

DÉNOMBREMENT ASYMPTOTIQUE DES GRAPHES AVEC UNE SUITE DE VOISINAGES

Charles Bordenave

CNRS & University of Toulouse

*Travail en collaboration avec **Pietro Caputo** (Univ. Roma 3)*

GRAPHES À SUITE DE DEGRÉS DONNÉE

GRAPHE AVEC UNE SUITE DE DEGRÉS DONNÉE

Une suite d'entiers $\mathbf{d} = (d_1, \dots, d_n)$ est **graphique** si il existe un graphe G avec sommets $[n] = \{1, \dots, n\}$ tel que pour tout i ,

$$\deg_G(i) = d_i.$$

Si graphique,

$$S = \sum_{i=1}^n d_i \in 2\mathbb{Z}.$$

Caractérisation des suites graphiques : *Erdős & Gallai (1960)*.

DÉNOMBREMENT ASYMPTOTIQUE

On considère une suite $\mathbf{d} = \mathbf{d}_n$ telle que pour tout k ,

$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}(d_i = k) \longrightarrow p_k.$$

et

$$\frac{1}{n} \sum_{i=1}^n d_i^2 \longrightarrow \sum_k k^2 p_k < \infty.$$

On s'intéresse aux graphes avec sommets $[n]$ et suite de degrés \mathbf{d} .

DÉNOMBREMENT ASYMPTOTIQUE

Soient $\kappa_1 = \sum_k k p_k$ et $\kappa_2 = \sum_k k(k-1)p_k$, on pose

$$\alpha = e^{-\frac{\kappa_2}{2\kappa_1} - \left(\frac{\kappa_2}{2\kappa_1}\right)^2}.$$

Théorème (Bender and Canfield (1978), Bollobás (1979))

Le nombre de graphes avec sommets $[n]$ et suite de degrés \mathbf{d} est équivalent à

$$|\mathcal{G}(\mathbf{d})| \sim \alpha \sqrt{2} \frac{(S/e)^{\frac{S}{2}}}{\prod_i d_i!}.$$

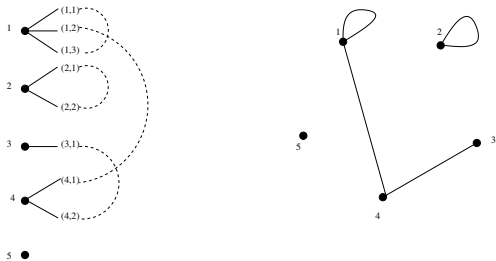
avec $S = \sum_i d_i$ pair.

MODÈLE DE CONFIGURATION

Ensemble des **demi-arêtes** :

$$E = \{(i, j) : i \in [n], 1 \leq j \leq d_i\}$$

Un **couplage** de E = une involution sans point fixe



Cela construit une application $G : \sigma \mapsto G(\sigma)$ de l'espace des couplages dans l'espace des **multi-graphes** avec suite de degrés \mathbf{d} .

MODÈLE DE CONFIGURATION

Pour chaque graphe $H \in \mathcal{G}(\mathbf{d})$, le nombre de préimage vaut

$$|G^{-1}(H)| = \prod_i d_i!.$$

Le nombre de couplages est égal à

$$(S-1)(S-3)\cdots 3.1 = \frac{S!}{2^{\frac{S}{2}} \left(\frac{S}{2}\right)!} \sim \sqrt{2} \left(\frac{S}{e}\right)^{\frac{S}{2}}.$$

Soit σ un couplage tiré uniformément et

$$\alpha_n = \mathbb{P}(G(\sigma) \in \mathcal{G}(\mathbf{d})) = \mathbb{P}(G(\sigma) \text{ est un graphe}).$$

On a donc

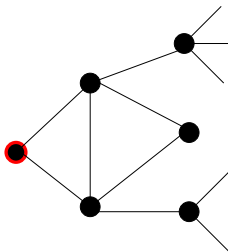
$$|\mathcal{G}(\mathbf{d})| = \alpha_n \frac{S!}{2^{\frac{S}{2}} \left(\frac{S}{2}\right)! \prod_i d_i!}.$$

La preuve consiste à montrer que $\alpha_n = \alpha + o(1)$.

GRAPHES À SUITE DE VOISINAGES DONNÉE

h -VOISINAGE

On considère \mathcal{G}_h^* , l'espace des graphes finis (sans label) enracinés de rayon h : on inclut les sommets à distance $h - 1$ et seul le nombre d'arêtes des sommets à distance $h - 1$ est dévoilé.



Exemple pour $h = 3$.

Pour $h = 1$: le degré de la racine.

SUITE DE VOISINAGES

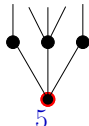
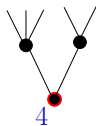
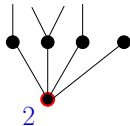
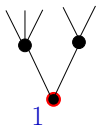
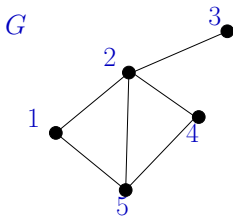
Une suite $\mathbf{g} = (g_1, \dots, g_n)$ d'éléments dans \mathcal{G}_h^* est **graphique** si il existe un graphe G avec sommets $[n]$ tel que pour tout i ,

$$(G, i)_h = g_i,$$

où $(G, i)_h$ est le h -voisinage du graphe G enraciné en i .

Pour $h \geq 2$, pas de caractérisation connue des suites graphiques !

UN EXEMPLE AVEC $h = 2$



SUITE DE VOISINAGES ADMISSIBLES

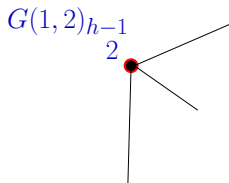
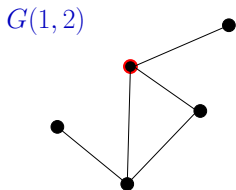
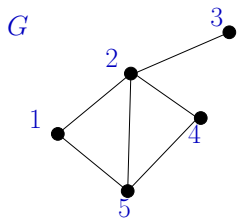
Si $G = (V, E)$ est un graphe et $\{u, v\} \in E$, on définit

$$G(u, v) = (G \setminus \{u, v\}, v),$$

graphe enraciné en v où l'arête $\{u, v\}$ a été enlevée.

Le $(h - 1)$ -voisinage de $G(u, v)$ est noté $G(u, v)_{h-1}$.

UN EXEMPLE AVEC $h = 2$



SUITE DE VOISINAGES ADMISSIBLES

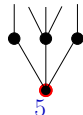
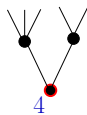
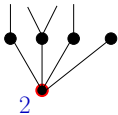
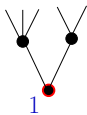
Pour tout $g, g' \in \mathcal{G}_{h-1}^*$ et $o \in V$ on pose,

$$S_{(G,o)}(g, g') = \left| \left\{ v : v \stackrel{\mathcal{G}}{\sim} o, G(o, v)_{h-1} = g, G(v, o)_{h-1} = g' \right\} \right|.$$

On trouve

$$\begin{aligned} \sum_i S_{(G,i)}(g, g') &= \sum_i S_{(G,i)}(g', g) \\ \sum_i S_{(G,i)}(g, g) &\in 2\mathbb{Z}. \end{aligned}$$

UN EXEMPLE AVEC $h = 2$



$$S_1(3, 1) = 1 \quad S_2(1, 3) = 2 \quad S_3(3, 0) = 1 \quad S_4(3, 1) = 1 \quad S_5(1, 2) = 2$$

$$S_1(2, 1) = 1 \quad S_2(2, 3) = 1 \quad S_4(2, 1) = 1 \quad S_5(3, 2) = 1$$

$$S_2(0, 3) = 1$$

SUITE DE VOISINAGES ADMISSIBLES

Une suite $\mathbf{g} = (g_1, \dots, g_n)$ dans \mathcal{G}_h^* est **admissible** si elle satisfait pour tout g, g' dans \mathcal{G}_{h-1}^* ,

$$S_{\mathbf{g}}(g, g') = \sum_i S_{g_i}(g, g') = \sum_i S_{g_i}(g', g) = S_{\mathbf{g}}(g', g)$$

$$S_{\mathbf{g}}(g, g) = \sum_i S_{g_i}(g, g) \in 2\mathbb{Z}.$$

C'est une **condition nécessaire** pour être graphique.

DÉNOMBREMENT ASYMPTOTIQUE

Soit $\mathcal{T}_h^* \subset \mathcal{G}_h^*$ l'ensemble des arbres enracinés de rayon h .

On a $\mathcal{T}_1^* = \mathcal{G}_1^*$.

Soit $\mathbf{t} = \mathbf{t}_n = (t_1, \dots, t_n)$ une suite admissible dans \mathcal{T}_h^* .

On suppose que pour tout $t \in \mathcal{T}_h^*$,

$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}(t_i = t) \longrightarrow p_t$$

et (pour simplifier) pour tout $1 \leq i \leq n$,

$$\deg_{t_i}(o) \leq \theta.$$

DÉNOMBREMENT ASYMPTOTIQUE

On fixe un ordre sur \mathcal{T}^* et on pose pour m pair

$$m!! = \frac{m!}{2^{m/2}(m/2)!} = \text{nombre de couplage à } m \text{ éléments.}$$

Théorème

Le nombre de graphes avec sommets $[n]$ et suite de h -voisinage \mathbf{t} est équivalent à

$$|\mathcal{G}(\mathbf{t})| \sim \alpha \frac{\prod_{t < t' \in \mathcal{T}_{h-1}^*} S_{\mathbf{t}}(t, t')! \prod_{t \in \mathcal{T}_{h-1}^*} S_{\mathbf{t}}(t, t)!!}{\prod_{t, t' \in \mathcal{T}_{h-1}^*} \prod_i S_{\mathbf{t}_i}(t, t')!},$$

où $\alpha = \dots > 0$ s'exprime en fonction de $(p_t)_{t \in \mathcal{T}_h^*}$.

SUITES GRAPHIQUES

Comme $\alpha > 0$, la suite t_n est graphique pour n suffisamment grand.

Il doit être possible d'étendre le critère d'Erdős-Gallai pour des h -voisinages qui sont des arbres (?).

MODÈLE DE CONFIGURATION COLORÉE

On pose $c = (g, g')$, ensemble des **demi-arêtes colorées** :

$$E_c = \{(i, j) : i \in [n], 1 \leq j \leq S_{t_i}(c)\},$$

$$\mathcal{C}_< = \{(g, g') : g < g' \in \mathcal{T}_{h-1}^*\},$$

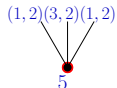
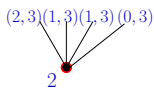
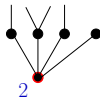
$$\mathcal{C}_= = \{(g, g) : g \in \mathcal{T}_{h-1}^*\}.$$

Si $\bar{c} = (g', g)$, la condition d'admissibilité se réécrit comme

$$|E_c| = |E_{\bar{c}}| \quad \text{pour tout } c \in \mathcal{C}_< ,$$

$$|E_c| \in 2\mathbb{Z} \quad \text{pour tout } c \in \mathcal{C}_= .$$

UN EXEMPLE AVEC $h = 2$



MODÈLE DE CONFIGURATION COLORÉE

On se donne une configuration colorée

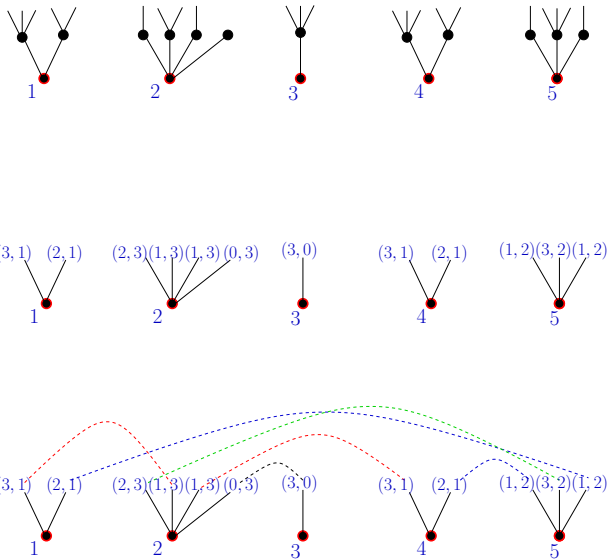
$$\sigma = (\sigma_c)_{c \in \mathcal{C}_{\leq}},$$

avec pour $c \in \mathcal{C}_{=}$, σ_c couplage de E_c et pour $c \in \mathcal{C}_{<}$, σ bijection de E_c dans $E_{\bar{c}}$.

On construit ainsi une application $\sigma \mapsto G(\sigma)$ dans l'espace des multi-graphes avec des arêtes proprement colorées, c.à.d.

- (i) si (u, v) a couleur c , (v, u) a couleur \bar{c} ,
- (ii) le sommet i a $S_{t_i}(c)$ arêtes issues de i de couleur c .

UN EXEMPLE AVEC $h = 2$



MODÈLE DE CONFIGURATION COLORÉE

Fait crucial : si $G = G(\sigma)$ est tel que $(G, i)_h$ est un arbre pour tout i alors

$$((G, 1)_h, \dots, (G, n)_h) = \mathbf{t} = (t_1, \dots, t_n).$$

On en déduit le dénombrement asymptotique en calculant pour $n \rightarrow \infty$,

$$\mathbb{P} (G(\sigma) \text{ n'a pas de cycle de longueur } \leq 2(h - 1))$$

lorsque $\sigma = (\sigma)_{c \in \mathcal{C}_{\leq}}$ est tiré uniformément.

ENTROPIE MICRO-CANONIQUE

DISTANCE EN VARIATION TOTALE

Soient \mathcal{X} un espace discret et P, Q deux probabilités sur \mathcal{X} .

$$d_{\text{TV}}(P, Q) = \frac{1}{2} \sum_{x \in \mathcal{X}} |P(x) - Q(x)| = \inf \mathbb{P}(X \neq Y),$$

où le sup est sur les lois de $(X, Y) \in \mathcal{X}^2$ telles que $X \stackrel{d}{\sim} P$ et $Y \stackrel{d}{\sim} Q$ (couplage des mesures).

MESURE EMPIRIQUE

À $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{X}^n$, on associe sa mesure empirique

$$U(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$$

c.à.d pour tout $x \in \mathcal{X}$,

$$U(\mathbf{x})(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(x_i = x).$$

ENTROPIE

Entropie de P

$$H(P) = - \sum_x P(x) \log P(x).$$

L'entropie admet une définition **micro-canonique**

$$H(P) = \lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \log |\{\mathbf{x} \in \mathcal{X}^n : d_{\text{TV}}(U(\mathbf{x}), P) \leq \varepsilon\}|.$$

L'entropie est maximisée pour la mesure uniforme.

ENTROPIE POUR LES VOISINAGES

GRAPHES À NOMBRE D'ARÊTES FIXÉS

Soit $\mathcal{G}_{n,m}$ l'ensemble des graphes avec sommets $[n]$ et m arêtes.

On prend une suite $m = m(n)$ telle que

$$\frac{2m}{n} \longrightarrow d.$$

On a

$$\log |\mathcal{G}_{n,m}| = \log \binom{\frac{n(n-1)}{2}}{m} = m \log n + ns(d) + o(n),$$

avec

$$s(d) = \frac{d}{2} - \frac{d}{2} \log d.$$

MESURE EMPIRIQUE

À $G \in \mathcal{G}_{n,m}$ on associe sa mesure empirique des h -voisinages

$$U(G)_h = \frac{1}{n} \sum_{i=1}^n \delta_{(G,i)_h}.$$

C'est donc une mesure de probabilité sur \mathcal{G}_h^* .

On veut définir une entropie micro-canonique pour les mesures empiriques de voisinages.

MESURE ADMISSIBLE

Soit P est une mesure de probabilités sur \mathcal{G}_h^* et (G, o) de loi P .

Rappel, pour tout $g, g' \in \mathcal{G}_{h-1}^*$,

$$S_{(G,o)}(g, g') = \left| \left\{ v : v \stackrel{\mathcal{G}}{\sim} o, G(o, v)_{h-1} = g, G(v, o)_{h-1} = g' \right\} \right|.$$

On pose

$$e_P(g, g') = \mathbb{E}_P S_{(G,o)}(g, g').$$

On dit que P est **admissible** si pour tout $g, g' \in \mathcal{G}_{h-1}^*$,

$$e_P(g, g') = e_P(g', g).$$

Par construction $U(G)_h$ est admissible.

ENTROPIE POUR LES VOISINAGES

On s'intéresse aux graphes dont les voisinages sont proches de P ,

$$\mathcal{G}_{n,m}(P, \varepsilon) = \{G \in \mathcal{G}_{n,m} : d_{\text{TV}}(U(G)_h, P) \leq \varepsilon\}.$$

Sous réserve que les limites soient définies,

$$J_h(P) = \lim_{\varepsilon \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{\log |\mathcal{G}_{n,m}(P, \varepsilon)| - m \log n}{n},$$

peut être vu comme une entropie micro-canonique.

ENTROPIE POUR LES VOISINAGES

Si P n'est pas admissible ou $\mathbb{E}_P \deg_G(o) \neq d$ ou $\text{supp}(P) \not\subset \mathcal{T}_h^*$.

$$J_h(P) = -\infty.$$

Simon, on peut définir la mesure de probabilité sur $\mathcal{T}_{h-1}^* \times \mathcal{T}_{h-1}^*$

$$\hat{P}(g, g') = \frac{e_P(g, g')}{d}$$

et

$$J_h(P) = -s(d) + H(P) - \frac{d}{2}H(\hat{P}) - \frac{d}{2} \sum_{t, t' \in \mathcal{T}_{h-1}^*} \mathbb{E}_P \log |S_{(G, o)}(t, t')|!$$

(se déduit à partir du théorème de dénombrement).

CAS DES DEGRÉS

Pour $h = 1$, on a simplement,

$$J_1(P) = s(d) - H(P|\text{Poi}(d)),$$

où

$$H(P|Q) = \sum_x P(x) \log \left(\frac{P(x)}{Q(x)} \right)$$

est l'entropie relative.

En particulier, l'unique maximum de J_1 est $\text{Poi}(d)$.

MAXIMUM DE L'ENTROPIE

Pour $1 \leq \ell < h$, si P_ℓ est la restriction de P à \mathcal{T}_ℓ^* , c.à.d la loi de $(G, o)_\ell$ on a

$$J_h(P) \leq J_\ell(P_\ell)$$

et égalité si et seulement si P a une forme explicite.

On sait résoudre pour Q mesure de probabilité sur \mathcal{T}_ℓ^* ,

$$\arg \max \{J_h(P) : P \text{ proba sur } \mathcal{T}_h^* \text{ telle que } P_\ell = Q\}.$$

ENTROPIE POUR LA TOPOLOGIE DE BENJAMINI-SCHRAMM

TOPOLOGIE DE BENJAMINI-SCHRAMM

Soient ρ et σ deux mesures de probabilité sur \mathcal{G}^* (identifié à l'espace projectif de $\mathcal{G}_1^* \times \mathcal{G}_2^* \times \dots$).

Si ρ_h est la restriction de ρ à \mathcal{G}_h^* , pour $\ell < h$, $\rho_\ell = (\rho_h)_\ell$ et on pose

$$d_{\text{BS}}(\rho, \sigma) = \sum_h 2^{-h} d_{\text{TV}}(\rho_h, \sigma_h).$$

Deux mesures sont proches si les lois de leurs h -voisinages sont proches pour h grand.

On dit que ρ est **unimodulaire** si pour tout h , ρ_h est admissible.

MESURE EMPIRIQUE

À $G \in \mathcal{G}_{n,m}$ on associe la suite de ses mesures empiriques des h -voisinages

$$U(G) = (U(G)_1, U(G)_2, \dots).$$

C'est équivalent à se donner la mesure empirique

$$U(G) = \frac{1}{n} \sum_{i=1}^n \delta_{(G,i)}.$$

C'est une mesure de probabilité unimodulaire sur \mathcal{G}^* .

ENTROPIE POUR BENJAMINI-SCHRAMM

On s'intéresse aux graphes dont tous les voisinages sont proches de la mesure ρ ,

$$\mathcal{G}_{n,m}(\rho, \varepsilon) = \{G \in \mathcal{G}_{n,m} : d_{\text{BS}}(U(G), \rho) \leq \varepsilon\}.$$

Sous réserve que les limites soient définies,

$$\Sigma(\rho) = \lim_{\varepsilon \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{\log |\mathcal{G}_{n,m}(\rho, \varepsilon)| - m \log n}{n},$$

peut être vu comme une entropie micro-canonique.

ENTROPIE POUR BENJAMINI-SCHRAMM

On prend une suite $m = m(n)$ telle que

$$\frac{2m}{n} \longrightarrow d.$$

Théorème

$$\Sigma(\rho) = \lim_{h \rightarrow \infty} \downarrow J_h(\rho_h).$$

MAXIMUM DE L'ENTROPIE

Pour $h = 1$, l'unique maximum de l'entropie est l'arbre de Galton-Watson avec distribution des descendants $\text{Poi}(d)$.

Pour tout $h \geq 1$ et P admissible mesure de probabilité sur \mathcal{T}_h^* ,

$$\arg \max \{ \Sigma(\rho) : \rho \text{ proba sur } \mathcal{T}^* \text{ telle que } \rho_h = P \}$$

admet pour unique maximum $\text{UGW}_h(P)$ l'arbre de Galton-Watson unimodulaire avec distribution des h -voisinages P (construit explicitement).

ARBRE DE GALTON-WATSON UNIMODULAIRE

On trouve

$$\Sigma(\text{UGW}_h(P)) = J_h(P).$$

Si $\ell > h$,

$$\text{UGW}_\ell(\text{UGW}_h(P)_\ell) = \text{UGW}_h(P).$$

REMARQUES FINALES

QUELQUES CONSÉQUENCES

Permet de montrer un **principe de grandes déviations** pour $U(G_n)$ où G_n est un graphe d'Erdős-Rényi de paramètres $(n, \lambda/n)$ ou un graphe uniforme avec suite de degrés donnée.

Permet de remonter que toute mesure unimodulaire ρ sur \mathcal{T}^* admet une suite de graphes finis G_n telle que $d_{BS}(U(G_n), \rho) \rightarrow 0$.

(Bowen (2003), Elek (2010), Benjamini-Schramm-Lyons (2012)).

Problème ouvert : est-ce que toute mesure unimodulaire ρ admet une suite de graphes finis G_n telle que $d_{BS}(U(G_n), \rho) \rightarrow 0$?

RETOUR SUR L'ENTROPIE

Avec

$$s(d) = \frac{d}{2} - \frac{d}{2} \log d,$$

$$\begin{aligned} J_h(P) &= -s(d) + H(P) - \frac{d}{2} H(\hat{P}) - \frac{d}{2} \sum_{t, t' \in \mathcal{T}_{h-1}^*} \mathbb{E}_P \log |S_{(G,o)}(t, t')|! \\ &\leq s(d). \end{aligned}$$

Pas de preuve directe ! Comment trouver une **formulation alternative** de cette entropie ? $J_1(P) = s(d) - H(P|\text{Poi}(d))$.

AU DELÀ DES ARBRES

$$\Sigma(\rho) = \lim_{\varepsilon \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{\log |\mathcal{G}_{n,m}(\rho, \varepsilon)| - m \log n}{n}.$$

Avec une renormalisation différente, on peut définir une entropie qui est ne vaut pas $-\infty$ sur les mesures qui ne sont pas supportées sur les arbres enracinés (?).

MERCI DE VOTRE ATTENTION !

*C. Bordenave & P. Caputo. Large deviations of empirical neighborhood distribution in sparse random graphs. *Probability Theory and Related Fields* (2014).*