

Etude des graphes aléatoires

Jean-Benoît Rossel

Projet de semestre

Professeur responsable : Thomas Mountford
Chaire de Processus stochastiques

6 février 2006

Résumé

Ce projet de semestre a pour but d'étudier quelques problèmes liés aux graphes aléatoires. C'est une théorie qui fut initialement développée par deux Hongrois, Paul Erdős (1913 – 1996) et Alfréd Rényi (1921 – 1970). Nous allons tout d'abord rappeler quelques fondements de la théorie des graphes et des probabilités, puis présenter quelques résultats. Nous nous intéresserons par exemple à la probabilité qu'un graphe aléatoire contienne une copie d'un graphe donné, ou encore au comportement de ce graphe aléatoire lorsque son nombre de sommets tend vers l'infini. Nous évoquerons enfin un problème qui n'est pas encore résolu à l'heure actuelle.

Table des matières

1	Introduction	2
1.1	Rappels et notations	2
1.2	Présentation des modèles de base	6
2	Outils mathématiques	8
2.1	La méthode des moments	8
2.2	La méthode de Stein	10
2.3	Le concept des valeurs critiques	13
3	Le problème de l'existence d'une copie d'un graphe donné	15
3.1	Exemple fondamental	15
3.2	Le théorème de Bollobás	16
3.3	Cas intermédiaire	18
4	Le problème du recouvrement	22
4.1	L'enracinement d'un graphe	22
4.2	Notations et théorème	23
5	Couplages et G-facteurs	27
5.1	Le théorème de Hall	27
5.2	Etude des couplages parfaits	29
5.3	Quelques mots sur les G -facteurs	31
	Bibliographie	33

Chapitre 1

Introduction

La notion de graphe aléatoire est apparue pour la première fois dans un article d'Erdős daté de 1947. Son modèle consistait à choisir équitablement un graphe parmi les $2^{\binom{n}{2}}$ graphes que l'on peut former à partir de n sommets. En d'autres termes, son modèle se basait sur l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, où l'ensemble Ω contient tous les graphes possédant n sommets, la tribu \mathcal{F} contient tous les sous-ensembles de Ω , et la probabilité \mathbb{P} est définie de manière à ce que pour tout $\omega \in \Omega$, on ait

$$\mathbb{P}(\omega) = 2^{-\binom{n}{2}}. \quad (1.1)$$

Aujourd'hui on considère essentiellement deux modèles de graphes aléatoires, qui sont le modèle binomial et le modèle uniforme. Nous allons les présenter à la suite de quelques rappels et notations.

1.1 Rappels et notations

Théorie des graphes

On définit un graphe G par deux ensembles notés V et E . Le premier est l'ensemble des sommets de G , et le second est celui de ses arêtes. Pour définir complètement un graphe il faudrait encore considérer la fonction d'incidence $\varphi : E \rightarrow \mathcal{P}_2(V)$ qui associe à chaque arête un couple de sommets. Mais dans ce projet on ne considérera que des graphes non orientés et simples, et donc on pourra négliger cette fonction φ , car il n'y aura pas de risque de confusion. En effet, pour tous sommets i et j distincts il y aura au maximum une seule arête (i, j) .

Soit alors $G = (V, E)$ un graphe quelconque. Les notations suivantes seront utilisées tout au long de ce projet :

- $v_G = v(G) = |V| = |V(G)|$ est le nombre de sommets de G .
- $e_G = e(G) = |E| = |E(G)|$ est le nombre d'arêtes de G , et $e_G(A, B)$ est le nombre d'arêtes de G ayant une extrémité dans $A \subseteq V$ et l'autre dans $B \subseteq V$.
- $d(G) = e_G/v_G$ est la *densité* de G .
- $m(G) = \max_{H \subseteq G} d(H)$ est la *densité maximale* de G .
- $N(v) = N_G(v) = \{w \in V | (v, w) \in E\}$ est le *voisinage du sommet* v .
- $N(S) = N_G(S) = \{\bigcup_{v \in S} N(v)\} \setminus S$ est le *voisinage* de $S \subseteq V$.
- $\overline{N}(v) = N(v) \cup \{v\}$ et $\overline{N}(S) = N(S) \cup S$ sont les *voisinages clos* respectifs de v et S .
- $deg(v) = deg_G(v) = |N(v)|$ est le degré du sommet v . On note aussi $\delta_G = \min_{v \in G} deg(v)$ le degré minimal de G , et $\Delta_G = \max_{v \in G} deg(v)$ le degré maximal de G .

- On note par $\overline{G} = (V, \mathcal{P}_2(V) \setminus E)$ le *graphe complémentaire* de G .
- K_n est le graphe possédant n sommets, avec une arête entre chaque paire de sommets distincts. On l'appelle *graphe complet sur n sommets*. Son graphe complémentaire est alors un graphe *vide*, possédant n sommets et aucune arête. De plus chaque sommet de K_n a un degré égal à $n - 1$.
- G est *biparti* si l'on peut écrire $V = V_1 \cup V_2$, avec V_1 et V_2 non vides et disjoints, et toute arête contient une extrémité dans V_1 et l'autre dans V_2 .
- $K_{m,n}$ est le graphe biparti avec $|V_1| = m$, $|V_2| = n$, et une arête entre chaque sommet de V_1 et chaque sommet de V_2 . C'est un *graphe biparti complet*. On appelle *étoile à n branches* le graphe $K_{1,n}$.
- \emptyset est le graphe ne possédant aucun sommet, et par conséquent aucune arête.
- C_k désigne un cycle de longueur k , avec donc k arêtes et k sommets.
- P_k désigne un chemin de longueur k , avec donc k arêtes et $k + 1$ sommets.
- Pour $W \subseteq V$, on note $G[W] = (W, E \cap \mathcal{P}_2(W))$ le *sous-graphe de G induit par W* .
- Pour $F \subseteq E$, on note $G[F] = (V, F)$ le *graphe partiel de G induit par F* .
- On dit qu'un graphe G' *contient une copie induite du graphe G* s'il existe un sous-graphe de G' qui est isomorphe à G . Et on dira que G' *contient une copie de G* s'il existe un sous-graphe partiel de G' isomorphe à G .
- On note enfin $\text{aut}(G)$ le nombre d'automorphismes de G .

L'ensemble des automorphismes de G est un groupe, dont la loi est la composition d'applications, notée \circ . En effet, chaque élément de cet ensemble est une application bijective de G dans lui-même, et donc tout élément possède un inverse. L'identité fait office d'élément neutre, et enfin la composition d'applications est toujours associative. Elle l'est donc en particulier dans cet ensemble.

Pour un graphe G donné, un automorphisme est en fait une application bijective $\sigma : V \rightarrow V$ telle que pour toute paire de sommets voisins i et j on a $\sigma(i)$ voisin de $\sigma(j)$. Ainsi un automorphisme de G préserve les degrés de chaque sommet de G . Par exemple, dans la figure 1.1, un automorphisme du graphe considéré doit envoyer la sommet a vers un autre sommet de degré 3, c'est-à-dire le a ou le c . Ensuite il y a également deux possibilités de fixer le sommet b . Et une fois que les sommets a et b ont été fixés, il n'y a plus qu'une seule possibilité pour fixer les deux autres. Au final on obtient dans ce cas $\text{aut}(G) = 4$.

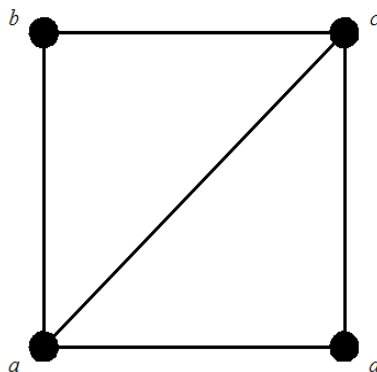


FIG. 1.1 – Exemple de graphe.

Probabilités

Les principales définitions, comme par exemple celle d'espace de probabilité, de variable aléatoire, ou encore celle d'espérance, peuvent être trouvées dans le livre [2]. Les variables aléatoires qui vont principalement nous intéresser sont les binomiales, les normales, celles de Poisson et les uniformes. On écrira par exemple $X \sim \mathcal{B}(n, p)$ pour dire que la variable X suit une loi binomiale de paramètres n et p . On rappelle alors que dans ce cas, $\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ représente la probabilité d'obtenir k succès ($k \in [n] = \{1, 2, \dots, n\}$) lors de n expériences indépendantes, sachant que chacune d'elles a une probabilité $p \in [0, 1]$ de succès.

On écrira aussi $X \sim \mathcal{P}(\lambda)$ pour dire que X suit une loi de Poisson de paramètre λ , autrement dit $\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$. Cette loi nous sera utile pour approximer des binomiales. En effet, soit $X_n \sim \mathcal{B}(n, p_n)$, avec $\lim_{n \rightarrow \infty} p_n = 0$ et $\lim_{n \rightarrow \infty} np_n = \lambda$. Il existe alors une suite (λ_n) avec $p_n = \lambda_n/n$ pour tout $n \in \mathbb{N}$, et $\lim_{n \rightarrow \infty} \lambda_n = \lambda$. De plus,

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(X_n = k) &= \lim_{n \rightarrow \infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} \\ &= \lim_{n \rightarrow \infty} \frac{n!}{k! (n-k)!} p_n^k (1-p_n)^{n-k} \\ &= \frac{1}{k!} \lim_{n \rightarrow \infty} n^k p_n^k \left(1 - \frac{\lambda_n}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} e^{-\lambda}, \quad \forall k \in \mathbb{N}. \end{aligned}$$

Autrement dit, $\lim_{n \rightarrow \infty} \mathbb{P}(X_n = k) = \mathbb{P}(Y = k)$, où $Y \sim \mathcal{P}(\lambda)$.

Finalement, l'écriture $X \sim \mathcal{N}(\mu, \sigma^2)$ nous dit que X suit une loi normale d'espérance μ et de variance σ^2 . Cette loi sert aussi à approximer des binomiales, par le biais de la loi des grands nombres et du théorème central limite que l'on retrouve dans le livre [2]. On peut la définir grâce à sa fonction de densité $f_X(x) = (\sqrt{2\pi}\sigma)^{-1} \exp\{(x - \mu)^2/2\}$. Quant à la loi uniforme que l'on écrira $\mathcal{U}(a, b)$, elle sera utilisée dans le modèle du même nom, et sa fonction de densité est $f(x) = 1/(b-a)$ si $x \in]a, b[$ et 0 sinon. Et de manière plus générale on écrira $\mathcal{L}(X)$ la loi de la variable aléatoire X .

Nous aurons encore besoin dans tout ce projet de quelques résultats de la théorie des probabilités, dont les preuves sont exposées dans les livres [2] et [3]. Il y a par exemple l'*inégalité de Chebyshev*, qui dit que si la variance de la variable aléatoire X existe, alors

$$\forall \varepsilon > 0, \quad \mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2}. \quad (1.2)$$

De plus, si $X \geq 0$ presque sûrement¹, alors on a l'*inégalité de Markov* :

$$\forall \varepsilon > 0, \quad \mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}(X)}{\varepsilon}. \quad (1.3)$$

¹Cela signifie que $\mathbb{P}(X < 0) = 0$

Considérons maintenant un ensemble fini quelconque Γ , avec $|\Gamma| = N$, et notons Γ_{p_1, \dots, p_N} l'ensemble aléatoire où chaque élément i de Γ apparaît avec une probabilité p_i , et de manière indépendante des autres. Soit ensuite \mathcal{S} une famille de sous-ensembles non-vides de Γ , et pour tout $A \in \mathcal{S}$, soit $I_A = \mathbb{1}_{[A \subseteq \Gamma_{p_1, \dots, p_N}]}$.² Posons ensuite $X = \sum_{A \in \mathcal{S}} I_A$ le nombre de sous-ensembles de \mathcal{S} contenus dans Γ_{p_1, \dots, p_N} . On a alors

$$\mathbb{P}(X = 0) \geq \exp \left\{ -\frac{\mathbb{E}(X)}{1 - \max_i p_i} \right\}. \quad (1.4)$$

En posant ensuite $\lambda = \mathbb{E}(X)$ et $\Delta = \frac{1}{2} \sum_{A \neq B} \sum_{A \cap B \neq \emptyset} \mathbb{E}(I_A I_B)$, on a

$$(i) \quad \mathbb{P}(X = 0) \leq \exp(-\lambda + \Delta), \quad (1.5)$$

$$(ii) \quad \mathbb{P}(X = 0) \leq \exp \left(-\frac{\lambda^2}{\lambda + 2\Delta} \right) = \exp \left(-\frac{\lambda^2}{\sum \sum_{A \cap B \neq \emptyset} \mathbb{E}(I_A I_B)} \right). \quad (1.6)$$

Comportement asymptotique

Dans ce projet, nous allons souvent étudier comment un graphe se comporte lorsque son nombre de sommets tend vers l'infini, raison pour laquelle nous devons introduire les notations qui suivront. Soient deux suites (a_n) et (b_n) , avec $b_n > 0$ pour n suffisamment grand. On écrit alors :

- $a_n = O(b_n)$ s'il existe des constantes $C > 0$ et $n_0 \in \mathbb{N}$ telles que $|a_n| \leq Cb_n$ pour tout $n \geq n_0$.
- $a_n = \Omega(b_n)$ s'il existe des constantes $c > 0$ et $n_0 \in \mathbb{N}$ telles que $a_n \geq cb_n$ pour tout $n \geq n_0$. Si $a_n \geq 0$, ceci revient donc à dire que $b_n = O(a_n)$.
- $a_n = \Theta(b_n)$ s'il existe des constantes $C, c > 0$ et $n_0 \in \mathbb{N}$ telles que $cb_n \leq a_n \leq Cb_n$ pour tout $n \geq n_0$. Ceci revient à dire que $a_n = O(b_n)$ et $a_n = \Omega(b_n)$. On dit alors que a_n et b_n ont le même ordre de grandeur. On peut aussi l'écrire $a_n \asymp b_n$.
- $a_n \sim b_n$ si $a_n/b_n \rightarrow 1$.
- $a_n = o(b_n)$ si $a_n/b_n \rightarrow 0$, autrement dit pour tout $\varepsilon > 0$, il existe $n_\varepsilon \in \mathbb{N}$ tel que $|a_n| < \varepsilon b_n$ pour tout $n \geq n_\varepsilon$.
- $a_n \ll b_n$ ou $b_n \gg a_n$ si $a_n \geq 0$ et $a_n = o(b_n)$.

On peut aussi utiliser ces mêmes notations avec des suites de variables aléatoires.

On dira finalement qu'un événement \mathcal{E}_n arrive *asymptotiquement presque sûrement* (abrégé *a.p.s.*) si $\mathbb{P}(\mathcal{E}_n) = 1$ lorsque n tend vers l'infini.

²L'écriture $\mathbb{1}_{\mathcal{E}}$ désigne l'indicatrice de l'événement \mathcal{E} , qui vaut 1 si \mathcal{E} est réalisé et 0 sinon.

1.2 Présentation des modèles de base

Le graphe aléatoire binomial

On se donne un nombre réel p compris entre 0 et 1. Le *graphe aléatoire binomial*, noté $\mathbb{G}(n, p)$, peut être défini en prenant pour Ω l'ensemble de tous les graphes possédant n sommets, et en posant, pour tout $G \in \Omega$,

$$\mathbb{P}(G) = p^{e_G} (1-p)^{\binom{n}{2}-e_G}. \quad (1.7)$$

Il peut donc être vu comme le résultat de $\binom{n}{2}$ expériences indépendantes, une pour chaque paire de sommets distincts, ayant chacune une probabilité de succès égale à p . On peut remarquer qu'en prenant $p = 1/2$, on retrouve le modèle de 1947 et l'équation (1.1).

De manière plus générale, on peut prendre pour Ω l'ensemble des graphes partiels d'un graphe F donné, et on aurait alors $\mathbb{P}(G) = p^{e_G} (1-p)^{e_F-e_G}$. En prenant $F = K_n$, on retrouve l'équation (1.7). On peut aussi considérer le *graphe aléatoire biparti* $\mathbb{G}(m, n, p)$ en posant $F = K_{m,n}$.

L'avantage de ce modèle est que la présence ou non d'une arête est un événement indépendant de celle des autres arêtes. Mais l'inconvénient est que le nombre d'arêtes n'est pas fixé. Il varie aléatoirement selon une loi binomiale dont l'espérance vaut $\binom{n}{2}p$.

Le graphe aléatoire uniforme

Soit M un nombre entier compris entre 0 et $\binom{n}{2}$. Le *graphe aléatoire uniforme*, noté $\mathbb{G}(n, M)$, peut alors être défini en prenant pour Ω l'ensemble des graphes ayant n sommets et M arêtes, et en posant, pour tout $G \in \Omega$,

$$\mathbb{P}(G) = \left(\frac{\binom{n}{2}}{M} \right)^{-1}. \quad (1.8)$$

Cette probabilité est donc identique pour tous les graphes appartenant à Ω . On peut d'ailleurs étendre ce modèle à toute famille \mathcal{F} de graphes ayant une même propriété. Par exemple on peut considérer l'ensemble de tous les graphes r -réguliers (i.e. les graphes dont chaque sommet est de degré r), ou encore l'ensemble de tous les arbres à n sommets.

On peut constater que si l'on suppose $\binom{n}{2}p$ proche de M , alors les deux modèles sont asymptotiquement équivalents. Or dans ce projet nous allons essentiellement étudier les cas où le nombre de sommets tend vers l'infini. On va alors se concentrer sur le modèle binomial en supposant p proche de 0.

Les graphes aléatoires évolutifs

On s'intéresse ici à un processus stochastique où l'on part d'un graphe vide, et où l'on rajoute au fur et à mesure des arêtes. On obtient alors une famille de graphes $\{\mathbb{G}(t)\}_t$, où le paramètre t peut être interprété comme le temps. C'est un processus de Markov qui peut être soit à temps discret, soit à temps continu.

Erdős et Rényi étudièrent le processus à temps discret en 1959, et le notèrent $\{G(n, M)\}_M$. On part donc du graphe vide, et ensuite la M -ème étape peut être identifiée au graphe aléatoire

uniforme $\mathbb{G}(n, M)$. Ce processus permet alors d'étudier $\mathbb{G}(n, M)$ lorsque M va de 0 à $\binom{n}{2}$. On pourrait par exemple étudier le nombre d'étapes nécessaire afin d'obtenir un graphe connexe.

Dans le cas où le temps est continu, on associe une variable aléatoire T_e à chaque arête e du graphe K_n . On suppose que ces $\binom{n}{2}$ variables sont indépendantes et identiquement distribuées. Ainsi $\{\mathbb{G}(t)\}_t$ comprend toutes les arêtes e avec $T_e \leq t$. Donc pour un temps fixé t_0 , $\{\mathbb{G}(t)\}_{t_0}$ correspond au graphe aléatoire binomial $\mathbb{G}(n, p)$, où $p = \mathbb{P}(T_e \leq t_0)$. On peut aussi définir la variable aléatoire $T_{(i)}$, pour représenter le moment auquel la i -ème arête est ajoutée. On retrouve alors le graphe aléatoire uniforme $\mathbb{G}(n, i)$, et de plus le processus $\{\mathbb{G}(T_{(i)})\}$, avec i allant de 1 à $\binom{n}{2}$ est équivalent au processus $\{\mathbb{G}(n, M)\}_M$ décrit ci-dessus. Par conséquent, le processus à temps continu permet de généraliser les graphes aléatoires binomial et uniforme, ainsi que le processus à temps discret.

Chapitre 2

Outils mathématiques

Avant d'entrer dans le vif du sujet, nous avons besoin de présenter ici les outils que nous allons utiliser tout au long du projet. Il s'agit de la méthode des moments, de la méthode de Stein, ainsi que de divers résultats présents dans le livre [3]. Ensuite nous présenterons le concept des valeurs critiques que nous utiliserons beaucoup par la suite.

2.1 La méthode des moments

Soit X une variable aléatoire non-négative et à valeurs entières. On a alors, d'après l'inégalité de Markov (1.3),

$$\mathbb{P}(X > 0) \leq \mathbb{E}(X). \quad (2.1)$$

En effet, X est à valeurs dans \mathbb{N} , et ensuite le résultat ci-dessus correspond au cas où l'on pose $\varepsilon = 1$ dans l'inégalité de Markov. La *méthode du premier moment* consiste alors à montrer que pour une suite (X_n) de variables aléatoires, on a $\mathbb{E}(X_n) = o(1)$. Ainsi, l'inégalité (2.1) implique que dans ce cas, $X_n = 0$ *a.p.s.*

Considérons maintenant une variable aléatoire X quelconque, dont l'espérance est positive et dont la variance existe. On a

$$\begin{aligned} \mathbb{P}(X = 0) &= 1 - \mathbb{P}(X \neq 0) \\ &= 1 - \mathbb{P}(-|X| < 0) \\ &\leq 1 - \mathbb{P}(\mathbb{E}(X) - |X| \leq \mathbb{E}(X)) \\ &\leq 1 - \mathbb{P}(|X - \mathbb{E}(X)| \leq \mathbb{E}(X)) \\ &= \mathbb{P}(|X - \mathbb{E}(X)| > \mathbb{E}(X)). \end{aligned}$$

En posant ensuite $\varepsilon = \mathbb{E}(X)$ dans l'inégalité de Chebyshev (1.2), on obtient

$$\mathbb{P}(X = 0) \leq \frac{\text{Var}(X)}{(\mathbb{E}(X))^2}. \quad (2.2)$$

La *méthode du deuxième moment* consiste alors à montrer que pour une suite de variables aléatoires non-négatives (X_n) , le terme de droite de l'inégalité (2.2) converge vers 0. Ceci implique donc que $X_n > 0$ *a.p.s.*

On peut encore remarquer que l'inégalité (2.2) peut être améliorée, dans le sens où il est possible de diminuer la valeur du terme de droite. En effet, on a $X = X \cdot \mathbb{1}_{[X \neq 0]}$, et ensuite l'inégalité de Cauchy-Schwarz permet de dire que $(\mathbb{E}(X))^2 \leq \mathbb{E}(X^2) \mathbb{P}(X \neq 0)$, ou encore

$$\mathbb{P}(X \neq 0) \geq \frac{(\mathbb{E}(X))^2}{\mathbb{E}(X^2)}, \quad (2.3)$$

puis finalement

$$\mathbb{P}(X = 0) \leq 1 - \frac{(\mathbb{E}(X))^2}{\mathbb{E}(X^2)} = \frac{\text{Var}(X)}{\mathbb{E}(X^2)} = \frac{\text{Var}(X)}{(\mathbb{E}(X))^2 + \text{Var}(X)}. \quad (2.4)$$

De manière générale, la méthode des moments permet de démontrer des convergences en distribution. C'est d'ailleurs l'une des plus vieilles méthodes qui permette de le faire, et on l'utilise toujours aujourd'hui car elle est conceptuellement assez simple et bien adaptée aux problèmes de combinatoire. Son désavantage réside dans le fait qu'elle nécessite souvent de longs calculs. De nombreux théorèmes ont été d'abord prouvés par le biais de cette méthode, avant que d'autres preuves plus élégantes fussent trouvées.

On appelle *moments d'une variable aléatoire* X les nombres $\mathbb{E}(X^k)$, pour $k \geq 1$. On suppose ici que ces nombres existent, et donc que $\mathbb{E}(|X^k|) < \infty$ pour tout $k > 0$. On dit alors que X a des moments finis. Il est clair que les moments d'une variable aléatoire sont définis par sa distribution. La réciproque n'est pas vraie en général, mais elle l'est dans la plupart des cas importants¹. On dira alors que la distribution d'une variable aléatoire X est *déterminée par ses moments* si elle a des moments finis et chaque variable aléatoire ayant des moments égaux à ceux de X possède la même distribution.

Une condition suffisante pour que X soit déterminée par ses moments est que la *fonction génératrice de ses moments*, notée $\varphi_X(t) = \mathbb{E}(e^{tX})$, soit finie pour tout t dans un voisinage de 0. Ceci est en particulier le cas si X suit une loi normale ou poissonnienne. La méthode des moments peut alors être résumée dans le théorème qui suit. Une preuve se trouve dans le livre [1].

Théorème 2.1. *Soit Z une variable aléatoire dont la distribution est déterminée par ses moments. Soient X_1, X_2, \dots des variables aléatoires ayant des moments finis et telles que pour tout entier $k \geq 1$, $\mathbb{E}(X_n^k)$ tend vers $\mathbb{E}(Z^k)$ lorsque n tend vers l'infini. Alors X_n converge en distribution vers Z .*

Pour une variable aléatoire X ayant des moments finis on peut aussi définir son k -ème moment factoriel par le nombre $\mathbb{E}(X)_k = \mathbb{E}[X(X-1) \cdots (X-k+1)]$, pour tout $k \in \mathbb{N}$. On pose par convention $\mathbb{E}(X)_0 = 1$. Cette définition est notamment utile pour le cas où X suit une loi de Poisson. En effet, il est facile de voir que le k -ème moment factoriel de $X \sim \mathcal{P}(\lambda)$ vaut λ^k . On peut aussi reformuler le théorème 2.1 de la manière suivante.

Théorème 2.2. *Soit X une variable aléatoire dont la distribution est déterminée par ses moments. Soient X_1, X_2, \dots des variables aléatoires ayant des moments finis et telles que pour tout entier $k \geq 1$, $\mathbb{E}(X_n)_k$ tend vers $\mathbb{E}(X)_k$ lorsque n tend vers l'infini. Alors X_n converge en distribution vers X .*

¹Voir le livre [1] pour plus de détails

2.2 La méthode de Stein

Cas poissonien

Une méthode pour montrer la convergence en distribution vers une loi normale fut étudiée par Charles Stein. Nous la présenterons plus tard. Nous allons tout d'abord nous intéresser au cas poissonien, qui fut étudié par Louis H. Y. Chen en se basant sur les travaux de Stein. De manière générale leur méthode consiste non seulement à montrer des convergences en distribution, mais aussi à estimer la distance séparant la distribution d'une variable aléatoire et celle d'une loi poissonienne ou normale. Plusieurs distances peuvent être définies, mais dans le cas poissonien on utilisera surtout la *distance de variation totale*, qui vaut

$$d_{TV}(X, Y) = \sup_A |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)|,$$

où X et Y sont des variables aléatoires quelconques, et le supremum est pris sur tous les boréliens A . On peut voir que si l'on considère une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ et une suite de nombres positifs $(\lambda_n)_{n \in \mathbb{N}}$ qui converge vers λ , alors $d_{TV}(X_n, \mathcal{P}(\lambda_n)) \rightarrow 0$ si et seulement si X_n converge en distribution vers $\mathcal{P}(\lambda)$.

De manière générale la méthode de Stein dans le cas poissonien se base sur le théorème suivant. La démonstration peut se trouver dans les livres de Chen ou de Stein lui-même.

Théorème 2.3. *Soit X une variable aléatoire que l'on peut écrire sous la forme $X = \sum_{\alpha \in A} I_\alpha$, où les I_α sont des variables aléatoires indicatrices. Supposons que pour tout $\alpha \in A$ il existe une famille de variables aléatoires $J_{\beta\alpha}$, avec $\beta \in A \setminus \{\alpha\}$, telles que*

$$\mathcal{L}(\{J_{\beta\alpha}\}_\beta) = \mathcal{L}(\{I_\beta\}_\beta | I_\alpha = 1). \quad (2.5)$$

Posons ensuite $\pi_\alpha = \mathbb{E}(I_\alpha)$, et $\lambda = \mathbb{E}(X) = \sum_{\alpha \in A} \pi_\alpha$. Alors

$$d_{TV}(X, \mathcal{P}(\lambda)) \leq \min(\lambda^{-1}, 1) \cdot \sum_{\alpha \in A} \pi_\alpha \left(\pi_\alpha + \sum_{\beta \neq \alpha} \mathbb{E}(|J_{\beta\alpha} - I_\beta|) \right). \quad (2.6)$$

Une manière d'utiliser ce théorème sans construire explicitement les variables $J_{\beta\alpha}$ est de passer par un *graphe de dépendance*. Soit $\{X_i\}_{i \in \mathcal{I}}$ une famille de variables aléatoires définies sur un même espace de probabilité. Un graphe de dépendance pour $\{X_i\}$ est un graphe L dont l'ensemble des sommets est \mathcal{I} et dont les arêtes relient les variables qui sont dépendantes. De plus, si A et B sont deux sous-ensembles disjoints de \mathcal{I} tels que $e_L(A, B) = 0$, alors $\{X_i\}_{i \in A}$ et $\{X_i\}_{i \in B}$ sont mutuellement indépendantes.

Si la famille $\{I_\alpha\}$ a un graphe de dépendance L , alors il existe des variables aléatoires $J_{\beta\alpha}$ satisfaisant l'équation (2.5), et telles que $J_{\beta\alpha} = I_\beta$ lorsque $(\alpha, \beta) \notin E(L)$. Donc dans l'équation (2.6) il suffit de considérer les β qui sont adjacents à α dans L . On peut alors améliorer le théorème 2.3, en utilisant notamment les relations $|J_{\beta\alpha} - I_\beta| \leq J_{\beta\alpha} + I_\beta$, et $\pi_\alpha \mathbb{E}(J_{\beta\alpha}) = \pi_\alpha \mathbb{E}(I_\beta | I_\alpha = 1) = \mathbb{P}(I_\beta = I_\alpha = 1) = \mathbb{E}(I_\alpha I_\beta)$.

On arrive alors au corollaire suivant.

Corollaire 2.4. *Supposons que $X = \sum_{\alpha \in A} I_\alpha$, où les I_α sont des variables aléatoires indicatrices avec un graphe de dépendance L . En utilisant les mêmes notations qu'au théorème 2.3 on obtient*

$$\begin{aligned} d_{TV}(X, \mathcal{P}(\lambda)) &\leq \min(\lambda^{-1}, 1) \cdot \left(\sum_{\alpha \in A} \pi_\alpha^2 + \sum_{\alpha, \beta: (\alpha, \beta) \in E(L)} (\mathbb{E}(I_\alpha I_\beta) + \mathbb{E}(I_\alpha)\mathbb{E}(I_\beta)) \right) \\ &= \min(\lambda^{-1}, 1) \cdot \left(\text{Var}(X) - \mathbb{E}(X) + 2 \sum_{\alpha, \beta: (\alpha, \beta) \in E(L)} \pi_\alpha \pi_\beta + 2 \sum_{\alpha \in A} \pi_\alpha^2 \right). \end{aligned}$$

On peut obtenir un cas encore plus simple du théorème 2.3 lorsque que $J_{\beta\alpha} - I_\beta$ garde toujours le même signe. On dit que les variables aléatoires indicatrices $(I_\alpha)_{\alpha \in A}$ sont *positivement corrélées* si pour tout $\alpha \in A$ il existe des variables aléatoires $J_{\beta\alpha}$ dont la distribution satisfait l'équation (2.5), et telles que $J_{\beta\alpha} \geq I_\beta$ pour tout $\beta \neq \alpha$. Dans le cas contraire les (I_α) sont dites *négativement corrélées*.

Dans le cas où elles sont positivement corrélées on obtient l'expression $\pi_\alpha \mathbb{E}(|J_{\beta\alpha} - I_\beta|) = \mathbb{E}(I_\alpha I_\beta) - \pi_\alpha \pi_\beta$, ce qui nous amène au théorème suivant, sur lequel se base la méthode de Stein-Chen.

Théorème 2.5. *Supposons que $X = \sum_{\alpha \in A} I_\alpha$, où les I_α sont des variables aléatoires indicatrices qui sont positivement corrélées. Alors en reprenant les notations du théorème 2.3 on obtient*

$$\begin{aligned} d_{TV}(X, \mathcal{P}(\lambda)) &\leq \min(\lambda^{-1}, 1) \cdot \left(\text{Var}(X) - \mathbb{E}(X) + 2 \sum_{\alpha \in A} \pi_\alpha^2 \right) \\ &\leq \frac{\text{Var}(X)}{\mathbb{E}(X)} - 1 + 2 \max_{\alpha \in A} \pi_\alpha. \end{aligned}$$

Autrement dit, pour qu'une somme de variables aléatoires positivement corrélées converge en distribution vers une loi de Poisson, il suffit que la probabilité de chaque élément de cette somme tende vers 0, et que sa variance soit asymptotiquement égale à son espérance. Ceci est une condition très naturelle dans le cas poissonien, car on sait que si $X \sim \mathcal{P}(\lambda)$, alors $\mathbb{E}(X) = \text{Var}(X) = \lambda$.

Le problème majeur de la méthode de Stein consiste donc à trouver cette famille de variables positivement corrélées. Ensuite elle nécessite beaucoup moins de calculs que la méthode des moments.

Cas normal

Le cas normal représente la version "originale" de la méthode de Stein. On se base à nouveau sur une distance entre distributions notée cette fois-ci $d_1(X, Y)$ et définie par

$$d_1(X, Y) = \sup\{ |\mathbb{E}(h(X)) - \mathbb{E}(h(Y))| : \sup_{x \in \mathbb{R}} |h(x)| + \sup_{x \in \mathbb{R}} |h'(x)| \leq 1 \}.$$

Et comme dans le cas précédent on remarque que $d_1(X_n, Y) \rightarrow 0$ implique que X_n converge en distribution vers Y . La méthode de Stein dans le cas normal se base alors sur le théorème suivant.

Théorème 2.6. Soit W une variable aléatoire que l'on peut décomposer comme suit. On considère deux ensembles finis d'indices A et B_α , où $\alpha \in A$, et des variables aléatoires $X_\alpha, W_\alpha, Z_\alpha, Z_{\alpha\beta}, W_{\alpha\beta}$ et $V_{\alpha\beta}$ dont le carré est intégrable au sens de Lebesgue pour tous $\alpha \in A, \beta \in B_\alpha$, et telles que

$$\begin{aligned} W &= \sum_{\alpha \in A} X_\alpha; \\ W &= W_\alpha + Z_\alpha, \quad \alpha \in A; \\ Z_\alpha &= \sum_{\beta \in B_\alpha} Z_{\alpha\beta}, \quad \alpha \in A; \\ W_\alpha &= W_{\alpha\beta} + V_{\alpha\beta}, \quad \alpha \in A, \beta \in B_\alpha. \end{aligned}$$

On suppose de plus que $\mathbb{E}(X_\alpha) = 0$, que W_α est indépendant de X_α , et que $W_{\alpha\beta}$ est indépendant de la paire $(X_\alpha, Z_{\alpha\beta})$. Et on pose ensuite $\sigma^2 = \text{Var}(W)$ et $\widetilde{W} = \sigma^{-1}W$. Alors il existe une constante C telle que

$$\begin{aligned} d_1(\widetilde{W}, \mathcal{N}(0, 1)) &\leq C\sigma^{-3} \left(\sum_{\alpha \in A} \mathbb{E}(|X_\alpha|Z_\alpha^2) \right. \\ &\quad \left. + \sum_{\alpha \in A} \sum_{\beta \in B_\alpha} (\mathbb{E}|X_\alpha Z_{\alpha\beta} V_{\alpha\beta}| + \mathbb{E}|X_\alpha Z_{\alpha\beta}| \cdot \mathbb{E}|Z_\alpha + V_{\alpha\beta}|) \right) \end{aligned}$$

De manière générale la méthode de Stein dans le cas normal consiste alors à trouver une décomposition ad hoc de la variable aléatoire que l'on considère, et de rendre le terme de droite de l'inégalité ci-dessus le plus petit possible. Mais dans la pratique on préfère utiliser le théorème qui suit, et dont la démonstration se trouve dans le livre [3].

Théorème 2.7. Soit $(S_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires. Pour tout n on suppose qu'il existe un ensemble A_n tel que $S_n = \sum_{\alpha \in A_n} X_{n\alpha}$, où $\{X_{n\alpha}\}_\alpha$ est une famille de variables aléatoires avec graphe de dépendance L_n . On suppose ensuite qu'il existe des nombres M_n et Q_n tels que pour tous $\alpha_1, \alpha_2 \in A_n$ on ait

$$\sum_{\alpha \in A_n} \mathbb{E}(|X_{n\alpha}|) \leq M_n, \quad \text{et} \quad \sum_{\alpha \in \overline{N}_{L_n}(\alpha_1, \alpha_2)} \mathbb{E}(|X_{n\alpha}| | X_{n\alpha_1}, X_{n\alpha_2}) \leq Q_n.$$

Notons ensuite $\sigma_n^2 = \text{Var}(S_n)$ et $\widetilde{S}_n = \sigma_n^{-1}(S_n - \mathbb{E}(S_n))$. Alors

$$d_1(\widetilde{S}_n, \mathcal{N}(0, 1)) = O\left(\frac{M_n Q_n^2}{\sigma_n^3}\right).$$

En particulier, si le terme de droite ci-dessus tend vers 0, alors \widetilde{S}_n converge en distribution vers une loi normale $\mathcal{N}(0, 1)$.

On remarque donc que, tout comme dans le cas poissonien, la méthode de Stein dans le cas normal nécessite moins de calculs que la méthode des moments. Le problème consiste cette fois-ci à trouver les valeurs M_n et Q_n , et à appliquer le théorème ci-dessus.

2.3 Le concept des valeurs critiques

Une des découvertes les plus intrigantes d'Erdős et Rényi dans l'étude des graphes aléatoires fut le phénomène des valeurs critiques. Dans la plupart des cas, la probabilité qu'un graphe satisfasse une propriété passe très vite de 0 à 1 et vice-versa selon l'évolution de son nombre d'arêtes. Dans le cas d'un graphe aléatoire binomial, on dit que $\hat{p}(n)$ est une valeur critique pour la propriété \mathcal{Q} si

$$\mathbb{P}(\mathbb{G}(n, p) \in \mathcal{Q}) \rightarrow \begin{cases} 0 & \text{si } p \ll \hat{p}(n), \\ 1 & \text{si } p \gg \hat{p}(n). \end{cases} \quad (2.7)$$

On traite de manière analogue le cas du graphe aléatoire uniforme, en disant que $\hat{M}(n)$ est un valeur critique pour la propriété \mathcal{Q} si

$$\mathbb{P}(\mathbb{G}(n, M) \in \mathcal{Q}) \rightarrow \begin{cases} 0 & \text{si } M \ll \hat{M}(n), \\ 1 & \text{si } M \gg \hat{M}(n). \end{cases} \quad (2.8)$$

La plupart des résultats que nous étudierons dans les prochains chapitres seront de cette forme. On appellera la première ligne de (2.7) et de (2.8) *cas 0*, tandis que la deuxième sera appelée *cas 1*.

Il faut encore noter que ces valeurs critiques ne sont pas uniques. En effet, soit \hat{p} une valeur critique, et supposons qu'il existe \hat{p}' tel que $\hat{p}' \asymp \hat{p}$. Alors \hat{p}' est aussi une valeur critique pour la propriété étudiée. Par conséquent on pourra toujours supprimer les constantes multiplicatives pour définir des valeurs critiques.

Nous allons maintenant étudier l'importance de ces valeurs critiques à travers l'exemple qui suit.

L'exemple des progressions arithmétiques

On s'intéresse à la présence d'une progression arithmétique de longueur k dans le sous-ensemble aléatoire $[n]_p$, où chaque élément de $[n] = \{1, 2, \dots, n\}$ apparaît avec une probabilité p , et indépendamment des autres éléments. Notons alors X_k la variable aléatoire définissant le nombre de ces progressions arithmétiques.

Pour utiliser la méthode du premier moment, il faut calculer l'espérance de X_k . Pour ce faire, nous avons besoin de connaître le nombre $f(n, k)$ de progressions arithmétiques de longueur k dans $[n]$. En fait on ne s'intéresse qu'à son ordre de grandeur, qui vaut n^2 car toute progression arithmétique est entièrement déterminée par ses deux premiers éléments. On peut alors numérotter ces progressions par $1, \dots, f(n, k)$, et pour chacune d'entre elles on définit une variable indicatrice $\mathbb{1}_i$ qui vaut 1 si la i -ème progression est entièrement présente dans $[n]_p$, et 0 sinon. Avec ces notations on a $X_k = \sum_{i=1}^{f(n, k)} \mathbb{1}_i$. L'espérance de la variable $\mathbb{1}_i$ est égale à la probabilité que la i -ème progression soit dans $[n]_p$, et par linéarité on obtient :

$$\mathbb{E}(X_k) = f(n, k) \mathbb{E}(\mathbb{1}_i) = f(n, k) \mathbb{P}(i) = f(n, k) p^k = \Theta(n^2 p^k).$$

Par conséquent, on remarque que si $p \ll n^{-2/k}$, alors $\lim_{n \rightarrow \infty} pn^{2/k} = 0$, et ceci implique que l'espérance de X_k tend vers 0. On peut alors utiliser la méthode des premiers moments et l'équation (2.1) pour conclure que dans ce cas, $\mathbb{P}(X_k > 0) = o(1)$.

Le cas $p \gg n^{-2/k}$ est plus complexe, car le fait d'avoir $\mathbb{E}(X_k)$ qui tend vers l'infini n'implique pas forcément que $\mathbb{P}(X_k > 0)$ tend vers 1. Il faut alors utiliser la méthode du deuxième moment, et donc calculer la variance de X_k . On remarque tout d'abord que les variables $\mathbb{1}_i$ et $\mathbb{1}_j$ sont indépendantes si les progressions i et j n'ont aucun élément en commun. Dans ce cas on a alors $\text{Cov}(\mathbb{1}_i, \mathbb{1}_j) = 0$.

Pour les autres cas il faut utiliser l'inégalité $\text{Cov}(\mathbb{1}_i, \mathbb{1}_j) \leq \mathbb{E}(\mathbb{1}_i \mathbb{1}_j)$. Si les progressions i et j ont un terme en commun, alors on a $\mathbb{E}(\mathbb{1}_i \mathbb{1}_j) = p^{2k-1}$. Le nombre de couples (i, j) satisfaisant cette propriété est de l'ordre $O(n^3)$. Il y a en effet n choix pour le premier terme de la progression i , n pour le premier terme de la progression j , et enfin $O(n)$ choix pour le terme qu'elles ont en commun.

Il est aussi possible que les progressions i et j ont deux termes ou plus en commun. Dans ce cas on a $\mathbb{E}(\mathbb{1}_i \mathbb{1}_j) \leq p^k$, et il n'y a plus que $O(n^2)$ couples (i, j) qui ont cette propriété. Car une fois que l'on a choisi les termes initiaux des deux progressions, le nombre de possibilités pour les construire ne dépend pas de n , mais plutôt des choix que l'on a faits.

On peut maintenant estimer la variance de X_k comme suit :

$$\text{Var}(X_k) = \sum_{i=1}^{f(n,k)} \sum_{j=1}^{f(n,k)} \text{Cov}(\mathbb{1}_i, \mathbb{1}_j) = O\left(n^3 p^{2k-1} + n^2 p^k\right).$$

En appliquant alors l'inégalité (2.2) et en posant $p \gg n^{-2/k}$, on obtient

$$\mathbb{P}(X_k = 0) = O\left(\frac{1}{np} + \frac{1}{n^2 p^k}\right) = o(1).$$

Avec les résultats obtenus et les expressions (2.7), on peut conclure que la valeur critique pour l'existence d'une progression arithmétique à k termes dans $[n]_p$ est $n^{-2/k}$.

Chapitre 3

Le problème de l'existence d'une copie d'un graphe donné

Le premier problème étudié par Erdős et Rényi en 1960 concernait l'existence d'au moins une copie d'un graphe G donné dans le graphe aléatoire binomial $\mathbb{G}(n, p)$ défini à la section 1.2. Ils trouvèrent la valeur critique de cette propriété pour une famille de graphes appelés *graphes équilibrés*. En 1981, Béla Bollobás traita le problème dans le cas général. Et en 1985, Ruciński et Vince trouvèrent une preuve plus simple, basée sur la méthode des premiers et deuxièmes moments, que nous allons présenter ici. Nous allons tout d'abord traiter un exemple, de la même manière qu'avec les progressions arithmétiques à la fin du chapitre précédent.

3.1 Exemple fondamental

Notons par X_G le nombre de copies d'un graphe G donné dans $\mathbb{G}(n, p)$, v le nombre de sommets de G et e son nombre d'arêtes. Il y a exactement $f(n, G) = \binom{n}{v} v! / \text{aut}(G)$ copies de G dans le graphe complet K_n . En effet, il faut tout d'abord choisir v sommets parmi n pour construire G . Ensuite il faut multiplier $\binom{n}{v}$ par le nombre de permutations qu'il est possible de faire avec ces v sommets, à savoir $v!$. Mais si l'on considère chacune de ces permutations, certaines correspondront exactement au même graphe. Par exemple, si l'on reprend la figure 1.1 et que l'on y permute les sommets a et c et les sommets b et d , on comptera deux fois le même graphe. Ceci est la raison de la division par $\text{aut}(G)$ dans l'expression ci-dessus.

Pour chaque copie G' de G dans K_n , on définit, comme dans l'exemple des progressions arithmétiques, la variable indicatrice $\mathbb{1}_{G'} = \mathbb{1}_{[\mathbb{G}(n, p) \supseteq G']}$, et ensuite

$$\mathbb{E}(X_G) = f(n, G) p^e = \Theta(n^v p^e) \rightarrow \begin{cases} 0 & \text{si } p \ll n^{-v/e}, \\ \infty & \text{si } p \gg n^{-v/e}. \end{cases}$$

Notons alors que la quantité $v! / \text{aut}(G)$ ne joue aucun rôle lorsque n tend vers l'infini. La méthode du premier moment permet ensuite de dire que

$$\mathbb{P}(X_G > 0) = o(1) \quad \text{si } p \ll n^{-v/e}. \quad (3.1)$$

La question est alors de savoir si l'on a $\mathbb{P}(X_G > 0) = 1 - o(1)$ si $p \gg n^{-v/e}$. Pour répondre à cette question, considérons l'exemple qui suit.

Soit H_0 le graphe ayant 4 sommets et 5 arêtes présenté à la figure 1.1, et soit G_0 un graphe obtenu en prenant un cinquième sommet et en le reliant à un sommet de H_0 . Il y a certes plusieurs possibilités non isomorphes d'obtenir G_0 , mais peu importe laquelle on choisit ici. En effet, si l'on reprend la figure 1.1 on constate qu'il est possible par exemple de relier le sommet rajouté au sommet a ou au sommet b , et que les deux graphes ainsi obtenus ne sont pas isomorphes, car dans le premier cas le sommet a serait de degré 4, et qu'aucun sommet n'aurait un tel degré dans le second cas.

Soit ensuite p tel que $n^{-5/6} \ll p \ll n^{-4/5}$, comme par exemple $p = n^{-9/11}$. On applique alors le résultat (3.1) à G_0 , puis à H_0 . On trouve que $\mathbb{E}(X_{G_0})$ tend vers l'infini, mais que *a.p.s.* il n'y a pas de copie de H_0 dans $\mathbb{G}(n, p)$, et donc qu'il n'y a pas non plus de copie de G_0 ! Le cas des graphes est donc plus complexe que celui des progressions arithmétiques. En fait le cas de G_0 est trompeur car il contient un sous-graphe plus dense que lui, à savoir H_0 . Par conséquent la valeur critique correcte pour le cas ci-dessus est $n^{-4/5}$ et non $n^{-5/6}$. Tout ceci fut confirmé par Bollobás en 1981 dans le théorème présenté à la section suivante.

3.2 Le théorème de Bollobás

Ce théorème est utile pour trouver la valeur critique permettant de dire si $\mathbb{G}(n, p)$ contient une copie d'un graphe G donné. Avant de présenter et de démontrer ce théorème, il nous faut les définitions et le lemme qui suivent. On rappelle tout d'abord que pour un graphe G , sa *densité maximale* est donnée par

$$m(G) = \max \left\{ \frac{e_H}{v_H} \mid H \subseteq G, v_H > 0 \right\}.$$

On définit ensuite l'expression Φ_G par

$$\Phi_G = \Phi_G(n, p) = \min \{ \mathbb{E}(X_H) \mid H \subseteq G, e_H > 0 \},$$

en notant que $\Phi_G \asymp \min \{ n^{v_H} p^{e_H} \mid H \subseteq G, e_H > 0 \}$.

Lemme 3.1. *Soit G un graphe ayant au moins une arête. Alors les assertions suivantes sont équivalentes.*

- (i) $np^{m(G)} \rightarrow \infty$,
- (ii) $n^{v_H} p^{e_H} \rightarrow \infty$ pour tout $H \subseteq G$ avec $v_H > 0$,
- (iii) $\mathbb{E}(X_H) \rightarrow \infty$ pour tout $H \subseteq G$ avec $v_H > 0$,
- (iv) $\Phi_G \rightarrow \infty$.

Démonstration. D'après la définition de $m(G)$, (i) est vérifié si et seulement si np^{e_H/v_H} tend vers l'infini pour tout $H \subseteq G$ avec $v_H > 0$. Et étant donné que $\mathbb{E}(X_H) \asymp n^{v_H} p^{e_H} = (np^{e_H/v_H})^{v_H}$, ceci est équivalent à (ii) et à (iii). Finalement, on voit que d'après la définition de Φ_G la condition (iv) revient à dire que $\mathbb{E}(X_H)$ tend vers l'infini pour tout $H \subseteq G$ avec $e_H > 0$. Ceci est équivalent à (iii), le cas où $v_H > 0$ et $e_H = 0$ étant trivial. □

Nous avons alors les armes nécessaires pour présenter et démontrer ce théorème.

Théorème 3.2. *Pour tout graphe G ayant au moins une arête, on a*

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathbb{G}(n, p) \supset G) = \begin{cases} 0 & \text{si } p \ll n^{-1/m(G)}, \\ 1 & \text{si } p \gg n^{-1/m(G)}. \end{cases}$$

Démonstration. Montrons tout d'abord le cas 0. Supposons alors que l'on ait $p \ll n^{-1/m(G)}$, et soit H' un sous-graphe partiel de G tel que $d(H') = m(G)$. Alors d'après (3.1), il n'y a *a.p.s.* aucune copie de H' dans $\mathbb{G}(n, p)$, et par conséquent aucune copie de G .

Le cas 1 se démontre grâce à la méthode du deuxième moment. Il faut alors borner supérieurement la variance de X_G . De manière analogue à celle utilisée dans l'exemple des progressions arithmétiques, on considère la variable indicatrice $\mathbb{1}_{G'}$, qui représente le fait que la copie G' de G soit présente ou non dans $\mathbb{G}(n, p)$. Les variables $\mathbb{1}_{G'}$ et $\mathbb{1}_{G''}$ sont indépendantes si $E(G') \cap E(G'') = \emptyset$. D'autre part, pour tout sous-graphe partiel H de G , il y a $\Theta\left(\binom{n}{v_H} \binom{n}{v_G - v_H}^2\right) = \Theta\left(n^{v_H} n^{2(v_G - v_H)}\right) = \Theta\left(n^{2v_G - v_H}\right)$ couples (G', G'') de copies de G dans le graphe complet K_n avec $G' \cap G''$ isomorphe à H . On obtient alors

$$\begin{aligned} \text{Var}(X_G) &= \sum_{G', G''} \text{Cov}(\mathbb{1}_{G'}, \mathbb{1}_{G''}) \\ &= \sum_{E(G') \cap E(G'') \neq \emptyset} [\mathbb{E}(\mathbb{1}_{G'} \mathbb{1}_{G''}) - \mathbb{E}(\mathbb{1}_{G'}) \mathbb{E}(\mathbb{1}_{G''})] \\ &\asymp \sum_{H \subseteq G, e_H > 0} n^{2v_G - v_H} (p^{2e_G - e_H} - p^{2e_G}) \\ &\asymp \sum_{H \subseteq G, e_H > 0} n^{2v_G - v_H} p^{2e_G - e_H} (1 - p) \\ &\asymp (1 - p) \sum_{H \subseteq G, e_H > 0} \frac{n^{2v_G} p^{2e_G}}{n^{v_H} p^{e_H}} \\ &\asymp (1 - p) \max_{H \subseteq G, e_H > 0} \frac{(\mathbb{E}(X_G))^2}{\mathbb{E}(X_H)} \\ &\asymp (1 - p) \frac{(\mathbb{E}(X_G))^2}{\Phi_G}. \end{aligned}$$

On remarque ensuite que si $p \gg n^{-1/m(G)}$, alors d'après le lemme 3.1 on a $\Phi_G \rightarrow \infty$. Par conséquent, l'inégalité (2.2) nous donne

$$\mathbb{P}(\mathbb{G}(n, p) \not\supset G) = \mathbb{P}(X_G = 0) \leq \frac{\text{Var}(X_G)}{(\mathbb{E}(X_G))^2} = O\left(\frac{1}{\Phi_G}\right) = o(1).$$

□

Comme nous l'avons mentionné au tout début de ce chapitre, Erdős et Rényi avaient déjà démontré ce théorème dans le cas particulier où G est un *graphe équilibré*, c'est-à-dire un graphe tel que $m(G) = e_G/v_G$. Ils différencièrent aussi les graphes équilibrés des graphes

strictement équilibrés, en disant que dans le premier cas, le graphe peut avoir un sous-graphe aussi dense que lui, mais pas dans le second. Leur valeur critique correspondait alors à celle donnée dans l'exemple de la section 3.1. Ainsi il leur fut impossible d'étudier un cas comme celui que nous avons présenté à la fin de ladite section. Ruciński reprit alors leurs idées en supposant que tout graphe peut être considéré comme un sous-graphe partiel d'un graphe équilibré, ce qu'il démontra par la suite.

Comme application de ce théorème on peut voir par exemple que la valeur critique permettant de dire si $\mathbb{G}(n, p)$ contient un triangle, c'est-à-dire le graphe K_3 , est $\hat{p} = 1/n$. En effet on a $m(K_3) = 1$. Autrement dit, on a le résultat suivant :

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathbb{G}(n, p) \supset K_3) = \begin{cases} 0 & \text{si } np \rightarrow 0, \\ 1 & \text{si } np \rightarrow \infty. \end{cases}$$

3.3 Cas intermédiaire

Le théorème 3.2 permet de traiter le cas où $np^{m(G)}$ tend vers 0 ou ∞ . Nous allons maintenant voir ce qui se passe lorsque cette quantité tend vers une constante c positive. On a vu dans la preuve de ce théorème que la quantité Φ_G joue un rôle important. En reprenant cette preuve on peut remarquer que l'on a les inégalités

$$1 - \Phi_G \leq \mathbb{P}(\mathbb{G}(n, p) \not\supset G) \leq O(1/\Phi_G).$$

Nous allons voir maintenant que les deux bornes de ces inégalités peuvent être améliorées.

Lemme 3.3. *Soit G un graphe avec au moins une arête. Alors pour tout $p = p(n) < 1$, on a les inégalités*

$$\exp \left\{ -\frac{1}{1-p} \Phi_G \right\} \leq \mathbb{P}(\mathbb{G}(n, p) \not\supset G) \leq \exp \{-\Theta(\Phi_G)\}.$$

Démonstration. L'inégalité de gauche résulte directement de l'inégalité (1.4), avec X remplacé par $X_{H'}$ où H' est un sous-graphe de G tel que $\mathbb{E}(X_{H'}) = \Phi_G$. L'autre inégalité provient de (1.6), où $S = X_G$ et les I_A sont remplacés par $I_G = \mathbf{1}_{[G \subseteq \mathbb{G}(n, p)]}$. Le dénominateur de l'exposant, dans l'inéquation (1.6), devient alors

$$\sum_{H \subseteq G, e_H > 0} \sum_{G'} \sum_{G'' \cap G' = H} p^{2e_G - e_H} = \Theta \left(\frac{(\mathbb{E}(X_G))^2}{\Phi_G} \right),$$

d'où l'on obtient l'inégalité de droite. Le lemme est ainsi démontré. □

Supposons à présent que nous avons $p = \Theta(n^{-1/m(G)})$. Alors par conséquent $\Phi_G = \Theta(1)$, et d'après le lemme 3.3, nous obtenons

$$0 < \liminf_{n \rightarrow \infty} \mathbb{P}(\mathbb{G}(n, p) \supset G) \leq \limsup_{n \rightarrow \infty} \mathbb{P}(\mathbb{G}(n, p) \supset G) < 1.$$

Or le calcul de ces limites n'est pas facile. Néanmoins on va voir dans le théorème suivant que dans le cas où le graphe G est strictement équilibré, il est possible de non seulement la calculer, mais aussi de déterminer la distribution asymptotique de la variable aléatoire X_G . C'est un résultat qui fut démontré par Bollobás en 1981, ainsi que par Ruciński et Karoński en 1983.

Théorème 3.4. *Si G est un graphe strictement équilibré et $np^{m(G)}$ tend vers $c > 0$, alors X_G converge en distribution vers une loi de Poisson de paramètre $\lambda = c^{v_G}/\text{aut}(G)$.*

Démonstration. La preuve de ce théorème est un exemple d'utilisation de la méthode des moments, et plus précisément du théorème 2.2. On considère le moment factoriel de X_G , défini par $\mathbb{E}(X_G)_k = \mathbb{E}[X_G(X_G - 1)\cdots(X_G - k + 1)]$ pour tout $k = 1, 2, \dots$. Il faut alors montrer que

$$\mathbb{E}(X_G)_k = \sum_{G_1, \dots, G_k} \mathbb{P}(\mathbf{1}_{G_1} \cdots \mathbf{1}_{G_k} = 1) = E'_k + E''_k, \quad (3.2)$$

où la sommation est effectuée sur tous les k -tuples de copies disjointes de G dans K_n , et E'_k représente les k -tuples où l'on a $V(G_i) \cap V(G_j) = \emptyset$ pour tout $i \neq j$.

Le cas où $k = 1$ est trivial. Supposons alors que l'équation ci-dessus est vraie pour un certain $k \in \mathbb{N}$. On a alors

$$\begin{aligned} (X_G)_{k+1} &= (X_G)_k (X_G - k) \\ &= \left(\sum_{G_1, \dots, G_k} \mathbf{1}_{G_1} \cdots \mathbf{1}_{G_k} \right) \cdot \left(\left(\sum_{G_{k+1}} \mathbf{1}_{G_{k+1}} \right) - k \right) \\ &= \underbrace{\sum_{G_1, \dots, G_{k+1}} \mathbf{1}_{G_1} \cdots \mathbf{1}_{G_{k+1}}}_{\text{si } G_{k+1} \notin (G_1, \dots, G_k)} + k \cdot \underbrace{\sum_{G_1, \dots, G_k} \mathbf{1}_{G_1} \cdots \mathbf{1}_{G_k}}_{\text{si } G_{k+1} \in (G_1, \dots, G_k)} - k \cdot \sum_{G_1, \dots, G_k} \mathbf{1}_{G_1} \cdots \mathbf{1}_{G_k} \\ &= \sum_{G_1, \dots, G_{k+1}} \mathbf{1}_{G_1} \cdots \mathbf{1}_{G_{k+1}}. \end{aligned}$$

En prenant ensuite les espérances à gauche et à droite, et selon le principe de la récurrence, l'égalité (3.2) est démontrée pour tout $k \in \mathbb{N}$. Analysons maintenant les termes E'_k et E''_k . On rappelle que $E'_k = \sum_{G_1, \dots, G_k} \mathbb{P}(\mathbf{1}_{G_1} \cdots \mathbf{1}_{G_k} = 1)$, où la somme s'effectue sur tous les k -tuples de copies disjointes de G dans K_n telles que $V(G_i) \cap V(G_j) = \emptyset$ pour tout $i \neq j$. Par conséquent, on a $E'_k \sim (\mathbb{E}(X_G))^k$. De plus, on a vu dans l'exemple au début de ce chapitre que $\mathbb{E}(X_G) = \Theta(n^{v_G} p^{e_G})$. Il faut toutefois diviser cette quantité par le nombre d'automorphismes de G , afin d'éviter qu'une copie de G dans $\mathbb{G}(n, p)$ soit comptée plusieurs fois. On obtient alors

$$E'_k \sim (\mathbb{E}(X_G))^k \sim \left(\frac{c^{v_G}}{\text{aut}(G)} \right)^k, \quad \forall k \in \mathbb{N}.$$

D'après la méthode des moments, ceci implique que X_G converge en distribution vers une loi de Poisson de paramètre $\lambda = c^{v_G}/\text{aut}(G)$ si $E''_k = o(1)$, car le k -ème moment factoriel d'une loi $\mathcal{P}(\lambda)$ vaut λ^k . Il reste donc à montrer que cette dernière égalité est vraie.

Notons e_t le nombre minimal d'arêtes présentes dans une intersection de k copies disjointes de G comportant t sommets. Il faut alors montrer que pour tout $k \geq 2$ et $k \leq t < kv_G$, on a $e_t > tm(G)$. Une fois cette propriété établie, on pourra conclure en disant

$$E''_k = \sum_{t=k}^{kv_G-1} O(n^t p^{e_t}) = o(1).$$

Pour un graphe F quelconque, on définit $f_F = m(G)v_F - e_F$. On remarque que $f_G = 0$, et que comme G est strictement équilibré, on a $f_H > 0$ pour tout sous-graphe H de G . Prenons

alors pour F la réunion de k copies de G en supposant qu'il existe i et j dans $\{1, 2, \dots, k\}$ avec $V(G_i) \cap V(G_j) \neq \emptyset$. Il faut alors montrer que $f_F < 0$, en procédant par récurrence et en utilisant l'égalité $f_{F_1 \cup F_2} = f_{F_1} + f_{F_2} - f_{F_1 \cap F_2}$. Posons alors $F = \bigcup_{i=1}^k G_i$, où les G_i sont numérotés de manière à avoir $G_1 \cap G_2 \neq \emptyset$. Pour $k = 2$, on obtient $f_{G_1 \cup G_2} = -f_{G_1 \cap G_2} < 0$ car $G_1 \cap G_2$ est un sous-graphe de G .

Dans le cas où $k \geq 3$, posons $F' = \bigcup_{i=1}^{k-1} G_i$ et $H = F' \cap G_k$. Il se peut que $H = \emptyset$, mais quoi qu'il en soit on aura toujours $f_H \geq 0$. On a aussi, d'après l'hypothèse de récurrence, $f_{F'} < 0$. Ainsi

$$f_F = f_{F'} + f_{G_k} - f_H < 0, \quad \forall k \geq 2.$$

□

Bollobás et Wierman démontrèrent en 1989 que le théorème 3.4 ne marche pas si G est équilibré mais pas strictement. On peut se référer à l'exemple 3.21 du livre [3]. Il faut aussi noter que si $p \gg n^{-1/m(G)}$, alors X_G converge en distribution vers une loi normale. Ceci sera l'objet du théorème suivant.

Théorème 3.5. *Posons $\tilde{X}_G = (X_G - \mathbb{E}(X_G)) / (\text{Var}(X_G))^{1/2}$. Soit G un graphe quelconque ayant au moins une arête. Supposons que si $n \rightarrow \infty$, $p = p(n)$ est tel que $np^{m(G)} \rightarrow \infty$ et $n^2(1-p) \rightarrow \infty$. Alors \tilde{X}_G converge en distribution vers une loi normale $\mathcal{N}(0, 1)$.*

Démonstration. Nous allons utiliser la méthode de Stein dans le cas normal, et plus particulièrement le théorème 2.7. Soit $\{G_\alpha\}_{\alpha \in A_n}$ la famille de tous les sous-graphes partiels de K_n isomorphes à G , et posons $I_\alpha = \mathbb{1}_{[G_\alpha \subset \mathbb{G}(n,p)]}$ et $X_\alpha = I_\alpha - \mathbb{E}(I_\alpha)$. On définit ensuite le graphe L_n dont l'ensemble des sommets est A_n , et dont les arêtes relient les indices α et β tels que les graphes G_α et G_β ont une arête commune. Alors L_n est un graphe de dépendance pour $\{X_\alpha\}$, et tout ceci permet en fait de formaliser ce qui a déjà été dit au début de la preuve du cas 1 du théorème 3.2. Il faut maintenant vérifier toutes les hypothèses du théorème 2.7.

On peut tout d'abord remarquer que $\mathbb{E}(|X_\alpha|) = 2\mathbb{E}(I_\alpha)(1 - \mathbb{E}(I_\alpha))$. On obtient alors $\mathbb{E}(|X_\alpha|) = O((1-p)\mathbb{E}(I_\alpha))$, et ainsi $\sum_{\alpha \in A_n} \mathbb{E}(|X_\alpha|) = O((1-p)\mathbb{E}(X_G))$. Par conséquent on peut choisir $M_n = O((1-p)\mathbb{E}(X_G))$.

Soient ensuite $\alpha_1, \alpha_2 \in A_n$. On définit $F = G_{\alpha_1} \cup G_{\alpha_2}$, et $F_\alpha = G_\alpha \cap F$ pour tout $\alpha \in A_n$. Alors $\alpha \in \bar{N}_{L_n}(\alpha_1, \alpha_2)$ si et seulement si $e(F_\alpha) \neq 0$. Le nombre de ces sous-graphes partiels de F est moins que $2^{v_F} \leq 2^{2v_G}$. Et pour chaque sous-graphe partiel $H \subseteq F$ il y a $O(n^{v_G - v_H})$ choix de α tel que $F_\alpha = H$, avec

$$\mathbb{E}(|X_\alpha| | X_{\alpha_1}) \leq \mathbb{E}(I_\alpha | X_{\alpha_1}) + \mathbb{E}(I_\alpha) \leq 2p^{e_G - e_H}.$$

De plus chaque F_α est isomorphe à un sous-graphe partiel de G . Il existe alors une constante B dépendant de G telle que

$$\sum_{\alpha \in \bar{N}_{L_n}(\alpha_1, \alpha_2)} \mathbb{E}(|X_\alpha| | X_{\alpha_1}) \leq B \cdot \sup_{H \subseteq G, e_H \geq 1} \frac{\mathbb{E}(X_G)}{\mathbb{E}(X_H)} = O\left(\frac{\mathbb{E}(X_G)}{\Phi_G}\right).$$

On peut par conséquent choisir $Q_n = O(\mathbb{E}(X_G)/\Phi_G)$. Et durant la démonstration du théorème 3.2 on a vu que l'on peut choisir $\sigma_n^2 = \text{Var}(X_G) \asymp (1-p)(\mathbb{E}(X_G))^2/\Phi(G)$. On obtient alors

$$\frac{M_n Q_n^2}{\sigma_n^3} = O\left((1-p)^{-1/2} \Phi_G^{-1/2}\right). \quad (3.3)$$

Analysons maintenant les autres hypothèses du théorème que nous sommes en train de démontrer. Le lemme 3.1 nous permet de dire que $np^{m(G)} \rightarrow \infty$ est équivalent à $\Phi_G \rightarrow \infty$. Et ceci implique que $(1-p)\Phi_G \rightarrow \infty$ si $p \leq 1/2$. D'autre part, si $p > 1/2$ on a $\Phi_G \asymp n^2$, et ainsi $(1-p)\Phi_G \asymp n^2(1-p) \rightarrow \infty$ par hypothèse.

On a finalement $X_G = \sum_{\alpha} I_{\alpha}$, et donc $X_G - \mathbb{E}(X_G) = \sum_{\alpha} X_{\alpha}$. L'équation (3.3) et le théorème 2.7 permettent alors de dire que \tilde{X}_G converge en distribution vers une loi normale $\mathcal{N}(0, 1)$. □

Chapitre 4

Le problème du recouvrement

On va étudier dans ce chapitre la probabilité que chaque sommet d'un graphe aléatoire appartienne à une copie d'un graphe G donné. Cette propriété sera notée COV_G . On peut remarquer que si G contient un sommet isolé¹, alors la propriété COV_G est égale à la propriété COV_H , où H est le sous-graphe de G ne contenant que les sommets de degré non-nul. On pourra alors supposer dans tout ce chapitre que le degré minimal de G vaut au moins 1.

4.1 L'enracinement d'un graphe

Soit v un sommet particulier d'un graphe aléatoire. Ce sommet peut avoir diverses positions dans une copie de G qui le recouvre, raison pour laquelle on introduit la notion de *graphe enraciné* (v, G) , où $v \in V(G)$ est la *racine* du graphe G . Prenons par exemple un triangle et ajoutons un quatrième sommet, que l'on relie ensuite à l'un des sommets du triangle. On note le graphe obtenu K_3^+ . Il y a alors trois manières non-isomorphes d'enraciner K_3^+ , comme le montre la figure 4.1.

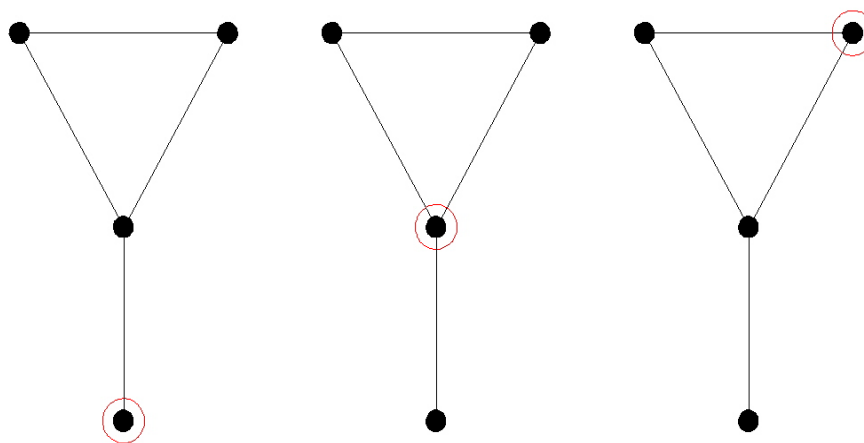


FIG. 4.1 – Les trois versions enracinées de K_3^+ . Les racines sont entourées en rouge.

¹Un sommet isolé est un sommet dont le degré est nul, autrement dit c'est un sommet qui n'est relié à aucun autre.

Pour un graphe enraciné quelconque (v, G) avec $v_G > 1$, on note $d(v, G) = e_G / (v_G - 1)$, et $m(v, G) = \max\{d(v, H) \mid v \in H \subseteq G\}$. En général $m(v, G)$ dépend de v , au contraire de $d(v, G)$. Le graphe enraciné (v, G) sera alors appelé *équilibré* si ces deux quantités sont égales, et *strictement équilibré* si $d(v, H) < d(v, G)$ pour tout sous-graphe H de G contenant le sommet v . On peut alors remarquer que parmi les trois versions de K_3^+ présentées dans la figure 4.1, seule celle de gauche est strictement équilibrée, alors que les deux autres ne sont même pas équilibrées. En effet on a dans ce cas $d(v, G) = 4/3$, et dans les deux versions de droite on peut considérer le sous-graphe K_3 , qui contient la racine et dont la densité vaut $3/2 > 4/3$. Donc dans ces deux cas on a $m(v, G) > d(v, G)$.

4.2 Notations et théorème

Nous avons encore besoin d'autres notations pour étudier ce problème du recouvrement. Pour $i \in \{1, \dots, n\}$ et $v \in V(G)$, on note $U_i(v)$ le nombre de copies de (v, G) contenues dans le graphe aléatoire $\mathbb{G}(n, p)$, et dans lesquelles le sommet i joue le rôle de la racine v . Et on note U_i le nombre total de copies de G contenant i . Ruciński et Vince démontrèrent en 1986, et de manière similaire aux problèmes du chapitre précédent, que pour n'importe quel sommet $i \in V$ la valeur critique pour la propriété " $U_i(v) > 0$ " est $p = n^{-1/m(v, G)}$, et que celle pour la propriété " $U_i > 0$ " est $p = n^{-1/\min_{v \in G} m(v, G)}$. Donc si l'on reprend l'exemple ci-dessus, on remarque que lorsque $p \gg n^{-3/4}$, alors un sommet fixé appartient *a.p.s.* à une copie de K_3^+ . Et c'est seulement lorsque $p \gg n^{-2/3}$ qu'il va appartenir à un triangle.

En fait dans ce chapitre nous serons plutôt intéressés par la variable aléatoire

$$W_G = |\{i \in [n], U_i = 0\}| = \sum_{i=1}^n \mathbf{1}_{[U_i=0]},$$

qui compte le nombre de sommets de $\mathbb{G}(n, p)$ qui ne sont pas recouverts par des copies de G . Par conséquent il y a une équivalence entre la propriété COV_G et la propriété " $W_G = 0$ ". Le théorème suivant va alors nous donner une valeur critique pour cette propriété, qui dépend bien sûr de la structure de G .

Pour un graphe G , on note $m_* = \min_{v \in G} m(v, G)$ et l'ensemble des sommets de G pour lesquels $m(v, G) = m_*$ est noté $M(G)$. Ensuite, pour un sommet $v \in M(G)$, on écrit :

- \mathcal{C}_v la famille de tous les sous-graphes H de G qui contiennent v et satisfont les conditions $d(v, H) = m_*$ et $N_H(v) \neq \emptyset$,²
- $s_v = \min_{H \in \mathcal{C}_v} e(H)$, et l'on pose $s_v = \infty$ si $\mathcal{C}_v = \emptyset$,
- $s = \max_{v \in M(G)} s_v$,
- $a = |M(G)| / \text{aut}(G)$.

Avec ces notations nous pouvons dès à présent attaquer ce théorème.

²Cette deuxième condition veut simplement dire que v n'est pas un sommet isolé de H .

Théorème 4.1. *Soit G un graphe dont le degré minimal vaut au moins 1.*

- (i) *Si pour tout sommet $v \in M(G)$, le graphe enraciné (v, G) est strictement équilibré, alors*

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathbb{G}(n, p) \in \text{COV}_G) = \begin{cases} 0 & \text{si } an^{v_G-1}p^{e_G} - \log(n) \rightarrow -\infty, \\ 1 & \text{si } an^{v_G-1}p^{e_G} - \log(n) \rightarrow \infty. \end{cases}$$

De plus, si $an^{v_G-1}p^{e_G} - \log(n) \rightarrow c \in \mathbb{R}$, alors W_G converge en distribution vers une loi de Poisson de paramètre e^{-c} , et par conséquent on a $\mathbb{P}(\mathbb{G}(n, p) \in \text{COV}_G) \rightarrow \exp(-e^{-c})$.

- (ii) *Si $s < \infty$, alors il existe des constantes $c, C > 0$ telles que*

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathbb{G}(n, p) \in \text{COV}_G) = \begin{cases} 0 & \text{si } p \leq c(\log(n))^{1/s} n^{-1/m_*}, \\ 1 & \text{si } p \geq C(\log(n))^{1/s} n^{-1/m_*}. \end{cases}$$

- (iii) *Enfin, si $s = \infty$ alors*

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathbb{G}(n, p) \in \text{COV}_G) = \begin{cases} 0 & \text{si } p \ll n^{-1/m_*}, \\ 1 & \text{si } p \gg n^{-1/m_*}. \end{cases}$$

Autrement dit, la valeur critique pour recouvrir tous les sommets de $\mathbb{G}(n, p)$ par des copies de G est dans ce cas égale, d'après le théorème 3.2, à la valeur critique pour l'existence d'une seule copie.

Démonstration. Nous nous contenterons de démontrer la partie (i) de ce théorème. Les deux autres parties nécessitent des outils qui ne seront pas présentés dans ce projet. Pour la partie (i), nous allons supposer que

$$n^{v_G-1}p^{e_G} = \Theta(\log(n)). \quad (4.1)$$

On remarque ensuite que pour tout sommet $v \notin M(G)$, on a $\mathbb{P}(U_i(v) > 0) = o(1)$.³ Ainsi le rôle décisif pour recouvrir les sommets de $\mathbb{G}(n, p)$ est joué par les graphes enracinés (v, G) , où $v \in M(G)$.

On note ensuite, pour chaque sommet $i \in \mathbb{G}(n, p)$, $\bar{U}_i = \sum_{j=1}^l U_i(v_j)$, où les $\{v_j\}_{j=1, \dots, l}$ forment une collection maximale de sommets de $M(G)$ pour lesquels les graphes enracinés (v_j, G) ne sont pas isomorphes deux à deux. On a alors

$$\sum_{i=1}^n \bar{U}_i = |M(G)|X_G.$$

Pour des raisons de symétrie on obtient que pour chaque sommet $i \in \mathbb{G}(n, p)$, l'espérance de la variable \bar{U}_i vaut $\mathbb{E}(\bar{U}_i) = \frac{|M(G)|}{n} \mathbb{E}(X_G) = \frac{a}{n} n^{v_G} p^{e_G} + o(1) = an^{v_G-1}p^{e_G} + o(1)$. En particulier pour le sommet numéro 1 que l'on va traiter par la suite, sans perte de généralité.

On choisit maintenant un sommet $v \notin M(G)$, ainsi qu'un sous-graphe $H_v \subseteq G$ le plus petit possible contenant v et tel que $d(v, H_v) = m(v, G) > m_*$. On note ensuite $U_i^*(v)$ le nombre de copies de (v, H_v) dans $\mathbb{G}(n, p)$ enracinées en i , et $U_i^* = \sum_{v \notin M(G)} U_i^*(v)$. Alors $\mathbb{E}(U_1^*) = o(1)$.

³Ce résultat se base sur celui de Ruciński et Vince cité au début de cette section.

Comme $\mathbb{E}(W_G) = n\mathbb{P}(U_1 = 0)$, nous avons besoin d'étudier le comportement asymptotique de $\mathbb{P}(U_1 = 0)$. Le fait d'avoir $\bar{U}_1 + U_1^* = 0$ implique $U_1 = 0$, et ensuite l'équation (1.4) permet de dire

$$\mathbb{P}(U_1 = 0) \geq \mathbb{P}(\bar{U}_1 + U_1^* = 0) \geq \exp\left\{-\frac{\mathbb{E}(\bar{U}_1 + U_1^*)}{1-p}\right\} = \exp\{-\mathbb{E}(\bar{U}_1) + o(1)\}.$$

Essayons à présent de démontrer l'autre inégalité. Soit \mathcal{S} la famille de tous les ensembles d'arêtes des copies enracinées au sommet 1 de (v_j, G) , pour $j = 1, \dots, l$, dans le graphe complet K_n .⁴ Pour chaque $A \in \mathcal{S}$, notons $I_A = \mathbb{1}_{[A \subseteq \mathbb{G}(n,p)]}$. On a alors

$$\begin{aligned} \sum_{A \neq B} \sum_{A \cap B \neq \emptyset} \mathbb{E}(I_A I_B) &= O\left(\sum_{s,t} n^{2v_G - s - 1} p^{2e_G - t}\right) \\ &= O\left(n^{2v_G - 2} p^{2e_G} \sum_{s,t} n^{-(s-1)} p^{-t}\right), \end{aligned}$$

où s et t représentent respectivement le nombre de sommets communs et d'arêtes communes d'une paire de copies de G enracinées au sommet 1. Cet enracinement explique d'ailleurs le terme -1 qui se trouve à l'exposant de n . Une telle intersection est un sous-graphe de (v, G) , et comme celui-ci est strictement équilibré, on aura toujours

$$\frac{t}{s-1} < \frac{e_G}{v_G - 1}.$$

Avec la supposition (4.1), ceci implique que $n^{s-1} p^t \geq n^\varepsilon$ pour $\varepsilon > 0$. Et avec l'inégalité (1.5), on obtient

$$\mathbb{P}(U_1 = 0) \leq \mathbb{P}(\bar{U}_1 = 0) \leq \exp\{-\mathbb{E}(\bar{U}_1) + o(1)\}.$$

Par conséquent,

$$\mathbb{E}(W_G) = n\mathbb{P}(U_1 = 0) = n \cdot \exp\{-\mathbb{E}(\bar{U}_1) + o(1)\} = n \cdot \exp\{-an^{v_G-1} p^{e_G} + o(1)\}.$$

On voit ainsi que lorsque $an^{v_G-1} p^{e_G} - \log(n) \rightarrow \infty$, alors $\log(\mathbb{E}(W_G)) \rightarrow -\infty$, et donc $\mathbb{E}(W_G) = o(1)$. Par la méthode du premier moment, ceci implique que $\mathbb{P}(W_G > 0) = o(1)$. D'autre part, si $an^{v_G-1} p^{e_G} - \log(n) \rightarrow -\infty$, alors $\mathbb{E}(W_G) \rightarrow \infty$ et nous appliquons ensuite la méthode du deuxième moment.

La variance de W_G peut être écrite sous la forme $\text{Var}(W_G) = \mathbb{E}(W_G(W_G - 1)) + \mathbb{E}(W_G) - (\mathbb{E}(W_G))^2$. En considérant la famille de tous les ensembles d'arêtes appartenant à des copies de (v, G) enracinées au sommet 1 ou au sommet 2, on peut appliquer l'inégalité (1.5) et obtenir

$$\begin{aligned} \mathbb{E}(W_G(W_G - 1)) &= n(n-1)\mathbb{P}(U_1 = U_2 = 0) \leq n(n-1)\mathbb{P}(\bar{U}_1 = \bar{U}_2 = 0) \\ &\leq n(n-1) \cdot \exp\{-2\mathbb{E}(\bar{U}_1) + o(1)\}. \end{aligned}$$

Ainsi lorsque $\mathbb{E}(W_G) \rightarrow \infty$, on a

$$\mathbb{P}(W_G = 0) \leq \frac{\text{Var}(W_G)}{(\mathbb{E}(W_G))^2} = \frac{\mathbb{E}(W_G(W_G - 1))}{(\mathbb{E}(W_G))^2} + \frac{1}{\mathbb{E}(W_G)} - 1 \rightarrow 0.$$

⁴On suppose que le sommet 1 est dans $M(G)$. Dans le cas contraire on peut toujours trouver une autre numérotation des sommets afin que le sommet 1 soit bel et bien dans $M(G)$.

On peut remarquer maintenant que pour le cas 0 la supposition (4.1) est toujours vérifiée. Considérons maintenant un graphe quelconque vérifiant l'hypothèse du cas 1. On peut alors remplacer la probabilité $p(n)$ définissant $\mathbb{G}(n, p)$ par $p'(n) = \min(p(n), p^*(n))$ pour tout n , où $p^*(n)$ satisfait la condition (4.1). Alors par monotonie, le fait que la propriété COV_G soit vérifiée pour $\mathbb{G}(n, p'(n))$ implique qu'elle l'est également pour $\mathbb{G}(n, p(n))$. Par conséquent il sera toujours possible de se ramener au cas traité jusqu'ici.

Pour terminer la preuve de la partie (i) de ce théorème, il reste maintenant à étudier le cas où $an^{v_G-1}p^{e_G} - \log(n) \rightarrow c \in \mathbb{R}$, et donc $\mathbb{E}(W_G) \rightarrow e^{-c}$. Il faudrait alors démontrer que le k -ème moment factoriel de W_G , qui vaut $\mathbb{E}[W_G(W_G - 1) \cdots (W_G - k + 1)]$, converge vers e^{-ck} pour tout $k \geq 1$. Alors d'après la méthode des moments, ceci permettrait de conclure que W_G converge en distribution vers une loi de Poisson de paramètre e^{-c} . Il serait aussi possible d'utiliser la méthode de Stein dans le cas poissonien. □

Appliquons alors ce théorème aux graphes K_3 et K_3^+ . Les deux peuvent être enracinés de manière à ce qu'ils soient strictement équilibrés. Donc d'après le cas (i) de ce théorème, les valeurs critiques pour les propriétés COV_{K_3} et $\text{COV}_{K_3^+}$ sont respectivement $(\log(n))^{1/3} n^{-2/3}$ et $(\log(n))^{1/4} n^{-3/4}$. On remarque en effet que par exemple dans le cas du triangle K_3 , si $p \ll (\log(n))^{1/3} n^{-2/3}$, alors $\frac{n^2 p^3}{\log(n)} \rightarrow 0$, et par conséquent $n^2 p^3 \rightarrow c \in \mathbb{R}$. Donc finalement $n^2 p^3 - \log(n) \rightarrow -\infty$, et *a.p.s.* il existe des sommets qui ne sont pas contenus dans un triangle.

Posons $(\log(n))^{1/4} n^{-3/4} \ll p(n) \ll n^{-2/3}$. Alors *a.p.s.* chaque sommet appartient à une copie de K_3^+ . Mais comme $p \ll n^{-2/3}$, ceci implique que $p^3 n^2 \rightarrow 0$, et le nombre moyen de triangles vaudra $\mathbb{E}(X_{K_3}) = \Theta(n^3 p^3) = o(n)$. Donc la majorité des sommets joueront, dans la copie de K_3^+ à laquelle ils appartiennent, le rôle du sommet en dehors du triangle.

Chapitre 5

Couplages et G -facteurs

On appelle *couplage* dans un graphe G un ensemble d'arêtes disjointes. Un couplage est dit *parfait* lorsque les arêtes considérées permettent de recouvrir chaque sommet de G . Ainsi une condition nécessaire pour qu'un graphe possède un couplage parfait est d'avoir un nombre pair de sommets. On dira alors de manière plus générale qu'un graphe satisfait la propriété PM s'il possède un couplage recouvrant ses sommets de manière à ce qu'il y ait finalement au plus un sommet isolé. On étudiera alors dans ce chapitre la probabilité que le graphe aléatoire $\mathbb{G}(n, p)$ satisfasse cette propriété.

5.1 Le théorème de Hall

Nous avons tout d'abord besoin de quelques notions, ainsi que d'un résultat fondamental de la théorie des graphes appelé *théorème de Hall*. On dit qu'un couplage est *maximal* dans un graphe quelconque s'il est impossible de rajouter une arête à ce couplage. D'un autre côté on dira qu'un couplage est *maximum* s'il contient le plus grand nombre possible d'arêtes dans le graphe considéré. Il faut faire très attention lorsqu'on utilise ces deux définitions. En effet, un couplage maximum est toujours maximal, mais la réciproque n'est pas vraie ! Considérons par exemple le chemin P_4 représenté à la figure 5.1. L'arête rouge définit un couplage maximal, car il n'existe aucune arête dans ce graphe qui soit disjointe à cette arête rouge. Pourtant ce couplage n'est pas maximum, car il ne possède qu'une seule arête et il existe un autre couplage en possédant deux. Il s'agit du couplage formé par les arêtes bleues.



FIG. 5.1 – Couplage dans le chemin P_4 .

Si on se donne un couplage M , on dira qu'un chemin est *M -alterné* si chacune de ses arêtes appartenant à M se trouve entre deux autres arêtes qui n'y sont pas, et vice-versa. Et si les deux extrémités d'un chemin M -alterné ne sont adjacentes à aucune arête de M , on dit que c'est un chemin *pouvant augmenter M* . Avec ces quelques notions nous pouvons maintenant énoncer et démontrer le théorème de Hall.

Théorème 5.1. Soit $G = (V, E)$ un graphe biparti où $V = X \cup Y$ et chaque arête de G relie un sommet de X à un sommet de Y . Alors G possède un couplage recouvrant chaque sommet de X si et seulement si $|N(S)| \geq |S|$ pour tout $S \subseteq X$.

Démonstration. La nécessité est évidente. Pour montrer la suffisance il faut raisonner avec la contraposée, en montrant que si M est un couplage maximum ne recouvrant pas tous les sommets de X , alors il existe $S \subseteq X$ tel que $|N(S)| < |S|$.

Soit alors $u \in X$ un sommet qui n'est adjacent à aucune arête de M . Soit W l'ensemble de tous les sommets de V atteignables depuis u par des chemins M -alternés, et posons $S = W \cap X$, $T = W \cap Y$, comme dans la figure 5.2. On a aussi $u \in S$.

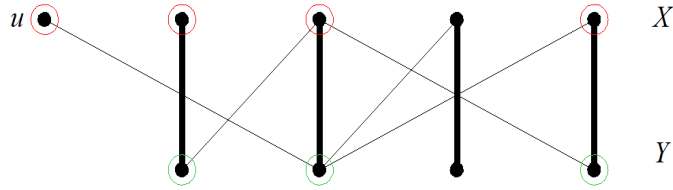


FIG. 5.2 – Graphe permettant de comprendre la démonstration du théorème de Hall. Les arêtes en gras forment le couplage maximum M , et les sommets de S sont entourés en rouge, tandis que ceux de T sont entourés en vert.

Les chemins M -alternés partant de u atteignent les sommets de Y par des arêtes qui ne sont pas dans M , puis reviennent dans X via des arêtes de M . Par conséquent chaque sommet de $S \setminus \{u\}$ est atteint par une arête de M depuis un sommet de T . De plus on sait que M est maximum, et donc G ne possède aucun chemin pouvant augmenter M . En effet, s'il en existait un on pourrait considérer le "couplage complémentaire" de ce chemin, et il aurait une arête de plus que le premier couplage considéré. C'est ce qui arrive par exemple si l'on considère le couplage rouge de la figure 5.1, et qu'on le remplace par le bleu.

Cette inexistence de chemin pouvant augmenter M montre alors que tous les sommets de T sont adjacents à une arête de M . Donc chaque chemin M -alterné atteignant $y \in T$ peut être prolongé jusqu'à S tout en restant M -alterné. Les arêtes de M entre T et $S \setminus \{u\}$ définissent alors une bijection entre ces deux ensembles, et par conséquent $|T| = |S \setminus \{u\}|$.

De plus ce couplage entre T et $S \setminus \{u\}$ nous donne $T \subseteq N(S)$. On peut alors montrer qu'en fait on a $T = N(S)$. Supposons en effet que le sommet $y \in Y \setminus T$ ait un voisin $v \in S$. L'arête (v, y) ne peut pas être dans M , car sinon il y aurait deux arêtes de M incidentes au sommet v , contredisant ainsi la définition de couplage. Et si (v, y) n'est pas dans M , alors cette arête peut être ajoutée à un chemin M -alterné atteignant v . Le chemin considéré serait alors toujours M -alterné, et ainsi on devrait avoir $y \in T$. Donc on arrive là aussi à une contradiction. Autrement dit l'arête (v, y) n'existe pas, et on a bel et bien $T = N(S)$.

Au final on obtient $|N(S)| = |T| = |S \setminus \{u\}| = |S| - 1 < |S|$, ce qui termine la preuve de ce théorème de Hall. □

5.2 Etude des couplages parfaits

Le théorème de Hall va être utile pour étudier la probabilité que le graphe aléatoire $\mathbb{G}(2n, p)$ satisfasse la propriété PM définie au tout début de ce chapitre. On considère alors une bipartition des $2n$ sommets de ce graphe en deux ensembles notés V_1 et V_2 , de taille n . Si l'on suppose que $\mathbb{G}(2n, p)$ ne possède aucun couplage parfait, cela signifie qu'il existe un ensemble $S \subset V_i$, pour $i = 1, 2$, qui viole la condition de Hall. Autrement dit on aurait $|S| > |N(S)|$. Supposons alors que S soit le plus petit ensemble ne vérifiant pas la condition de Hall. S doit par conséquent satisfaire les propriétés suivantes :

- (i) $|S| = |N(S)| + 1$,
- (ii) $|S| \leq \lceil n/2 \rceil$,
- (iii) chaque sommet de $N(S)$ doit être adjacent à au moins deux sommets de S .

Posons ensuite $s = |S|$. Si $s = 1$, alors S est un sommet isolé, car la condition (i) ci-dessus nous conduit à avoir $N(S) = \emptyset$. Si $s = 2$, alors S est constitué de deux sommets, appartenant à V_i , de degré 1 et qui sont adjacents à un même sommet de V_j , où $i \neq j$. Un tel graphe est appelé une *cerise*. Soit alors X la variable aléatoire qui compte le nombre de cerises dans $\mathbb{G}(2n, p)$. Sachant que pour tout x on a $(1 - x) \leq e^{-x}$, et en supposant que $\log(n) - \log(\log(n)) \leq np \leq 2 \log(n)$, on a alors

$$\begin{aligned} \mathbb{E}(X) &= O(n^3 p^2 (1-p)^{2n-2}) = O(n^3 p^2 e^{-2np}) \\ &= O\left(n^3 \left(\frac{\log(n)}{n}\right)^2 \left(\frac{\log(n)}{n}\right)^2\right) \\ &= o(1). \end{aligned}$$

Ainsi $\mathbb{G}(2n, p)$ ne possède *a.p.s.* aucune cerise si l'hypothèse ci-dessus est satisfaite. Ce résultat peut également être trouvé lorsque s est plus grand que 2. Notons par \mathcal{A} l'événement qui décrit la présence d'un sous-ensemble minimal $S \subset V_1$ violant la condition de Hall et possédant au moins 3 sommets. En utilisant les trois conditions ainsi que l'hypothèse ci-dessus et la formule de Stirling qui nous dit que $n! \sim \sqrt{2\pi} \cdot e^{-n} n^{n+1/2}$, on obtient

$$\begin{aligned} \mathbb{P}(\mathcal{A}) &\leq \sum_{s=3}^{\lceil \frac{n}{2} \rceil} \binom{n}{s} \binom{n}{s-1} \binom{s}{2}^{s-1} p^{2s-2} (1-p)^{s(n-s+1)} \\ &\leq \sum_{s \geq 3} \frac{n^s}{s!} \frac{n^{s-1}}{(s-1)!} s^{2(s-1)} p^{2(s-1)} \exp(-pms + s^2 p - sp) \\ &\leq \sum_{s \geq 3} \left(\frac{en}{s}\right)^s \left(\frac{en}{s-1}\right)^{s-1} s^{2(s-1)} \left(\frac{2 \log(n)}{n}\right)^{2(s-1)} \exp\left(-\frac{\log(n)}{2}s + \frac{\log(\log(n))}{2}s\right) \\ &\leq \sum_{s \geq 3} en \left(\frac{ens \cdot 2 \log(n)}{ns}\right)^{2(s-1)} \left(\frac{1}{\sqrt{n}}\right)^s (\log(n))^{s/2} \\ &< \frac{n}{(2 \log(n))^2} \sum_{s \geq 3} \left(\frac{4e^2 \log^{5/2}(n)}{\sqrt{n}}\right)^s \\ &= \frac{16 e^6 \log^{11/2}(n)}{\sqrt{n} - 4e^2 \log^{5/2}(n)} = O\left(\frac{\log^{11/2}(n)}{\sqrt{n}}\right) = o(1). \end{aligned}$$

La dernière ligne de calcul du raisonnement ci-dessus a été obtenue grâce au programme *Mathematica*. On obtient donc qu'avec l'hypothèse ci-dessus, les seuls sous-graphes de $\mathbb{G}(2n, p)$ ne remplissant pas la condition de Hall sont *a.p.s.* ses sommets isolés. De manière générale il y a donc équivalence entre l'existence d'un couplage parfait et l'absence de sommets isolés dans $\mathbb{G}(n, p)$. Ceci veut dire que si l'on veut connaître la probabilité que $\mathbb{G}(n, p)$ contienne un couplage parfait, on doit tout d'abord connaître la probabilité qu'il ne contienne aucun sommet isolé. Tel est l'objet du théorème ainsi que du corollaire qui suivent.

Théorème 5.2. *Soit T la variable aléatoire définissant le nombre de sommets isolés dans $\mathbb{G}(n, p)$. Alors on a*

$$\mathbb{P}(T > 0) \rightarrow \begin{cases} 0 & \text{si } np - \log(n) \rightarrow +\infty, \\ 1 & \text{si } np - \log(n) \rightarrow -\infty. \end{cases}$$

De plus, si $np - \log(n)$ tend vers $c \in \mathbb{R}$, alors T converge en distribution vers une loi de Poisson de paramètre e^{-c} .

Démonstration. On a $T \sim \mathcal{B}(n, (1-p)^{n-1})$, et donc son espérance vaut $\mathbb{E}(T) = n(1-p)^{n-1}$. En utilisant le fait que $\log(1-p) \approx -p$ pour p suffisamment petit, on obtient

$$\begin{aligned} \log(\mathbb{E}(T)) &= \log(n) + (n-1)\log(1-p) \\ &\approx \log(n) - (n-1)p \\ &= p - np + \log(n). \end{aligned}$$

Par conséquent, on remarque que si $np - \log(n) \rightarrow +\infty$, alors $\log(\mathbb{E}(T)) \rightarrow -\infty$, et ainsi l'espérance de T tend vers 0. Par la méthode des premiers moments et l'inégalité (2.1), ceci implique que $T = 0$ *a.p.s.* Et dans le cas où $np - \log(n) \rightarrow -\infty$, on va utiliser l'inégalité (2.2). On a

$$\begin{aligned} \log\left(\frac{\text{Var}(T)}{(\mathbb{E}(T))^2}\right) &= \log\left(\frac{n(1-p)^{n-1}(1-(1-p)^{n-1})}{n^2[(1-p)^{n-1}]^2}\right) \\ &= \log\left(\frac{1-(1-p)^{n-1}}{n(1-p)^{n-1}}\right) \\ &\approx -(1-p)^{n-1} - \log(n) + (n-1)p \\ &\approx -\underbrace{(1-p)^{n-1}}_{\geq 0} - \underbrace{\log(n) + np - p}_{\rightarrow -\infty} \rightarrow -\infty. \end{aligned}$$

Par conséquent, on obtient que le terme de droite de l'inégalité (2.2) tend vers 0. Ainsi $\mathbb{P}(T = 0)$ tend vers 0 lorsque $np - \log(n)$ tend vers $-\infty$.

Le cas intermédiaire peut être démontré grâce à la méthode de Stein dans le cas poissonien. On peut écrire $T = \sum_{i=1}^n I_i$, où I_i est une variable indicatrice qui vaut 1 si le sommet i est isolé et 0 sinon. Pour appliquer la méthode de Stein, et plus précisément le théorème 2.5, il faut voir que ces variables sont positivement corrélées. Notons \mathbb{G}_i le sous-graphe de $\mathbb{G}(n, p)$ obtenu en supprimant toutes les arêtes partant du sommet i , et considérons ensuite la variable indicatrice J_{ji} qui vaut 1 si le sommet j est isolé dans \mathbb{G}_i et 0 sinon. On a alors clairement $J_{ji} \geq I_j$, car le fait d'avoir supprimé toutes les arêtes adjacentes au sommet i augmente la probabilité que j soit isolé. Et si i est lui-même isolé, alors on a $J_{ji} = I_j$, et donc en général on a $\mathcal{L}(J_{ji}) = \mathcal{L}(I_j | I_i = 1)$. Ainsi les $\{I_i\}$ sont effectivement positivement corrélées.

Comme on suppose que $np - \log(n)$ tend vers $c \in \mathbb{R}$, les calculs précédents permettent de dire que $\mathbb{E}(T)$ tend vers e^{-c} . On a de plus

$$\begin{aligned} \log\left(\frac{\text{Var}(T)}{\mathbb{E}(T)}\right) &= \log\left(\frac{n(1-p)^{n-1}(1-(1-p)^{n-1})}{n(1-p)^{n-1}}\right) \\ &= \log(1-(1-p)^{n-1}) \\ &\approx -(1-p)^{n-1} \\ &\approx -e^{-np}. \end{aligned}$$

Or on doit avoir $\lim_{n \rightarrow \infty} np = \infty$, car sinon il serait impossible d'avoir $np - \log(n)$ qui tend vers une constante. Ceci implique que le logarithme de $\text{Var}(T)/\mathbb{E}(T)$ tend vers 0, et par conséquent $\text{Var}(T)/\mathbb{E}(T)$ tend vers 1. Les hypothèses du théorème 2.5 sont ainsi toutes satisfaites, et ceci permet de conclure que T converge en distribution vers une loi de Poisson de paramètre e^{-c} . \square

Corollaire 5.3. *Pour un graphe aléatoire binomial quelconque, on a*

$$\mathbb{P}(\mathbb{G}(n, p) \in \text{PM}) \rightarrow \begin{cases} 0 & \text{si } np - \log(n) \rightarrow -\infty, \\ e^{-e^{-c}} & \text{si } np - \log(n) \rightarrow c \in \mathbb{R}, \\ 1 & \text{si } np - \log(n) \rightarrow +\infty. \end{cases}$$

Démonstration. Ce corollaire est une conséquence directe de l'équivalence entre l'absence de sommets isolés et la présence d'un couplage parfait dans $\mathbb{G}(n, p)$. \square

5.3 Quelques mots sur les G -facteurs

On se donne un graphe G quelconque. On appelle alors *G -facteur partiel* toute union de copies disjointes de G . Si $G = K_2$, autrement dit G est une simple arête, cette notion est équivalente à celle de couplage. On appellera ensuite *G -facteur* une union disjointe de G recouvrant chaque sommet d'un graphe aléatoire. Dans le cas de K_2 , on retrouve la notion de couplage parfait.

Trouver la valeur critique permettant de dire si $\mathbb{G}(n, p)$ possède *a.p.s.* un G -facteur ou non est un problème très difficile. D'ailleurs il existe des cas de graphes très simples pour lesquels cette valeur critique est encore inconnue à l'heure actuelle. C'est le cas par exemple pour $G = K_3$, autrement dit pour un triangle. On ne sait toujours pas avec certitude quelle est la valeur critique permettant de dire si le graphe aléatoire $\mathbb{G}(n, p)$ peut être recouvert par une union disjointe de triangles.

Il existe tout de même quelques résultats intéressants concernant ces G -facteurs. Supposons par exemple que G soit un arbre ayant plus de 2 sommets. Alors Łuczak et Ruciński démontrèrent en 1991 que dans ce cas il y a correspondance, tout comme dans le cas des couplages parfaits, entre l'absence de sommets isolés et la présence d'un G -facteur dans $\mathbb{G}(n, p)$. Ainsi le théorème 5.2 permet d'énoncer le corollaire suivant de manière analogue au corollaire 5.3. Il est ensuite possible d'étendre ce résultat à des forêts.

Corollaire 5.4. *Soit B un arbre possédant $b \geq 2$ sommets. En supposant que n est divisible par b on a*

$$\mathbb{P}(\mathbb{G}(n, p) \text{ possède un } B\text{-facteur}) \rightarrow \begin{cases} 0 & \text{si } np - \log(n) \rightarrow -\infty, \\ e^{-e^{-c}} & \text{si } np - \log(n) \rightarrow c \in \mathbb{R}, \\ 1 & \text{si } np - \log(n) \rightarrow +\infty. \end{cases}$$

La valeur critique pour l'existence de G -facteurs est également connue pour des graphes remplissant certaines conditions. Au chapitre 4 nous avons défini $d(v, G) = e_G/(v_G - 1)$, et $m(v, G) = \max\{d(v, H) \mid v \in H \subseteq G\}$. Et dans le cas où le degré minimal de G est inférieur à $m(v, G)$, il existe des constantes c et C telles que¹

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathbb{G}(vn, p) \text{ possède un } G\text{-facteur}) = \begin{cases} 0 & \text{si } p \leq cn^{-1/m(v, G)}, \\ 1 & \text{si } p \geq Cn^{-1/m(v, G)}, \end{cases}$$

où v représente le nombre de sommets de G . Ce résultat ne peut pas s'appliquer à un triangle, car si $G = K_3$, alors le degré minimal de G vaut $\delta(G) = 2$, et d'autre part on a $m(v, G) = 3/2$. Les hypothèses ne sont donc pas satisfaites. D'ailleurs elles ne le sont pour aucun graphe complet K_n avec $n \geq 3$. On voit ainsi un côté assez paradoxal de la théorie des graphes aléatoires. En effet il existe des résultats qui sont démontrables pour des graphes aussi compliqués que possible en satisfaisant l'hypothèse ci-dessus, alors qu'ils ne le sont pas pour des graphes très simples, à l'image du triangle. Ceci permet de mettre en valeur la complexité de cette théorie, qui par conséquent a encore de beaux jours devant elle.

¹Voir le livre [3] pour plus de détails.

Bibliographie

- [1] Richard Durrett. *Probability : Theory and Examples*. Duxbury Press, 2005.
- [2] Sheldon M. Ross. *Initiation aux probabilités*. Presses polytechniques et universitaires romandes, 2002.
- [3] Svante Janson, Tomasz Łuczak, Andrzej Ruciński. *Random Graphs*. Wiley Interscience, 2000.
- [4] Douglas B. West. *Introduction to Graph Theory - Second Edition*. Prentice Hall, 2001.