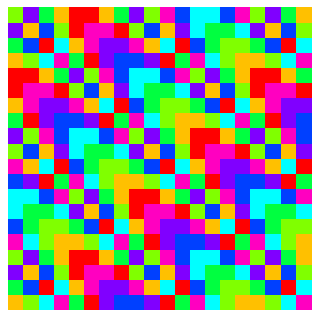
Coefficients binomiaux de mots

Michel Rigo

http://www.discmath.ulg.ac.be/
16 septembre 2012

http://orbi.ulg.ac.be/handle/2268/201779



La notion de coefficient binomial de mots est classique en COW. Voir, par exemple, Sakarovitch & Simon, Lothaire.

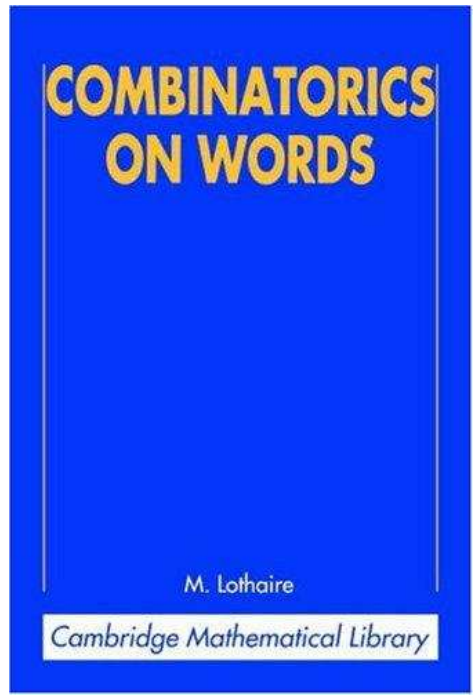
$\binom{w}{x}$ nombre de fois où x apparaît comme sous-mot de w

i.e., x apparaît comme sous-suite de w

On compte les applications
$\varphi : \{1, \ldots, |x|\} \to \{1, \ldots, |w|\}$ telles que

$$\varphi(1) < \cdots < \varphi(|x|)$$

$$w_{\varphi(1)} \cdots w_{\varphi(|x|)} = x$$



$$\binom{aabbab}{ab} = 7$$

Généralise la notion usuelle de coefficient binomial d'entiers

$$\binom{a^m}{a^n} = \binom{m}{n}, \quad m, n \in \mathbb{N}$$

On a $\binom{w}{a} = |w|_a, \quad a \in A$

Ces coefficients se calculent aisément :

$$\binom{w}{\varepsilon} = 1, \qquad \binom{w}{x} = 0, \quad \text{if } |w| < |x|$$

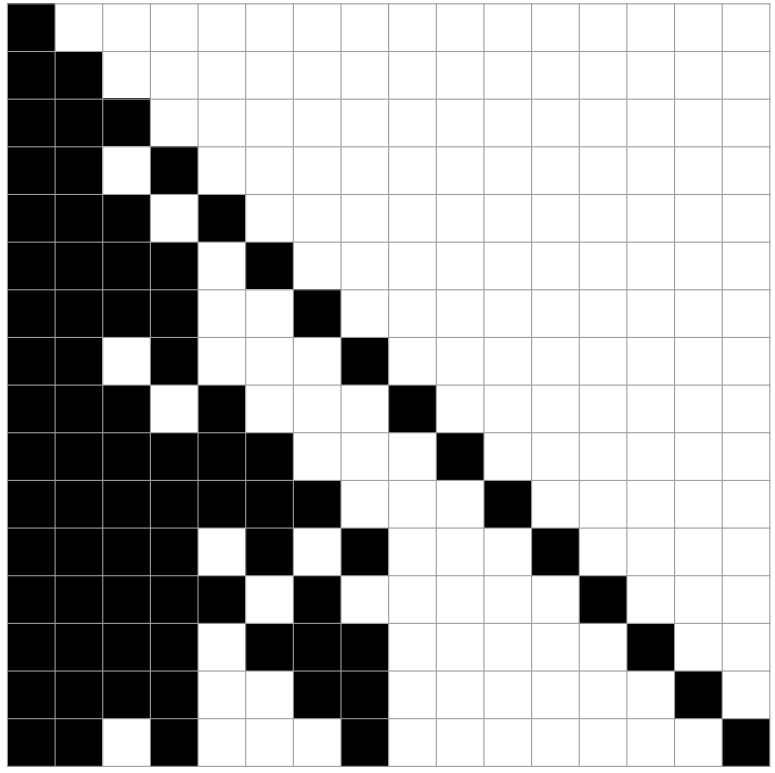
$$u, v \in A^*, a, b \in A, \quad \binom{ua}{vb} = \binom{u}{vb} + \delta_{a,b}\binom{u}{v}$$

```
coeff[u_, v_] := coeff[u, v] =
   If[Length[v] == 0, 1,
    If[Length[u] < Length[v], 0,
     coeff[Drop[u, -1], v]
      + ((Last[u] == Last[v]) /. {True -> 1, False -> 0})
         coeff[Drop[u, -1], Drop[v, -1]]
      ]
     ]
```

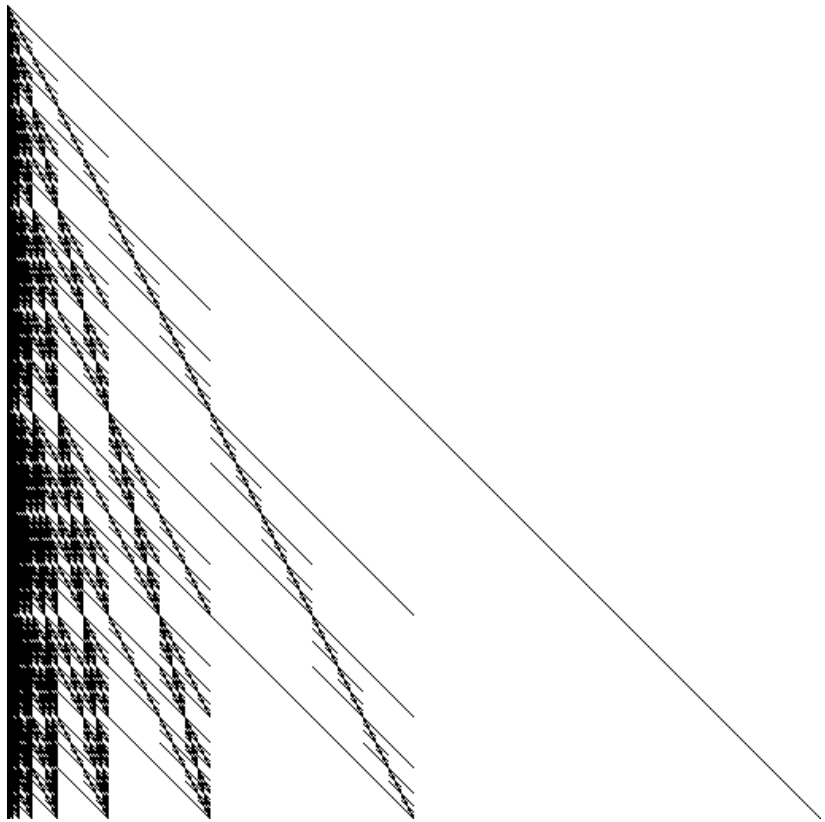
	ε	1	10	11	100	101	110	111	1000	$\cdots$	1100
ε											
1											
10											
11											
100											
101											
110				1			1				
111											
1000											
$\vdots$											
1100							2				
$\vdots$											

$$\binom{ua}{vb} = \binom{u}{vb} + \delta_{a,b}\binom{u}{v}$$

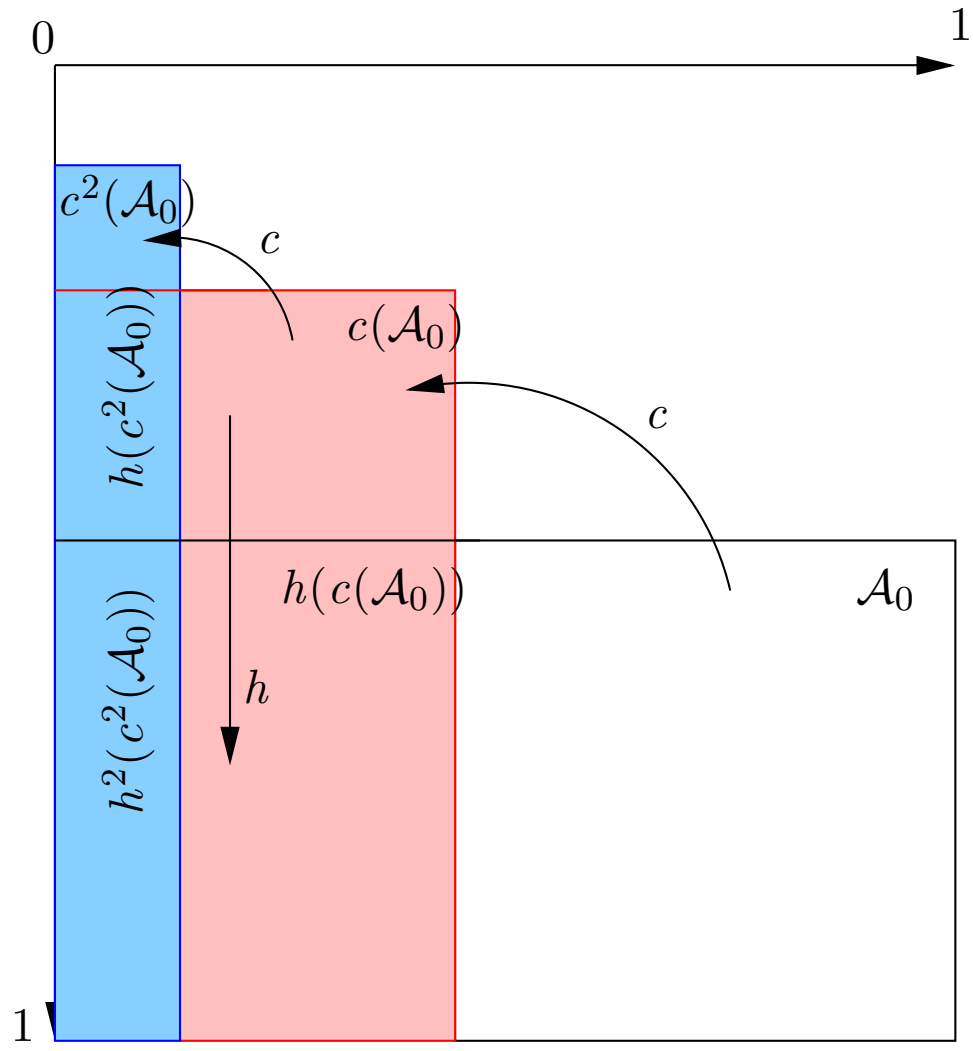
	ε	1	10	11	100	101	110	111
ε	**1**	0	0	0	0	0	0	0
1	**1**	**1**	0	0	0	0	0	0
10	1	1	1	0	0	0	0	0
11	**1**	**2**	0	**1**	0	0	0	0
100	1	1	2	0	1	0	0	0
101	1	2	1	1	0	1	0	0
110	1	2	2	1	0	0	1	0
111	**1**	**3**	0	**3**	0	0	0	**1**



Triangle de Pascal (modulo 2) généralisé à $1\{0,1\}^*$



J. Leroy, M. R., M. Stipulanti, Generalized Pascal triangle for binomial coefficients of words, *Adv. in Appl. Math.* **80** (2016), 24—47.



c homothétie de centre $(0,0)$ et rapport $1/2$; $h : (x,y) \mapsto (x,2y)$.

Definition

Soit $k \geq 1$. Deux mots u, v sont *k-binomialement équivalents* $u \equiv_k v$ si et seulement si

$$\binom{u}{x} = \binom{v}{x} \quad \forall x \in A^{\leq k}.$$

Remarque : équivalence 1-binomiale $=$ *équivalence abélienne*.

On trouve aussi la notion de *k-spectre* d'un mot u.
C'est un polynôme (formel) de $\mathbb{N}\langle A^* \rangle$ de degré k

$$\mathrm{Spec}_{u,k} = \sum_{x \in A^{\leq k}} \binom{u}{x} x.$$

Deux mots u, v sont k-binomialement équivalents SSI ils ont le même k-spectre. $\rightsquigarrow$ information complète.

Exemple

Le 2-spectre du mot $u = abbab$ est

$$\mathrm{Spec}_{u,2} = 1\varepsilon + 2a + 3b + aa + 4ab + 2ba + 3bb.$$

Le 3-spectre de ce mot est

$$\mathrm{Spec}_{u,3} = \mathrm{Spec}_{u,2} + aab + 2aba + 3abb + 2bab + bba + bbb.$$

Notez que le k-spectre contient

$$\frac{(\#A)^{k+1} - 1}{(\#A) - 1} \text{ coefficients (éventuellement nuls).}$$

$\rightsquigarrow$ croissance exponentielle en k.

$$2 + 3 = \binom{5}{1},\ 1 + 4 + 2 + 3 = \binom{5}{2},\ 1 + 2 + 3 + 2 + 1 + 1 = \binom{5}{3}$$

$ababbba$, $abbabab$, $baabbab$, $babaabb$ sont 2-binomialement équivalents

$$|w|_a = 3, \quad |w|_b = 4, \quad \binom{w}{aa} = \binom{3}{2} = 3, \quad \binom{w}{bb} = \binom{4}{2} = 6$$

$$\binom{w}{ab} = 7, \quad \binom{w}{ba} = 5$$

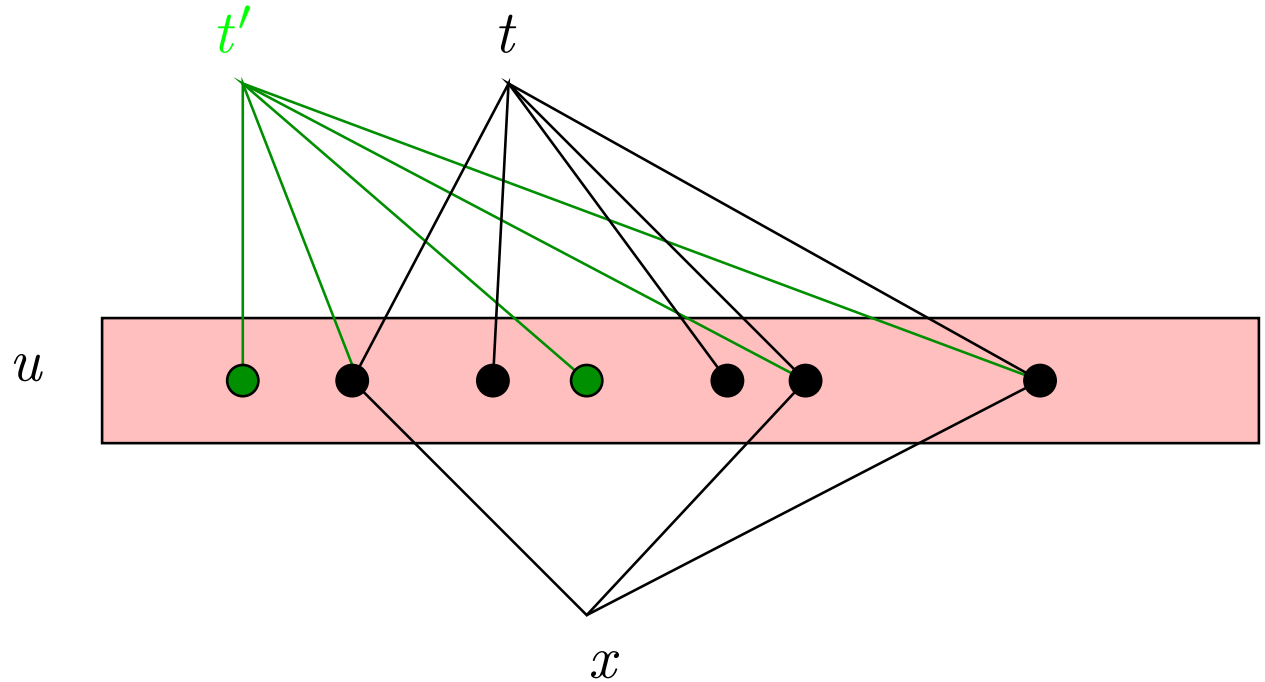
mais pas 3-binomialement équivalents

$$\binom{ababbba}{aab} = 3, \quad \binom{abbabab}{aab} = 4.$$

M. Dudik, L. J. Schulman 2003

Si $|u| \geq k \geq |x|$, alors on a

$$\binom{|u|-|x|}{k-|x|}\binom{u}{x} = \sum_{t \in A^k} \binom{u}{t}\binom{t}{x}.$$



problèmes de reconstruction

$u \equiv_k v$ si et seulement si

$$\binom{u}{x} = \binom{v}{x} \quad \forall x \in A^{\leq k}.$$

Corollaire

Soient $u, v \in A^{\geq k}$. On a $u \equiv_k v$ SSI $\binom{u}{t} = \binom{v}{t}$ pour tout mot t de longueur k.

En COW, il existe un zoologie de relations d'équivalence :

- équivalence abélienne (depuis Erdős en 1961)

$$abbacba \sim_{ab} cababba$$

- équivalence k-abélienne (Karhumäki, Saarela, Zamboni 2013)

$$|u|_x = |v|_x \quad \forall x \in A^{\leq k}$$

- équivalence cyclique ou en termes de sous-groupes de permutations (Cassaigne 2014, Charlier, Puzynina, Zamboni 2015)
- équivalence k-binomiale
- (Parikh) matrix equivalence (Salomaa *et al.* 2000)
- congruence de Simon (1975, Karandikar, Schnoebelen 2015)

$$Supp(\mathrm{Spec}_{u,k}) = Supp(\mathrm{Spec}_{v,k})$$

applications aux "piecewise testable languages"

Liens avec les matrices de Parikh.

$A = \{a_1, \ldots, a_k\}$. Le "*Parikh matrix mapping*"

$$\psi_k : A^* \to \mathbb{N}^{(k+1)\times(k+1)}$$

est le morphisme défini par la condition :
si $\psi_k(a_q) = (m_{i,j})_{1\leq i,j\leq k+1}$, alors pour tout $i \in \{1, \ldots, k+1\}$,

$$m_{i,i} = 1, \quad m_{q,q+1} = 1,$$

tous les autres éléments de la matrice $\psi_k(a_q)$ étant 0.

DÉFINITION

Deux mots sont M*-équivalents*, ou *matrice-équivalents*,
s'ils possèdent la même matrice de Parikh.

Exemple, $\#A = 2$

Considérons $A = \{a, b\}$. On a

$$\psi_2(a) = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \ \psi_2(b) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

et

$$\psi_2(abbab) = \psi_2(a)\psi_2(b)\psi_2(b)\psi_2(a)\psi_2(b) = \begin{pmatrix} 1 & 2 & 4 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix}.$$

Les matrices de Parikh, pour un alphabet de taille k, encodent

$$k(k+1)/2$$

coefficients binomiaux d'un mot w pour les sous-mots de longueur $\leq k$.

Théorème (A. Mateescu, A. Salomaa, K. Salomaa, S. Yu 2001)

Soit $A = \{a_1, \ldots, a_k\}$ un alphabet ordonné.
Soient w un mot fini et $\psi_k(w) = (m_{i,j})_{1 \leq i,j \leq k+1}$.
Alors

$$m_{i,j+1} = \binom{w}{a_i \cdots a_j}$$

pour tous i, j tels que $1 \leq i \leq j \leq k$.

$\rightsquigarrow$ information partielle : $\mathcal{O}(k^2)$ vs. $\Omega((\#A)^k)$

Exemple sur $A = \{a, b, c\}$

$$\psi_3(w) = \begin{pmatrix} 1 & \binom{w}{a} & \binom{w}{ab} & \binom{w}{abc} \\ 0 & 1 & \binom{w}{b} & \binom{w}{bc} \\ 0 & 0 & 1 & \binom{w}{c} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\psi_3(wb) = \begin{pmatrix} 1 & \binom{w}{a} & \binom{w}{ab} & \binom{w}{abc} \\ 0 & 1 & \binom{w}{b} & \binom{w}{bc} \\ 0 & 0 & 1 & \binom{w}{c} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Par exemple,

$$\binom{wb}{ab} = \binom{w}{a} + \binom{w}{ab}$$

Aussi, les *matrices de Parikh* **généralisées** ψ_u à tout mot $u \in A^*$,

Soit $u = u_1 \cdots u_\ell$.
Si $\psi_u(a) = (m_{i,j})_{1 \leq i,j \leq \ell+1}$,
alors pour tout $i \in \{1, \ldots, \ell+1\}$, $m_{i,i} = 1$,
et pour tout $i \in \{1, \ldots, \ell\}$,

$$m_{i,i+1} = \delta_{a,u_i},$$

les autres éléments de la matrice $\psi_u(a)$ étant nuls.

Remarque

On retrouve les matrices de Parikh 'classiques' avec

$$u = a_1\, a_2 \cdots a_k$$

si $A = \{a_1, \ldots, a_k\}$.

Il vient

$$\psi_{abba}(a) = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \ \psi_{abba}(b) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Généralisation naturelle du théorème de Mateescu et al.

Théorème (Şerbănuţă 2004)

Soient $u = u_1 \cdots u_\ell$ et w un mot. Soit $\psi_u(w) = (m_{i,j})_{1 \leq i,j \leq \ell+1}$. Alors, pour tout $1 \leq i \leq j \leq \ell$,

$$m_{i,j+1} = \binom{w}{u_i \cdots u_j}.$$

En particulier, la première ligne de $\psi_u(w)$ contient les coefficients correspondant aux préfixes de u :

$$\binom{w}{\varepsilon}, \binom{w}{u_1}, \binom{w}{u_1 u_2}, \ldots, \binom{w}{u_1 \cdots u_{\ell-1}}, \binom{w}{u}.$$

De même, la dernière colonne de $\psi_u(w)$ contient les coefficients correspondant aux suffixes de u :

$$\binom{w}{u}, \binom{w}{u_2 \cdots u_\ell}, \ldots, \binom{w}{u_1}, \binom{w}{\varepsilon}.$$

Exemple

$$\psi_{abba}(w) = \begin{pmatrix} 1 & \binom{w}{a} & \binom{w}{ab} & \binom{w}{abb} & \binom{w}{abba} \\ 0 & 1 & \binom{w}{b} & \binom{w}{bb} & \binom{w}{bba} \\ 0 & 0 & 1 & \binom{w}{b} & \binom{w}{ba} \\ 0 & 0 & 0 & 1 & \binom{w}{a} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Lien entre l'équivalence k-binomiale et la M-équivalence

Proposition

Pour un alphabet binaire, deux mots sont 2-binomialement équivalents SSI ils ont la même matrice de Parikh.

$$\psi_2(w) = \begin{pmatrix} 1 & \binom{w}{a} & \binom{w}{ab} \\ 0 & 1 & \binom{w}{b} \\ 0 & 0 & 1 \end{pmatrix}$$

$\Rightarrow$ clair !

$\Leftarrow$

$$\binom{w}{aa} = \binom{|w|_a}{2}$$

$$\binom{w}{aa} + \binom{w}{ab} + \binom{w}{ba} + \binom{w}{bb} = \binom{|w|}{2}$$

Malheureusement, on n'a pas mieux.

Deux mots sur $\{a, b, c\}$,

$$u = abcbabcbabcbab \text{ and } v = bacabbcabbcbba$$

- non 3-binomialement équivalents : $\binom{u}{abb} = 34$ et $\binom{v}{abb} = 36$,
- MAIS même matrice de Parikh $\psi_3(u) = \psi_3(v)$.

Note : ils n'ont pas la même matrice de Parikh *généralisée*

$$\psi_{abb}(u) \neq \psi_{abb}(v).$$

En effaçant les c's, on obtient deux mots sur $\{a, b\}$

$$u' = abbabbabbab \text{ et } v' = baabbabbbba$$

- non 3-binomialement équivalents : $\binom{u'}{abb} = 34$, $\binom{v'}{abb} = 36$
- MAIS même matrice de Parikh

$$\begin{pmatrix} 1 & 4 & 16 \\ 0 & 1 & 7 \\ 0 & 0 & 1 \end{pmatrix}$$

En effet, l'équivalence 3-binomiale est un rafinement de l'équivalence 2-binomiale.

Enfin, deux mots sur $\{a, b, c\}$

$$u = bccaa \text{ et } v = cacab$$

- non 2-binomialement équivalents : $\binom{u}{ca} = 4$ et $\binom{v}{ca} = 3$,
- MAIS avec la même matrice de Parikh $\psi_3(u) = \psi_3(v)$.

$$\begin{pmatrix} 1 & 2 & 0 & 0 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Théorème (A. Salomaa 2010)

Sur un alphabet binaire A, deux mots ont la même matrice de Parikh SSI l'un s'obtient à partir de l'autre par une suite finie de transformations de la forme

$$x\,ab\,y\,ba\,z \to x\,ba\,y\,ab\,z$$

où $a, b \in A$ et $x, y, z \in A^*$.

Valide aussi pour l'équivalence 2-binomiale.

$$1011001001011 \equiv_2 1101001000111 \equiv_2 1100110000111$$

$$\#[0\cdots01\cdots1]_{\equiv_2} = 1$$

$$\#(\{a,b\}^n/\equiv_2) \;=\; \frac{n^3+5n+6}{6}$$

Remarque

Si $x \equiv_{k-1} y$, alors

$$p\textcolor{red}{x}q\textcolor{red}{y}r \equiv_k p\textcolor{red}{y}q\textcolor{red}{x}r$$

Cependant, il n'est pas clair que le résultat précédent puisse être généralisé

Sur 3 lettres :

$$2100221 \equiv_2 0221102$$

mais 2100221 ne peut pas être factorisé en $pxqyr$ avec $x \equiv_{ab} y$.

Questions

La notion d'évitement est classique en COW (Thue début XX^e).

- $\#A = 2$, tout mot de longueur ≥ 4 contient un carré uu
- $\#A = 2$, les cubes (et les chevauchements) peuvent être évités

$$abbabaabbaababbabaababbaabbabaab\cdots$$

- $\#A = 3$, les carrés peuvent être évités

$$(abb)(ab)(a)(abb)(a)(ab)(abb)(ab)(a)(ab)(abb)(a)(abb)(ab)\cdots$$

$$0 \mapsto 012,\ 1 \mapsto 02,\ 2 \mapsto 1$$

- $\#A = 3$, les carrés abéliens sont inévitables
- $\#A = 4$, les carrés abéliens peuvent être évités (V. Keränen)
- $\#A = 3$, les cubes abéliens peuvent être évités (F. M. Dekking)

Questions

On définit un carré 2-binomial uv où $u \equiv_2 v$

"carré abélien $\prec$ carré 2-binomial $\prec$ $\cdots$ $\prec$ carré"

- les carrés sont évitables sur 3 lettres
- les carrés abéliens sont évitables sur 4 lettres

$\rightsquigarrow$ les carrés 2-binomiaux sont-ils évitables sur 3 lettres ?

$$0 \mapsto 012,\ 1 \mapsto 02,\ 2 \mapsto 1$$

Remarque : les carrés k-binomiaux sont évitables sur 3 lettres, $\forall k \geq 2$.

Questions

On définit un carré 2-binomial uv où $u \equiv_2 v$

$$\text{“carré abélien } \prec \text{ carré 2-binomial } \prec \ \cdots \ \prec \text{ carré”}$$

- les carrés sont évitables sur 3 lettres
- les carrés abéliens sont évitables sur 4 lettres

$\rightsquigarrow$ les carrés 2-binomiaux sont-ils évitables sur 3 lettres ?

$$0 \mapsto 012,\ 1 \mapsto 02,\ 2 \mapsto 1$$

Remarque : les carrés k-binomiaux sont évitables sur 3 lettres, $\forall k \geq 2$.

Questions

On définit un cube 2-binomial uvw où $u \equiv_2 v$, $v \equiv_2 w$

$$abbabaabbaab$$

"cube abélien $\prec$ cube 2-binomial $\prec$ $\cdots$ $\prec$ cube"

- les cubes sont évitables sur 2 lettres
- les cubes abéliens sont évitables sur 3 lettres

$\rightsquigarrow$ les cubes 2-binomiaux sont-ils évitables sur 2 lettres ?

$$0 \mapsto 001,\ 1 \mapsto 011$$

M. Rao, M. R., P. Salimov, Avoiding 2-binomial squares and cubes, *Theoret. Comput. Sci.* **572** (2015), 83–91.

Questions

On définit un cube 2-binomial uvw où $u \equiv_2 v$, $v \equiv_2 w$

$$abbabaabbaab$$

"cube abélien $\prec$ cube 2-binomial $\prec$ $\cdots$ $\prec$ cube"

- les cubes sont évitables sur 2 lettres
- les cubes abéliens sont évitables sur 3 lettres

$\rightsquigarrow$ les cubes 2-binomiaux sont-ils évitables sur 2 lettres ?

$$0 \mapsto 001,\ 1 \mapsto 011$$

M. Rao, M. R., P. Salimov, Avoiding 2-binomial squares and cubes, *Theoret. Comput. Sci.* **572** (2015), 83–91.

Questions

Sakarovitch et Simon demandaient déjà d'avoir une meilleure connaissance de $\#(A^n/\sim_k)$ où $\sim_k$ est la congruence de Simon.

- Etant donnés $k \geq 1$ et deux mots u, v de longueur n

  decider, en temps polynomial en n, k, si $u \equiv_k v$.

- Etant donnés $k \geq 1$ et deux mots w, x

  trouver, en temps polynomial, toutes les occurrences des facteurs de w qui sont k-binomialement équivalents à x.

- Etant donnés deux mots u, v de longueur n,

  trouver le plus grand k tel que $u \equiv_k v$.

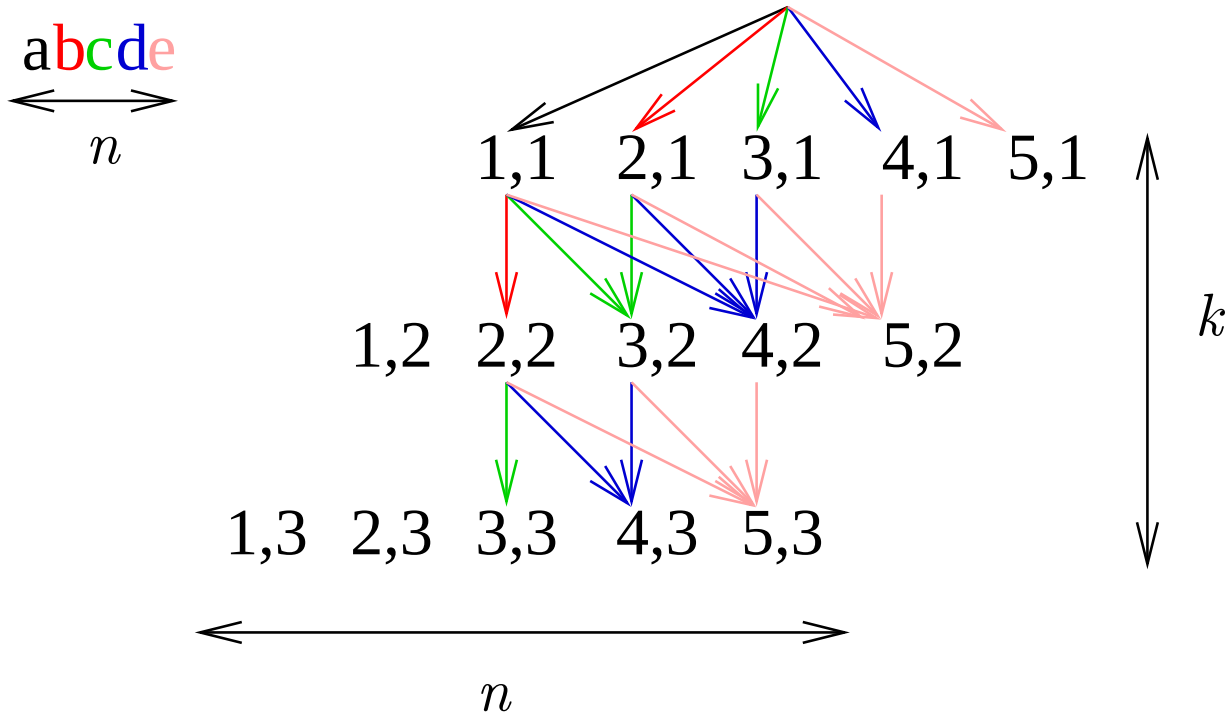
cf. aussi *k-abelian pattern matching*, T. Ehlers, F. Manea, R. Mercas, D. Nowotka, DLT 2014. (en temps linéaire)

Idées principales du papier
'Testing k-binomial equivalence'
`arXiv:1509.00622`
D. Freydenberger *et al.*

On considérera uniquement la première question.

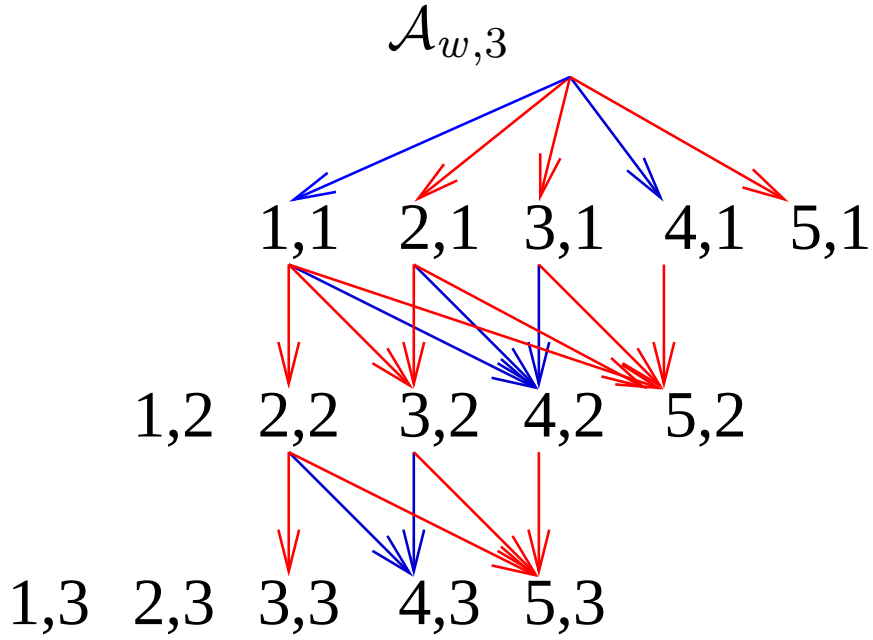
Première réponse, étant donnés un mot w de longueur n et un entier k

$\rightsquigarrow$ construire un AFND $\mathcal{A}_{w,k}$ ayant $nk+1$ états



- tous les états sont finals,
- accepte exactement les sous-mots de w de longueur $\leq k$
- un sous-mot x est accepté $\binom{w}{x}$ fois !

$w = abbab,\ k = 3$



$$\binom{w}{abb} = 3, \quad \binom{w}{ba} = 2$$

Deux automates sont **équivalents** s'ils
acceptent le *même langage* avec *les mêmes multiplicitiés*.

Etant donnés deux mots u, v

- construire $\mathcal{A}_{u,k}$ et $\mathcal{A}_{v,k}$
- $u \equiv_k v$ se réduit à '$\mathcal{A}_{u,k}$ et $\mathcal{A}_{v,k}$ sont-ils équivalents ?'

W. Tzeng, SIAM J. Computing 1992

$\rightsquigarrow$ algorithme polynomial, au moins en n^3...

Du résumé du papier de Tzeng :

Two probabilistic automata are equivalent if for any string x, the two automata accept x with equal probability. This paper presents an $\mathcal{O}((n_1 + n_2)^4)$ algorithm for determining whether two probabilistic automata U_1 and U_2 are equivalent, where n_1 and n_2 are the number of states in U_1 and U_2, respectively.

• S. Kiefer, A. S. Murawski, et al. *On the complexity of the equivalence problem for probabilistic automata*, LNCS **7213** (2012), 467–481.

• M.-P. Schützenberger, *On the definition of a family of automata*, Inf. and Control, 245–270, 1961. (minimisation d'automates pondérés)

Seconde réponse, un algorithme probabiliste

Définition

Soient un mot $w \in \{0,1\}^*$ de longueur n et un entier k,

$$Q_{w,k}(X) := \sum_{v \in A^{\leq k}} \binom{w}{v} X^{val_2(1v)}$$

$$Q_{0010,2}(X) = X + 3X^2 + X^3 + 3X^4 + X^5 + X^6$$

Comme le k-spectre, contient l'information complète.

Exemple

Le 2-spectre du mot $abbab$ est

$$1\underbrace{\varepsilon}_{1}+2\underbrace{a}_{10}+3\underbrace{b}_{11}+\underbrace{aa}_{100}+4\underbrace{ab}_{101}+2\underbrace{ba}_{110}+3\underbrace{bb}_{111}.$$

$$Q_{01101,2}(X) = X + 2X^2 + 3X^3 + X^4 + 4X^5 + 2X^6 + 3X^7.$$

Remarque

$Q_{w,k}$ est de degré

$$val(1\underbrace{1\cdots 1}_{k \text{ fois}}) = 2^{k+1} - 1$$

$\rightsquigarrow$ croît exponentiellement avec k.

Remarque

Deux mots u, v sont k-binomialement équivalents SSI

$$Q_{u,k}(X) = Q_{v,k}(X).$$

A première vue, il est nécessaire de calculer tous les coefficients ! (au moins la moitié d'entre eux)

Soit p un grand nombre premier (bien choisi),
$Q_{u,k}(X)$ et $Q_{v,k}(X)$ sont vus comme des polynômes de $\mathbb{F}_p[X]$.

Si $u \not\equiv_k v$, alors $Q_{u,k}(X) - Q_{v,k}(X)$ est un polynôme non nul de degré d ayant au plus d zéros. Si on choisit $\alpha \in \mathbb{F}_p$ aléatoirement,

$$\mathbb{P}((Q_{u,k} - Q_{v,k})(\alpha) = 0) \leq d/p.$$

Si $u \equiv_k v$, alors $Q_{u,k}(X) - Q_{v,k}(X) = 0$.
Pour tout $\alpha \in \mathbb{F}_p$, $Q_{u,k} - Q_{v,k}(\alpha) = 0$

Remarque

Deux mots u, v sont k-binomialement équivalents SSI

$$Q_{u,k}(X) = Q_{v,k}(X).$$

A première vue, il est nécessaire de calculer tous les coefficients ! (au moins la moitié d'entre eux)

Soit p un grand nombre premier (bien choisi),
$Q_{u,k}(X)$ et $Q_{v,k}(X)$ sont vus comme des polynômes de $\mathbb{F}_p[X]$.

Si $u \not\equiv_k v$, alors $Q_{u,k}(X) - Q_{v,k}(X)$ est un polynôme non nul de degré d ayant au plus d zéros. Si on choisit $\alpha \in \mathbb{F}_p$ aléatoirement,

$$\mathbb{P}((Q_{u,k} - Q_{v,k})(\alpha) = 0) \leq d/p.$$

Si $u \equiv_k v$, alors $Q_{u,k}(X) - Q_{v,k}(X) = 0$.
Pour tout $\alpha \in \mathbb{F}_p$, $Q_{u,k} - Q_{v,k}(\alpha) = 0$

Un algorithme probabiliste

En *supposant* que d/p est '*petit*', on choisit aléatoirement $\alpha \in \mathbb{F}_p[X]$.
En *supposant* que l'on calcule '*facilement*' $Q_{u,k}(\alpha)$ et $Q_{v,k}(\alpha)$.

- Si $Q_{u,k}(\alpha) \neq Q_{v,k}(\alpha)$, alors $u \not\equiv_k v$.
  $\rightsquigarrow$ L'algorithme renvoie $u \not\equiv_k v$.
- Si $Q_{u,k}(\alpha) = Q_{v,k}(\alpha)$, alors *presque sûrement* $u \equiv_k v$.
  $\rightsquigarrow$ L'algorithme renvoie $u \equiv_k v$.

On a $Q_{u,k}(\alpha) = Q_{v,k}(\alpha)$ et $u \not\equiv_k v$, uniquement si on a tiré un zéro du polynôme non nul $(Q_{u,k} - Q_{v,k})(X)$.

$\rightsquigarrow$ On obtient une conclusion erronée $u \equiv_k v$ alors que $u \not\equiv_k v$, avec une probabilité d'au plus d/p.

Choix de p ?

Les coefficients in $Q_{w,k} \in \mathbb{F}_p[X]$ sont inférieurs à n^k, en effet

$$\binom{a^n}{a^k} = \binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!} < n^k$$

Choisir un nombre premier
$p \in [n^k, 2n^k]$

Ce n'est pas un problème pour obtenir un algorithme polynomial :

- AKS est polynomial en $\log(n)$
- test probabiliste de Miller–Rabin,
  déterministe si hypothèse de Riemann OK.

$Q_{w,k}(X)$ est de degré $2^{k+1}-1$ et $p \geq n^k$

$$\text{probabilité d'erreur : } \frac{d}{p} \leq \frac{2^{k+1}-1}{n^k} \overset{n\to+\infty}{\longrightarrow} 0$$

Pour des mots suffisamment longs u, v, on croira volontiers l'algorithme quand il renvoie '$u \equiv_k v$'.

Résultat principal

Soit w un mot de longueur n. Soit $\alpha \in \mathbb{F}_p$.
La valeur $Q_{w,k}(\alpha)$ peut être calculée en $\mathcal{O}(k^2 n)$ opérations.

$$Q_{w,k}(X) = \sum_{|v| \leq k} \binom{w}{v} X^{val_2(1v)} = \sum_{\ell=1}^{k} X^{2^\ell} \left(\underbrace{\sum_{|v|=\ell} \binom{w}{v} X^{val_2(v)}}_{=: R_{w,\ell}(X)} \right)$$

$\rightsquigarrow$ Il faut déterminer $R_{w,\ell}(\alpha)$ pour tout $\ell \in \{1, \ldots, k\}$

$w = w_1 \cdots w_n \quad w[i,n] = w_i \cdots w_n$

On utilise la programmation dynamique pour la table $k \times n$ et les valeurs

$$R_{w[i,n],t}(\alpha), \quad i \in \{1,\ldots,n\}, t \in \{1,\ldots,k\}$$

$R_{w,k}$	$R_{w[2,n],k}$	$R_{w[3,n],k}$	$\cdots$	$\cdots$	$R_{w[n,n],k}$	0
$R_{w,k-1}$	$R_{w[2,n],k-1}$	$R_{w[3,n],k-1}$	$\cdots$	$\cdots$	$R_{w[n,n],k-1}$	0
$R_{w,k-2}$	$R_{w[2,n],k-2}$	$R_{w[3,n],k-2}$	$\cdots$	$\cdots$	$R_{w[n,n],k-2}$	0
$\vdots$	$\vdots$	$\vdots$			$\vdots$	$\vdots$
			$R_{w[i,n],t}$	$R_{w[i+1,n],t}$ $R_{w[i+1,n],t-1}$		
$\vdots$	$\vdots$	$\vdots$			$\vdots$	$\vdots$
$R_{w,1}$	$R_{w[2,n],1}$	$R_{w[3,n],1}$	$\cdots$	$\cdots$	$R_{w[n,n],1}$	0
1	1	1	$\cdots$	$\cdots$	1	1

$R_{\underbrace{w[n+1,n]}_{=\varepsilon},t} = 0$ si $t > 0$; $R_{w[i,n],0} = 1$ pour tout $1 \leq i \leq n+1$

$R_{w[i,n],t}$, $i \leq n$, $t \geq 1$,
dépend uniquement de $R_{w[i+1,n],t}$ and $R_{w[i+1,n],t-1}$

Soient $i \leq n$, $t \geq 1$, on a

$$R_{w[i,n],t}(X) = R_{w[i+1,n],t}(X) + R_{w[i+1,n],t-1}(X), \text{ si } w_i = 0$$

$$R_{w[i,n],t}(X) = R_{w[i+1,n],t}(X) + X^{2^t} R_{w[i+1,n],t-1}(X), \text{ si } w_i = 1$$

On se souvient que

$$R_{w[i,n],t}(X) = \sum_{|v|=t} \binom{w_i \cdots w_n}{v} X^{val_2(v)}$$

$$\underbrace{R_{w[i,n],t}(X)}_{\downarrow} = R_{w[i+1,n],t}(X) + R_{w[i+1,n],t-1}(X), \text{ si } w_i = 0$$

$\sum_{|v|=t} \binom{0w_{i+1}\cdots w_n}{v} X^{val_2(v)}$ v débute avec 0 ou 1

$= \sum_{|u|=t-1} \binom{0w_{i+1}\cdots w_n}{0u} X^{val_2(0u)} + \sum_{|u|=t-1} \binom{0w_{i+1}\cdots w_n}{1u} X^{val_2(1u)}$

$= \sum_{|u|=t-1} \binom{w_{i+1}\cdots w_n}{u} X^{val_2(u)} + \sum_{|u|=t-1} \binom{w_{i+1}\cdots w_n}{0u} X^{val_2(0u)}$
$+ \sum_{|u|=t-1} \binom{w_{i+1}\cdots w_n}{1u} X^{val_2(1u)}$

$$= \overbrace{\sum_{|u|=t-1} \binom{w_{i+1}\cdots w_n}{u} X^{val_2(u)}}^{R_{w[i+1,n],t-1}(X)}$$
$$+ \underbrace{\sum_{|u|=t-1} \binom{w_{i+1}\cdots w_n}{0u} X^{val_2(0u)} + \sum_{|u|=t-1} \binom{w_{i+1}\cdots w_n}{1u} X^{val_2(1u)}}_{R_{w[i+1,n],t}(X)}$$

En résumé,

- Calculer un élément $R_{w[i,n],t}(\alpha)$ de la table
  est *une addition* dans $\mathbb{F}_p$ et $p \sim n^k$.
  Cela nécessite $\mathcal{O}(\log p) = \mathcal{O}(k \log n)$ — arithmétique des corps finis
- On doit calculer $k \times n$ éléments de ce type
  $\rightsquigarrow \mathcal{O}(k^2 n \log n)$
- Enfin, on calcule

$$Q_{w,k}(\alpha) = \sum_{\ell=1}^{k} \alpha^{2^\ell} R_{w,\ell}(\alpha)$$

  k produits, chacun nécessitant $\mathcal{O}(\log^2 p) = \mathcal{O}(k^2 \log^2 n)$
  $\rightsquigarrow \mathcal{O}(k^3 \log^2 n)$

References

- P. Karandikar, M. Kufleitner, Ph. Schnoebelen. On the index of Simon's congruence for piecewise testability, *Information Processing Letters* **15** (2015), 515–519.
- J. Maňuch, Characterization of a word by its subwords, in : G. Rozenberg, W. Thomas (Eds.), Developments in Language Theory, World Scientific Publ. Co., Singapore, 2000, pp. 210–219.
- A. Mateescu, A. Salomaa, K. Salomaa, Yu Sheng, A Sharpening of the Parikh Mapping, *RAIRO-Theoretical Informatics and Applications* **35** (2001), 551–564.
- M. Rigo, P. Salimov, Another generalization of abelian equivalence : Binomial complexity of infinite words, *Theoret. Comput. Sci.* **601** (2015), 47—57.

References

- M. Rigo, Relations on words, Arxiv/1602.03364
- J. Sakarovitch, I. Simon, Subwords, in : M. Lothaire (Ed.), Combinatorics on Words, Addison-Wesley, Reading, MA, 1983, pp. 105–142.
- A. Salomaa, Counting (scattered) subwords, *EATCS Bull.* **81** (2003) 165–179.
- A. Salomaa, Connections between subwords and certain matrix mappings, *Theoret. Comput. Sci.* **340** (2005) 188–203.
- A. Salomaa, Criteria for the matrix equivalence of words, *Theoret. Comput. Sci.* **411** (2010) 1818–1827.
- T.-F. Şerbănuţă, Extending Parikh matrices, *Theoret. Comput. Sci.* **310** (2004), 23–246.

4th CANT School & Conference — CIRM, Marseille

Combinatorics, Automata and Number Theory

November 28th – December 2nd, 2016

http ://www.cant.ulg.ac.be/cant2016/

http ://scientific-events.weebly.com/1502.html

21th international conference DLT

Developments in Language Theory

7 – 11 August 2017

http ://www.cant.ulg.ac.be/dlt/