

Noyau de Karp-Sipser et convergence locale

Nicolas Nelson

April 2024

Table des matières

1	Introduction mathématique	1
2	Déroulement du stage	2
3	L'algorithme de Karp-Sipser	3
3.1	Définitions	3
3.2	Le noyau	3
3.3	Multigraphes aléatoires à distribution des degrés fixée	4
4	Un détour en théorie des jeux	4
4.1	Jeu sur un arbre aléatoire	5
4.2	Preuve du théorème 1	6
4.3	Exemples de transitions de phase	8
4.3.1	Loi de Poisson	8
4.3.2	Transition de phase discontinue	8
5	Convergence locale et taille du noyau	9
5.1	Convergence locale de multigraphes aléatoires	9
5.1.1	Espace métrique des graphes enracinés	9
5.1.2	Convergence locale de graphes déterministes	10
5.1.3	Graphes aléatoires	11
5.2	Noyau de Karp-Sipser d'une suite localement convergente	11
6	Autres questions abordées pendant le stage	15

1 Introduction mathématique

L'algorithme de Karp-Sipser prend un graphe localement fini G en entrée. Il consiste à repérer les feuilles (sommets de degré 1) et les retirer, ainsi leurs uniques voisins, en itérant le procédé jusqu'à ce qu'il ne reste plus de feuilles. Le sous-graphe restant est appelé noyau de Karp-Sipser de G . Les travaux de Karp et Sipser dans [KS81] ont établi l'existence d'une transition de phase pour ce noyau lorsque le graphe suit le modèle d'Erdős-Renyi

$G(n, \lambda/n)$. Pour $\lambda < e$, le noyau est de taille $O(1)$ lorsque $n \rightarrow +\infty$, et pour $\lambda > e$, il est de taille $\Theta(n)$ (voir [AFP98]).

Ces résultats possèdent une interprétation locale surprenante. Le voisinage d'un sommet choisi uniformément au hasard dans $G(n, \lambda/n)$ converge localement vers un arbre de Galton-Watson, dont le nombre d'enfants par sommet suit une loi de Poisson de paramètre λ . Appliquer l'algorithme de Karp-Sipser à un tel arbre revient, à peu de chose près, à déterminer le vainqueur d'un jeu combinatoire sur cet arbre. En particulier, l'existence d'un noyau non vide est directement lié à la probabilité d'obtenir un match nul.

Or, les résultats récents de Martin et Holroyd dans [HM21] résolvent entièrement la question de la probabilité qu'un tel jeu soit un match nul, pour un arbre de Galton-Watson quelconque. Nous avons donc cherché à généraliser la transition de phase obtenue pour le graphe d'Erdős-Rényi à des graphes aléatoires à distribution de degrés quelconques, qui convergent localement vers des arbres. Cependant, l'algorithme de Karp-Sipser est essentiellement non local dans son comportement. Pour l'instant, nous n'avons donc montré que la borne inférieure : dans le cas (explicitement connu) où la probabilité de faire match nul dans l'arbre limite est 0, le noyau de Karp-Sipser est de taille $o(n)$ lorsque $n \rightarrow +\infty$. Nous conjecturons également que dans le cas supercritique où cette probabilité est non nulle, le noyau est de taille $\Theta(n)$.

2 Déroulement du stage

Mon stage s'est déroulé du 1 février au 15 juin 2024, au Department of Statistics de l'université d'Oxford, sous la supervision de James Martin. Il m'a partagé certains projets sur lesquels il travaillait, notamment en lien avec l'algorithme de Karp-Sipser appliqué à des graphes aléatoires. Nous nous sommes vus régulièrement pour en discuter, ainsi que pour des réunions hebdomadaires de son équipe de recherche (comprenant Christina Goldschmidt et plusieurs doctorants et post-doctorants). J'ai passé la majorité de mon temps à lire des articles ou livres mathématiques pertinents, dont certains se trouvent dans la bibliographie. J'ai aussi cherché, souvent sur des idées de James, à démontrer des résultats nouveaux. Je n'ai pas reconstruit beaucoup de succès dans cette recherche proprement dite, mais il m'a proposé de continuer à échanger sur le sujet, ce que je ferai avec plaisir.

J'ai aussi eu l'occasion d'assister régulièrement à des séminaires de probabilité, dont un hebdomadaire qui est interne aux doctorants et post-docs du département. Par ailleurs, James m'a permis de travailler dans un bureau au même étage que le sien, que j'ai partagé avec plusieurs doctorants.

Je remercie vivement James et toute l'équipe du troisième étage pour leur accueil, Cristina Toninelli pour m'avoir mis en contact avec lui, et le DMA pour m'avoir donné l'occasion et les moyens de réaliser ce stage. Ce fut un formidable apprentissage mathématique et humain, qui m'a donné un aperçu clair et concret du monde de la recherche.

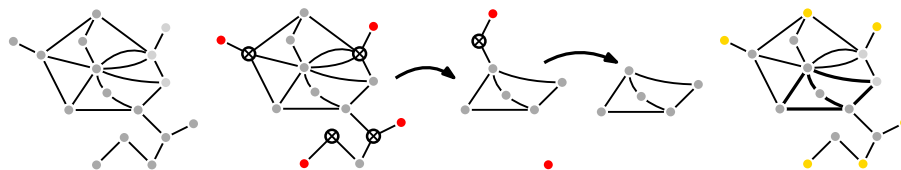


FIGURE 1 – Illustration de l’algorithme de Karp-Sipser. Les quatre premières figures montrent le graphe initial, ainsi que les sommets retirés (en rouge) et leurs voisins (les croix). Lorsqu’il ne reste plus de feuilles, on obtient le noyau de Karp-Sipser à la quatrième figure. La dernière figure représente le graphe initial avec le noyau de Karp-Sipser en gras et l’ensemble indépendant obtenu en jaune.

3 L’algorithme de Karp-Sipser

3.1 Définitions

Soit G un multigraphe fini. L’algorithme de Karp-Sipser [KS81] consiste à retirer de G les sommets de degré 1 (appelés feuilles) ainsi que leurs uniques voisins, jusqu’à ce que G ne contienne plus aucun sommet de degré 1 (voir figure 1). Les sommets isolés sont également supprimés. Le sous-graphe restant est appelé noyau de Karp-Sipser.

La motivation initiale de Karp et Sipser était l’étude des appariements de sommets, c’est-à-dire des ensembles d’arêtes deux à deux non adjacentes. L’ensemble des arêtes reliant une feuille retirée à son plus proche voisin forme un tel appariement, qui est optimal lorsque le noyau est vide. La figure 1 illustre une autre motivation, citée dans [BCC22] : l’ensemble des feuilles retirées et des sommets isolés forme un ensemble indépendant à l’intérieur de G , également optimal si le noyau est vide.

3.2 Le noyau

Une propriété remarquable de ce processus est la commutativité : quel que soit l’ordre dans lequel on retire les feuilles de G , le noyau obtenu à la fin de l’algorithme est toujours le même, ce qui justifie sa bonne définition. Pour une preuve de ce résultat, voir l’annexe de [BG01]. Dans la suite, on notera $N_{KS}(G)$ le noyau de Karp-Sipser de G .

Lorsque le multigraphe G est aléatoire, $N_{KS}(G)$ est également une variable aléatoire, fonction déterministe de G . Je vais présenter dans la suite divers résultats et conjectures liés au comportement de cette variable. On s’intéresse notamment à l’asymptotique de $|N_{KS}(G)|$ lorsque $|G| \rightarrow +\infty$.

3.3 Multigraphes aléatoires à distribution des degrés fixée

Soit n un entier naturel non nul, et $d = (d_1, \dots, d_n)$ un vecteur d'entiers naturels non nuls tel que $m = \sum_{v=1}^n d_v$ soit pair. Soit alors G un multigraphe à n sommets numérotés de 1 à n , où le sommet v est de degré d_v . On voit le sommet v comme ayant d_v demi-arêtes qui lui sont rattachées. Le nombre total de demi-arêtes m est pair : on peut donc les apparier deux à deux uniformément au hasard. La loi du multigraphe obtenu est notée $\mathbf{CM}(d)$, du nom anglais « configuration model ».

Ce modèle de graphe aléatoire, introduit par Bollobás dans [Bol80], reste moins étudié que celui d'Erdős-Rényi, mais est plus flexible puisqu'il permet de choisir a priori les degrés de chaque sommet. Par ailleurs, on peut généraliser le modèle en choisissant les degrés eux-mêmes de manière aléatoire. Des articles tels que [BR19] motivent l'étude de ce modèle par sa capacité à mieux décrire des réseaux physiques ou numériques, puisque les degrés peuvent être choisis inhomogènes à souhait.

Lorsqu'on étudie une suite $(G_n)_{n \in \mathbb{N}}$ de tels multigraphes, on impose souvent une condition de régularité sur les vecteurs $d^{(n)}$ associés. Soit X_n le degré d'un sommet de G_n choisi uniformément au hasard. Cette variable aléatoire a pour fonction de répartition F_n , avec

$$F_n(x) = \frac{1}{n} \sum_{v=1}^n \mathbf{1}_{\{d_v^{(n)} \leq x\}}$$

qui est la distribution empirique des degrés. On va supposer les conditions suivantes sur la suite (F_n) .

Condition 1. — (a) **Convergence faible de la distribution de degrés** : il existe une fonction de répartition F telle que

$$X_n \xrightarrow[n \rightarrow +\infty]{d} X$$

où X_n et X ont pour fonctions de répartition respectives F_n et F . De plus, on suppose que $F(0) = 0$, c'est-à-dire $X \geq 1$ presque sûrement.

— (b) **Convergence du degré moyen** :

$$\mathbb{E}[X_n] \xrightarrow[n \rightarrow +\infty]{} \mathbb{E}[X] < +\infty$$

— (c) **Convergence du second moment** :

$$\mathbb{E}[X_n^2] \xrightarrow[n \rightarrow +\infty]{} \mathbb{E}[X^2] \in (0, +\infty)$$

4 Un détour en théorie des jeux

Cette section sera consacrée à certains résultats de Martin et Holroyd dans [HM21], datant de 2021.

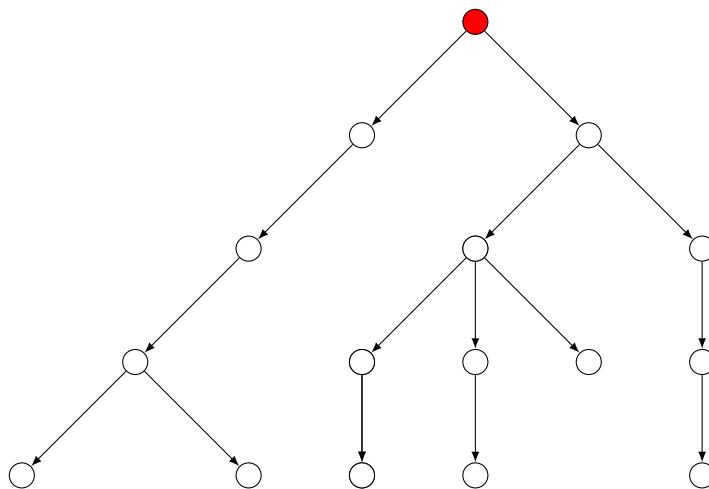


FIGURE 2 – Un arbre de jeu, avec le jeton représenté en rouge sur la racine

4.1 Jeu sur un arbre aléatoire

On s'intéresse à un jeu combinatoire à deux joueurs, dont les positions et les coups possibles sont décrits par un arbre orienté \mathcal{T} . Un jeton se trouve sur un sommet de \mathcal{T} . Chacun leur tour, les deux joueurs déplacent le jeton sur un nouveau sommet en suivant une arête orientée. Dès qu'un joueur ne peut pas se déplacer, c'est-à-dire dès qu'il se trouve sur une feuille, il a perdu, et son adversaire a gagné. Le joueur dont c'est le tour est appelé joueur **suivant**, et l'autre, même s'il n'a pas encore joué, est appelé joueur **précédent**.

On s'intéresse uniquement à la stratégie optimale. Une **stratégie** est une fonction qui à chaque sommet associe un coup légal, c'est-à-dire l'un des enfants du sommet, s'il en existe un. En partant d'un sommet donné (la racine en général), une stratégie est dite **gagnante** si, quelle que soit la stratégie de l'autre joueur, elle mène à une victoire. Si \mathcal{T} est fini, la partie termine toujours, et l'un des deux joueurs dispose nécessairement d'une stratégie gagnante en commençant à la racine : on dit alors que l'arbre est **gagnant** pour ce joueur et **perdant** pour son adversaire. Lorsque \mathcal{T} est infini, il est cependant possible qu'aucun joueur ne dispose d'une stratégie gagnante, l'arbre est alors dit **égal**. Soit \mathcal{T} un arbre de Galton-Watson. Notons $\mathbf{p} = (p_0, p_1, \dots)$ la distribution du nombre d'enfants de chaque sommet. Bien que l'arbre soit aléatoire, on suppose que sa réalisation est connue des deux joueurs. Soient $N = N(\mathbf{p})$ et $P = P(\mathbf{p})$ les probabilités respectives que l'arbre soit gagnant pour le premier joueur ("Next") et pour le second joueur ("Previous"), et $D = 1 - N - P$ la probabilité que l'arbre soit nul.

Il est bien connu que les arbres de Galton-Watson admettent une transition de phase : l'arbre est presque sûrement fini si et seulement si $\sum_i i p_i \leq 1$ et $p_1 < 1$. Cependant, il ne suffit pas que l'arbre soit infini pour obtenir un match nul :

il peut y avoir une branche infinie, mais dont un joueur peut dévier afin de gagner la partie. Martin et Holroyd ont montré l'existence d'une transition de phase pour D lorsque \mathbf{p} varie, mais elle n'a pas lieu au même endroit que celle pour la survie du processus en général. Leur analyse repose sur des fonctions génératrices. Soit

$$G(x) = \sum_{k=0}^{+\infty} p_k x^k$$

la fonction génératrice de la distribution du nombre d'enfants, et soit $H = 1 - G$. On note $H^2 = H \circ H$, et pour f une fonction définie sur $[0, 1]$, on note $\mathbf{FP}(f) = \{x \in [0, 1] \mid f(x) = x\}$. Leur premier théorème est le suivant.

Théorème 1. $D = \max \mathbf{FP}(H^2) - \min \mathbf{FP}(H^2)$; $N = \min \mathbf{FP}(H^2)$; $1 - P = \max \mathbf{FP}(H^2)$.

En particulier, $D > 0$ si et seulement si H^2 a plusieurs points fixes dans $[0, 1]$. Holroyd et Martin donnent plusieurs exemples de distributions avec, à chaque fois, la condition pour avoir $D > 0$, dont certains sont illustrés dans la section 4.3.

4.2 Preuve du théorème 1

Soit \mathcal{N} l'ensemble des sommets v de \mathcal{T} tels que le jeu soit gagnant pour le joueur suivant lorsque le jeton commence en v . Soit \mathcal{P} l'ensemble correspondant pour les victoires du joueur précédent, et \mathcal{D} pour les matchs nuls. Ainsi, \mathcal{N} , \mathcal{P} et \mathcal{D} forment une partition de l'ensemble des sommets. Pour un arbre de Galton-Watson, si o est la racine, on a alors $N = \mathbb{P}(o \in \mathcal{N})$, et de même pour P et D .

Pour $n \geq 0$, on définit \mathcal{N}_n l'ensemble des sommets à partir desquels le joueur suivant dispose d'une stratégie gagnante qui assure une victoire en strictement moins de n coups (en comptant les coups des deux joueurs). Soit \mathcal{P}_n l'ensemble des sommets à partir duquel le joueur précédent peut assurer une victoire en au plus n coups, et $\mathcal{D}_n = V \setminus (\mathcal{N}_n, \mathcal{P}_n)$. En particulier, $\mathcal{N}_0 = \mathcal{P}_0 = \emptyset$. On peut voir \mathcal{D}_n comme l'ensemble des sommets à partir duquel le jeu est un match nul, avec la convention qu'une partie est déclarée nulle après n coups.

Soit v un sommet et Γ l'ensemble de ses enfants. En considérant le premier coup, on a la relation de récurrence suivante :

- $v \in \mathcal{N}_{n+1}$ ssi $\Gamma \cap \mathcal{P}_n \neq \emptyset$
- $v \in \mathcal{P}_{n+1}$ ssi $\Gamma \subseteq \mathcal{N}_n$

et $v \in \mathcal{D}_{n+1}$ sinon. On en déduit directement que $\mathcal{N}_1 = \emptyset$, et par récurrence, que pour tout entier k , $\mathcal{N}_{2k+1} = \mathcal{N}_{2k}$ et $\mathcal{P}_{2k} = \mathcal{P}_{2k-1}$. Pour relier \mathcal{N} , \mathcal{P} et \mathcal{D} à \mathcal{N}_n , \mathcal{P}_n et \mathcal{D}_n , on a besoin de la proposition suivante, analogue au lemme de König montré dans [Kön27].

Lemme 1. Soit \mathcal{T} un arbre d'arité finie. Alors $\mathcal{N} = \bigcup_{n=0}^{+\infty} \mathcal{N}_n$ et de même pour \mathcal{P} .

Démonstration. Soient $\mathcal{N}' = \mathcal{N} \setminus \bigcup_{n=0}^{+\infty} \mathcal{N}_n$ et $\mathcal{P}' = \mathcal{P} \setminus \bigcup_{n=0}^{+\infty} \mathcal{P}_n$. On veut montrer que $\mathcal{N}' = \mathcal{P}' = \emptyset$.

Si $v \in \mathcal{N}'$, alors l'ensemble $\Gamma(v)$ de ses enfants contient au moins un élément de \mathcal{P} , mais aucun élément de $\mathcal{P} \setminus \mathcal{P}'$: sinon, le joueur Suivant pourrait gagner en un nombre fini n de coups. Si $v \in \mathcal{P}'$, alors $\Gamma(v) \subseteq \mathcal{N}$, et nous affirmons qu'au moins un élément de $\Gamma(v)$ est dans \mathcal{N}' . En effet, sinon, chaque élément w de $\Gamma(v)$ est dans un certain $\mathcal{N}_{m(w)}$. Mais alors $M = \max\{m(w), w \in \Gamma(v)\}$ est fini, et $v \in \mathcal{P}_{M+1}$, contredisant l'hypothèse.

Nous affirmons maintenant qu'en partant d'un sommet de \mathcal{N}' , le joueur Précédent dispose d'une stratégie assurant une victoire ou un match nul. Si le joueur Suivant se déplace sur un sommet de $\mathcal{N} \cup \mathcal{P}$, le joueur Précédent peut alors gagner ou faire égalité comme d'habitude. Et si le joueur Suivant se déplace sur un sommet de \mathcal{P}' , le joueur Précédent dispose d'au moins un coup qui emmène le jeton sur un sommet de \mathcal{N}' . La partie ne s'achève donc pas. La même stratégie permet aussi au joueur Suivant de faire égalité à partir d'un sommet de \mathcal{P}' . Il n'y a donc pas de tels sommets. \square

Revenons à la démonstration du théorème 1. Pour $n \in \mathbb{N}$, on pose $N_n = \mathbb{P}(o \in \mathcal{N}_n)$ et de même pour P_n, D_n . On rappelle que $N = \mathbb{P}(o \in \mathcal{N})$ et de même pour P, D .

Corollaire 1. *Pour une distribution \mathbf{p} quelconque du nombre d'enfants, on a $N_n \rightarrow N$ lorsque $n \rightarrow +\infty$, et de même $P_n \rightarrow P, D_n \rightarrow D$.*

Démonstration. Le lemme 1 et la croissance des suites d'ensembles (\mathcal{N}_n) et (\mathcal{P}_n) entraînent que $N_n \nearrow N$ et $P_n \nearrow P$ lorsque $n \rightarrow +\infty$. Puisque $D = 1 - N - P$ et $D_n = 1 - N_n - P_n$ pour $n \geq 0$, on en déduit que $D_n \searrow D$. \square

Appliquons maintenant la formule de récurrence entre \mathcal{N}_{n+1} et \mathcal{P}_n vue plus haut à la racine o .

$$\begin{aligned}
N_{n+1} &= \mathbb{P}(o \in \mathcal{N}_{n+1}) \\
&= \sum_{k=0}^{+\infty} \mathbb{P}(o \in \mathcal{N}_{n+1} \mid |\Gamma(o)| = k) \mathbb{P}(|\Gamma(o)| = k) \\
&= \sum_{k=0}^{+\infty} \mathbb{P}(\Gamma(o) \subseteq \mathcal{P}_n \mid |\Gamma(o)| = k) p_k \\
&= \sum_{k=0}^{+\infty} (1 - P_n)^k p_k \\
&= G(1 - P_n)
\end{aligned} \tag{1}$$

Un calcul analogue donne $1 - N_{n+1} = G(1 - P_n)$. En rappelant la définition $H = 1 - G$, on en déduit $N_{n+2} = H^2(N_n)$ et $1 - P_{n+2} = H^2(1 - P_n)$ pour tout entier n . De plus, $N_0 = P_0 = 0$. Ainsi, puisque H^2 est croissante et continue,

$$N = \lim_{n \rightarrow +\infty} H^{2n}(0) = \min \mathbf{FP}(H^2)1 - P = \lim_{n \rightarrow +\infty} H^{2n}(1) = \max \mathbf{FP}(H^2)$$

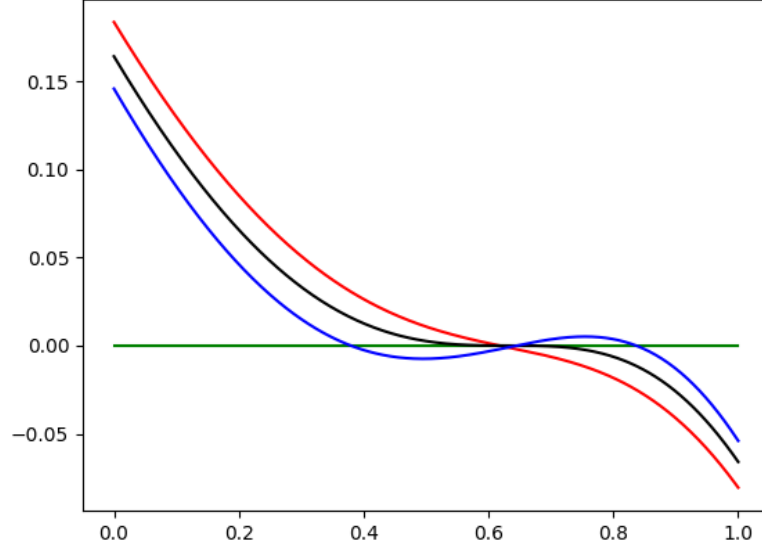


FIGURE 3 – Les fonctions $H^2(x) - x$ associées à une loi de Poisson de paramètre $e - 0, 2, e$ et $e + 0, 2$

et par conséquent,

$$D = 1 - N - P = \max \mathbf{FP}(H^2) - \min \mathbf{FP}(H^2)$$

4.3 Exemples de transitions de phase

4.3.1 Loi de Poisson

Lorsque \mathbf{p} est une loi de Poisson de paramètre λ , $G(x) = e^{-\lambda(1-x)}$, $H(x) = 1 - e^{-\lambda(1-x)}$ et $H^2(x) = 1 - e^{-\lambda e^{-\lambda(1-x)}}$. Des calculs détaillés dans [HM21] montrent que H^2 a un unique point fixe dans $[0, 1]$ lorsque $\lambda \leq e$, et trois points fixes lorsque $\lambda > e$.

De plus, la transition de phase est continue : D tend vers 0 lorsque λ tend vers e .

4.3.2 Transition de phase discontinue

À titre d'illustration, voici un exemple de distribution qui admet une transition de phase discontinue. Soit $G(x) = (1 - t) + t(0, 5x^2 + 0, 5x^{10})$, et \mathbf{p} la loi de probabilité de fonction génératrice G . Il existe alors un réel $t_c \simeq 0,9791$ tel que, pour $t < t_c$, l'équation $F^2(x) = x$ admet une unique solution et $D = 0$. À

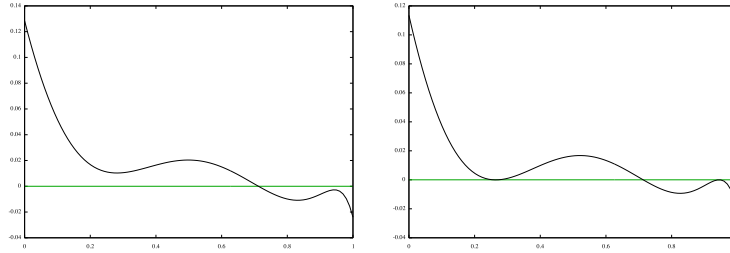


FIGURE 4 – La fonction $H^2(x) - x$ pour la famille de distributions décrite en 4.3.2, avec $t = 0,976$ et $t = 0,9791$

$t = t_c$, deux nouvelles solutions apparaissent, en $x_- \simeq 0,264$ et $x_+ \simeq 0,681$, qui sont des racines doubles. Pour $t > t_c$, l'équation admet cinq racines.

5 Convergence locale et taille du noyau

5.1 Convergence locale de multigraphes aléatoires

Soit $(G_n)_{n \in \mathbb{N}}$ une suite de multigraphes aléatoires. On s'intéresse au comportement du voisinage d'un sommet uniformément choisi du multigraphe. Notons donc, pour n entier, (G_n, o_n) le graphe enraciné aléatoire obtenu lorsque o_n est un sommet choisi uniformément au hasard parmi les sommets de G_n . Plaçons-nous dans un contexte où $|V(G_n)| \rightarrow +\infty$ lorsque $n \rightarrow +\infty$: le plus souvent, on a $|V(G_n)| = n$ pour tout n . La **convergence locale**, introduite par Benjamini et Schramm dans [BS11], permet de formaliser le comportement asymptotique du voisinage de o_n . La présentation suivante de cette théorie est due à van der Hofstad dans [Van24], avec quelques ajustements pour traiter le cas des multigraphes.

5.1.1 Espace métrique des graphes enracinés

Définition 1. Un *multigraphe enraciné* est un couple (G, o) dans lequel $G = (V(G), (x_{u,v})_{u,v \in V(G)})$ est un multigraphe et $o \in V(G)$. Ici, $x_{u,v}$ désigne le nombre d'arêtes entre u et v , et $x_{u,u}$ désigne le nombre de boucles adjacentes à u . Un multigraphe enraciné est *localement fini* lorsque chaque sommet est de degré fini, c'est-à-dire

$$\forall u \in V(G) \quad x_{u,u} + \sum_{v \in V(G)} x_{u,v} < +\infty$$

On se restreint à des multigraphes non orientés, où $x_{u,v} = x_{v,u}$ pour tous $u, v \in V(G)$. Pour l'instant, les définitions sont déterministes et non aléatoires.

Définition 2. Pour un multigraphe enraciné (G, o) , on note $B_r^{(G)}(o)$ le sous-multigraphe enraciné de tous les sommets à distance au plus r de o , où la

distance considérée est la distance de graphe sur G .

Définition 3. Deux multigraphes enracinés (G_1, o_1) et (G_2, o_2) sont *isomorphes* s'il existe une bijection ϕ de $V(G_1)$ dans $V(G_2)$ telle que pour tous $u, v \in V(G_1)$, $x_{\phi(u), \phi(v)} = x_{u, v}$, et $\phi(o_1) = o_2$. On note alors $(G_1, o_1) \simeq (G_2, o_2)$.

On note \mathcal{G}^* l'espace des multigraphes enracinés connexes à isomorphisme près. Par abus de notation, on note souvent $(G, o) \in \mathcal{G}^*$.

Définition 4. Soient (G_1, o_1) et (G_2, o_2) deux multigraphes enracinés **connexes**. On pose

$$R_* = \sup\{r \in \mathbb{N} \mid B_r^{(G_1)}(o_1) \simeq B_r^{(G_2)}(o_2)\}$$

et on définit la distance

$$d_{\mathcal{G}^*}((G_1, o_1), (G_2, o_2)) = \frac{1}{R_* + 1}$$

Lorsque $R_* = +\infty$, les sous-multigraphes $B_r^{(G_1)}(o_1)$ et $B_r^{(G_2)}(o_2)$ sont isomorphes pour tout r , ce qui entraîne un isomorphisme entre (G_1, o_1) et (G_2, o_2) : une preuve détaillée se trouve dans l'annexe de [Van24] au lemme A.11.

La distance $d_{\mathcal{G}^*}$ fait de \mathcal{G}^* un espace métrique polonais : une famille dénombrable dense est donnée par l'ensemble des multigraphes enracinés finis.

5.1.2 Convergence locale de graphes déterministes

Pour appliquer des arguments de convergence locale à un multigraphe non connexe (G, o) , on considère la composante connexe $\mathcal{C}(o)$ de la racine. Pour $h : \mathcal{G}^* \rightarrow \mathbb{R}$, on étend alors la définition aux multigraphes non connexes en posant par convention $h(G, o) := h(\mathcal{C}(o))$. Cela permet de définir ensuite la convergence locale (faible) de multigraphes finis. On s'intéresse essentiellement au cas où la racine o est choisie uniformément au hasard parmi les sommets du multigraphe.

Définition 5. Pour $n \in \mathbb{N}$, soit $G_n = (V(G_n), E(G_n))$ un multigraphe fini, pas nécessairement connexe. Soit (G_n, o_n) le multigraphe enraciné obtenu en choisissant $o_n \in V(G_n)$ uniformément au hasard, et en restreignant G_n à la composante connexe $\mathcal{C}(o_n)$ de o_n dans G_n . On dit alors que la suite $((G_n, o_n))_{n \in \mathbb{N}}$ converge *localement faiblement* vers (G, o) , un élément potentiellement aléatoire de \mathcal{G}^* de loi μ , lorsque pour toute fonction $h : \mathcal{G}^* \rightarrow \mathbb{R}$ continue bornée,

$$\mathbb{E}[h(G_n, o_n)] \xrightarrow{n \rightarrow +\infty} \mathbb{E}_\mu[h(G, o)]$$

où l'espérance à gauche porte sur la racine aléatoire o_n , et à droite sur le multigraphe de loi μ . On note alors $(G_n, o_n) \xrightarrow{d} (G, o)$.

Remarque. La continuité de h est définie avec la topologie engendrée par $d_{\mathcal{G}^*}$ au départ.

5.1.3 Graphes aléatoires

Van der Hofstad définit trois notions différentes de convergence locale pour une suite $(G_n)_{n \in \mathbb{N}}$ de graphes aléatoires finis : convergence locale faible, en probabilité et presque sûre. Nous nous contenterons ici de définir la deuxième notion, celle qui nous intéresse le plus ici.

Définition 6. La suite (G_n) converge *localement en probabilité* vers $(G, o) \in \mathcal{G}^*$ de loi μ lorsque pour toute fonction $h : \mathcal{G}^* \rightarrow \mathbb{R}$ continue bornée,

$$\mathbb{E}[h(G_n, o_n) \mid G_n] \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \mathbb{E}_\mu[h(G, o)]$$

Remarque. Puisque le membre de gauche dans cette convergence est une variable aléatoire, il est possible que le membre de droite le soit aussi, en quel cas μ est une mesure de probabilité aléatoire sur \mathcal{G}^* . Cependant, dans les exemples considérés, μ sera déterministe.

Les deux théorèmes suivants sont donnés à titre d'exemples. Chacun d'entre eux est démontré dans [Van24].

Théorème 2. Pour $n \in \mathbb{N}$, soit G_n un graphe aléatoire d'Erdős-Rényi de paramètres $(n, \lambda/n)$, où λ est une constante donnée. Soit o_n un sommet de G_n choisi uniformément au hasard. La suite $(G_n, o_n)_{n \in \mathbb{N}}$ converge localement en probabilité vers un arbre de Galton-Watson dont le nombre d'enfants suit une loi de Poisson de paramètre λ .

Théorème 3. Pour $n \in \mathbb{N}$, soit G_n un graphe aléatoire de loi $\mathbf{CM}(d^{(n)})$. On suppose les conditions **1.(a)**, **(b)** vérifiées. La suite (G_n) converge alors localement en probabilité vers un arbre de Galton-Watson dont :

- le nombre d'enfants de la racine suit la loi de X
- le nombre d'enfants de chaque autre sommet suit la loi X^* donnée par

$$p_k^* = \frac{(k+1) \mathbb{P}(X = k+1)}{\mathbb{E}[D]}$$

pour tout $k \in \mathbb{N}$.

5.2 Noyau de Karp-Sipser d'une suite localement convergente

Dans cette partie, (G, o) est un multigraphe enraciné localement fini, aléatoire, connexe, de loi μ . Soit $((G_n, o_n))_{n \in \mathbb{N}}$ une suite de multigraphes finis aléatoires, convergeant localement vers (G, o) . A-t-on alors

$$\frac{|N_{KS}(G_n)|}{n} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \mu(o \in N_{KS}(G)) \quad ?$$

Au cours du stage, nous n'avons réussi à apporter une réponse que partielle, à cause du caractère non local (donc discontinu) du noyau de Karp-Sipser. Nous

avons montré que la borne supérieure est toujours vraie, avec une preuve assez courte. La borne inférieure n'est pas toujours vérifiée, mais elle l'est sous une condition de "presque localité" sur le noyau de Karp-Sipser de G_n . Cependant, nous n'avons pas réussi à montrer le résultat espéré, à savoir que $G(n, \lambda/n)$ vérifie cette condition, ainsi que $\mathbf{CM}(d^{(n)})$ sous des hypothèses peu restrictives. Ci-dessous le résultat obtenu sur la borne supérieure.

Théorème 4. *Soient $\zeta = \mu(o \in N_{KS}(G))$ et $\epsilon > 0$ fixé. Alors $\mathbb{P}(|N_{KS}(G_n)| \leq n(\zeta + \epsilon)) \xrightarrow{n \rightarrow +\infty} 1$.*

Démonstration. On va appliquer une forme tronquée de l'algorithme de Karp-Sipser. Soit v un sommet de G_n , et soit k un entier naturel. L'algorithme k -tronqué consiste à appliquer l'algorithme de Karp-Sipser à $B_k^{(G_n)}(v)$, mais sans jamais supprimer de sommet à distance exactement k de v . De manière équivalente, cela revient à partir de $B_k^{(G_n)}(v)$ et à rajouter un chemin infini au voisinage de chaque sommet qui n'est pas dans $B_{k-1}^{(G_n)}(v)$, puis appliquer l'algorithme de Karp-Sipser. On note alors $v \in N_{k-KS}(G, v)$ si v n'est pas supprimé pendant ce processus. Autrement dit, $v \in N_{k-KS}(G, v)$ si et seulement si, pour supprimer v , l'algorithme de Karp-Sipser doit d'abord supprimer un sommet à distance au moins k de v . On pose alors

$$Y_k^{(n)} = \sum_{v \in V(G_n)} \mathbf{1}_{\{v \in N_{k-KS}(G, v)\}} = n \mathbb{E}[\mathbf{1}_{\{o_n \in N_{k-KS}(o_n)\}} \mid G_n]$$

La fonction $(G, o) \mapsto \mathbf{1}_{\{o \in N_{k-KS}(G, o)\}}$ est continue pour la topologie de \mathcal{G}^* , puisqu'elle ne dépend que des sommets à distance au plus k de o . Elle est également bornée. La convergence locale entraîne donc que :

$$\frac{Y_k^{(n)}}{n} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \mu(o \in N_{k-KS}(G, o)) \quad (2)$$

Cependant, on a l'égalité d'événements

$$\{o \in N_{KS}(G)\} = \bigcap_{k=1}^{+\infty} \{o \in N_{k-KS}(G, o)\}$$

car l'algorithme de Karp-Sipser supprime o si et seulement si il le supprime en un nombre fini d'étapes, ce qui met en jeu un nombre fini de sommets chacun à une distance finie de o . La même propriété est vraie pour (G_n, o_n) à n fixé. Ceci montre d'une part que pour tous entiers k, n , on a $|N_{KS}(G_n)| \leq Y_k^{(n)}$. D'autre part, l'union à droite étant décroissante, on a

$$\mu(o \in N_{k-KS}(G, o)) \xrightarrow[k \rightarrow +\infty]{} \mu(o \in N_{KS}(G))$$

Posons $\zeta_k = \mu(o \in N_{k-KS}(G, o))$. On a donc $\zeta_k \xrightarrow[k \rightarrow +\infty]{} \zeta$. Soit k tel que $\zeta_k \leq$

$\zeta + \epsilon/2$. Dès lors,

$$\mathbb{P}(|N_{KS}(G_n)| > n(\zeta + \epsilon)) \leq \mathbb{P}(Y_k^{(n)} > n(\zeta + \epsilon)) \quad (3)$$

$$\leq \mathbb{P}(Y_k^{(n)} > n(\zeta_k + \frac{\epsilon}{2})) \quad (4)$$

$$= o(1) \quad (5)$$

□

Ce résultat entraîne deux corollaires, le premier étant connu depuis Karp et Sipser, et le second étant a priori nouveau.

Corollaire 2. *Si G_n suit la loi du graphe d'Erdős-Rényi $G(n, \lambda/n)$ avec $\lambda \leq e$, alors*

$$\frac{|N_{KS}(G_n)|}{n} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0.$$

Corollaire 3. *Si G_n suit $\mathbf{CM}(d^{(n)})$ et vérifie les conditions **1.(a)**, **(b)**, et si la fonction génératrice G^* de D^* est telle que $(1 - G^*) \circ (1 - G^*)$ n'a qu'un seul point fixe dans $[0, 1]$, alors*

$$\frac{|N_{KS}(G_n)|}{n} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0.$$

Démonstration. Ces deux corollaires découlent du résultat précédent, des exemples de convergence locale donnés à la partie 5.1.3 et du théorème 1. En effet, sur un arbre de jeu, le noyau de Karp-Sipser est exactement l'ensemble des sommets à partir duquel le jeu est un match nul.

Pour le voir, on peut d'abord montrer, par récurrence sur k , que pour tout entier k , tout sommet de \mathcal{P}_{2k} est supprimé par l'algorithme de Karp-Sipser en tant que feuille, et tout sommet de \mathcal{N}_{2k} est supprimé en tant que voisin d'une feuille. Pour $k = 0$, c'est trivial puisque $\mathcal{N}_0 = \mathcal{P}_0 = \emptyset$. Pour l'hérédité, on constate qu'un sommet de $\mathcal{P}_{2k+2} = \mathcal{P}_{2k+1}$ a tous ses enfants dans \mathcal{N}_{2k} . Lorsqu'ils sont tous supprimés par l'algorithme de Karp-Sipser en tant que voisins de feuilles, le sommet en question devient une feuille et est supprimé. Un sommet de \mathcal{N}_{2k+2} a au moins un enfant dans \mathcal{P}_{2k+1} qui est supprimé en tant que feuille, supprimant avec lui son parent.

On montre ensuite, par une récurrence analogue sur le nombre d'itérations de l'algorithme, que tout sommet supprimé par l'algorithme de Karp-Sipser est gagnant ou perdant. Par conséquent, sur un arbre où la probabilité de match nul est 0, la probabilité que la racine soit dans le noyau de Karp-Sipser est également 0.

Dans les cas de ces deux corollaires, d'après le théorème 1, la distribution limite μ est celle d'un arbre aléatoire où la probabilité d'un match nul est 0. Par conséquent, $\mu(o \in N_{KS}(G)) = 0$, et le théorème 4 permet de conclure. □

Pour se rapprocher d'une réciproque, on donne la condition suffisante suivante.

Théorème 5. Avec les notations du théorème 4, supposons la condition de presque localité :

$$\lim_{k \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P}(o_n \in N_{k-KS}(G_n, o_n), o_n \notin N_{KS}(G_n)) = 0.$$

Alors

$$\frac{|N_{KS}(G_n)|}{n} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \zeta.$$

Démonstration. L'inégalité de Markov donne, pour $\epsilon > 0$ quelconque,

$$\begin{aligned} \mathbb{P}\left(\sum_{u \notin N_{KS}(G_n)} \mathbf{1}_{\{u \in N_{k-KS}(G_n, u)\}} \geq \epsilon n\right) &\leq \frac{1}{\epsilon n} \mathbb{E}\left[\sum_{u \notin N_{KS}(G_n)} \mathbf{1}_{\{u \in N_{k-KS}(G_n, u)\}}\right] \\ &= \frac{1}{\epsilon} \mathbb{P}(o_n \notin N_{KS}(G_n), o_n \in N_{k-KS}(G_n, o_n)) \end{aligned} \quad (6)$$

$$(7)$$

Ainsi,

$$\lim_{k \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P}\left(\sum_{u \notin N_{KS}(G_n)} \mathbf{1}_{\{u \in N_{k-KS}(G_n, u)\}} \geq \epsilon n\right) = 0$$

De plus, on a

$$Y_k^{(n)} = |N_{KS}(G_n)| + \sum_{u \notin N_{KS}(G_n)} \mathbf{1}_{\{u \in N_{k-KS}(G_n, u)\}}$$

ce qui entraîne

$$\lim_{k \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P}\left(\left|\frac{Y_k^{(n)}}{n} - \frac{|N_{KS}(G_n)|}{n}\right| \geq \epsilon\right) = 0$$

et ce pour tout $\epsilon > 0$. Soit maintenant $\eta > 0$. On a, pour k quelconque,

$$\limsup_{n \rightarrow +\infty} \mathbb{P}\left(\left|\zeta - \frac{|N_{KS}(G_n)|}{n}\right| \geq \eta\right) \leq \limsup_{n \rightarrow +\infty} \left(\mathbb{P}\left(|\zeta - \zeta_k| \geq \frac{\eta}{3}\right)\right) \quad (8)$$

$$+ \mathbb{P}\left(\left|\zeta_k - \frac{Y_k^{(n)}}{n}\right| \geq \frac{\eta}{3}\right) \quad (9)$$

$$+ \mathbb{P}\left(\left|\frac{Y_k^{(n)}}{n} - \frac{|N_{KS}(G_n)|}{n}\right| \geq \frac{\eta}{3}\right) \quad (10)$$

Pour k suffisamment grand, le premier terme est nul ; d'après l'équation (2), le second terme est de limite supérieure arbitrairement petite ; et par le résultat précédent, le troisième terme aussi. Le terme de gauche étant indépendant de k , il est donc nul, ce qui montre

$$\frac{|N_{KS}(G_n)|}{n} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \zeta \quad (11)$$

□

Nous conjecturons que cette condition suffisante est vérifiée par $G(n, \lambda/n)$, et pour $\mathbf{CM}(d^{(n)})$ lorsque la distribution X n'est pas critique (en un certain sens défini dans l'article [HM21]). Karp et Sipser affirment ce résultat dans leur article [KS81] pour le cas de $G(n, \lambda/n)$, mais sans en donner une preuve complète.

6 Autres questions abordées pendant le stage

De concert avec James, j'ai travaillé sur plusieurs autres questions en lien avec l'algorithme de Karp-Sipser. Vers le début du stage, j'ai par exemple cherché à établir un couplage entre $G(n, M)$, le graphe aléatoire uniforme à n sommets et M arêtes, et $G(n, M+1)$, qui rend le noyau de Karp-Sipser croissant presque sûrement. Je n'ai pas trouvé de tel couplage, mais des vérifications numériques pour de petites valeurs de n n'ont soulevé aucun contre-exemple.

Pendant le dernier mois du stage, suite à la publication de l'article [Gla+24], James m'a proposé d'utiliser la convergence locale pour établir plus directement un TLC pour la taille de l'appariement maximal d'un graphe aléatoire, soit dans le cas de $G(n, \lambda/n)$ que traite déjà cet article, soit dans le cas sous-critique pour $\mathbf{CM}(d^{(n)})$. Nous ne sommes pas arrivés au bout de la question, mais il m'a proposé de continuer à correspondre avec lui à ce sujet.

Références

- [AFP98] Jonathan ARONSON, Alan FRIEZE et Boris G PITTEL. « Maximum matchings in sparse random graphs: Karp–Sipser revisited ». In : *Random Structures & Algorithms* 12.2 (1998), p. 111-177.
- [BR19] AD BARBOUR et Adrian RÖLLIN. « Central limit theorems in the configuration model ». In : *The Annals of Applied Probability* 29.2 (2019), p. 1046-1069.
- [BG01] M BAUER et O GOLINELLI. « Core percolation in random graphs: a critical phenomena analysis ». In : *The European Physical Journal B-Condensed Matter and Complex Systems* 24 (2001), p. 339-352.
- [BS11] Itai BENJAMINI et Oded SCHRAMM. « Recurrence of distributional limits of finite planar graphs ». In : *Selected Works of Oded Schramm* (2011), p. 533-545.
- [Bol80] Béla BOLLOBÁS. « A probabilistic proof of an asymptotic formula for the number of labelled regular graphs ». In : *European Journal of Combinatorics* 1.4 (1980), p. 311-316.
- [BCC22] Thomas BUDZINSKI, Alice CONTAT et Nicolas CURIEN. « The critical Karp–Sipser core of random graphs ». In : *arXiv preprint arXiv:2212.02463* (2022).

- [Gla+24] Margalit GLASGOW et al. « A central limit theorem for the matching number of a sparse random graph ». In : *arXiv preprint arXiv:2402.05851* (2024).
- [HM21] Alexander E HOLROYD et James B MARTIN. « Galton–watson games ». In : *Random Structures & Algorithms* 59.4 (2021), p. 495-521.
- [KS81] Richard M KARP et Michael SIPSER. « Maximum matching in sparse random graphs ». In : *Proceedings of the 22nd Annual Symposium on Foundations of Computer Science*. 1981, p. 364-375.
- [Kön27] Dénes KÖNIG. « Über eine Schlussweise aus dem Endlichen ins Unendliche ». In : *Acta Sci. Math. (Szeged)* 3.2-3 (1927), p. 121-130.
- [Van24] Remco VAN DER HOFSTAD. *Random graphs and complex networks*. Cambridge university press, 2024.