

# FONCTIONNELLES DE GRAPHS ALÉATOIRES ÉCHANTILLONNÉS À PARTIR D'UN GRAPHON

## 2.1 Graphes

**Notations préliminaires** Dans tout ce chapitre,  $F$  et  $G$  désignent des graphes finis simples. Pour  $n \in \mathbb{N}^*$ , on note  $[n] := \{1, \dots, n\}$ . On note  $|B|$  le cardinal de l'ensemble  $B$ . On désigne par  $\mathcal{N}(m, \sigma^2)$  une loi normale de moyenne  $m$  et de variance  $\sigma^2$ .

### 2.1.1 Notions sur les graphes

Un graphe simple (non orienté)  $G$  est un couple  $(V(G), E(G))$  formé d'un ensemble  $V(G)$  de  $v(G)$  sommets et d'un sous-ensemble  $E(G)$  d'arêtes choisies dans la collection des  $\binom{v(G)}{2}$  paires non ordonnées de sommets. On note  $e(G)$  le nombre d'arêtes du graphe  $G$ . Rappelons qu'un graphe simple est un graphe sans boucles (un sommet n'est jamais relié à lui même) et sans arêtes multiples (deux sommets sont reliés par au plus une arête). Un graphe est fini si son nombre de sommets (et donc d'arêtes) est fini. Dans ce cas, en numérotant les sommets de 1 à  $v(G)$ , on pourra identifier  $V(G)$  à  $[v(G)]$ . On considérera dans la suite des graphes finis simples (et non orientés). On note  $\mathcal{F}$  l'ensemble des graphes finis simples (et non orientés).

Un graphe fini simple  $G$  peut être caractérisé par son ensemble de sommets  $V(G) = [v(G)]$  et sa matrice d'adjacence  $A$  de dimension  $v(G) \times v(G)$  définie par :

$$A(i, j) = \begin{cases} 1 & \text{si } \{i, j\} \in E(G) \\ 0 & \text{sinon.} \end{cases}$$

Ainsi,  $A$  est une matrice symétrique binaire avec des zéros sur la diagonale. On peut également associer à une matrice d'adjacence d'un graphe  $G$  son image pixelisée (« pixel picture » en anglais) : il s'agit du carré unité que l'on subdivise en petits carrés de longueur  $1/v(G)$  ; les 0 sont remplacés par des carrés blancs et les 1 sont remplacés par des carrés noirs.

Nous verrons que l'image pixelisée peut donner une intuition de la convergence d'une suite de grands graphes aléatoires, voir partie [2.1.3](#).

Le nombre maximal d'arêtes dans un graphe simple  $G$  à  $v(G)$  sommets est  $\binom{v(G)}{2}$ . La densité  $\rho$  d'arêtes du graphe  $G$  est la fraction de ces arêtes qui sont réellement présentes :

$$\rho = \frac{e(G)}{\binom{v(G)}{2}} = \frac{2e(G)}{v(G)(v(G) - 1)}.$$

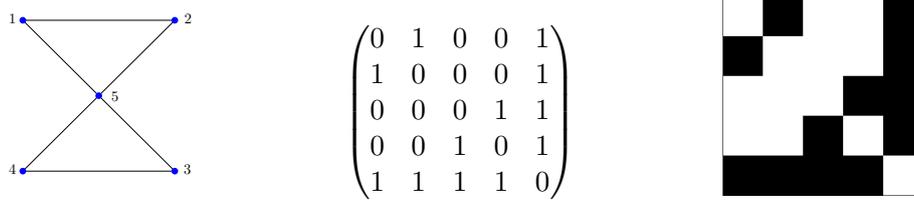


FIGURE 2.1 – Un graphe à 5 sommets, sa matrice d’adjacence et son image pixelisée.

On dira qu’une suite de graphes simples finis  $(G_n : n \in \mathbb{N}^*)$  est dense (resp. creuse) si la suite de densités de ces graphes tend vers une constante strictement positive (resp. vers 0) quand  $n$  tend vers l’infini.

### 2.1.2 Graphes aléatoires

Nous allons maintenant considérer des graphes aléatoires. Les modèles de graphes aléatoires les plus simples sont ceux d’Erdős et Rényi, introduits en 1959 – 1960. Rappelons qu’un graphe d’Erdős-Rényi  $G_n(\mathbf{p})$  à  $n$  sommets et de paramètre  $\mathbf{p}$  avec  $0 < \mathbf{p} < 1$ , est un graphe aléatoire ayant pour ensemble des sommets  $[n] = \{1, \dots, n\}$  et dont les arêtes existent de manière indépendante avec probabilité  $\mathbf{p}$ . A partir des années 2000, d’autres modèles de graphes, plus complexes, ont été développés afin d’obtenir des graphes ayant des propriétés plus proches des réseaux réels. Parmi ces modèles figurent, par exemple, le modèle de configuration (modèle de graphes aléatoires où les degrés des sommets sont fixés) ou encore le modèle d’attachement préférentiel (modèle de graphes aléatoires dynamique, construits de manière récursive), voir Durrett [72] et van der Hofstad [187] pour de récents ouvrages de références.

### 2.1.3 Heuristique de la convergence d’une suite de graphes aléatoires

Dans cette partie, nous allons voir de manière informelle, quelle peut être la limite d’une suite de graphes au travers de deux exemples. Les exemples que nous développerons sont tirés de Borgs, Chayes, Lovász, Sós et Vesztergombi [37]. On considère une suite de graphes finis simples denses  $(G_n : n \in \mathbb{N}^*)$ . Quelle peut-être la limite de cette suite ?

Commençons par le modèle d’Erdős-Rényi. Soit  $(G_n(1/2) : n \in \mathbb{N}^*)$  une suite de graphes où  $G_n(1/2)$  est un graphe d’Erdős-Rényi de paramètre  $1/2$  à  $n$  sommets. Si l’on regarde la suite d’images pixelisées associées aux matrices d’adjacences, on peut remarquer une convergence « graphique » de cette suite d’images vers le carré unité uniformément gris, comme l’illustre l’image 2.2 ci-dessous. L’origine est placée dans le coin en haut à gauche afin d’être en accord avec la convention de numérotation des éléments matriciels.

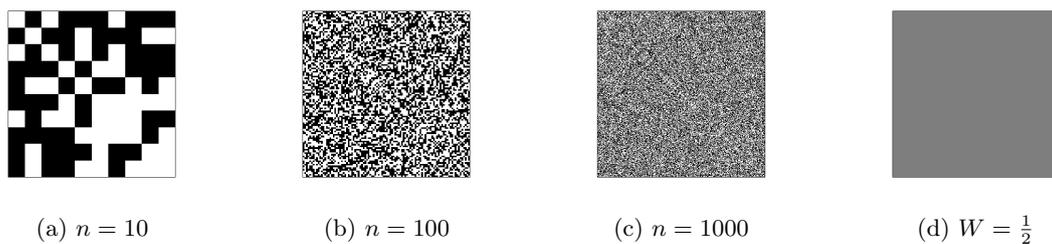


FIGURE 2.2 – Graphes d’Erdős-Rényi de paramètre  $\mathbf{p} = \frac{1}{2}$  et leur limite.

Ainsi, on peut conjecturer que la suite de graphes  $(G_n(1/2) : n \in \mathbb{N}^*)$  « converge », en un sens que nous définirons dans la suite, vers le graphe infini dont la matrice d'adjacence correspond à la fonction constante égale à  $1/2$  sur  $[0, 1]^2$ .

Intéressons nous à un deuxième exemple. Considérons le modèle de graphes aléatoires d'attachement uniforme croissant que l'on notera  $(GUA_n : n \in \mathbb{N}^*)$ . Le graphe  $GUA_n$  ayant pour ensemble de sommets  $[n]$  se construit de manière itérative. On commence par un seul sommet. À chaque étape un nouveau sommet est créé et chaque paire de sommets non encore connectés à l'étape précédente est reliée avec probabilité  $1/k$  où  $k$  est le nombre de sommets à l'étape actuelle (voir figure 2.3). On remarque encore une fois une convergence graphique des images pixelisées quand le nombre de sommets augmente, vers la fonction de deux variables  $W$  définie par  $W(x, y) = 1 - \max(x, y)$  pour tout  $x, y \in [0, 1]$ .

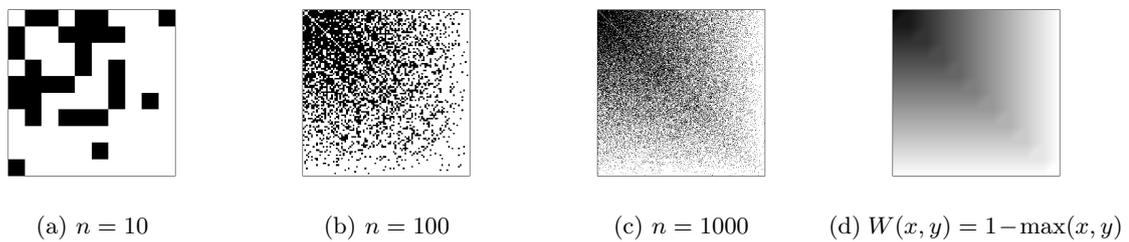


FIGURE 2.3 – Graphes aléatoires d'attachement uniforme croissant et leur limite

## 2.2 Des graphes aux graphons

### 2.2.1 Densités d'homomorphismes

Nous allons définir la notion d'homomorphismes qui permettra de donner une caractérisation de la convergence d'une suite de graphes denses. De nombreux paramètres de graphes peuvent s'exprimer à partir du nombre d'homomorphismes, voir Freedman, Lovasz et Schrijver [95] pour des exemples d'applications.

Soient  $F, G \in \mathcal{F}$  deux graphes simples tels que  $p = v(F)$  et  $n = v(G)$ . Dans la suite, il faut voir  $F$  comme un petit graphe et  $G$  comme un grand graphe. Un homomorphisme  $\varphi$  de  $F$  dans  $G$  est une application de  $V(F) = [p]$  dans  $V(G) = [n]$  qui préserve l'adjacence des arêtes. Autrement dit, c'est une application de  $V(F)$  dans  $V(G)$  telle que si  $\{i, j\} \in E(F)$  alors  $\{\varphi(i), \varphi(j)\} \in E(G)$ . On note  $\text{Hom}(F, G)$  l'ensemble des homomorphismes de  $F$  dans  $G$ . Par exemple, si  $F$  est un triangle alors  $|\text{Hom}(F, G)|$  est le nombre de triangles dans  $G$  multiplié par 6 (permutations des sommets). On peut ainsi définir la densité d'homomorphismes de  $F$  dans  $G$  comme la quantité normalisée :

$$t(F, G) = \frac{|\text{Hom}(F, G)|}{n^p},$$

qui est la probabilité qu'une application de  $V(F)$  dans  $V(G)$  choisie au hasard soit un homomorphisme.

De manière similaire, on note  $\text{Inj}(F, G)$  l'ensemble de homomorphismes injectifs et on lui associe sa densité :

$$t_{\text{inj}}(F, G) = \frac{|\text{Inj}(F, G)|}{A_n^p},$$

où pour tout  $n \geq k \geq 1$ ,  $A_n^k = n!/(n-k)!$ . Les notations utilisées suivent celles de Borgs, Chayes, Lovász, Sós et Vesztergombi [40] (dans Bollobás et Riordan [32],  $t_{\text{inj}}$  est notée  $t$ ).

Enfin, on définit l'ensemble des homomorphismes induits de  $F$  dans  $G$ , noté  $\text{Ind}(F, G)$ . Rappelons qu'un homomorphisme induit de  $F$  dans  $G$  est un homomorphisme injectif qui préserve également les relations de non-adjacence entre les arêtes. Autrement dit, c'est une application injective  $\varphi$  de  $V(F)$  dans  $V(G)$  telle que  $\{i, j\} \in E(F)$  si et seulement si  $\{\varphi(i), \varphi(j)\} \in E(G)$ . On définit la densité d'homomorphismes induits par :

$$t_{\text{ind}}(F, G) = \frac{|\text{Ind}(F, G)|}{A_n^p}.$$

Il est facile de voir que les différentes notions d'homomorphismes sont reliées, voir par exemple Lovász et Szegedy [139], paragraphe 2.4. Pour  $F, G \in \mathcal{F}$  deux graphes finis simples, on a :

$$|t_{\text{inj}}(F, G) - t(F, G)| \leq \frac{1}{n} \binom{p}{2}, \quad (2.1)$$

ainsi que

$$t_{\text{inj}}(F, G) = \sum_{F' \geq F} t_{\text{ind}}(F', G) \quad \text{et} \quad t_{\text{ind}}(F, G) = \sum_{F' \geq F} (-1)^{e(F')-e(F)} t_{\text{inj}}(F', G), \quad (2.2)$$

où  $F' \geq F$  signifie que  $V(F) = V(F')$  et  $E(F) \subset E(F')$ , i.e.  $F'$  parcourt l'ensemble des graphes simples obtenus à partir de  $F$  en ajoutant des arêtes.

**Comptes de sous-graphes** De nombreux auteurs étudient les comptes de sous-graphes plutôt que les densités d'homomorphismes. Deux graphes  $F_1$  et  $F_2$  sont isomorphes (ou  $F_2$  est une copie de  $F_1$ ) s'il existe un homomorphisme induit bijectif  $\varphi$  de  $V(F_1)$  dans  $V(F_2)$ . L'application  $\varphi$  est aussi appelée un isomorphisme. Pour deux graphes  $F$  et  $G$ , on définit le nombre de copies de  $F$  dans  $G$ , i.e. le nombre de sous-graphes de  $G$  qui sont isomorphiques à  $F$  :

$$\chi_F(G) = |\{F' \subset G : F' \text{ est une copie de } F\}|, \quad (2.3)$$

où  $F' \subset G$  signifie que  $V(F) \subset V(G)$  et  $E(F) \subset E(G)$ . Un automorphisme d'un graphe  $F$  est un isomorphisme de graphe de  $F$  dans lui-même. On note  $|\text{Aut}(F)| = \chi_F(F)$  son cardinal. On note  $K_n$  le graphe complet à  $n$  sommets. On a les relations suivantes entre les comptes de sous-graphes, les nombres d'homomorphismes injectifs et les densités d'homomorphismes, voir Bollobás et Riordan [32], Section 2.1 : soient  $F, G \in \mathcal{F}$  deux graphes simples ayant  $p$  et  $n$  sommets respectivement et tels que  $n \geq p \geq 1$ . On a les égalités suivantes :

$$\chi_F(K_n) = \frac{A_n^p}{|\text{Aut}(F)|} \quad \text{et} \quad \chi_F(G) = \frac{A_n^p}{|\text{Aut}(F)|} t_{\text{inj}}(F, G).$$

### 2.2.2 Graphons

Dans cette partie, nous allons introduire la notation de graphons défini par Lovász et Szegedy [139] en 2006 qui apparaîtra dans la suite, comme l'objet limite adapté pour décrire les limites de suites de graphes denses. Nous étendrons également les définitions d'homomorphismes de graphes aux graphons.

On appelle graphon toute fonction  $W$  mesurable symétrique de  $[0, 1]^2$  dans  $[0, 1]$ . On note  $\mathcal{W}$  l'espace des graphons.

**Construction d'un graphon à partir d'un graphe :** Pour tout graphe simple  $G$  à  $n$  nœuds, il existe une manière naturelle de construire un graphon  $W_G$  associé au graphe  $G$ , voir Borgs, Chayes, Lovász, Sós et Vesztergombi [40], paragraphe 3.1. Pour ce faire, on divise l'intervalle  $[0, 1]$  en  $n$  intervalles  $(J_i : 1 \leq i \leq n)$  de longueur  $1/n$  où pour tout  $i \in \{1, \dots, n\}$ ,  $J_i = [\frac{i-1}{n}, \frac{i}{n})$  et pour tout  $x \in J_i, y \in J_j$ , on définit  $W_G$  par :

$$W_G(x, y) = \begin{cases} 1 & \text{si } \{i, j\} \in E(G) \\ 0 & \text{sinon.} \end{cases}$$

Ainsi, le graphe de  $W_G$  est l'image pixelisée du graphe  $G$ . La fonction  $W_G$  est aussi appelée graphon empirique (« empirical graphon » en anglais).

**Densités d'homomorphismes associées aux graphons :** La correspondance entre les graphes simples et les graphons suggère qu'il est possible d'étendre de manière naturelle les densités d'homomorphismes de graphes aux graphons, voir Borgs, Chayes et Lovász [36]. Pour tout graphe simple  $F$  et pour tout graphon  $W \in \mathcal{W}$ , on définit

$$t(F, W) = t_{\text{inj}}(F, W) = \int_{[0,1]^{V(F)}} \prod_{\{i,j\} \in E(F)} W(x_i, x_j) \prod_{k \in V(F)} dx_k \quad (2.4)$$

et

$$t_{\text{ind}}(F, W) = \int_{[0,1]^{V(F)}} \prod_{\{i,j\} \in E(F)} W(x_i, x_j) \prod_{\{i,j\} \notin E(F)} (1 - W(x_i, x_j)) \prod_{k \in V(F)} dx_k. \quad (2.5)$$

Ces formules peuvent être interprétées comme le « nombre » normalisé d'homomorphismes injectifs (resp. homomorphismes induits) de  $F$  dans un graphe ayant pour ensemble de sommets  $[0, 1]$  et de poids entre les arêtes donnés par  $W$ .

Pour tout graphe simple  $F \in \mathcal{F}$  et tout graphe simple  $G$ , on a :

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

Les densités d'homomorphismes injectifs  $t_{\text{inj}}$  coïncident dans le cadre des graphons avec les densités d'homomorphismes  $t$  puisque l'on a  $t_{\text{inj}}(F, W) = t(F, W)$ , pour tout  $W \in \mathcal{W}$  et  $F \in \mathcal{F}$ . En effet, intuitivement, les tirages de sommets avec ou sans remise coïncident quand le nombre de sommets est infini.

Pour terminer cette partie, nous introduisons la fonction de degré normalisée  $D$  d'un graphon  $W$  qui interviendra dans nos résultats du chapitre 4. Elle est définie pour tout  $x \in [0, 1]$  par :

$$D(x) = \int_0^1 W(x, y) dy.$$

En notant  $K_2$  le sous-graphe complet à deux sommets (autrement dit, une arête), on remarque que l'on a :

$$t_{\text{inj}}(K_2, W) = \int_0^1 D(x) dx.$$

### 2.2.3 Convergence d'une suite de graphes denses

Dans cette section, nous avons choisi d'introduire deux notions de convergence de suite de graphes denses : la convergence associée aux densités d'homomorphismes (celle qui sera l'objet de notre étude) et la convergence métrique pour la distance de coupe. D'autres convergences équivalentes aux deux précédentes ont également été établies, voir Borgs, Chayes et Gamarnik [34] et Borgs, Chayes, Lovász, Sós et Vesztergombi [41].

Notons également que la théorie de limites de graphes est aussi intimement reliée à la théorie des tableaux partiellement échangeables de variables aléatoires étudiée par Aldous [12] et Hoover [110]. Le lecteur pourra se référer à Diaconis et Janson [61] et Lovász [138], section 11.3.3 pour plus de détails.

### Densités d'homomorphismes et convergence

Dans ce paragraphe, nous allons définir la première notion de convergence de suites de graphes denses introduite par Lovász et Szegedy [139] en 2006. Elle fait intervenir les densités d'homomorphismes définies dans la section 2.2.1. Cette convergence est également appelée convergence à gauche, voir Lovász [138], chapitre 11.

Lovász et Szegedy [139] ont énoncé la définition suivante :

**Définition.** *Une suite de graphes finis simples  $(G_n : n \in \mathbb{N}^*)$  est dite convergente si la suite  $(t_{\text{inj}}(F, G_n) : n \in \mathbb{N}^*)$  admet une limite pour tout  $F \in \mathcal{F}$ .*

Par les formules qui relient  $t$ ,  $t_{\text{inj}}$  et  $t_{\text{ind}}$ , c'est équivalent à demander que la suite  $(t(F, G_n) : n \in \mathbb{N}^*)$  converge pour tout  $F \in \mathcal{F}$  ou que la suite  $(t_{\text{ind}}(F, G_n) : n \in \mathbb{N}^*)$  converge pour tout  $F \in \mathcal{F}$ .

Le résultat principal de Lovász et Szegedy [139] est le suivant :

**Theorem** (Lovász et Szegedy (2006)). *Pour toute suite de graphes convergente  $(G_n : n \in \mathbb{N}^*)$ , il existe un graphon  $W \in \mathcal{W}$  tel que*

$$\lim_{n \rightarrow +\infty} t(F, G_n) = t(F, W), \quad \forall F \in \mathcal{F}. \quad (2.6)$$

La preuve originale de Lovász et Szegedy de ce théorème utilise le lemme de régularité de Szemerédi [182] et le théorème de convergence des martingales. Deux autres preuves ont également été établies par Elek et Szegedy [73] en utilisant la théorie des ultra produits et ultra limites et par Diaconis et Janson [61] grâce à la théorie des variables échangeables.

Borgs, Chayes et Lovász [36] ont montré l'unicité de la limite  $W$  à une bijection préservant la mesure près. Autrement dit, pour toute autre fonction limite  $W' \in \mathcal{W}$ , il existe une application bijective  $\varphi : [0, 1] \rightarrow [0, 1]$  telle que  $\varphi$  et son inverse préservent la mesure et telle que  $W(x, y) = W'(\varphi(x), \varphi(y))$  pour presque tout  $x, y \in [0, 1]$ .

On dira qu'une suite de graphes  $(G_n : n \in \mathbb{N}^*)$  converge vers un graphon  $W \in \mathcal{W}$  si (2.6) a lieu.

Lovász et Szegedy ont également établi la réciproque du théorème précédent :

**Proposition** (Lovász et Szegedy (2006)). *Pour tout graphon  $W \in \mathcal{W}$ , il existe une suite de graphes  $(G_n : n \in \mathbb{N}^*)$  telle que*

$$\lim_{n \rightarrow +\infty} t(F, G_n) = t(F, W), \quad \forall F \in \mathcal{F}.$$

Nous verrons dans la sous-section 2.3 un exemple de construction d'une telle suite de graphes  $(G_n : n \in \mathbb{N}^*)$ .

### Distance de coupe et convergence

La deuxième notion de convergence proposée par Borgs, Chayes, Lovász, Sós et Vesztergombi [38, 40] est une convergence métrique définie à partir de la distance de coupe. La

distance de coupe est basée sur une norme de coupe notée  $\|\cdot\|_{\square}$  introduite par Frieze et Kannan [96] en 1999. Pour  $A$  une matrice réelle carrée de taille  $n \times n$  avec  $n \in \mathbb{N}^*$ , on définit sa norme de coupe par :

$$\|A\|_{\square} = \frac{1}{n^2} \max_{S, T \subseteq [n]} \left| \sum_{i \in S, j \in T} A(i, j) \right|.$$

Il est naturel d'étendre cette norme aux fonctions mesurables symétriques et bornées de  $[0, 1]^2$  dans  $\mathbb{R}$ . On définit la norme de coupe notée  $\|\cdot\|_{\square}$ , pour toute fonction mesurable symétrique et bornée  $U$  de  $[0, 1]^2$  dans  $[0, 1]$  par :

$$\|U\|_{\square} = \sup_{S, T \subseteq [0, 1]} \left| \int_{S \times T} U(x, y) dx dy \right|,$$

où le supremum est pris sur l'ensemble des sous-ensembles mesurables  $S$  et  $T$  de  $[0, 1]$ .

La distance de coupe sur l'espace des graphons, notée  $\delta_{\square}$ , est définie pour tous graphons  $W_1, W_2 \in \mathcal{W}$  par :

$$\delta_{\square}(W_1, W_2) = \inf_{\phi: [0, 1] \rightarrow [0, 1]} \|W_1 - W_2^{\phi}\|_{\square},$$

où l'infimum est pris sur toutes les fonctions bijectives  $\phi : [0, 1] \rightarrow [0, 1]$  telle que  $\phi$  et son inverse préservent la mesure et où  $W_2^{\phi}$  est défini par  $W_2^{\phi}(x, y) = W_2(\phi(x), \phi(y))$ , pour tout  $x, y \in [0, 1]$ . La distance  $\delta_{\square}$  est en fait une pseudo-distance. En effet, deux graphons différents peuvent être à distance nulle pour la distance de coupe. En identifiant tous les noyaux dont la distance de coupe est nulle, on obtient l'espace quotient noté  $\tilde{\mathcal{W}}$  qui est l'ensemble des graphons non étiquetés.

Nous énonçons maintenant le théorème établi par Borgs, Chayes, Lovász, Sós et Vesztergombi [38] en 2006 qui donne l'équivalence entre la convergence métrique et la convergence au sens des densités d'homomorphismes, justifiant ainsi l'introduction de la distance de coupe : une suite de graphes simples finis ( $G_n : n \in \mathbb{N}^*$ ) est convergente si et seulement si elle est de Cauchy pour la distance de coupe  $\delta_{\square}$ . De plus, la suite de graphes simples finis ( $G_n : n \in \mathbb{N}^*$ ) converge vers  $W \in \mathcal{W}$  si et seulement si  $\lim_{n \rightarrow +\infty} \delta_{\square}(W_{G_n}, W) = 0$ .

Nous terminons ce paragraphe sur une propriété de l'espace  $\tilde{\mathcal{W}}$  muni de la distance de coupe. L'espace  $(\tilde{\mathcal{W}}, \delta_{\square})$  est compact, voir Lovász et Szegedy [139] ou Lovász [138], partie 9.3. En particulier,  $(\tilde{\mathcal{W}}, \delta_{\square})$  est un espace métrique complet. En combinant ce fait et le résultat précédent sur l'équivalence des convergences, on en déduit que l'espace des graphons est l'espace complété des graphes finis avec la distance de coupe.

## 2.3 Échantillonnage de graphes aléatoires à partir d'un graphon

Soit  $W \in \mathcal{W}$ . On définit un modèle de graphes aléatoires générés à partir du graphon  $W$ . On génère un graphe aléatoire  $G_n(W)$  à  $n$  sommets avec  $n \in \mathbb{N}^*$  à partir de  $W$  de la manière suivante : on échantillonne une suite de variables aléatoires  $X = (X_i : i \in \mathbb{N}^*)$  indépendantes de loi uniforme sur  $[0, 1]$  puis, étant donnée cette suite, on met une arête entre les sommets  $i$  et  $j$  avec  $i, j \in [n]$  avec probabilité  $W(X_i, X_j)$ . Pour une suite donnée  $X$ , on répète l'opération de manière indépendante pour chacune des paires  $(i, j) \in [n]^2$  avec  $i < j$ . Le processus ainsi décrit repose donc sur deux niveaux d'aléa.

Remarquons que les graphes aléatoires  $G_n(W)$  sont une généralisation des graphes aléatoires d'Erdős-Rényi  $G_n(\mathbf{p})$  obtenus en prenant le graphon constant égale à  $\mathbf{p}$  avec  $0 < \mathbf{p} < 1$ .

Il a été prouvé par Lovász et Szegedy [139] en 2006, que la suite de graphes aléatoires  $(G_n(W) : n \in \mathbb{N}^*)$  converge presque sûrement vers le graphon choisi  $W$ . Autrement dit :

**Proposition** (Lovász et Szegedy (2006)). *Pour tout graphe simple  $F \in \mathcal{F}$ , on a presque sûrement :*

$$\lim_{n \rightarrow +\infty} t_{\text{inj}}(F, G_n(W)) = t(F, W). \quad (2.7)$$

La preuve repose sur l'égalité  $\mathbb{E}[t_{\text{inj}}(F, G_n(W))] = t(F, W)$ , un lemme de concentration sur les graphons (conséquence directe de l'inégalité d'Azuma) et le lemme de Borel-Cantelli, voir Lovász [138], partie 11.4.1.

Dans le cas d'Erdős-Rényi, i.e. quand le graphon  $W$  est constant, on retrouve la conjecture graphique énoncée dans la sous-section 2.1.3 : la suite de graphes d'Erdős-Rényi de paramètre  $1/2$  (resp.  $\mathfrak{p}$  avec  $0 < \mathfrak{p} < 1$ ) converge vers le graphon constant  $W \equiv 1/2$  (resp.  $W \equiv \mathfrak{p}$ ).

Nous allons maintenant nous intéresser aux fluctuations associées à ce principe d'invariance. Nous étudierons dans le chapitre 4, le cas où  $W$  est non constant, mais historiquement les fluctuations ont d'abord été montrées dans le cas où  $W$  est constant (autrement dit, pour le modèle d'Erdős-Rényi). C'est ce dernier cas que nous allons commencer par développer.

### Cas particulier du graphon constant

Au tournant des années 80 et 90, Nowicki [159] et Janson et Nowicki [118] ont été les premiers à montrer que la vitesse de convergence associée à la convergence presque sûre est d'ordre  $n$ . Leurs preuves s'appuyaient sur la théorie des U-statistiques.

**Proposition** (Nowicki (1989), Janson et Nowicki (1991)). *Pour tout  $F \in \mathcal{F}$  ayant  $e$  arêtes, on a la convergence en loi suivante :*

$$n(t_{\text{inj}}(F, G_n(\mathfrak{p})) - \mathfrak{p}^e) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}(0, 2e^2 \mathfrak{p}^{2e-1} (1 - \mathfrak{p})).$$

Janson et Nowicki ont aussi obtenu la convergence des densités d'homomorphismes induits ainsi que la normalité asymptotique de vecteurs de comptes de sous-graphes et de comptes de sous-graphes induits. Dans le cas particulier de la loi jointe des comptes des arêtes, triangles et des graphes étoiles à 2 feuilles (« two-stars » en anglais), Reinert et Röllin [167], proposition 2, ont obtenu des bornes sur l'approximation. Par des méthodes de calculs de Malliavin, Krokowski et Thäle [132] ont généralisé le résultat de Reinert et Röllin (dans un espace de probabilité différent) et ont obtenu la vitesse de convergence associée au théorème central limite multidimensionnel de Janson et Nowicki [118]. Féray, Méliot et Nikeghbali [83] se sont également intéressés aux densités d'homomorphismes en lien avec la convergence « mod-gaussian ».

Chatterjee et Varadhan [46] ont quant à eux, développé une théorie de grandes déviations pour les graphes d'Erdős-Rényi en lien avec les graphons, voir aussi Chatterjee [44].

L'étude asymptotique des comptes normalisés de sous-graphes a également été faite dans le cas où le paramètre  $\mathfrak{p}$  des graphes d'Erdős-Rényi dépend de  $n$ . Grâce à la méthode des moments, Ruciński [174] a déterminé toutes les valeurs de  $\mathfrak{p}$  telles que les comptes normalisés de sous-graphes convergent en loi vers une loi normale. Nowicki et Wierman [161] ont appliqué la théorie des U-statistiques pour obtenir des résultats similaires mais dans un cadre plus faible. En utilisant la méthode de Stein, Barbour, Karoński et Ruciński [18] ont complété ces résultats en donnant des bornes sur l'erreur dans le théorème central limite. Dans le cas particulier du nombre de triangles, Gilmer et Kopparty [100] ont prouvé un théorème central limite local.

### Cas où $W$ n'est pas constant

Dans le cas où  $W$  est non constant, Féray, Méliot et Nikeghbali [84], théorème 21, ont établi fin 2017, alors que nous travaillions déjà sur ce sujet, que la vitesse de convergence associée au principe d'invariance est d'ordre  $\sqrt{n}$ .

**Proposition** (Féray, Méliot et Nikeghbali (2017)). *Pour tout  $F \in \mathcal{F}$ , on a la convergence en loi suivante :*

$$\sqrt{n} (t_{\text{inj}}(F, G_n) - t(F, W)) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}(0, \sigma(F)^2),$$

où

$$\sigma(F)^2 = \sum_{q, q' \in V(F)} t((F \bowtie F)(q, q'), W) - v(F)^2 t(F, W)^2$$

et  $(F \bowtie F')(q, q')$  est l'union disjointe de deux graphes simples finis  $F$  and  $F'$  où l'on a identifié les sommets  $q \in F$  et  $q' \in F'$  (voir figure 2.4).

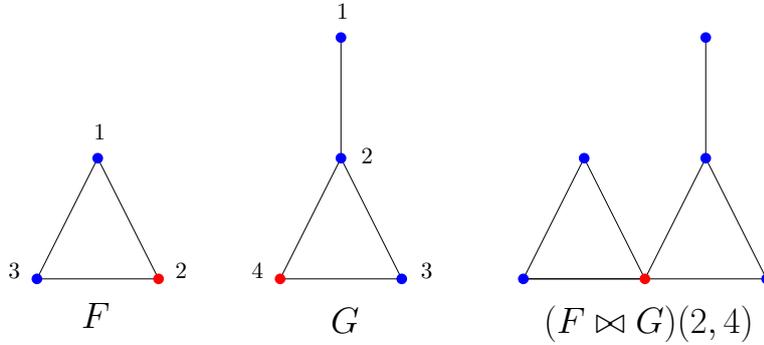


FIGURE 2.4 – Exemple de deux graphes connectés par deux sommets.

*Remarque.* Quand le graphon  $W \equiv \mathbf{p}$  est constant, il est facile de voir que la variance du théorème central limite énoncé par Féray, Méliot et Nikeghbali [84] est nulle, confirmant ainsi que la vitesse en  $\sqrt{n}$  n'est plus pertinente dans ce cas.

### Autres résultats de convergence

En utilisant la méthode de Stein, Fang et Röllin [81] ont obtenu la vitesse de convergence du théorème central limite multidimensionnel associé à la loi jointe des comptes normalisés d'arêtes et de cycles de longueur 4. Cela leur a notamment permis d'établir un intervalle de confiance pour tester si un graphe donné  $G$  provient du modèle d'Erdős-Rényi ou bien d'un modèle de  $W$ -graphes aléatoires avec un  $W$  non constant. Maugis, Priebe, Olhede et Wolfe [143] ont donné un théorème central limite pour les comptes de sous-graphes observés parmi une collection finie de  $W$ -graphes aléatoires générés à partir d'un même graphon  $W$  ou de plusieurs graphons et quand le nombre d'observations de l'échantillon augmente mais que le nombre de sommets de chaque graphe observé reste fini. Ceci leur permet de tester si les graphes observés proviennent d'un graphon particulier.

On peut également s'intéresser à des suites de graphons qui tendent vers 0. Cela mène à des approximations de Poisson des comptes de sous-graphes. Ainsi, Coulson, Gaunt et Reinert [51], corollaire 4.1, ont utilisé la méthode de Stein pour établir des approximations de Poisson pour la loi du nombre de sous-graphes qui sont isomorphiques à un graphe strictement équilibré donné.

## 2.4 Résultats du Chapitre 4

Dans cette partie, nous allons synthétiser les résultats du Chapitre 4 qui contient l'article [53] reporté sans modifications.

### Contexte

Soit  $W \in \mathcal{W}$  un graphon et  $(G_n(W) : n \in \mathbb{N}^*)$  la suite de  $W$ -graphes aléatoires associée à  $W$ , voir partie 2.3.

### Espaces fonctionnels

Soit  $I = [0, 1]$ . On note  $\mathcal{B}(I)$  (resp.  $\mathcal{B}^+(I)$ ) l'espace des fonctions mesurables de  $I$  dans  $\mathbb{R}$  (resp. dans  $\mathbb{R}^+$ ). On note  $\mathcal{C}(I)$  l'espace des fonctions continues de  $I$  dans  $\mathbb{R}$ . Pour  $f \in \mathcal{B}(I)$ , on note  $\|f\|_\infty$  la norme infinie de  $f$  sur  $I$ . On note  $\mathcal{C}^k(I)$  l'ensemble des fonctions de  $I$  dans  $\mathbb{R}$  qui ont leur  $k$ -ième dérivée continue sur  $I$ .

Nous allons dans un premier temps, nous intéresser au comportement asymptotique de la fonction de répartition empirique (FDR) des degrés associée à cette suite de graphes. Enfin, nous considérerons les asymptotiques des densités d'homomorphismes de la suite  $(G_n(W) : n \in \mathbb{N}^*)$ .

#### 2.4.1 Fonction de répartition empirique des degrés

La littérature concernant les propriétés des graphes et en particulier des degrés de leurs sommets est très importante, voir par exemple Newman, Barabási et Watts [155], chapitre 3, Newman [158], chapitre 3 ou Bollobás [29]. Pour étudier la structure de grands graphes aléatoires, de nombreuses caractéristiques peuvent être analysées : nombre et taille des composantes, points isolés, degré d'un sommet, distance entre les sommets, étude des petits graphes (par exemple des triangles, arbres, cliques, ...) ou encore des notions autour de la centralité (« betweenness centrality » en anglais).

La suite des degrés (ou encore loi des degrés) est une des propriétés fondamentales des graphes, c'est celle qui nous intéressera dans la suite. On rappelle que le degré d'un sommet est égal au nombre d'arêtes qui lui sont rattachées.

### Étude des degrés, graphes et graphons

Commençons par rappeler la notion de loi des degrés d'un graphe. Soit  $G \in \mathcal{F}$ . On définit  $p_k$  comme la fraction de sommets du graphe  $G$  qui sont de degrés  $k$ , pour tout  $k \in \llbracket 0, v(G) - 1 \rrbracket$ . Les quantités  $p_k$  représentent la loi des degrés du graphe  $G$ . La valeur  $(p_k : k \in \llbracket 0, v(G) - 1 \rrbracket)$  peut également être vue comme la probabilité qu'un sommet choisi au hasard dans le graphe soit de degré  $k$ . Une autre notion qui contient essentiellement la même information que la loi des degrés est la suite des degrés d'un graphe. Insistons sur le fait que la loi des degrés (ou la suite des degrés) ne donne pas, en règle générale, la structure complète d'un graphe. On rappelle également que les degrés des réseaux réels (comme internet ou les réseaux sociaux) suivent une loi de puissance (« power-law » en anglais) : la plupart des sommets ont des petits degrés mais certains sommets ont des très grands degrés, voir Bollobás [29] ou Newman [154]. Ce n'est pas le cas des premiers modèles de graphes aléatoires, comme les graphes d'Erdős-Rényi dont la loi du degré de chaque sommet est une loi binomiale. Le modèle d'Erdős-Rényi apparaît donc inadapté pour modéliser les réseaux réels.

Pour cette raison entre autres, d'autres modèles de graphes, plus sophistiqués, ont été créés afin d'obtenir non plus des graphes où les degrés sont proches de lois binomiales mais de lois arbitraires données. En 1980, Bollobás [28] a été le premier à construire de tels graphes,

voir aussi Molloy et Reed [147], Newman, Strogatz and Watts [156, 157] ainsi que les ouvrages plus généraux de Bollobás [29], Durrett [72] (chapitre 3) et van der Hofstad [187]. Plusieurs algorithmes existent pour créer de tels graphes. Citons par exemple le modèle de configuration introduit en 1995 par Molloy et Reed [147] qui consiste à apparier des paires de sommets de manière uniforme, les algorithmes basés sur les listes d'adjacence et le théorème d'Havel-Hakami, voir Tinhofer [184], les algorithmes par méthode de Monte-Carlo par chaînes de Markov ou encore les algorithmes séquentiels, voir Blitzstein et Diaconis [26].

Afin d'obtenir des modèles encore plus proches des réseaux réels, on peut également construire des modèles statistiques plus généraux de graphes vérifiant non plus seulement une propriété (comme le nombre de sommets ou la loi des degrés) mais un ensemble de propriétés. C'est le cas du modèle de graphes exponentiels (modèle statistique qui donne une plus grande probabilité aux graphes qui correspondent le mieux aux caractéristiques observées) introduit par Holland et Leinhardt [107] en 1981, voir aussi Strauss [180], Frank et Strauss [94], Anderson, Wasserman et Crouch [14], Park et Newman [162] ou encore Robins, Pattison, Kalish et Lusher [170]. Il est également possible de construire des graphes par partitionnement des données (« clustering » en anglais), où l'ensemble des sommets des graphes est divisé en un certain nombre de blocs ou classes. Un des plus populaires est le modèle à blocs stochastiques (« stochastic bloc model » (SBM) en anglais) introduit en 1983 par Holland, Laskey et Leinhardt [106] et développé ensuite par Nowicki and Snijders [160]. Il décrit la probabilité d'apparition d'une arête entre deux nœuds uniquement en fonction des classes auxquelles ils appartiennent. Ce modèle est adapté pour analyser des petits graphes mais ne permet pas d'analyser les propriétés fines de très grands graphes. Des modèles non paramétriques de graphes aléatoires ont été développés comme le modèle de graphes aléatoires générés par un graphon. Ce dernier peut d'ailleurs être vu comme une généralisation du modèle à blocs stochastiques en prenant un graphon constant par morceaux. Ce lien est utilisé de manière déterminante pour l'estimation non-paramétrique des graphons.

L'estimation des graphons, consiste à retrouver, à partir de l'observation d'un ou plusieurs graphes, le modèle de graphon à partir duquel est engendré le ou les graphes. La littérature autour de la modélisation non-paramétrique des graphons est très riche, voir par exemple Airoidi, Costa et Chan [5], Wolfe et Olhede [194], Borgs, Chayes et Smith [39], Gao, Lu et Zhou [97], Latouche et Robin [133], Klopp, Tsybakov et Verzelen [124] ou Klopp et Verzelen [125].

Par ailleurs, des études paramétriques ont été développées en se basant sur la loi des degrés observée. Ainsi, Chatterjee, Diaconis et Sly [45] ont obtenu la limite de graphes aléatoires ayant une suite de degrés donnée vers un graphon dont la fonction degré est la limite de la suite de degrés. Ils ont utilisé le modèle de graphes exponentiels introduit par Holland et Leinhardt [107] en 1981. Bickel, Chen et Levina [25] ont montré, entre autres, que la FDR des degrés de graphes aléatoires échantillonnés à partir d'un graphon  $W$  converge en loi vers la fonction degré du graphon  $W$ .

### Comportement asymptotique de la FDR empirique des degrés

Nous allons maintenant énoncer nos résultats concernant l'étude asymptotique de la FDR empirique des degrés d'une suite de graphes échantillonnés à partir d'un graphon. On rappelle que la fonction degré  $D = (D(x) : x \in [0, 1])$  du graphon  $W$  est définie par :

$$D(x) = \int_0^1 W(x, y) dy.$$

*Remarque.* Comme un graphon est défini à une bijection préservant la mesure près, il existe toujours une version équivalente au graphon pour lequel la fonction degré est croissante. Si la fonction degré est croissante alors cette version est unique dans  $L^1$ , voir Bickel et Chen [24]. C'est cette version dite « canonique » que nous considérerons dans la suite.

On considère la FDR empirique  $\Pi_n = (\Pi_n(y) : y \in [0, 1])$  de la suite des degrés normalisée du graphe  $G_n(W)$  définie par

$$\Pi_n(y) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{D_i^{(n)} \leq D(y)\}},$$

où  $nD_i^{(n)}$  est le degré du sommet  $i$  dans  $G_n(W)$ .

*Remarque.* Lorsque  $D$  est strictement croissante,  $D^{-1}$  est bien définie et pour tout  $y \in [0, 1]$ , on a :

$$\Pi_n(D^{-1}(y)) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{D_i^{(n)} \leq y\}}.$$

On note  $\text{Id}$  l'application identité sur  $[0, 1]$ . À l'image du théorème de Glivenko-Cantelli qui fournit la convergence presque sûre uniforme de la suite des fonctions de répartition empiriques d'un échantillon de variables aléatoires indépendantes et identiquement distribuées, nous obtenons presque sûrement, la convergence uniforme de la FDR empirique des degrés vers la fonction identité sur  $[0, 1]$ .

**Théorème A.** *On suppose que  $D$  est strictement croissante sur  $[0, 1]$ . Alors on a la convergence presque sûre suivante :*

$$\|\Pi_n - \text{Id}\|_\infty \xrightarrow[n \rightarrow +\infty]{p.s.} 0.$$

Dans l'esprit du théorème de Kolmogorov-Smirnov qui fournit, sous une hypothèse supplémentaire, la vitesse de convergence associée au théorème de Glivenko-Cantelli, nous obtenons les fluctuations associées au théorème A en ajoutant des conditions sur  $W$  et  $D$  :

$$W \in \mathcal{C}^3([0, 1]^2), D' > 0, W < 1 \text{ et } D > 0. \quad (2.8)$$

**Théorème B.** *On suppose que  $W$  satisfait la condition (2.8). Alors on a la convergence des lois finies dimensionnelles suivante :*

$$(\sqrt{n}(\Pi_n(y) - y) : y \in (0, 1)) \xrightarrow[n \rightarrow +\infty]{(lfd)} \chi,$$

où  $\chi = (\chi_y : y \in (0, 1))$  est un processus gaussien centré défini, pour tout  $y \in (0, 1)$  par :

$$\chi_y = \int_0^1 (\rho(y, u) - \bar{\rho}(y)) dB_u,$$

avec  $B = (B_u, u \geq 0)$  un mouvement brownien standard, et  $(\rho(y, u) : u \in [0, 1])$  et  $\bar{\rho}(y)$  défini pour  $y \in (0, 1)$  par :

$$\rho(y, u) = \mathbf{1}_{[0, y]}(u) - \frac{W(y, u)}{D'(y)} \quad \text{et} \quad \bar{\rho}(y) = \int_0^1 \rho(y, u) du.$$

*Remarque.* On obtient également l'expression explicite du noyau de covariance du processus gaussien  $\chi$  (voir remarque 4.24) ainsi que celle de sa variance donnée pour tout  $y \in (0, 1)$ , par la formule suivante :

$$\Sigma(y, y) = y(1 - y) + \frac{1}{D'(y)^2} \left( \int_0^1 W(y, x)^2 dx - D(y)^2 \right) + \frac{2}{D'(y)} \left( D(y)y - \int_0^y W(y, x) dx \right).$$

La preuve de ce théorème repose sur des développements uniformes d'Edgeworth pour des variables binomiales (voir Bhattacharya et Rao [22], Nagaev, Chebotarev et Zolotukhin [150] ou Uspensky [186]) et sur la méthode de Stein pour des vecteurs de variables binomiales de dimension 2 (voir Bentkus [20]).

Pour la convergence du processus  $(\sqrt{n}(\Pi_n(y) - y) : y \in (0, 1))$  pour la topologie de Skorohod, voir partie 2.4.3.

### 2.4.2 Densités d'homomorphismes de graphes partiellement étiquetés

Dans cette partie, nous allons étendre la notion de densité d'homomorphismes de graphes à des graphes partiellement étiquetés. Nous nous intéresserons ensuite aux résultats asymptotiques de mesures aléatoires construites à partir de ces densités d'homomorphismes pour des graphes aléatoires générés à partir d'un graphon.

**Notations préliminaires** Soit  $n \in \mathbb{N}^*$  et  $k \in [n]$ . On définit l'ensemble  $\mathcal{S}_{n,k}$  des  $[n]$ -mots de longueur  $k$  dont tous les caractères sont distincts, voir (4.7) pour une définition exacte. On a  $|\mathcal{S}_{n,k}| = A_n^k = n!/(n-k)!$ .

#### Homomorphismes de graphes partiellement étiquetés

La notion d'homomorphismes de graphes s'étend de manière naturelle aux graphes partiellement étiquetés. Un graphe partiellement étiqueté est un graphe simple fini dont certains de ses sommets sont étiquetés par différents entiers.

Soient  $F, G \in \mathcal{F}$  deux graphes simples tels que  $V(F) = [p]$  et  $V(G) = [n]$ . On suppose que  $n \geq p > k \geq 1$ . Soient  $\ell \in \mathcal{S}_{p,k}$  l'ensemble des  $k$  sommets étiquetés de  $F$  et  $\alpha \in \mathcal{S}_{n,k}$  l'ensemble des  $k$  sommets étiquetés de  $G$ . On définit l'ensemble  $\text{Inj}(F^\ell, G^\alpha)$  des homomorphismes injectifs  $f$  de  $F$  dans  $G$  tels que  $f(\ell_i) = \alpha_i$  pour tout  $i \in [k]$ , ainsi que sa densité :

$$t_{\text{inj}}(F^\ell, G^\alpha) = \frac{|\text{Inj}(F^\ell, G^\alpha)|}{A_{n-k}^{p-k}}.$$

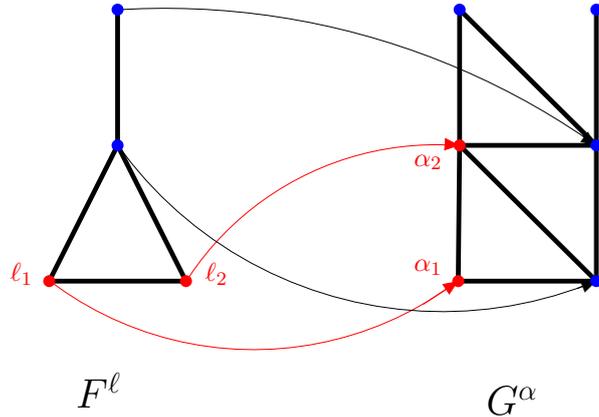


FIGURE 2.5 – Exemple d'homomorphisme injectif pour des graphes partiellement étiquetés..

*Remarque.* Il est facile de voir, en sommant sur tous les étiquetages de  $k$  sommets de  $G$ , que l'on a l'égalité suivante :

$$t_{\text{inj}}(F, G) = \frac{1}{|\mathcal{S}_{n,k}|} \sum_{\alpha \in \mathcal{S}_{n,k}} t_{\text{inj}}(F^\ell, G^\alpha). \quad (2.9)$$

### Une probabilité aléatoire

On définit la probabilité aléatoire  $\Gamma_n^{F,\ell}$  sur  $([0, 1], \mathcal{B}([0, 1]))$ , où  $\mathcal{B}([0, 1])$  est la tribu borélienne sur  $[0, 1]$ , par : pour tout  $g \in \mathcal{B}^+([0, 1])$  :

$$\Gamma_n^{F,\ell}(g) = \frac{1}{|\mathcal{S}_{n,k}|} \sum_{\alpha \in \mathcal{S}_{n,k}} g\left(t_{\text{inj}}(F^\ell, G_n^\alpha)\right). \quad (2.10)$$

*Remarque.* Afin de simplifier les résultats de convergence, la mesure a été définie pour des fonctions  $g$  d'une variable mais on peut également la définir pour des fonctions  $g$  à  $d$  variables avec  $d \geq 1$  en considérant non plus un seul graphe simple  $F$  mais une suite de  $d$  graphes simples  $F = (F_m : 1 \leq m \leq d) \in \mathcal{F}^d$ . Les résultats multidimensionnels se trouvent dans le chapitre 4.

### Résultats asymptotiques

Nous avons la convergence presque sûre suivante :

**Théorème C.** *La suite  $(\Gamma_n^{F,\ell}(dx) : n \in \mathbb{N}^*)$  de mesures aléatoires sur  $[0, 1]$  converge presque sûrement pour la topologie de la convergence étroite de mesures sur  $[0, 1]$  vers la probabilité déterministe  $\Gamma^{F,\ell}(dx)$ .*

Afin de simplifier la lecture, nous avons choisi de ne pas donner l'expression exacte de la probabilité limite  $\Gamma^{F,\ell}$  qui s'exprime en fonction des densités d'homomorphismes de graphes partiellement étiquetés dans des graphons (qui sont des généralisations naturelles des densités d'homomorphismes de graphes simples dans des graphons, voir la partie 4.2.3 du chapitre 4 pour une définition précise), voir l'équation (4.42) du chapitre 4.

En particulier, ce théorème permet d'obtenir les résultats suivants :

- Par le théorème de Portmanteau, on a presque sûrement, pour toute fonction  $g \in \mathcal{C}([0, 1])$ ,  $\lim_{n \rightarrow +\infty} \Gamma_n^{F,\ell}(g) = \Gamma^{F,\ell}(g)$ .
- Si on prend  $g = \text{Id}$  dans (2.10), on retrouve la convergence donnée par (2.7) puisque l'on a d'après (2.9) :

$$t_{\text{inj}}(F, G_n) = \frac{1}{|\mathcal{S}_{n,k}|} \sum_{\alpha \in \mathcal{S}_{n,k}} t_{\text{inj}}(F^\ell, G_n^\alpha),$$

et  $\Gamma^{F,\ell}(\text{Id}) = t(F, W)$ , voir remarque 4.8 du chapitre 4.

- Si on prend  $g = \mathbf{1}_{[0,D(y)]}$  avec  $y \in (0, 1)$  et  $F = K_2$  dans (2.10), on a, grâce à l'expression de  $\Gamma^{F,\ell}$  donnée dans la remarque 4.8 du chapitre 4, avec  $\bullet$  un des sommets de  $K_2$ , que :

$$\Gamma_n^{K_2,\bullet}(g) = \Pi_n(y) \quad \text{et} \quad \Gamma^{K_2,\bullet}(g) = y.$$

Si on suppose que  $D$  est strictement croissante sur  $(0, 1)$  alors le théorème C implique la convergence presque sûre de  $\Pi_n(y)$  vers  $y$  pour tout  $y \in (0, 1)$ . En utilisant le théorème de Dini, on obtient la convergence presque sûre de  $\Pi_n$  vers la fonction identité sur  $[0, 1]$  pour la norme uniforme, voir remarque 4.22 du chapitre 4.

Nous obtenons également les fluctuations associées à cette convergence presque sûre.

**Théorème D.** *Pour toute fonction  $g \in \mathcal{C}^2([0, 1])$ , on a la convergence en loi suivante :*

$$\sqrt{n} \left( \Gamma_n^{F,\ell}(g) - \Gamma^{F,\ell}(g) \right) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N} \left( 0, \sigma^{F,\ell}(g)^2 \right),$$

avec  $\sigma^{F,\ell}(g)^2 = \text{Var}(\mathcal{U}_g^{F,\ell})$  et  $\mathcal{U}_g^{F,\ell}$  définie par l'équation (4.43) du chapitre 4.

Pour  $g = \text{Id}$ , on retrouve le théorème central limite énoncé par Féray, Méliot et Nikeghbali [84].

Par contre, comme  $g = \mathbf{1}_{[0, D(y)]}$  n'appartient pas à l'ensemble  $\mathcal{C}^2([0, 1])$ , on ne peut pas appliquer directement le théorème précédent (avec  $F = K_2$  et  $k = 1$ ) pour obtenir la convergence en loi du processus  $(\sqrt{n}(\Pi_n(y) - y) : y \in [0, 1])$  vers  $(\chi(y) : y \in [0, 1])$  donnée dans le théorème B. C'est pour cette raison que d'autres techniques ont dû être mises en place pour démontrer le théorème B.

Comme corollaire immédiat du théorème D énoncé en dimension supérieure, voir le théorème 4.11 du chapitre 4, on établit la convergence en lois finies-dimensionnelles (en fait, du processus puisque l'ensemble  $\mathcal{F}$  est dénombrable) suivante :

**Proposition E.** *On a la convergence en lois finies-dimensionnelles :*

$$(\sqrt{n}(t_{\text{inj}}(F, G_n) - t(F, W)) : F \in \mathcal{F}) \xrightarrow[n \rightarrow \infty]{(lfd)} \Theta_{\text{inj}},$$

où  $\Theta_{\text{inj}} = (\Theta_{\text{inj}}(F) : F \in \mathcal{F})$  est un processus gaussien centré de fonction de covariance  $K_{\text{inj}}$  donnée, pour  $F, F' \in \mathcal{F}$ , par :

$$K_{\text{inj}}(F, F') = \sum_{q \in V(F)} \sum_{q' \in V(F')} t((F \bowtie F')(q, q'), W) - v(F)v(F')t(F, W)t(F', W).$$

En particulier, on retrouve le théorème central limite donné par Féray, Méliot et Nikeghbali [84]. Ce dernier résultat nous permet aussi d'obtenir le théorème central limite pour les densités d'homomorphismes de graphes quantiques (voir la convergence (4.52)) et pour les densités d'homomorphismes induits, voir le corollaire 4.15 du chapitre 4.

### 2.4.3 Extensions possibles

#### Convergence du processus du théorème B

Dans le théorème B, nous avons établi la convergence au sens des lois finies dimensionnelles du processus  $(\sqrt{n}(\Pi_n(y) - y) : y \in (0, 1))$  vers le processus  $(\chi_y : y \in (0, 1))$ . On aimerait pouvoir montrer la convergence du processus pour la topologie de Skorokhod. Pour ce faire, un argument de tension est nécessaire. On pourrait par exemple calculer les moments croisés d'ordre 4 du processus. Les techniques développées dans le chapitre 4 devraient pouvoir fonctionner en utilisant des développements d'Edgeworth pour des sommes de vecteurs aléatoires indépendants de variables de Bernoulli corrélées. Cependant, cette approche semble très technique et il nous apparaît préférable de trouver une méthode alternative.

#### Résultats asymptotiques pour des FDR empiriques associées à d'autres petits graphes comme les triangles

On pourrait établir un résultat plus général sur les asymptotiques de FDR empiriques associées non plus seulement aux degrés de la suite de  $W$ -graphes aléatoires  $(G_n(W) : n \in \mathbb{N}^*)$ , i.e. au graphe complet à deux sommets  $K_2$  mais à des graphes simples finis quelconques  $F$ . On pense notamment aux triangles ou aux graphes étoilés à deux arêtes (i.e. un sommet connecté par deux arêtes à deux autres sommets, « two-stars » en anglais). Pour  $F \in \mathcal{F}$  tel que  $V(F) = [p]$  et  $\ell \in \mathcal{S}_{p,k}$ , on pourrait s'intéresser au processus suivant :

$$\left( \frac{1}{|\mathcal{S}_{n,k}|} \sum_{\alpha \in \mathcal{S}_{n,k}} \mathbf{1}_{\{t_{\text{inj}}(F^\ell, G_n^\alpha) \leq t_x(F^\ell, W)\}} : x \in (0, 1)^k \right),$$

où  $t_x(F^\ell, W) = \mathbb{E}[t_{\text{inj}}(F^\ell, G_n^{[k]}) | (X_1, \dots, X_k) = x]$  d'après (4.31) et la seconde égalité de (4.37), dans le chapitre 4.

