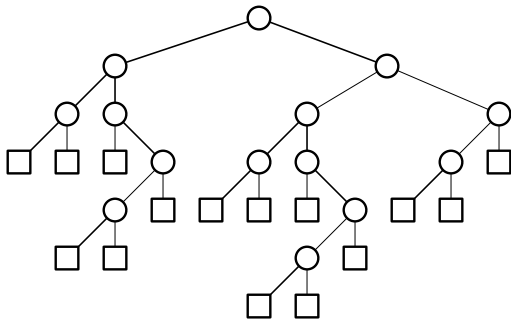


## Les arbres binaires compactés possèdent un exponentiel étiré

Wenjie Fang, Université Paris-Est Marne-la-Vallée  
Travaux joints avec Andrew Elvey Price et Michael Wallner

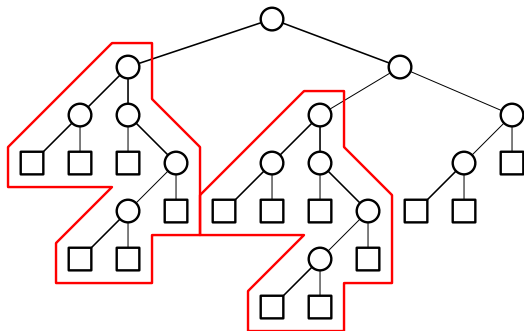
09 avril 2019, Séminaire Combinatoire, LIX, École polytechnique

# Compacter un arbre binaire



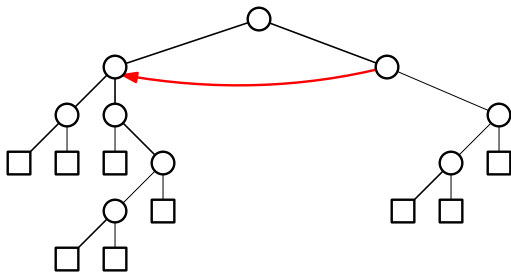
On essaie de compresser un arbre binaire ...

# Compacter un arbre binaire



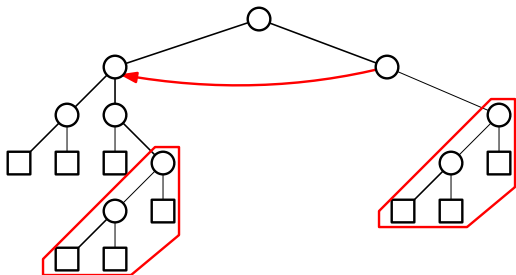
... en identifiant les sous-arbres identiques ...

# Compacter un arbre binaire



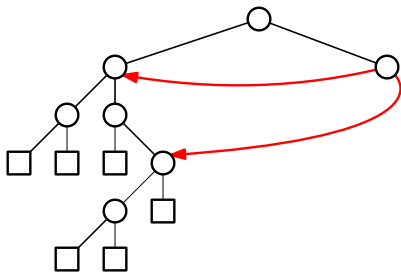
... et en les stockant une seule fois ...

# Compacter un arbre binaire



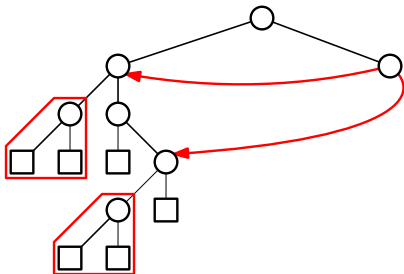
... en identifiant les sous-arbres identiques ...

# Compacter un arbre binaire



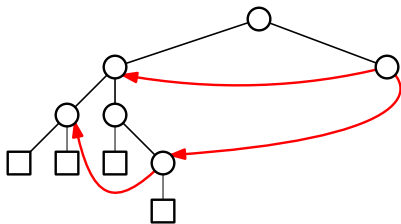
... et en les stockant une seule fois ...

# Compacter un arbre binaire



... en identifiant les sous-arbres identiques ...

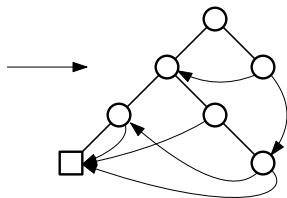
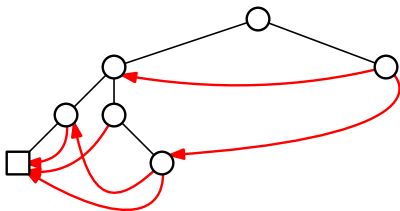
# Compacter un arbre binaire



... et en les stockant une seule fois ...

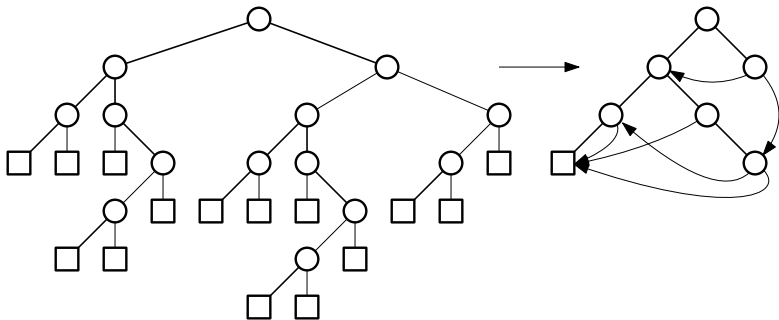


# Compacter un arbre binaire



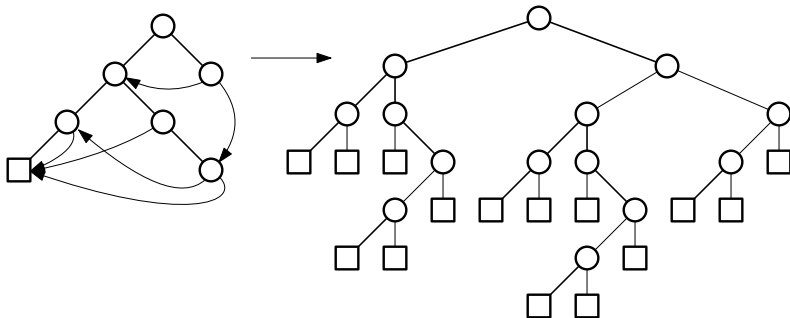
... et en les stockant une seule fois ...

# Compacter un arbre binaire



Les **arbres compactés** sont les arbres avec pointeurs ainsi obtenus.

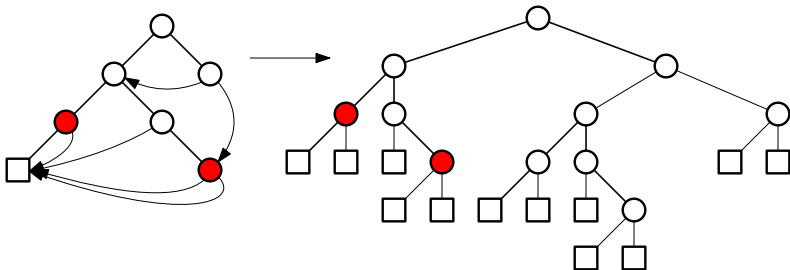
# Arbres compactés



Un **arbre compacté** est un arbre binaire

- dont chaque feuille (sauf la première) est un pointeur, ...
- ... vers un nœud précédent en ordre postfixe,
- dont chaque nœud possède un arbre “décomprimé” unique.

# Arbres relaxés



Un **arbre relaxé** est un arbre binaire

- dont chaque feuille (sauf la première) est un pointeur, ...
- ... vers un nœud précédent en ordre postfixe.
- dont chaque nœud possède un arbre “décomprimé” unique.

# Ce que l'on sait et ce que l'on veut

- (Flajolet, Sipala, Steyaert 1990)
  - Algorithme linéaire pour compacter un arbre binaire de taille  $n$
  - Espérance de la taille compactée :  $O(n/\log n)$
- (Genitrini, Gittenberger, Kauers, Wallner 2019)
  - $n$  nœuds, avec la hauteur à droite  $\leq k$ 
    - Arbres relaxés :

$$\gamma_k n! \left( 4 \cos \left( \frac{\pi}{k+3} \right) \right)^n n^{-k/2}$$

- Arbres compactés :

$$\gamma_k n! \left( 4 \cos \left( \frac{\pi}{k+3} \right) \right)^n n^{-\frac{k}{2} - \frac{1}{k+3} - \left(\frac{1}{4} - \frac{1}{k+3}\right) \frac{1}{\cos^2\left(\frac{\pi}{k+3}\right)}}$$

Et l'asymptotique pour  $n$  nœuds avec  $n \rightarrow \infty$ , **sans d'autres restrictions?**

# Notre résultat

- $c_n$  : nombre des arbres compactés à  $n$  nœuds
- $r_n$  : nombre des arbres relaxés à  $n$  nœuds

## Théorème (Elvey Price, F., Wallner 2019+)

Pour  $n \rightarrow \infty$ , on a

$$c_n = \Theta\left(n!4^n e^{3a_1 n^{1/3}} n^{3/4}\right), \quad r_n = \Theta\left(n!4^n e^{3a_1 n^{1/3}} n\right).$$

Ici,  $a_1$  est la plus grande racine de la fonction d'Airy  $\text{Ai}(x)$ , définie par  $\text{Ai}''(x) = x\text{Ai}(x)$  et  $\text{Ai}(x) \rightarrow 0$  quand  $x \rightarrow +\infty$ .

**Pas la constante multiplicative !**

Exponentiel étiré :  $e^{3a_1 n^{1/3}}$

Probabilité qu'un arbre relaxé de taille  $n$  est compacté :  $\Theta(n^{1/4})$ .

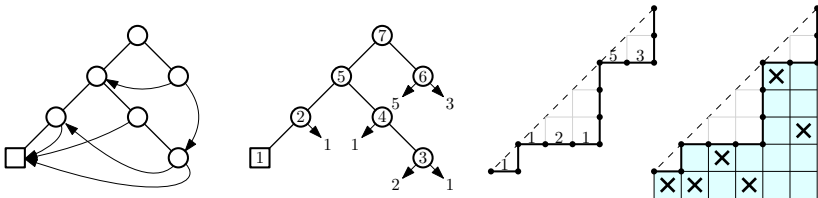
# Comment c'est fait ?

- **Bijection** avec les chemins de Dyck décorés
- **Réurrence** à deux paramètres
- **Heuristique** pour le comportement typique
- **Troncature** de heuristique  $\Rightarrow$  preuve des bornes

Basée seulement sur la récurrence, la méthode est bien simple.

# Encodage chemins de Dyck décorés (relaxé)

Traitons d'abord les **arbres relaxés** :



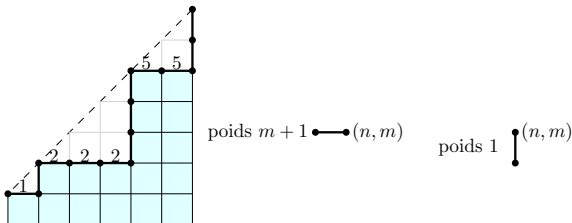
D'un arbre relaxé à un chemin de Dyck décoré :

- Numéroter les nœuds en ordre **postfixe**, et les pointeurs
- Tracer le chemin de Dyck : pointeur  $\Rightarrow H$ , terminant un nœud  $\Rightarrow V$
- Mettre les numéros des pointeurs
- Croiser les cases



# Une récurrence pour les arbres relaxés

Pour un chemin :  $\prod_{i=1}^n (m_i + 1)$ , où  $m_i$  est la coord. du  $i$ -ème pas  $H$



## Proposition

Soit  $r_{n,m}$  le nombre des chemins décorés terminant à  $(n, m)$ . Alors,

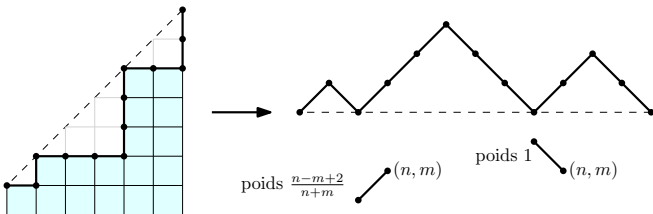
$$\begin{aligned}
 r_{n,m} &= (m + 1)r_{n-1,m} + r_{n,m-1}, & \text{pour } n \geq m \geq 1, \\
 r_{n,m} &= 0, & \text{pour } n < m, \\
 r_{n,0} &= 1, & \text{pour } n \geq 0.
 \end{aligned}$$

Le nombre des *arbres relaxés* à  $n$  nœuds est  $r_{n,n}$ .

# Une transformation

Changement de coordonnées :  $(n, m) \rightarrow (n + m, n - m)$

On divise par  $n!$  en posant  $d_{n+m, n-m} = r_{n, m}/n!$ . Alors  $r_n = n!d_{2n, 0}$ .



Réurrence :

$$d_{n, m} = \frac{n - m + 2}{n + m} d_{n-1, m-1} + d_{n-1, m+1}$$

Plus un pas est élevé, moins est son poids !

# Quelques observations

Pour les [arbres relaxés](#) :

$$r_n = \Theta \left( n! 4^n e^{3a_1 n^{1/3}} n \right)$$

- $4^n$  vient des chemins de Dyck.
- $n!$  vient des poids.

D'où vient l'exponentiel étiré ?

# Une première heuristique

Considérons les chemins de Dyck de taille  $2n$  de hauteur  $\leq n^\alpha$ ,  $\alpha < 1/2$ .

## Proposition (Kousha 2012)

*La probabilité d'avoir un tel chemin par un tirage uniforme vérifie*

$$\log(\mathbb{P}[\text{hauteur} \leq n^\alpha]) \sim -\pi^2 n^{1-2\alpha}$$

Dans ces chemins, un pas montant typique a de poids

$$\frac{\Theta(n) - \Theta(n^\alpha)}{\Theta(n) + \Theta(n^\alpha)} = 1 - \Theta(n^{\alpha-1}).$$

Il y en a typiquement  $\Theta(n)$  tels pas, qui donne en total un poids de

$$(1 - \Theta(n^{\alpha-1}))^{\Theta(n)} = \exp(-\Theta(n^\alpha)).$$

Prenant en compte de la probabilité, la contribution de tels chemins est

$$\exp(-\Theta(n^\alpha) - \Theta(n^{1-2\alpha})),$$

qui est maximisée à  $\alpha = 1/3$ , donnant l'**exponentiel étire**  $\exp(-\Theta(n^{1/3}))$ .

# La bonne mise en échelle

Trop heuristique ... Mais ça indique que la bonne hauteur est  $n^{1/3}$  !

*Ansatz* :

$$d_{n,m} \sim h(n)f(n^{-1/3}(m+1)),$$

$$s(n) = \frac{h(n)}{h(n-1)} = 2 + cn^{-2/3} + O(n^{-1}).$$

- $h(n)$  : croissance générale en  $n$ , environ  $2^n \rho^{n^{1/3}}$  pour un certain  $\rho$
- $f(x)$  : mise en échelle avec hauteur typique  $n^{1/3}$

Supposons que  $m = \kappa n^{1/3} - 1$ .

*Ansatz* + récurrence :

$$f(\kappa)s(n) = \frac{n - \kappa n^{1/3} + 1}{n + \kappa n^{1/3} - 1} f\left((n-1)^{-1/3}(\kappa n^{1/3} - 2)\right) + f\left((n-1)^{-1/3}\kappa n^{1/3}\right)$$

On en déduit :

$$0 = (c + 2\kappa)f(\kappa) - f''(\kappa) + O(n^{-1/3}).$$

# La première estimation

$$0 = (c + 2\kappa)f(\kappa) - f''(\kappa) + O(n^{-1/3}).$$

C'est l'équation de la **fonction d'Airy** !

Comme  $f(\kappa) \rightarrow 0$  pour  $\kappa \rightarrow \infty$ , on a

$$f(\kappa) \approx b \text{Ai} \left( \frac{c + 2\kappa}{2^{2/3}} \right).$$

$f(\kappa) \rightarrow 0$  pour  $\kappa \rightarrow 0 \Rightarrow c = 2^{-2/3}a_1$ .

Par le comportement asymptotique de  $\text{Ai}(x)$  à  $x \rightarrow 0$ , on a :

$$r_n = n!d_{2n,0} = n!4^n \exp \left( 3a_1 n^{1/3} + \dots \right).$$

# Heuristique raffinée

*Ansatz* d'ordre 2 :

$$d_{n,m} \sim h(n) \left( f(n^{-1/3}(m+1)) + n^{-1/3} g(n^{-1/3}(m+1)) \right),$$

$$s(n) = 2 + cn^{-2/3} + dn^{-1} + O(n^{-4/3}).$$

Avec la même méthode, on a le terme polynomial :

$$r_n \approx n!4^n \exp\left(3a_1 n^{1/3}\right) n.$$

*Ansatz* en général :

$$d_{n,m} \approx h(n) \sum_{j=0}^k f_j(n^{-1/3}(m+1)) n^{-j/3},$$

$$s(n) = 2 + \gamma_2 n^{-2/3} + \gamma_3 n^{-1} + \dots + \gamma_k n^{-k/3} + o(n^{-k/3}).$$

Une troncature suffisante nous donne les bornes.

# Gendarmes l'asymptotique

Supposons qu'il existe  $(s_n)_n \geq 0$  et  $(X_{n,m})_{n,m} \geq 0$  tels que

$$X_{n,m} s_n \leq \frac{n-m+2}{n+m} X_{n-1,m-1} + X_{n-1,m+1},$$

pour  $n$  assez grand, pour tout  $0 \leq m \leq n$ .

Soit  $h_n = \prod_{i=1}^n s_i$ , alors  $X_{n,m} h_n \leq b_0 d_{n,m}$  pour une constante  $b_0$ .

Borne inférieure !

En renversant l'inégalité, on a la borne supérieure !



# Borne inférieure - *ansatz* et expansion

Prenons

$$X_{n,m} = \left(1 - \frac{2m^2}{3n} + \frac{m}{2n}\right) \text{Ai} \left(a_1 + \frac{2^{1/3}(m+1)}{n^{1/3}}\right),$$

$$s_n = 2 + \frac{2^{2/3}a_1}{n^{2/3}} + \frac{8}{3n} - \frac{1}{n^{7/6}}.$$

La différence est

$$P_{n,m} = -X_{n,m}s_n + \frac{n-m+2}{n+m}X_{n-1,m-1} + X_{n-1,m+1}.$$

La zone au-delà  $n^{2/3-\varepsilon}$  est négligeable.

Il suffit de prouver  $P_{n,m} \geq 0$  pour  $m < n^{2/3-\varepsilon}$ .

Alors par substitution et développement en  $n$ , on a

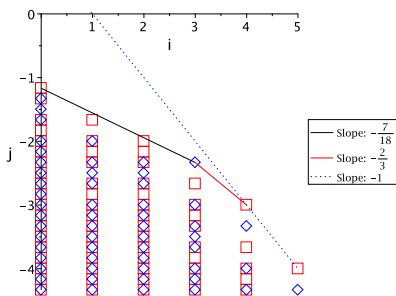
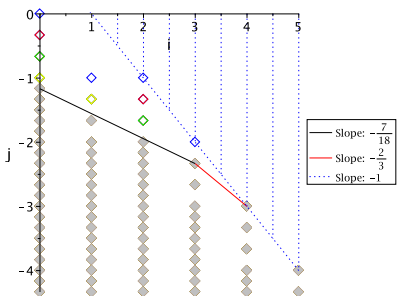
$$P_{n,m} = p_0(n, m)\text{Ai}(\alpha) + p_1(n, m)\text{Ai}'(\alpha), \text{ avec } \alpha = a_1 + \frac{2^{1/3}m}{n^{1/3}}.$$

$p_0(n, m), p_1(n, m)$  : séries de  $n^{-1/6}$  à coeffs polynomiaux en  $m$ .

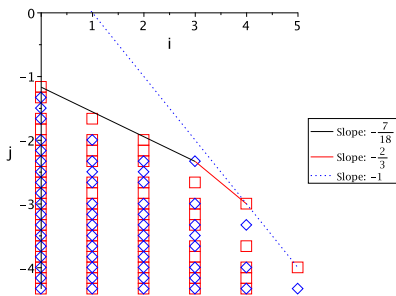
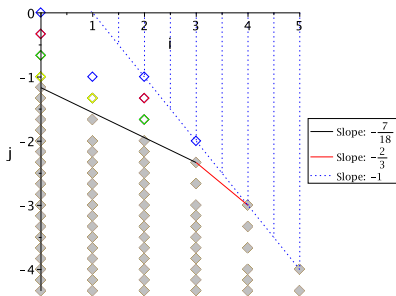
## Borne inférieure - polygone

$$P_{n,m} = \text{Ai}(\alpha) \left( -\frac{\sigma_4}{n^{7/6}} - \frac{2^{5/3}a_1m}{3n^{5/3}} - \frac{41m^2}{9n^2} - \frac{2^{8/3}a_1m^3}{3n^{8/3}} - \frac{34m^4}{9n^3} + \dots \right) +$$

$$\text{Ai}'(\alpha) \left( \frac{2^{1/3}}{n^{3/2}} - \frac{8a_1m}{9n^2} - 19\frac{2^{1/3}m^2}{9n^{7/3}} - \frac{2^{13/3}m^3}{9n^{7/3}} + \dots \right).$$



# Borne inférieure - analyse par cas



- $m \leq x_0(n/2)^{1/3}$ , où  $Ai(a_1 + x)Ai'(a_1 + x)$  change de signe,
- $x_0(n/2)^{1/3} < m \leq n^{7/18}$ ,
- $n^{7/18} < m < n^{2/3-\epsilon}$ .

En travaillant sur la fonction d'Airy, on arrive à montrer tout est négatif.

# Borne supérieure

Identique, avec un ansatz différent :

$$\hat{X}_{n,m} = \left( 1 - \frac{2m^2}{3n} + \frac{m}{2n} + \frac{3m^4}{10n^2} \right) \text{Ai} \left( a_1 + \frac{2^{1/3}(m+1)}{n^{1/3}} \right),$$
$$\hat{s}_n = 2 + \frac{2^{2/3}a_1}{n^{2/3}} + \frac{8}{3n} + \frac{1}{n^{7/6}}.$$

Encore une discussion des cas...

$$r_n = \Theta \left( n! 4^n e^{3a_1 n^{1/3}} n \right).$$

# Lemme de cerise

Sur les arbres compactés, on a cette observation :



## Lemme (de cerise)

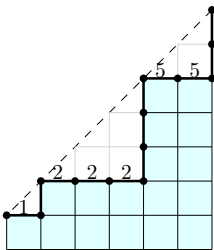
*Pour un arbre relaxé  $T$ , si les cerises  ne répètent pas les nœuds précédents, alors  $T$  est compacté.*


$T$  n'est pas compacté  $\Rightarrow$  deux nœuds  $u, v$  avec les mêmes arbres décomprimés, donc aussi pour leurs enfants correspondants


Descente jusqu'à une cerise

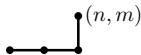
# Encodage chemins de Dyck décorés (compacté)

Lemme de cerise   $\Rightarrow$  éviter certaines décorations de 



poids  $m + 1$    $(n, m)$

poids 1   $(n, m)$



retirer  $(m - 1)$  possibilités

## Proposition

Soit  $c_{n,m}$  le nombre des chemins décorés stricts à  $(n, m)$ . Alors,

$$c_{n,m} = (m + 1)c_{n-1,m} + c_{n,m-1} - (m - 1)c_{n-2,m-1}, \quad \text{pour } n \geq m \geq 1,$$

$$c_{n,m} = 0, \quad \text{pour } n < m,$$

$$c_{n,0} = 1, \quad \text{pour } n \geq 0.$$

Le nombre des **arbres compactés** à  $n$  nœuds est  $c_n = c_{n,n}$ .

# Arbres compactés

Récurrance pour les arbres compactés :

$$e_{n,m} = \frac{n-m+2}{n+m} e_{n-1,m-1} + e_{n-1,m+1} - \frac{2(n-m-2)}{(n+m)(n+m-2)} e_{n-3,m-1}.$$

Terme **néga**tif, pas très bien...

Mais on peut le borner par deux récurrances à terme positif.

Avec deux *ansätze* appropriés, on a

$$c_n = \Theta \left( n! 4^n e^{3a_1 n^{1/3}} n^{3/4} \right).$$

# Un changement de terme polynomial

*Ansatz* pour la borne inférieure :

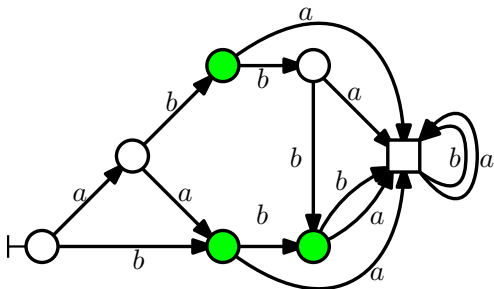
$$\hat{X}_{n,m} = \left(1 - \frac{2m^2}{3n} + \frac{m}{4n}\right) \text{Ai} \left(a_1 + \frac{2^{1/3}(m+1)}{n^{1/3}}\right),$$
$$\hat{s}_n = 2 + \frac{2^{2/3}a_1}{n^{2/3}} + \frac{13}{6n} - \frac{1}{n^{7/6}}.$$

Seule différence en  $\hat{s}_n \Rightarrow$  changement du terme polynomial



# Automate minimal d'un langage fini

Un langage  $\Rightarrow$  un **unique** automate minimal

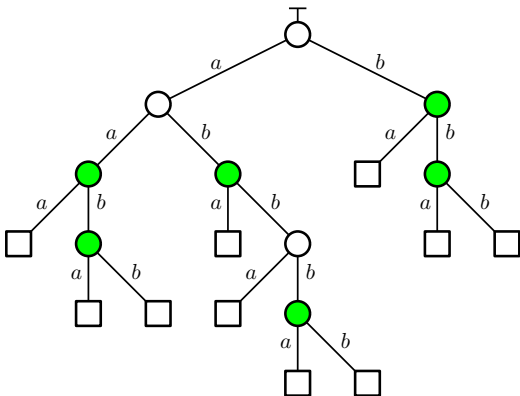


Nb d'états de l'automate minimal  $\approx$  complexité du langage

**Question** : combien d'automates minimaux à  $n$  états reconnaissant un **language fini** ?

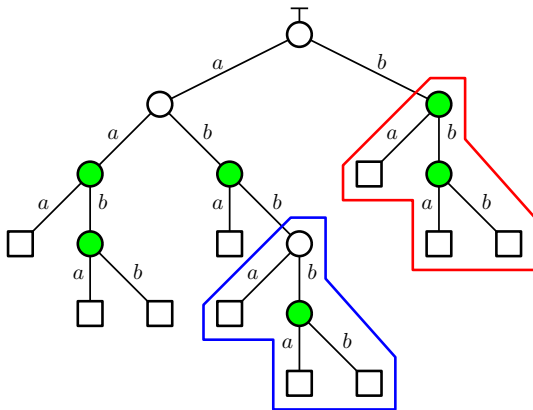
Très "arbres compactés" !

# Minimiser un trie



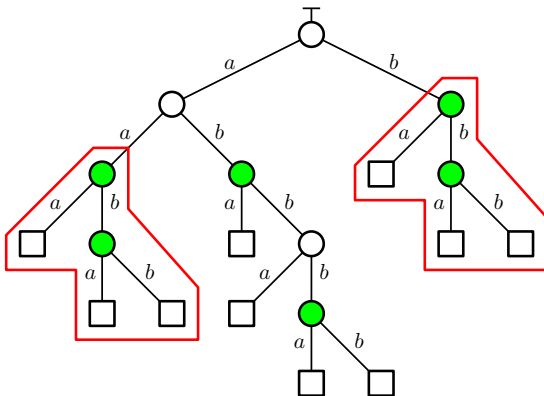
Prenons un trie pour le compacter ...

# Minimiser un trie



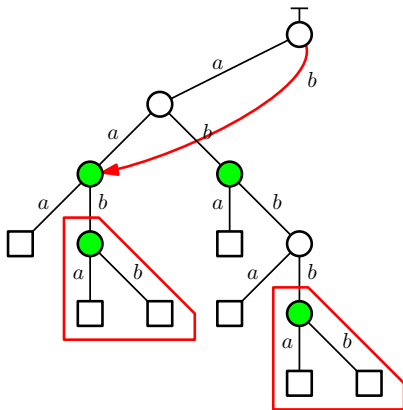
... avec l'arbre et les couleurs identiques ...

# Minimiser un trie



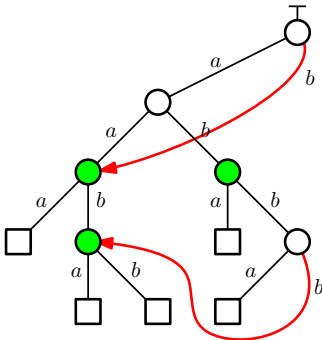
... avec l'arbre et les couleurs identiques ...

# Minimiser un trie



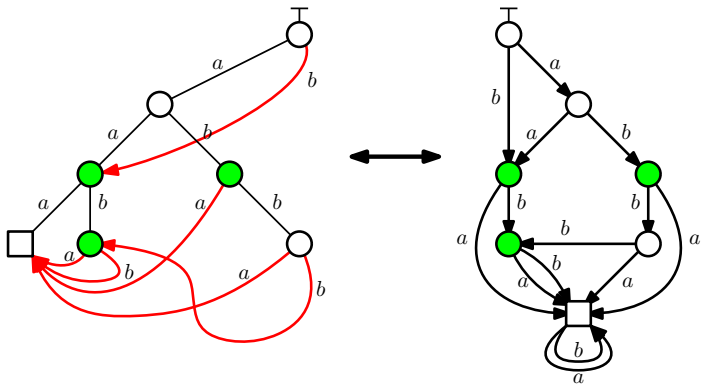
... en épuisant tous les possibilités ...

# Minimiser un trie



... en épuisant tous les possibilités ...

# Minimiser un trie



... et on arrive à un automate minimal.

# Énumération des automates minimaux des langages finis

Fixons  $A = \{a, b\}$  de taille 2 (autres tailles possibles).

**Théorème (Elvey Price, F., Wallner 2020+)**

*Le nombre  $m_{2,n}$  d'automates à  $n$  états minimaux pour un langage fini à  $A = \{a, b\}$  est asymptotiquement*

$$m_{2,n} = \Theta\left(n! 8^n e^{3a_1 n^{1/3}} n^{7/8}\right).$$

- Codage par chemins de Dyck décorés
- Un “lemme de cerise” !
- Récurrence similaire
- Méthode identique



# Résumé

## Ce qui est bien :

- À partir de seulement une récurrence (assez simple) ;
- Sans trop d'étude de la fonction génératrice ;
- Relativement simple, donc potentiellement généralisable.

## À travailler :

- Quel type de récurrence ? Les termes négatifs éventuel ...
- Il faut quand même partir d'une heuristique ...
- Et la constante multiplicative.

## Déjà une application !

Michael Fuchs, Guan-Ru Yu, Louxin Zhang, *On the Asymptotic Growth of the Number of Tree-Child Networks*, arXiv 2003.08049.

# Résumé

## Ce qui est bien :

- À partir de seulement une récurrence (assez simple) ;
- Sans trop d'étude de la fonction génératrice ;
- Relativement simple, donc potentiellement généralisable.

## À travailler :

- Quel type de récurrence ? Les termes négatifs éventuel ...
- Il faut quand même partir d'une heuristique ...
- Et la constante multiplicative.

## Déjà une application !

Michael Fuchs, Guan-Ru Yu, Louxin Zhang, *On the Asymptotic Growth of the Number of Tree-Child Networks*, arXiv 2003.08049.

**Merci de votre attention !**