

MOTIFS EN COMBINATOIRE

Sergey Kirgizov

Université de Bourgogne

Assemblée générale du LIB, 4 Juillet '24

MOTS

Au commencement était un mot...

MOTS

Au commencement était un mot...

MAMA

MOTS

Au commencement était un mot...

MAMA

EUPHONIE

CAERDYDD

ÉGRÉGORE

VERLAN

MOTS

Au commencement était un mot...

MAMA	ABRACADABRA
EUPHONIE	10011101011
CAERDYDD	((().[[])
ÉGRÉGORE	12364545481
VERLAN	UUUDDFFLRD

...

MOTS

Au commencement était un mot...

MAMA

ABRACADABRA

EUPHONIE

10011101011

CAERDYDD

((().[.]

...

ÉGRÉGORE

12364545481

VERLAN

UUUDDFFLRD

Combient des mots y a-t-il
dans le monde?

MOTS

Au commencement était un mot...

MAMA	ABRACADABRA	
EUPHONIE	10011101011	
CAERDYDD	((().[].))	...
ÉGRÉGORE	12364545481	
VERLAN	UUUDDFFLRD	

Combient des mots y a-t-il
dans le monde ?

Combinatoire naïve

Alphabet de k lettres.

Nous avons $\underbrace{k \times k \times \cdots \times k}_{n \text{ fois}} = k^n$ mots différent de longueur n .

MOTS

Au commencement était un mot...

MAMA	ABRACADABRA	
EUPHONIE	10011101011	
CAERDYDD	((().[[])	...
ÉGRÉGORE	12364545481	
VERLAN	UUUDDFFLRD	

Combient des mots y a-t-il
dans le monde ?

Combinatoire naïve

Alphabet de k lettres.

Nous avons $\underbrace{k \times k \times \cdots \times k}_{n \text{ fois}} = k^n$ mots différent de longueur n .

Linguistique naïve

Compter les mots dans un dictionnaire

MOTS

Au commencement était un mot...

MAMA	ABRACADABRA	
EUPHONIE	10011101011	
CAERDYDD	((().[[])	...
ÉGRÉGORE	12364545481	
VERLAN	UUUDDFFLRD	

Combient des mots y a-t-il
dans le monde?

Linguistique un peu plus élaborée

Tenir compte du développement de la langue, de sa dynamique, de son histoire

MOTS

Au commencement était un mot...

MAMA	ABRACADABRA	
EUPHONIE	10011101011	
CAERDYDD	((()))..[[]]	...
ÉGRÉGORE	12364545481	
VERLAN	UUUDDFFLRD	

Combient des mots y a-t-il
dans le monde ?

Linguistique un peu plus élaborée

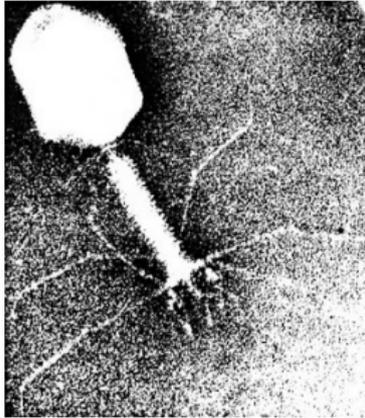
Tenir compte du développement de la langue, de ça dynamique, de son histoire

Combinatoire un peu plus élaborée

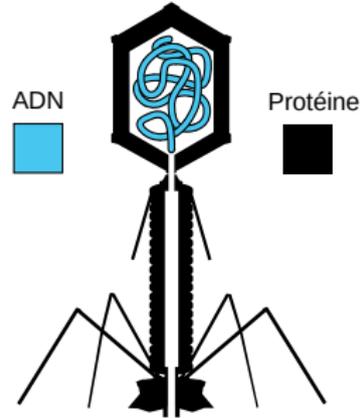
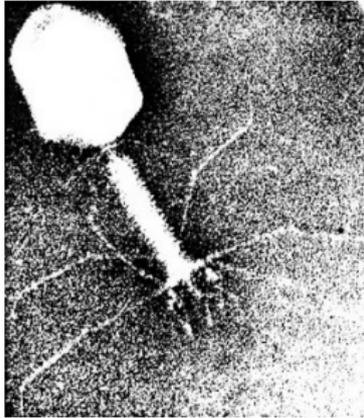
Ne considérer que les mots ayant une certaine propriété

POUR EFFECTUER UN CALCUL
CORRECT, IL FAUT RÉPONDRE
À LA QUESTION - QU'EST-CE
EXACTEMENT QU'UN "MOT" ?

ADNs VUS COMME LES MOTS

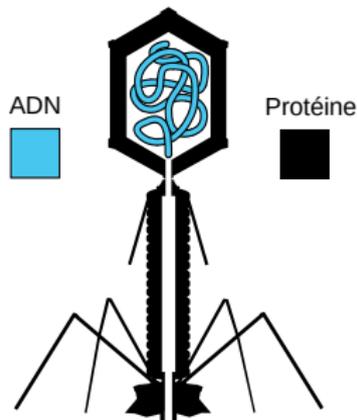
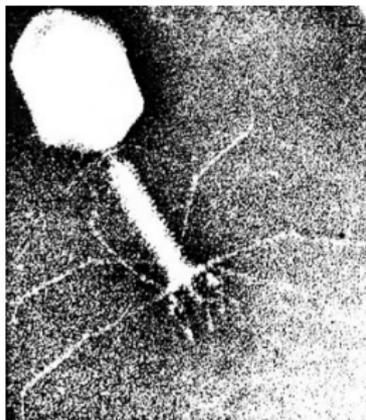


ADNs VUS COMME LES MOTS



T4 phage virion

ADNs VUS COMME LES MOTS

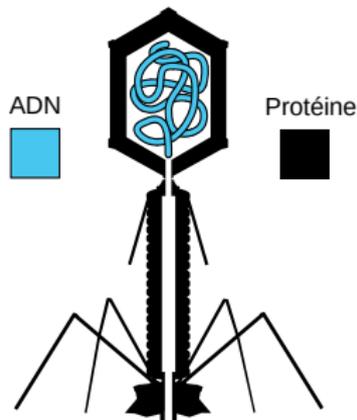
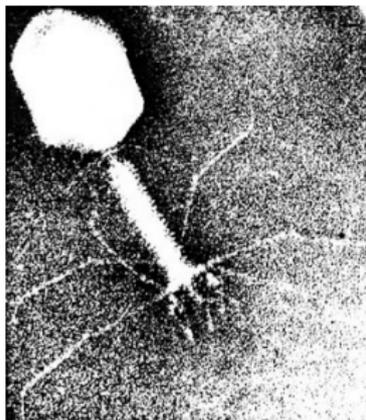


T4 phage virion

Alphabet : { G, C, A, T }

ADN : AATTTTCCTTATTAGGCCGCAAGGGCCTTC...

ADNs VUS COMME LES MOTS



T4 phage virion

Alphabet : { G, C, A, T }

ADN : AATTTTCCTTATTAGGCCGCAAGGGCCTTC...

Toutes les séquences sont-elles possibles ?

Fig. à droite par Adenosine de Wikipedia (CC BY-SA 2.5)

NULLOMERS DANS LES ADN

Séquences du génome :

AATCGTTT

ATGCAAGG

TGGGCCCA

GGTCCCCC

CCCCCCCC

NULLOMERS DANS LES ADN

Séquences du génome :

AATCGTTT

ATGCAAGG

TGGGCCCA

GGTCCCCC

CCCCCCCC

Motifs consecutifs :

A présent

CA présent

AC absent

ACCC absent

GCC présent

NULLOMERS DANS LES ADN

Séquences du génome :

AATCGTTT

ATGCAAGG

TGGGCCCA

GGTCCCCC

CCCCCCCC

Motifs consecutifs :

A présent

CA présent

AC absent

ACCC absent

GCC présent

Les nullomers sont des sous-séquence consecutive les plus courtes qui n'apparaissent pas dans le génome d'une espèce.

[Greg Hampikian and Tim Andersen.
Absent sequences : nullomers and primes, 2007]

**Aucune occurrence
dans le génome
humain**

**CGCTCGACGTA,
GTCCGAGCGTA,
CGACGAACGGT,
CCGATACGTCG**

MOTS ÉVITANT LES MOTIFS EN COMBINATOIRE

010
000 Mots binaires évitant 11
01 001
0 00 101 ...
1 10 100

2 3 5 8

MOTS ÉVITANT LES MOTIFS EN COMBINATOIRE

010
000 Mots binaires évitant 11
01 001
0 00 101 ...
1 10 100

2 3 5 8

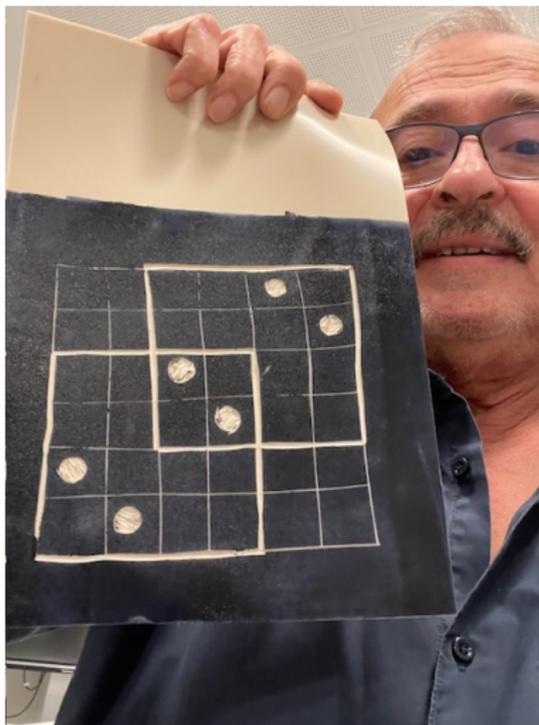
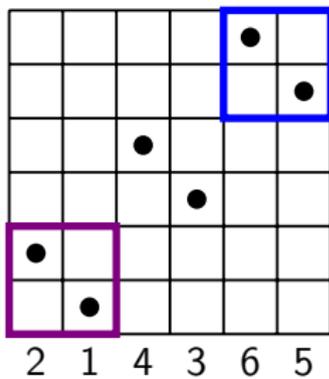
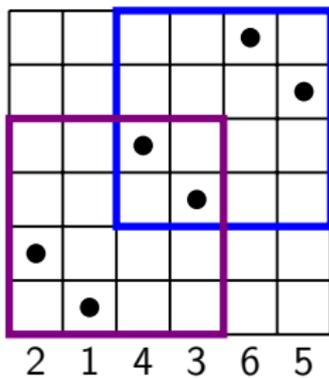
Mots de Knuth-Fibonacci sont des mots binaires ne contenant aucune occurrence du facteur 1^k . Ils sont énumérés par des nombres de Fibonacci généralisés.

- Évitant 11 : Fibonacci, $a_n = a_{n-1} + a_{n-2}$
- Évitant 111 : Tribonacci, $a_n = a_{n-1} + a_{n-2} + a_{n-3}$

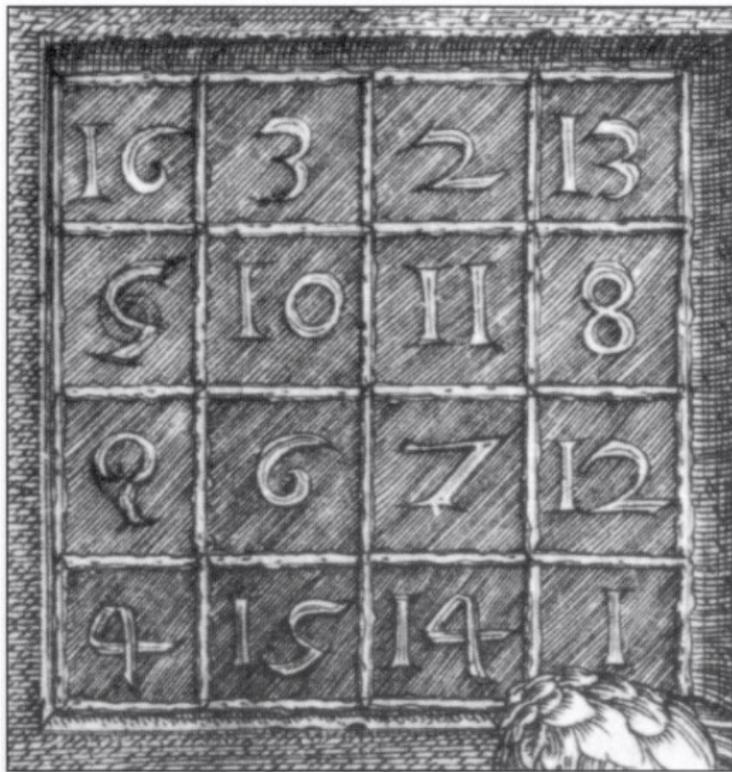
 The Art of Computer Programming, Volume 3
2nd ed., page 286, 1998, Donald Knuth

IL EXISTE ÉGALEMENT
D'AUTRES OBJETS DISCRETS -
ARBRES, GRAPHES, CARTES,
TRIANGULATIONS, CHEMINS,
TREILLIS, ...

... PERMUTATIONS ...

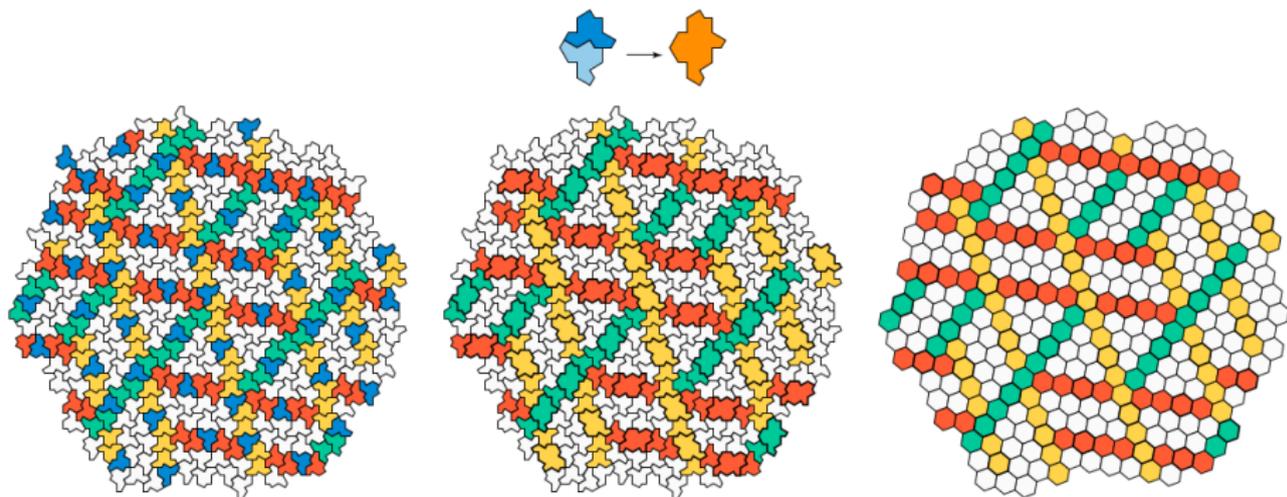


... CARRÉS MAGIQUES ...



[Melencolia I. Albrecht Dürer, 1514]

... PAVAGES ...



[An aperiodic monotile.
David Smith, Joseph Samuel Myers,
Craig S. Kaplan, and Chaim Goodman-Strauss, 2023]

CONFÉRENCE À DIJON

Permutation Patterns 2023



Nous avons organisé une conférence autour de motifs dans les permutations et autres objets combinatoires

PROJET PICS - PATTERNS IN COMBINATORICS

Programme “Jeunes Chercheuses et Jeunes Chercheurs”
de l’agence nationale de la recherche.

À partir de 2022, pour une durée de quatre ans.

Participants et co-auteurs :

Armen Petrossian

Célia Biane

Daniel Pinson

Diego Villamizar

Greg Hampikian

Jean-Luc Baril

José Luis Ramírez

Justin Masson

Khaydar Nurligareev

Matteo Cervetti

Mehdi Naima

Nathanaël Hassler

Olivier Togni

Richard Genestier

Rémi Maréchal

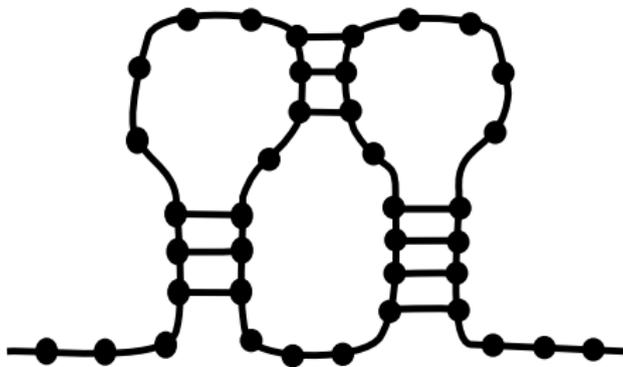
Sergey Kirgizov

Vincent Vajnovszki

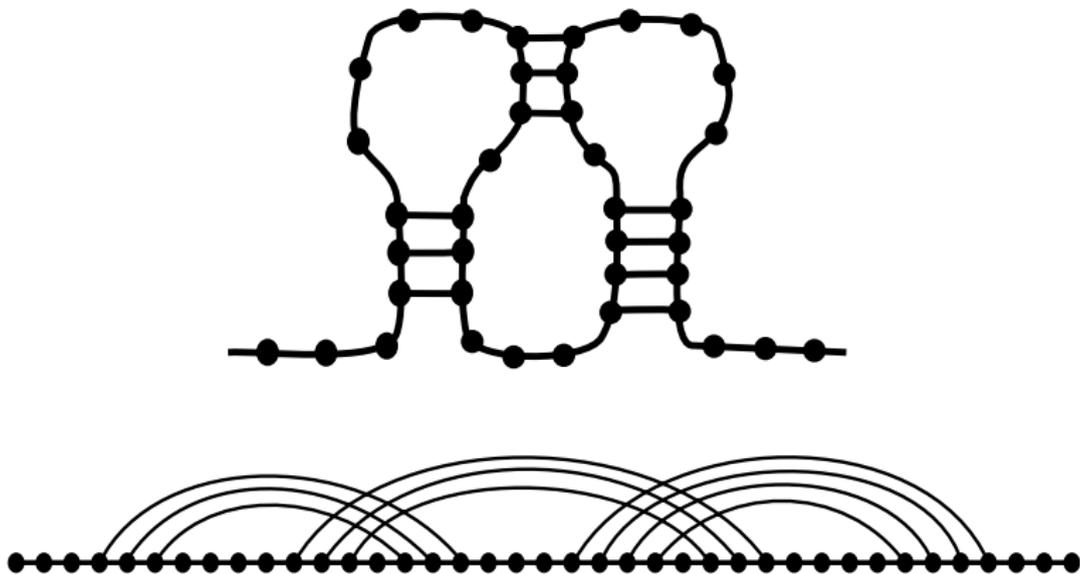
Yann Ponty

...

LA STRUCTURE SECONDAIRE D'ARN ET SES COUPLAGES



LA STRUCTURE SECONDAIRE D'ARN ET SES COUPLAGES



MOTIFS COLLEX DANS LES COUPLAGES PARFAITS

collex = collés aux extrémités

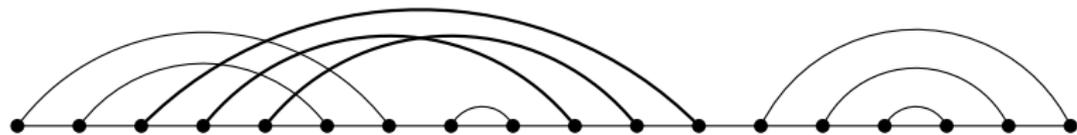
MOTIFS COLLEX DANS LES COUPLAGES PARFAITS

collex = collés aux extrémités



MOTIFS COLLEX DANS LES COUPLAGES PARFAITS

collex = collés aux extrémités



Ce couplage parfait contient

1 occurrence de motif

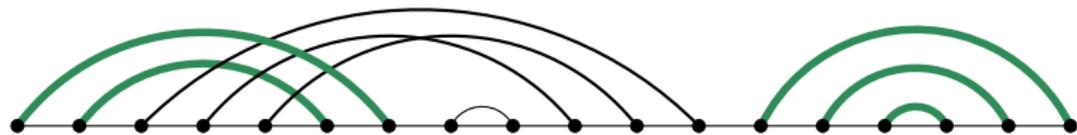


3 occurrences de motif



MOTIFS COLLEX DANS LES COUPLAGES PARFAITS

collex = collés aux extrémités



Ce couplage parfait contient

1 occurrence de motif



3 occurrences de motif



DISTRIBUTION DE DANS LES COUPLAGES PARFAITS

$n \backslash k$	1	2	3	4	5	6	7	8	9	OEIS
0	1	2	10	68	604	6584	85048	1269680	21505552	A165968
1	0	1	4	30	272	3020	39504	595336	10157440	A179540
2	0	0	1	6	60	680	9060	138264	2381344	
3	0	0	0	1	8	100	1360	21140	368704	
4	0	0	0	0	1	10	150	2380	42280	
5	0	0	0	0	0	1	12	210	3808	
6	0	0	0	0	0	0	1	14	280	
7	0	0	0	0	0	0	0	1	16	
8	0	0	0	0	0	0	0	0	1	

Combien de couplages différents avec n liens ont exactement k occurrences du motif  ?

DISTRIBUTION DE DANS LES COUPLAGES PARFAITS

$n \backslash k$	1	2	3	4	5	6	7	8	9	OEIS
0	1	2	10	68	604	6584	85048	1269680	21505552	A165968
1	0	1	4	30	272	3020	39504	595336	10157440	A179540
2	0	0	1	6	60	680	9060	138264	2381344	
3	0	0	0	1	8	100	1360	21140	368704	
4	0	0	0	0	1	10	150	2380	42280	
5	0	0	0	0	0	1	12	210	3808	
6	0	0	0	0	0	0	1	14	280	
7	0	0	0	0	0	0	0	1	16	
8	0	0	0	0	0	0	0	0	1	

Combien de couplages différents avec n liens ont exactement k occurrences du motif ? **La réponse :** $a_{n,k}$

DISTRIBUTION DE DANS LES COUPLAGES PARFAITS

$n \backslash k$	1	2	3	4	5	6	7	8	9	OEIS
0	1	2	10	68	604	6584	85048	1269680	21505552	A165968
1	0	1	4	30	272	3020	39504	595336	10157440	A179540
2	0	0	1	6	60	680	9060	138264	2381344	
3	0	0	0	1	8	100	1360	21140	368704	
4	0	0	0	0	1	10	150	2380	42280	
5	0	0	0	0	0	1	12	210	3808	
6	0	0	0	0	0	0	1	14	280	
7	0	0	0	0	0	0	0	1	16	
8	0	0	0	0	0	0	0	0	1	

Combien de couplages différents avec n liens ont exactement k occurrences du motif ? **La réponse :** $a_{n,k}$

Soit $b_{n,k} = a_{n+1,k}$, la fonction génératrice exponentielle est

$$\sum_{n=0}^{\infty} \sum_{k=0}^n b_{n,k} \frac{z^n}{n!} u^k = \frac{e^{z(u-1)}}{\sqrt{(1-2z)^3}}$$

DISTRIBUTION DE DANS LES COUPLAGES PARFAITS

$n \backslash k$	1	2	3	4	5	6	7	8	9	OEIS
0	1	2	10	68	604	6584	85048	1269680	21505552	A165968
1	0	1	4	30	272	3020	39504	595336	10157440	A179540
2	0	0	1	6	60	680	9060	138264	2381344	
3	0	0	0	1	8	100	1360	21140	368704	
4	0	0	0	0	1	10	150	2380	42280	
5	0	0	0	0	0	1	12	210	3808	
6	0	0	0	0	0	0	1	14	280	
7	0	0	0	0	0	0	0	1	16	
8	0	0	0	0	0	0	0	0	1	

Combien de couplages différents avec n liens ont exactement k occurrences du motif ? **La réponse :** $a_{n,k}$

Équivalences asymptotiques

$$a_{n,k} \sim \frac{1}{2^k k!} \left(\frac{2}{e}\right)^{n+1/2} n^n \quad \frac{a_{n,k}}{a_{n,k+1}} \sim 2(k+1).$$

DISTRIBUTION DE  ?

COLLEX TWIST

Collex twist gauche

Toutes les séries de points de départ consécutifs sont inversées.

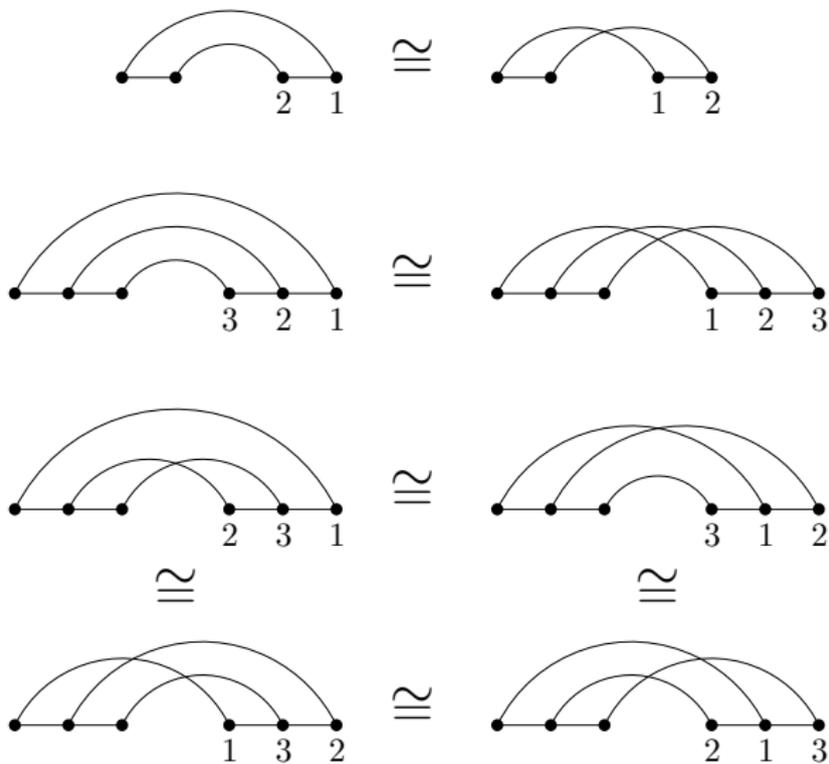
Collex twist droit

Toutes les séries de points d'arrivée consécutives sont inversées.



GRÂCE AU COLLEX TWIST
A LA MÊME DISTRIBUTION QUE !

ÉQUIVALENCE DE MOTIFS SELON COLLEX TWISTS



Équivalence = même distribution

EXPLORATION DE MOTIFS DANS LES ARN

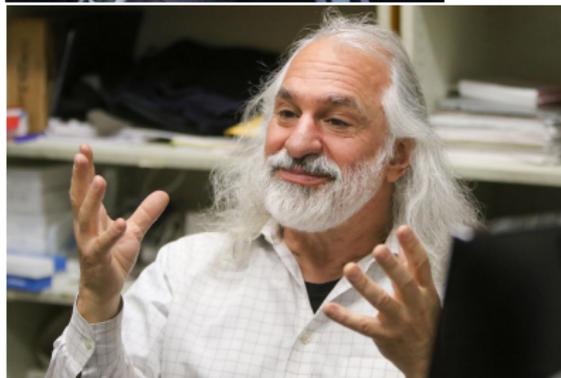
Site web interactif conçu par Daniel Pinson.



APERÇU DE NOS RÉSULTATS SUR LES ARN

- Étude combinatoire des copulages parfaits
 - ▶ modèle théorique des structures secondaires d'ARN
 - ▶ énumération directe
 - ▶ bijections
 - ▶ **pour tout motif collex μ** nous savons calculer la distribution à l'aide d'une méthode modifiée de Goulden-Jackson (méthode des clusters)
 - ▶ équivalences asymptotiques à la Borinsky
- Exploration de données réelles, venant de la base de données des structures d'ARN
 - ▶ comparaison avec les données théoriques
 - ▶ application web
- Étude des permutations auto-chevauchant et autres objets...

CÉLIA, KHAYDAR, DANIEL ET GREG



- 📄 Endhered patterns in matchings and RNA.
Célia Biane, Greg Hampikian, Sk, Khaydar Nurligareev
<https://arxiv.org/abs/2404.18802>
- 📄 Asymptotics of self-overlapping permutations
Sk and Khaydar Nurligareev
<https://arxiv.org/abs/2311.11677>
- 📄 Clusters of endhered patterns in permutations and matchings. In preparation.

Autres papiers sur arXiv et mon site web

PERSPECTIVES ?

ACTIONS EN COURS ET PERSPECTIVES

- Recherche sur les structures discrètes et leurs motifs
- Interaction entre la combinatoire et d'autres domaines : linguistique, biologie, sociologie...
- La science et l'art
- Déposer un projet ERC (Conseil Européen de la Recherche) ?
- Nouveau cours de master lié à l'étude des structures biologiques ?



*Les ARN et leurs motifs,
L'info et maths...
Je les admire !*