Compter les sous-graphes

Danièle Gardy
Laboratoire DAVID (UVSQ)

Travail en collaboration avec Gwendal Collet,
Bernhard Gittenberger, Élie de Panafieu, Vlady Ravelomanana

Réunion Métaconc, Caen, 30 janvier 2019

Graphes aléatoires

Il y a 60 ans : Modèles d'Erdős-Rényi

Un graphe G est défini par la donnée des ensembles de ses sommets $V(G)$ et de ses arêtes $E(G)$

– Graphe simple : pas d'arêtes multiples, pas de boucles
– Multigraphe : arêtes multiples et boucles autorisées

Deux modèles classiques de graphes aléatoires à n sommets

- $G(n, m)$: tirer aléatoirement m arêtes distinctes parmi les $\frac{n(n-1)}{2}$ possibles
- $G(n, p)$: tirer chaque arête avec probabilité p

$G(n, p)$ lorsque $n^2 p \to +\infty$ et $G(n, m)$ pour $m = \binom{n}{2} p$ ont le même comportement asymptotique

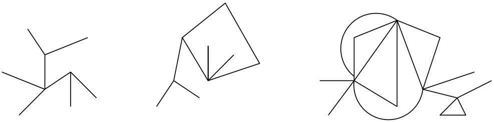
[Erdős-Rényi'59 et '60]

Naissance de la composante géante

Comportement des composantes connexes d'un graphe G lorsque $m, n \to +\infty$?

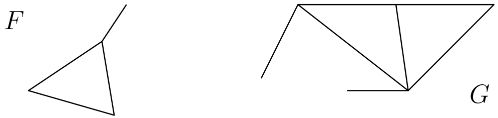
Densité $d(G) = \frac{|E(G)|}{|V(G)|} \Rightarrow$ phénomène de seuil

- Au début, arbres et unicycles isolés
- Apparition du premier cycle pour $m \sim \frac{n}{3}$
- Apparition de la composante géante pour $m \sim \frac{n}{2}$



Sous-graphes

$$F \text{ est un sous-graphe de } G$$
$$\Updownarrow$$
$$V(F) \subset V(G) \text{ et } E(F) \subset E(G)$$



- Seuil d'apparition de F quand G évolue ?
- $G[F]$ nombre de sous-graphes de G isomorphes au graphe F
  Loi de $G[F]$?

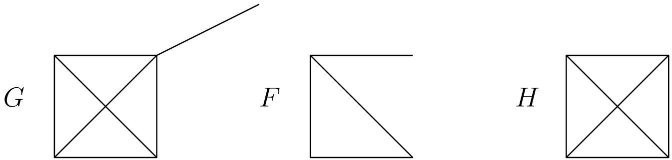
Seuil d'apparition d'un sous-graphe

Dans le modèle $G(n, m)$, si $m = o\left(n^{2-\frac{1}{d(F)}}\right)$, alors

$$Pr(G[F] \geq 1) = \mathcal{O}\left(\left(\frac{m}{n^{2-\frac{1}{d(F)}}}\right)^{m(F)}\right) \to 0$$

[Erdős-Rényi'60]

Densité et équilibre



$$d(F) = 1 < d(G) = \frac{7}{5} < d(H) = \frac{6}{4}$$

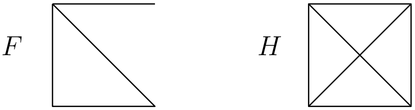
Densité essentielle de G : $d^\star(G) = \max_{H \subset G} d(H)$

G est **équilibré** $\Leftrightarrow d(G) = d^\star(G)$

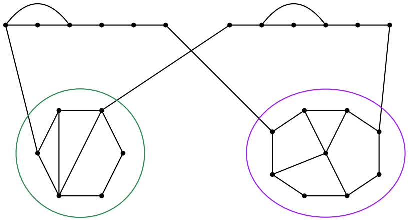
- G **strictement** équilibré $\Leftrightarrow d(G) > \max_{H \subsetneq G} d(H)$
- G **tout juste** équilibré $\Leftrightarrow d(G) = \max_{H \subsetneq G} d(H)$

Graphe équilibré

Ici H est strictement équilibré mais F est tout juste équilibré



Un plus gros exemple de graphe tout juste équilibré



Seuil d'apparition d'un sous-graphe

Modèle $G(n, p)$

- [Bollobas'81] Seuil d'apparition d'un sous-graphe

$$p = \frac{1}{n^{d^\star(F)}}$$

- [Rucinski'88]
  – Si $n^2(1-p) \to +\infty$ et $np^{d^\star(F)} \to +\infty$: $G[F]$ suit une loi normale
  – Si $np^{d^\star(F)} = \Theta(1)$: $G[F]$ suit une loi de Poisson $\Leftrightarrow$ F est strictement équilibré

Sous-graphes

Depuis les années 60, beaucoup de résultats sur le seuil d'apparition dans G d'un sous-graphe F et son nombre d'occurrences $G[F]$

- Résultats plus précis (F fortement équilibré, ...)
- Contrainte sur le graphe G (régulier, à degrés contraints, ...)
- Contrainte sur le sous-graphe F (régulier, ...)
- Autres modèles de graphe (à invariance d'échelle, ...)
- Étude du coefficient de clustering (relié au nombre de triangles)
- ...

La plupart de ces résultats ont été obtenus par des méthodes probabilistes

Combinatoire analytique des graphes

Cete partie de l'histoire commence il y a 30 ans...

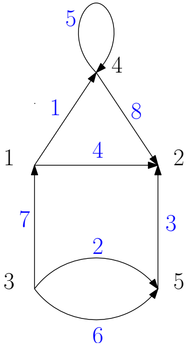
- 1989 : Flajolet, Knuth et Pittel, premiers cycles d'un graphe (fonction génératrice bivariée, configuration d'arrêt)
- 1990 : Mc Kay et Wormald, comptage des graphes à séquence de degrés donnée (fortement multi-dimensionnel : une variable par sommet du graphe)
- 1993 : Janson, Knuth, Luczak et Pittel, naissance de la composante géante (fonctions génératrices bivariées, graphes simples et multigraphes)
- 2009 : Gimenez et Noy, énumération de graphes planaires (décomposition des graphes planaires en composants connexes, 2-connexes, 3-connexes, ...)
- Depuis une décennie : Drmota, Noy, Ravelomanana, Rué, ... pour différentes classes de graphes et divers paramètres
- 2014 : de Panafieu, approche générale des *multigraphes*

Multigraphes

$$G = (V(G), E(G))$$

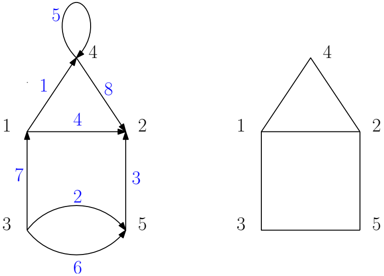
- Les sommets sont étiquetés
- Les arêtes sont étiquetées et orientées
- Les arêtes multiples et boucles sont autorisées
- Une arête est un triplet (v, w, e) (origine, but, étiquette)
- Les étiquettes des sommets et des arêtes sont des entiers
- Étiquetage *canonique* d'un multigraphe de n sommets à m arêtes
  - Sommets étiquetés de 1 à n
  - Arêtes étiquetées de 1 à m

Multigraphes



Ce multigraphe a 5 sommets et 8 arêtes, dont une boucle et une arête double ; il est étiqueté canoniquement

Multigraphes vs. graphes simples



Pour passer d'un multigraphe à un graphe simple :

- Enlever les boucles et les arêtes multiples
- Oublier l'orientation des arêtes

Dans ce qui suit, « graphe » = multigraphe

Compter les (multi)graphes canoniques

- Une arête = un couple (x, y) de sommets (ordre important)
- Un (multi)graphe = une suite de m arêtes, i.e., de $2m$ sommets

$$M_{n,m} = n^{2m}$$

Fonction génératrice exponentielle

$$\begin{aligned} M(z,w) &= \sum_{n,m\geq 0} M_{n,m} \frac{z^n}{n!} \frac{w^m}{2^m\, m!} \\ &= \sum_{n\geq 0} \frac{z^n}{n!}\, e^{n^2 \frac{w}{2}} \end{aligned}$$

(Pour tenir compte de l'orientation, chaque arête a un poids $\frac{1}{2}$)

Compter une famille $\mathcal{F}$ de graphes

Fonction génératrice exponentielle

$$F(z,w) = \sum_{G \in \mathcal{F}} \frac{w^{m(G)}}{2^{m(G)} m(G)!} \frac{z^{n(G)}}{n(G)!}$$

Nombre de graphes de $\mathcal{F}$ à n sommets et m arêtes

$$F_{n,m} = n!\, 2^m m!\, [z^n w^m] F(z,w)$$

Nombre d'occurrences d'un sous-graphe F

Variable supplémentaire u pour marquer les occurrences de F

$$\mathrm{G}^F(z, w, u) = \sum_{G \in \mathrm{G}} u^{G[F]} \frac{w^{m(G)}}{2^{m(G)} m(G)!} \frac{z^{n(G)}}{n(G)!}$$

Nombre de graphes à n sommets, m arêtes, et exact. t copies de F

$$\mathrm{G}^F_{n,m,t} = n! 2^m m! \, [z^n w^m u^t] \, \mathrm{G}^F(z, w, u)$$

Comment obtenir $\mathrm{G}^F(z, w, u)$?

Patchworks de graphes

- Plusieurs occurrences d'un graphe F peuvent **s'intersecter** : elles forment un **patchwork**
- Un patchwork est lui-même un graphe et peut apparaître comme sous-graphe

Comment obtenir $G^F(z, w, u)$?

- Savoir compter les graphes avec *une* occurrence *distinguée* d'un sous-graphe H
- Prendre pour H un *patchwork* de t occurrences du sous-graphe F
- On sait donc compter les graphes avec *au moins* t occurrences de F
- Inclusion-exclusion $\Rightarrow$ *exactement* t occurrences de F

Sous-graphe distingué H

Pour former un graphe contenant une copie distinguée de H :

- Choisir une copie de $H \leftrightarrow H(z,w)$
- Ajouter des sommets non dans $H \leftrightarrow e^z$
- Si on a en tout n sommets, choisir des arêtes dans les n^2 possibles $\leftrightarrow e^{n^2 w/2}$

Nombre de graphes avec une copie distinguée de H (et peut-être d'autres copies...)

$$G_{n,m}^{[H]} = n!\, 2^m\, m!\, [z^n w^m] H(z,w)\, e^z\, e^{n^2 w/2}$$

[CdePGGR'18, Th. 1]

Patchworks de graphes

Un F-patchwork est un ensemble fini de graphes isomorphes à F

$$P = \{(V_1, E_1), \dots, (V_{|P|}, E_{|P|})\}$$

Ces graphes peuvent partager des sommets et des arêtes

- si deux sommets appartenant à deux graphes distincts ont la même étiquette, on les fusionne
- si deux arêtes appartenant à deux graphes distincts ont la même étiquette, on les fusionne

$\Rightarrow$ deux arêtes de même étiquette doivent porter sur les mêmes sommets, et avoir la même orientation

Patchwork P d'un graphe F

- Taille $|P| = t$ de P : nombre de copies de F
- Sommets et arêtes de P

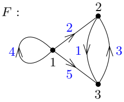
$$V(P) = \bigcup_{i=1}^{t} V_i, \quad \text{and} \quad E(P) = \bigcup_{i=1}^{t} E_i,$$

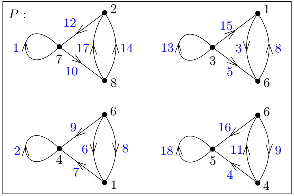
- Cardinalités $n(P)$ et $m(P)$
- Densité d'un patchwork P

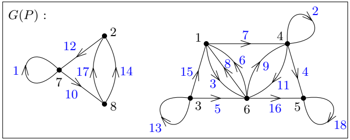
$$d(P) := d(G(P)) = \frac{E(P)}{V(P)}$$

où $G(P)$ est le graphe obtenu en fusionnant les sommets et arêtes partagés

Exemple de patchwork







- F est le graphe initial
- P est un patchwork de 4 copies de F
- $G(P)$ est le graphe correspondant à P

Comment obtenir $G^F(z, w, u)$?

- ✓ Savoir compter les graphes avec *une* occurrence *distinguée* d'un sous-graphe H
- ✓ Prendre pour H un *patchwork* de t occurrences du sous-graphe F
- ✓ On sait donc compter les graphes avec *au moins* t occurrences de F
- Inclusion-exclusion $\Rightarrow$ *exactement* t occurrences de F

Graphes avec t occurrences de F

- u marque le nombre de copies de F dans le patchwork
- $\mathrm{Patch}_{[F]}(z, w, u)$ fonction génératrice des patchworks de F
- Les graphes avec un patchwork distingué H de taille t ont *au moins* t copies de F
- Pour avoir *exactement* t copies, inclusion–exclusion.
  Traduction sur la fonction génératrice : $u \rightsquigarrow u - 1$

Nombre de graphes à n sommets, m arêtes, et t copies de F :

$$G^{[H]}_{n,m,t} = n!\, 2^m\, m!\, [z^n w^m u^t]\, \mathrm{Patch}_{[F]}(z, w, u-1)\, e^z\, e^{n^2 w/2}$$

[CdePGGR'18, Th. 3]

Sous-graphes

Seuil d'apparition d'un sous-graphe F dans G

n fixé, m croît à partir de 0
Comportement asymptotique de $G[F]$ quand $m \nearrow$?

- Au début, pas possible d'avoir F
- Quand m/n^2 atteint $d(F)$, F apparaît
- Ensuite, « beaucoup » de copies de F

Seuil d'apparition

- Sous-graphe F dense
  Dans le modèle $G(n, m)$, si $m = o\left(n^{2-\frac{1}{d(F)}}\right)$, alors

$$Pr(G[F] \geq 1) = \mathcal{O}\left(\left(\frac{m}{n^{2-\frac{1}{d(F)}}}\right)^{m(F)}\right) \to 0$$

  [Erdős-Rényi '60]

- Graphes simples, sous-graphe F de densité essentielle $d^\star$
  Soit $m = \mathcal{O}(n^\alpha)$ pour $0 < \alpha < 2$

$$\begin{cases} G[F] = 0 \text{ a.a.s} & \text{pour } \alpha < 2 - 1/d^\star, \\ E(G[F]) = \mathcal{O}(n^{m(F)(\alpha-2+1/d^\star)}) & \text{pour } \alpha \geq 2 - 1/d^\star. \end{cases}$$

  [Erdős-Rényi '60, Bollobás'81]

Pour aller plus loin

Les résultats précédents ne faisaient intervenir que l'espérance du nombre de copies à n et m connus

Peut-on avoir plus d'information ?

$$G_{n,m,t}^{[F]} = n!\, 2^m\, m!\, [z^n w^m u^t]\, \mathrm{Patch}_{[F]}(z, w, u-1)\, e^z\, e^{n^2 w/2}$$

Calcul de $\mathrm{Patch}_{[F]}(z, w, u)$ compliqué !

F strictement équilibré

Supposons F **strictement équilibré**

$\Updownarrow$

P patchwork de deux copies non disjointes de F : $d(P) > d(F)$

$\Downarrow$

Lorsque F apparaît, seuls les patchworks *disjoints* contribuent

$\Downarrow$

Pour étudier la transition : remplacer $\mathrm{Patch}_{[F]}(z, w, u)$ par $e^{uF(z,w)}$

Lois de Poisson

F strictement équilibré, de densité $d(F)$; $\alpha = 2 - 1/d(F)$, $m \sim cn^{\alpha}$

$G[F]$ suit une loi limite Poisson de paramètre $\lambda(F)$, avec

- (multigraphes)

$$\lambda(F) = \frac{c^{m(F)}}{m(F)!\, n(F)!}$$

  [CdePGGR'18]

- (Graphes simples, F avec $|\,\mathrm{Aut}(F)|$ automorphismes)

$$\lambda(F) = \frac{(2c)^{m(F)}}{|\,\mathrm{Aut}(F)|}$$

  [Bollobás'81]

Graphes sans copie de F

F strictement équilibré, de densité $d(F)$; $\alpha = 2 - 1/d(F)$, $m \sim cn^{\alpha}$

$$Proba(G[F] = 0) = e^{-\lambda(F)}$$

avec

$$\begin{cases} \lambda(F) := \frac{c^{m(F)}}{n(F)!m(F)!} & \text{(multigraphes)} \\ \\ \lambda(F) := \frac{(2c)^{\ell}}{|\operatorname{Aut}(F)|} & \text{(graphes simples, } F \text{ a } \ell \text{ arêtes)} \end{cases}$$

Restrictions sur les degrés

Attention : ce n'est plus un modèle $G(n,m)$ ou $G(n,p)$

δ_d poids d'un sommet de degré d, $\Delta(x) = \sum_d \delta_d \frac{x^d}{d!}$

- Graphes d-réguliers

$$\delta_d = 1 \text{ et pour } i \neq d, \delta_i = 0$$

- Graphes eulériens (degrés pairs)

$$\Delta(x) = \cosh x$$

- Graphes à loi de puissance : $Pr(deg(v) = d) \approx \lambda\, d^{-\beta}$

$$\Delta(x) = \sum_d d^{-\beta} x^d$$

[cf. l'exposé de Collet, Aléa'18]

Restrictions sur les degrés

- F est strictement équilibré et de fonction génératrice $F(z, w, \mathbf{y}) = \sum_{G \in \mathcal{F}} \left(\prod_{v \in V(G)} y_{\deg(v)} \right) \frac{w^{m(G)}}{2^{m(G)} m(G)!}$
- Conditions de régularité sur $\Delta(x)$ et $F(z, w)$
- $m = \Theta(n^\alpha)$ pour $\alpha = 2 - 1/d(F)$

Alors $G[F]$ suit une loi limite Poisson de paramètre

$$\lambda(F) = \lim_{n \to +\infty} F\left(n, \frac{1}{2m}, \left(\frac{\chi^d \Delta^{(d)}(\chi)}{\Delta(\chi)} \right)_{d \geq 0} \right)$$

avec $\chi := \chi(n)$ unique solution positive de $\frac{\chi \Delta'(\chi)}{\Delta(\chi)} = \frac{2m}{n}$
[CdePGGR'18, th. 9]

Si l'ensemble Δ est fini

- $\Delta(x)$ polynôme et apériodique : $\gcd\{d_1 - d_2 \mid \delta_{d_1} \neq 0, \delta_{d_2} \neq 0\} = 1$
- n, m t.q. $\lim \frac{2n}{m} \in]\min(\mathrm{Support}(\Delta(x))), \max(\mathrm{Support}(\Delta(x)))[$

i) F ni arbre ni unicycle $\Rightarrow \quad G[F] = 0$ a.a.s.

ii) T arbre à k sommets et $|\mathrm{Aut}(T)|$ automorphismes

$$E(G[T]) \sim \frac{1}{|\mathrm{Aut}(T)|} \frac{n^k}{(2m)^{k-1}} \prod_{v \in V(T)} \frac{\chi^{\deg(v)} \Delta^{(\deg(v))}(\chi)}{\Delta(\chi)}.$$

iii) C_ℓ cycle de longueur ℓ : $G[C_\ell]$ suit asympt. une loi de Poisson de moyenne

$$\frac{1}{2\ell} \left(\frac{n}{m} \frac{\chi^2 \Delta''(\chi)}{\Delta(\chi)} \right)^\ell .$$

[CdePGGR'18, th. 8]

En conclusion

Qu'avons-nous obtenu ?

- Une méthode générale pour aborder les questions d'énumération et de distribution des sous-graphes
- Les sous-graphes fortement équilibrés sont raisonnablement bien compris
- C'est nettement plus complexe pour les sous-graphes tout juste équilibrés
- Mais il est possible de traiter systématiquement au moins une classe importante de graphes

Bibliographie

- *Threshold functions for small subgraphs : an analytic approach*
  Collet, G., Gittenbergerg, de Panafieu, Ravelomanana
  [Eurocomb 2017, Vienne]
- *Threshold functions for small subgraphs in simple graphs and multigraphs*
  Collet, G., Gittenberger, de Panafieu, Ravelomanana
  [2018, ArXiV report 1807.05772, à paraître]