# Limite de grands graphes aléatoires denses, et extension aux graphes pondérés

Introduction au domaine de recherche

Julien Weibel, sous la direction de Romain Abraham et Jean-François Delmas

5 octobre 2020

## 1 Introduction

Un graphe ou réseau est composé d'un ensemble d'éléments discrets et distincts (les sommets), et dont certains d'entre eux interagissent par paire (modélisé par des arêtes). De nombreux phénomènes intéressants peuvent être modélisés en termes de réseaux. C'est par exemple le cas des interactions entre particules en physique statistique, entre neurones dans le cerveau, ou encore entre individus dans un réseau social. Le point commun entre ces exemples, et bien d'autres, et que les réseaux qui interviennent pour les modéliser sont de grands réseaux denses : des graphes comportant un très grand nombre de nœuds ainsi qu'un important nombre d'arêtes.

Le problème est que la taille trop importante de ces graphes empêche de les représenter en machine. Un moyen de contourner ce problème consiste à choisir aléatoirement un sous-graphe de taille plus raisonnable. La loi de ce sous-graphe est caractériser par les nombres d'homomorphismes de graphes associés au graphe initial. De plus, la loi de ce sous-graphe permet de calculer la plupart des propriétés intéressantes du graphe initial.

Une autre problématique soulevée par ces grands graphes est leur approximation par ajouts successifs de sommets, c'est-à-dire par une suite de graphes qui se déduisent les uns des autres par l'ajout d'un ou plusieurs sommets (et des arêtes associées). La question qui se pose alors est celle de l'existence d'un objet limite pour une suite de graphes dont le nombre de sommets est croissant (et tend donc vers l'infini). Cet objet limite existe et s'appelle un *graphon*, qui est une généralisation continue de la matrice d'adjacence d'un graphe. Chaque graphe fini peut également se représenter comme un graphon. Les nombres d'homomorphismes de graphes se généralisent en quantité d'homomorphismes d'un graphe dans un graphon, donnant des analogues continues plus simple à calculer que leurs versions combinatoires. Enfin, il y a un lien entre la convergence d'une suite de graphes vers un graphon, et la convergence des nombres d'homomorphismes de graphes associés aux graphes de cette suite. Ce qui justifie l'introduction de la notion de graphon.

Cette introduction au domaine de recherche est composée de trois parties. Dans la première partie, nous nous intéressons aux graphes et aux liens entre homomorphismes de graphes et échantillonnage de sous-graphe. Dans la deuxième partie, nous introduisons les graphons et montrons qu'ils sont les objets limites recherchés et en quoi ils généralisent la notion de graphe. Enfin, dans la dernière partie, nous présenterons notre travail de recherche sur l'extension de ces outils aux graphes pondérés.

# 2 Graphes et nombres d'homomorphismes

Dans cette section, on présente brièvement la théorie des homomorphismes de graphes finis. On définit les nombres d'homomorphismes de graphes, et on montre qu'ils caractérisent un graphe. On montre également le lien entre les nombres d'homomorphismes de graphes et l'échantillonnage de sous-graphes.

## 2.1 Définition d'un graphe, et notations

Un graphe (non-orienté) $G = (V, E)$ est la donnée d'un ensemble de sommets $V$, et d'un ensemble d'arêtes $E$ qui est un sous-ensemble symétrique de $V \times V$. Dans tout ce qui suit, nous aurons en permanence besoin de considérer des graphes. Nous introduisons donc quelques notations qui seront utilisées tout au long de ce document. On note $V(G)$ l'ensemble des nœuds (ou sommets) du graphe $G$, et $v(G) = |V(G)|$ le nombre de ses sommets. On note $E(G)$ l'ensemble des arêtes du graphe $G$, et $e(G) = |E(G)|$ le nombre de ses arêtes. Si $G$ est un graphe pondéré (resp. décoré), alors pour toute arête $(i, j) \in E(G)$, on note $e_G(i, j)$ le poids (resp. la décoration) associé à cette arête.

## 2.2 Les nombres d'homomorphismes

Dans cette sous-section, nous introduisons les nombres d'homomorphismes de graphes, qui comptent le nombre d'application entre deux graphes préservant la structure de graphe. Nous montrons également que les nombres d'homomorphismes de graphes permettent de caractériser entièrement un graphe.

### 2.2.1 Définition des nombres d'homomorphismes

Un *homomorphisme* d'un graphe $F$ dans un graphe $G$ est une application $h : V(F) \to V(G)$ préservant la relation d'adjacence définie par les arêtes : $ij \in E(F)$ implique $h(i)h(j) \in E(G)$. On note $\mathtt{hom}(F, G)$ le nombre d'homomorphisme du graphe $F$ dans le graphe $G$.

On peut également imposer que les copies des nœuds de $F$ dans le graphe $G$ soient distinctes, ce qui revient à imposer que l'homomorphisme $h : V(F) \to V(G)$ soit injectif. On note $\mathtt{inj}(F, G)$ le nombre d'homomorphisme injectif du graphe $F$ dans le graphe $G$. Enfin, on peut également imposer que l'homomorphisme $h : V(F) \to V(G)$ préserve la relation d'adjacence et de non-adjacence : $ij \in E(F)$ équivaut à $h(i)h(j) \in E(G)$. On note $\mathtt{ind}(F, G)$ le nombre d'homomorphisme injectif du graphe $F$ dans le graphe $G$ préservant aussi la relation de non-adjacence.

La définition de nombre d'homomorphisme peut également être étendue aux graphes pondérés. Soit $G$ un graphe pondéré dont on note $\alpha_u(G)$ le poids du sommet $u \in V(G)$ et $beta_{uv}(G)$ le poids de l'arête $uv \in E(G)$. Alors pour toute fonction $\varphi : V(F) \to V(G)$, on assigne les poids :

$$\alpha_\varphi = \prod_{u \in V(F)} \alpha_{\varphi(u)}(G), \quad \text{et} \quad \mathtt{hom}_\varphi(F, G) = \prod_{uv \in E(F)} \beta_{\varphi(u)\varphi(v)}(G) \ .$$

On définit alors le nombre d'homomorphisme du graphe simple $F$ dans le graphe pondéré $G$ par :

$$\mathtt{hom}(F, G) = \sum_{\varphi : V(F) \to V(G)} \alpha_\varphi \mathtt{hom}_\varphi(F, G) \ .$$

De même, on définit le nombre d'homomorphisme injectif du graphe simple $F$ dans le graphe pondéré $G$ par :

$$\mathtt{inj}(F, G) = \sum_{\substack{\varphi : V(F) \to V(G) \\ \varphi \text{ injective}}} \alpha_\varphi \mathtt{hom}_\varphi(F, G) \ .$$

### 2.2.2 Définition des densités homomorphiques

Ces nombres d'homomorphismes grandissent avec la taille du graphe $G$, il est donc souvent plus parlant de les normaliser pour obtenir des *densités homomorphiques*. En écrivant $n = v(G)$ et $k = v(F)$, on définit la densité homomorphique de $F$ dans $G$ par :

$$t(F, G) = \frac{\mathtt{hom}(F, G)}{n^k} \ ,$$

qui est la probabilité qu'une fonction $V(F) \to V(G)$ choisi aléatoirement soit un homomorphisme. On note $(n)^k = n(n-1)\cdots(n-k+1)$ le nombre d'injection de $[k]$ dans $[n]$ (et donc aussi le nombre d'injection de $V(F)$ dans $V(G)$). On définit de même

$$t_{\mathtt{inj}}(F, G) = \frac{\mathtt{inj}(F, G)}{(n)_k} \ ,$$

qui est la probabilité qu'une fonction injective $V(F) \to V(G)$ soit un homomorphisme ; et

$$t_{\mathtt{ind}}(F, G) = \frac{\mathtt{ind}(F, G)}{(n)_k} \ ,$$

qui est la probabilité qu'une fonction injective $V(F) \to V(G)$ soit un homomorphisme préservant l'adjacence et la non-adjacence. Si $G$ est un graphe pondéré, alors on définit

$$\alpha_G = \sum_{v \in V(G)} \alpha_v(G), \quad \text{et} \quad t(F, G) = \frac{\mathtt{hom}(F, G)}{\alpha_G^{v(F)}} \ .$$

### 2.2.3 Relations entre les nombres d'homomorphismes

Les différents nombres d'homomorphismes sont des quantités extrêmement liées entre elles. Pour deux graphes simples $F$ et $G$, on a :

$$\mathtt{inj}(F, G) = \sum_{F' \supseteq F} \mathtt{ind}(F', G) \ ,$$

où $F'$ décrit l'ensemble des graphes simples obtenus à partir du graphe $F$ en rajoutant des arêtes. De plus, on a :

$$\mathtt{hom}(F, G) = \sum_{P} \mathtt{inj}(F/P, G) \ ,$$

où $P$ décrit l'ensemble des partitions $V(F)$, et $F/P$ est le quotient simple de graphe.

Il est également possible d'échanger le rôle des nombres d'homomorphismes dans les équations précédentes en utilisant le principe d'inclusion-exclusion (ou la méthode d'inversion de Möbius) :

$$\mathtt{ind}(F, G) = \sum_{F' \supseteq F} (-1)^{e(F') - e(F)} \mathtt{inj}(F', G) \ ,$$

et

$$\mathtt{inj}(F, G) = \sum_{P} \mu_P \ \mathtt{hom}(F/P, G) \ ,$$

où $F'$ et $P$ décrive les mêmes ensembles que précédemment, et où les $\mu_P$ sont des coefficients qui peuvent être décris explicitement. Ainsi, la connaissance des $(\mathtt{hom}(F, G) : F$ graphe simple$)$ est équivalente à la connaissance des $(\mathtt{inj}(F, G) : F$ graphe simple$)$, et aussi équivalente à la connaissance des ($\mathtt{ind}(F,G)$ : $F$ graphe simple). Ceci est particulièrement intéressant car les nombres d'homomorphismes ($\mathtt{hom}(F,G)$ : $F$ graphe simple) sont plus simple à calculer.

Ces relations peuvent également s'étendre aux densités homomorphiques $t_{\mathtt{inj}}$ et $t_{\mathtt{ind}}$, ce qui donne :

$$t_{\mathtt{inj}}(F,G) = \sum_{F' \supseteq F} t_{\mathtt{ind}}(F',G) \ ,$$

et

$$t_{\mathtt{ind}}(F,G) = \sum_{F' \supseteq F} (-1)^{e(F')-e(F)} t_{\mathtt{inj}}(F',G) \ .$$

Cependant, pour les densités homomorphiques $t$ et $t_{\mathtt{inj}}$, la relation n'est pas simple à cause de la différence de constante de normalisation. Comme nous sommes principalement intéressés par de grands graphes, l'inégalité suivante est suffisante en générale pour relier ces deux quantités :

$$|t_{\mathtt{inj}}(F,G) - t(F,G)| \leq \frac{1}{v(G)} \binom{v(F)}{2} \ .$$

### 2.2.4 Caractérisation des graphes par les nombres d'homomorphismes

Les nombres d'homomorphismes de graphes permettent de caractériser les graphes simples, au sens des deux propositions suivantes.

**Proposition 1.** *Soit $G$ et $G'$ deux graphes simples tels que* $\mathtt{hom}(F,G) = \mathtt{hom}(F,G')$ *pour tout graphe simple $F$. Alors $G = G'$.*

**Proposition 2.** *Soit $G$ et $G'$ deux graphes simples tels que* $\mathtt{hom}(G,F) = \mathtt{hom}(G',F)$ *pour tout graphe simple $F$. Alors $G = G'$.*

Notons que grâce aux relations entre les différents types de nombres d'homomorphismes que nous venons de voir dans la sous-section précédente, les résultats des propositions 1 et 2 restent vrais si l'on remplace `hom` par `inj` ou `ind`.

Pour un graphe simple $G$ et un entier $p$, le graphe $G(p)$ est obtenu à partir de $G$ en remplaçant chaque nœud par $p$ nœuds jumeaux. le graphe simple C'est-à-dire, $G(p)$ est le graphe dont l'ensemble des sommets contient $p$ copies de chacun des sommets de $G$, et dont deux sommets sont reliés par une arête si et seulement si les sommets dont ils sont des copies sont reliés par une arête dans le graphe $G$. Les densités homomorphiques $t(\cdot,G)$ ne caractérisent plus le graphe $G$ comme dans les propositions précédentes, car $t(F,G(p)) = t(F,G)$.

Cependant, les densités homomorphiques caractérisent les graphes simples au sens plus faible de la proposition suivante.

**Proposition 3.** *Soit $G_1$ et $G_2$ deux graphes simples tels que $t(F,G_1) = t(F,G_2)$ pour tout graphe simple $F$. Alors il existe un troisième graphe simple $G$ et deux entiers $p_1$ et $p_2$ tels que $G_1 \cong G(p_1)$ et $G_2 \cong G(p_2)$.*

## 2.3 Lien avec l'échantillonnage de sous-graphes

Comme dans nos problèmes on considère de grands graphes denses, il n'est pas possible de les appréhender dans leur ensemble. Il faut donc réduire la taille du graphe $G$ à observer, en regardant le sous-graphe induit par un sous-ensemble des sommets choisi aléatoirement.

La loi de ce sous-graphe induit est caractérisée par les *nombres d'homomorphismes* (ou plutôt par les densités homomorphiques) d'un graphe $F$ dans le graphe $G$, pour tous les choix du graphe simple $F$.

**Proposition 4.** *Pour deux graphes simples $F$ et $G$, $t_{\mathtt{ind}}(F, G)$ est la probabilité qu'en échantillonnant $v(F)$ nœuds de $G$ (ordonnés et sans répétition), ces nœuds induisent $F$ comme sous-graphe de $G$.*

# 3 Les objets limites des suites de graphes : les graphons

Dans cette section, nous présentons les outils principaux pour l'étude des graphons. Les *graphons* sont les objets limites des suites de graphes convergentes. On peut étendre certaines notions comme les densités homomorphiques aux graphons. Ainsi, les graphons peuvent être vus comme une généralisation continue des graphes finis.

L'article [Gla16] fournit une introduction brève mais efficace sur ce qu'est un graphon. La monographie [Lov12] écrite par Lovász rassemble les résultats les plus importants obtenus sur les graphons.

## 3.1 Les graphons

### 3.1.1 Définition d'un graphon

On définit un *noyau* comme une fonction bornée symétrique mesurable $W : [0, 1]^2 \to \mathbb{R}$. Si la fonction $W$ est à valeurs dans $[0, 1]$, on dit alors que $W$ est un *graphon*. On identifie les noyaux ou graphons qui sont égaux presque partout. On peut considérer un graphon qui prend uniquement les valeurs 0 et 1 comme un graphe simple dont l'ensemble des nœuds est l'intervalle $[0, 1]$.

Les noyaux généralise les graphes pondérés dans le sens suivant. Un noyau $W$ est appelé une *step-function* s'il existe une partition $S_1 \cup \cdots \cup S_k$ de $[0, 1]$ en ensembles mesurables telle que $W$ est constant sur les ensembles produits $S_i \times S_j$. Les ensembles $S_i$ sont appelés les *marches* de $W$. Pour tout graphe pondéré $G$ avec $n$ sommets, on définit le noyau step-function $W_H$ comme suit : on divise $[0, 1]$ en $n$ intervalles $J_1, \cdots , J_n$ intervalles de longueur respectives $\lambda(J_i) = \alpha_i/\alpha_G$, et pour tout $x \in J_i$ et $y \in J_j$, on pose $W_G(x, y) = \beta_{ij}(G)$. Remarquons que la fonction $W_G$ dépend de la façon dont les sommets de $G$ sont étiquetés.

Réciproquement, toute step-function $W$ correspond à un graphe pondéré : soit $S_1, \cdots , S_k$ les marches de $W$, alors ce graphe pondéré est défini sur l'ensemble de sommets $[k] = \{1, \cdots , k\}$, et l'arête $ij$ porte le poids $W(x, y)$ où $x \in S_i$ et $y \in S_j$.

Remarquons que nous pouvons voir tout graphe simple (non-pondéré) $G$ comme un graphe complet pondéré dont le poids de chaque arête vaut 1 si l'arête est présente dans le graphe $G$ et 0 si l'arête est absente.

### 3.1.2 Généralisation des nombres d'homomorphismes

Les densités homomorphiques dans les graphes peuvent être étendues aux graphons, et plus généralement aux noyaux. Pour un noyau $W$ et un multi-graphe $F = (V, E)$ (sans boucle, i.e. sans arête de la forme $(v, v)$ pour $v \in V$), on définit la *densité homomorphique* de $F$ dans $W$ par :

$$t(F, W) = \int_{[0,1]^V} \prod_{ij \in E} W(x_i, x_j) \prod_{i \in V} \mathrm{d}x_i \ .$$

On peut penser $W$ comme un graphe dont l'ensemble des sommets est l'intervalle $[0, 1]$ et dont le poids de l'arête $xy$ est $W(x, y)$. Alors, la formule ci-dessus peut être vue comme un analogue infini des densités homomorphiques de graphes pondérés. On a même que cette définition en est une généralisation, au sens où pour un graphe pondéré $G$, on a :

$$t(F, G) = t(F, W_G) \ .$$

Dans ce contexte, la notion de densité injective $t_{\mathtt{inj}}$ n'a pas d'impact. Cela vient du fait qu'une affectation $i \mapsto x_i$ ($i \in V(F)$, $x_i \in [0,1]$) est injective avec probabilité 1. Ainsi, on a $t_{\mathtt{inj}}(F, W) = t(F, W)$ pour tout noyau $W$ et tout graphe $F$. Cependant, la densité de sous-graphe induit reste intéressante à définir, et peut être exprimée via une formule intégrale relativement simple :

$$t_{\mathtt{ind}}(F, W) = \int_{[0,1]^V} \prod_{ij \in E} W(x_i, x_j) \prod_{ij \in \mathcal{P}_2(V) \setminus E} (1 - W(x_i, x_j)) \prod_{i \in V} \mathrm{d}x_i \ .$$

On dispose également d'équivalent des formules reliant les différents types de nombres d'homomorphismes :

$$t(F, W) = \sum_{F' \supseteq F} t_{\mathtt{ind}}(F', W) \ ,$$

et

$$t_{\mathtt{ind}}(F, W) = \sum_{F' \supseteq F} (-1)^{e(F') - e(F)} t(F', W) \ ,$$

où $F'$ décrit l'ensemble des graphes simples obtenus à partir de $F$ en rajoutant des arêtes.

### 3.1.3 Isomorphisme faible

Les graphes que nous cherchons à étudier sont des graphes non-étiquetés. Hors les noyaux et graphons que nous venons de définir correspondent aux graphes étiquetés. On introduit donc une relation d'isomorphisme faible permettant de "désétiqueter" les graphons.

On dit que deux noyaux $U$ et $W$ sont *isomorphes à un ensemble de mesure nulle près* s'il existe une fonction $\varphi : [0,1] \to [0,1]$ inversible et préservant la mesure telle que $U^\varphi(x, y) = U(\varphi(x), \varphi(y)) = W(x, y)$ presque partout. On considère également une version légèrement plus faible de cette relation d'isomorphisme où on autorise la fonction $\varphi$ à ne pas être inversible. Ceci est motivé par le fait qu'une fonction préservant la mesure n'est pas nécessairement inversible. Cette définition de relation d'isomorphisme faible est justifiée par le résultat suivant.

**Proposition 5.** *Soit $W$ un noyau, et $\varphi : [0,1] \to [0,1]$ fonction préservant la mesure. Alors pour tout multi-graphe $F$, on a $t(F, W^\varphi) = t(F, W)$.*

On voudrait dire que les noyaux $W$ et $W^\varphi$ sont "faiblement isomorphe". Cependant, comme les fonctions préservant la mesure ne sont pas forcément inversible, la relation d'isomorphisme à un ensemble de mesure nulle près n'est a priori pas symétrique. On dit donc que deux noyaux $U$ et $W$ sont *faiblement isomorphes* si $t(F, U) = t(F, W)$ pour tout graphe simple $F$. Nous verrons plus loin, grâce à des résultats d'échantillonnage sur les graphons, que ces deux notions d'isomorphismes sont en réalité équivalentes.

## 3.2 La distance de coupe

### 3.2.1 La distance de coupe pour les graphes

On commence par définir la distance de coupe entre deux graphes. Pour un graphe $G$, et $S, T \subset V(G)$ deux sous-ensembles de sommets, on note $e_G(S, T) = \sum_{i \in S, j \in T} e_G(i, j)$ la somme des poids des arêtes entre $S$ et $T$. Pour deux graphes $G$ et $G'$ définit sur le même ensemble de $n$ nœuds, on définit la distance de coupe :

$$d_\square(G,G') = \max_{S,T\subset V(G)} \frac{|e_G(S,T) - e_{G'}(S,T)|}{n^2} \ .$$

Le fait de normaliser par $n^2$ au lieu de $|S|\times|T|$ permet d'éviter de favoriser les choix de petits ensembles.

Pour deux graphes $G$ et $G'$ non-étiquetés définis sur des ensembles de sommets différents mais de même cardinal $n$, on définit la distance de coupe par :

$$\hat{\delta}_\square(G,G') = \min_{\hat{G},\hat{G}'} d_\square(\hat{G},\hat{G}') \ ,$$

où $\hat{G}$ et $\hat{G}'$ parcourent l'ensemble des étiquetages des graphes $G$ et $G'$ respectivement.

On définit un *recouvrement fractionnel* (en anglais *fractionnal overlay*) des graphes $G$ et $G'$ comme une matrice $n\times n'$ (où $n = v(G)$ et $n' = v(G')$) à coefficients positifs $X = (X_{iu} : i\in V(G), u\in V(G'))$ telle que pour tout $i\in V(G)$, $\sum_{u=1}^{n'} X_{iu} = \frac{1}{n}$, et pour tout $u\in V(G')$, $\sum_{i=1}^{n} X_{iu} = \frac{1}{n}$. Par exemple, si $n = n'$ et que $\sigma : V(G)\to V(G')$ est une bijection, alors $X_{iu} = \frac{1}{n}\mathbb{1}(\sigma(i) = u)$ définit un recouvrement fractionnel. On note $\mathcal{X}(G,G')$ l'ensemble des recouvrement fractionnel des graphes $G$ et $G'$.

Pour deux graphes $G$ et $G'$ définis sur deux ensembles de nœuds (potentiellement) distincts, on définit la distance de coupe étiquetées pour un recouvrement fractionnel $X\in\mathcal{X}(G,G')$ par :

$$d_\square(G,G',X) = \max_{Q,R\subset V\times V'} \left| \sum_{iu\in Q, jv\in R} X_{iu}X_{jv}(\mathbb{1}(ij\in E(G)) - \mathbb{1}(uv\in E(G'))) \right| \ .$$

Puis, on définit la distance de coupe non-étiquetées comme une optimisation sur le choix du recouvrement fractionnel $X\in\mathcal{X}(G,G')$ :

$$\delta_\square(G,G') = \min_{X\in\mathcal{X}(G,G')} d_\square(G,G',X) \ .$$

Pour des graphes pondérés $G$ et $G'$, on écrit les poids normalisés des arêtes $\alpha_i = \alpha_i(G)$ et $\alpha'_i = \alpha_i(G')$, et les poids des arêtes $\beta_{ij} = \beta_{ij}(G)$ et $\beta'_{ij} = \beta_{ij}(G')$. Un *recouvrement fractionnel* de $G$ et $G'$ est définit comme une matrice $n\times n'$ à coefficients positifs $X$ telle que $\sum_{u=1}^{n'} X_{iu} = \alpha_i$ et $\sum_{i=1}^{n} X_{iu} = \alpha'_i$. On définit alors la distance de coupe étiquetées :

$$d_\square(G,G',X) = \max_{Q,R\subset V\times V'} \left| \sum_{iu\in Q, jv\in R} X_{iu}X_{jv}(\beta_{ij} - \beta'_{uv}) \right| \ .$$

On définit de même que pour les graphes simples, la distance de coupe non-étiquetées par :

$$\delta_\square(G,G') = \min_{X\in\mathcal{X}(G,G')} d_\square(G,G',X) \ .$$

Pour deux graphes $G$ et $G'$ possédant le même nombre de nœuds, les distances $\delta_\square$ et $\hat{\delta}_\square$ ne donnent pas forcément la même valeur, néanmoins on a :

$$\delta_\square(G,G') \le \hat{\delta}_\square(G,G') \ .$$

Enfin, remarquons que la distance de coupe $\delta_\square$ est en réalité une pseudo-distance : elle peut être nulle pour deux graphes $G$ et $G'$ distincts. C'est par exemple le cas si $G' = G(k)$ pour $k > 1$ entier.

### 3.2.2 La distance de coupe pour les graphons et noyaux

Pour un noyau $W$, on définit sa norme de coupe par :

$$\| W \|_\square = \sup_{S,T\subset[0,1]} \left| \int_{S\times T} W(x,y) \, \mathrm{d}x\mathrm{d}y \right| \ .$$

De plus, un résultat important est que le supremum dans cette définition est atteint. Remarquons la comparaison entre la norme de coupe et la norme $L_1$ :

$$\| W \|_\square \leq \| W \|_1 = \int_{[0,1]^2} |W(x,y)| \, \mathrm{d}x\mathrm{d}y \ .$$

La norme de coupe $\| \cdot \|_\square$ induit la distance de coupe étiquetés $d_\square$ sur l'ensemble des noyaux, et définie par $d_\square(U,W) = \| U - W \|_\square$.

On définit $\bar{S}_{[0,1]}$ l'ensemble des fonctions $[0,1] \to [0,1]$ préservant la mesure, et $S_{[0,1]}$ l'ensemble des fonctions inversibles $[0,1] \to [0,1]$ préservant la mesure. La distance de coupe non-étiquetées entre deux noyaux $U$ et $W$ est définie par :

$$\delta_\square(U,W) = \inf_{\varphi\in S_{[0,1]}} d_\square(U, W^\varphi) \ .$$

Cette définition de la distance de coupe (non-étiquetées) pour les noyaux prolonge celle pour les graphes pondérés, au sens où pour $G$ et $G'$ deux graphes pondérés, on a :

$$\delta_\square(G,G') = \delta_\square(W_G, W_{G'}) \ .$$

De plus, l'infimum dans la définition de la distance de coupe $\delta_\square$ est atteint, au sens du résultat suivant.

**Théorème 6.** *Soient $U$ et $W$ deux noyaux. Alors on a*

$$\delta_\square(U,W) = \inf_{\varphi\in S_{[0,1]}} d_\square(U,W^\varphi) = \inf_{\varphi\in \bar{S}_{[0,1]}} d_\square(U,W^\varphi) = \min_{\varphi,\psi\in \bar{S}_{[0,1]}} d_\square(U^\varphi, W^\psi) \ .$$

Et donc, la distance de coupe $\delta_\square$ est une vraie distance (vérifie la condition de séparation) sur l'espace des noyaux non-étiquetés (c'est-à-dire où l'on identifie les noyaux faiblement isomorphes). Le théorème suivant montre que la distance de coupe $\delta_\square$ vérifie même mieux.

**Théorème 7.** *L'espace des graphons non-étiquetés muni de la distance de coupe $\delta_\square$ est compact.*

Enfin, on dispose du lemme suivant montrant que les noyaux step-function sont denses dans l'espace des noyaux pour la norme de coupe, et donnant même une borne concrète sur la qualité de l'approximation.

**Lemme 8.** *Pour tout $W$ noyau et $k \geq 1$ entier, il existe un noyau step-function $U$ avec $k$ marches tel que*

$$\| W - U \|_\square \leq \frac{2}{\sqrt{\log k}} \| W \|_2 \ .$$

## 3.3 Échantillonnage de graphons

### 3.3.1 $W$-graphe aléatoire

Soit $S = (x_1, \cdots, x_n)$ ensemble ordonné comportant $n$ éléments de $[0,1]$, et soit $W$ graphon. On définit le graphe pondéré (étiqueté) $\mathbb{H}(S, W)$ d'ensemble de sommets $[n] = \{1, \cdots, n\}$, et dont l'arête $ij$ porte le poids $W(x_i, x_j)$.

Soit $H$ un graphe dont les arêtes sont pondérés avec des poids dans $[0,1]$. On définit le graphe aléatoire (non-pondéré) $\mathbb{G}(H)$ de même ensemble de sommets que $H$, et dont les arêtes sont indépendantes, l'arête $ij$ suivant une loi de Bernoulli de paramètre $e_H(i,j) \in [0,1]$ (la valeur 1 représentant une arête présente, et 0 pour une arête absente). Ceci permet de définir le graphe aléatoire $\mathbb{G}(S, W) = \mathbb{G}(\mathbb{H}(S, W))$.

Pour $n \geq 1$ entier, on définit $S = (X_1, \cdots, X_n)$ ensemble ordonné aléatoire de $n$ éléments de $[0,1]$, les $X_i$ étant indépendants et identiquement distribués de loi uniforme sur $[0,1]$. Alors, les graphes aléatoires échantillonnés à partir du graphon $W$ sont définis comme $\mathbb{H}(n, W) = \mathbb{H}(S, W)$ et $\mathbb{G}(n, W) = \mathbb{G}(S, W)$.

### 3.3.2 Estimation de la distance de coupe par échantillonnage

Dans cette sous-section, on présente différents résultats permettant de contrôler la distance de coupe entre graphons et graphes échantillonnés à partir de ces graphons.

Nous commençons avec un lemme montrant que le graphe $\mathbb{G}(H)$ est proche du graphe $H$ avec grande probabilité.

**Lemme 9.** *Soit un graphe $H$ comportant $q$ sommets, et dont les arêtes sont pondérés avec des poids dans $[0,1]$. Alors pour tout $\varepsilon \geq 10/\sqrt{q}$, on a*

$$\mathbb{P}(d_\square(\mathbb{G}(H), H) > \varepsilon) \leq e^{-\varepsilon^2 q^2/100} \ .$$

Nous avons ensuite le premier lemme d'échantillonnage montrant que la distance de coupe entre deux graphons peut être approximer avec grande probabilité par échantillonnage de ces graphons.

**Lemme 10** (Premier lemme d'échantillonnage)**.** *Soient $U$ et $W$ deux graphons, et soit $k \geq 1$ entier. Soit $S = (X_1, \cdots, X_k)$ un ensemble ordonné aléatoire de $k$ éléments, les $X_i$ étant indépendants et identiquement distribués de loi uniforme sur $[0,1]$.*

1. *Avec probabilité au moins $1 - 4e^{-\sqrt{k}/10}$, on a*

$$|d_\square(\mathbb{H}(S, U), \mathbb{H}(S, W)) - \| U - W \|_\square| \leq \frac{8}{k^{1/4}} \ .$$

2. *Avec probabilité au moins $1 - 5e^{-\sqrt{k}/10}$, on a*

$$|d_\square(\mathbb{G}(S, U), \mathbb{G}(S, W)) - \| U - W \|_\square| \leq \frac{10}{k^{1/4}} \ .$$

Nous avons enfin un lemme montrant qu'avec grande probabilité un graphe échantillonné est proche de son graphon modèle pour la distance de coupe.

**Lemme 11** (Second lemme d'échantillonnage)**.** *Soit $W$ un graphon, et soit $k \geq 1$ entier.*

1. *Avec probabilité au moins $1 - 2\exp(-k/(2\log k))$, on a*

$$\delta_\square(\mathbb{H}(k, W), W) \leq \frac{20}{\sqrt{\log k}} \ .$$

2. *Avec probabilité au moins* $1 - 2\exp(-k/(2\log k))$, *on a*

$$\delta_\square(\mathbb{G}(k,W),W) \leq \frac{22}{\sqrt{\log k}} \ .$$

### 3.3.3 Lemmes de comptage

Dans cette sous-section, nous présentons les deux lemmes de comptage qui permettent de relier la distance de coupe aux densités homomorphiques.

Le lemme de comptage permet de contrôler l'écart entre les densités homomorphiques de deux graphons à partir de leur écart pour la distance de coupe.

**Lemme 12** (Lemme de comptage)**.** *Soient $U$ et $W$ deux graphons, et $F$ graphe simple, alors*

$$|t(F,U) - t(F,W)| \leq e(F)\delta_\square(U,W) \ .$$

Une conséquence du lemme de comptage est que pour tout graphe simple $F$, la fonction $W \mapsto t(F,W)$ est Lipschitzienne.

Réciproquement, le lemme de comptage inverse permet de contrôler la distance de coupe à partir d'un contrôle sur les densités homomorphiques.

**Lemme 13** (Lemme de comptage inverse)**.** *Soient $U$ et $W$ deux graphons, et $k \geq 1$ entier. Supposons que pour tout graphe simple $F$ à $k$ sommets, on ait $|t(F,U) - t(F,W)| \leq 2^{-k^2}$. Alors,*

$$\delta_\square(U,W) \leq \frac{50}{\sqrt{\log k}} \ .$$

Enfin, les deux lemmes de comptage permettent d'obtenir la caractérisation suivante permettant de relier la distance de coupe et l'isomorphisme faible de noyaux.

**Corollaire 14.** *Soient $U$ et $W$ deux noyaux. Il y a équivalence entre :*

1. $\delta_\square(U,W) = 0$
2. *les noyaux $U$ et $W$ sont faiblement isomorphes (i.e. $t(F,U) = t(F,W)$ pour tout graphe simple $F$)*
3. *il existe $\varphi, \psi \in \bar{S}_{[0,1]}$ telles que $U^\varphi = W^\psi$ presque partout*

## 3.4 Convergence de suite de grands graphes denses

On souhaite définir une notion de convergence pour des suites de grands graphes denses. Comme nos graphes sont grands, on souhaite définir cette notion de convergence à partir de l'échantillonnage de sous-graphe induit : on choisit aléatoirement et uniformément un sous-ensemble de $k$ sommets de $V(G)$, et on retourne le sous-graphe induit par ces sommets. Comme dit en proposition 4, la probabilité que le sous-graphe observé soit $F$ est la quantité $t_{ind}(F,G)$. On dit qu'une suite de graphes $(G_n)$ avec $v(G_n) \to \infty$ est *convergente* si les densités des sous-graphes induits $t_{ind}(F,G_n)$ convergent pour tout graphe fini $F$.

Il est plus pratique de travailler avec les densités homomorphiques $t(F,G_n)$ ou les densités de sous-graphes $t_{inj}(F,G_n)$, ce qui ne modifie pas la notion de convergence que nous venons de définir. Les relations entre $t_{inj}$ et $t_{ind}$ montrée précédemment montre que ces quantités peuvent s'écrire comme des combinaisons linéaires les une des autres. Donc, les $t_{inj}(F,G_n)$, pour tout $F$ graphe simple, convergent lorsque $n \to \infty$ si et seulement si les $t_{ind}(F,G_n)$ convergent également lorsque $n \to \infty$. Pour les densités homomorphiques, l'argument est un peu plus complexe : on sait que $t(F,G) - t_{inj}(F,G) = O(1/v(G))$, et donc cette différence tend vers 0 si $v(G) \to \infty$. Et donc, les $t(F,G)$ tendent vers une limite, pour tout graphe simple $F$, lorsque $n \to \infty$ si et seulement si les $t_{inj}(F,G_n)$ font de même.

Les résultats suivants montre le lien entre cette définition de convergence de suite de graphes, la distance de coupe et les graphons.

Le théorème suivant montre que la distance de coupe $\delta_\square$ est "la bonne" pour étudier la convergence des suites de graphes.

**Théorème 15.** *Une suite de graphes simples $(G_n)$ avec $v(G_n) \to \infty$ est convergente si et seulement si elle est de Cauchy pour la distance $\delta_\square$.*

Le théorème précédent peut de plus être étendu au graphons, ce qui donne le théorème suivant.

**Théorème 16.** *Soit $(W_n)$ une suite de graphons, et soit $W$ un graphon. Alors, les $t(F,W_n)$ convergent pour tout graphe simple $F$ lorsque $n \to \infty$ si et seulement si la suite $(W_n)$ est de Cauchy pour la distance $\delta_\square$. De plus, $t(F,W_n) \to t(F,W)$ pour tout graphe simple $F$ si et seulement si $\delta_\square(W_n,W) \to 0$.*

La compacité de l'espace des graphons (non-étiquetés) permet de montrer le théorème suivant.

**Théorème 17.** *Pour toute suite convergente de graphes simples $(G_n)$, il existe un graphon $W$ tel que $t(F,G_n) \to t(F,G)$ pour tout graphe simple $F$.*

Ainsi, on peut associer à toute suite de graphes convergente $(G_n)$ un objet limite qui est un graphon $W$. On dit que $W$ est la *limite* de la suite de graphes $(G_n)$, et on note $G_n \to W$. Ce théorème justifie donc l'introduction des graphons pour construire des objets limites explicites et agréables à manipuler.

Enfin, le théorème suivant montre que la distance de coupe $\delta_\square$ est "la bonne" distance à considérer pour caractériser la convergence des suites de graphes vers leurs graphons limites.

**Théorème 18.** *Pour une suite de graphes $(G_n)$ avec $v(G_n) \to \infty$, on a $G_n \to W$ si et seulement si $\delta_\square(W_{G_n},W) \to 0$.*

### 3.5 Quelques exemples d'applications des graphons

Les graphons peuvent être utilisés pour modéliser de nombreux problèmes de réseaux, nous en citons ici deux exemples.

Un premier exemple d'utilisation des graphons est donné par l'article [BC17], qui montre comment faire de l'estimation non-paramétrique de graphons, avec une application au cas du filtrage collaboratif sur des réseaux sociaux.

Un deuxième exemple d'utilisation des graphons est donné par l'article [DDZ20], présentant un modèle épidémiologique. Ce modèle cherche à estimer la dynamique du nombre de personnes infectées par une maladie parmi une population d'individus dont on souhaite tenir compte des spécificités individuelles, les relations de contact entre les individus étant modélisées par un graphon.

## 4 Domaine de recherche : extension au cas des graphes pondérés

Dans cette section, nous présentons le positionnement du travail de recherche effectué et qui reste à poursuivre. Nous commençons par présenter les graphes pondérés que nous souhaiterions étudier. Puis, nous introduisons une généralisation des graphons comme candidat pour les limites de suites de graphes pondérés : les graphons de lois. Enfin, nous donnons une idée des résultats qui peuvent être obtenus avec les graphons de lois.

## 4.1 Les graphes pondérés aléatoires

Nous souhaitons étudier des graphes pondérés aléatoires. Les graphons permettaient certes de générer des graphes pondérés, mais présentant une source d'aléa relativement faible : les choix des sommets dans $[0,1]$ étaient aléatoires, puis les poids des arêtes étaient donnés par un graphon $W$, ce qui imposait une contrainte importante sur la corrélation entre les poids des différentes arêtes. Nous souhaitons plutôt étudier des graphes pondérés aléatoires dont les poids des arêtes seraient indépendants. Dans le travail effectué lors du mémoire de M2, nous avons considérer uniquement des graphes dont seules les arêtes sont pondérés (les sommets étant donc tous de même poids), c'est donc dans ce cadre que nous plaçons dans la suite de ce document. Une direction futur pour poursuivre ce travail serait donc d'étendre le cadre des résultats au cas des graphes pondérés aléatoires dont les sommets sont également pondérés.

## 4.2 Les graphons de lois

Dans le cas des graphons classiques $W : [0,1]^2 \to [0,1]$, la valeur de $W(x,y)$ représentait la probabilité de garder l'arête entre $x$ et $y$. Donc, si on attribue aux arêtes présentes un poids 1 et aux arêtes absentes un poids 0, on peut considérer tout graphe simple comme un graphe pondéré. L'arête entre $x$ et $y$ peut alors être vu comme une arête pondérée dont le poids est donné par une variable aléatoire de distribution $W(x,y)\delta_1 + (1 - W(x,y))\delta_0$.

Partant de ce constat, on cherche à étendre le cadre des graphons à celui de limite de grands graphes pondérés denses dont les poids des arêtes sont des variables aléatoires indépendantes les unes des autres. On introduit donc des généralisation des graphons, de la forme $W : [0,1]^2 \to \mathrm{Proba}(\mathbb{R})$ que l'on appellera *graphons de lois*. Un graphon de lois $W$ représente donc un noyau de probabilité $W(x,y,\mathrm{d}z)$ qui doit vérifier certaines hypothèses supplémentaires (H) :

— $W$ est *symétrique* : $\forall (x,y) \in [0,1]^2, \quad W(x,y) = W(y,x)$

— $W$ est *bi-mesurable* en $x$ et $y$ : pour tout $A$ mesurable, $W(x,y,A)$ est une fonction mesurable de $x$ et de $y$

— $W$ est un *noyau de probabilité* : $\forall (x,y) \in [0,1]^2, \quad W(x,y,\mathrm{d}z)$ est une loi de probabilité en $z$.

On introduit ensuite l'espace des graphons de lois

$$\mathcal{W} = \left\{ W : [0,1]^2 \to \mathrm{Proba}(\mathbb{R}) \mid W \text{ vérifiant (H)} \right\}.$$

On a de même que dans le cas des graphons classiques, que les graphons de lois step-function représente les graphes pondérés aléatories finis.

## 4.3 Adaptation des densités homomorphiques

Dans le cas des graphons classiques, on avait exploité les nombres d'homomorphismes comme paramètres de graphes pour caractériser la valeur (à isomorphisme près) des graphons. Cependant, dans le cas des graphons classiques les arêtes sont bi-valuées (de poids 0 ou 1 selon la présence ou non de l'arête), mais nos graphons de lois contiennent des lois générales plus complexes à caractériser. Il nous faut donc prendre une classe plus large de paramètres de graphes. En plus de caractériser la valeur de nos graphons de lois, cette famille de paramètres de graphes doit également être de taille dénombrable pour pouvoir bien se comporter vis-à-vis des permutations entre p.s. et $\forall$ (quantificateur universelle).

On choisit donc nos paramètres de graphes sous la forme de *densité homomorphique* :

$$t(F^f, W) = \int_{[0,1]^k} \int_{\mathbb{R}^{E(F)}} \prod_{ij \in E(F)} f_{ij}(z_{ij})\ W(x_i, x_j, \mathrm{d}z_{ij}) \prod_{i \in V(F)} \mathrm{d}x_i \tag{1}$$

où $W \in \mathcal{W}$ est un graphon de lois, et $F^f$ est un graphe $F$ comportant $k$ sommets et dont les arêtes sont décorées par une famille de fonction mesurable $f = (f_{ij})_{ij \in E(F)}$ appelées *fonctions de pondération*.

Pour un graphe pondéré fini $G$ comportant $n$ sommets, on définit la *densité homomorphique* :

$$t(F^f, G) = \frac{1}{n^k} \sum_{x_i \in V(G): 1 \leq i \leq k} \prod_{ij \in E(F)} f_{ij}(e_G(x_i, x_j))\ . \tag{2}$$

Si on voit $G$ comme un graphe pondéré aléatoire dont chaque arête $ij$ est décorée par une masse de Dirac $\delta_{e_G(i,j)}$, alors on a par un calcul immédiat que $t(F^f, G) = t(F^f, W_G)$. On peut adapter de la même manière la *densité homomorphique induite* (ou *injective*) $t_{ind}$.

### 4.4 Généralisation des résultats aux graphons de lois

Nous venons de définir les graphons de lois, et de leurs adapter les densités homomorphiques. L'objectif suivant est donc de généraliser aux graphons de lois les résultats présentés dans la section 3 : les résultats d'approximation des $W$-graphes aléatoires, et les résultats reliant les densités homomorphiques et la convergence de sous-graphes échantillonnés. Ce travail a déjà été grandement entamé lors du mémoire de M2, au cours duquel la majorité des résultats de la section 3 ont pu être adaptés.

Pour poursuivre ce travail, plusieurs directions sont à considérer : adapter les derniers résultats encore manquants de la section 3 ; étendre le cadre déjà développé aux cas des graphes aléatoires dont les arêtes et les sommets sont pondérés aléatoirement ; étudier l'estimation de ces graphons de lois.

## Références


[BC17] Christian Borgs and Jennifer T. Chayes. Graphons : A nonparametric method to model, estimate, and design algorithms for massive networks, 2017.

[DDZ20] Jean-François Delmas, Dylan Dronnier, and Pierre-André Zitt. An infinite-dimensional sis model, 2020.

[Gla16] Daniel Glasscock. What is a graphon ?, 2016.

[Lov12] László Lovász. *Large networks and graph limits*, volume 60. American Mathematical Society, Colloquium Publications, 2012.