Encadrement :
50% Frédérique Bassino (LIPN, Université Sorbonne Paris Nord) - courriel bassino@lipn.fr
50% Lucas Gerin (CMAP, École Polytechnique)

Graphes et structures aléatoires avec motifs interdits

Cette proposition de thèse se situe à l'interface de la *combinatoire, de la théorie des graphes et des probabilités discrètes*.

Certains problèmes algorithmiques ou combinatoires sur des objets discrets (graphes, permutations, mots, réseaux,...) sont parfois très difficiles ou mal posés dans le cas général. La philosophie générale du domaine est que paradoxalement on peut dire des choses assez précises sur ces problèmes si l'on considère un objet *typique*. On va supposer cet objet aléatoire, une difficulté vient du fait qu'il n'est pas tiré uniformément au hasard mais est souvent soumis à certaines contraintes (sous-motif interdit ou imposé, structure géométrique ou topologique,...).

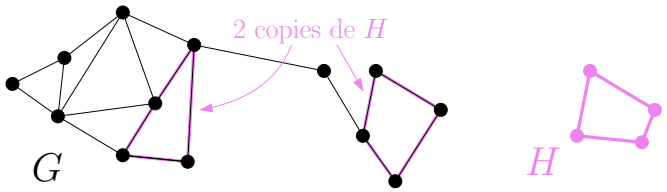


FIGURE 1. Un graphe G de taille 13, qui contient 2 occurrences H (en tant que sous-graphe induit). On note $\mathrm{Occ}(H, G) = 2$.

Le projet vise plus précisément à étudier un certain modèle de graphes aléatoires, contraints d'éviter un sous-graphe donné en tant que sous-graphe induit (voir Fig.1). Dans l'article [BBF+22] (voir également [Stu21]) les auteurs ont étudié les *cographes*. Un graphe est un cographe s'il ne contient pas P_4 comme sous-graphe induit, $P_4 =$ ●–●–●–● étant le graphe constitué d'une ligne de 4 sommets.

La genre de question que l'on se pose est la suivante :

Si un grand graphe G est contraint à éviter P_4, quels sont ses sous-graphes induits "typiques" ?

Le choix de P_4 n'est pas arbitraire, c'est l'exemple non-trivial le plus simple juste après le cas des graphes qui évitent le triangle K_3 (les questions sont alors assez différentes, voir par exemple [Łuc00] ou [Cha16] pour une théorie plus générale).

On peut commencer par regarder ce qui se passe dans un graphe aléatoire uniforme $\mathcal{G}_n$ sans aucune contrainte. Pour tout graphe fixé H on a un comportement déterministe du nombre d'occurrences de H : asymptotiquement avec grande probabilité,

$$\mathrm{occ}(H, \mathcal{G}_n) \sim \#\{k\text{-uplets possibles}\} \times \mathbb{P}(\text{un } k\text{-uplet fixé induit le graphe } H) \sim \binom{n}{k} 2^{-\binom{k}{2}}.$$

Pour le cas d'un graphe où le motif P_4 est interdit la situation est très différente :

Théorème (*Bassino-Bouvel-Féray-Gerin-Maazoun-Pierrot [BBF+22]*)

Soit $\boldsymbol{G_n}$ un cographe aléatoire uniforme de taille n. Pour tout $k \geq 1$ et H graphe fini de taille k, il existe une variable aléatoire Γ_H telle que, lorsque $n \to +\infty$,

$$\frac{\mathrm{occ}(H, \boldsymbol{G_n})}{\binom{n}{k}} \xrightarrow{(loi)} \Gamma_H.$$

Si H n'est pas un cographe, alors $\mathrm{occ}(H, \boldsymbol{G_n}) = 0$ et $\Gamma_H \equiv 0$ mais si H est un cographe, alors

- *Presque-sûrement $\Gamma_H > 0$;*
- *La variable Γ_H est "vraiment" aléatoire : $\mathrm{Var}(\Gamma_H) > 0$.*

Ce Théorème peut s'énoncer en disant que $\boldsymbol{G_n}$ admet une limite au sens des *graphons*. Intuitivement, les graphons sont les matrices d'adjacence de graphes infinis indexés par l'intervalle $[0, 1]$, à ré-étiquetage près. Le cadre théorique des graphons pour l'étude des limites de graphes denses a été posé par Lovász et Szegedy [LS06]. Le résultat principal est donc l'existence d'un *cographon brownien*, aux propriétés fractales semblables à celles du permuton séparable brownien (voir une simulation en Fig.2).

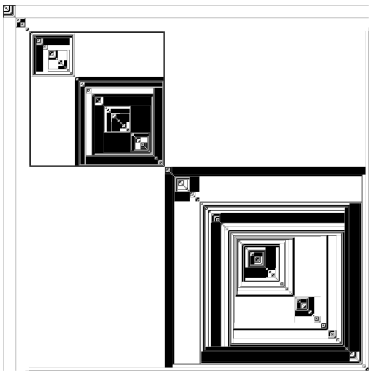
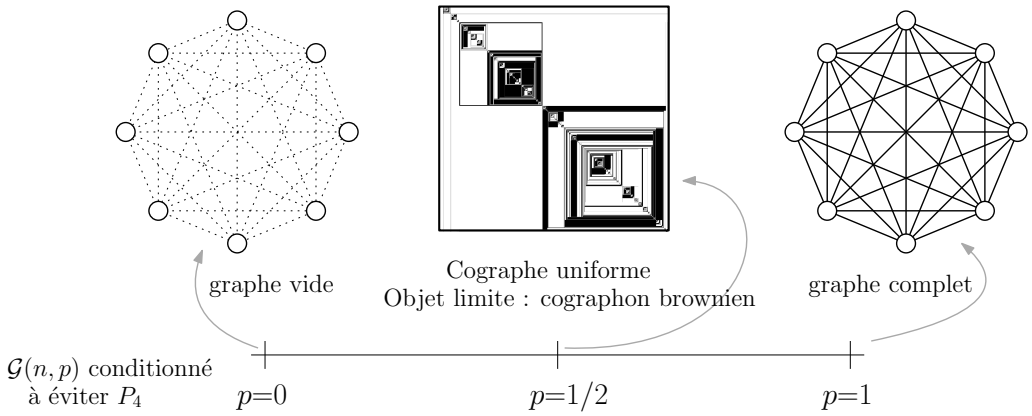


FIGURE 2. La matrice d'adjacence d'un cographe aléatoire uniforme de taille $n = 4500$. On peut voir cette simulation comme l'approximation du cographon brownien introduit dans [BBF+22].

Piste 1 : Modèle d'Erdös-Rényi conditionné à éviter un motif (limite d'échelle)

Une piste pour généraliser le résultat de [BBF+22] est la suivante. Puisqu'un graphe uniforme de taille n peut se voir comme une réalisation d'Erdös-Rényi $\mathcal{G}(n, p)$ avec $p = 1/2$, ce résultat peut se voir comme le fait qu'un $\mathcal{G}(n, 1/2)$ conditionné à éviter P_4 converge vers le cographon brownien. On peut maintenant étudier les possibles transitions de phase lorsque p passe de 0 (graphe vide) à $p = 1/2$ puis à $p = 1$ (graphe complet) :



Un problème naturel pour approfondir la compréhension des graphes aléatoires denses sous contraintes est donc de déterminer, en fonction de p, s'il existe une limite d'échelle à $\mathcal{G}(n,p)$ conditionné à éviter P_4 (ou éventuellement d'autres motifs).

Ce sujet, à la fois récent et prometteur, ouvre de nombreuses perspectives de recherche. Pour mener des calculs explicites, il faudra vraisemblablement combiner plusieurs approches complémentaires, telles que l'étude des arbres conditionnés, la combinatoire analytique, ou encore la dynamique sur les graphes.

Si jamais le cas de $\mathcal{G}(n,p)$ est inaccessible, on peut également envisager de construire d'autres modèles de graphes aléatoires biaisés par la taille et conditionnés à éviter P_4. Si les modèles sont bien choisis ils devraient avoir la même limite d'échelle que les $\mathcal{G}(n,p)$ conditionnés.

Piste 1bis : Modèle d'Erdös-Rényi conditionné à éviter un motif (statistiques extrémales)

Parallèlement à l'existence d'une limite d'échelle, on peut chercher à étudier quelques statistiques naturelles sur $\mathcal{G}(n,p)$ conditionné à éviter P_4 : diamètre, nombre chromatique, plus grande clique, plus grand ensemble indépendant,... Une motivation vient de [LRS+10, KMRS14] qui introduisent une version probabiliste de la *conjecture d'Erdős-Hajnal*.

Un sous-ensemble des sommets d'un graphe est appelé *homogène* si c'est soit une clique, soit un *independent set*. On sait (voir [Chu14]) que tout graphe de taille n possède un ensemble homogène de taille au moins logarithmique en n (et c'est optimal à une constante près). La conjecture d'Erdős et Hajnal affirme que, si un graphe de taille n évite un ou plusieurs sous-graphe induits, alors des ensembles homogènes de taille polynomiale en n existent forcément. Plus précisément, pour tout H, il existe une constante $\varepsilon = \varepsilon(H) > 0$ telle que tout graphe évitant H possède un ensemble homogène de taille n^{ε}.

La conjecture d'Erdős-Hajnal reste largement ouverte (voir [Chu14] pour l'état de l'art). Un angle d'attaque naturel consiste à remplacer "tout graphe évitant H" dans la conjecture par "presque tout graphe évitant H". Cette version plus faible a été établie dans [LRS+10]. Ce résultat a été amélioré par Kang, McDiarmid, Reed and Scott dans [KMRS14] : pour une grande famille de motifs H, un graphe aléatoire uniforme évitant H a avec forte probabilité un ensemble homogène de taille linéaire en n.

Dans [BBD+22] nous avons montré que les cographes aléatoires ne vérifient pas cette propriété d'Erdős-Hajnal de linéarité asymptotique. L'étude du problème analogue pour $\mathcal{G}(n,p)$ conditionné devrait permettre de mieux comprendre la spécificité des cographes.

Piste 2 : Autres structures avec motifs exclus

L'étude de la littérature et les différentes techniques que l'étudiant-e aura du maîtriser pour les Pistes 1 et 1bis devraient lui permettre d'étudier également d'autres structures discrètes aléatoires.

Ainsi pour les matrices à signes alternants (ASM pour *alternating sign matrices)* qui sont des objets très importants en combinatoire et physique statistique, plusieurs notions de "motifs interdits" ont été introduites [BES+24, JL07] et il serait très intéressant de développer une approche probabiliste pour ces objets.

Présentation des encadrants

Frédérique Bassino (50%)

Situation actuelle : Professeure à l'Université Paris-Nord (Laboratoire LIPN)

Thèse d'Informatique fondamentale (1996), HDR (2005)
6 Thèses encadrées (5 soutenues, 1 en cours).

Lucas Gerin (50%)

Situation actuelle : Professeur assistant à l'École Polytechnique (Laboratoire CMAP)

Thèse de Probabilités (2008), HDR (2018)
2 Thèse encadrées (en cours)

La Thèse commencera dans le contexte très stimulant de l'ANR PRC LOUCCOUM (Large Objects Under Combinatorial Constraints and Outside Uniform Models) qui a démarré en Janvier 2025, autour des structures aléatoires avec contraintes, dont Lucas Gerin est le coordinateur principal.

Bibliographie

[BBD+22] Frédérique Bassino, Mathilde Bouvel, Michael Drmota, Valentin Féray, Lucas Gerin, Mickaël Maazoun, and Adeline Pierrot. Linear-sized independent sets in random cographs and increasing subsequences in separable permutations. *Combinatorial Theory*, 2(3), 2022.

[BBF+22] Frédérique Bassino, Mathilde Bouvel, Valentin Féray, Lucas Gerin, Mickaël Maazoun, and Adeline Pierrot. Random cographs : Brownian graphon limit and asymptotic degree distribution. *Random Structures & Algorithms*, 60(2) :166–200, 2022.

[BES+24] Mathilde Bouvel, Eric S Egge, Rebecca N Smith, Jessica Striker, and Justin M Troyka. Enumeration of pattern-avoiding alternating sign matrices : An asymptotic dichotomy. *arXiv preprint arXiv :2411.07662*, 2024.

[Cha16] Sourav Chatterjee. An introduction to large deviations for random graphs. *Bulletin of the American Mathematical Society*, 53(4) :617–642, 2016.

[Chu14] Maria Chudnovsky. The Erdös-Hajnal conjecture - a survey. *J. Graph Theory*, 75(2) :178–190, 2014.

[JL07] Robert Johansson and Svante Linusson. Pattern avoidance in alternating sign matrices. *Annals of Combinatorics*, 11 :471–480, 2007.

[KMRS14] Ross J. Kang, Colin McDiarmid, Bruce Reed, and Alex Scott. For most graphs H, most H-free graphs have a linear homogeneous set. *Random Structures Algorithms*, 45(3) :343–361, 2014.

[LRS+10] Martin Loebl, Bruce Reed, Alex Scott, Andrew Thomason, and Stéphan Thomassé. Almost all H-free graphs have the Erdös-Hajnal property. In *An irregular mind*, volume 21 of *Bolyai Soc. Math. Stud.*, pages 405–414. János Bolyai Math. Soc., Budapest, 2010.

[LS06] László Lovász and Balázs Szegedy. Limits of dense graph sequences. *Journal of Combinatorial Theory, Series B*, 96(6) :933–957, 2006.

[Łuc00] Tomasz Łuczak. On triangle-free random graphs. *Random Structures & Algorithms*, 16(3) :260–276, 2000.

[Stu21] Benedikt Stufler. Graphon convergence of random cographs. *Random Structures & Algorithms*, 59(3) :464–491, 2021.