

Travaux d'Etudes et de Recherches
Composante géante dans le graphe d'Erdős-Rényi

Pierre BOUTAUD et Jérémy DAVAIL
M1 MFA, Université Paris-Sud
dirigé par Pascal MAILLARD

2015

Sommaire

1	Graphe d'Erdős-Rényi et notions essentielles	1
1.1	Le graphe d'Erdős-Rényi	1
1.2	Processus de branchement et résultats essentiels	1
1.3	Preuves des résultats essentiels	4
1.4	Exploration des composantes du graphe	11
1.5	Probabilité de survie des processus de branchement	12
2	Régime sur-critique	14
2.1	Existence et unicité de la composante géante	14
2.2	Preuve des résultats sur-critiques	17
3	Régime sous-critique	27
3.1	Ordre des composantes maximales sous-critiques	27
3.2	Preuves des résultats sous-critiques	28
4	Régime critique	32
5	Simulations	35
5.1	Code R pour les simulations	35
5.2	Présentation de simulations	37

Introduction

L'objectif de ces travaux d'études et de recherches est l'étude du comportement des composantes connexes de grandes tailles (c'est-à-dire maximale) d'un graphe aléatoire et de leur éventuelle unicité. On se place dans le cas particulier du graphe d'Erdős-Rényi avec n sommets et distribution d'arêtes p en prenant n arbitrairement grand. C'est un graphe où il existe une arête entre deux sommets quelconques du graphes avec probabilité p , indépendamment pour chacun des n sommets. Ce modèle de graphes aléatoires a été l'objet d'études de Paul Erdős et Alfréd Rényi [6] durant les années 1960. Ils se consacrent notamment à l'étude des composantes connexes du graphe.

Il s'agira donc par la suite de présenter proprement le graphe d'Erdős-Rényi ainsi que les processus de branchement, outils essentiels pour l'exploration d'un graphe et donc l'étude de ses composantes connexes.

On rappellera aussi quelques résultats généraux qui s'avèreront utiles pour notre développement.

On s'intéressera alors au cas où n est grand et où $p = \lambda/n$ pour un certain $\lambda > 0$. Suivant les valeurs de ce λ , les composantes de taille maximale vont alors avoir des propriétés fort intéressantes.

En effet, trois cas principaux ont été distingués lors des études d'Erdős et Rényi :

- *le régime sur-critique* lorsque $\lambda > 1$: la taille de la composante maximale est d'ordre n .
- *le régime critique* lorsque $\lambda = 1$: la taille de la composante maximale est d'ordre $n^{2/3}$.
- *le régime sous-critique* lorsque $\lambda < 1$: la taille de la composante maximale est d'ordre $\log(n)$.

Nous allons énoncer et prouver ces résultats en suivant Van der Hofstadt [1] mais nous allons généraliser ces trois cas en regardant l'évolution de la taille de la composante maximale lorsque $\lambda > 1$ (resp $\lambda < 1$) tend vers 1 , il s'agira de la fenêtre critique par le haut (resp fenêtre critique par le bas). Nous pousserons et adapterons donc les preuves de Van der Hofstadt pour approcher la fenêtre critique. Les preuves que nous ferons utiliseront des calculs de premier et second moments, ainsi que des couplages.

Enfin, nous ferons quelques simulations afin de concrétiser et vérifier les résultats énoncés.

Chapitre 1

Graphe d'Erdős-Rényi et notions essentielles

1.1 Le graphe d'Erdős-Rényi

Définition 1.1 (Graphe d'Erdős-Rényi). *Le graphe aléatoire d'Erdős-Rényi à n sommets dans $[n]$ et distribution d'arêtes p , noté $ER_n(p)$, est un graphe aléatoire où tous sommets v et v' distincts sont reliés avec probabilité p de manière indépendante, i.-e. :*

$$\forall v, v' \in [n], \mathbb{P}(v \leftrightarrow v') = p \quad (1.1)$$

Définition 1.2 (Matrice de correspondance). *La matrice de correspondance du graphe $ER_n(p)$ est*

$$(B_{i,j})_{1 \leq i, j \leq n} \quad (1.2)$$

où les $B_{i,j}$ sont des variables aléatoires de loi $Ber(p)$ valant 1 si $i \leftrightarrow j$ et 0 sinon. Elles sont indépendantes.

Remarque. *Les $(B_{i,j})_{1 \leq i < j \leq n}$ sont i.i.d. mais pas les $(B_{i,j})_{1 \leq i, j \leq n}$ car $\forall i, j, B_{i,j} = B_{j,i}$.*

Dans la suite, on désignera par $\mathcal{C}(v)$ la composante d'un sommet v du graphe et $|\mathcal{C}_{max}|$ la taille d'une composante maximale.

1.2 Processus de branchement et résultats essentiels

Afin d'obtenir la taille de toutes les composantes de notre graphe $ER_n(p)$, nous allons devoir utiliser une certaine vision des processus de branchement : les processus d'exploration. Nous commencerons par quelques rappels et autres énoncés sur les processus de branchement en général avant d'aller plus avant en *explorant* $ER_n(p)$.

Dans le cadre que nous nous sommes fixé pour l'étude du graphe d'Erdős-Rényi, les processus considérés sont tout simplement des processus de branchement de Galton-Watson avec loi de reproduction binomiale $\text{Bin}(n, p)$.

Définition 1.3 (Processus de Galton-Watson). *Un processus Z de Galton-Watson avec loi de reproduction μ est tel que si on se donne une famille de variables aléatoires i.i.d. $(X_{n,i})_{n,i \geq 1} \sim X$ de loi μ alors*

$$Z_0 = 1 \text{ et } \forall n \geq 1, Z_n = \sum_{i=1}^{Z_{n-1}} X_{n,i} \quad (1.3)$$

On désignera la probabilité d'extinction d'un processus de Galton-Watson Z par

$$\eta = \mathbb{P}(\exists n : Z_n = 0) \quad (1.4)$$

et la probabilité de survie par

$$\zeta = 1 - \eta \quad (1.5)$$

Pour un processus de branchement de Galton-Watson quelconque Z admettant un moment d'ordre 1 fini, on a le résultat suivant :

Théorème 1.4. *Considérons un processus de Galton-Watson Z de loi de reproduction μ en employant les notations de la définition 1.3.*

Si $\mathbb{E}[X] < 1$ alors $\eta = 1$.

Si $\mathbb{E}[X] > 1$ alors $\eta < 1$.

Si $\mathbb{E}[X] = 1$ et $\mathbb{P}(X = 1) < 1$ alors $\eta = 1$.

De plus, η est le plus petit point fixe dans $[0, 1]$ de la fonction génératrice G_x de X .

Nous allons maintenant énoncer quelques résultats bien connus donnant des bornes qui nous serviront souvent dans la suite :

Théorème 1.5 (Inégalité de Markov).

Soit X une variable aléatoire positive telle que $\mathbb{E}[X] < \infty$. Alors,

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}. \quad (1.6)$$

Théorème 1.6 (Inégalité de Chebychev). *Soit X une variable aléatoire intégrable. Alors,*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\text{Var}(X)}{a^2}. \quad (1.7)$$

Proposition 1.7 (Borne de Chernoff). *Soit X une variable aléatoire intégrable, $a \in \mathbb{R}$. Alors,*

$$\mathbb{P}(X \geq a) \leq \exp(-\phi^*(a)) \quad (1.8)$$

où

$$\phi^*(a) = \sup_{\lambda \geq 0} (\lambda a - \log(\mathbb{E}[e^{\lambda X}])) \quad (1.9)$$

Nous allons maintenant énoncer et démontrer la formule de Kemperman qui interviendra plus d'une fois dans la suite.

Soit $(Y_i)_{i \geq 1}$ une suite de variables aléatoires i.i.d. de loi à support dans $\{-1, 0, 1, 2, 3, \dots\}$. On note \mathbb{P}_k pour la loi d'une marche aléatoire partant de k . Soit $S_n = k + Y_1 + \dots + Y_n$ la position de la marche partant de k après n étapes. Enfin notons

$$H_0 = \inf \{n : S_n = 0\} \quad (1.10)$$

le premier temps d'atteinte de 0 de la marche aléatoire.

Théorème 1.8 (Théorème du Temps d'Atteinte ou Formule de Kemperman). *Avec les notations ci-dessus, on a*

$$\mathbb{P}_k(H_0 = n) = \frac{k}{n} \mathbb{P}_k(S_n = 0). \quad (1.11)$$

On démontrera ce théorème dans la partie qui suit.

Aussi, voici la formule de Stirling qui s'avère toujours bien utile.

Théorème 1.9 (Formule de Stirling).

$$n! \sim \frac{n^n}{e^n} \sqrt{2\pi n} \quad (1.12)$$

Voici maintenant quelques résultats sur les processus binomiaux et de Poisson :

Lemme 1.10 (Grande déviation pour la loi binomiale).

Soit X_n une variable aléatoire de loi binomiale de paramètres n et p . Alors, pour tout $a \in (p, 1]$,

$$\mathbb{P}(X_n \geq na) \leq e^{-nI(a)} \quad (1.13)$$

où

$$I(a) = a \log\left(\frac{a}{p}\right) + (1-a) \log\left(\frac{1-a}{1-p}\right) \quad (1.14)$$

De plus, en posant

$$I_p(a) = p - a - a \log(p/a), \quad (1.15)$$

on a

$$I(a) \geq I_p(a) \quad (1.16)$$

Proposition 1.11 (Loi de la progéniture totale d'un processus de branchement de Poisson). *La progéniture totale T^* d'un processus de branchement de Poisson de paramètre λ est telle que pour tout n ,*

$$\mathbb{P}_\lambda^*(T^* = n) = \frac{(\lambda n)^{n-1}}{n!} e^{-\lambda n}. \quad (1.17)$$

Théorème 1.12 (Processus de branchement de Poisson et binomial). *On considère un processus de branchement binomial de paramètres n et p et un processus de branchement de Poisson de paramètre $\lambda = np$. Notons T et T^* les progénitures totales respectives de ces processus. On a pour tout $k \geq 1$,*

$$\mathbb{P}_{n,p}(T \geq k) = \mathbb{P}_\lambda^*(T^* \geq k) + e_n(k), \quad (1.18)$$

où

$$|e_n(k)| \leq \frac{\lambda^2}{n} \sum_{s=1}^{k-1} \mathbb{P}_\lambda^*(T^* \geq s). \quad (1.19)$$

En particulier, $|e_n(k)| \leq k\lambda^2/n$.

1.3 Preuves des résultats essentiels

Preuve du théorème 1.4.

Notons

$$\eta_n = \mathbb{P}(Z_n = 0). \quad (1.20)$$

Par croissance des événements $(\{Z_n = 0\})_n$, on a $\eta_n \uparrow \eta$. Soit

$$\forall s, G_n(s) = \mathbb{E}[s^{Z_n}], \quad (1.21)$$

la fonction génératrice du nombre d'individus de la n -ième génération. Comme Z_n est une variable aléatoire entière, on a

$$\eta_n = G_n(0). \quad (1.22)$$

En conditionnant par rapport à Z_1 , on a

$$\forall s, G_n(s) = \mathbb{E}[s^{Z_n}] = \sum_{i=0}^{\infty} \mu(\{i\}) \mathbb{E}[s^{Z_n} | Z_1 = i] = \sum_{i=0}^{\infty} \mu(\{i\}) G_{n-1}(s)^i \quad (1.23)$$

Alors, en notant $G_X = G_1$ la fonction génératrice de $X_{1,1}$, il vient

$$\forall n \in \mathbb{N}^*, \forall s, G_n(s) = G_X(G_{n-1}(s)). \quad (1.24)$$

En particulier, pour $s = 0$, on obtient

$$\forall n \in \mathbb{N}^*, \eta_n = G_X(\eta_{n-1}) \quad (1.25)$$

En passant à la limite quand $n \rightarrow \infty$ et par continuité de G_X

$$\eta = G_X(\eta). \quad (1.26)$$

Si $\mathbb{P}(X = 1) = 1$ alors $Z_n = 1$ p.s. donc le résultat est immédiat. Si $\mathbb{P}(X \leq 1) = 1$ et $\mathbb{P}(X = 0) > 0$ alors $\mathbb{P}(Z_n = 0) = 1 - (1 - \mathbb{P}(Z_n = 0))^n \rightarrow 1$ et le résultat est immédiat. On se contentera donc dans la suite de la preuve de supposer que $\mathbb{P}(X \leq 1) < 1$.

Soit $\psi \in [0, 1]$ un autre point fixe de G_X et montrons que $\forall n \in \mathbb{N}, \eta_n \leq \psi$ donc $\eta \leq \psi$ en passant à la limite quand $n \rightarrow \infty$. Procédons par récurrence. On a déjà que $\eta_0 = 0 \leq \psi$. Puis si $\eta_{n-1} \leq \psi$ alors par croissance de G_X et comme ψ est un point fixe de G_X on a bien

$$\eta_n = G_X(\eta_{n-1}) \leq G_X(\psi) = \psi, \quad (1.27)$$

ce qui achève la récurrence. Donc η est le plus petit point fixe dans $[0, 1]$ de G_X .

On remarque que G_X est une fonction croissante et deux fois dérivable. On a d'après le théorème de dérivabilité sous le signe intégral $\forall s \geq 0$

$$G_X''(s) = \mathbb{E}[X(X-1)s^{X-2}] \geq 0, \quad (1.28)$$

ainsi G_X est convexe sur \mathbb{R}^+ .

De plus, quand $\mathbb{P}(X \leq 1) < 1$, G_X est strictement croissante et strictement convexe sur \mathbb{R}_+^* . Ainsi il y a au plus deux points fixes de G_X dans $[0, 1]$ et 1 est toujours un point fixe de G_X . Comme de plus $G_X(0) > 0$, il y a exactement un point fixe quand $G_X'(1) = \mathbb{E}[X] < 1$ donc $\eta = 1$ et deux quand $G_X'(1) > 1$ donc $\eta < 1$. Enfin, quand $G_X'(1) = 1$, si $\mathbb{P}(X = 1) < 1$, il y a exactement un point fixe. Ceci termine la preuve. \square

Preuve du théorème 1.5. L'inégalité 1.6 provient directement de

$$a\mathbb{P}(X \geq a) \leq \mathbb{E}[X\mathbf{1}_{\{X \geq a\}}] \leq \mathbb{E}[X]. \quad (1.29)$$

\square

Preuve du théorème 1.6. Il suffit de remarquer que

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) = \mathbb{P}((X - \mathbb{E}[X])^2 \geq a^2). \quad (1.30)$$

On conclut ensuite par l'inégalité de Markov 1.6. \square

Preuve de la proposition 1.7.

D'après Markov, on a :

$$\begin{aligned} \forall \lambda \in \mathbb{R}^+, \mathbb{P}(X \geq a) &= \mathbb{P}(e^{\lambda X} \geq e^{\lambda a}) \\ &\leq \mathbb{E}[e^{\lambda X}]e^{-\lambda a} \\ &\leq \inf_{\lambda \geq 0} \exp(\log(\mathbb{E}[e^{\lambda X}]) - \lambda a) \\ &\leq \exp\left(\inf_{\lambda \geq 0} (\log(\mathbb{E}[e^{\lambda X}]) - \lambda a)\right) \\ &\leq \exp(-\phi^*(a)) \end{aligned}$$

\square

Pour démontrer le théorème 1.8, nous aurons besoin du lemme suivant (cf. [2] page 122) :

Lemme 1.13 (Temps d'atteinte par rotations cycliques). *Soit $Y = (Y_1, \dots, Y_n)$ une suite de variables aléatoires de loi à support dans $\{-1, 0, 1, 2, 3, \dots\}$ telle que $\sum_{i=1}^n Y_i = -k$ pour un certain $k \in \{1, \dots, n\}$. Alors il existe exactement k termes dans $\{1, \dots, n\}$ notés i_1, \dots, i_k tels que n soit le Premier Temps d'Atteinte de $-k$ pour $Y^{(i_j)} \forall j \in \{1, \dots, k\}$.*

Preuve du théorème 1.8.

On note, pour tout $i \in \mathbb{N}$ et $l \in \{1, \dots, n\}$, $Y_l^{(i)} = Y_{l+i} \mathbb{1}_{i+l \leq n} + Y_{l+i-n} \mathbb{1}_{i+l > n}$.

Aussi, on note $S_m^{(i)} = k + Y_1^{(i)} + \dots + Y_m^{(i)}$.

On a

$$\begin{aligned} \mathbb{P}_k(H_0 = n) &= \frac{1}{n} \sum_{i=1}^n \mathbb{P}_k(H_0^{(i)} = n) \\ &= \frac{1}{n} \mathbb{E}_k[\sum_{i=1}^n \mathbb{1}_{H_0^{(i)} = n}] \\ \text{(d'après le lemme 1.13)} &= \frac{1}{n} \mathbb{E}_k[k \cdot \mathbb{1}_{S_n = 0}]. \end{aligned}$$

On a donc

$$\mathbb{P}_k(H_0 = n) = \frac{k}{n} \mathbb{P}_k(S_n = 0) \quad (1.31)$$

ce qui démontre le théorème 1.8. \square

Preuve du lemme 1.13. Considérons $s_k = \sum_{i=1}^k Y_i$ et notons $s_k^{(l)} = \sum_{i=1}^k x_{l+i}$.

Alors notons $i = m$ le plus petit i tel que $s_i = \min_{1 \leq j \leq n} s_j$

Tout d'abord, remplaçons Y par $Y^{(m)} = (Y_{m+1}, \dots, Y_n, Y_1, \dots, Y_m)$.

On a donc que n est le Premier temps d'Atteinte de $-k$ pour $Y^{(m)}$.

En effet, supposons que $\exists j < n$ tel que $s_j^{(m)} = -k$. Alors supposons dans un premier temps que $j \leq n - m$.

Dans ce cas :

$$\begin{aligned} s_j^{(m)} &= s_n - s_m - (s_n - s_{m+j}) \\ &= s_{m+j} - s_m \\ &\geq 0 \end{aligned}$$

car $s_m = \min_{1 \leq j \leq n} s_j$.

Ceci est une contradiction car $s_j^{(m)} = -k$.

Supposons maintenant que $j > n - m$.

Alors

$$\begin{aligned} s_j^{(m)} &= s_n - s_m + s_{j-n+m} \\ &= -k - s_m + s_{j-n+m} \\ &> -k \end{aligned}$$

car m est le plus petit indice tel que $s_m = \min_{1 \leq j \leq n} s_j$ or $j - n + m < m$ donc $s_m < s_{j-n+m}$. Ceci est aussi une contradiction car $s_j^{(m)} = -k$.

Donc n est bien le Premier temps d'Atteinte de $-k$ pour $Y^{(m)}$.

Montrons maintenant qu'il y a exactement k rotations cycliques telles que le Premier temps d'Atteinte de $-k$ est n . Soit $Y = (Y_1, \dots, Y_n)$ une suite telle que n soit le Premier temps d'Atteinte de $-k$. Notons $\alpha_1, \dots, \alpha_k$ les Premiers temps d'Atteinte respectifs de $-1, \dots, -k$. Nous avons par conséquent $n = \alpha_k$.

Montrons que pour tout $1 \leq t < \alpha_1$, le Premier temps d'Atteinte de $-k$ pour $Y_{(t)}$ est plus petit que n mais que n est le Premier temps d'Atteinte de $-k$ pour $Y^{(\alpha_1)}$.

Soit $1 \leq t < \alpha_1$. On a $s_t \geq 0$ d'où $s_{n-t}^{(t)} = s_n - s_t \leq -k$. Donc $s^{(t)}$ atteint $-k$ avant le temps n . On a bien que pour tout $1 \leq t < \alpha_1$, n n'est pas le Premier temps d'Atteinte de $-k$ pour $Y^{(t)}$.

Soit $m \leq n - \alpha_1$. Alors

$$s_m^{(\alpha_1)} = s_{\alpha_1+m} - s_{\alpha_1} \geq -k + 1 \quad (1.32)$$

car $\alpha_1 + m \leq n$ donc $s_{\alpha_1+m} \geq -k$ et $s_{\alpha_1} = -1$.

Soit $n - \alpha_1 \leq m < n$. Alors

$$s_m^{(\alpha_1)} = s_{n-\alpha_1}^{(\alpha_1)} + s_{m-(n-\alpha_1)} \geq -k + 1 \quad (1.33)$$

car $s_{n-\alpha_1}^{(\alpha_1)} = -k + 1$ et $s_{m-(n-\alpha_1)} \geq 0$.

On a donc que pour tout $m < n$, $s_m^{(\alpha_1)} \geq -k + 1$. Donc n est bien le Premier temps d'Atteinte de $-k$ pour $Y^{(\alpha_1)}$.

Maintenant il suffit de réitérer le procédé en montrant que pour tout $1 \leq i \leq k$, n est le Premier temps d'Atteinte de $-k$ pour $Y^{(\alpha_i)}$ et que pour tout $\alpha_i < m < \alpha_{i+1}$, n n'est pas le Premier temps d'Atteinte de $-k$ pour $Y^{(m)}$.

On a donc bien qu'il existe exactement k rotations cycliques (qui sont les rotations de α_i pour tout $1 \leq i \leq k$) telles que n soit le Premier temps d'Atteinte de $-k$ pour la suite.

□

Preuve du lemme 1.10.

En reprenant la preuve de la proposition 1.7, et avec $X_1 \sim \text{Ber}(p)$,

$$\begin{aligned} I(a) &= \sup_{t \geq 0} (ta - \log(\mathbb{E}[e^{tX_1}])) = \sup_{t \geq 0} (ta - \log(pe^t + (1-p))) \\ &= a \log\left(\frac{a}{p}\right) + (1-a) \log\left(\frac{1-a}{1-p}\right). \end{aligned} \quad (1.34)$$

On remarque que pour $t \geq 0$,

$$pe^t + (1-p) = 1 + p(e^t - 1) \leq e^{p(e^t - 1)}, \quad (1.35)$$

donc

$$I(a) \geq \sup_{t \geq 0} (ta - p(e^t - 1)) = p - a - a \log(p/a) = I_p(a). \quad (1.36)$$

On conclut par Chernoff (proposition 1.7). \square

Preuve de la proposition 1.11.

Cette proposition découle du théorème 1.8 en prenant $k = 1$ et $\forall i, Y_i = X_i^* - 1$. On a

$$\begin{aligned} \mathbb{P}(T^* = n) &= \mathbb{P}_1(H_0 = n) \\ &= \frac{1}{n} \mathbb{P}_1(S_n = 0) \\ &= \frac{1}{n} \mathbb{P}_1(1 + Y_1 + \dots + Y_n = 0) \\ &= \frac{1}{n} \mathbb{P}(1 + X_1^* + \dots + X_n^* - n = 0) \\ &= \frac{1}{n} \mathbb{P}(X_1^* + \dots + X_n^* = n - 1). \end{aligned} \quad (1.37)$$

Or, les X_i^* étant des variables i.i.d. de loi de Poisson de paramètre λ , leur somme suit une loi de Poisson de paramètre λn . Donc

$$\mathbb{P}(T^* = n) = \frac{1}{n} \frac{(\lambda n)^{n-1}}{(n-1)!} e^{-\lambda n} = \frac{(\lambda n)^{n-1}}{n!} e^{-\lambda n}. \quad (1.38)$$

\square

Afin de prouver le théorème 1.12, énonçons la proposition suivante, que l'on démontrera un peu plus loin.

Proposition 1.14 (Limite de Poisson pour des variables binomiales). *De manière générale, soient des variables aléatoires indépendantes $(I_i)_{1 \leq i \leq n}$ telles que $I_i \sim \text{Ber}(p_i)$ et $\lambda = \sum_{i=1}^n p_i$. Soit*

$X = \sum_{i=1}^n I_i$ et Y une variable aléatoire de loi de Poisson de paramètre λ . Il existe alors un couplage (\hat{X}, \hat{Y}) de X et Y tel que

$$\mathbb{P}(\hat{X} \neq \hat{Y}) \leq \sum_{i=1}^n p_i^2. \quad (1.39)$$

En particulier, cela s'applique lorsque X suit une loi binomiale de paramètres n et λ/n . En ce cas :

$$\mathbb{P}(\hat{X} \neq \hat{Y}) \leq \frac{\lambda^2}{n}. \quad (1.40)$$

Preuve du théorème 1.12.

Nous allons faire une preuve par couplage grâce aux arguments de la proposition 1.14 (prouvée ci-après). On décrit les processus de branchement par leur loi de reproduction, à savoir des variables binomiales et de Poisson respectivement. On applique le couplage de la proposition 1.14 pour chacune des variables X_i et X_i^* qui caractérisent les processus de branchement, avec $X_i \sim \text{Bin}(n, \lambda/n)$, $X_i^* \sim \text{Poi}(\lambda)$ et

$$\mathbb{P}(X_i \neq X_i^*) \leq \frac{\lambda^2}{n}. \quad (1.41)$$

Les vecteurs $(X_i, X_i^*)_{1 \leq i}$ sont i.i.d.. \mathbb{P} désignera la loi jointe du couplage obtenu ci-dessus.

On observe en premier lieu que

$$\mathbb{P}_{n,p}(T \geq k) = \mathbb{P}(T \geq k, T^* \geq k) + \mathbb{P}(T \geq k, T^* < k), \quad (1.42)$$

et

$$\mathbb{P}_\lambda^*(T^* \geq k) = \mathbb{P}(T \geq k, T^* \geq k) + \mathbb{P}(T < k, T^* \geq k). \quad (1.43)$$

En soustrayant, il vient

$$|\mathbb{P}_{n,p}(T \geq k) - \mathbb{P}_\lambda^*(T^* \geq k)| \leq \max\{\mathbb{P}(T \geq k, T^* < k), \mathbb{P}(T < k, T^* \geq k)\}. \quad (1.44)$$

On va alors appliquer la proposition 1.14 et le fait que $\{T \geq k\}$ ne dépend que de X_1, \dots, X_{k-1} . Ainsi T est un temps d'arrêt pour la filtration $\sigma(X_1, \dots, X_k)$, de même, T^* est un temps d'arrêt pour la filtration $\sigma(X_1^*, \dots, X_k^*)$.

Quand $T \geq k$ et $T^* < k$, ou quand $T < k$ et $T^* \geq k$, il existe nécessairement un $s < k$ tel que $X_s \neq X_s^*$. On obtient alors une borne en considérant le premier tel s :

$$\mathbb{P}(T \geq k, T^* < k) \leq \sum_{s=1}^{k-1} \mathbb{P}(X_i = X_i^* \forall i \leq s-1, X_s \neq X_s^*, T \geq k). \quad (1.45)$$

On observe que quand $X_i = X_i^*$ pour tout $i \leq s-1$ et $T \geq k$ alors pour tout $i \leq s-1$, $X_1^* + \dots + X_i^* \geq i$ d'où $T^* \geq s$. De plus, T^* étant un temps d'arrêt pour $\sigma(X_1^*, \dots, X_k^*)$, $\{T^* \geq s\}$ est indépendant de $\{X_s \neq X_s^*\}$. On obtient alors

$$\begin{aligned} \mathbb{P}(T \geq k, T^* < k) &\leq \sum_{s=1}^{k-1} \mathbb{P}(T^* \geq s, X_s \neq X_s^*) \\ &= \sum_{s=1}^{k-1} \mathbb{P}(T^* \geq s) \mathbb{P}(X_s \neq X_s^*). \end{aligned} \quad (1.46)$$

D'après la proposition 1.14, on a donc

$$\mathbb{P}(T \geq k, T^* < k) \leq \frac{\lambda^2}{n} \sum_{s=1}^{k-1} \mathbb{P}(T^* \geq s). \quad (1.47)$$

Par un raisonnement similaire, il vient

$$\begin{aligned} \mathbb{P}(T < k, T^* \geq k) &\leq \sum_{s=1}^{k-1} \mathbb{P}(T^* \geq s) \mathbb{P}(X_s \neq X_s^*) \\ &\leq \frac{\lambda^2}{n} \sum_{s=1}^{k-1} \mathbb{P}(T^* \geq s). \end{aligned} \quad (1.48)$$

On applique ceci à (1.44)

$$|\mathbb{P}_{n,p}(T \geq k) - \mathbb{P}_\lambda^*(T^* \geq k)| \leq \frac{\lambda^2}{n} \sum_{s=1}^{k-1} \mathbb{P}_\lambda^*(T^* \geq s). \quad (1.49)$$

Ceci achève la preuve. □

Preuve de la proposition 1.14.

Soient $(I_i)_{1 \leq i \leq n}$ des variables indépendantes de loi $Ber(p_i)$ et $(J_i)_{1 \leq i \leq n}$ des variables indépendantes de loi $Poi(p_i)$. On abrégera

$$p_{i,x} = \mathbb{P}(I_i = x) \text{ et } q_{i,x} = \mathbb{P}(J_i = x). \quad (1.50)$$

Soit (\hat{I}_i, \hat{J}_i) un couplage maximal (i.-e. maximisant $\mathbb{P}(\hat{I}_i = \hat{J}_i)$) de I_i et J_i , $(\hat{I}_i, \hat{J}_i)_i$ indépendants. Voir [1] page 59-60 pour de plus amples explications sur ce couplage *maximal*. On a

$$\begin{aligned} \mathbb{P}(\hat{I}_i = \hat{J}_i = x) &= \min \{p_{i,x}, q_{i,x}\} \\ &= (1 - p_i) \mathbb{1}_{\{x=0\}} + p_i e^{-p_i} \mathbb{1}_{\{x=1\}} \end{aligned} \quad (1.51)$$

car pour $x = 0$, $1 - p_i \leq e^{-p_i}$. On va maintenant utiliser le fait que $1 - e^{-p_i} \leq p_i$:

$$\mathbb{P}(\hat{I}_i \neq \hat{J}_i) = 1 - \mathbb{P}(\hat{I}_i = \hat{J}_i) = 1 - (1 - p_i) - p_i e^{-p_i} = p_i(1 - e^{-p_i}) \leq p_i^2. \quad (1.52)$$

On pose ensuite $\hat{X} = \sum_{i=1}^n \hat{I}_i$ et $\hat{Y} = \sum_{i=1}^n \hat{J}_i$. \hat{X} a même loi que X et \hat{Y} a même loi que $Y \sim Poi(p_1 + \dots + p_n)$. Par σ -additivité et l'inégalité précédente, il vient

$$\mathbb{P}(\hat{X} \neq \hat{Y}) \leq \mathbb{P}\left(\bigcup_{i=1}^n \{\hat{I}_i \neq \hat{J}_i\}\right) \leq \sum_{i=1}^n \mathbb{P}(\hat{I}_i \neq \hat{J}_i) \leq \sum_{i=1}^n p_i^2. \quad (1.53)$$

□

1.4 Exploration des composantes du graphe

On s'inspirera ici de [4].

Nous sommes maintenant en mesure de décrire le processus d'exploration dans le graphe $ER_n(p)$ qui, on le verra plus loin, pourra être couplé au processus de branchement dans l'arbre binomial de paramètres n et p . Il y a en fait deux processus d'exploration possibles : le Breadth-First Search (BFS) et le Depth-First Search (DFS), ou en français recherche en largeur et recherche en profondeur. Les deux processus sont assez similaires, à ceci près que le premier repose sur une structure de liste (First In First Out) et le second sur une structure de pile (First In Last Out).

Dans les deux cas, on commence par définir 3 états possibles pour les sommets du graphe lors de son exploration : *inactif*, *actif* et *neutre*. On désignera par I_t , A_t et N_t les ensembles des sommets respectivement inactifs (explorés), actifs et neutres à l'étape $t \in \mathbb{N}$ de l'exploration. Ces trois ensembles sont toujours disjoints et recouvrent $[n]$.

Considérons maintenant le cas DFS et explorons la composante d'un sommet $v \in [n]$, par exemple le sommet 1. Ici, A_t sera vu comme une pile. Au temps $t = 0$, I_0 est vide, $A_0 = \{1\}$ et $N_0 = \{2, \dots, n\}$. Pour passer de l'étape t à l'étape $t + 1$, on considère le dernier sommet v_t ajouté à A_t , appelé sommet courant, et on teste pour des voisins dans N_t . Si v_t admet des voisins dans N_t , alors on les déplace dans A_t et on passe à l'étape suivante, sinon on déplace v_t dans I_t . On ne passe à l'étape $t + 1$ que quand I_t comporte t sommets. L'exploration de la composante s'arrête quand A_t est vide, elle est alors de taille $T = \inf\{t : |A_t| = 0\}$ (T est un temps d'arrêt).

Remarque. Il est important d'observer qu'à la fin de l'étape t , juste avant de passer à l'étape $t + 1$, I_t comporte t sommets et on est assuré que la composante contient au moins $t + |A_t|$ sommets. Dans la suite, on notera souvent $S_t = |A_t|$ à la fin de l'étape t .

Pour explorer le graphe entièrement avec la méthode DFS, il suffit de prendre un nouveau sommet neutre (par exemple le plus petit restant pour l'ordre défini sur $[n]$) et de le placer dans A_t , on explore alors la composante de ce sommet et on recommence jusqu'à ce que A_t et N_t soient tous les deux vides. On explore ainsi une à une chaque composante connexe du graphe en profondeur. C'est-à-dire que quand un sommet a plusieurs enfants, on regarde une par une la descendance de chaque enfant plutôt que de faire un comptage génération par génération.

La méthode BFS explore elle aussi la composante connexe d'un sommet v donné mais en regardant ce qui se passe génération par génération : i.-e. on explore entièrement l'ensemble des sommets à distance de graphe k de v avant de regarder les sommets plus éloignés. L'initialisation est la même que dans le cas DFS ; ensuite, pour passer de l'étape t à $t + 1$, on regarde les voisins dans N_t du premier élément v_t de la liste A_t , on les ajoute à la queue de la liste A_t , puis on enlève v_t de A_t pour le placer dans I_t . Ainsi, là encore I_t contient exactement t sommets à la fin de la t -ième étape et la composante de v est de taille $T = \inf\{t : |A_t| = 0\}$.

Pour explorer tout le graphe, on relance l'algorithme BFS sur un nouveau sommet neutre jusqu'à épuiser tous les sommets.

Dans les deux cas, quand on dira à l'avenir "à l'étape t ", on sous-entendra en fait "à la fin de l'étape t juste avant de passer à $t + 1$ ".

Le choix de la méthode DFS ou BFS n'importe pas car la donnée intéressante est le nombre de sommets actifs à chaque étape et il est le même peu importe la méthode. Dans la suite, quand nous ferons référence à l'exploration du graphe, nous ne préférons donc pas une méthode à l'autre, à l'exception des simulations où il a fallu faire un choix pour le codage de l'exploration.

1.5 Probabilité de survie des processus de branchement

On suivra [3].

Pour la démonstration de la taille de la composante maximale dans la fenêtre critique, nous aurons besoin de résultats sur les processus de branchement de Galton-Watson avec une distribution Z sur la progéniture d'un arbre construit comme suit : Nous avons un unique sommet à la génération 0. Chaque sommet de la génération t a un nombre aléatoire d'enfants à la génération $t+1$ avec la distribution Z . Le nombre d'enfants de chaque élément est indépendant des autres éléments de l'arbre. On sait que si $\mathbb{E}[Z] > 1$ alors la probabilité de survie q est l'unique solution dans $(0,1]$ de $1 - q = g_Z(1 - q)$ où g_Z est la fonction génératrice de Z . Lorsque $\mathbb{E}[Z] < 1$ l'espérance du nombre total de sommets dans le processus de branchement est

$$1 + \mathbb{E}[Z] + \mathbb{E}[Z]^2 + \dots = \frac{1}{1 - \mathbb{E}[Z]}, \quad (1.54)$$

et en particulier, la probabilité de survie du processus est 0. Notons maintenant $\mathcal{T}_{n,p}$ le processus de branchement binomial de paramètres n et p , c'est à dire un processus de branchement où la progéniture suit une loi $\text{Bin}(n,p)$. Alors la fonction génératrice $g_{\text{Bin}(n,p)}$ satisfait à

$$g_{\text{Bin}(n,p)}(x) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} x^k = (1 - p(1-x))^n,$$

lorsque $np > 1$ la probabilité de survie $q = q_{n,p}$ satisfait à

$$1 - q = (1 - pq)^n.$$

On vérifie alors que si $\varepsilon = np - 1 \sim 0$ avec $\varepsilon > 0$ alors

$$q \sim 2\varepsilon \quad (1.55)$$

De même, la probabilité de survie ζ_λ d'un processus de branchement de Poisson de paramètre λ est équivalent à 2ϵ quand $\lambda = 1 + \epsilon$ ou $\lambda = 1 - \epsilon$ et $0 < \epsilon = \epsilon(n) \xrightarrow{n \rightarrow +\infty} 0$.

Démontrons ce dernier résultat :

la probabilité d'extinction η_λ d'un processus de branchement de Poisson de paramètre λ est l'unique point fixe dans $[0, 1[$ de la fonction génératrice d'où

$$\begin{aligned}\eta_\lambda &= e^{-\lambda(1-\eta_\lambda)} \\ 1 - \zeta_\lambda &= e^{-\lambda\zeta_\lambda}.\end{aligned}$$

Or quand $\lambda \rightarrow 1$, $\zeta_\lambda \rightarrow 0$ donc on peut faire un développement limité de l'exponentielle :

$$\begin{aligned}1 - \zeta_\lambda &= 1 - \lambda\zeta_\lambda + \frac{\lambda^2\zeta_\lambda^2}{2} + o(\lambda^2\zeta_\lambda^2) \\ \lambda - 1 &= \frac{\lambda^2\zeta_\lambda}{2}(1 + o(1)) \\ \epsilon &= \frac{\lambda^2\zeta_\lambda}{2}(1 + o(1))\end{aligned}$$

et comme $\lambda \rightarrow 1$, on a bien

$$\zeta_\lambda \sim 2\epsilon \tag{1.56}$$

Chapitre 2

Régime sur-critique

2.1 Existence et unicité de la composante géante

On suivra [1].

On note $\zeta_\lambda = 1 - \eta_\lambda$ la probabilité de survie d'un processus de branchement de Poisson de paramètre λ . Dans tout ce chapitre, on prendra $\lambda = \lambda(n) = 1 + \epsilon(n) > 1$ tel que $\epsilon \gg n^{-1/3}(\log(n))^{1/2}$, avec la notation $f \gg g$ pour $\frac{f}{g} \rightarrow \infty$ quand $n \rightarrow \infty$.

Théorème 2.1 (Loi des grands nombres pour la composante géante).

$$\frac{|C_{max}|}{n\zeta_\lambda} \xrightarrow{\mathbb{P}} 1 \tag{2.1}$$

Un sommet quelconque du graphe ayant une probabilité ζ_λ d'être dans une grande composante (i-e de taille maximale), le nombre de sommets dans une grande composante est de l'ordre de $\zeta_\lambda n$. Le Théorème 2.1 implique alors que tous les sommets qui appartiennent à des grandes composantes sont en fait dans une *unique* composante connexe, que l'on appelle la *composante géante*.

Avant de prouver le théorème 2.1, crucial pour l'étude du régime sur-critique, nous allons énoncer les propositions (temporairement admises) suivantes qui représentent les points essentiels de la démonstration.

Posons

$$I_\lambda = \lambda - 1 - \log(\lambda). \tag{2.2}$$

On remarquera que

$$I_\lambda \sim \frac{\epsilon^2}{2} \quad (n \rightarrow \infty) \tag{2.3}$$

si $\epsilon(n) \rightarrow 0$.

Soit

$$J(\alpha, \lambda) = I_{(1-e^{-\lambda\alpha})/\alpha} \quad (2.4)$$

On remarquera que quand $\epsilon \rightarrow 0$, comme $\zeta_\lambda \sim 2\epsilon$, en prenant $\alpha = c'\zeta$, $c' < 1$,

$$\begin{aligned} J(\alpha, \lambda) &= \frac{1}{2} \left(\frac{1 - e^{-\lambda\alpha}}{\alpha} - 1 \right)^2 + o\left(\frac{1 - e^{-\lambda\alpha}}{\alpha} - 1 \right) \\ &\sim \frac{(1 - c')^2 \epsilon^2}{2} \end{aligned} \quad (2.5)$$

Théorème 2.2 (Il n'y a pas de composante intermédiaire). Prenons $k_n = \frac{c}{I_\lambda} \log(n)$ et $\alpha = c'\zeta_\lambda$ où $c > 1$ et $c' < 1$ sont tels que $c(1 - c')^2 \geq 1$, et posons $\delta = 1 - \frac{c}{I_\lambda} J(\alpha; \lambda)$ où $J(\alpha; \lambda)$ est défini en (2.4). Alors il existe $\delta' < \delta$ qui ne dépend pas de λ tel qu'il n'y a pas de composante connexe de taille comprise entre k_n et αn avec probabilité supérieure à $1 - O(n^{-\delta'})$.

Remarque. Cela implique donc que, en notant \mathcal{C}_2 la deuxième plus grande composante du graphe, $\mathbb{P}_\lambda(k_n \leq |\mathcal{C}_2|) \rightarrow 0$ quand $n \rightarrow \infty$.

Lemme 2.3 (Couplage). (1) Pour tous n et p , les arbres aléatoires \mathcal{T}_v et $\mathcal{T}_{n,p}$ correspondant respectivement au processus d'exploration de $\mathcal{C}(v)$ et à l'arbre binomial de paramètres n et p peuvent être couplés de sorte que $\mathcal{T}_v \subset \mathcal{T}_{n,p}$.

(2) Pour tous n, k et p , il existe un couplage entre les variables entières $|\mathcal{C}(v)|$ et $|\mathcal{T}_{n-k,p}|$ de telle sorte que soit $|\mathcal{C}(v)| \geq |\mathcal{T}_{n-k,p}|$, soit les deux sont de taille au moins k .

Théorème 2.4 (Borne inférieure pour la queue de la composante connexe). Pour tout $k \in [n]$,

$$\mathbb{P}_{n,p}(|\mathcal{C}(1)| \geq k) \geq \mathbb{P}_{n-k,p}(T^\leq \geq k), \quad (2.6)$$

où T^\leq est la progéniture totale d'un processus de branchement avec distribution binômiale de paramètres $n - k$ et $p = \frac{\lambda}{n}$.

Commençons par montrer que $|\mathcal{C}(v)| \geq k_n$, pour $k_n = \frac{c}{I_\lambda} \log(n)$ avec $c > 1$, est proche de la probabilité de survie d'un processus de branchement de Poisson de paramètre λ .

Proposition 2.5 (La queue de la composante est la probabilité de survie du processus de branchement). Soit $c > 1$, posons $k_n = \frac{c}{I_\lambda} \log(n)$. Pour n assez grand, on a :

$$\mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k_n) = \zeta_\lambda + O(k_n/n) \quad (2.7)$$

Notons

$$\chi_{<k}(\lambda) = \mathbb{E}[|\mathcal{C}(v)| \mathbf{1}_{\{|\mathcal{C}(v)| < k\}}] \quad (2.8)$$

et remarquons que cette quantité ne dépend pas du sommet v choisi.

Définissons

$$Z_{\geq k} = \sum_{v \in [n]} \mathbb{1}_{\{|\mathcal{C}(v)| \geq k\}}. \quad (2.9)$$

Un argument de second moment va être nécessaire dans la preuve du théorème 2.1, c'est pourquoi il nous faut contrôler la variance de $Z_{\geq k}$:

Proposition 2.6 (Une estimation sur la variance de $Z_{\geq k}$). *Pour tout n et $k \in [n]$,*

$$\text{Var}_\lambda(Z_{\geq k}) \leq (\lambda k + 1)n\chi_{<k}(\lambda) \quad (2.10)$$

Corollaire 2.7 (Concentration du nombre de sommets dans les grandes composantes). *Soit $c > 1$, posons $k_n = \frac{c}{I_\lambda} \log(n)$ et soit $\nu \in (\frac{2}{3}, 1)$. Alors pour chaque $\delta < 2\nu - 1$, pour n grand qui ne dépend pas de λ ,*

$$\mathbb{P}_\lambda(|Z_{\geq k_n} - n\zeta_\lambda| > n^\nu) = O(n^{-\delta}) \quad (2.11)$$

Proposition 2.8 (Borne exponentielle pour les composantes sur-critiques de taille inférieure à $\zeta_\lambda n$). *Soit $c > 1$, $c' < 1$, tels que $c(1 - c')^2 \geq \frac{2}{3}$. Posons $k_n = \frac{c}{I_\lambda} \log(n)$ et $\alpha = c'\zeta_\lambda$. Alors, pour $n \geq 2\lambda$,*

$$\mathbb{P}_\lambda(k_n \leq |\mathcal{C}(v)| \leq \alpha n) \leq e^{-k_n J(\alpha, \lambda)} / [1 - e^{-J(\alpha, \lambda)}]. \quad (2.12)$$

De plus, si $\epsilon \rightarrow 0$ quand $n \rightarrow \infty$, alors la borne tend vers 0 plus vite que $\frac{n^{2/3 - c(1 - c')^2}}{\log(n)}$.

Pour la suite, fixons $\nu \in [\frac{2}{3}, 1]$, $\alpha = c'\zeta_\lambda$ avec $c' \in [\frac{1}{2}, 1[$ et prenons $k_n = \frac{c}{I_\lambda} \log(n)$ avec $c > 1$, de telle sorte que $c(1 - c')^2 \geq 1$. Soit \mathcal{E}_n l'évènement tel que :

- (1) $|Z_{\geq k_n} - n\zeta_\lambda| \leq n^\nu$
- (2) Il n'existe pas de $v \in [n]$ tel que $k_n \leq |\mathcal{C}(v)| \leq \alpha n$.

Lemme 2.9 (La taille de la composante géante est égale à $Z_{\geq k_n}$). *L'évènement \mathcal{E}_n a lieu avec grande probabilité, i.e., si δ_0 est le minimum entre δ' du Théorème 2.2 et δ du Corollaire 2.7, alors $\mathbb{P}_\lambda(\mathcal{E}_n^c) = O(n^{-\delta_0})$. De plus, $|\mathcal{C}_{\max}| = Z_{\geq k_n}$.*

Soit S_t le nombre de sommets actifs dans le Graphe d'Erdős-Rényi. Posons, comme pour le processus de branchement,

$$S_0 = 1, \quad S_t = S_{t-1} + X_t - 1, \quad (2.13)$$

où X_t est le nombre de sommets qui deviennent actifs à la $t^{\text{ième}}$ étape après l'exploration du sommet v_t , celui-ci devenant inactif.

Nous avons donc à la $t^{\text{ième}}$ étape dans un graphe d'Erdős-Rényi à n sommets, S_t sommets actifs, t sommets inactifs et $n - t - S_t$ sommets neutres.

Chaque couple de sommets ayant une probabilité p d'être relié, on a donc :

$$X_t \sim \text{Bin}(n - (t - 1) - S_{t-1}, p). \quad (2.14)$$

Proposition 2.10 (La loi de S_t). *Pour tout $t \in [n]$, et $(S_t)_{t \geq 0}$ satisfaisant les relations 2.13 et 2.14, on a :*

$$S_t + (t - 1) \sim \text{Bin}(n - 1, 1 - (1 - p)^t). \quad (2.15)$$

On restreindra l'usage de cette proposition au cas où $|\mathcal{C}(v)| \geq t$, car on ne s'intéresse pas à ce qu'il advient une fois l'exploration terminée.

Proposition 2.11 (Unicité de la solution dans l'équation de la survie de la probabilité de Poisson). *L'unique solution dans $(0,1]$ de l'équation $1 - e^{-\lambda\alpha} = \alpha$ est $\alpha = \zeta_\lambda$ où ζ_λ est la probabilité de survie pour un processus de Poisson de paramètre λ .*

2.2 Preuve des résultats sur-critiques

Preuve du théorème 2.1.

Ce théorème se montre en quelques étapes découlant chacune des propositions et autres corollaires énoncés plus hauts. Tout repose ici sur $Z_{\geq k}$, défini en (2.9), le nombre de sommets dans les composantes de taille au moins k .

Soient $c > 1$ et $\frac{1}{2} \leq c' < 1$ tels que $c(1 - c')^2 \geq 1$. Dans un premier temps, on prend $k_n = \frac{c}{I_\lambda} \log(n)$ et on calcule

$$\mathbb{E}_\lambda[Z_{\geq k_n}] = n\mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k_n). \quad (2.16)$$

On évalue alors $\mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k_n)$ à l'aide la proposition 2.5 :

$$\begin{aligned} \mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k_n) &= \zeta_\lambda + O(k_n/n) \\ &= \zeta_\lambda \left(1 + O\left(\frac{c \log(n)}{I_\lambda n}\right)\right) \\ &= \zeta_\lambda (1 + o(1)) \end{aligned} \quad (2.17)$$

car $\frac{\log(n)}{I_\lambda n} \ll n^{-1/3}$.

Dans un deuxième temps, on utilise le corollaire 2.7 pour contrôler l'écart entre $Z_{\geq k_n}$ et son espérance :

$$\forall \nu \in \left(\frac{2}{3}, 1\right), \forall \delta < 2\nu - 1, \mathbb{P}_\lambda(|Z_{\geq k_n} - n\zeta_\lambda| > n^\nu) = O(n^{-\delta}). \quad (2.18)$$

Cette quantité tend donc vers 0 quand $n \rightarrow +\infty$ On a donc avec grande probabilité pour

$$\forall v \in \left(\frac{2}{3}, 1\right), |Z_{\geq k_n} - n\zeta_\lambda| \leq n^\nu, \quad (2.19)$$

or $\sum_{v \in [n]} \mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k_n) = \mathbb{E}_\lambda[Z_{\geq k_n}] = n\zeta_\lambda(1 + o(1))$, donc, avec grande probabilité :

$$\forall v \in \left(\frac{2}{3}, 1\right), |Z_{\geq k_n} - \mathbb{E}_\lambda[Z_{\geq k_n}]| \leq n^\nu. \quad (2.20)$$

Dans un troisième temps, on montre à l'aide du théorème 2.2 qu'avec grande probabilité il n'y a pas de composante de taille intermédiaire, i.e. pour $k_n = \frac{c}{l_\lambda} \log(n)$, $\alpha = c'\zeta_\lambda$ avec $c > 1$, $c' \in [\frac{1}{2}, 1[$ tels que $c(1 - c')^2 \geq 1$, il n'y a pas de composante connexe de taille comprise entre k_n et αn .

Dans un quatrième temps, on applique le lemme 2.9, c'est-à-dire que conditionnellement à l'événement \mathcal{E}_n défini dans ce lemme,

$$Z_{\geq k_n} = |\mathcal{C}_{max}|. \quad (2.21)$$

Ainsi, en combinant les résultats obtenus à chacune des quatre étapes de la preuve, on obtient bien le résultat du théorème 2.1. \square

Preuve du lemme 2.3.

(1) Notons

$$S_0 = 1 \text{ et } \forall t \in \mathbb{N}^*, S_t = S_{t-1} + \sum_{i=1}^{n-(t-1)-S_{t-1}} B_{t,i} - 1 \quad (2.22)$$

le nombre de sommets actifs dans le processus d'exploration où les $B_{t,i}$ sont des variables aléatoires de Bernoulli de paramètre p codant pour l'existence d'une arête entre le sommet courant v_t du processus d'exploration et le i -ème sommet neutre à l'étape t . Les $B_{t,i}$ sont indépendantes car l'existence d'une arête entre v_t et le i -ème sommet neutre à l'étape t ne dépend ni des autres sommets actifs ou inactifs, ni des autres sommets neutres. On peut donc coupler ce processus au processus d'exploration de l'arbre $\mathcal{T}_{n,p}$ et le résultat est alors immédiat.

(2) D'après le point précédent, on peut également coupler S_t avec $S_t^{(k)}$, nombre de sommets actifs dans l'exploration de $\mathcal{T}_{n-k,p}$:

$$S_0^{(k)} = 1 \text{ et } \forall t \in \mathbb{N}^*, S_t^{(k)} = S_{t-1}^{(k)} + \sum_{i=1}^{n-k} B_{t,i} - 1 \quad (2.23)$$

Pour prouver (2), il suffit alors de montrer que pour tout t , on a presque sûrement soit $S_t^{(k)} \leq S_t$ soit $S_t + t \geq k$ et $S_t^{(k)} + t \geq k$. Procédons par récurrence sur t .

Le cas $t = 1$ est clair.

Supposons que notre prédicat est vérifié pour un certain t fixé. Notons $A_t = \{S_t^{(k)} \leq S_t\}$,

$B_t = \{S_t + t \geq k\}$ et $C_t = \{S_t^{(k)} + t \geq k\}$.

Sur l'événement $B_t \cap C_t$ on a

$$S_{t+1}^{(k)} + t + 1 = S_t^{(k)} + t + \sum_{i=1}^{n-k} B_{t,i} \geq k \quad (2.24)$$

$$S_{t+1} + t + 1 = S_t + t + \sum_{i=1}^{n-(t-1)-S_{t-1}} B_{t,i} \geq k \quad (2.25)$$

donc B_{t+1} et C_{t+1} sont vérifiés.

Sur $A_t \cap \mathbb{C}B_t$ on somme moins de Bernoulli dans l'expression de $S_{t+1}^{(k)}$ que dans celle de S_{t+1} donc on obtient A_{t+1} .

Reste à voir ce qui se passe sur $A_t \cap B_t \cap \mathbb{C}C_t$. En ce cas, on a immédiatement B_{t+1} presque sûrement (conditionnellement à $A_t \cap B_t \cap \mathbb{C}C_t$) de par l'expression de S_{t+1} . Si on se place sur l'événement $\mathbb{C}A_{t+1}$, on obtient alors C_{t+1} presque sûrement donc le prédicat est vérifié au rang $t + 1$. Dans le cas contraire, on est sur A_{t+1} ce qui donne trivialement le prédicat au rang $t + 1$.

Dans tous les cas, on a réussi à montrer l'hérédité, donc le prédicat est vrai pour tout t ce qui achève la preuve de ce lemme. \square

Preuve du théorème 2.4.

Pour prouver ce théorème, nous allons utiliser une approche par couplage. Soit N_t le nombre de sommets non explorés au temps t . Définissons un temps d'arrêt \mathcal{T}_k

$$\mathcal{T}_k = \min\{t : N_t \leq n - k\}. \quad (2.26)$$

Puisque $N_{k-1} \leq n - (k - 1) - 1 = n - k$ on a $\mathcal{T}_k \leq k - 1$. On a donc trivialement

$$\mathbb{P}_{n,p}(|\mathcal{C}(1)| \geq k) = \mathbb{P}_{n,p}(S_t > 0 \ \forall t \leq \mathcal{T}_k). \quad (2.27)$$

Soit $(X_i^{\leq})_{i \geq 1}$ une suite de variables aléatoires i.i.d. (indépendantes et identiquement distribuées) de loi $\text{Bin}(n - k, p)$. Pour $i \leq \mathcal{T}_k$, et conditionnellement à N_{i-1} , soit $Y_i \sim \text{Bin}(N_{i-1} - (n - k), p)$ distribuée indépendamment de toutes les autres variables aléatoires évoquées. Définissons

$$X_i = X_i^{\leq} + Y_i. \quad (2.28)$$

De manière évidente on a $X_i \leq X_i^{\leq}$ presque sûrement pour tout $i \leq \mathcal{T}_k$, tandis que, conditionnellement à N_{i-1} , $X_i \sim \text{Bin}(N_{i-1}, p)$ ce qui est nécessaire pour la relation 2.14.

Soit

$$S_i^{\leq} = X_1^{\leq} + \dots + X_i^{\leq} - (i - 1). \quad (2.29)$$

Alors pour tout $t \leq \mathcal{T}_k$ on a $S_t^{\leq} \leq S_t$. En utilisant le couplage ci-dessus et le fait que $\mathcal{T}_k \leq k - 1$, on a

$$\{S_t > 0 \ \forall t \leq \mathcal{T}_k\} \supseteq \{S_t^{\leq} > 0 \ \forall t \leq \mathcal{T}_k\} \supseteq \{S_t^{\leq} > 0 \ \forall t \leq k - 1\} = \{T^{\leq} \geq k\}, \quad (2.30)$$

où $T^{\leq} = \min\{t : S_t^{\leq} = 0\}$ est la progéniture totale d'un processus de branchement avec distribution binômiale de paramètres $n - k$ et p . Pour prouver le théorème 2.4, il suffit maintenant de passer aux probabilités dans l'équation 2.30. \square

Preuve de la proposition 2.5.

Pour la borne supérieure de $\mathbb{P}_{\lambda}(|\mathcal{C}(v)| \geq k_n)$, on utilise le lemme 2.3 ainsi que le théorème 1.12 pour obtenir que

$$\mathbb{P}_{\lambda}(|\mathcal{C}(v)| \geq k_n) \leq \mathbb{P}_{n, \lambda/n}(T \geq k_n) \leq \mathbb{P}_{\lambda}^*(T^* \geq k_n) + k_n \lambda^2 / n, \quad (2.31)$$

or $\lambda^2 = (1 + \epsilon^2)$, donc il existe une constante C qui ne dépend pas de λ telle que

$$\mathbb{P}_{\lambda}(|\mathcal{C}(v)| \geq k_n) \leq \mathbb{P}_{\lambda}^*(T^* \geq k_n) + Ck_n/n, \quad (2.32)$$

où T et T^* désignent respectivement la progéniture totale d'un processus de branchement binomial de paramètres n et λ/n et d'un processus de branchement de Poisson de paramètre λ respectivement.

On précise cette majoration avec une inégalité exponentielle sur la loi de la progéniture totale découlant de la proposition 1.11 :

$$\begin{aligned} \mathbb{P}_{\lambda}^*(T^* \geq k_n) &= \mathbb{P}_{\lambda}^*(T^* = \infty) + \mathbb{P}_{\lambda}^*(k_n \leq T^* < \infty) \\ &= \zeta_{\lambda} + O(e^{-k_n I_{\lambda}}) \\ &= \zeta_{\lambda} + O(n^{-a I_{\lambda}}) \\ &= \zeta_{\lambda} + o(1/n) \end{aligned} \quad (2.33)$$

car $k_n = \frac{c}{I_{\lambda}} \log(n)$. A noter que le n grand, ne dépendra pas de λ , car il provient de la formule de Stirling, que et que l'on a supposé $\epsilon \gg n^{-1/3}(\log(n))^{1/2}$. En réinjectant ceci dans l'égalité précédente, on obtient la borne supérieure de la Proposition 2.5.

Pour la borne inférieure, en utilisant les mêmes théorèmes que précédemment ainsi que le théorème 2.4 et en posant $\lambda_n := \lambda(1 - k_n/n)$, il vient :

$$\mathbb{P}_{\lambda}(|\mathcal{C}(v)| \geq k_n) \geq \mathbb{P}_{n-k_n, \lambda/n}(T \geq k_n) \geq \mathbb{P}_{\lambda_n}^*(T^* \geq k_n) - C \frac{k_n}{n}, \quad (2.34)$$

où T et T^* sont respectivement les progénitures totales d'un processus de branchement binomial de paramètres $n - k_n$ et λ/n et un processus de branchement de Poisson de paramètre λ_n respectivement. On sait que

$$\mathbb{P}_{\lambda_n}^*(T^* \geq k_n) = \zeta_{\lambda_n} + O(e^{-k_n I_{\lambda_n}}) = \zeta_{\lambda_n} + o(1/n), \quad (2.35)$$

pour n grand, ne dépendant pas de λ . Par le Théorème des accroissements finis, on obtient alors :

$$\eta_{\lambda_n} = \eta_{\lambda} + (\lambda_n - \lambda) \frac{d}{d\lambda} \eta_{\lambda} |_{\lambda=\lambda_n^*} = \eta_{\lambda} + O(k_n/n), \quad (2.36)$$

pour un $\lambda_n^* \in (\lambda_n, \lambda)$, le grand ne dépendant pas de λ .

Ainsi, puisque $\zeta_\lambda = 1 - \eta_\lambda$, on a aussi $\zeta_{\lambda_n} = \zeta_\lambda + O(k_n/n)$. En rassemblant ces résultats, on obtient la borne inférieure de la Proposition 2.5.

La borne inférieure et supérieure ainsi obtenues achèvent la preuve de cette proposition. \square

Preuve de la proposition 2.6.

Définissons

$$Z_{<k} = \sum_{v \in [n]} \mathbb{1}_{\{|\mathcal{C}(v)| < k\}} \quad (2.37)$$

Puis, comme $Z_{<k} = n - Z_{\geq k}$,

$$\text{Var}_\lambda(Z_{\geq k}) = \text{Var}_\lambda(Z_{<k}). \quad (2.38)$$

Ceci prouve que $\text{Var}_\lambda(Z_{<k}) \leq (\lambda k + 1)n\chi_{<k}(\lambda)$. En effet, il suffit d'observer que

$$\begin{aligned} \text{Var}_\lambda(Z_{<k}) &= \mathbb{E}_\lambda[Z_{<k}^2] - \mathbb{E}_\lambda[Z_{<k}]^2 \\ &= \sum_{i,j \in [n]} \mathbb{P}(|\mathcal{C}(i)| < k, |\mathcal{C}(j)| < k) - \sum_{i,j \in [n]} \mathbb{P}(|\mathcal{C}(i)| < k)\mathbb{P}(|\mathcal{C}(j)| < k) \\ &= \sum_{i,j \in [n]} [\mathbb{P}_\lambda(|\mathcal{C}(i)| < k, |\mathcal{C}(j)| < k) - \mathbb{P}_\lambda(|\mathcal{C}(i)| < k)^2] \end{aligned} \quad (2.39)$$

On fait alors une disjonction de cas dépendant de la communication entre i et j :

$$\begin{aligned} \text{Var}_\lambda(Z_{<k}) &= \sum_{i,j \in [n]} [\mathbb{P}_\lambda(|\mathcal{C}(i)| < k, |\mathcal{C}(j)| < k, i \leftrightarrow j) - \mathbb{P}_\lambda(|\mathcal{C}(i)| < k)^2] \\ &\quad + \sum_{i,j \in [n]} \mathbb{P}_\lambda(|\mathcal{C}(i)| < k, |\mathcal{C}(j)| < k, i \not\leftrightarrow j) \end{aligned} \quad (2.40)$$

De plus, on sait que quand $i \leftrightarrow j$ on a $|\mathcal{C}(i)| = |\mathcal{C}(j)|$, ce qui permet de calculer :

$$\begin{aligned} \sum_{i,j \in [n]} \mathbb{P}_\lambda(|\mathcal{C}(i)| < k, |\mathcal{C}(j)| < k, i \leftrightarrow j) &= \sum_{i,j \in [n]} \mathbb{E}_\lambda[\mathbb{1}_{\{|\mathcal{C}(i)| < k\}} \mathbb{1}_{\{i \leftrightarrow j\}}] \\ &= \sum_{i \in [n]} \mathbb{E}_\lambda[\mathbb{1}_{\{|\mathcal{C}(i)| < k\}} \sum_{j \in [n]} \mathbb{1}_{\{i \leftrightarrow j\}}] \\ &= \sum_{i \in [n]} \mathbb{E}_\lambda[|\mathcal{C}(i)| \mathbb{1}_{\{|\mathcal{C}(i)| < k\}}] \\ &= n\chi_{<k}(\lambda). \end{aligned} \quad (2.41)$$

Simplifions maintenant la somme correspondant au cas où $i \not\leftrightarrow j$ et écrivons que, pour $l < k$, on a :

$$\begin{aligned} \mathbb{P}_\lambda(|\mathcal{C}(i)| = l, |\mathcal{C}(j)| < k, i \not\leftrightarrow j) \\ = \mathbb{P}_\lambda(|\mathcal{C}(i)| = l) \mathbb{P}_\lambda(i \not\leftrightarrow j \mid |\mathcal{C}(i)| = l) \mathbb{P}_\lambda(|\mathcal{C}(j)| < k \mid i \not\leftrightarrow j, |\mathcal{C}(i)| = l) \end{aligned} \quad (2.42)$$

Quand $|\mathcal{C}(i)| = l$ et $i \leftrightarrow j$, la loi de $|\mathcal{C}(j)|$ est la même que celle de $|\mathcal{C}(1)|$ dans un graphe aléatoire à $n - l$ sommets et probabilités d'arêtes $p = \lambda/n$, i.e.,

$$\mathbb{P}_{n,\lambda}(|\mathcal{C}(j)| < k | i \leftrightarrow j, |\mathcal{C}(i)| = l) = \mathbb{P}_{n-l,\lambda}(|\mathcal{C}(1)| < k), \quad (2.43)$$

où $\mathbb{P}_{m,\lambda}$ désigne la probabilité d'existence des arêtes du graphe $ER_m(\lambda/n)$. Ainsi,

$$\begin{aligned} \mathbb{P}_{n,\lambda}(|\mathcal{C}(j)| < k | i \leftrightarrow j, |\mathcal{C}(i)| = l) & \quad (2.44) \\ = \mathbb{P}_{n-l,\lambda}(|\mathcal{C}(1)| < k) & = \mathbb{P}_{n,\lambda}(|\mathcal{C}(1)| < k) + \mathbb{P}_{n-l,\lambda}(|\mathcal{C}(1)| < k) - \mathbb{P}_{n,\lambda}(|\mathcal{C}(1)| < k) \end{aligned}$$

On peut coupler $ER_{n-l}(p)$ et $ER_n(p)$ en rajoutant les l sommets manquants $\{n - l + 1, \dots, n\}$ et en occupant les arêtes st avec $s \in \{n - l + 1, \dots, n\}$ et $t \in [n]$ indépendamment avec probabilité p .

On remarque que dans ce couplage $\mathbb{P}_{n-l,\lambda}(|\mathcal{C}(1)| < k) - \mathbb{P}_{n,\lambda}(|\mathcal{C}(1)| < k)$ est égal à la probabilité de l'événement $\{|\mathcal{C}(1)| < k\}$ dans $ER_{n-l}(p)$, mais aussi à la probabilité de l'événement $\{|\mathcal{C}(1)| \geq k\}$ dans $ER_n(p)$.

Si $|\mathcal{C}(1)| < k$ dans $ER_{n-l}(p)$, mais $|\mathcal{C}(1)| \geq k$ dans $ER_n(p)$, alors au moins un des sommets $\{n - l + 1, \dots, n\}$ est relié presque sûrement à un des au plus k sommets de $\mathcal{C}(1)$ dans $ER_{n-l}(p)$. Ceci avec probabilité au plus lkp :

$$\mathbb{P}_\lambda(|\mathcal{C}(j)| < k, i \leftrightarrow j | |\mathcal{C}(i)| = l) - \mathbb{P}_\lambda(|\mathcal{C}(j)| < k) \leq lk\lambda/n \quad (2.45)$$

Ainsi,

$$\begin{aligned} & \sum_{i,j \in [n]} [\mathbb{P}_\lambda(|\mathcal{C}(i)| < k, |\mathcal{C}(j)| < k, i \leftrightarrow j) - \mathbb{P}_\lambda(|\mathcal{C}(i)| < k)\mathbb{P}_\lambda(|\mathcal{C}(j)| < k)] \\ & \leq \sum_{l=1}^{k-1} \sum_{i,j \in [n]} \frac{\lambda kl}{n} \mathbb{P}_\lambda(|\mathcal{C}(i)| = l) = \frac{\lambda k}{n} \sum_{i,j \in [n]} \mathbb{E}_\lambda[|\mathcal{C}(i)| \mathbf{1}_{\{|\mathcal{C}(i)| < k\}}] \\ & = nk\lambda\chi_{<k}(\lambda) \end{aligned} \quad (2.46)$$

Ceci, combiné avec (2.40) et (2.41) conclut la preuve. □

Preuve du corollaire 2.7.

D'après la Proposition 2.5,

$$\mathbb{E}_\lambda[Z_{\geq k_n}] = n\mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k_n) = n\zeta_\lambda + O(k_n), \quad (2.47)$$

et ainsi, pour n assez grand (qui ne dépend pas de λ) et puisque $k_n = o(n^\nu)$, il vient :

$$\{|Z_{\geq k_n} - \mathbb{E}_\lambda[Z_{\geq k_n}]\} \leq n^\nu/2 \subseteq \{|Z_{\geq k_n} - n\zeta_\lambda| \leq n^\nu\} \quad (2.48)$$

Comme $\chi_{<k_n}(\lambda) \leq k_n$, d'après l'inégalité de Chebychev (Théorème 1.6) et la Proposition 2.6,

$$\begin{aligned} \mathbb{P}_\lambda(|Z_{\geq k_n} - n\zeta_\lambda| > n^\nu) &\leq \mathbb{P}_\lambda(|Z_{\geq k_n} - \mathbb{E}_\lambda[Z_{\geq k_n}]| > n^\nu/2) \\ &\leq 4n^{-2\nu} \text{Var}_\lambda(Z_{\geq k_n}) \\ &\leq 4n^{1-2\nu} (\lambda k_n^2 + k_n) \\ &\leq n^{-\delta}, \end{aligned} \tag{2.49}$$

ceci pour tout $\delta < 2\nu - 1$ et n assez grand (qui ne dépend pas de λ), étant donné que $k_n = \frac{c}{1-\lambda} \log(n)$. □

Preuve de la proposition 2.8.

On a :

$$\mathbb{P}_\lambda(k_n \leq |\mathcal{C}(v)| \leq \alpha n) = \sum_{t=k_n}^{\alpha n} \mathbb{P}_\lambda(|\mathcal{C}(v)| = t) \leq \sum_{t=k_n}^{\alpha n} \mathbb{P}_\lambda(S_t = 0), \tag{2.50}$$

où $S_t = S_{t-1} + X_t - 1$. On a $S_t \sim \text{Bin}(n-1, 1 - (1-p)^t) + 1 - t$, où $p = \frac{\lambda}{n}$, d'après la Proposition 2.10. Par conséquent, pour $p = \frac{\lambda}{n}$,

$$\mathbb{P}_\lambda(S_t = 0) = \mathbb{P}(\text{Bin}(n-1, 1 - (1-p)^t) = t-1), \tag{2.51}$$

où, par abus de langage, $\mathbb{P}(\text{Bin}(m, q) = s)$ désigne la probabilité qu'une variable aléatoire de loi Binomiale de paramètres m et q prenne une valeur s .

Pour comprendre l'apparition d'exponentielles dans l'inégalité (2.12), remarquons que pour $p = \frac{\lambda}{n}$ et $t = \lfloor \alpha n \rfloor$,

$$1 - (1-p)^t = 1 - \left(1 - \frac{\lambda}{n}\right)^{\lfloor \alpha n \rfloor} = (1 - e^{-\lambda\alpha})(1 + o(1)) \quad \text{pour } n \geq 2\lambda. \tag{2.52}$$

En effet, pour $n \geq 2\lambda$, $\left(1 - \frac{\lambda}{n}\right)^{\lfloor \alpha n \rfloor} \geq e^{-\lambda\alpha - \frac{\lambda^2\alpha}{n}}$. De plus, l'unique solution dans $(0,1]$ de l'équation $1 - e^{-\lambda\alpha} = \alpha$ est $\alpha = \zeta_\lambda$ (Proposition 2.11).

Si $\alpha < \zeta_\lambda$, alors $\alpha < 1 - e^{-\lambda\alpha}$, et donc $g(\alpha; \lambda) := (1 - e^{-\lambda\alpha})/\alpha < 1$. Par conséquent, $J(\alpha; \lambda) = I_{g(\alpha; \lambda)} > 0$, et donc, la probabilité dans l'égalité (2.51) est exponentiellement petite.

Par l'égalité (2.51) et en remarquant que $1 - p \leq e^{-p}$ et donc que $1 - (1-p)^t \geq 1 - e^{-pt}$,

$$\begin{aligned} \mathbb{P}_\lambda(S_t = 0) &= \mathbb{P}_\lambda(\text{Bin}(n-1, 1 - (1-p)^t) = t-1) \\ &\leq \mathbb{P}_\lambda(\text{Bin}(n-1, 1 - (1-p)^t) \leq t-1) \\ &\leq \mathbb{P}_\lambda(\text{Bin}(n, 1 - (1-p)^t) \leq t) \leq \mathbb{P}_\lambda(\text{Bin}(n, 1 - e^{-pt}) \leq t) \\ &= \mathbb{P}_\lambda(e^{-s\text{Bin}(n, 1 - e^{-pt})} \geq e^{-st}) \end{aligned}$$

On a donc :

$$\mathbb{P}_\lambda(S_t = 0) \leq \mathbb{P}_\lambda(e^{-s \text{Bin}(n, 1-e^{-pt})} \geq e^{-st}) \quad (2.53)$$

et ceci quelque soit $s > 0$.

On a de plus grâce à l'inégalité de Markov (Théorème 1.5) :

$$\begin{aligned} \mathbb{P}_\lambda(S_t = 0) &\leq e^{st} \mathbb{E}[e^{s \text{Bin}(n, 1-e^{-pt})}] = e^{st} ((1 - e^{-pt})e^{-s} + e^{-pt})^n \\ &= e^{st} (1 + (1 - e^{-pt})(e^{-s} - 1))^n \leq e^{st+n(1-e^{-pt})(1-e^{-s})} \\ &= e^{st+n(1-e^{-\lambda t/n})(1-e^{-s})}, \end{aligned}$$

Où on utilise le fait que $1 - x \leq e^{-x}$ dans la dernière égalité. Donc on a :

$$\mathbb{P}_\lambda(S_t = 0) \leq e^{st+n(1-e^{-\lambda t/n})(1-e^{-s})}, \quad (2.54)$$

Soit

$$s^* = \log(n(1 - e^{-\lambda t/n})/t), \quad (2.55)$$

le minimisateur de la fonction $s \mapsto st + n(1 - e^{-\lambda t/n})(1 - e^{-s})$.

Soit $t = \beta n$ et $g(\beta; \lambda) = (1 - e^{-\lambda \beta})/\beta$. Remarquons que $\lim_{\beta \downarrow 0} g(\beta; \lambda) = \lambda > 1$, et que $g(\zeta_\lambda; \lambda) = 1$ d'après la Proposition 2.11.

De plus, la fonction $\beta \mapsto g(\beta; \lambda)$ est décroissante, donc :

$$\frac{\partial}{\partial \beta} g(\beta; \lambda) = \frac{g(\beta; \lambda)}{\beta} [(\beta \lambda) e^{-\beta \lambda} - 1] < 0. \quad (2.56)$$

Par conséquent, $s^* \geq 0$ lorsque $t = \lfloor \alpha n \rfloor$ avec $\alpha \leq \zeta_\lambda$. L'inégalité 2.54 pour $s = s^* = \log(n(1 - e^{-\lambda t/n})/t)$ nous donne :

$$\mathbb{P}_\lambda(S_t = 0) \leq e^{-t(\log g(t/n; \lambda) - 1 - g(t/n; \lambda))} = e^{-t I_{g(t/n; \lambda)}}. \quad (2.57)$$

De plus, comme $\lambda \mapsto I_\lambda$ décroît et que $t/n \leq \alpha < \zeta_\lambda$, alors :

$$\mathbb{P}_\lambda(S_t = 0) \leq e^{-t I_{g(\alpha; \lambda)}} = e^{-t J(\alpha; \lambda)}. \quad (2.58)$$

On a donc pour finir :

$$\begin{aligned} \mathbb{P}_\lambda(k_n \leq |\mathcal{C}(v)| \leq \alpha n) &\leq \sum_{t=k_n}^{\alpha n} \mathbb{P}_\lambda(S_t = 0) \leq \sum_{t=k_n}^{\alpha n} e^{-t J(\alpha; \lambda)} \\ &\leq e^{-k_n J(\alpha; \lambda)} / [1 - e^{-J(\alpha; \lambda)}]. \end{aligned}$$

On a bien :

$$\mathbb{P}_\lambda(k_n \leq |\mathcal{C}(v)| \leq \alpha n) \leq e^{-k_n J(\alpha; \lambda)} / [1 - e^{-J(\alpha; \lambda)}]. \quad (2.59)$$

Supposons maintenant que $\epsilon \rightarrow 0$ quand $n \rightarrow \infty$ et observons le comportement de la borne.

On sait que $J(\alpha, \lambda) \sim \frac{(1-c')^2 \epsilon^2}{2}$ et que $I_\lambda \sim \frac{\epsilon^2}{2}$, donc

$$e^{-k_n J(\alpha, \lambda)} = e^{-\frac{c}{I_\lambda} \log(n) J(\alpha, \lambda)} \sim n^{-c(1-c')^2}. \quad (2.60)$$

D'autre part,

$$1 - e^{-J(\alpha, \lambda)} \sim J(\alpha, \lambda) \sim \epsilon^2. \quad (2.61)$$

Ainsi, la borne est équivalente à $n^{-c(1-c')^2} \epsilon^{-2}$. Or on sait que $\epsilon \gg n^{-1/3} (\log(n))^{1/2}$, donc $n^{-c(1-c')^2} \epsilon^{-2} \ll \frac{n^{2/3-c(1-c')^2}}{\log(n)}$.

Ceci conclut la démonstration de la Proposition 2.8. \square

Preuve du théorème 2.2.

Par union et indépendance des sommets, on a :

$$\mathbb{P}_\lambda(\exists v : k_n \leq |\mathcal{C}(v)| \leq \alpha n) \leq n \mathbb{P}_\lambda(k_n \leq |\mathcal{C}(1)| \leq \alpha n) \leq \frac{n e^{-k_n J(\alpha, \lambda)}}{1 - e^{-J(\alpha, \lambda)}}, \quad (2.62)$$

la deuxième inégalité découlant de la Proposition 2.8.

Si $\lambda > 1$ ne tend pas vers 1, alors $J(\alpha, \lambda)/I_\lambda$ est borné. Le terme de droite étant dans ce cas un $O(n^{-\delta})$ où $\delta = \frac{c}{I_\lambda} J(\alpha, \lambda) - 1$, il existe alors un $0 < \delta' < \delta$, ne dépendant pas de λ tel que le terme de droite soit un $O(n^{-\delta'})$. Si maintenant $\epsilon \rightarrow 0$, i.e. $\lambda \rightarrow 1$, on a

$$\frac{n e^{-k_n J(\alpha, \lambda)}}{1 - e^{-J(\alpha, \lambda)}} \sim n^{1-c(1-c')^2}. \quad (2.63)$$

Donc on a bien le résultat dans ce cas-ci également. \square

Preuve du lemme 2.9.

Pour montrer que \mathcal{E}_n a lieu presque sûrement, notons \mathcal{E}_n^c le complémentaire de \mathcal{E}_n et montrons que celui-ci n'arrive jamais avec grande probabilité.

D'après le Corollaire 2.7, $\mathbb{P}_\lambda(|Z_{\geq k_n} - n\zeta_\lambda| > n^\nu) = O(n^{-\delta})$ et d'après le Théorème 2.2, $\mathbb{P}_\lambda(\exists v \in [n] : k_n \leq |\mathcal{C}(v)| \leq \alpha n) \leq O(n^{-\delta})$.

Ces deux majorations impliquent donc $\mathbb{P}_\lambda(\mathcal{E}_n^c) = O(n^{-\delta})$.

Il s'agit maintenant de prouver que $|\mathcal{C}_{max}| = Z_{\geq k_n}$ sur l'événement \mathcal{E}_n . Pour ce faire remarquons d'abord que $\{|Z_{\geq k_n} - \zeta_\lambda| \leq n^\nu\} \subseteq \{Z_{\geq k_n} \leq 1\}$. Ceci implique que $|\mathcal{C}_{max}| \leq Z_{\geq k_n}$ lorsque l'événement \mathcal{E}_n a lieu. De plus, $|\mathcal{C}_{max}| < Z_{\geq k_n}$ implique qu'il existe deux composantes connexes de taille au moins k_n .

En outre, \mathcal{E}_n implique qu'il n'existe pas de composante connexe de taille comprise entre k_n et αn . Par conséquent, il existe deux composantes connexes de taille au moins αn .

On a donc $Z_{\geq k_n} \geq 2\alpha n$. Lorsque $2\alpha > \zeta_\lambda$ et n suffisamment grand, on a alors une contradiction avec $Z_{\geq k_n} \leq \zeta_\lambda n + n^\nu$. On en déduit alors que $|\mathcal{C}_{max}| = Z_{\geq k_n}$. \square

Preuve de la proposition 2.10. Soit N_t le nombre de sommets neutres au temps $t \geq 0$. On a donc

$$N_t = n - t - S_t. \quad (2.64)$$

Remarquons que $X \sim \text{Bin}(m, p)$ lorsque $Y = m - X \sim \text{Bin}(m, 1 - p)$. Pour montrer la Proposition 2.10, il suffit alors de montrer que pour tout $t \geq 0$,

$$N_t \sim \text{Bin}(n - 1, (1 - p)^t). \quad (2.65)$$

Pour prouver la relation (2.65), nous pouvons voir que chaque sommet de $\{2, \dots, n\}$ a une probabilité $(1 - p)^t$ de rester neutre après t étapes (en supposant que l'on commence par le sommet 1). Par la relation (2.14) on a, conditionnellement à S_{t-1} , $X_t \sim \text{Bin}(n - (t - 1) - S_{t-1}, p) = \text{Bin}(N_{t-1}, p)$. Sachant que $N_0 = n - 1$, on a :

$$N_t = n - t - S_t = n - t - S_{t-1} - X_t + 1 = N_{t-1} - X_t \sim \text{Bin}(N_{t-1}, 1 - p). \quad (2.66)$$

On conclut maintenant facilement par récurrence sur t . □

Preuve de la proposition 2.11. La fonction génératrice d'une loi de Poisson de paramètre λ est :

$$\begin{aligned} g_{\text{Pois}}(s) &= \sum_{k=0}^{\infty} s^k \mathbb{P}(X = k) \\ &= \sum_{k=0}^{\infty} s^k e^{-\lambda} (\lambda)^k / k! \\ &= e^{-\lambda} \sum_{k=0}^{\infty} (s\lambda)^k / k! \\ &= e^{-\lambda} e^{s\lambda} \\ &= e^{\lambda(s-1)} \end{aligned}$$

Or on sait que la probabilité d'extinction η_λ d'un processus de branchement de Poisson de paramètre λ est un point fixe de la fonction génératrice. Donc

$$\eta_\lambda = e^{\lambda(\eta_\lambda - 1)} \quad (2.67)$$

Donc si on note $\zeta_\lambda = 1 - \eta_\lambda$ la probabilité de survie, on a

$$\begin{aligned} 1 - \zeta_\lambda &= e^{\lambda(1 - \zeta_\lambda - 1)} \\ \zeta_\lambda &= 1 - e^{\lambda(\zeta_\lambda)}. \end{aligned}$$

Ce qui conclut la démonstration. □

Chapitre 3

Régime sous-critique

3.1 Ordre des composantes maximales sous-critiques

On se réfère ici à [1].

Après avoir étudié pleinement le régime sur-critique $\lambda > 1$, nous nous tournons maintenant vers le régime sous-critique $\lambda = 1 - \epsilon < 1$, où $\epsilon \gg n^{-1/3} \log(n)^{1/2}$. Si dans le cas sur-critique il existait une unique composante de taille maximale, géante, il n'en est rien dans le cas sous-critique. Nous allons voir que l'on a en fait le résultat suivant :

Théorème 3.1 (Ordre des composantes maximales sous-critiques).

On a

$$\frac{I_\lambda |\mathcal{C}_{max}|}{\log(n)} \xrightarrow{\mathbb{P}} 1 \quad (3.1)$$

où I_λ est défini en (2.2).

Les composantes maximales sont donc d'ordre $\log(n)$, ce qui est bien moins impressionnant que l'ordre n de la composante géante du régime sur-critique.

La preuve de ce théorème découle des deux théorèmes qui suivent :

Théorème 3.2 (Borne supérieure sur les composantes sous-critiques maximales).

Pour tout $c > 1$, il existe $\delta > 0$ tel que :

$$\mathbb{P}_\lambda(|\mathcal{C}_{max}| \geq \frac{c}{I_\lambda} \log(n)) = O(n^{-\delta}). \quad (3.2)$$

Théorème 3.3 (Borne inférieure sur les composantes sous-critiques maximales).

Pour tout $c' < 1$, il existe $\delta' > 0$ tel que :

$$\mathbb{P}_\lambda(|\mathcal{C}_{max}| \leq \frac{c'}{I_\lambda} \log(n)) = O(n^{-\delta'}). \quad (3.3)$$

Proposition 3.4 (Une autre estimation sur la variance de $Z_{\geq k}$). *Pour tout $n, k \in [n]$ et pour tout $\lambda > 0$,*

$$\text{Var}_\lambda(Z_{\geq k}) \leq n\chi_{\geq k}(\lambda), \quad (3.4)$$

où

$$\chi_{\geq k}(\lambda) = \mathbb{E}[|\mathcal{C}(v)|\mathbb{1}_{\{|\mathcal{C}(v)| \geq k\}}]. \quad (3.5)$$

Evidemment, par indépendance, $\chi_{\geq k}(\lambda)$ ne dépend pas du sommet v choisi.

Admettons temporairement ces résultats afin de prouver le théorème 3.1 dans la section suivante.

3.2 Preuves des résultats sous-critiques

Preuve du théorème 3.1.

Soit $\iota > 0$. On a

$$\begin{aligned} \mathbb{P}_\lambda\left(\left|\frac{|\mathcal{C}_{max}|}{\log(n)} - \frac{1}{I_\lambda}\right| \geq \iota\right) &= \mathbb{P}_\lambda(|\mathcal{C}_{max}| - \log(n)/I_\lambda \geq \iota \log(n)) \\ &= \mathbb{P}_\lambda(|\mathcal{C}_{max}| - \log(n)/I_\lambda \geq \iota \log(n)) \\ &\quad + \mathbb{P}_\lambda(|\mathcal{C}_{max}| - \log(n)/I_\lambda \leq -\iota \log(n)) \\ &= \mathbb{P}_\lambda(|\mathcal{C}_{max}| \geq (\iota + 1/I_\lambda) \log(n)) \\ &\quad + \mathbb{P}_\lambda(|\mathcal{C}_{max}| \leq (1/I_\lambda - \iota) \log(n)) \\ &\leq \mathbb{P}_\lambda(|\mathcal{C}_{max}| \geq (\iota + 1/I_\lambda) \log(n)) \\ &\quad + \mathbb{P}_\lambda(|\mathcal{C}_{max}| \leq (1/I_\lambda + \iota) \log(n)) \end{aligned}$$

Par les théorèmes 3.2 et 3.3, ces deux quantités sont respectivement $O(n^{-\delta})$ et $O(n^{-\delta'})$ pour $\delta > 0$ et $\delta' > 0$ bien choisis, donc ceci tend bien vers 0 quand $n \rightarrow \infty$. \square

Preuve du théorème 3.2.

Comme dans les sections précédentes, on considère la variable aléatoire $Z_{\geq \frac{c}{I_\lambda} \log(n)}$ pour $c > 1$. On pose $k_n = \frac{c}{I_\lambda} \log(n)$ et on a

$$\mathbb{E}_\lambda[Z_{\geq k_n}] = n\mathbb{P}_\lambda(|\mathcal{C}(1)| \geq k_n) \quad (3.6)$$

On utilise alors le lemme 2.3 qui nous donne que

$$\mathbb{P}_\lambda(|\mathcal{C}(1)| \geq k_n) \leq \mathbb{P}_{n, \lambda/n}(T \geq k_n) \quad (3.7)$$

où T désigne la progéniture totale d'un processus de branchement binomial de paramètres $n, \lambda/n$.

On considère le processus $\hat{S}_t = \hat{X}_1 + \dots + \hat{X}_t - (t-1)$ où $(\hat{X}_i)_{i \geq 1}$ est une famille de variables i.i.d. de loi binomiale de paramètres n et λ/n . Il vient alors :

$$\mathbb{P}_{n,\lambda/n}(T > t) \leq \mathbb{P}_{n,\lambda/n}(\hat{S}_t > 0) = \mathbb{P}_{n,\lambda/n}(\hat{X}_1 + \dots + \hat{X}_t \geq t) \leq e^{-tI_\lambda}, \quad (3.8)$$

l'inégalité découlant du fait que $\hat{X}_1 + \dots + \hat{X}_t \sim \text{Bin}(nt, \lambda/n)$ et du lemme 1.10.

On a alors

$$\mathbb{P}_\lambda(|\mathcal{C}(1)| > k_n) \leq e^{-k_n I_\lambda} \quad (3.9)$$

et donc on obtient

$$\mathbb{P}_\lambda(|\mathcal{C}_{max}| > \frac{c}{I_\lambda} \log(n)) \leq ne^{-\frac{c}{I_\lambda} \log(n) I_\lambda} = n^{1-c}. \quad (3.10)$$

On pose alors $\delta = c - 1 > 0$. □

Preuve du théorème 3.3.

On vérifie aisément que $\mathbb{P}_\lambda(|\mathcal{C}_{max}| < k) = \mathbb{P}_\lambda(Z_{\geq k} = 0)$. Il nous suffit alors de démontrer que $\mathbb{P}_\lambda(Z_{\geq k_n} = 0) = O(n^{-\delta'})$, où $k_n = \frac{c'}{I_\lambda} \log(n)$ avec $c' < 1$. Nous allons dans ce but appliquer l'inégalité de Chebychev.

Commençons par alléger les notations

$$P_{\geq k}(\lambda) = \mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k). \quad (3.11)$$

On a alors

$$\mathbb{E}_\lambda[Z_{\geq k}] = nP_{\geq k}(\lambda) \quad (3.12)$$

Notons T un processus de branchement binomial de paramètres $n - k_n$ et $p = \lambda/n$. D'après le lemme 2.3, on a

$$P_{\geq k_n}(\lambda) \geq \mathbb{P}_{n-k_n}(T \geq k_n). \quad (3.13)$$

Notons T^* la progéniture totale d'un processus de branchement de Poisson de paramètre $\lambda_n = \lambda(n - k_n)/n$. Par le théorème 1.12 on a

$$\mathbb{P}_{n-k_n, p}(T \geq \frac{c'}{I_\lambda} \log(n)) = \mathbb{P}_{\lambda_n}^*(T^* \geq \frac{c'}{I_\lambda} \log(n)) + O\left(\frac{c' \lambda^2 \log(n)}{I_\lambda n}\right). \quad (3.14)$$

De plus, par la proposition 1.11,

$$\mathbb{P}_{\lambda_n}^*(T^* \geq \frac{c'}{I_\lambda} \log(n)) = \sum_{k=\frac{c'}{I_\lambda} \log(n)}^{\infty} \mathbb{P}_{\lambda_n}^*(T^* = k) = \sum_{k=\frac{c'}{I_\lambda} \log(n)}^{\infty} \frac{(\lambda_n k)^{k-1}}{k!} e^{-\lambda_n k}. \quad (3.15)$$

La formule de Stirling nous donne

$$k! = \left(\frac{k}{e}\right)^k \sqrt{2\pi k} (1 + o(1)), \quad (3.16)$$

donc en utilisant les définitions de I_λ et λ_n , on observe que $I_{\lambda_n} = I_\lambda + o(1)$ puis

$$\mathbb{P}_{\lambda_n}^*(T^* \geq \frac{c'}{I_\lambda} \log(n)) = \frac{1}{\lambda} \sum_{k=\frac{c'}{I_\lambda} \log(n)}^{\infty} \frac{1}{\sqrt{2\pi k^3}} e^{-I_{\lambda_n} k} (1 + o(1)) \geq e^{-I_\lambda \frac{c'}{I_\lambda} \log(n)(1+o(1))}. \quad (3.17)$$

Il vient alors qu'avec $k_n = \frac{c'}{I_\lambda} \log(n)$ et pour tout $0 < \alpha < 1 - c'$,

$$\mathbb{E}_\lambda[Z_{\geq k_n}] = nP_{\geq k_n}(\lambda) \geq n^{(1-c')(1+o(1))} \geq n^\alpha, \quad (3.18)$$

pour n assez grand (qui ne dépend pas de λ car ceci vient de Stirling).

On s'attaque maintenant à la variance de $Z_{\geq k_n}$ en utilisant la proposition 3.4. Pour n'importe quelle variable aléatoire X à valeurs entières, on a

$$\mathbb{E}[X \mathbb{1}_{\{X \geq k\}}] = k\mathbb{P}(X \geq k) + \sum_{t=k+1}^{\infty} \mathbb{P}(X = t). \quad (3.19)$$

D'après l'inégalité 3.9 et par définition de $\chi_{\geq k_n}(\lambda)$,

$$\begin{aligned} \chi_{\geq k_n}(\lambda) &= k_n P_{\geq k_n}(\lambda) + \sum_{t=k_n+1}^n P_{\geq t}(\lambda) \leq k_n e^{-(k_n-1)I_\lambda} \sum_{t=k_n}^n e^{-I_\lambda(t-1)} \\ &\leq k_n e^{-(k_n-1)I_\lambda} + \frac{e^{-k_n I_\lambda}}{1 - e^{-I_\lambda}} = O(k_n n^{-c'}). \end{aligned} \quad (3.20)$$

En appliquant la proposition 3.4, on obtient

$$\text{Var}_\lambda(Z_{\geq k_n}) \leq n\chi_{\geq k_n}(\lambda) \leq O(k_n n^{1-c'}), \quad (3.21)$$

d'autre part, pour n assez grand,

$$\mathbb{E}_\lambda[Z_{\geq k_n}] \geq n^\alpha. \quad (3.22)$$

Ainsi, par l'inégalité de Chebychev, pour n assez grand,

$$\mathbb{P}_\lambda(Z_{\geq k_n}) \leq \frac{\text{Var}_\lambda(Z_{\geq k_n})}{\mathbb{E}_\lambda[Z_{\geq k_n}]^2} \leq O(k_n n^{1-c'-2\alpha}). \quad (3.23)$$

On remarque que $2\alpha - (1 - c') > 0$ quand $0 < \alpha < 1 - c'$ donc

$$\forall 0 < \delta' < 2\alpha - (1 - c'), \quad \mathbb{P}_\lambda(Z_{\geq k_n}) = O(n^{-\delta'}). \quad (3.24)$$

On conclut la preuve en utilisant le fait que

$$\mathbb{P}_\lambda(|\mathcal{C}_{max}| < k_n) = \mathbb{P}_\lambda(Z_{\geq k_n} = 0). \quad (3.25)$$

□

Preuve de la proposition 3.4.

La preuve ressemble assez à celle de l'estimation dans le cas sur-critique. Nous avons

$$\text{Var}_\lambda(Z_{\geq k}) = \sum_{i,j \in [n]} [\mathbb{P}_\lambda(|\mathcal{C}(i)| \geq k, |\mathcal{C}(j)| \geq k) - \mathbb{P}_\lambda(|\mathcal{C}(i)| \geq k)^2]. \quad (3.26)$$

On conditionne par $i \leftrightarrow j$

$$\begin{aligned} \mathbb{P}_\lambda(|\mathcal{C}(i)| \geq k, |\mathcal{C}(j)| \geq k) &= \mathbb{P}_\lambda(|\mathcal{C}(i)| \geq k, i \leftrightarrow j) \\ &\quad + \mathbb{P}_\lambda(|\mathcal{C}(i)| \geq k, |\mathcal{C}(j)| \geq k, i \not\leftrightarrow j). \end{aligned} \quad (3.27)$$

D'autre part,

$$\begin{aligned} \mathbb{P}_\lambda(|\mathcal{C}(i)| = l, |\mathcal{C}(j)| \geq k, i \leftrightarrow j) \\ = \mathbb{P}_\lambda(|\mathcal{C}(i)| = l, i \leftrightarrow j) \mathbb{P}_\lambda(|\mathcal{C}(j)| \geq k | |\mathcal{C}(i)| = l, i \leftrightarrow j). \end{aligned} \quad (3.28)$$

Quand $|\mathcal{C}(i)| = l$ et $i \leftrightarrow j$, tous les sommets dans des composantes autres que $\mathcal{C}(i)$, dont j en particulier, forment un graphe $ER_{n-l}(\lambda/n)$. Comme la probabilité de l'événement $\{|\mathcal{C}(j)| \geq k\}$ dans $ER_n(p)$ croît avec n ,

$$\mathbb{P}_\lambda(|\mathcal{C}(j)| \geq k | |\mathcal{C}(i)| = l, i \leftrightarrow j) \leq \mathbb{P}_\lambda(|\mathcal{C}(j)| \geq k). \quad (3.29)$$

Pour tout $l \geq 1$, il vient alors

$$\mathbb{P}_\lambda(|\mathcal{C}(i)| = l, |\mathcal{C}(j)| \geq k, i \leftrightarrow j) - \mathbb{P}_\lambda(|\mathcal{C}(i)| = l) \mathbb{P}_\lambda(|\mathcal{C}(j)| \geq k) \leq 0, \quad (3.30)$$

d'où

$$\text{Var}_\lambda(Z_{\geq k}) \leq \sum_{i,j=1}^n \mathbb{P}_\lambda(|\mathcal{C}(i)| \geq k, i \leftrightarrow j). \quad (3.31)$$

Ainsi,

$$\begin{aligned} \text{Var}_\lambda(Z_{\geq k}) &\leq \sum_{i \in [n]} \sum_{j \in [n]} \mathbb{E}_\lambda[\mathbb{1}_{\{|\mathcal{C}(i)| \geq k\}} \mathbb{1}_{\{j \in \mathcal{C}(i)\}}] \\ &= \sum_{i \in [n]} \mathbb{E}_\lambda[\mathbb{1}_{\{|\mathcal{C}(i)| \geq k\}} \sum_{j \in [n]} \mathbb{1}_{\{j \in \mathcal{C}(i)\}}]. \end{aligned} \quad (3.32)$$

Comme $\sum_{j \in [n]} \mathbb{1}_{\{j \in \mathcal{C}(i)\}} = |\mathcal{C}(i)|$ et par indépendance des sommets,

$$\text{Var}_\lambda(Z_{\geq k}) \leq \sum_{i \in [n]} \mathbb{E}_\lambda[|\mathcal{C}(i)| \mathbb{1}_{\{|\mathcal{C}(i)| \geq k\}}] = n \mathbb{E}_\lambda[|\mathcal{C}(1)| \mathbb{1}_{\{|\mathcal{C}(1)| \geq k\}}] = n \chi_{\geq k}(\lambda). \quad (3.33)$$

Ceci achève la preuve. □

Chapitre 4

Régime critique

En vue des chapitres précédents, on est amené à croire que pour le cas critique, $\lambda = 1$, la plus grande composante est d'ordre $n^{2/3}$, et de plus, qu'il existe une fenêtre critique de $\lambda = 1 + O(n^{-1/3})$ dans laquelle cela reste vrai. Nous allons donc dans ce chapitre donner des arguments heuristiques dans cette direction. Pour des arguments complets, on peut se référer à D. Aldous [5] qui étudie en détail la fenêtre critique.

Reprenons le graphe $ER_n(p)$ avec maintenant $\lambda = 1$ soit $p = 1/n$. Nous avons alors le théorème suivant :

Théorème 4.1 (Taille de la composante maximale dans le cas critique). *Soit $ER_n(p)$ le graphe d'Erdős-Rényi à n sommets avec une distribution d'arêtes de probabilité $p=1/n$. Alors*

$$|\mathcal{C}_{max}| = \Theta(n^{2/3}). \quad (4.1)$$

Preuve du théorème 4.1.

Considérons une marche $S_n = 1 + Y_1 + \dots + Y_n$ où $Y_i \in \{-1, 0, 1, \dots\} \forall i \in \mathbb{N}$. On a donc $\forall i \in \mathbb{N}, Y_i \sim (Bin(n, 1/n) - 1)$ et donc $S_m \sim (Bin(nm, 1/n) - m + 1)$. Or lorsque n est grand, d'après le théorème 1.12, on peut assimiler une loi Binomiale de paramètres n et p à une loi de Poisson de paramètre np .

D'où

$$\begin{aligned} \mathbb{P}_1(H_0 \geq k) &= \sum_{m \geq k} \mathbb{P}_1(H_0 = m) \\ (d'après la prop 1.8) &= \sum_{m \geq k} \frac{1}{m} \mathbb{P}_1(S_m = 0) \\ &= \sum_{m \geq k} \frac{1}{m} \mathbb{P}(Bin(nm, 1/n) = m - 1) \\ &\sim \sum_{m \geq k} \frac{1}{m} \mathbb{P}(Poiss(m) = m - 1). \end{aligned}$$

Or

$$\frac{1}{m} \mathbb{P}(Poiss(m) = m - 1) = \frac{1}{m} e^{-m} \frac{m^{m-1}}{(m-1)!}. \quad (4.2)$$

Et d'après le théorème 1.9, on a

$$\begin{aligned} \frac{1}{m} \mathbb{P}(\text{Pois}(m) = m-1) &= \frac{1}{m} e^{-m} \frac{m^{m-1}}{(m-1)^{m-1} e^{-(m-1)} \sqrt{2\pi m}} \\ &\sim \frac{1}{m} e^{-m} \frac{m^{m-1}}{(m)^{m-1} e^{-m} \sqrt{2\pi m}} \\ &= \frac{1}{m\sqrt{2\pi m}} \end{aligned}$$

Donc

$$\mathbb{P}_1(H_0 \geq k) \sim \sum_{m \geq k} \frac{1}{m} \frac{1}{\sqrt{2\pi m}} \sim \sum_{m \geq k} m^{-3/2} \sim k^{-1/2}. \quad (4.3)$$

On cherche maintenant de manière non rigoureuse un α tel que $|\mathcal{C}_{max}| = \Theta(n^\alpha)$.
Calculons donc $\mathbb{E}[Z_{\geq n^\alpha}]$. On a d'après le lemme 2.3

$$\mathbb{E}[Z_{\geq n^\alpha}] = \mathbb{E}\left[\sum_{v \in [n]} \mathbb{1}_{\{|\mathcal{C}(v)| \geq n^\alpha\}}\right] \sim \mathbb{E}\left[\sum_{v \in [n]} \mathbb{1}_{\{H_0 \geq n^\alpha\}}\right] = n \mathbb{P}_1(H_0 > n^\alpha) = n \cdot n^{-\alpha/2}.$$

Or on veut $\mathbb{E}[Z_{\geq n^\alpha}] = \Theta(n^\alpha)$ soit :

$$\begin{aligned} n \cdot n^{-\alpha/2} &\sim n^\alpha \\ n^{(3/2)\alpha} &\sim n \\ \alpha &\sim \frac{2}{3} \end{aligned}$$

Finalement on a bien

$$|\mathcal{C}_{max}| = \Theta(n^{2/3}). \quad (4.4)$$

On voit donc bien apparaître un $n^{2/3}$. Il nous reste maintenant à le montrer de manière plus rigoureuse. Pour cela, il suffit de majorer et de minorer $|\mathcal{C}(v)|$.

Commençons par la majoration.

Montrons que $\forall \varepsilon > 0, \exists C < \infty$ tel que

$$\mathbb{P}(\exists v, |\mathcal{C}(v)| > Cn^{2/3}) < \varepsilon \quad (4.5)$$

Soit $\varepsilon > 0$. Prenons un $C < \infty$ de manière arbitraire

$$\begin{aligned} \mathbb{P}(\exists v, |\mathcal{C}(v)| > Cn^{2/3}) &= \mathbb{P}(\#\mathcal{C}_{\geq Cn^{2/3}} \geq 1) \\ (d'après Markov) &\leq \mathbb{E}(\#\mathcal{C}_{\geq Cn^{2/3}}) \\ &\leq \mathbb{E}(Z_{\geq Cn^{2/3}}) / Cn^{2/3} \end{aligned}$$

où $\#\mathcal{C}_{\geq Cn^{2/3}}$ est le nombre de composantes de taille supérieure ou égale à $Cn^{2/3}$.

Or il existe C' tel que

$$\mathbb{E}(Z_{\geq Cn^{2/3}}) \leq C'n^{2/3} \quad (4.6)$$

On a donc

$$\mathbb{P}(\exists v, |\mathcal{C}(v)| > Cn^{2/3}) \leq \frac{C'}{C}. \quad (4.7)$$

On peut donc choisir C afin de rendre $\frac{C'}{C} < \varepsilon$ et donc nous donner la majoration souhaitée.

Pour la minoration, il faut se référer à [1].

□

Chapitre 5

Simulations

5.1 Code R pour les simulations

La fonction `Bernoulli` génère un k -échantillon de Bernoulli de paramètre p .

```
Bernoulli=function(k,p)
{
  u=runif(k)
  res=(u<p)
  return(res)
}
```

La fonction `SimuleER` génère la matrice du graphe $ER(n, \lambda/n)$.

```
SimuleER=function(n, lambda)
{
  ER=diag(n)
  X=Bernoulli(n*(n-1)/2, lambda/n)
  ER[lower.tri(ER,diag=F)]<-X
  ER=(ER+t(ER))-diag(n)
  return(ER)
}
```

`decoupefamille` est une fonction esclave pour la fonction `TailleCompoDFS`, elle sert à supprimer les arêtes éventuelles dans ER reliant les sommets de *famille*.

```
decoupefamille=function(ER, famille)
{
  M=ER
  nb=length(famille)
  for (i in 1:(nb-1))
  {
```

```

    for (j in min(i+1,nb):nb)
    {
        M[famille[i],famille[j]]=0
        M[famille[j],famille[i]]=0
    }
}
return (M)
}

```

La fonction `TailleCompoDFS` utilise la méthode Depth-First Search pour explorer la composante du sommet k étant donné la matrice ER du graphe. Cette fonction retourne une liste de 3 éléments : la taille de $\mathcal{C}(k)$, la matrice ER actualisée pour la dernière étape de la récursion (ceci a une utilité uniquement dans le programme codant la fonction) et la liste des sommets présents dans $\mathcal{C}(k)$.

```

TailleCompoDFS=function(k, ER)
{
    X=ER-diag(dim(ER)[1])
    if (sum(X[k,]==1)==0)
    {
        return(list(1, ER, k))
    }
    else
    {
        offsprings=which(X[k,]==1)
        nboffsprings=length(offsprings)
        DFS=1
        present=k
        ERd=decoupefamille(ER, c(k, offsprings))
        for (i in offsprings) #récursion
        {
            if (!is.element(i, present))
            {
                TCDFS=TailleCompoDFS(i, ERd)
                DFS=DFS+TCDFS[[1]]
                ERd=TCDFS[[2]]
                present=c(present, TCDFS[[3]])
            }
        }
        return(list(DFS, ERd, present))
    }
}

```

La fonction `TailleComposantes` rend une matrice où la première colonne donne un représentant de chaque composante et la deuxième la taille de chacune de ces com-

posantes.

```
TailleComposantes=function(ER)
{
  n=dim(ER)[1]
  M=NULL
  dejavu=NULL
  for (i in 1:n)
  {
    if (!is.element(i,dejavu))
    {
      TCDFS=TailleCompoDFS(i,ER)
      M=rbind(M,c(i,TCDFS[[1]]))
      dejavu=c(dejavu,TCDFS[[3]])
    }
  }
  M=M[order(-M[,2]),]
  return(M)
}
```

On se contentera ici d'une approche DFS pour déterminer la taille des composantes de $ER_n(p)$, les résultats obtenus avec la méthode BFS seraient, rappelons-le, évidemment les mêmes.

5.2 Présentation de simulations

Quelques exemples de données retournées par le programme pour des valeurs différentes de n et λ . On observe bien les régimes sous-critique, critique et sur-critique décrits plus haut de plus en plus prononcés quand n est grand. Notons que les tailles indiquées ici peuvent être répétées car il s'agit de ce que retourne les deux premières tailles de `TailleComposantes` donc si il existe deux composantes disjointes de même taille, il y aura répétition (voir dernière ligne du tableau par exemple).

λ	$n = 100$		$n = 1000$	
	$ C_{max} $	$ C_2 $	$ C_{max} $	$ C_2 $
2	81	2	785	5
1.2	37	13	396	24
1.1	31	28	124	101
1	12	11	58	34
0.9	8	7	33	25
0.5	5	4	9	9

Bibliographie

- [1] "Random Graphs and Complex Networks. Vol. I", Remco Van Der Hofstad, October 20, 2014, Chapitres 2 à 5.
- [2] "Combinatorial Stochastic Processes", J. Pitman, Ecole d'Été de Probabilités de Saint-Flour XXXII, 2002.
- [3] "A simple branching process approach to the phase transition in $G_{n,p}$ ", Béla Bollobás and Oliver Riordan, July 24, 2012.
- [4] "The phase transition in random graphs - a simple proof", Michael Krivelevich and Benny Sudakov, September 25, 2012.
- [5] "Brownian excursions, critical random graphs and the multiplicative coalescent", D. Aldous, 1997.
- [6] "On the evolution of random graphs", P.Erdős et A.Rényi, 1960.