copie de cette licence est disponible à l'adresse suivante :

<https://creativecommons.org/licenses/by-nc-sa/4.0/>





1	Graphes d’Erdős–Rényi	1
1.1	Définition et généralités	1
1.2	Méthode du premier et du second moment	2
1.2.1	Transition de phase pour l’existence de nœuds isolés.	2
1.2.2	Transition de phase pour l’existence de nœuds de degré ≤ 1	3
1.2.3	Une transition de phase pour le diamètre.	4
1.3	Première obstruction : exemple de la transition de phase pour la connectivité	6
1.4	Emergence de la composante géante	7
1.4.1	Le régime sous-critique	8
1.4.2	Le régime sur-critique	9
1.4.3	Une autre preuve du cas sous-critique.	12
1.4.4	Existence de chemins arbitrairement longs	13
1.5	Exposition multiple	14
2	Arbres de Galton–Watson	15
2.1	Arbres discrets	15
2.2	Arbres de Galton–Watson	17
2.2.1	Propriétés de base	17
2.2.2	Transition de phase pour la probabilité d’extinction.	18
2.2.3	Taille de l’arbre dans la cas sous-critique	18
2.3	Loi de l’arbre conditionné à l’extinction	19
2.4	Décomposition en épine de l’arbre conditionné à survivre	20
2.4.1	Survie à l’instant n	20
2.4.2	Survie à $n = \infty$ dans le cas sur-critique	22
2.4.3	Survie à $n = \infty$ dans le cas sous-critique	23

3	Un aperçu de quelques autres modèles de graphes aléatoires....	27
3.1	Lois de puissance	27
3.1.1	Le modèle de configuration	27
3.1.2	Le modèle d'attachement préférentiel.	27
3.2	Phénomène de petit monde	27
3.3	Stochastic block model	28



1.1 Définition et généralités

Le graphe d'Erdős-Rényi est un modèle de graphe aléatoire sur l'ensemble $\mathcal{G}(n)$ des graphes simples, labellisés à n nœuds. Il s'agit donc d'une variable aléatoire sur un ensemble fini. La cardinalité de $\mathcal{G}(n)$ est calculé en comptant que chaque arête peut être présente ou non. Puisqu'il y a $\binom{n}{2} = \frac{n(n-1)}{2}$ arêtes possibles, on a donc

$$|\mathcal{G}(n)| = 2^{n(n-1)/2}.$$

Définition 1.1. Pour $p \in [0, 1]$, on notera $G(n, p) \in \mathcal{G}(n)$ un graphe aléatoire tiré dans $\mathcal{G}(n)$ de la manière suivante : chaque arête est présente avec probabilité p indépendamment les unes des autres.

Lemme 1.2. La loi de $G(n, p)$ est donnée par

$$\mathbb{P}(G(n, p) = H) = p^m (1 - p)^{\frac{n(n-1)}{2} - m}, \quad H \in \mathcal{G}(n),$$

où m est le nombre d'arêtes de H .

Démonstration. On parcourt chaque arête, qui par définition est dans le graphe avec probabilité p et n'y est pas avec probabilité $1 - p$. □

Corollaire 1.3. $G(n, 1/2)$ est la mesure uniforme sur l'ensemble des graphes simples à n sommets.

Démonstration. Pour $p = 1/2$ on a donc

$$\mathbb{P}(G(n, p) = H) = \left(\frac{1}{2}\right)^{n(n-1)/2} = \frac{1}{|\mathcal{G}(n)|}.$$

□

Corollaire 1.4. Conditionné par son nombre d'arêtes m , $G(n, p)$ est uniformément réparti sur l'ensemble des graphes de $\mathcal{G}(n)$ à m arêtes.

Démonstration. Pour $H \in \mathcal{G}(n)$ avec m arêtes, on a

$$\mathbb{P}(G(n, p) = H \mid m \text{ arêtes}) = \frac{\mathbb{P}(G(n, p) = H)}{\mathbb{P}(m \text{ arêtes})} = \frac{p^m (1-p)^{\frac{n(n-1)}{2} - m}}{\sum_{H \in \mathcal{G}(n): m \text{ arêtes}} p^m (1-p)^{\frac{n(n-1)}{2} - m}}.$$

□

Lemme 1.5. Pour tout $v \in V$, $\mathcal{L}(\deg(v)) = \text{Bin}(n-1, p)$. En particulier, $\mathbb{E}(\deg(v)) = (n-1)p$.

Donc degré borné $\Leftrightarrow p \leq \lambda/n$, régime particulièrement riche où le degré converge vers une loi de Poisson de paramètre λ .

Questions typiques : plus long cycle, graphe hamiltonien, nombre chromatique, diamètre, connectivité, etc. Encore de nombreuses questions ouvertes ! Notamment les phénomènes de transition de phase.

1.2 Méthode du premier et du second moment

Proposition 1.6. Si $X \geq 0$, alors

$$\mathbb{P}(X = 0) \leq \frac{\text{Var}(X)}{\mathbb{E}(X^2)}.$$

Si $X \in \mathbb{N}$, alors

$$1 - \mathbb{E}(X) \leq \mathbb{P}(X = 0).$$

En particulier, si (X_n) est une suite de variables aléatoires à valeurs dans \mathbb{N} , alors :

- $\mathbb{P}(X_n = 0) \rightarrow 1$ si $\mathbb{E}(X_n) \rightarrow 0$;
- $\mathbb{P}(X_n = 0) \rightarrow 0$ si $\mathbb{E}(X_n) \rightarrow \infty$ et $\mathbb{E}(X_n^2) \sim [\mathbb{E}(X_n)]^2$.

Démonstration. $\mathbb{P}(X \geq 1) \leq \mathbb{E}(X)$ par Markov, et

$$\mathbb{E}(X) = \mathbb{E}(X; X > 0) \leq \sqrt{\mathbb{E}(X^2)\mathbb{P}(X > 0)}$$

par Cauchy-Schwarz. □

1.2.1 Transition de phase pour l'existence de nœuds isolés

On dit qu'un nœud est isolé si son degré est nul, i.e., qu'il n'a pas de voisin.

Proposition 1.7. Dans le régime $p = c \ln n / n$, on a

$$\mathbb{P}(\exists \text{ nœud isolé}) \rightarrow \begin{cases} 1 & \text{si } c < 1, \\ 0 & \text{si } c > 1. \end{cases}$$

Démonstration. En écrivant $S = \sum X_i$ avec $X_i = \mathbb{1}\{\deg(i) = 0\}$ on a $\mathbb{P}(\exists \text{ nœud isolé}) = 1 - \mathbb{P}(S = 0)$ et il faut donc montrer que $\mathbb{P}(S = 0) \rightarrow \mathbb{1}\{c > 1\}$, et pour cela on va utiliser la méthode du premier et du second moment. On a

$$\mathbb{E}(S) = n(1-p)^{n-1} \sim n(1-p)^n = n \exp(n \ln(1-p)) = n \exp(-np + O(p^2 n)) \sim n^{1-c}.$$

Donc $\mathbb{E}(S) \rightarrow 0$ pour $c > 1$, et donc $\mathbb{P}(S = 0) \rightarrow 1$ pour $c > 1$ par la méthode du premier moment. On suppose maintenant $c < 1$, il suffit de montrer par la méthode du second moment que $\mathbb{E}(S_n^2) \sim [\mathbb{E}(S)]^2$. On a

$$\mathbb{E}(S^2) = \mathbb{E}(S) + n(n-1)\mathbb{P}(\deg(1) = \deg(2) = 0).$$

Puisque

$$\mathbb{P}(\deg(1) = \deg(2) = 0) = (1-p)^{2n-3}$$

on a

$$n(n-1)\mathbb{P}(\deg(1) = \deg(2) = 0) \sim n^2(1-p)^{2n} \sim (\mathbb{E}(S))^2$$

ce qui donne le résultat. \square

1.2.2 Transition de phase pour l'existence de nœuds de degré ≤ 1

Dans la section précédente on a étudié l'existence de nœuds isolés, i.e., de degré 0. On regarde maintenant le nombre de nœuds de degré ≤ 1 .

Proposition 1.8. Dans le régime

$$p = \frac{\ln n + \ln \ln n + \omega_n}{n}$$

on a

$$\mathbb{P}(\exists i : \deg(i) \leq 1) \rightarrow \begin{cases} 1 & \text{si } \omega_n \rightarrow -\infty \\ 0 & \text{si } \omega_n \rightarrow +\infty. \end{cases}$$

Démonstration. On définit maintenant $S = \sum_{k=1}^n \mathbb{1}\{\deg(i) \leq 1\}$ et il s'agit donc de montrer que $\mathbb{P}(S = 0) \rightarrow 1$ si $\omega_n \rightarrow -\infty$ et $\mathbb{P}(S = 0) \rightarrow 0$ si $\omega_n \rightarrow +\infty$. Pour cela on utilise la méthode du premier et du second moment. On a

$$\mathbb{E}(S) = n((1-p)^{n-1} + np(1-p)^{n-2}) \approx e^{\ln n - np} + e^{2 \ln n - np + \ln p}$$

Pour $np = \ln n + \ln \ln n + \omega_n$ on a donc

$$\mathbb{E}(S) \approx e^{-\ln \ln n} + e^{2 \ln n - \ln n - \ln \ln n - \omega_n + \ln \ln n - \ln n}$$

et donc

$$\mathbb{E}(S) \rightarrow \begin{cases} 0 & \text{si } \omega_n \rightarrow +\infty \\ \infty & \text{si } \omega_n \rightarrow -\infty \end{cases}$$

On a donc bien $\mathbb{P}(S = 0) \rightarrow 1$ pour $\omega_n \rightarrow +\infty$. Pour $\omega_n \rightarrow -\infty$, on étudie donc le second moment. On a

$$\mathbb{E}(S^2) = \mathbb{E}(S) + n(n-1)\mathbb{P}(\deg(1) \leq 1, \deg(2) \leq 1)$$

et

$$\begin{aligned} \mathbb{P}(\deg(1) = 0, \deg(2) = 0) &= (1-p)^{2n-3} \\ \mathbb{P}(\deg(1) = 0, \deg(2) = 1) &= (1-p)^{n-1}(n-2)p(1-p)^{n-3} \\ \mathbb{P}(1 \sim 2, \deg(1) = \deg(2) = 1) &= p(1-p)^{2n-4} \\ \mathbb{P}(1 \not\sim 2, \deg(1) = \deg(2) = 1) &= (1-p)[(n-2)p(1-p)^{n-3}]^2 \end{aligned}$$

et donc

$$\begin{aligned} \mathbb{E}(S^2) &= n(1-p)^{n-1} \\ &\quad + n(n-1)(1-p)^{2n-3} \\ &\quad + n(n-1)(n-2)p(1-p)^{2n-4} \\ &\quad + n(n-1)p(1-p)^{2n-4} \\ &\quad + n(n-1)(n-2)^2p^2(1-p)^{2n-5} \end{aligned}$$

On a

$$\begin{aligned} \frac{n(1-p)^{n-1}}{n^4p^2(1-p)^{2n}} &\sim \frac{1}{n^3p^2(1-p)^n} = e^{-3\ln n - 2\ln \ln n + 2\ln n - np} \rightarrow 0 \\ \frac{n(n-1)(1-p)^{2n-3}}{n^4p^2(1-p)^{2n}} &\sim \frac{1}{n^2p^2} \rightarrow 0 \\ \frac{n(n-1)(n-2)p(1-p)^{2n-4}}{n^4p^2(1-p)^{2n}} &\sim \frac{1}{np} \rightarrow 0 \\ \frac{n(n-1)p(1-p)^{2n-4}}{n^4p^2(1-p)^{2n}} &\sim \frac{1}{n^2p} \rightarrow 0 \\ \frac{n(n-1)(n-2)^2p^2(1-p)^{2n-5}}{n^4p^2(1-p)^{2n}} &\rightarrow 1 \end{aligned}$$

ce qui montre bien que $\mathbb{E}(S^2) \sim (\mathbb{E}S)^2$ et donc $\mathbb{P}(S = 0) \rightarrow 0$ pour $\omega_n \rightarrow -\infty$. □

1.2.3 Une transition de phase pour le diamètre

Théorème 1.9. Dans le régime $p = c\sqrt{2\ln n/n}$, on a

$$\mathbb{P}(\text{diam}(G) \leq 2) \rightarrow \begin{cases} 0 & \text{si } c < 1, \\ 1 & \text{si } c > 1. \end{cases}$$

Démonstration. On dit qu'une arête $e = (i, j)$ est mauvaise si la distance entre i et j est > 2 , i.e., elle n'est pas dans le graphe et que ses extrémités n'ont pas de voisin en commun. Soit x_e l'indicatrice de si l'arête est mauvaise :

$$x_{ij} = \mathbb{1} \{i \not\sim j \text{ et } \forall k : i \not\sim k \text{ ou } k \not\sim j\}$$

Alors le graphe est de diamètre ≤ 2 si et seulement si aucune arête est mauvaise, et donc si $X = \sum_e x_e$ on a

$$\mathbb{P}(\text{diam}(G) \leq 2) = \mathbb{P}(X = 0)$$

On va donc appliquer la méthode des premier et second moment. On a

$$\mathbb{E}(x_e) = (1-p)(1-p^2)^{n-2} \sim e^{-np^2} = n^{-2c^2}$$

et donc

$$\mathbb{E}(X) = \binom{n}{2} (1-p)(1-p^2)^{n-2} \sim \frac{1}{2} n^{2-2c^2}.$$

Donc $\mathbb{E}(X) \rightarrow 0$ pour $c > 1$ et donc $\mathbb{P}(X = 0) = \mathbb{P}(\text{diam}(G) \leq 2) \rightarrow 1$ pour $c > 1$. On calcule maintenant le second moment. Pour cela, il faut calculer $\mathbb{E}(x_e x_{e'})$. On distingue trois cas.

1. si e et e' n'ont aucune extrémité en commun, alors x_e et $x_{e'}$ sont indépendantes, et donc $\mathbb{E}(x_e x_{e'}) = (\mathbb{E}(x_e))^2$. Puisqu'il y a $\binom{n}{2} \binom{n-2}{2}$ couples d'arêtes avec aucune extrémité en commun, cela donne une contribution

$$\binom{n}{2} \binom{n-2}{2} [(1-p)(1-p^2)^{n-2}]^2 \sim \frac{1}{4} n^{4-4c^2};$$

2. si e et e' ont deux extrémités en commun, i.e., $e = e'$, alors $\mathbb{E}(x_e x_{e'}) = \mathbb{E}(x_e)$. Puisqu'il y a $\binom{n}{2}$ couples d'arêtes avec deux extrémités en commun, cela donne une contribution

$$\binom{n}{2} (1-p)(1-p^2)^{n-2} \sim \frac{1}{2} n^{2-2c^2};$$

3. si e et e' ont une extrémité en commun, disons $e = (i, j)$ et $e' = (i, j')$ avec $i \neq j$, $j \neq j'$ et $i \neq j'$, alors $x_e x_{e'} = 1$ si et seulement si e et e' sont mauvaises, i.e. :

- les arêtes (i, j) et (i, j') sont absentes, ce qui arrive avec probabilité $(1-p)^2$;
- pour tout $k \neq i, j, j'$, k n'est pas un voisin commun de i et j et n'est pas un voisin commun de i et j' , i.e., soit $k \not\sim i$ ou $k \sim i$, $k \not\sim j$ et $k \not\sim j'$, ce qui arrive avec probabilité $1-p+p(1-p)^2 = 1-2p^2+p^3$.

Puisqu'il y a $\binom{n}{2} 2(n-2)$ couples d'arêtes avec une extrémité en commun, cela donne une contribution

$$\binom{n}{2} 2(n-2)(1-p)^2 (1-2p^2+p^3)^{n-3} \sim n^3 e^{-2np^2} = n^{3-4c^2}.$$

Le terme dominant est donc le premier, qui est $\sim [\mathbb{E}(X)]^2$, et on peut donc appliquer la méthode du second moment pour obtenir que $\mathbb{P}(X = 0) \rightarrow 0$ pour $c < 1$. \square

1.3 Première obstruction : exemple de la transition de phase pour la connectivité

Le résultat suivant montre que la transition de phase pour la connectivité arrive au même moment que la transition de phase pour le nombre de nœuds isolés, ce qui constitue un exemple du phénomène de première obstruction.

On pourrait montrer que la transition de phase pour l'existence de chemins hamiltoniens a lieu dans le régime

$$p = \frac{\ln n + \ln \ln n + \omega_n}{n}$$

et que donc, au vu de la Proposition 1.8, le phénomène de première obstruction a aussi lieu pour l'existence de chemins hamiltoniens, cf. [2, Théorèmes 8.9 et 8.11].

Théorème 1.10. *Dans le régime $p = c \ln n / n$, on a*

$$\mathbb{P}(G \text{ connecté}) \rightarrow \begin{cases} 0 & \text{si } c < 1, \\ 1 & \text{si } c > 1. \end{cases}$$

En fait, un résultat beaucoup plus fin existe. Si on ajoute les arêtes une par une, le moment où le dernier nœud isolé disparaît correspond exactement (avec probabilité qui tend vers 1) au moment où le graphe devient connecté, cf. Théorème 7.4 de Bollobas [2]. Par ailleurs, le résultat suivant (admis) décrit le comportement asymptotique de la probabilité que $G(n, p)$ soit connecté pour $c = 1$.

Théorème 1.11. *Si $p = \frac{\ln n + c}{n}$, alors $\mathbb{P}(G \text{ connecté}) \rightarrow \exp(-e^{-c})$.*

Démonstration du Théorème 1.10. On suppose $c < 1$. Alors d'après la Proposition 1.7 on a

$$\mathbb{P}(G(n, p) \text{ connecté}) \leq \mathbb{P}(\nexists \text{ nœud isolé}) \rightarrow 0.$$

On étudie maintenant le cas difficile $c > 1$: on sait qu'il n'y a pas de nœuds isolés, et on veut montrer que cela implique que le graphe est connecté. L'idée est qu'il ne peut y avoir deux composantes géantes. Si G n'est pas connecté, alors il existe $A \subset V$ avec $|A| = k \leq n/2$ telle qu'il n'y a aucune arête entre A et $V \setminus A$. On a donc

$$\begin{aligned} \mathbb{P}(G \text{ non connecté}) &\leq \mathbb{P}(\exists A \subset V : |A| \leq n/2 \text{ et pas d'arête entre } A \text{ et } V \setminus A) \\ &\leq \sum_{k=1}^{n/2} \mathbb{P}(\exists A \subset V : |A| = k \text{ et pas d'arête entre } A \text{ et } V \setminus A) \\ &\leq \sum_{k=1}^{n/2} \binom{n}{k} \mathbb{P}(\text{pas d'arête entre } \{1, \dots, k\} \text{ et } \{k+1, \dots, n\}) \\ &= \sum_{k=1}^{n/2} \binom{n}{k} (1-p)^{k(n-k)} \\ &\leq \sum_{k=1}^{n/2} \left(\frac{ne}{k}\right)^k e^{-pk(n-k)} \end{aligned}$$

où l'on a utilisé $\ln(1-x) \leq -x$ et

$$\binom{n}{k} \leq (ne/k)^k \quad (1.1)$$

pour la dernière inégalité. En effet,

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1} \leq \frac{n^k}{k^k} \times \frac{k^k}{k!} \leq \left(\frac{en}{k}\right)^k.$$

Si l'on définit $f(x) = x(1-x)$ et $g(x) = x \ln(1/x)$, on a donc en utilisant $p = c \ln n/n$

$$\mathbb{P}(G \text{ non connecté}) \leq \sum_{k=1}^{n/2} \left(\frac{ne}{k}\right)^k e^{-cf(k/n)n \ln n} = \sum_{k=1}^{n/2} \exp(-cf(k/n)n \ln n + ng(k/n)).$$

L'idée est que le terme en $n \ln n$ va dominer et donner un terme qui décroît suffisamment vite vers 0. Le seul problème est pour les valeurs de k/n petite. En effet, comme f est croissante sur $[0, 1/2]$ et que $\sup g = g(1/2) = 1/e$, on a pour tout $x \in [\varepsilon, 1/2]$ que

$$-cf(x)n \ln n + ng(x) \leq -cf(\varepsilon)n \ln n + ng(1/2)$$

ce qui donne

$$\sum_{k=\varepsilon n}^{n/2} \exp(-cf(k/n)n \ln n + ng(k/n)) \leq n \exp(-cf(\varepsilon)n \ln n + ng(1/2))$$

qui tend vers 0. Il reste donc à contrôler les autres termes pour $k = 1, \dots, \varepsilon n$. Pour cela, on réécrit la borne

$$\sum_{k=1}^{\varepsilon n} \left(\frac{ne}{k}\right)^k e^{-cf(k/n)n \ln n} = \sum_{k=1}^{\varepsilon n} \left(\frac{e}{k}\right)^k \exp(-(cf(k/n) - k/n)n \ln n).$$

Pour $x \rightarrow 0$, on a $f(x) \sim x$ et donc comme $c > 1$, il existe $\alpha > 0$ tel que $cf(x) - x \geq \alpha x$ pour $x \leq \varepsilon$: on a alors

$$\sum_{k=1}^{x^* n} \left(\frac{e}{k}\right)^k e^{-n \ln n (cf(\frac{k}{n}) - \frac{k}{n})} \leq \sum_{k=1}^{x^* n} \left(\frac{e}{k}\right)^k e^{-\alpha k \ln n} \leq \frac{1}{n^\alpha} \sum_{k \geq 1} \left(\frac{e}{k}\right)^k$$

qui tend bien vers 0. □

1.4 Emergence de la composante géante

On s'intéresse maintenant au régime $p = c/n$, qui correspond au régime de l'émergence d'une composante géante, i.e., de taille $\propto n$. On note dans la suite \mathcal{C}_k la k -ième plus grande composante connexe de $G(n, p)$. Le résultat suivant (admis) décrit de manière assez fine ce qui se passe.

Théorème 1.12. *Soit \mathcal{C}_k la k -ième plus grande composante connexe de $G(n, p)$. Dans le régime $np = \lambda \in]0, \infty[$, on a :*

- si $\lambda < 1$ alors $|\mathcal{C}_1|/\ln n \xrightarrow{L} (\lambda - 1 - \ln \lambda)^{-1}$;
- si $\lambda > 1$ alors $|\mathcal{C}_1|/n \xrightarrow{L} 1 - \rho$ avec ρ l'unique solution dans $]0, 1[$ de $x = e^{\lambda(x-1)}$, et il existe $c \in]0, \infty[$ avec $\mathbb{P}(|\mathcal{C}_2| \geq c \ln n) \rightarrow 0$.

Dans le cadre de ce cours, nous ne démontrerons que le résultat plus faible (mais plus accessible) suivant. Il montre bien que dans le régime sous-critique, les composantes connexes sont de taille au plus logarithmique, alors que dans le régime sur-critique, il y a au moins une composante connexe de taille macroscopique.

Théorème 1.13. *Dans le régime $np = \lambda \in]0, \infty[$, on a*

- si $\lambda < 1$, alors $\mathbb{P}(|\mathcal{C}_1| \geq a \ln n) \rightarrow 0$ pour tout $a > 1/(\lambda - 1 - \ln \lambda)$;
- si $\lambda > 1$, alors il existe $\varepsilon > 0$ tel que $\mathbb{P}(\exists \text{ chemin de longueur } > \varepsilon n) \rightarrow 1$.

1.4.1 Le régime sous-critique

On s'intéresse ici au régime sous-critique $np = \lambda < 1$. L'idée est de faire un parcours en largeur de la composante connexe

ALGORITHME 1.14 – Exploration de la composante connexe de v

$T_0 = v$, v est vert et tous les autres nœuds sont blancs. Tant qu'il y a un nœud vert dans T_t :

- on prend la première feuille verte dans le parcours en profondeur, il devient bleu, on rajoute toutes les arêtes entre cette feuille et tous ses voisins dans le graphe qui ne sont pas déjà dans l'arbre, et ces voisins deviennent verts.

Lemme 1.15. *Soit \mathcal{C} la composante connexe dont fait partie v . L'Algorithme 1.14 renvoie un arbre couvrant de \mathcal{C} , et si V_k désigne le nombre de nœuds verts à l'itération k , alors*

$$|\mathcal{C}| = \min \{k \geq 0 : V_k = 0\}. \quad (1.2)$$

Si $X_t = V_t - V_{t-1}$, dans le cas d'un graphe d'Erdős–Rényi on a par définition

$$X_{t+1} \mid T_t \sim \text{Binom}(n - |T_t|, p).$$

En rajoutant des arêtes fictives, on peut donc écrire $V_t \leq V_t^+$ où les incréments de V_t^+ sont i.i.d. de loi commune $\text{binom}(n, p)$. Cela revient à majorer la taille de la composante connexe par la taille d'un arbre de Galton–Watson, et donc on obtient le résultat suivant.

Lemme 1.16. *Si \mathcal{C} est la composante connexe dont fait partie v , alors*

$$\mathbb{P}(|\mathcal{C}| > z) \leq e^{-z\psi(\lambda)}$$

avec $\psi(\lambda) = -\ln((n-1)p) + n \ln(1 - 1/n) - n \ln(1-p) + \ln(1-p)$.

Démonstration. On utilise la Proposition 2.9. Dans ce cas, la loi de reproduction est $\text{binom}(n, p)$ et donc

$$\mathbb{E}(e^{\theta \xi}) = (pe^\theta + 1 - p)^n$$

et donc

$$\varphi(\theta) = \theta - n \ln(pe^\theta + 1 - p)$$

et donc

$$\varphi'(\theta) = 1 - n \frac{pe^\theta}{pe^\theta + 1 - p} = 0 \iff \theta = \ln\left(\frac{1-p}{(n-1)p}\right)$$

et donc

$$\sup \varphi = \ln\left(\frac{1-p}{(n-1)p}\right) - n \ln\left(\frac{1-p}{n-1} + 1 - p\right) = \ln\left(\frac{1-p}{(n-1)p}\right) - n \ln\left(\frac{n(1-p)}{n-1}\right).$$

□

On obtient alors directement la preuve du Théorème 1.13 dans la cas sous-critique.

Démonstration du Théorème 1.13 dans le cas $\lambda < 1$. Si $\mathcal{C}[v]$ dénote la composante connexe de v , on a

$$\mathbb{P}(|\mathcal{C}_1| \geq a \ln n) = \mathbb{P}(\exists v : |\mathcal{C}[v]| \geq a \ln n)$$

et donc la borne de l'union donne

$$\mathbb{P}(|\mathcal{C}_1| \geq a \ln n) \leq n \mathbb{P}(|\mathcal{C}[1]| \geq a \ln n)$$

et donc le lemme précédent donne

$$\mathbb{P}(|\mathcal{C}_1| \geq a \ln n) \leq n \exp(-a\psi \ln n)$$

qui tend vers 0 pour $a\psi > 1$, ce qui donne le résultat puisque $\psi \rightarrow \lambda - 1 - \ln \lambda$. □

1.4.2 Le régime sur-critique

Autre exploration de la composante connexe

On pourrait étudier le régime sur-critique en comparant à nouveau la composante connexe de v à un arbre de Galton–Watson, l'intuition étant que le cas $\lambda > 1$ correspond à un arbre de Galton–Watson sur-critique, et donc infini avec probabilité > 0 . Moralement, si l'arbre de Galton–Watson correspondant à v est infini, alors v va faire partie d'une composante connexe de taille macroscopique et cela explique la constante $1 - \rho$ qui apparaît dans le Théorème 1.12 : cette constante est la probabilité (et donc la proportion de nœuds) que v génère un arbre de Galton–Watson infini. Néanmoins, formaliser cet argument est très subtil, cf. [6]. A la place, nous suivrons l'approche de [5] qui repose sur une autre exploration de la composante connexe d'un nœud. La différence principale est qu'on découvre les voisins un par un, au lieu de “en paquet” comme dans l'Algorithme 1.14.

A chaque étape de cet algorithme, on a une partition N_t, A_t, I_t de l'ensemble des nœuds :

- $N = \{\text{nœuds neutres}\}$;
- $A = \{\text{nœuds actifs}\}$, c'est une pile $\rightarrow s(A) = \text{sommet de } A$, i.e., dernier nœud ajouté ;
- $I = \{\text{nœuds inactifs}\}$.

Tous les nœuds commencent neutres, et l'exploration en profondeur transforme les nœuds : neutres \rightarrow actifs \rightarrow inactifs. On suppose que les nœuds sont ordonnés.

ALGORITHME 1.17 – Autre exploration de la composante connexe de v

- (1) $t = 0, A_0 = I_0 = \emptyset, N_0 = V$;
- (2) si $I_t \neq V$:
 - (2.1) si $A_t = \emptyset$: $N_{t+1} = N_t \setminus \{\min(N_t)\}, A_{t+1} = \{\min(N_t)\}, t = t + 1 \rightarrow (2)$;
 - (2.2) sinon :
 - (2.2.1) si $N_t = \emptyset$, on passe successivement tous les nœuds de A dans I ;
 - (2.2.2) sinon, soit $v = \min(N_t)$: si $(s(A_t), v) \in E, N_{t+1} = N_t \setminus \{v\}, A_{t+1} = A_t \cup \{v\}, t = t + 1, \rightarrow (2.2)$;
 - (2.2.3) sinon, $v = \text{suivant}(v, s(A_t), N_t)$, i.e., le nœud suivant v dans N_t qui n'a pas encore été exploré par $s(A_t)$ si $N_t \neq \emptyset, t = t + 1 \rightarrow (2.2.1)$;
 - (2.2.4) si on est arrivés à la fin de la liste, $A_{t+1} = A_t \setminus \{s(A_t)\}, I_{t+1} = I_t \cup \{s(A_t)\}$;
- (3) explorer les arêtes non explorées.

Lemme 1.18. *Pour tout $t \geq 0$, il existe un chemin C tel que $A_t \subset C$.*

Démonstration. Récurrence : induction ok quand on enlève un nœud, mais aussi quand on en ajoute un. \square

On voit l'étape 2.2.1 comme poser une question, et on définit $X_i = 1$ si la réponse à la i -ième question est oui et 0 sinon. On définit alors Q_t le nombre de questions posées jusqu'à l'instant t .

Proposition 1.19. *Le graphe de départ est $G(n, p)$ si et seulement si les (X_i) sont i.i.d. $\sim Be(p)$.*

Démonstration. Dans les deux cas, chaque arête est questionnée une unique fois et est déclarée présente avec proba p . \square

Intérêt : on transforme un problème avec de la géométrie en un problème sur une suite de variables aléatoires i.i.d..

Lemme 1.20. *Pour tout $t \geq 0$ on a $|I_t| |N_t| \leq t$.*

Démonstration. Soit $v \in I_t$: nécessairement, toutes les paires (v, w) avec $w \in N_t$ ont été questionnées (et ont déclaré être absentes). Il y a $|I_t| |N_t|$ telles paires, et chacune a nécessairement été questionné une fois, ce qui a pris une unité de temps à chaque fois. \square

Lemme 1.21. *Pour tout $t \geq 0$ on a $|A_t| \leq 1 + \sum_{k=1}^t X_k$.*

Démonstration. On distingue les nœuds qui deviennent actifs suite à une question (verts) des nœuds qui deviennent actifs car ils commencent l'exploration d'une composante (bleus). Le nombre de verts est borné par le nombre de réponses positives à l'instant t , à savoir $\sum_{k=1}^{Q_t} X_k$, et le nombre de bleus vaut un (c'est le nœud tout en bas de la pile) : cela donne $|A_t| \leq 1 + \sum_{k=1}^{Q_t} X_k \leq 1 + \sum_{k=1}^t X_k$ \square

Lemme 1.22. *Pour tout $t \geq 0$ on a $|A_t| + |I_t| \geq \sum_{k=1}^t X_k$.*

Démonstration. On remarque que

$$|A_t \cup I_t| \geq \sum_{k=1}^{Q_t} X_k + \# \text{ composantes connexes dont l'exploration a commencé avant } t.$$

En effet, si le membre de droite augmente, le membre de gauche aussi, et le membre de gauche ne peut décroître. En outre, le nombre de composantes connexes dont l'exploration a commencé avant t est donné par $t - Q_t \geq \sum_{k=1}^t X_k - \sum_{k=1}^{Q_t} X_k$ ce qui donne le résultat. \square

Application au cas sur-critique

Lemme 1.23. *$T = \varepsilon n^2/2$: si $|\sum_{i=1}^T X_i - \frac{\varepsilon(1+\varepsilon)n}{2}| \leq n^{2/3}$ alors $|A_T| \geq \varepsilon^2 n/5$.*

Démonstration. On suppose que $|\sum_{i=1}^T X_i - \frac{\varepsilon(1+\varepsilon)n}{2}| \leq n^{2/3}$ et on montre par l'absurde que $|A_T| \geq \varepsilon^2 n/5$.

On montre d'abord que $|I_T| < n/3$. En utilisant les lemmes 1.20 et 1.21 il vient pour tout $t \leq T$

$$\begin{aligned} T &\geq t \\ &\geq |I_t| |N_t| \\ &= |I_t| (n - |I_t| - |A_t|) \\ &\geq |I_t| (n - |I_t| - \sum_{k=1}^t X_k - 1) \\ &\geq |I_t| (n - |I_t| - 1 - \frac{\varepsilon(1+\varepsilon)n}{2} - n^{2/3}) \end{aligned}$$

et donc pour ε suffisamment petit et tout $t \leq T$,

$$|I_t| (2n/3 - |I_t|) \leq T = \frac{\varepsilon n^2}{2}$$

ce qui implique $|I_T| < n/3$ car sinon, il existerait $t \leq T$ avec $|I_t| = n/3$.

Supposons donc $|A_T| \leq \varepsilon^2 n/5$: le lemme 1.22 donne

$$|I_T| \geq \sum_{k=1}^T X_k - |A_T| \geq \frac{\varepsilon(1+\varepsilon)n}{2} - n^{2/3} - \frac{\varepsilon^2 n}{5} = \frac{\varepsilon n}{2} - n^{2/3} + \frac{3\varepsilon^2 n}{10}.$$

Donc

$$\frac{\varepsilon n^2}{2} = T \geq |I_T| |N_T| = |I_T| (n - |A_T| - |I_T|)$$

et puisque $|I_T| < n/3 < (n - |A_T|)/2$ et $x \mapsto x(n - |A_T| - x)$ est croissante sur $[0, (n - |A_T|)/2]$, on obtient

$$\begin{aligned} \frac{\varepsilon n^2}{2} &\geq \left(\frac{\varepsilon n}{2} - n^{2/3} + \frac{3\varepsilon^2 n}{10} \right) \left(n - \frac{\varepsilon^2 n}{5} - \left(\frac{\varepsilon n}{2} - n^{2/3} + \frac{3\varepsilon^2 n}{10} \right) \right) \\ &= \left(\frac{\varepsilon n}{2} - n^{2/3} + \frac{3\varepsilon^2 n}{10} \right) \left(n - \frac{\varepsilon^2 n}{2} - \frac{\varepsilon n}{2} + n^{2/3} \right) \\ &= \frac{\varepsilon n^2}{2} + \left(-\frac{1}{4} + \frac{3}{10} \right) \varepsilon^2 n^2 + O(\varepsilon^3) n^2 + o(n^2). \end{aligned}$$

Comme $1/4 > 3/10$, on obtient une contradiction. \square

Lemme 1.24. $\mathbb{P}(|\sum_{i=1}^T X_i - \frac{\varepsilon(1+\varepsilon)n}{2}| \leq n^{2/3}) \rightarrow 1$, avec $\varepsilon = \lambda - 1$.

Démonstration. Bienaymé-Tchebychev. \square

1.4.3 Une autre preuve du cas sous-critique

On peut utiliser cette autre exploration de la composante connexe pour prouver le résultat suivant d'une autre manière que via une comparaison avec Galton-Watson.

Proposition 1.25. Si $\lambda > 1$, alors il existe $c > 0$ tel que $\mathbb{P}(|\mathcal{C}_1| \geq \varepsilon \ln n) \rightarrow 0$.

Pour cela on a besoin du résultat supplémentaire suivant.

Lemme 1.26. Soit \mathcal{C} explorée entre t et t' , et pour $s \in [t, t']$

$$E_s = \{(v, w) \in V^2 : v \text{ ou } w \in A_s \cup (I_s \cap \mathcal{C}) \text{ et } (v, w) \text{ a été questionnée}\}.$$

Alors pour tout $t \leq s < t'$, on a $s - t = |E_s|$.

Démonstration. Tant qu'on explore une composante connexe, on pose toujours une question à une paire dont l'une des extrémités est soit active, soit inactive, et cette extrémité le reste pour toujours (soit active, soit inactive). \square

Ainsi que du résultat suivant.

Lemme 1.27. Pour tout $0 < a < 1$ on a $\mathbb{P}(\text{Bin}(n, p) > (1+a)np) \leq e^{-a^2 np/3}$.

Lemme 1.28. Si $|\mathcal{C}_1| \geq 7 \ln n / \varepsilon^2$ alors il existe i avec $\sum_{j=i}^{i+kn} X_j \geq k$.

Démonstration. Soit \mathcal{C} une composante connexe de G de taille $> k$, t l'instant où son exploration a commencé, t' le $k+1$ -ième instant où un nœud de \mathcal{C} devient actif, $I' = I_{t'-1} \cap \mathcal{C}$ et $A' = A_{t'-1}$. Alors $|I' \cup A'| = k$ et donc l'algorithme a eu k réponses positives entre t et t' : il reste à montrer que $t' - t \leq kn$. Le lemme 1.26 donne

$$\begin{aligned}
t' - t &= |E_{t'}| \\
&\leq |\{(v, w) \in V^2 : v \text{ ou } w \in A' \cup I'\}| \\
&\leq \underbrace{\binom{|A' \cup I'|}{2}}_{\text{arêtes internes}} + \underbrace{|I' \cup A'| (n - |I' \cup A'|)}_{\text{arêtes avec une seule extrémité dans } I' \cup A'} \\
&= \frac{k(k-1)}{2} + k(n-k) \\
&\leq kn.
\end{aligned}$$

□

Lemme 1.29. $\mathbb{P}(\exists i : \sum_{j=i}^{i+kn} X_j \geq k) \rightarrow 0$.

Démonstration. On a

$$\mathbb{P}(\text{intervalle de longueur } kn \text{ contenant au moins } k \text{ un}) \leq n^2 \mathbb{P}(\text{Bin}(kn, p) \geq k)$$

En utilisant le lemme on obtient

$$\mathbb{P}(\text{Bin}(kn, p) \geq k) \leq \exp\left(-\frac{1}{3} \left(\frac{1}{np} - 1\right)^2 knp\right) = e^{-\frac{7}{3(1-\varepsilon)} \ln n}$$

qui donne finalement

$$\mathbb{P}\left(|\mathcal{C}_1| \geq \frac{7 \ln n}{\varepsilon^2}\right) \leq \exp\left(-\frac{7}{3(1-\varepsilon)} \ln n + 2 \ln n\right) \rightarrow 0.$$

parce que ε a été choisi de telle sorte que $\frac{7}{3(1-\varepsilon)} > 2$. □

1.4.4 Existence de chemins arbitrairement longs

On peut utiliser l'Algorithme 1.17 pour prouver l'existence de chemins arbitrairement longs.

Lemme 1.30. Soit $k < n$ entiers et $G = (V, E)$ avec n nœuds tel que tous S, T disjoints avec $|S|, |T| = k$ ont au moins une arête entre S et T . Alors G contient un chemin de longueur $n - 2k + 1$.

Démonstration. DFS : on considère t où $|I| = |N|$. Par construction, ils n'ont pas d'arêtes et donc, par hypothèse sur le graphe, $|I| \leq k - 1$ et donc $|A| = n - |I| - |N| \geq n - 2k + 2$. □

Théorème 1.31. *Pour tout $\varepsilon > 0$ suffisamment petit, il existe C tel que si $G \sim G(n, p)$ avec $p = C/n$, alors $\mathbb{P}(G \text{ contient un chemin de longueur } \geq (1 - \varepsilon)n) \rightarrow 1$.*

Démonstration. Soit $k = \varepsilon n/2$: il suffit de montrer que pour tout A, B disjoints avec $|A| = |B| = k$ ils ont au moins une arête en commun. Pour A et B disjoints et fixés de taille k , la probabilité qu'ils n'aient pas d'arêtes en commun est $(1 - p)^{k^2}$ et donc

$$\mathbb{P}(\text{pour tout } A, B \dots) \leq \binom{n}{k}^2 (1 - p)^{k^2} \leq \left(\frac{en}{k}\right)^{2k} e^{-pk^2}$$

en utilisant (1.1). □

1.5 Exposition multiple

Proposition 1.32 (Exposition multiple/arrosage). *Si $G_i \sim G(n, p_i)$ alors $G_1 \cup G_2 \sim G(n, q)$ avec $q = 1 - (1 - p_1)(1 - p_2)$.*

Cela permet de montrer des propriétés de monotonie évidentes.

Lemme 1.33. *La fonction $\mathbb{P}(G(n, p) \text{ est connecté})$ est croissante en p .*

Cette technique permet aussi de montrer des résultats plus fins. Par exemple, une fois que l'on sait que l'on a un chemin grâce au Théorème 1.13, on peut montrer qu'il y a en fait un cycle.

Lemme 1.34. *On suppose $np = \lambda$ avec $\lambda > 1$. Soit ε la constante qui apparaît dans le Théorème 1.13. Alors $\mathbb{P}(\exists \text{ cycle de longueur } \geq \varepsilon n) \rightarrow 1$.*

Démonstration. On utilise l'arrosage multiple et on écrit $1 - p = (1 - p_1)(1 - p_2)$ avec $p_2 = \varepsilon/(2n)$ si bien que $p_1 \geq (1 + \varepsilon/2)/n$ et donc G_1 contient avec grande proba un chemin de longueur $\propto n$. On conclut en utilisant le fait que

$$\mathbb{P}(G_2 \text{ contient une arête entre les } n^{2/3} \text{ premiers et derniers nœuds de } P) \rightarrow 1.$$

En effet, 1 moins cette proba est $\leq (1 - p_2)^{n^{4/3}}$ puisqu'il y a $n^{4/3}$ arêtes à considérer (et que les nœuds du chemin sont indépendants de l'arrosage). □



2.1 Arbres discrets

Un arbre discret est un arbre enraciné et plongé dans le plan, si bien qu'il y a une notion de droite et de gauche. Formellement, il est défini de la manière suivante. On note $\mathcal{U} = \{\emptyset\} \cup \bigcup_{n \geq 1} (\mathbb{N}^*)^n$ l'ensemble des individus, i.e., les suites finies d'entiers, avec \emptyset la racine. Pour un individu $u \in \mathcal{U}$ on note $|u| \in \mathbb{N}$ sa génération, i.e., l'unique entier n tel que $u \in (\mathbb{N}^*)^n$, avec $|u| = 0$ si $u = \emptyset$.

Définition 2.1. (Notation de Neveu–Ulam–Harris) Un arbre discret t est un sous-ensemble de \mathcal{U} (i.e., un ensemble d'individus), tel que :

1. $\emptyset \in t$
 2. $(u_1, \dots, u_n) \in t \Rightarrow (u_1, \dots, u_{n-1}) \in t$
 3. pour tout $u = (u_1, \dots, u_n) \in t$, il existe $\exists N_u(t) \in \mathbb{N}$ tel que $(u_1, \dots, u_n, k) \in t \Leftrightarrow k \in \{1, \dots, N_u(t)\}$ (avec $\{1, \dots, N_u(t)\} = \emptyset$ pour $N_u(t) = 0$).
- On note $\mathcal{T} = \{\text{arbres discrets}\}$ l'ensemble des arbres discrets.

Cette définition autorise les arbres discrets infinis. Par ailleurs, on note d'emblée que l'ensemble \mathcal{T} n'est pas dénombrable, ce qui pose en théorie des difficultés de mesurabilité, que l'on passera allègrement sous le tapis dans le cadre de ce cours. On aura besoin de plusieurs fonctionnelles définies sur les arbres. Pour $t \in \mathcal{T}$ et $n \geq 0$, on considérera notamment :

- $|t| \in \mathbb{N}^* \cup \{\infty\}$ la taille de l'arbre t :

$$|t| = \sum_{u \in \mathcal{U}} \mathbb{1}\{u \in t\};$$

- $t|_n \in \mathcal{T}$ l'arbre t tronqué au niveau n :

$$t|_n = \{u \in t : |u| \leq n\};$$

- $Z_n(t) \in \mathbb{N}$ le nombre d'individus à la génération n dans l'arbre t :

$$Z_n(t) = \sum_{u \in t} \mathbb{1}\{|u| = n\};$$

- $h(t) \in \mathbb{N}^* \cup \{\infty\}$ la hauteur de l'arbre t :

$$h(t) = \inf\{n \geq 1 : Z_n(t) = 0\} \in \mathbb{N}^* \cup \{\infty\};$$

- $\theta_u(t) \in \mathcal{T}$ l'arbre enraciné à $u \in t$:

$$\theta_u(t) = \{v \in \mathcal{U} : uv \in t\}.$$

On a en particulier $h(t \mid_n) \leq n + 1$.

Lemme 2.2. On a

$$\sum_{u \in t, |u| < n} (N_u(t) - 1) = Z_n(t) - 1.$$

On présente deux preuves. La première preuve, directe et plus simple, et une deuxième preuve par récurrence, plus compliquée mais qui permet de se familiariser avec les techniques que l'on utilisera par la suite.

Démonstration directe. On a

$$\sum_{u \in t, |u| < n} (N_u(t) - 1) = \sum_{i=0}^{n-1} \sum_{u \in t, |u|=i} (N_u(t) - 1)$$

et puisque

$$\sum_{u \in t, |u|=i} N_u(t) = Z_{i+1}(t)$$

on obtient

$$\sum_{u \in t, |u| < n} (N_u(t) - 1) = \sum_{i=0}^{n-1} \sum_{u \in t, |u|=i} (Z_{i+1}(t) - Z_i(t))$$

ce qui donne le résultat puisque $Z_0(t) = 1$. □

Démonstration par récurrence. Pour $n = 1$ la propriété est vraie, et pour $n \geq 2$ on a

$$\begin{aligned} \sum_{u \in t, |u| < n} (N_u(t) - 1) &= N_\emptyset(t) - 1 + \sum_{u \in t, 1 \leq |u| < n} (N_u(t) - 1) \\ &= Z_1(t) - 1 + \sum_{u \in t, |u|=1} \sum_{v \in \theta_u(t), |v| < n-1} (N_v(\theta_u(t)) - 1) \\ &= Z_1(t) - 1 + \sum_{u \in t, |u|=1} (Z_{n-1}(\theta_u(t)) - 1) \\ &= Z_1(t) - 1 + Z_n(\theta_u(t)) - Z_1(t) \end{aligned}$$

□

2.2 Arbres de Galton–Watson

2.2.1 Propriétés de base

Un arbre de Galton–Watson est un arbre discret aléatoire : c’est donc une variable aléatoire dans l’ensemble \mathcal{T} . On rappelle que cet ensemble n’est pas dénombrable, mais on passera sous silence les subtilités de mesurabilité que cela engendre. Un arbre de Galton–Watson est caractérisée par une loi de reproduction que l’on identifiera avec la loi d’une variable aléatoire ξ . On définit un arbre de Galton–Watson par la propriété de branchement suivante.

Définition 2.3 (Arbre de Galton–Watson). T variable aléatoire à valeurs dans \mathcal{T} est un arbre de Galton–Watson de loi de reproduction ξ si :

1. $Z_1(T)$ est distribué selon ξ ;
2. conditionnellement à $Z_1(T), \theta_1(T), \dots, \theta_k(T)$ sont i.i.d. et suivent la même loi que T .

De manière plus formelle, T est un arbre de Galton–Watson si et seulement si pour tout $k \geq 0$ et tout $A_1, \dots, A_k \subset \mathcal{T}$ mesurables, on a

$$\mathbb{P}(Z_1 = k, \theta_i \in A_i) = \mathbb{P}(\xi = k) \prod_{i=1}^k \mathbb{P}(T \in A_i)$$

avec la convention $\prod_{i=1}^k \mathbb{P}(T \in A_i) = 1$ pour $k = 0$.

Attention : T peut être infini c’est la raison pour laquelle on considère des ensembles généraux A_i . En effet, de même que pour une variable absolument continue X on a $\mathbb{P}(X = x) = 0$, ici on a $\mathbb{P}(T = t) = 0$ pour $t \in \mathcal{T}$ infini. Le résultat suivant, qui décrit la loi d’un arbre de Galton–Watson, est plus proche de l’intuition, qui nous dit qu’un arbre de Galton–Watson est tel que les nombres d’enfants sont des variables i.i.d. distribuées selon la loi de reproduction. Comme T peut a priori être infini, on considère dans le résultat suivant la loi de l’arbre tronqué.

Proposition 2.4. Si T est un arbre de Galton–Watson de loi de reproduction ξ , alors pour tout $t \in \mathcal{T}$ et $n \geq 0$ on a

$$\mathbb{P}(T \mid_n = t) = \prod_{u \in t, |u| < n} \mathbb{P}(\xi = N_u(t)).$$

Démonstration. Par récurrence :

$$\begin{aligned} \mathbb{P}(T \mid_n = t) &= \mathbb{P}(N_\emptyset(T) = N_\emptyset(t), \theta_u(T) \mid_{n-1} = \theta_u(t) \mid_{n-1}, u \in t, |u| = 1) \\ &= \mathbb{P}(\xi = N_\emptyset(t)) \prod_{u \in t, |u|=1} \mathbb{P}(T \mid_{n-1} = \theta_u(t) \mid_{n-1}) \\ &= \mathbb{P}(\xi = N_\emptyset(t)) \prod_{u \in t, |u|=1} \prod_{v \in \theta_u(t) \mid_{n-1}} \mathbb{P}(\xi = N_v(t)) \end{aligned}$$

□

Lemme 2.5. Si T est un arbre de Galton–Watson de loi de reproduction ξ , alors $\mathbb{E}(Z_n(T)) = (\mathbb{E}\xi)^n$.

Démonstration. Par récurrence :

$$\mathbb{E}(Z_{n+1}(T)) = \mathbb{E}\left(\sum_{u \in T: |u|=1} Z_n(\theta_u(T))\right) = \mathbb{E}(\xi)\mathbb{E}(Z_n(T))$$

avec la deuxième égalité qui vient de la définition d'un arbre de Galton–Watson. \square

2.2.2 Transition de phase pour la probabilité d'extinction

On s'intéresse à la probabilité d'extinction de T , i.e., la probabilité que T soit fini que l'on note $q = \mathbb{P}(|T| < \infty)$. On va prouver le résultat de transition de phase suivant.

Théorème 2.6. Si $\mathbb{E}(\xi) \leq 1$ et $\mathbb{P}(\xi = 0) > 0$, alors $q = 1$. Sinon, $q < 1$.

On a la classification suivante :

Cas sous-critique : $\mathbb{E}\xi < 1$;

Cas critique : $\mathbb{E}\xi = 1$;

Cas sur-critique : $\mathbb{E}\xi > 1$.

Ce résultat est en fait un corollaire du résultat suivant. Dans la suite on note $q_n = \mathbb{P}(h < n) = \mathbb{P}(Z_n = 0)$ la probabilité que T survive jusqu'à la génération n . Par monotonie, on a $q_n \uparrow q = \mathbb{P}(h < \infty)$. Dans la suite, on note ϕ la fonction génératrice de ξ : $\phi(s) = \mathbb{E}(s^\xi)$ pour $s \in [0, 1]$.

Théorème 2.7. On a $q_{n+1} = \phi(q_n)$. En particulier, q est le plus petit point fixe de ϕ dans $[0, 1]$.

Démonstration. La propriété de branchement donne $q_{n+1} = \phi(q_n)$ et donc q est bien un point fixe de ϕ . Par ailleurs, on a $\phi'(x) = \mathbb{E}(\xi s^{\xi-1})$ et $\phi''(x) = \mathbb{E}(\xi(\xi-1)s^{\xi-2})$ et donc ϕ est convexe. En outre, $\phi(0) = \mathbb{P}(\xi = 0)$, $\phi(1) = 1$ et $\phi'(1) = m$ et donc si $\phi'(0) < 1$ alors on a équivalence entre $\mathbb{E}\xi > 1$ et l'existence d'un point fixe < 1 . Puisque $q_0 = 0$ et $q_{n+1} = \phi(q_n)$, q est en effet le plus petit point fixe de ϕ dans ce cas. En effet, si \underline{q} est le plus petit point fixe, $q_n < \underline{q}$ implique $q_{n+1} = \phi(q_n) < \phi(\underline{q}) = \underline{q}$ et donc $q \leq \underline{q}$. Puisque $\phi'(0) = \mathbb{P}(\xi = 1)$ on obtient le résultat. \square

2.2.3 Taille de l'arbre dans la cas sous-critique

On utilise les notations de l'algorithme 1.14. On a le résultat fondamental suivant lorsque l'on explore un arbre de Galton–Watson.

Lemme 2.8. Si T est un arbre de Galton–Watson, alors V est une marche aléatoire, i.e., les incréments $V_{t+1} - V_t$ sont i.i.d. et suivent la loi de reproduction ξ .

Ce résultat nous permet de montrer le résultat suivant.

Proposition 2.9. Si T est un arbre de Galton–Watson de loi de reproduction ξ sous-critique, alors

$$\mathbb{P}(|T| > z) \leq \exp\left(-z \sup_{\theta} \varphi(\theta)\right)$$

avec $\varphi(\theta) = \theta - \ln \mathbb{E}(e^{\theta \xi})$.

Démonstration. Puisque la composante connexe de la racine est l'arbre tout entier, on a d'après (1.2)

$$\mathbb{P}(|T| > z) = \mathbb{P}(V_0, \dots, V_z > 0) \leq \mathbb{P}(V_z > 0) \leq \mathbb{E}[e^{\theta(\xi-1)}]^z = \exp\left(-z\theta + z \ln \mathbb{E}(e^{\theta \xi})\right).$$

□

2.3 Loi de l'arbre conditionné à l'extinction

Théorème 2.10. Soit T un arbre de Galton–Watson de loi de reproduction ξ . Alors T conditionné à s'éteindre est égal en distribution à un arbre de Galton–Watson de loi de reproduction ξ^* donnée par

$$\mathbb{P}(\xi^* = k) = q^{k-1} \mathbb{P}(\xi = k), \quad k \in \mathbb{N}.$$

Démonstration. Soit $t \in \mathcal{T}$ un arbre fini. Alors

$$\mathbb{P}(T \mid_n = t \mid |T| < \infty) = \frac{1}{\mathbb{P}(|T| < \infty)} \mathbb{P}(T \mid_n = t) \mathbb{P}(|T| < \infty \mid T \mid_n = t).$$

La propriété de branchement implique $\mathbb{P}(|T| < \infty \mid T \mid_n = t) = q^{Z_n(t)}$ et donc la Proposition 2.4 donne

$$\mathbb{P}(T \mid_n = t \mid |T| < \infty) = q^{Z_n(t)-1} \prod_{u \in t, |u| < n} \mathbb{P}(\xi = N_u(t)).$$

On obtient donc le résultat grâce au Lemme 2.2. □

REMARQUE 2.11 n vérifie facilement que l'on a bien défini une loi :

$$\sum_{k \geq 0} q^{k-1} \mathbb{P}(\xi = k) = \frac{1}{q} \phi(q) = 1.$$

Par ailleurs, on doit avoir $\mathbb{E}(\xi^*)$ puisque l'arbre de Galton–Watson de loi de reproduction ξ^* s'éteint presque sûrement, par construction, ce que l'on vérifie facilement aussi :

$$\mathbb{E}(\xi^*) = \sum_{k \geq 0} k q^{k-1} \mathbb{P}(\xi = k) = \mathbb{E}(\xi q^{\xi-1}) = \phi'(q).$$

On vérifie sur un dessin que nécessairement $\phi'(q) \leq 1$ par convexité, ce qui donne bien le résultat. ■

EXEMPLE 2.12 (Loi de Poisson). Quand ξ suit une loi de Poisson de paramètre λ , alors

$$\mathbb{P}(\xi^* = k) = q^{k-1} \mathbb{P}(\xi = k) = q^k e^{\lambda(1-q)} \frac{e^{-\lambda} \lambda^k}{k!} = q^k e^{\lambda(1-q)} \frac{e^{-\lambda} \lambda^k}{k!}$$

et l'on reconnaît la loi de Poisson de paramètre $q\lambda = \mu$, solution de $\mu e^{-\mu} = \lambda e^{-\lambda}$ avec $\mu < 1 < \lambda$. Dans ce cas (λ, μ) est appelé une paire conjuguée. ■

2.4 Décomposition en épine de l'arbre conditionné à survivre

2.4.1 Survie à l'instant n

On suit la présentation de [4]. Pour $t \in \mathcal{T}$ et $n \geq 0$ on définit $R_{n+1}(t) \in \mathbb{N}_* \cup \{\infty\}$ le rang du premier enfant (dans l'ordre lexicographique, i.e., le plus à gauche) de la racine qui a un descendant à la génération $n+1$:

$$R_{n+1}(t) = \min \{1 \leq i \leq Z_1(t) : Z_n(\theta_i(t)) > 0\}$$

avec la convention $\min \emptyset = \infty$.

Théorème 2.13. Soit $1 \leq j \leq k < \infty$. Conditionnellement à $\{R_{n+1} = j, Z_1 = k\}$, les sous-arbres $(\theta_i, i = 1, \dots, k)$ sont indépendants avec

$$\mathcal{L}(\theta_i | R_{n+1} = j, Z_1 = k) = \begin{cases} \mathcal{L}(T | Z_n = 0) & 1 \leq i \leq j-1, \\ \mathcal{L}(T | Z_n > 0) & i = j, \\ \mathcal{L}(T) & j+1 \leq i \leq k. \end{cases} \quad (2.1)$$

La distribution jointe de R_{n+1} et Z_1 conditionnellement à $Z_{n+1} > 0$ est

$$\mathbb{P}(R_{n+1} = j, Z_1 = k | Z_{n+1} > 0) = c_n q_n^{j-1} \mathbb{P}(\xi = k) \quad \text{où} \quad c_n = \frac{1 - q_n}{1 - q_{n+1}}. \quad (2.2)$$

Démonstration. Soit

$$A_i = \begin{cases} \{t \in \mathcal{T} : Z_n(t) = 0\} & \text{si } 1 \leq i < j, \\ \{t \in \mathcal{T} : Z_n(t) > 0\} & \text{si } i = j, \\ \mathcal{T} & \text{si } j < i \leq k. \end{cases}$$

Alors

$$\begin{aligned} \mathbb{P}(R_{n+1} = j, Z_1 = k) &= \mathbb{P}(Z_1 = k, \theta_1 \in A_1, \dots, \theta_k \in A_k) \\ &= \mathbb{P}(\xi = k) \prod_{i=1}^k \mathbb{P}(T \in A_i) \\ &= \mathbb{P}(\xi = k) q_n^{j-1} (1 - q_n) \end{aligned}$$

et puisque

$$\mathbb{P}(R_{n+1} = j, Z_1 = k \mid Z_{n+1} > 0) = \frac{1}{1 - q_{n+1}} \mathbb{P}(R_{n+1} = j, Z_1 = k)$$

on obtient bien (2.2). De plus, pour $\tilde{A}_i \subset \mathcal{T}$, on a

$$\begin{aligned} \mathbb{P}(\theta_1 \in \tilde{A}_1, \dots, \theta_k \in \tilde{A}_k \mid R_{n+1} = j, Z_1 = k) \\ &= \frac{\mathbb{P}(\theta_1 \in \tilde{A}_1, \dots, \theta_k \in \tilde{A}_k, R_{n+1} = j, Z_1 = k)}{\mathbb{P}(R_{n+1} = j, Z_1 = k)} \\ &= \frac{\mathbb{P}(\theta_1 \in A'_1, \dots, \theta_k \in \tilde{A}'_k, Z_1 = k)}{\mathbb{P}(R_{n+1} = j, Z_1 = k)} \end{aligned}$$

avec $A'_i = A_i \cap \tilde{A}_i$. Cela donne

$$\mathbb{P}(\theta_1 \in \tilde{A}_1, \dots, \theta_k \in \tilde{A}_k \mid R_{n+1} = j, Z_1 = k) = \frac{1}{(1 - q_n)q_n^{j-1}} \prod_{i=1}^k \mathbb{P}(T \in A'_i)$$

et puisque

$$\mathbb{P}(T \in A'_i) = \begin{cases} \mathbb{P}(T \in \tilde{A}_i \mid Z_n = 0)q_n & \text{si } 1 \leq i < j, \\ \mathbb{P}(T \in \tilde{A}_i \mid Z_n > 0)(1 - q_n) & \text{si } i = j, \\ \mathbb{P}(T \in \tilde{A}_i) & \text{si } j < i \leq k, \end{cases}$$

on obtient

$$\begin{aligned} \mathbb{P}(\theta_1 \in \tilde{A}_1, \dots, \theta_k \in \tilde{A}_k \mid R_{n+1} = j, Z_1 = k) \\ &= \prod_{i=1}^{j-1} \mathbb{P}(T \in \tilde{A}_i \mid Z_n = 0) \times \mathbb{P}(T \in \tilde{A}_j \mid Z_n > 0) \times \prod_{i=j+1}^k \mathbb{P}(T \in \tilde{A}_i) \end{aligned}$$

ce qui prouve le résultat. \square

Soit (V_{n+1}, Y_{n+1}) , $n \geq 0$, une suite de variables aléatoires indépendantes avec la loi (2.2) :

$$\mathbb{P}(V_{n+1} = j, Y_{n+1} = k \mid Z_{n+1} > 0) = c_n p_k q_n^{j-1}$$

et soit \tilde{T}_0 un arbre de Galton–Watson indépendant. On construit de manière récursive \tilde{T}_{n+1} de la manière suivante :

1. la première génération de \tilde{T}_{n+1} est Y_{n+1} ;
2. on prend \tilde{T}_n le sous-arbre enraciné par le V_{n+1} -ème enfant de la première génération de \tilde{T}_{n+1} ;
3. les $V_{n+1} - 1$ frères à la gauche de l'enfant distingué forment des arbres de Galton–Watson conditionnés à mourir à la génération n ;
4. les $Y_{n+1} - V_{n+1}$ frères à la droite de l'enfant distingué donnent naissance à des arbres de Galton–Watson non conditionnés.

Théorème 2.14. $\mathcal{L}(\tilde{T}_n) = \mathcal{L}(T \mid Z_n > 0)$.

Démonstration. Par récurrence. \square

2.4.2 Survie à $n = \infty$ dans le cas sur-critique

On décrit maintenant l'arbre sur-critique (i.e., on suppose que $\mathbb{E}(\xi)m > 1$) conditionné à la survie.

Proposition 2.15. *Conditionnellement à $h(T) = \infty$, soit \tilde{T} le sous-arbre de T consistant des nœuds avec une lignée infinie. Alors \tilde{T} est un arbre de Galton–Watson de loi de reproduction $\tilde{\xi}$ de fonction génératrice*

$$\tilde{\phi}(s) = \frac{\phi((1-q)s+q) - q}{1-q}.$$

Démonstration. En admettant que $\mathcal{L}(\tilde{T}_n) \rightarrow \mathcal{L}(\tilde{T})$ et $\mathcal{L}(T | Z_n > 0) \rightarrow \mathcal{L}(T | h(T) = \infty)$ (en un certain sens), il suit du Théorème 2.14 que conditionnellement à $h(T) = \infty$, T peut être décrit comme suit :

1. on tire

$$\mathbb{P}(V_\infty = j, Y_\infty = k) = p_k q^{j-1}, 1 \leq j \leq k < \infty.$$

2. la racine a Y_∞ enfants ;

3. les $V_\infty - 1$ premiers enfants génèrent un arbre de Galton–Watson de loi de reproduction ξ^* ;

4. l'enfant V_∞ génère un arbre de Galton–Watson de loi $T | h(T) = \infty$;

5. les $Y_\infty - V_\infty$ enfants restants génèrent des arbres de loi T .

Il suit de cette description de T^∞ que \tilde{T} est en effet un arbre de Galton–Watson, et que la fonction génératrice de sa loi de reproduction est donnée par

$$N_\infty = 1 + \sum_{k=V_\infty+1}^{Y_\infty} \mathbb{1}\{\theta_k \text{ est infini}\}.$$

Nous n'avons donc qu'à montrer que $\mathbb{E}(s^{N_\infty}) = \tilde{\phi}(s)$. Soit B_k des variables aléatoire i.i.d. de Bernoulli, indépendantes de $\Delta_\infty = Y_\infty - V_\infty$ et avec paramètre $1 - q$: la propriété de branchement implique alors que

$$\mathbb{E}(s^{N_\infty}) = \mathbb{E}\left(s^{1+\sum_{u=1}^{\Delta_\infty} B_k}\right) = s\mathbb{E}(x^{\Delta_\infty}) \text{ with } x = (1-q)s + q$$

et par ailleurs,

$$\begin{aligned}
\mathbb{E}(s^{N_\infty}) &= s \sum_{\delta \geq 0} x^\delta \mathbb{P}(\Delta_\infty = \delta) \\
&= s \sum_{\delta \geq 0} x^\delta \sum_{k \geq 1} \mathbb{P}(V_\infty = k - \delta, Y_\infty = k) \\
&= s \sum_{\delta \geq 0} x^\delta \sum_{k \geq \delta+1} p_k q^{k-\delta-1} \\
&= s \sum_{\delta \geq 0} \frac{x^\delta}{q^{\delta+1}} \mathbb{E}(q^\xi; \xi \geq \delta + 1) \\
&= \frac{s}{q} \mathbb{E} \left(\sum_{\delta \geq 0} \frac{x^\delta}{q^\delta} q^\xi \mathbb{1}_{\{\delta \leq \xi - 1\}} \right) \\
&= \frac{s}{q} \mathbb{E} \left(q^\xi \sum_{\delta=0}^{\xi-1} (x/q)^\delta \right) \\
&= \frac{s}{q} \mathbb{E} \left(q^\xi \frac{1 - (x/q)^\xi}{1 - x/q} \right) \\
&= \frac{s}{q-x} \mathbb{E} \left(q^\xi - x^\xi \right)
\end{aligned}$$

ce qui donne le résultat puisque $\mathbb{E}(q^\xi) = q$. □

2.4.3 Survie à $n = \infty$ dans le cas sous-critique

Préliminaires : distribution biaisée par la taille

Lemme 2.16. Soit G un graphe à n nœuds de suite de degrés $(d_i, i = 1, \dots, n)$. Soit D la loi du degré d'un nœud tiré uniformément au hasard et D^* la loi du degré d'un nœud choisi uniformément au hasard parmi les extrémités d'une arête tirée uniformément au hasard. Alors

$$\mathbb{P}(D^* = k) = \frac{k \mathbb{P}(D = k)}{\mathbb{E}(D)}.$$

Démonstration. On a

$$\mathbb{P}(D = k) = \frac{1}{n} \sum_{v \in V} \mathbb{1}_{\{d_v = k\}} \quad \text{et} \quad \mathbb{E}(D) = \frac{1}{n} \sum_v d_v.$$

Puisque $\#\text{arêtes} = \frac{1}{2} \sum_v d_v$, on a par ailleurs

$$\mathbb{P}(D^* = k) = \frac{1}{\frac{1}{2} \sum_v d_v} \sum_{e \in E} \sum_{u \in e} \frac{1}{2} \mathbb{1}_{\{d_u = k\}} = \frac{1}{n \mathbb{E}(D)} \sum_{u \in V} \mathbb{1}_{\{d_u = k\}} \sum_{e \ni u} = \frac{k}{\mathbb{E}(D)} \mathbb{P}(D = k)$$

□

Un autre exemple où les distributions biaisées arrivent est le suivant. On considère T tables, N_k tables ont k individus. On choisit une table uniformément au hasard, et on note X le nombre d'individus à cette table. Une autre manière de faire est de choisir une personne au hasard et de noter X^* le nombre de gens à sa table. On a alors

$$\mathbb{P}(X^* = k) = \frac{\#\text{personnes à la table } k}{\#\text{personnes}} \propto k \times \#\text{tables de } k \propto k\mathbb{P}(X = k).$$

Arbre sous-critique conditionné à survivre

Théorème 2.17. Si $\mathbb{E}\xi \leq 1$ et $\mathbb{P}(\xi = 1) < 1$, si bien que $q_n \rightarrow 1$, alors

$$\frac{1 - q_{n+1}}{1 - q_n} \rightarrow \mathbb{E}\xi \quad \text{et en particulier,} \quad \frac{1}{n} \log(1 - q_n) \rightarrow \log \mathbb{E}\xi.$$

Démonstration. On a vu que $q_{n+1} = \mathbb{E}(q_n^\xi)$ et donc

$$1 - q_{n+1} = \mathbb{E}(1 - (1 - (1 - q_n))^\xi)$$

i.e.,

$$\frac{1 - q_{n+1}}{1 - q_n} = \mathbb{E}(f(1 - q_n, \xi)) \quad \text{avec} \quad f(x, k) = \frac{1 - (1 - x)^k}{x}$$

Donc $f(x, k) \rightarrow k$ lorsque $x \rightarrow 0$ et donc par domination, on obtient $\frac{1 - q_{n+1}}{1 - q_n} \rightarrow \mathbb{E}(\xi)$. \square

Contrairement au cas sur-critique, on ne peut pas directement conditionner un arbre de Galton–Watson sous-critique à survivre. Pour donner un sens à ce conditionnement dégénéré, on conditionne l'arbre à survivre à l'instant n et on fait tendre $n \rightarrow \infty$. Dans la théorie des processus de Markov, cela revient à considérer le Q -processus.

Dans la suite on considère la loi de reproduction biaisée par la taille :

$$\mathbb{P}(\hat{\xi} = k) = \frac{k\mathbb{P}(\xi = k)}{\mathbb{E}\xi}.$$

Définition 2.18. \hat{T} est l'arbre biaisé par la taille : $Z_1(\hat{T})$ est distribué selon $\hat{\xi}$ et conditionnellement à $Z_1(\hat{T}) = k$:

1. on tire U uniformément au hasard dans $\{1, \dots, k\}$;
2. les arbres θ_i sont indépendants ;
3. θ_i pour $i \neq U$ est distribué selon T ;
4. θ_U est distribué selon \hat{T} .

Lemme 2.19. On a

$$\mathbb{P}(\hat{T} |_{m=t}) = \frac{p_k}{m} \sum_{j=1}^k \prod_{i \neq j} \mathbb{P}(T |_{m-1} = \theta_i(t)) \mathbb{P}(\hat{T} |_{m-1} = \theta_j(t))$$

Théorème 2.20. *Si $m \leq 1$, alors T conditionné par $Z_n > 0$ converge faiblement lorsque $n \rightarrow \infty$ vers l'arbre biaisé par la taille. Techniquement, cela veut dire que pour tout $m \geq 0$ et $t \in \mathcal{T}$ avec $Z_n(t) = 0$, on a*

$$\mathbb{P}(T |_{m=t} | Z_n > 0) \rightarrow \mathbb{P}(\widehat{T} |_{m=t}).$$

Démonstration. On a vu que $c_n \rightarrow 1/m$ et donc

$$\mathbb{P}(V_{n+1} = j, Y_{n+1} = k | Z_{n+1} > 0) = c_n p_k q_n^{j-1} \rightarrow \frac{p_k}{m} = \frac{k p_k}{m} \times \frac{1}{k}$$

On procède par récurrence sur m . Pour $m = 1$ on a

$$\mathbb{P}(T |_{1=t} | Z_n > 0) = \mathbb{P}(N_{\emptyset}(T) = N_{\emptyset}(t) | Z_n > 0) = \sum_j c_n q_n^{j-1} p_k \rightarrow \frac{k p_k}{m} = \mathbb{P}(\widehat{T} |_{1=t}).$$

Pour $m \geq 2$:

$$\begin{aligned} \mathbb{P}(T |_{m=t} | Z_n > 0) &= \mathbb{P}(Z_1 = Z_1(t), \theta_i |_{m-1} = \theta_i(t) | Z_n > 0) \\ &= \sum_j \mathbb{P}(R_{n+1} = j, Z_1 = Z_1(t), \theta_i |_{m-1} = \theta_i(t) | Z_n > 0) \\ &= \sum_j \mathbb{P}(R_{n+1} = j, Z_1 = Z_1(t) | Z_n > 0) \\ &\quad \times \mathbb{P}(\theta_i |_{m-1} = \theta_i(t) | R_{n+1} = j, Z_1 = Z_1(t), Z_n > 0) \\ &= \sum_j c_n q_n^{j-1} p_k \\ &\quad \times \mathbb{P}(\theta_i |_{m-1} = \theta_i(t) | R_{n+1} = j, Z_1 = Z_1(t)) \\ &= \sum_j c_n q_n^{j-1} p_k \\ &\quad \times \prod_{i=1}^{j-1} \mathbb{P}(T |_{m-1} = \theta_i(t) | Z_n = 0) \mathbb{P}(T |_{m-1} = \theta_j(t) | Z_n > 0) \\ &\quad \times \prod_{i>j} \mathbb{P}(T |_{m-1} = \theta_i(t)) \\ &\rightarrow \frac{p_k}{m} \sum_j \prod_{i \neq j} \mathbb{P}(T |_{m-1} = \theta_i(t)) \mathbb{P}(\widehat{T} |_{m-1} = \theta_j(t)) \\ &= \mathbb{P}(\widehat{T} |_{m=t}) \end{aligned}$$

□



3.1 Lois de puissance

En pratique, de nombreux graphes exhibent des suites de degrés qui suivent une loi de puissance. Pour ces graphes, le modèle d'Erdős–Rényi, qui prédit des lois suites de degrés dont la queue de distribution décroît exponentiellement vite, n'est pas adéquat. Deux grandes familles de modèles de graphes aléatoires ont été proposés pour essayer de capturer cette loi de puissance : le modèle de configuration et le modèle d'attachement préférentiel.

3.1.1 Le modèle de configuration

3.1.2 Le modèle d'attachement préférentiel

Albert et Barabási [1] ont proposé un modèle très populaire, appelé modèle d'attachement préférentiel. On part d'un graphe quelconque, puis on rajoute les nœuds un par un. Chaque nouveau nœud choisit m anciens nœuds au hasard, et s'y connecte. Chaque ancien nœud est choisi avec une probabilité proportionnelle à son degré. On voit émerger une loi de puissance. Le degré du nœud i évolue de la manière suivante :

$$\frac{d}{dt}k_i(t) = \frac{k_i(t)}{2t} \iff k_i(t) \propto k(t/t_i)^{1/2}$$

Pourquoi le $\frac{1}{2t}$?

et donc si on choisit i au hasard et que $t_i = i$, le degré moyen est donné par

$$\mathbb{P}(k_i(t) > k) = \mathbb{P}(t_i < m^2 t / k^2) = m^2 t / k^2 / (t + m_0)$$

et donc

$$\mathbb{P}(k_i(t) = k) = \frac{2m^2}{k^3}$$

Robustesse vs. vulnérabilité [3].

3.2 Phénomène de petit monde

Watts et Strogatz [7].

3.3 Stochastic block model



Bibliography

- [1] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- [2] Béla Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.
- [3] Béla Bollobás and Oliver Riordan. Robustness and vulnerability of scale-free random graphs. *Internet Math.*, 1(1):1–35, 2003.
- [4] Jochen Geiger. Elementary new proofs of classical limit theorems for Galton-Watson processes. *J. Appl. Probab.*, 36(2):301–309, 1999.
- [5] Michael Krivelevich and Benny Sudakov. The phase transition in random graphs: a simple proof. *Random Structures Algorithms*, 43(2):131–138, 2013.
- [6] Remco van der Hofstad. *Random Graphs and Complex Networks*, volume 1. Cambridge Series in Statistical and Probabilistic Mathematics, 2017.
- [7] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of “small-world” networks. *Nature*, 393(6684):440–442, 06 1998.